В·П·КУЗНЕЦОВ

ИНТЕРВАЛЬНЫЕ СТАТИСТИЧЕСКИЕ МОДЕЛИ

«РАДИО И СВЯЗЬ»

В·П·КУЗНЕЦОВ

ИНТЕРВАЛЬНЫЕ СТАТИСТИЧЕСКИЕ МОДЕЛИ

Москва
«Радио и связь»
1991

УДК 621.391

Кузнецов В. П. Интервальные статистические модели. — М.: Радио и связь, 1991. — 352 с.: ил. **ISBN 5-256-00726-2.**

На базе новой аксиоматики развивается аппарат размытых математических моделей случайных явлений. Эти модели охватывают множественные, интервальные, нечеткие, и вообще любые неполные и отрывочные статистические описания характеристик явления, подходя к распределениям вероятностей как пределу изобилия данных. Сфера действия моделей простирается от неустойчивых, уникальных явлений до статистически устойчивых к повторам. В этих широких пределах освещаются и интерпретируются понятия интервальной вероятности и среднего, анализируются причинные связи, случайные преобразования, отношения зависимости и независимости, исследуются предельные законы, описываются случайные процессы и прочее другое.

Применительно к новым моделям вводятся критерии и разрабатываются универсальные методы синтеза оптимальных решающих правил (оценок, различения гипотез). Реализующие их устройства просты по структуре и способны эффективно работать в изменяющихся окружающих условиях, основанием для чего служит выбор надежных моделей. Доверие к моделям завоевывается вовлечением в них небольшого числа исходных вероятностей и средних, представленных в интервальном виде, отражающем нестабильность реальных явлений и дефицит исходных данных о нем. Рассматривается совместный синтез надежных моделей и решающих правил.

Для научных работников в области связи и управления; может быть полезна всем, кто интересуется математическими методами описания случайных явлений и задачами принятия решений при неопределенности.

Табл. 1. Ил. 37. Библиогр. 25 назв.

Рецензент: проф., докт. техн. наук Ф. П. ТАРАСЕНКО

Редакция литературы по радиотехнике и электросвязи

К $\frac{2303020000\text{-}040}{046(01)\text{-}91}$ 96-90

ISBN 5-256-00726-2

© Кузнецов В. П., 1991

Памяти матери
Кузнецовой Екатерины Ивановны
посвящается

ВВЕДЕНИЕ

Теория вероятностей есть не что иное, как математический язык описания случайных явлений. Привычка, навык к этому языку мешают задуматься над существованием других, быть может более удобных форм и описаний, не входящих в словарь общеупотребительного языка, но делающих простыми ситуации, столь трудные в современном «произношении». Разработка нового математического языка шире общепризнанного и его использование составляет суть предлагаемой книги.

Символьный язык — средство описания и способ общения, но в то же время это инструмент, с помощью которого можно что-то исследовать, создавать, обрабатывать, конструировать, а для вероятностно-статистических методов — получать решающие правила, алгоритмы, оценки. Последние реализуются работающими устройствами. Критерием жизненности, приемлемости нового символьного аппарата служит его надежность, адекватность, способность делать то, чего ранее не было, обрабатывать то, что не обрабатывалось, упрощать то, что было сложным. Именно эта цель преследовалась при введении интервальных моделей и старательно претворялась при разработке методов.

Базу книги закладывают интервальные вероятностно-статистические категории, дающие универсальный способ описания как имеющихся знаний, так и их отсутствия, т. е. незнания; под эти категории подводится аксиоматика. Получается новая теория, непривычная, наверно, с первого взгляда, но охватывающая огромное разнообразие явлений как устойчивых, определяемых вероятностями, так и неустойчивых, невероятностных, с небольшим числом неполно исследованных закономерностей, наконец и вовсе с неизвестными свойствами.

Покажем, что зерно интервального подхода уже скрыто лежит в недрах современных вероятностных построений, задача — взрастить его (первая часть книги) и «собрать урожай» (вторая часть).

Пример 1. Пусть модель случайной величины с исходами на числовой прямой $\mathscr{X}$ описывается плотностью распределения вероятностей. Интегрирование по ней дает вероятности $P(A)$ отрезков $A \in \mathscr{X}$ и их объединений (сумм) вплоть до счетных, составляющих набор $\mathscr{A}$ измеримых событий. В $\mathscr{A}$ не могут войти все события, как бы не размельчалось оно до борелевских и далее лебеговских множеств [1]. Всегда останутся так называемые неизмеримые события $B \notin \mathscr{A}$, для которых вероятности уже будут интервальные $\underline{P}(B)$, $\overline{P}(B)$, оп-

ределенные как внутренняя и внешняя меры формулами: $\underline{P}(B)=\sup_{A:\,B\supset A\in\mathscr{A}} P(A)$, $\overline{P}(B)=\inf_{A:\,B\subset A\in\mathscr{A}} P(A)$. Точно так же и с математическими ожиданиями (средними статистическими) от случайных величин — функций $g(x)$, $x\in\mathscr{X}$. Они оказываются точными Mg на классе $\mathscr{LA}$ измеримых (интегрируемых) функций и станут интервальными, определенными формулами $\underline{M}f=\sup_{g:\,f\geqslant g\in\mathscr{LA}} Mg$, $\overline{M}f=\inf_{g:\,f\leqslant g\in\mathscr{LA}} Mg$ для остальных, неизмеримых f, которых, в общем, великое множество.

Пример позволяет раскрыть следующую конструкцию современных вероятностных построений. На ядре $\mathscr{A}$, представляющем набор событий пространства исходов $\mathscr{X}$, первичными заданы точные вероятности, образующие распределение вероятностей. Стремятся закладывать в $\mathscr{A}$ как можно большее число событий, чтобы при продолжении вероятностей, а оно осуществляется интегрированием по вероятностному распределению, получались точные математические ожидания (средние) Mg у всех обозримых случайных величин. Последние определяются как измеримые функции $g(x)$, $x\in\mathscr{X}$ (обычно кусочно-непрерывные с возможными скачками первого рода) на исходах явления. Но вопреки стараниями сделать точными вероятности и средние для абсолютно всех событий и функций $f(x)$ не удается (кроме явлений с конечным и счетным числом исходов), так как все равно остаются называемые неизмеримыми события и функции, их много и из-за них при продолжении возникают интервальные вероятности и интервальные средние $\underline{M}f$, $\overline{M}f$. Это первое. А второе, что интервалы $\underline{M}f$, $\overline{M}f$ определены формально на всех $\forall f$, а точные значения $\overline{M}g$ есть частный случай интервальных при $\underline{M}g=\overline{M}g$ и их часть.

Можно возразить, зачем затрагивать практически бессмысленные неизмеримые функции? Ответ тот, что сужение ядра $\mathscr{A}$, вынуждаемое физической невозможностью измерять, да и вообще знать много вероятностей, существенно расширяет класс неизмеримых функций, делая средние неточными. А если допустить, что в ядре $\mathscr{A}$ вообще событий может быть мало, да еще первичные вероятности на $\mathscr{A}$ неточные, т. е. интервальные, то всюду как вероятности, так и средние станут интервальными. Проиллюстрируем их на примере семейства вероятностных распределений.

Пример 2. Пусть задано параметрическое семейство распределений вероятностей P_θ, $\theta\in\Theta$. Тогда средние всех измеримых функций будут интервальными, определенными как нижняя и верхняя грани

$$\underline{M}g=\inf_{\theta\in\Theta} M_\theta g,\quad \overline{M}g=\sup_{\theta\in\Theta} M_\theta g.$$

Подстановка на место $g(x)$ индикаторных функций событий A (равных 1, если аргумент x принадлежит A, и 0, если не принадлежит) ведет к интервальным вероятностям $\underline{P}(A)$, $\overline{P}(A)$ как составной части средних.

В связи со вторым примером напрашивается другой резонный вопрос: а разве семейства распределений вероятностей не служат уже универсальной формой отражения любых неточных знаний о явлении? Так вот, оказывается, что интервально-статистический подход дает более удобную форму для этих целей, в которую вписываются, как частный случай, и распределения вероятностей, и их семейства со всем богатством возможностей, а также многое-многое другое.

Интервальная модель на пространстве исходов $\mathscr{X}$ определяется формально совокупностью интервалов средних $\underline{M}f$, $\overline{M}f$, $\forall f$, связанных между собой аксиомами. Далее функции $f(x)$, $x\in\mathscr{X}$ называются признаками. Любую нашу модель можно задавать по стандартной схеме рис. В.1. Ядро модели формируется набором $\mathscr{G}$ признаков $g\in\mathscr{G}$, именуемых первичными, и указанием границ средних на них. Вероятности рассматриваются как составная часть средних. С $\mathscr{G}$ средние продолжаются на $\forall f$, образуя собой модель.

Отличительными для новой теории являются четыре ключевые момента.

1. Лишение вероятностей привилегий для задания модели и уравнивание в правах с более емким понятием среднего статистического числовых признаков[1]. Это означает отказ от распределений вероятностей как необходимой части модели и от алгебры событий как обязательной атрибутики ядра. Это раскрепощает ядро $\mathscr{G}$ и саму модель.

2. Почему обязательно вероятности и средние должны быть точными? Это всегда идеальный случай. Реальный подход — считать их интервальными — мгновенно «развязывает руки». В самом деле, отрезок $[0, 1]$ в качестве вероятности некоторого события означает полное отсутствие знания этой вероятности. И здесь не надо, что самое замечательное, задумываться о существовании точной вероятности, обязанной абсолютной статистической устойчивости явления. Пусть имеет место неустойчивость, характерная для многих приложений! Если неустойчивость не полная, а частичная, то приходим к интервальным вероятностям внутри $[0, 1]$. Наконец, если концы интервалов смыкаются, то получаются точные вероятности. Все сказанное переносится на средние, только областью их значений будут уже интервалы на всей оси чисел.

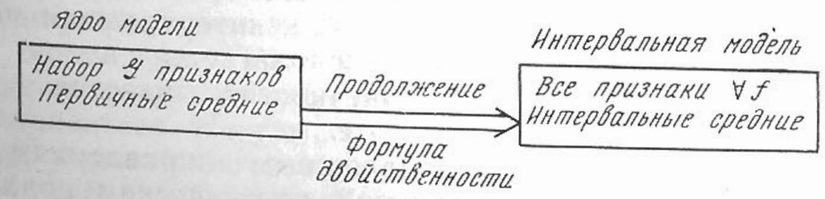


Рис. В.1. Структура интервальных моделей

[1] Попытка такого рода делалась в [2], но свелась лишь к другой расстановке акцентов в прежнем вероятностном языке.

3. Необычайная гибкость ядра $\mathcal{G}$ по форме и числу элементов и вольность задания на нем средних как в виде точных значений, так и интервальных или только одной из границ. Это «размораживает» структуру модели, позволяя в рамках единой конструкции с одной стороны универсально, а с другой (манипулируя составом $\mathcal{G}$) — крайне экономно вкладывать в модель лишь имеющиеся знания о явлении с поправками на их точность и с прицелом на простоту.

4. Средние с первичного ядра $\mathcal{G}$ продолжаются на все признаки Vf с помощью алгоритмических формул двойственности. Если изображать модель как тело, то формулы двойственности работают с его оболочкой, обходя множественное представление в виде семейств точек (распределений вероятностей, требующих громоздких формул интегрирования), «запрягая» в статистические проблемы дуальный подход.

Универсальность интервально-статистических моделей обеспечивается возможностью заключать в ядро почти любые по объему и форме знания о явлении. Взять процесс: если нет никаких данных о нем, то ядро пустое, а модель голая. Если стало известно среднее процесса, то это соответствует определенным признакам модели, формирующим грани ее тела. Если дополнительно дана средняя мощность, то это дополняет ядро еще одним признаком, дает ему еще одну грань. Если сюда добавить знания о вероятностях превышений, то состав ядра усложняется, модель усечется новыми гранями, сделается более точной; а если к тому же пополнить сведениями о корреляциях, то тем более, и т. д. Пределом знаний будет точка, т. е. вероятностная модель (число граней которой равно числу элементарных исходов). Возникает естественная иерархия от простого к сложному, как в живой природе при ее сотворении и эволюции. Сначала нет ничего, нет никаких знаний, и это соответствует самой простой модели. Любые сведения, какой бы размытый вид они ни обретали, формируют уже готовую худо-бедно работающую модель. По мере накопления сведений модель совершенствуется, усложняется. Наконец, если имеются точные данные о вероятностях событий, то получается вероятностная модель.

Связь моделей иллюстрируется рис. В.2, где стрелки указывают уровни совершенствования и усложнения моделей. Черные стрелки относятся к классическим моделям, где путь к приложениям повел от распределений вероятностей к их семействам все более сложного вида. Белые стрелки — к интервальным моделям, иерархическая система которых располагается вдоль дороги, тянущейся от нуля знаний к бесконечности ... навстречу классическому движению.

Двум гносеологически противоположным направлениям движения свойственны разные проблемы. В классическом подходе основная проблема — априорное незнание, неопределенность, вынуждающая усложнять модель переходом к семейству с целью облегчить «воз» из огромного числа вероятностей, нагружающий

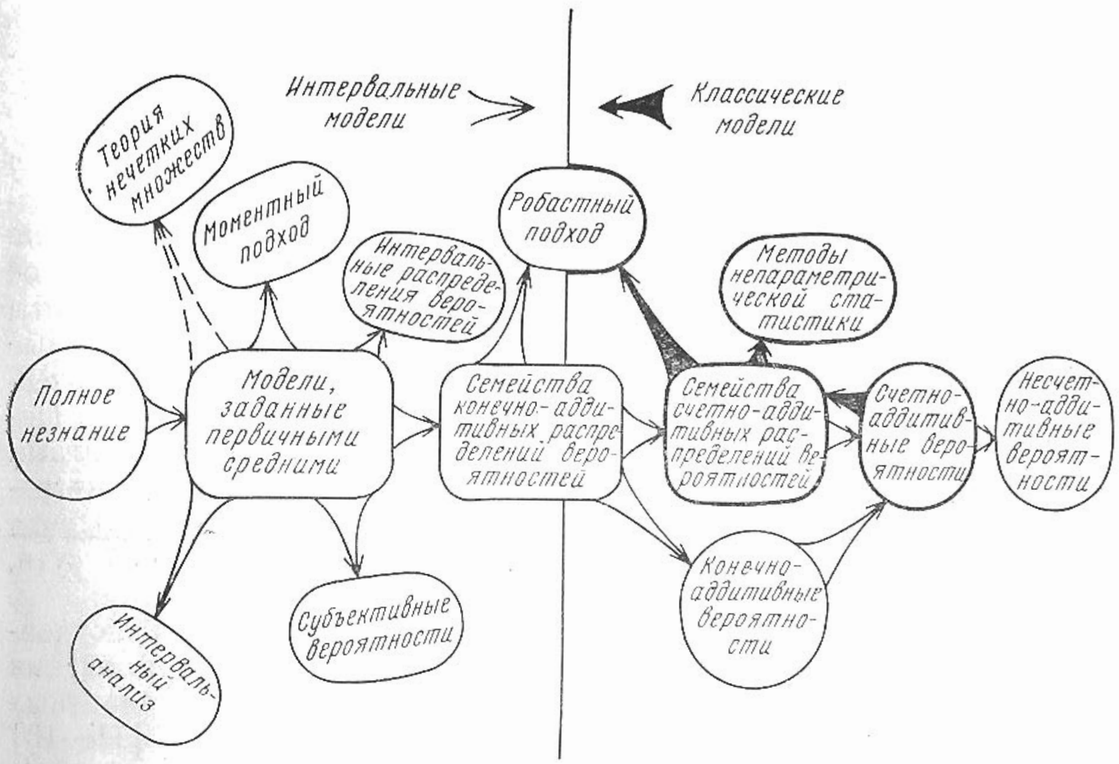


Рис. В.2. Иерархия моделей

современную модель. В нашем подходе «воз» заполняет постепенно: модели будут тем сложнее и точнее, чем больше знаний, и проблемы состоят в экономном их представлении, заключении в ядро, а если требуют упрощения, то отбрасывании всего несущественного, второстепенного.

Итак, в отличие от теории вероятностей, освещающей поточечную структуру моделей, исключающую иерархию по знаниям, мы будем мыслить модель ее внешними атрибутами, оболочкой. Знания вкладываем в грани, формируемые первичными средними, по их числу и составу модели усложняются, образуя иерархию.

Иерархия интервальных моделей переносится на получаемые из них оптимальные решающие правила. При малом числе исходных данных правила будут иметь плохие качественные характеристики, но зато устойчивы к внешним аномалиям, нестабильностям. При накоплении знаний о явлении качество повышается, правила, в общем, усложняются, становятся более избирательными к ситуации, так как настраиваются на одну или несколько из них. Здесь вроде бы качественным привилегиями обладают классические вероятностные модели, и только ими, казалось бы, надо пользоваться. Но увы, это самообман, так как необходимого для таких моделей багажа знаний почти всегда нет, а их весьма «вольный выбор» (тяготеющий к нормальному распределению в силу его относительной простоты) носит декларативный характер. Адекватные модели не должны вовлекать никаких других, кроме имеющихся проверенных знаний, всегда конечных, для надеж-

ности представленных в интервальной (размытой) доверительной форме. Это как раз та часть интервальных моделей, которая располагается на рис. В.2 слева от вертикальной линии. От них правилам будет передана по наследству надежность и простота.

Упорядоченность интервальных моделей по числу и составу данных позволяет помечтать о сводной библиотеке правил, в которой исследователь по имеющимся реальным данным о явлении смог бы отыскать модель и соответствующее ей оптимальное правило и по тому, устраивает его качество или нет, уже решал бы, нужно ли уточнять имеющиеся или собирать дополнительные сведения (экспериментом, более скурпулезным анализом физической структуры явления и т. д.), тем самым усложняя модель. Помыслы о библиотеке ограничиваются, увы, пока результатами этой монографии, а их желается много больше, для чего потребуются усилия всех, кто заинтересуется нашим подходом. Почва подготовлена, и методы теории, алгоритмические по своей сути, допускают привлечение ЭВМ.

История этой теории такова. Ее источником стала неудовлетворенность, развившаяся у автора от попыток использования для инженерно-исследовательских задач сначала инвариантных и непараметрических методов [3—10], затем робастных [11—13] ([12] содержит обзор, включающий некоторые работы автора), и вытекающее отсюда естественное желание искать что-то новое. «Забуксовал», так и не раскрывшись в полной мере, моментный подход; причину мы видим в близости его (см. рис. В.2) к предлагаемым здесь неклассическим интервально-статистическим методам. Не очень вписалась в классические методы и потому не получила должного распространения также теория субъективных вероятностей [14]. В отход вообще от вероятностей двинулась теория нечетких множеств Заде [15, 18]. Где-то стороной развивался интервальный анализ [16]. Не обрела заслуженную автономию теория обобщенных чебышевских неравенств [17]. На подготовленной этими исследованиями почве родилась идея, послужившая стержнем монографии: так выбрать фундамент, чтобы с единой платформы охватить все перечисленные нами направления с целью не только увидеть их в новом свете, но и значительно расширить возможности для приложений (распространение новой идеологии на интеграл можно найти в [19]).

Чтение настоящей книги, наверно, потребует от читателя большого терпения из-за отсутствия классических аналогов. Помощь могут оказать заключения, раскрывающие краткое содержание глав, связывающие их идейной канвой.

Книга писалась на кафедре вычислительной математики МЭИС. Рассчитана на инженеров-исследователей, аспирантов и студентов с соответствующей теоретической подготовкой.

Автор считает приятным долгом выразить глубокое признание всем тем, кто так или иначе способствовал становлению данной теории и выходу в свет книги. Ввиду новизны материала не исключаются огрехи, вину за которые автор берет на себя.

Глава 1.
ОПИСАНИЕ СЛУЧАЙНЫХ ЯВЛЕНИЙ

1.1. ИНТЕРВАЛЬНЫЕ ВЕРОЯТНОСТИ И СРЕДНИЕ

Пространство исходов. Мы живем в мире случайностей, в окружении непредвиденных действий и непредсказуемых до конца фактов. Корни случайностей разнообразны, они берут начало и от физических эффектов типа дробового шума, и от невозможности абсолютного предугадания течения процесса или поведения живого организма (в частности, индивида), и от нашего незнания (или нежелания знать) результата предстоящего (прошедшего) эксперимента и т. д. И как следствие, оказывается, что что-то может произойти, а может и нет, случиться или не случиться, быть или не быть. Эта неясность охватывается категорией случайности. Наша цель — ее описать.

Формально под *случайным явлением* понимается совокупность взаимно исключающих друг друга исходов, называемых *элементарными*, один из которых обязан произойти, но неизвестно какой.

Например, если бросается монета один раз, результатом будет герб Г (орел) или решка Р; если два раза, то элементарных исходов будет либо три (два герба ГГ, две решки РР и разной значимости), либо четыре (при разной значимости учитывается порядок следования ГР или РГ), как мы этого захотим. При бросании точки на числовую прямую $\mathscr{R}$ результатом будет число (случайная величина), а для случайного процесса, пусть шума, — реализация $x(t)$ как функция времени t.

Совокупность всех элементарных исходов формально есть некоторое абстрактное множество $\mathscr{X}$, называемое *пространством элементарных исходов*, тогда как любой из элементарных исходов — это точка x этого пространства, т. е. $x \in \mathscr{X}$. Вообще подмножество $\mathscr{X}$, обозначаемое заглавной буквой $A \subset \mathscr{X}$, называется *случайным событием*; оно произойдет, если выпадет любой входящий в него элементарный исход x.

Замечания. 1. Данное нами определение совпадает с классическим в плане введения пространства элементарных исходов $\mathscr{X}$, но не связывает случайность с вероятностями, и в этом смысле шире.

2. Может показаться, что введением пространства $\mathscr{X}$ и привязкой к нему мы как-то сузили класс охватываемых случайностей. На самом деле, пространство $\mathscr{X}$ есть элемент создаваемой нами математической модели, находится в наших руках и его можно делать сколь хотим широким, включая мыслимые и даже немыслимые исходы, если только это представляется удобным. Например (как это часто предлагают студенты), при бросании

монеты можно дополнительно включить исход, что монета встанет на ребро или что выигрыш в игре будет равен бесконечности: $\infty \in \mathscr{X}$. Можно вообще говорить о пространстве всего того, что может быть и не быть, но вряд ли это придаст описаниям экономность.

3. Элементарные исходы — не есть обязательно физическая реальность, так как появляться могут, например, нечеткие события (о которых пойдет речь чуть ниже). И тем не менее явление в нашем определении будет случайным, если хоть какое-то пространство $\mathscr{X}$ с абстрактными элементами удается с ним связать, что, как правило, возможно.

4. Детерминированные явления есть частный подкласс случайных, когда пространство элементарных исходов содержит всего один элемент.

Признаки явления. Любые измерения в своей сути есть перевод качественных показателей в количественные с представлением в числовой форме. Таким образом, в виде числа или набора чисел удобно характеризовать результаты случайного явления, отвлекаясь от конкретного содержания элементарных исходов и событий (которые в описательном смысле могут быть весьма сложными). Для этого достаточно иметь число $f(x)$ вместо каждого элементарного исхода x, т. е. поставить в соответствие $x \to f(x)$, $x \in \mathscr{X}$. Числовая функция $f(x)$ на пространстве элементарных исходов называется количественным признаком, или просто *признаком* случайного явления; обозначается сокращенно f; строго говоря, это отображение $\mathscr{X}$ в числовую прямую $\mathscr{R}$: $\mathscr{X} \xrightarrow{f} \mathscr{R}$.

Примеры признаков: f — число гербов при двух бросаниях монеты $f(\text{РР}) = 0$, $f(\text{РГ}) = f(\text{ГР}) = 1$, $f(\text{ГГ}) = 2$. Еще пример — случайный выигрыш в азартной игре, исходами которой могут быть довольно сложные комбинации условий, а f — число денег.

Разнообразных признаков сколь угодно много: столько, сколько возможно придумать различных функций $f(x)$. Признаки f вместе с $\mathscr{X}$ играют фундаментальную роль для всего дальнейшего изложения, потому что с их помощью можно описать очень многое, если не сказать все. В частности, любые события, предикаты, высказывания, связанные с явлением. В самом деле, элементарный исход x_1 тождественно описывается *дельта-признаком* $\delta_{x_1}(x)$, равным 1 при $x = x_1$ и 0 при $x \neq x_1$; а подмножество $A \subset \mathscr{X}$ (событие A) — признаком в виде индикаторной функции, обозначаемой $A(x)$ и равной 1 при $x \in A$ и 0 при $x \notin A$. Будем называть $A(x)$ *индикаторным признаком* события A; два разных значения 1 или 0 этого признака указывают, принадлежит ли это x к A или нет, и только (рис. 1.1).

Более общим является понятие нечеткого события с не столь категоричным заявлением о принадлежности к нему. Функцию $q(x)$, заключенную между 0 и 1: $0 \leqslant q(x) \leqslant 1$, $x \in \mathscr{X}$, будем называть *признаком нечеткого события*. Так как своего описания на языке случайного явления нечеткое событие может и не иметь, то

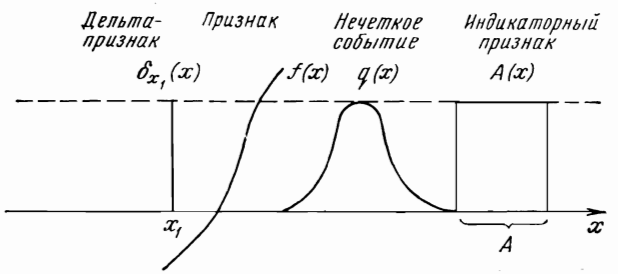


Рис. 1.1. Признаки, события

$q(x)$, заключенную между 0 и 1: $0 \leqslant q(x) \leqslant 1$, $x \in \mathscr{X}$, будем называть просто нечетким событием. Значение $q(x) = 1$ говорит, что x достоверно входит в событие q, значение $q(x) = 0$ — не входит, а $q(x) = 1/2$ равносильно балансу уверенности и сомнения: входит или не входит. Другие промежуточные между 0 и 1 значения $q(x)$ отдают предпочтение либо уверенности, что входит (при $q(x) > 1/2$), либо сомнению.

Итак, каждое событие отождествляется с эквивалентным признаком, а всем событиям соответствует подкласс признаков — это все функции со значениями от 0 до 1. Интересно и важно то, что логике высказываний, по правилам которой из одних событий логически образуются другие, соответствуют вполне определенные арифметические действия между признаками: «не q» $\Leftrightarrow 1 - q(x)$; «q_1 и q_2» $\Leftrightarrow q_1(x)\, q_2(x)$; «$q_1$ или q_2» $\Leftrightarrow \min\{1,\ q_1(x) + q_2(x)\}$.

Признаками описываются не только события, но и многие другие характеристики случайного явления, например, мощность, эффективное значение и т. д. Любое преобразование признаков: сложение, умножение и т. д. — ведет к новому признаку. Таким образом, всевозможные признаки вместе с привычными действиями между ними как действительными функциями переменной x дают универсальный аппарат описания всего того, что связано так или иначе с результатами случайного явления или может быть получено из них с помощью логических, арифметических и аналитических действий.

Если $g(x) \geqslant f(x)$, $\forall x \in \mathscr{X}$, то будем говорить, что признак g *мажорирует* признак f и писать $g \geqslant f$. Это означает, что какой бы исход не случился, признак g примет значение, не меньшее признака f. Не о любых двух выбранных признаках можно вынести такое суждение, что соответствует частичной их упорядоченности.

Средние значения признаков. Выделим отдельный признак f и рассмотрим, какие сведения могут иметься о нем априори (полагая, что явление еще не произошло или результат его не известен). Первое — это область возможных значений, определяемая видом $f(x)$. Областью значений может быть дискретный набор чисел, например для индикаторных признаков $A(x)$ это 0 или 1. Это может быть отрезок числовой прямой $[\min_x f,\ \max_x f]$ от минимального $\min_x f(x)$ до максимального $\max_x f(x)$ значения f. Для

общности, учитывая, что минимум и максимум не всегда достигаются, заменяем их на инфимум и супремум: $[\inf f, \sup f]$.

Второе — это данные о среднестатистических свойствах признака f. А именно, ожидаемое в среднем его значение Mf (здесь M от слова MEAN — средний), называемое *точным средним признака* f.

Как оно получается? Самое простое — на основании симметрии эксперимента. Например, игроки, скажем, в «орел — решку» хорошо понимают, что при равных ставках шансы на выигрыш и проигрыш одинаковы, т. е. в среднем они будут иметь ноль. В далекой древности бросали астрагалы — кости конечностей животных — и для обеспечения «равных шансов» каждый раз партнеры менялись местами.

Точные сведения могут находиться изучением внутренних механизмов явления, его природы, как это делается в статистической физике. Возможна оценка среднего по результатам предварительного опроса, проведением наблюдений, искусственно созданного обучающего эксперимента, испытаний. Если испытания независимы и проводятся в одинаковых условиях, то среднее будет пределом среднего арифметического наблюдаемых значений признака при неограниченном увеличении числа испытаний.

Наконец, просто по опыту, под которым понимается совокупность прямых и косвенных сведений о явлении, можно примерно знать, что ожидается в среднем. В самом деле, каждый из нас догадывается, сколько в среднем он потратит времени на дорогу к месту работы или в другое место, каков средний ожидаемый доход от намечаемого дела или расход от туристической поездки и т. д.

Частным случаем такого точного среднего, когда $f(x)=A(x)$ — индикаторный признак события A, является вероятность, обозначаемая $P(A)=MA(x)$ (P от слова PROBABILITY). *Вероятность* — это среднее ожидаемое число появлений A при независимых повторных испытаниях, деленное на число испытаний.

Ключевым во всех предыдущих рассуждениях является то, что в строгом смысле точные средние (и вероятности) — это параметры статистически устойчивого явления и достигаются они усреднением при неограниченном повторении того же самого явления в независимых и устойчивых условиях. Так как организовать устойчивое повторение подчас затруднительно, а неограниченное число раз просто невозможно, то часто подразумевается мыслимый или умозрительный повтор. Но чтобы «проиграть» явление сколько-то раз, проделав это в уме или с помощью ЭВМ, нужно более или менее точно знать физическую модель явления, его природу. Так, и в случае симметричной монеты, ее не обязательно подбрасывать, поскольку итак ясно, что в среднем число орлов должно равняться числу решек. Это классический пример для определения точного среднего.

Но практика не всегда вторит теории, а действительное — желаемому. Реальные явления часто таковы, что их внутренние механизмы до конца не поддаются исследованиям, опыты уникальны, их повторы неустойчивы, а предварительные наблюдения ограничены. В результате точное среднее остается как идеальное понятие, достигаемое в пределе, применение которого сопровождается многими «вот если бы» или «пусть...».

Интервальные средние и вероятности. Генеральная ваша мысль такова, что не только неустойчивость явлений, но и любая неабсолютность статистических знаний (недостаточность, неточность, ограниченность), свойственная почти всем реальным задачам, естественно вынуждает переход к интервальным понятиям. Расширим понятие среднего, отказавшись от его определения как числа.

Интервальным средним признака f называется отрезок $[\underline{M}f, \overline{M}f]$ с границами $\underline{M}f$ — *нижним средним* и $\overline{M}f$ — *верхним средним*. В частном случае равенства $\underline{M}f=\overline{M}f=Mf$ интервальное среднее переходит в *точное* и обозначается без черточек. Другой частный случай, когда границы интервального среднего совпадают с минимальным и максимальным значениями функции: $\underline{M}f=\inf f$, $\overline{M}f=\sup f$. Это означает, что о среднем ожидаемом значении признака f ничего неизвестно. Оно любое в промежутке значений f. Здесь уже не важно, устойчиво или неустойчиво явление, можно ли организовать повторы или нет: ничто есть ничто.

Таким образом, интервальное среднее $[\underline{M}f, \overline{M}f]$ дает широкий охват описания среднестатистических свойств признаков от полного незнания до точного знания среднего. Интерпретировать его можно по-разному. И как диапазон возможных значений существующего (но неизвестного) точного значения Mf в статистически устойчивом эксперименте и как введение защитного допуска на Mf от неустойчивости явления. И вообще, как более общее понятие, чем точное среднее, когда последнее в силу упомянутых ранее обстоятельств не определяется. Важно то, что в отличие от точного интервальное среднее всегда существует хотя бы потому, что всегда имеется возможность перехода к крайнему случаю полного названия этого среднего.

При $f(x)=A(x)$ интервальное среднее превращается в интервальную вероятность $[\underline{P}(A), \overline{P}(A)]$; $\underline{P}(A)=\underline{M}A$, $\overline{P}(A)=\overline{M}A$. На простейших примерах проиллюстрируем некоторые ее интерпретации.

Пример 1.1. Пусть имеется астрагал (кость конечности животного) с четырьмя состояниями при бросании. Понятно, что точные вероятности всех этих состояний существуют. Но чтобы найти их, требуется бесконечно долгий эксперимент. По ограниченному эксперименту можно оценить вероятности лишь приближенно в виде некоторых доверительных интервалов значений. Это и будут интервальные вероятности.

Пример 1.2. Имеются два одинаковых с виду кубика с нанесенными на гранях от 1 до 6 очков. Для симметричного кубика вероятность выпадения одной грани, скажем шестерки, равна 1/6. У другого же кубика центр тяжести

смещен в сторону, противоположную шестерке (в средние века по вполне понятной хитрости в кубики заливали куски металла), и вероятность выпадения шести очков стала больше 1/6: $p_6 > 1/6$. Неизвестно, какой из кубиков и в какой последовательности подставляется в игре. Тогда вероятность шестерки будет интервальная $[1/6, p_6]$. При увеличении повторов эксперимента относительная частота шестерок в разных сериях может сходиться к любому числу из этого интервала. Здесь неустойчивость вызвана вмешательством человеческого фактора в виде совершенно неизвестной стратегии подстановок. При известных стратегиях интервал сужается вплоть до точных значений (когда берется кубик одного типа).

Пример 1.3. Пусть монета не бросается, а одна или другая ее сторона показывается некоторым субъектом. Это будет демонстрацией статистической неустойчивости в своей природе человеческого фактора даже при полнейшем желании делать показ независимо а равновозможно. Если здесь и можно описать среднестатистический результат, то лишь интервальной вероятностью (зависящей от психических особенностей субъекта).

Математическая модель явления. Математическая модель — это единообразная удобная для разработчиков символьная форма описания результатов явления совместно с присущими им закономерностями. Предыдущими рассуждениями мы подошли к ее пониманию. Имеется пространство $\mathscr{X}$ элементарных исходов, а на нем определено неисчислимое множество числовых признаков, одни из которых мажорируют другие, мажорируются третьими или их линейными комбинациями и т. д. Для всех абсолютно ограниченных признаков f, класс которых обозначим $\mathscr{F}_{00} = \{f : \sup_x |f(x)| < \infty\}$, существуют интервальные средние $\underline{M}f$, $\bar{M}f$ внутри промежутка значений f. Но этого может оказаться еще мало.

Модели станут более универсальными, если избирательно допускать существование средних на некоторых неограниченных признаках. Для ряда признаков это естественно: так, если признак f не ограничен снизу, но ограничен сверху $\sup f = H < \infty$ числом H, то верхнее среднее $\bar{M}f$ не может быть больше H, поэтому существует. Итак, имеем $\bar{M}f < \infty$ на классе $\mathscr{F}_0 = \{f : \sup f < \infty\}$ всех ограниченных сверху функций. Аналогично, нижние средние $\underline{M}f$ существуют на классе $(-\mathscr{F}_0)$ (получается из $\mathscr{F}_0$ переменной знака функций) всех ограниченных снизу функций. Нужды практики при их строгом оформлении требуют в целом ряде задач существования средних на более широких классах неограниченных признаков.

Назовем класс $\mathscr{F}$ признаков, на котором определены $\bar{M}f < \infty$, *областью существования верхних средних*. Мы увидим далее, что перемена знака у признаков приведет к классу $(-\mathscr{F})$, на котором будут существовать нижние средние: $\underline{M}f$, $f \in -\mathscr{F}$, а на их пересечении будут и те, и другие, т. е. интервальные средние. Не исключается $\mathscr{F} = \mathscr{F}_0$, но это будет самый узкий возможный класс.

Если это более широкий класс, то $\mathscr{F}$ должен удовлетворять трем свойствам:

C1. $g \in \mathscr{F}$, $f \leqslant g \Rightarrow f \in \mathscr{F}$;

C2. $f \in \mathscr{F}$, c, $b^+ \in \mathscr{R}$, $b^+ \geqslant 0 \Rightarrow b^+ f + c \in \mathscr{F}$;

C3. f, $g \in \mathscr{F} \Rightarrow f + g \in \mathscr{F}$.

(Очевидно, для $\mathscr{F}_0$ все они выполняются.) Из перечисленных свойств следует: 1) $c \in \mathscr{F}$, где c есть признак, принимающий постоянное значение c (получается из **C**2 при $b^+ = 0$); 2) $\mathscr{F} \supset \mathscr{F}_0$ (из **C**1 и 1); 3) $f_i \in \mathscr{F}$, $i = 1, \ldots, k \Rightarrow c + \sum_{i=1}^{k} b^+_i f_i \in \mathscr{F}$, где плюс означает неотрицательность чисел $b^+_i \geqslant 0$. Последнее свойство называется *полулинейностью*, а класс $\mathscr{F}$ со свойствами **C**1, **C**2, **C**3 — *полулинейным*.

Математическая модель явления включает в себя: а) пространство $\mathscr{X}$ элементарных событий, б) полулинейный класс $\mathscr{F}$ признаков (если он шире $\mathscr{F}_0$), в) средние на нем; а все вместе это $\{\mathscr{X}, \mathscr{F}, \bar{M}, \underline{M}\}$ или $\{\mathscr{X}, \bar{M}, \underline{M}\}$ при $\mathscr{F} = \mathscr{F}_0$. Перейдем к основной части модели: требованиям к $\bar{M}$ и $\underline{M}$.

Аксиоматика. Средние $\underline{M}f$, $\bar{M}f$ разных признаков должны находиться в определенной взаимной пропорции (устанавливаемой, например, из примера 1 или 2 введения). При принятом нами подходе, в котором средние являются самостоятельными составляющими модели, отношения между средними должны постулироваться. Среди отношений нужно выделить основополагающие, называемые *аксиомами*, тогда другие становятся вытекающими из аксиом свойствами. Важно, сколько задать аксиом.

Если аксиоматически связать средние сразу слишком жесткими отношениями, то это сузит классы возможных моделей. Лучше создать как можно более свободную конструкцию, имея в виду, что усиление всегда достижимо внедрением дополнительных свойств внутрь конкретной модели или класса моделей. Таким образом, аксиом должно быть по возможности меньше, но чтобы остаться в рамках физической интерпретируемости модели (иначе возникнет новая теория).

Аксиомы средних. Для $\forall f, g \in \mathscr{F}$:

A1. $g \geqslant f \Rightarrow \bar{M}g \geqslant \bar{M}f$;

A2. $\bar{M}(b^+ f + c) = b^+ \bar{M}f + c$; $\forall b^+ \geqslant 0$, $c \in \mathscr{R}$;

A3. $\bar{M}(f + g) \leqslant \bar{M}f + \bar{M}g$;

A4. $\underline{M}(-f) = -\bar{M}f$.

Здесь стрелка $\Rightarrow$ заменяет слово «следует», а плюс в верхнем индексе указывает на неотрицательность числа. Обсудим аксиомы.

A1 — аксиома сохранения порядка: если признак g мажорирует f, то верхнее среднее у него не меньше, чем у f. Иначе и не может быть, так как значения g всегда будут меньше f.

A2 — аксиома переноса: умножение признака f на неотрицательное число и прибавление к нему любого постоянного

числа приводит к таким же операциям над $\overline{M}f$. Это понятно, ибо так преобразуется каждое значение признака.

A3 — аксиома полуаддитивности: верхнее среднее от суммы признаков не больше суммы их верхних средних. В самом деле, для суммы одинаковых признаков $f+f$ на основании **A2** при $c=0$, $b^+=2$ будет справедливо равенство: $\overline{M}(f+f)=\overline{M}f+\overline{M}f$. Это как раз случай «синфазного» сложения, когда положительная часть складывается с такой же положительной, отрицательная — с отрицательной. В общем, сложение разных признаков не будет синфазным, что и ведет к уменьшению $\overline{M}(f+g)$ по сравнению с $\overline{M}f+\overline{M}g$.

A4 — аксиома обращения: однозначно связывает нижние средние с верхними, для чего у признака меняется знак. Следует из того, что для любого интервала $[\underline{m}, \overline{m}]$ при перемене знака границы меняются местами: $[-\overline{m}, -\underline{m}]$. По этой аксиоме, коль скоро $\overline{M}f$ определены на $\mathcal{F}$, то $\underline{M}f$ будут определены на $-\mathcal{F}$.

Конечно же, дело «вкуса», какие из равносильных свойств средних брать в качестве аксиом, поэтому сделанный выбор не нуждается в обсуждении.

Определение интервальной модели средних, основные свойства. *Интервальной моделью средних* (сокращенно, ИМ) называется совокупность верхних средних $\overline{M}f$ на заданной области их существования $f\in\mathcal{F}$ (удовлетворяющей свойствам **C1**, **C2**, **C3**) и нижних средних $\underline{M}f$, $f\in-\mathcal{F}$, *согласованных* между собой в том смысле, что удовлетворяются аксиомы **A1** сохранения порядка, **A2** переноса, **A3** полуаддитивности и **A4** обращения. Интервальная модель средних обозначается $\mathcal{M}$ или $\langle\overline{M}\mathcal{F}\rangle$.

Непосредственно из аксиом выводятся такие свойства (полагаем признаки из соответствующих областей существования):

1. $\overline{M}(b^+f+c)=b^+\overline{M}f+c$ (следует из **A2** переменной знака и **A4**);

2. Для любых констант $\overline{M}c=c$ (из **A2** и **1** при $b^+=0$);

3. $\underline{M}(f+g)\geqslant\underline{M}f+\underline{M}g$ (из **A3** и **A4**);

4. $\overline{M}\sum\limits_1^k f_i\leqslant\sum\limits_1^k\overline{M}f_i$ — верхняя полуаддитивность (из **A3** по индукции);

5. $\underline{M}\sum f_i\geqslant\sum\underline{M}f_i$ — нижняя полуаддитивность (из **3** по индукции);

6. $\underline{M}f\leqslant\overline{M}f$ (**A3** и **A4**$\Rightarrow 0=M0=\overline{M}(f-f)\leqslant\overline{M}f-\underline{M}f$);

7. $f\geqslant g\Rightarrow\underline{M}f\geqslant\underline{M}g$ (из **A1** переменной знака функции и **A4**);

8. $\inf f\leqslant\underline{M}f$, $\overline{M}f\leqslant\sup f$ (из **7**, **A1** и **2**, так как $\inf f\leqslant f\leqslant\sup f$);

9. $|\overline{M}|f=\max\{|\underline{M}f|,\ |\overline{M}f|\}\leqslant\overline{M}|f|$ (так как $\pm f\leqslant|f|\Rightarrow\overline{M}f\leqslant\overline{M}|f|$, $-\underline{M}f\leqslant\overline{M}|f|$, здесь $|\overline{M}|$ — новое обозначение, называемое максимальным по модулю средним);

10. $\underline{M}(f+g)\leqslant\underline{M}f+\overline{M}g\leqslant\overline{M}(f+g)$ — псевдоаддитивность (из **3** и **A4** $\underline{M}f=\underline{M}[(f+g)-g]\geqslant\underline{M}(f+g)-\overline{M}g$; из **A3** и **A4** $\overline{M}g=\overline{M}[(g+f)-f]\leqslant\overline{M}(g+f)-\underline{M}f$);

11. Mg точное $\Rightarrow\overline{M}(f+g)=\overline{M}f+Mg$; $\underline{M}(f+g)=\underline{M}f+Mg$ (из **A3**, **3** и **10**);

12. $M\sum\limits_{i=1}^k f_i=\sum\limits_{i=1}^k Mf_i$ — конечная аддитивность точных средних Mf_i, $i=1, \ldots, k$;

13. Непрерывность средних по отношению к равномерной при $n\to\infty$ сходимости функций:

$$\sup_x|f_n(x)-f(x)|\to 0\Rightarrow\underline{M}f_n\to\underline{M}f,\ \overline{M}f_n\to\overline{M}f$$

(так как $|\underline{M}f_n-\underline{M}f|$ и $|\overline{M}f_n-\overline{M}f|$ не превышают $\sup|f_n-f|$).

Итак, из аксиом получены такие естественные свойства, что нижнее среднее $\underline{M}f$ не больше верхнего **6** и не меньше нижней грани функции **8**. Для $\underline{M}f$ также справедлив закон переноса **1**. Свойства **3**, **4**, **5**, **10** — подмена свойства аддитивности **12**, справедливого для точных средних.

Формально определив ИМ, мы пока не указали конкретно, что представляет собой область существования $\mathcal{F}$ верхних средних и насколько она может быть шире $\mathcal{F}_0$. Ограничились только ее свойствами полулинейности **C1**, **C2**, **C3**. Конкретизация $\mathcal{F}$ будет произведена после того, как мы познакомимся с универсальным способом задания ИМ, к которому и обратимся.

1.2. ПРОДОЛЖЕНИЕ ПЕРВИЧНЫХ СРЕДНИХ

Вступление. Утилитарные достоинства той или иной теории, тех или иных моделей и методов определяются по трем основным направлениям:

1) универсальность — нацеленность на работу с широким классом объектов или явлений;

2) удобство и гибкость аппарата, податливость к упрощениям, грубым прикидкам;

3) простота и естественность перевода «языка явлений» на язык адекватных им моделей и физическая интерпретируемость параметров модели.

Во многом направления перекликаются между собой, но скорее это три похожих зайца, бегущих в разные стороны. Задержим взгляд на последнем.

У модели должны быть выделены приоритетные, первичные параметры, связывающие ее с явлением, варьируя число и значения которых, можно достичь адекватности модели, подобно прибору, подкручивая ручки которого добиваются настройки. Мы покажем, что любой набор признаков с заданными для них средними (точными или размытыми, в виде интервала или одной границы) может играть роль первичных параметров ИМ. Этим результатом убиваются все три зайца: универсальность достигается за счет разнообразия выбора первичных признаков, а гибкость и интерпретируемость — выбором их числа и варьируемости границ интервальных средних, смысл которых нам уже известен.

Первичные признаки и средние. Пусть $\mathcal{G}^*$ — набор *первичных признаков*. Неважно, какой он, конечный или бесконечный, состоит из ограниченных признаков или нет — это вопрос приложений. Каждый из этих признаков $g \in \mathcal{G}^*$ есть функция $g(x)$ переменной x, пробегающей множество $\mathcal{X}$ элементарных исходов, поэтому они называются также первичными функциями. Для каждого первичного признака $g \in \mathcal{G}^*$ заданы интервальные первичные средние $\underset{\sim}{M}g$, $\tilde{M}g$ или одно из этих значений. Волнистая черта подчеркивает не столько то, что это первичные средние, сколько то, что они могут быть в общем не согласованы между собой в смысле выполнения аксиоматических свойств **A1**, **A2**, **A3** и **A4**. Необязательность контроля согласованности дает определенную вольность, упрощающую процедуру задания первичных средних, а значит, и самих моделей.

Для g может быть известно $\underset{\sim}{M}g$ и (или) $\tilde{M}g$, поэтому выделим из набора $\mathcal{G}^*$ два поднабора: верхний $\mathcal{G}_в$, на котором определены $\tilde{M}g$, и нижний $\mathcal{G}_н$ с $\underset{\sim}{M}g$. На их пересечении $\mathcal{G}_в \cap \mathcal{G}_н$ заданы и те, и другие, т. е. первичные интервалы $\underset{\sim}{M}g$, $\tilde{M}g$ средних.

Нижний первичный поднабор $\mathcal{G}_н$ моментально обращается в верхний. Для этого, опираясь на **A4**, определим $\tilde{M}(-g) = -\underset{\sim}{M}g$, и, таким образом, вместо g с заданным нижним средним $\underset{\sim}{M}g$ имеем функцию с противоположным знаком $g_1 = -g$ с заданным на ней уже верхним средним, равным $\tilde{M}g_1 = -\underset{\sim}{M}g$. Подвергая указанному обращению все $g \in \mathcal{G}_н$, исходный набор $\mathcal{G}^*$ переводим в эквивалентный верхний первичный набор $\mathcal{G} = \mathcal{G}_в \cup (-\mathcal{G}_н)$. Теперь заданным считается $\tilde{M}g$, $g \in \mathcal{G}$. Такое приведение, подчас неудобное с позиций естественной интерпретируемости параметров модели, оказывается тем не менее весьма удобным для унификации и упрощения записи формул.

Несмотря на отсутствие требования согласованности, первичные верхние средние $\tilde{M}g$, $g \in \mathcal{G}$, нельзя задавать совсем произвольно, так как это может привести к противоречию (скажем, $g(x) \geqslant 0$, а $\tilde{M}g < 0$).

Первичные средние $\tilde{M}g$, $g \in \mathcal{G}$, называются *непротиворечивыми*, если при любом конечном выборе $g_i \in \mathcal{G}$, чисел $c^+_i \geqslant 0$ и c, таком, что $c + \sum c^+_i g_i(x) \geqslant 0$, имеет место $c + \sum c^+_i \tilde{M}g_i \geqslant 0$. Признаки в левой части первого неравенства назовем *вторичными*.

Обозначим $\mathcal{L}^+\mathcal{G} = \{g(x) = c + \sum c^+_i g_i(x),\ g_i \in \mathcal{G}\}$ — класс всех возможных вторичных признаков, т. е. конечных линейных комбинаций первичных функций g_i с неотрицательными при них коэффициентами c^+_i и произвольным свободным членом c. Назовем $\mathcal{L}^+\mathcal{G}$ *полулинейной оболочкой* $\mathcal{G}$ (или полулинейной комбинацией признаков $\mathcal{G}$) и, пользуясь аксиомой **A2**, формально перенесем на нее первичные средние

$$\tilde{M}g = \tilde{M}\left(c + \sum c^+_i g_i\right) = c + \sum c^+_i \tilde{M}g_i,\ g \in \mathcal{L}^+\mathcal{G}. \qquad (1.1)$$

Теперь требование непротиворечивости формулируется следующим образом:

$$0 \leqslant g \in \mathcal{L}^+\mathcal{G} \Rightarrow \tilde{M}g \geqslant 0,$$

т. е. для вторичных признаков g, принимающих только неотрицательные значения, верхнее среднее (заданное, преобразованное из нижнего или перенесенное с $\mathcal{G}$ на $\mathcal{L}^+\mathcal{G}$) не должно быть отрицательным. Необходимость непротиворечивости не требует пояснений. Ниже на примерах будет показано, во что оно выливается.

Переходим к основному результату. Обозначим $\mathcal{F}_{\mathcal{G}} = \{f : f \leqslant g \in \mathcal{L}^+\mathcal{G}\}$ — класс признаков, таких, что каждый из них мажорируется хотя бы одним вторичным. Таким образом, имеем $f \in \mathcal{F}_{\mathcal{G}}$, если в $\mathcal{L}^+\mathcal{G}$ существует хотя бы один признак g, такой, что $f \leqslant g$. Очевидно, что в $\mathcal{F}_{\mathcal{G}}$ входят все ограниченные сверху признаки (мажорируются постоянными $c \in \mathcal{L}^+\mathcal{G}$), поэтому $\mathcal{F}_0 \subset \mathcal{F}_{\mathcal{G}}$, причем $\mathcal{F}_0 = \mathcal{F}_{\mathcal{G}}$, если все $g \in \mathcal{G}$ ограничены.

Назовем $\mathcal{F}_{\mathcal{G}}$ *классом $\mathcal{G}$-мажорируемых признаков*.

Теорема продолжения и согласования средних.

Теорема 1.1. *Если первичные средние $\tilde{M}g_i$, $g_i \in \mathcal{G}$, непротиворечивы, то по формуле*

$$\overline{M}f = \inf_{g:\, f(x) \leqslant g(x) \in \mathcal{L}^+\mathcal{G}} \tilde{M}g,\quad f \in \mathcal{F}_{\mathcal{G}};\quad \underline{M}f = -\overline{M}(-f),\quad f \in -\mathcal{F}_{\mathcal{G}}, \qquad (1.2)$$

они продолжаются, делаясь согласованными, на все $\mathcal{G}$-мажорируемые признаки $f \in \mathcal{F}_{\mathcal{G}}$, образуя ИМ $\mathcal{M}$ с областью существования верхних средних $\mathcal{F} = \mathcal{F}_{\mathcal{G}}$.

Иначе, также (1.2) записываем:

$$\overline{M}f = \inf\{\tilde{M}g : f(x) \leqslant g(x) \in \mathcal{L}^+\mathcal{G}\}.$$

Доказательство. Аксиомы **A1** и **A2** очевидны из (1.2), аксиома **A4** выполняется по определению. Остается проверить аксиому **A3**. Так как $f_1(x) \leqslant g^*(x) \in \mathcal{L}^+\mathcal{G}$, $f_2(x) \leqslant g^{**}(x) \in \mathcal{L}^+\mathcal{G} \Rightarrow f_1(x) + f_2(x) \leqslant g^*(x) + g^{**}(x) = g(x) \in \mathcal{L}^+\mathcal{G}$ и $\tilde{M}g^* + \tilde{M}g^{**} = \tilde{M}g$, то

$$\overline{M}(f_1 + f_2) = \inf\{\tilde{M}g : f_1 + f_2 \leqslant g \in \mathcal{L}^+\mathcal{G}\} \leqslant$$

$$\leqslant \inf\{[\tilde{M}g^* + \tilde{M}g^{**}] : f_1 \leqslant g^* \in \mathcal{L}^+\mathcal{G},\ f_2 \leqslant g^{**} \in \mathcal{L}^+\mathcal{G}\} = \overline{M}f_1 + \overline{M}f_2,$$

что и требовалось доказать.

Согласно (1.2) для нахождения $\overline{M}f$ из первичных составляются вторичные признаки $g(x) = c + \sum c^+_i g_i(x)$ так, чтобы они мажорировали $f(x)$, т. е. $g(x) \geqslant f(x)$. Каждый из этих $g(x)$ будет иметь в общем свое значение $\tilde{M}g$, определяемое (1.1). Берется «наилучшее» среди них, т. е. то, которое минимально.

В (1.2) заложен следующий принцип конструктивной математики: получать лишь результаты, к которым можно сколь угодно

приблизиться за конечное число операций. Именно поэтому вторичные признаки набираются как конечные суммы первичных. Действие этого принципа раскрывается вытекающим из (1.2) следствием.

Следствие. *Каждому f, верхнее среднее $\bar{M}f$ которого конечно, и заданному $\varepsilon>0$ всегда можно указать такую мажорирующую конечную линейную комбинацию $g(x)=c+\sum c^+_i g_i(x)\geqslant f(x)$ первичных признаков (т. е. вторичный признак), что*

$$\bar{M}f+\varepsilon\geqslant c+\sum c_i^+\bar{M}g_i=\bar{M}g.$$

Поскольку $g\geqslant f\Rightarrow\bar{M}g\geqslant\bar{M}f$, для следствия верно $|\bar{M}g-\bar{M}f|\leqslant\varepsilon$. При уменьшении ε понадобится вовлекать в g большее число первичных признаков для приближения к $\bar{M}f$, количество операций возрастает.

Отношения между определенными нами множествами признаков иллюстрируются рис. 1.2. Здесь каждая из верхних полусфер включает в себя все нижние. Согласно теореме 1.1 средние с первичного набора $\mathcal{G}$ продолжаются на линейные комбинации $\mathcal{L}^+\mathcal{G}$ первичных признаков (вторичные признаки) и уже через них распространяются на класс $\mathcal{F}_\mathcal{G}$ признаков, мажорируемых вторичными. Это и будет областью существования верхних средних; ядром ее является класс $\mathcal{F}_0$ всех ограниченных сверху признаков: $\mathcal{F}_0\subset\mathcal{F}_\mathcal{G}$. В $\mathcal{F}_\mathcal{G}$ войдут и неограниченные признаки, если неограниченные имеются среди первичных, а иначе, $\mathcal{F}_\mathcal{G}=\mathcal{F}_0$.

Согласованные первичные средние. Замечательно то, что (1.2) не только дает верхние средние $\bar{M}f$ для $\forall f\in\mathcal{F}_\mathcal{G}$ (в частности, $\underline{M}f$, $\bar{M}f$ для $\forall f\in\mathcal{F}_{00}$), но и попутно, подстановкой $g_i\in\mathcal{G}$ как части f, уточненные и согласованные между собой $\bar{M}g_i$. Так как g_i мажорирует сам себя как первичный признак, то согласно (1.2) $\bar{M}g_i\geqslant\bar{M}g_i$, $g_i\in\mathcal{G}$. Если получается $\bar{M}g_i=\bar{M}g_i$, то первичное значение $\bar{M}g_i$ само по себе уже согласовано с другими средними и обозначается $\bar{M}g_i$. Если же $\bar{M}g_i>\bar{M}g_i$, то найдутся вторичные признаки, мажорирующие g_i, уточняющие $\bar{M}g_i$, а само $\bar{M}g_i$ будет *несогласованным* и без ущерба может быть изъято из первичного набора. Таким образом, *могут влиять на вид ИМ только согласованные первичные средние*, да и то, как будет видно дальше,

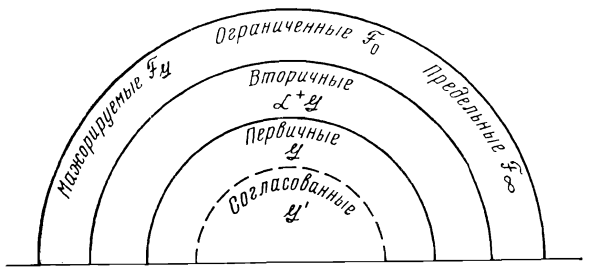


Рис. 1.2. Этапы продолжения средних

не обязательно все. Их минимальное число называется *размерностью* ИМ. Интервальная модель, первичными для которой являются $\bar{M}g$, $g\in\mathcal{G}$, обозначается $\langle\bar{M}\mathcal{G}\rangle$, а если все несогласованные первичные признаки исключены из набора, то $\langle\bar{M}\mathcal{G}'\rangle$, $\mathcal{G}'\subset\mathcal{G}$.

Расширение первичного набора включением в него средних $\bar{M}f$, найденных по (1.2), не меняет ИМ, поэтому получаем одно и то же, если брать за первичный набор $\mathcal{G}$, $\mathcal{G}'$ или все $\mathcal{F}=\mathcal{F}_\mathcal{G}:\mathcal{M}=\langle\bar{M}\mathcal{G}\rangle=\langle\bar{M}\mathcal{G}'\rangle=\langle\bar{M}\mathcal{F}\rangle$. Заметим при этом одну особенность, что несогласованные первичные признаки g_i, для которых $\bar{M}g_i>\bar{M}g_i$, обязательно должны быть мажорируемы хотя бы одним из вторичных признаков $g\in\mathcal{L}^+\mathcal{G}$, который и даст на основе (1.2) уточненное значение $\bar{M}g_i$. Таким образом, $\mathcal{L}^+\mathcal{G}'=\mathcal{L}^+\mathcal{G}$ и $\mathcal{F}_{\mathcal{G}'}=\mathcal{F}_\mathcal{G}$.

Из сказанного также следует, что первичные средние $\widetilde{M}g_i$, $g_i\in\mathcal{G}$, являются согласованными в том и только том случае, если справедливо: $g_i(x)\leqslant g(x)\in\mathcal{L}^+\mathcal{G}\Rightarrow\bar{M}g_i\leqslant\bar{M}g$ для всех $g_i\in\mathcal{G}$ и $g\in\mathcal{G}^+\mathcal{L}$, удовлетворяющих первому неравенству. Тогда $\bar{M}g_i=\bar{M}g_i$, $\forall g_i\in\mathcal{G}$.

Признаки случайных величин. В качестве примера рассмотрим тот случай, когда $\mathcal{X}=\mathcal{R}$ — числовая прямая. Такое явление имеет числовые исходы и называется *случайной величиной* (сокращенно с. в.). Признаками f с. в. могут быть любые числовые функции $f(X)$ на $\mathcal{R}$ (далее числовые исходы $X\in\mathcal{R}$ обозначаются заглавными буквами).

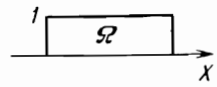
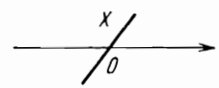
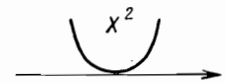
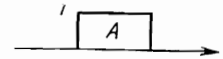
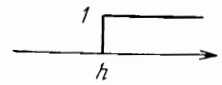
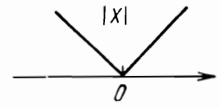
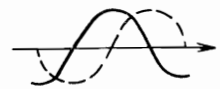
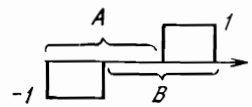
Случайная величина называется дискретной, если возможные ее значения составляют конечное или счетное множество Ω точек. Удобно считать для таких с. в. пространством исходов всю прямую $\mathcal{R}$, а тот факт, что $\Omega\subset\mathcal{R}$, отразить добавлением первичной вероятности $\underline{P}(\Omega)=1$ (напомним, что $\underline{P}(A)=MA(x)$), согласно которой Ω есть достоверное событие. Так как $\bar{P}(\Omega)\geqslant\underline{P}(\Omega)=1$, то можно написать $P(\Omega)=1$.

Некоторые характерные признаки с. в. упорядочены в табл. 1.1. Дадим два примера расчета их средних по теореме 1.1.

Пример 1.4. Пусть известно, что среднее с. в. равно m; записываем $MX=m$ (т. е. $\bar{M}(\pm X=\pm m)$. Оно согласовано и продолжается на основании свойства **12** на вторичные признаки $f(X)=c+c_1X$: $M(c+c_1X)=c+c_1MX$. Область существования $\mathcal{F}$ составляют функции, мажорируемые прямыми $c+c_1X$. Так как $c^2-2cX\geqslant-X^2$, то $-X^2\in\mathcal{F}$ и $\bar{M}(-X^2)=\min\limits_c(-2cm+c^2)=-m^2$, т. е. $MX^2=m^2$. Функция X^2 не мажорируется прямыми, поэтому $\bar{M}X^2=\infty$ и $X^2\notin\mathcal{F}$. Для любых индикаторных признаков $A(x)$ событий невозможно подыскать мажорирующую прямую $c+c_1X$, кроме $c=0$, $c_1=1$, поэтому $\underline{P}(A)\equiv0$, $\bar{P}(A)\equiv1$. Отсюда вывод, что первичное среднее, будучи в единственном числе, нетривиальных данных о событиях не несет.

Пример 1.5. Пусть первичным для с. в. X является верхнее среднеквадратическое (мощность с. в.) $\bar{M}X^2=\bar{b}$, что порождает ИМ размерности 1. Об-

Таблица 1.1

Характеристика случайной величины	Формальное обозначение	Признаки
Ω — множество значений с. в.	$\underline{P}(\Omega)=1$	
Среднее с.в. точно равно m лежит в интервале $\underline{m}$, $\overline{m}$	$MX=m$ $\underline{M}X=\underline{m}$, $\overline{M}X=\overline{m}$	
Среднеквадратическое с.в. не меньше $\underline{b}$ и не больше $\overline{b}$	$\underline{M}X^2=\underline{b}$, $\overline{M}X^2=\overline{b}$	
Вероятность попадания в отрезок A лежит в заданных пределах	$\underline{P}(A)$, $\overline{P}(A)$	
Вероятность «выброса» за уровень h не больше p_h	$\overline{P}(X>h)=p_h$	
Среднее модуля с.в. лежит в указанных пределах	$\underline{M}\lvert X\rvert$, $\overline{M}\lvert X\rvert$	
Средние гармонические признаков	$\underline{M}\cos uX$, $\overline{M}\cos uX$ $\underline{M}\sin uX$, $\overline{M}\sin uX$	
Вероятность попадания в отрезок A больше, чем в отрезок B (§ 1.4)	$\overline{M}[B(x)-A(x)]\leqslant 0$	

ласть существования верхних средних $\mathcal{F}_{\mathcal{G}}$ составляют все признаки $f(X)$, мажорируемые параболами вида $c+c_2{}^+X^2$, т. е. те f, для которых $\overline{\lim\limits_{|X|\to\infty}} f(X)/X^2<\infty$. При $a_1>0$, подбирая, как это видно из рис. 1.3, соответствующим образом мажорирующую параболу: $c=0$, $c^+{}_2=a_1{}^{-2}$, $\{a_1\leqslant X\leqslant a_2\}\leqslant X^2/a^2{}_1$, находим: $\overline{P}(a_1\leqslant X\leqslant a_2)\leqslant \overline{M}X^2/a^2{}_1=\overline{b}/a^2{}_1$ (при $a_2=\infty$ имеем аналог неравенства Чебышева). Равенство будет, когда правая часть меньше 1, т. е. при $a_1>\sqrt{\overline{b}}$ (при $a_2<0$ заменяется a_1 на a_2), иначе вероятность $\overline{P}$ тривиальна и равна 1. Таким образом, знание $\overline{b}$ делает нетривиальными верхние вероятности отрезков, удаленных от начала оси по крайней мере на расстояние, превышающее $\sqrt{\overline{b}}$.

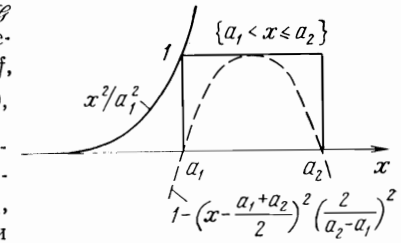


Рис. 1.3. Расчет вероятностей по мощности

Найдем при тех же исходных данных $\overline{M}X$. Мажорируя прямую X параболой: $X\leqslant c+c^+{}_2X^2$, $\forall X$, что эквивалентно $4c_0c^+{}_2\geqslant 1$, и минимизируя при последнем ограничении (замененном на равенство) среднее параболы $\overline{M}(c+c^+{}_2X^2)=c+c^+{}_2\overline{b}$, находим ее коэффициенты $c^+{}_2=1/(2\sqrt{\overline{b}})$, $c=1/4c^+{}_2=\sqrt{\overline{b}}/2$, подстановкой которых получаем $\overline{M}X=\sqrt{\overline{b}}$. Аналогичен путь нахождения $\overline{M}(X\pm m)^2$, для чего $(X\pm m)^2$ мажорируется параболой $c+c^+{}_2X^2$, откуда вытекает требование на ее коэффициенты: $c>c^+{}_2m^2/(c^+{}_2-1)$, минимизируя при этом ограничении среднее параболы, получаем коэффициенты $c^+{}_2=1+|m|/\sqrt{\overline{b}}$, $c=(\sqrt{\overline{b}}+|m|)|m|$, откуда $\overline{M}(X\pm m)^2=(|m|+\sqrt{\overline{b}})^2$.

Пусть теперь к среднеквадратическим в качестве первичных добавляется нулевое среднее $MX=0$ (что эквивалентно $\overline{M}(\pm X)=0$). Тогда ИМ сужается, ее размерность становится равной 3. Вторичными будут уже любые направленные вверх параболы со средними на них $\overline{M}[c+c^+{}_2(X-c_1)^2]=c+c^+{}_2\overline{b}+c^+{}_2c^2{}_1$, где использовалось свойство **11** § 1.1. В частности, отсюда $\overline{M}(X\pm m)^2=\overline{M}X_2+m^2=\overline{b}+m^2$.

Будем искать вероятности событий, а именно отрезков. Среди парабол, мажорирующих индикаторный признак отрезка: $\{a_1\leqslant X\leqslant a_2\}\leqslant c+c^+{}_2(X-c_1)^2$, а это будет при $c\geqslant 0$, $c_1\leqslant a_1$, $c+c^+{}_2(a_1-c_1)^2\geqslant 1$, нужно найти такую подбором коэффициентов c, c_1, $c^+{}_2$, у которой минимально верхнее среднее. Несложными вычислениями находим: $c=0$, $c_1=-\overline{b}/a_1$, $c^+{}_2=(a_1+\overline{b}/a_1)^{-2}$, где считалось $a_1>0$ (при $a_2<0$ нужно a_1 заменить на a_2), в результате вероятность равна минимальному среднему параболы: $\overline{P}(a_1\leqslant X\leqslant a_2)=(1+a^2{}_1/\overline{b})^{-1}$. Вероятность нетривиальна при любом $a_1>0$ (или $a_2<0$).

Нижнюю вероятность отрезка рассчитаем, «вписав» параболу (нанесена штриховой линией на рис. 1.3), с которой переносится на событие среднее

$$\underline{P}(a_1\leqslant X\leqslant a_2)\geqslant 1-\overline{M}\left(X-\frac{a_1+a_2}{2}\right)^2\left(\frac{2}{a_2-a_1}\right)^2=$$
$$=1-\frac{4\overline{b}+(a_1+a_2)^2}{(a_2-a_1)^2}.$$

Правая часть больше 0 при $-a_1a_2>\overline{b}$ (отсюда $a_1<0$ и $a_2>0$), и тогда получаем нижнюю вероятность (иначе, она 0). Итак, знание нулевого среднего,

уточняя верхние вероятности отрезков, отстоящих на $|a_1|$ от начала оси, делает нетривиальными нижние вероятности достаточно широких отрезков, включающих начало оси. Отметим, что дополнительное знание $\underline{M}X^2=\underline{b}$ вероятностей не меняет.

Признаки случайных процессов. Случайный процесс X_t есть нумерованная индексом t (называемым временем) последовательность с. в. Значениями t могут быть отрезок $[0, T]$ временной оси $\mathscr{R}$, вся ось, некоторые дискретные точки-отсчеты t_1, t_2, ..., t_n — на этой оси (тогда процесс становится вектором). Это сейчас неважно. Обозначим T — множество этих значений.

Пространство исходов $\mathscr{X}$ будут всевозможные реализации x_t, $t\in$T как функции времени t.

Чтобы описать процесс, нужно описать каждую с. в. X_t в отдельности своими признаками, как это было проделано, а также связь между X_t и $X_{t'}$ при различных t и $t'\in$T. Признаки этой связи, в частности, составляют произведения $X_tX_{t'}$, а их средние $MX_tX_{t'}=b_{t,t'}$, $\overline{M}X_tX_{t'}=\overline{b}_{t,t'}$ будут нижней и верхней *корреляционными функциями*. В более общем случае — это корреляции после преобразования каждой с. в. одной и той же функцией F (так называемые безынерционные преобразования), тогда первичными параметрами процесса будут $MF(X_t)F(X_{t'})$, $\overline{M}F(X_t)F(X_{t'})$. Так, если $F(X)=\{X>h\}$ — индикаторная функция превышений уровня h, то это будет корреляция превышений (при $h=0$ корреляция полярностей).

Отличными от указанных являются признаки в виде интегралов, такие как $\int\limits_{\mathrm{T}} F(X_t)\,dt$. В частности, $\frac{1}{T}\overline{M}\int\limits_0^T X_t^2\,dt$ будет верхней средней интегральной мощностью процесса на отрезке $\mathrm{T}=[0, T]$.

Таким образом, признаки и их средние являются универсальной формой описания любых явлений. Сложность описания диктуется не столько пространством $\mathscr{X}$ исходов, сколько размерностью интервальной модели, определяемой числом первичных средних.

Голая модель. Пусть первичным для ИМ является единственный факт: событие B является достоверным, и более ничего. Это соответствует первичной вероятности $\underline{P}(B)=1$ (или же $P(B)=1$, что то же самое). Такая модель называется *B-индикаторной* и обозначается $\mathscr{I}_B$. Ее описывают средние $\overline{M}f(x)=\sup\limits_{x\in B} f(x)$, а область существования составляет класс всех признаков, ограниченных сверху при $x\in B$.

Если $B=\mathscr{X}$, то модель называется *голой* и обозначается $\mathscr{I}$. Она описывается средними $\overline{M}f=\sup f$, и ей соответствует полное отсутствие данных о явлении, а область существования составляет класс $\mathscr{F}_0$ всех ограниченных сверху признаков.

Модифицированная формула продолжения. Замена первичных признаков $g_i\in\mathscr{G}$ и их средних $\overline{M}g_i$ (приведенных к верхним) на $g'_i(x)=c^+_ig_i(x)+c$ и на $\overline{M}g'_i=c^+_i\overline{M}g_i+c$, очевидно, носит эквивалентный характер. Центрируем первичные признаки, заменив g_i на $\overset{\circ}{g}_i(x)=g_i(x)-\overline{M}g_i$, и тогда $\overline{M}\overset{\circ}{g}_i=0$. *Центрированный* набор признаков обозначим $\overset{\circ}{\mathscr{G}}$ и $\overline{M}\overset{\circ}{\mathscr{G}}=0$ — все нулевые первичные значения. Для этого набора (1.2) преобразуется:

$$\overline{M}f=\inf\{[c+\sum_i c_i^+\overline{M}\overset{\circ}{g}_i]:\quad c+\sum_i c_i^+\overset{\circ}{g}_i(x)\geqslant f(x)\}=$$

$$=\inf\{c: c\geqslant f(x)-\sum_i c_i^+\overset{\circ}{g}_i(x)\}=$$

$$=\inf_{c_i^+}\sup_x[f(x)-\sum_i c_i^+\overset{\circ}{g}_i(x)]. \tag{1.3}$$

Очевидно, любой вторичный признак вида $\overset{\circ}{g}(x)=\sum c^+_i\overset{\circ}{g}_i(x)$ центрирован, т. е. его первичное верхнее среднее $\overline{M}\overset{\circ}{g}=0$ и, наоборот, любой центрированный признак представляется в указанном виде. Поэтому смысл формулы (1.3) состоит в отыскании такого из вторичных центрированных признаков, который наименьшим образом отклоняется вверх от функции $f(x)$, т. е. наилучшей, что ли, верхней аппроксимации функции $f(x)$ центрированными вторичными признаками. Вычисления по (1.3) поясним примером.

Пример 1.6. Пусть случайная величина (т. е. $\mathscr{X}=\mathscr{R}$) задана одним первичным значением $\overline{M}e^{-|X|}=\mu$. Требуется найти $\overline{P}(0\leqslant X\leqslant d)$. Задача по (1.3) сводится к нахождению величины $\overline{P}(0\leqslant X\leqslant d)=\min\limits_{c^+}\max\limits_X[\{0\leqslant X\leqslant d\}-c^+(e^{-|X|}-\mu)]=\min\limits_{c^+}\max\{1-c^+(e^{-d}-\mu),\ c^+\mu\}$ (для наглядности советуем нарисовать графики функций). Пусть $e^{-d}<\mu$, тогда минимум достигается при равенстве обеих частей под знаком максимума, откуда $c^+=e^d$, $\overline{P}(0\leqslant X\leqslant d)=e^d\mu$. При $e^{-d}\geqslant\mu$ минимум достигается при $c^+=0$, и тогда искомая вероятность равна 1.

Дополнения.

1. В область существования $\mathscr{F}$ полагались входящими признаки с $\overline{M}f=-\infty$. Интервальную модель можно полагать заданной на всех признаках f, назначая для тех f, для которых $\overline{M}f$ не существует, среднее равным бесконечности: $\overline{M}f=\infty$; аналогично $\underline{M}f=-\infty$ для f, у которых нет нижнего среднего. Аксиомы ИМ в этом случае в общем удовлетворяются, если считать $0\cdot\infty=0$ и учесть, что класс всех функций не замкнут относительно операции сложения, так как если $f(x_0)=\infty$, $g(x_0)=-\infty$ при некотором x_0, то совершенно неясно, чему будет равно $f(x_0)+g(x_0)=\infty-\infty$. Целесообразно аксиому А3 считать выполненной только для тех признаков, для которых сложение определено.

2. Если первичными являются как нижние средние $\underline{M}g$, $g\in\mathscr{G}_{\mathrm{н}}$, так и верхние $\overline{M}g'$, $g'\in\mathscr{G}_{\mathrm{в}}$, то вторичными признаками будут всевозможные конечные линейные комбинации вида

$$g(x)=c+\sum_i c_i^+g'_i(x)-\sum_i d_i^+g_i(x),\quad g'_i\in\mathscr{G}_{\mathrm{в}},\quad g_i\in\mathscr{G}_{\mathrm{н}},$$

с произвольным коэффициентом c и неотрицательными c^+_i и d^+_i. Формула продолжения (1.2) тогда примет вид

$$\overline{M}f = \inf\left\{\left[c + \sum c_i^+ \overline{M}g_i' - \sum d_i^+ \underset{\sim}{M}g_i\right]: \quad c + \sum c_i^+ g_i'(x) - \right.$$
$$\left. - \sum d_i^+ g_i(x) \geqslant f(x)\right\},$$

а (1.3) записывается в виде

$$\overline{M}f = \inf_{c_i^+,\, d_i^+} \sup_x \left[f(x) - \sum c_i^+ \left(g_i'(x) - \overline{M}g_i'\right) + \sum d_i^+ \left(g_i(x) - \underset{\sim}{M}g_i\right)\right],$$

где $g_i \in \mathcal{G}_н$, $g'_i \in \mathcal{G}_в$.

3. Полагая $\mathcal{G} = \mathcal{G}_н \cup \mathcal{G}_в$, всегда можно сделать так, чтобы на $\mathcal{G}$ были заданы интервалы средних $\underset{\sim}{M}g$, $\overline{M}g$, $g \in \mathcal{G}$. Для этого незаданные средние заменяются на соответствующие экстремальные значения функции g. Тогда вторичными будут всевозможные конечные линейные комбинации — линейная оболочка $\mathcal{G}$, обозначаемая $\mathcal{L}\mathcal{G} = \{g(x) = c + \sum c_i g_i(x)\}$, где $g_i \in \mathcal{G}$, а c и c_i произвольны. Формула продолжения средних запишется в виде

$$\overline{M}f = \inf\left\{\left[c + \sum \overline{M}c_i g_i\right]: c + \sum c_i g_i(x) \geqslant f(x)\right\},$$

$$\overline{M}f = \inf_{c_i} \sup_x \left[f(x) - \sum \left(c_i g_i(x) - \overline{M}c_i g_i\right)\right],$$

где $\overline{M}c_i g_i = c_i \overline{M}g_i$ при $c_i > 0$ и $\overline{M}c_i g_i = c_i \underset{\sim}{M}g_i$ при $c_i < 0$.

4. В формуле продолжения класс $\mathcal{L}^+\mathcal{G}$ по свойству **13** вполне может быть заменен на его замыкание $[\mathcal{L}^+\mathcal{G}]$ относительно равномерной сходимости. В этот класс кроме конечных полулинейных комбинаций функций из $\mathcal{G}$ входят их равномерно сходящиеся пределы. Сказанное имеет, конечно же, нетривиальный смысл лишь тогда, когда набор $\mathcal{G}$ бесконечный, иначе $\mathcal{L}^+\mathcal{G}$ и его замыкание совпадают.

5. Формулы продолжения аналогичны формулам двойственности, используемым в теории обобщенных чебышевских неравенств [17]. Только работают они здесь в другой аксиоматике.

6. То что вторичные признаки являются конечными суммами первичных, а не бесконечными, — это принципиальное в нашей теории «нежелание» добавлять «лишних» аксиом, ибо не доказуемо, что сумма бесконечного (счетного) числа нулей, рассматриваемая в целом, а не как предел, есть ноль.

7. Каждый признак $f(x)$ сам по себе есть с. в., поэтому ИМ $\langle \overline{M}\mathcal{F} \rangle$ можно рассматривать как совокупность согласованных средних значений всевозможных с. в., определенных на пространстве $\mathcal{X}$. Набор признаков g_τ, $\tau \in T$, в совокупности дает случайный вектор, если T дискретно, или процесс, если непрерывно.

8. Тот факт, что ИМ определяется средними $\overline{M}f$ необозримого множества $\mathcal{F}$ признаков, не является помехой применения. Возможность вычисления $\overline{M}f$ по (1.2) отнюдь не означает, что для каждой f эти вычисления должны быть проделаны. Совсем наоборот, как далее будет видно, нет нужды в прикладных задачах выходить за рамки вторичных признаков.

9. Введение любых свойств непрерывности средних, не следующих прямо из аксиоматики, выделяет подкласс моделей из общего их ансамбля, равно как другие дополнительные свойства.

10. Начальные моменты. Задают с. в., если первичными для нее являются средние степенных функций: $\underset{\sim}{M}X^j$, $\overline{M}X^j$, $j \in J$, где J — набор целочисленных неотрицательных индексов. Средние $\underset{\sim}{M}X^j$ называются нижними моментами j-го порядка, а $\overline{M}X^j$ — верхними. Первый момент $\underset{\sim}{M}X$, $\overline{M}X$ называется нижним и верхним средним самой с. в., а $\underset{\sim}{M}X^2$, $\overline{M}X^2$ нижней и верхней мощностью.

Абсолютными начальными моментами называются моменты модуля с. в. $\underset{\sim}{M}|X|^j$, $\overline{M}|X|^j$, $j > 0$. Очевидно, при целых четных неотрицательных j это есть просто начальные моменты. В общем, j может не быть целым. Для абсолютных моментов, если они согласованы, должны выполняться неравенства: при $r \geqslant s \geqslant 0$

$$(\overline{M}|X|^r)^{1/r} \geqslant (\overline{M}|X|^s)^{1/s}, \quad (\underset{\sim}{M}|X|^r)^{1/r} \geqslant (\underset{\sim}{M}|X|^s)^{1/s}$$

(доказывается по аналогии с [1], стр. 169).

1.3. ОТНОШЕНИЯ МЕЖДУ ИНТЕРВАЛЬНЫМИ МОДЕЛЯМИ

Здесь интервальные модели геометрически изображаются как некоторые выпуклые «тела» с характерным внешним контурным описанием, отношениями включения и операциями объединения и пересечения. Причем, так как ИМ полностью определяется своими средними, через них только и будем вводить ниже формальные отношения и операции между ИМ.

Геометрическая иллюстрация ИМ. Пусть возможных элементарных исходов конечное число: $\mathcal{X} = \{x_1, \ldots, x_r\}$, и введем вектор вероятностей $\mathbf{P} = (p_1, \ldots, p_r)$. Так как $\sum p_i = 1$, то размерность $\mathbf{P}$ на 1 меньше числа исходов r. Множество всех $\mathbf{P}$ обозначим

$$\mathcal{Y} = \left\{\mathbf{P}: p_i \geqslant 0, \ \sum_{i=1}^r p_i = 1\right\}.$$

Это есть подмножество r-мерного эвклидова пространства $\mathcal{R}^r$. На рис. 1.4 при $r = 3$ этим семейством является треугольник.

Для каждого фиксированного $\mathbf{P} \in \mathcal{Y}$ среднее значение признака $f(x)$, он же вектор $f = (f(x_1), \ldots, f(x_r))$, равно скалярному произведению f на $\mathbf{P}$:

$$M_{\mathbf{P}} f = \sum_{i=1}^r f(x_i)\, p_i.$$

Пусть $\mathcal{M}_0$ — семейство векторов, $\mathcal{M}_0 \subset \mathcal{Y}$. Тогда средние не будут уже точными, а для каждого f будут определяться своими границами

$$\underline{M}f = \inf_{\mathbf{P} \in \mathcal{M}_0} M_{\mathbf{P}} f, \quad \overline{M}f = \sup_{\mathbf{P} \in \mathcal{M}_0} M_{\mathbf{P}} f.$$

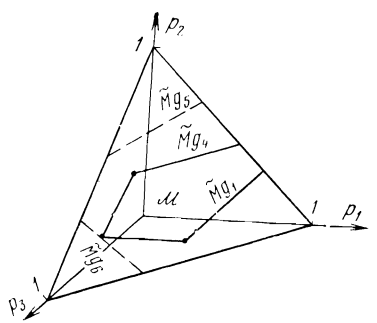


Рис. 1.4. Геометрия ИМ

Убедимся в том, что так определенные нижние и верхние средние удовлетворяют аксиомам ИМ:

A1: $g(x_i) \geqslant f(x_i) \Rightarrow M_{\mathbf{P}} g \geqslant M_{\mathbf{P}} f, \quad \forall \mathbf{P} \Rightarrow \overline{M} g \geqslant \overline{M} f.$

A2: $\overline{M}(b^+ f + c) = \sup\limits_{\mathbf{P} \in \mathscr{M}_0} b^+ M_{\mathbf{P}} f + c = b^+ \overline{M} f + c.$

A3: $\overline{M}(f+g) = \sup\limits_{\mathbf{P} \in \mathscr{M}_0} (M_{\mathbf{P}} f + M_{\mathbf{P}} g) \leqslant \sup\limits_{\mathbf{P} \in \mathscr{M}_0} M_{\mathbf{P}} f + \sup\limits_{\mathbf{P} \in \mathscr{M}_0} M_{\mathbf{P}} g =$

$= \overline{M} f + \overline{M} g.$

A4: $-\underline{M} f = -\inf\limits_{\mathbf{P} \in \mathscr{M}_0} M_{\mathbf{P}} f = \sup\limits_{\mathbf{P} \in \mathscr{M}_0} M_{\mathbf{P}} (-f) = \overline{M}(-f).$

Средние не изменятся, если семейство $\mathscr{M}_0$ заменить на его выпуклую оболочку — «тело» векторов **P**. Это и будет ИМ $\mathscr{M}$.

Таким образом, любое выпуклое «тело» $\mathscr{M}$ векторов вероятностей образует ИМ на дискретном пространстве $\mathscr{X}$. И наоборот, любой ИМ соответствует выпуклое «тело» $\mathscr{X}$. Как оно получается?

С каждым средним $\overline{M} f$ связывается полупространство векторов **P**, удовлетворяющее условию $M_{\mathbf{P}} f \leqslant \overline{M} f$, т. е. окаймленное с одной стороны гиперплоскостью **P**: $M_{\mathbf{P}} f = \overline{M} f$. Вид f дает направление гиперплоскости, а значение $\overline{M} f$ — ее положение. Модель как совокупность $\overline{M} f$, $f \in \mathscr{F}$, есть пересечение соответствующих разным $\overline{M} f$ полупространств, образующих выпуклое тело $\mathscr{M}$. Сами значения средних $\overline{M} f$ дают положения касающихся $\mathscr{M}$ гиперплоскостей.

Первичные средние $\overline{M} g_i$, $i = 1, \ldots, k$, определяют исходные первичные гиперплоскости **P**: $M_{\mathbf{P}} g_i = \overline{M} g_i$, задающие внешний вид $\mathscr{M}$. Если их число конечно, то ИМ будет многогранником $\mathscr{M}_k$ (как это видно из рис. 1.4), грани которого составляют согласованные значения $\overline{M} g_i = \overline{M} g_i$. Несогласованные же, такие как Mg_5, проходят вне $\mathscr{M}$ и никакого влияния не оказывают. Есть гиперплоскости **P**: $M_{\mathbf{P}} g = \overline{M} g$, которые проходят через вершины многогранника $\mathscr{M}$, но не совпадают ни с одной из его граней, как $\overline{M} g_6$. Хотя они и согласованы, но их исключение также не повлияет на вид $\mathscr{M}$, и в этом смысле они избыточны. Удаление всех избыточных элементов из первичного набора наглядно представляется лишь при конечном числе первичных признаков, и может стать непреодолимым при бесконечном. Сказанное относится не только к дискретным, но и к произвольным $\mathscr{X}$.

Число элементов безызбыточного первичного набора составляет *размерность* ИМ. Интервальные модели бывают конечной и бесконечной размерностей. В первом случае — это число граней многогранника $\mathscr{M}$, за исключением тривиальных, совпадающих с гранями $\mathscr{I}$.

Размерность ИМ характеризует тот минимальный объем данных, который нужен для ее задания. Конечно же, для практического применения целесообразными являются именно наборы из

конечного числа данных. Интервальные модели бесконечной размерности — это выпуклые тела, контуры которых задаются бесконечным числом касательных гиперплоскостей.

Обсуждение. Теперь следует остановиться, чтобы обсудить смысл принятого подхода. Что лучше, описывать ИМ «телом» $\mathscr{M}$ как множеством векторов **P** вероятностей, своего рода «атомов» модели, или только внешними контурами, т. е. $\overline{M} g$, $g \in \mathscr{G}$? Если $\mathscr{X}$ состоит из небольшого числа элементов, то, по-видимому, и так, и так, принципиального различия нет. Но если число элементов $\mathscr{X}$ растет, то также растет и размерность **P**, все мельче становятся атомы модели. Описание **P** несоразмерно усложняется при переходе к бесконечному числу исходов, в частности, к непрерывному пространству $\mathscr{X}$, например числовой прямой (и еще более — к процессам X_t). Тогда «атомы» и вовсе исчезают: становятся недосягаемыми. Для задания **P** приходится прибегать ко всевозможным ухищрениям, обсуждаемым в следующем параграфе. В то же время контурное описание, т. е. описание первичными параметрами, от «атомов» **P** никак не зависит, да и определяется, минуя **P**. Здесь все зависит от числа граней (размерности ИМ) и их положения (вида $g_i \in \mathscr{G}$), но отнюдь не от пространства $\mathscr{X}$. Почти как в геометрии: неважно, чем начинено тело, какой оно материи, а важно лишь внешнее его строение и пропорции.

Иерархия моделей. Для двух ИМ $\mathscr{M}_1$ и $\mathscr{M}_2$ на $\mathscr{X}$, если $\overline{M}_1 f \leqslant \overline{M}_2 f$, $\forall f$, то будем говорить, что $\mathscr{M}_1$ *включается* в $\mathscr{M}_2$ и писать $\mathscr{M}_1 \subset \mathscr{M}_2$. В указанном неравенстве, если f не ограничена сверху и $\overline{M} f$ для нее не существует, то формально считаем его бесконечным. Необходимым условием включения $\mathscr{M}_1 \subset \mathscr{M}_2$ является то, что область существования $\mathscr{F}_1$ средних $\overline{M}_1 f$ должна быть не уже области $\mathscr{F}_2$ существования $\overline{M}_2 f$: $\mathscr{F}_1 \supset \mathscr{F}_2$. Включение иллюстрируется рис. 1.5 как включение «тела» $\mathscr{M}_1$ в $\mathscr{M}_2$. Будем говорить также, что $\mathscr{M}_1$ более узкая, чем $\mathscr{M}_2$, а $\mathscr{M}_2$ — более широкая.

Добавление первичных параметров отсекает новые грани у «фигуры» $\mathscr{M}$ и приводит к ее сужению. К сужению приводит и уточнение средних: для $\overline{M} f$ — это уменьшение, а для $\underline{M} f$ — увеличение, поэтому включение $\mathscr{M}_1 \subset \mathscr{M}_2$ соответствует тому, что в $\mathscr{M}_1$ вложено больше данных или же они более точные. В результате $\mathscr{M}_1$ более подробная, чем $\mathscr{M}_2$.

Самой широкой среди всех является *голая* модель (на рис. 1.4 — это множество $\mathscr{I}$ всех векторов вероятностей), определяемая средними $\overline{M} f = \sup f$, $f \in \mathscr{F}_0$, на всех ограниченных сверху признаках и обозначаемая $\mathscr{I}$. Она получается, если никаких данных — «одежд», отличающих одну ИМ от другой, нет, они отсутствуют: $\mathscr{G} \equiv 0$, т. е. о явлении ничего неизвестно — своего рода «черный ящик» с абсолютно загадочной структурой, выход которого x наблюдается. Для $\mathscr{I}$ интервал средних $\underline{M} f$, $\overline{M} f$ каждого ограниченного признака $f \in \mathscr{F}_{00}$ совпадает с диапазоном $\inf f$, $\sup f$ его возможных значений.

Итак, самой неточной среди всех, включающей все остальные, и самой примитивной по способу задания и своей структуре является голая ИМ: $\mathscr{I} \supset \mathscr{M}$, $\forall \mathscr{M}$. Мы говорим, что какое-то x из $\mathscr{X}$ произойдет, не зная никаких закономерностей.

По мере накопления данных ИМ сужается. А существуют ли самые узкие ИМ? Для дискретного пространства $\mathscr{X}$ — да, это векторы **P** вероятностей. А в общем, ответ отрицательный, о чем говорилось в обсуждении выше.

Потребность логической замкнутости иерархического класса всех моделей приводит к необходимости введения *пустой модели* $\varnothing$. Это обозначение неправильного задания ИМ, когда первичные значения $\bar{M}g$ противоречивы, в результате границы среднего, формально найденные по формуле продолжения (1.2), будут не только «перепутаны»: нижняя больше верхней, но и убегают в бесконечности: $\underline{M}f = \infty$, $\bar{M}f = -\infty$. Этими средними на $\forall f$ и считаем определенную $\varnothing$ — единственную ИМ, своей несогласованностью выходящую из ансамбля всех остальных. Так как для любой ИМ $\underline{M}f \geqslant -\infty$, $\bar{M}f \leqslant \infty$, то $\varnothing \subset \mathscr{M}$, $\forall \mathscr{M}$.

Пусть $\mathscr{Q}$ — набор признаков. Назовем *$\mathscr{Q}$-расширением* $\mathscr{M}$ такую модель $\langle \bar{M}\mathscr{Q} \rangle$, у которой первичными являются соответствующие $\mathscr{M}$ значения $\bar{M}q$, $q \in \mathscr{Q}$ (они, очевидно, согласованы). Расширение иллюстрируется рис. 1.5. Это способ упрощения $\mathscr{M}$ за счет включения ее в многогранник со сторонами $\bar{M}q$, $q \in \mathscr{Q}$ — новыми первичными данными — и пренебрежения всеми остальными знаниями о $\mathscr{M}$. Конечно, чтобы не потерять при расширении слишком много, нужно специально выбирать направления граней (вид q), и целая проблема, какие грани экономно будет ввести, какие оставить и сколько их.

Набор $\mathscr{Q}$ признаков называется *определяющим* для модели $\mathscr{M}$, если ее $\mathscr{Q}$-расширение совпадает с самой моделью: $\langle \bar{M}\mathscr{Q} \rangle = \mathscr{M}$. Определяющие — это те признаки, которые в своей совокупности полностью обеспечивают ИМ всем необходимым. Ясно, что определяющим для ИМ всегда является набор $\mathscr{G}$ первичных признаков (или безызбыточный вариант этого набора), а тем более любой включающий $\mathscr{G}$ класс признаков, например вторичных.

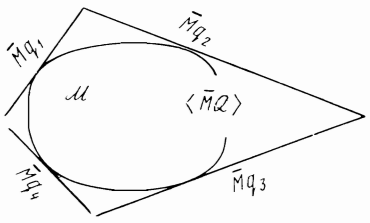


Рис. 1.5. Расширение ИМ бесконечной размерности до ИМ с четырьмя первичными значениями

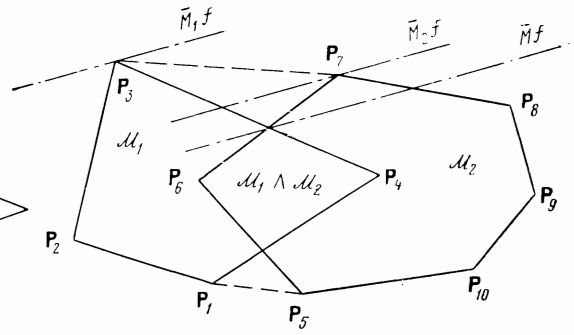


Рис. 1.6. Пересечение и объединение ИМ

Пересечение ИМ. Пусть $\mathscr{M}_1$ и $\mathscr{M}_2$ — две ИМ на одном и том же (произвольном) пространстве $\mathscr{X}$ и каждая из них полностью определяется своими согласованными средними $\bar{M}_1 f$, $f \in \mathscr{F}_1$ и $\bar{M}_2 f$, $f \in \mathscr{F}_2$, на областях $\mathscr{F}_1$ и $\mathscr{F}_2$ соответственно.

Пересечением $\mathscr{M}_1$ и $\mathscr{M}_2$ называется ИМ $\mathscr{M} = \mathscr{M}_1 \wedge \mathscr{M}_2$, определяемая средними $\bar{M}f = \min\{\bar{M}_1 f, \bar{M}_2 f\}$, $\forall f \in \mathscr{F}_1 \cup \mathscr{F}_2$. Здесь волнистая черта означает, что $\bar{M}f$, во-первых, могут оказаться несогласованными между собой, так как не выполняется аксиома **А3** полуаддитивности, и тогда должны подвергнуться согласованию. Это видно из рис. 1.6, где штриховой линией означены касательные, соответствующие средним $\bar{M}_1 f$, $\bar{M}_2 f$ и $\bar{M}f$, $\bar{M}f < \tilde{M}f = \min\{\bar{M}_1 f, \bar{M}_2 f\}$. А во-вторых, $\tilde{M}f$ продолжаются по формуле (1.3) на класс $\mathscr{L}^+(\mathscr{F}_1 \cup \mathscr{F}_2)$ признаков, составленных из полулинейных комбинаций $\mathscr{F}_1$ вместе с $\mathscr{F}_2$.

Пересечение означает, что верными являются данные, содержащиеся как в $\mathscr{M}_1$, так и в $\mathscr{M}_2$, и из них берутся наиболее точные. В результате сужаются как интервальные средние, так и ИМ: $\mathscr{M}_1 \wedge \mathscr{M}_2 \subset \mathscr{M}_1$, $\mathscr{M}_1 \wedge \mathscr{M}_2 \subset \mathscr{M}_2$ (см. рис. 1.6).

Первичными средними (гранями) пересечения будут первичные средние (грани) $\mathscr{M}_1$ и $\mathscr{M}_2$ и никакие другие, поэтому пересечение $\langle \tilde{M}_1 \mathscr{G}_1 \rangle \wedge \langle \tilde{M}_2 \mathscr{G}_2 \rangle = \langle \tilde{M}(\mathscr{G}_1 \cup \mathscr{G}_2) \rangle$ равносильно сложению двух первичных наборов между собой в единый $\mathscr{G}_1 \cup \mathscr{G}_2$ с сохранением первичных значений, из которых некоторые оказываются несогласованными и не влияют на вид пересечения. Областью существования пересечения будет $\mathscr{L}^+(\mathscr{F}_{\mathscr{G}_1} \cup \mathscr{F}_{\mathscr{G}_2})$.

Операция пересечения распространяется на произвольное число $\mathscr{M}_\theta = \langle \bar{M}_\theta \mathscr{F}_\theta \rangle$, $\theta \in \Theta$:

$$\mathscr{M} = \bigwedge_{\theta \in \Theta} \mathscr{M}_\theta \Leftrightarrow \tilde{M}f = \inf_{\theta \in \Theta} \bar{M}_\theta f, \quad \forall f \subset \mathscr{L}^+\left(\bigcup_\theta \mathscr{F}_\theta\right).$$

Используем этот факт.

Пример 1.7. Представление ИМ пересечением. Любую ИМ $\langle \tilde{M}\mathscr{G} \rangle$ можно представить в виде пересечения

$$\langle \tilde{M}\mathscr{G} \rangle = \bigwedge_{g \in \mathscr{G}} \langle \tilde{M}g \rangle$$

моделей $\langle \tilde{M}g \rangle$ размерности 1, определенные каждая одним первичным значением $\tilde{M}g$ признака g, заменяющего индекс θ. Тогда $\mathscr{F}_g = \mathscr{F}_\theta$ составляют признаки, мажорируемые g, а $\mathscr{L}^+(\cup \mathscr{F}_g) = \mathscr{F}_{\mathscr{G}}$ — их полулинейные комбинации.

Объединение ИМ. Пусть $\mathscr{M}_\theta = \langle \bar{M}\mathscr{F}_\theta \rangle$, $\theta \in \Theta$ — семейство ИМ на $\mathscr{X}$, индексированных параметров θ, пробегающим множество Θ. Определим их объединение (обозначается $\bigvee$) следующим образом:

$$\mathscr{M} = \bigvee_{\theta \in \Theta} \mathscr{M}_\theta \Leftrightarrow \overline{M}f = \sup_{\theta \in \Theta} \bar{M}_\theta f, \quad \forall f \in \bigcap_\theta \mathscr{F}_0.$$

Здесь средние $\overline{M}f$, полученные максимизацией $\overline{M}_\theta f$ по θ, согласованы, что легко проверяется и что иллюстрируется рис. 1.6, где объединение обведено штриховкой и соответствует выпуклой оболочке тел $\mathscr{M}_1$ и $\mathscr{M}_2$. Как видно из рис. 1.6, при объединении рождаются новые грани (первичные признаки), обозначенные штриховыми линиями, не совпадающими с гранями $\mathscr{M}_1$ и $\mathscr{M}_2$, изображенными сплошными. При этом грани объединения не будут выходить за рамки линейной оболочки граней составляющих, т. е., если быть более строгими, при $\mathscr{M}_1=\langle\overline{M}_1\mathscr{G}_1\rangle$ и $\mathscr{M}_2=\langle\overline{M}_2\mathscr{G}_2\rangle$ первичные признаки объединения будут располагаться в классе $\mathscr{L}^+(\mathscr{G}_1\cup\mathscr{G}_2)$.

Операция объединения $\mathscr{M}_1\vee\mathscr{M}_2$ — символическое отражение фразы: «Верна (правильно отражает явление) модель $\mathscr{M}_1$ или $\mathscr{M}_2$». Это сомнение, неуверенность, ведущая к расширению ИМ.

Пример 1.8. Представление ИМ объединением «вершин». Пусть пространство исходов $\mathscr{X}$ дискретно. Каждой ИМ конечной размерности можно указать вершины $\mathbf{P}_i$ — векторы вероятностей. На рис. 1.6 для $\mathscr{M}_1$ — это $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$, $\mathbf{P}_4$ и $\mathscr{M}_1$ является их оболочкой. При объединении двух ИМ вершины будут выбираться из вершин $\mathscr{M}_1$ и $\mathscr{M}_2$. Никакие другие появиться не могут. Для ИМ любой размерности, когда $\mathscr{M}_1=\bigvee\limits_\theta\mathbf{P}_{1\theta}$ есть оболочка некоторого семейства векторов $\mathbf{P}_{1\theta}$, задающих тело $\mathscr{M}_1$ (или только его контуры), и аналогично $\mathscr{M}_2=\bigvee\limits_\vartheta\mathbf{P}_{2\vartheta}$, верно $\mathscr{M}_1\vee\mathscr{M}_2=\bigvee\limits_\theta\mathbf{P}_{1\theta}\vee\bigvee\limits_\vartheta\mathbf{P}_{2\vartheta}$, т. е. производится объединение этих семейств в одно. Причем достаточно ограничиться теми элементами семейств, которые при объединении оказываются крайними (не входят в выпуклые оболочки других).

Свойства операций.

1. Идемпотентность: $\mathscr{M}\wedge\mathscr{M}=\mathscr{M}$, $\mathscr{M}\vee\mathscr{M}=\mathscr{M}$.
2. Коммутативность: $\mathscr{M}_1\wedge\mathscr{M}_2=\mathscr{M}_2\wedge\mathscr{M}_1$, $\mathscr{M}_1\vee\mathscr{M}_2=\mathscr{M}_2\vee\mathscr{M}_1$.
3. Ассоциативность: $(\mathscr{M}_1\wedge\mathscr{M}_2)\wedge\mathscr{M}_3=\mathscr{M}_1\wedge(\mathscr{M}_2\wedge\mathscr{M}_3)$, $(\mathscr{M}_1\vee\mathscr{M}_2)\vee\mathscr{M}_3=\mathscr{M}_1\vee(\mathscr{M}_2\vee\mathscr{M}_3)$.
4. $\mathscr{M}\wedge\mathscr{I}=\mathscr{M}$, $\mathscr{M}\vee\mathscr{I}=\mathscr{I}$.
5. $\mathscr{M}\wedge\varnothing=\varnothing$, $\mathscr{M}\vee\varnothing=\mathscr{M}$.

Эти свойства доказываются элементарно и распространяются на любое число операций. Очевидно также, что

$$\mathscr{M}_1\subset\mathscr{M}_2\Rightarrow\mathscr{M}_1\wedge\mathscr{M}_2=\mathscr{M}_1,\ \mathscr{M}_1\vee\mathscr{M}_2=\mathscr{M}_2.$$

Отсюда получается сразу же такое известное в алгебре свойство.

6. Закон поглощения:

$(\mathscr{M}_1\wedge\mathscr{M}_2)\vee\mathscr{M}_1=\mathscr{M}_1$, $(\mathscr{M}_1\vee\mathscr{M}_2)\wedge\mathscr{M}_1=\mathscr{M}_1$

(так как $\mathscr{M}_1\wedge\mathscr{M}_2\subset\mathscr{M}_1$, $\mathscr{M}_1\vee\mathscr{M}_2\supset\mathscr{M}_1$). Это есть замена привычного для булевских алгебр свойства дистрибутивности, которое для наших операций не выполняется. Причина в том, что объединение $\vee$ ИМ не обычное множественное, всем привычное, а выпуклое, так как в результате должны образоваться снова ИМ, которые по своей природе выпуклы. И, как следствие, невозможность определить дополнение к ИМ (или противоположную ИМ), так как к выпуклому телу дополнение не будет выпуклым.

Дополнения.

1. Для дискретных пространств $\mathscr{X}$ формула (1.2) вытекает из формулы двойственности линейного программирования, когда по ограничениям

$$\sum_{i=1}^{r}p_i=1,\quad\sum_{i=1}^{r}g_{ji}\,p_i\leqslant\widehat{M}g_j,\quad j=1,\ldots,k,$$

находится максимум линейных форм $\overline{M}f=\max\limits_{\mathbf{P}}\sum\limits_{i=1}^{r}f_ip_i$.

2. Операции над моделями могут быть определены, если ИМ заданы на разных пространствах исходов: $\mathscr{M}_1$ — на $\mathscr{X}_1$, а $\mathscr{M}_2$ — на $\mathscr{X}_2$. Тогда составляется их объединение $\mathscr{X}=\mathscr{X}_1\cup\mathscr{X}_2$ и $\mathscr{M}_1$ дополняется первичным значением $\mathbf{P}_1(\mathscr{X}_1)=1$, а $\mathscr{M}_2$ — значением $\mathbf{P}_2(\mathscr{X}_2)=1$. В результате $\mathscr{M}_1$ и $\mathscr{M}_2$ сводятся к одному $\mathscr{X}$ со всеми вытекающими отсюда возможностями.

3. Если говорится, что явление с исходами $\mathscr{X}$ описывается семейством моделей $\mathscr{M}_\theta$, $\theta\in\Theta$, то это фактически означает, что моделью является объединение $\bigvee\limits_\theta\mathscr{M}_\theta$.

4. Содержащееся в примере 1.8 представление ИМ как объединение вершин не является универсальным в силу невозможности для произвольных пространств $\mathscr{X}$ отделения «атомов» $\mathbf{P}$ модели.

1.4. ИНТЕРВАЛЬНЫЕ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Свойства интервальных вероятностей. Исторически сложилось, что в основу описания случайных явлений были положены вероятности, чем мы обязаны наглядности игровых примеров (монета, карты, кость), давших теории вероятностей начальный толчок и подпитывающих ее на протяжении развития. Эта же наглядность, а в дальнейшем отработанность и стройность теории привела и к фактическому игнорированию других подходов[1].

Наши интервальные модели в своем определении базируются на интервальных средних, а вероятностям — нижней $\underline{P}(A)$ и верхней $\overline{P}(A)$ отведена роль частного случая, когда признаками являются индикаторные функции $A(x)$ событий: $\underline{P}(A)=\underline{M}A(x)$; $\overline{P}(A)=\overline{M}A(x)$, $A\subset\mathscr{X}$. Для любой ИМ вероятности $\underline{P}(A)$, $\overline{P}(A)$ определены для $\forall A$ (так как $A(x)\in\mathscr{F}_{00}$).

Свойства вероятностей непосредственно вытекают из согласованности средних (см. § 1.1). Обозначая знаком плюс $A+B$ и символом сумма ΣA_i объединение непересекающихся событий (в отличие от общего обозначения объединения $\cup$), имеем:

1. $\underline{P}(\mathscr{X})=\overline{P}(\mathscr{X})=P(\mathscr{X})=1$ — пространство $\mathscr{X}$ является всегда достоверным событием (доказывается $\underline{M}1=P(\mathscr{X})=1$).

2. Обращение вероятностей: $\underline{P}(A)=1-\overline{P}(A^c)$, где A^c — дополнение к событию A (так как $\underline{MA}(x)=1-\overline{M}(1-A(x))$.

[1] В [2] в основу теории положены точные средние, но они наделяются столь жесткими свойствами, что это полностью свело получаемые в результате модели к вероятностным.

3. Верхняя полуаддитивность: $AB=\varnothing \Rightarrow \bar{P}(A+B) \leqslant \bar{P}(A)+\bar{P}(B)$ (так как $\overline{M}[A(x)+B(x)] \leqslant \overline{M}A(x)+\overline{M}B(x)$).

4. Нижняя полуаддитивность: $AB=\varnothing \Rightarrow \underline{P}(A+B) \geqslant \underline{P}(A)+\underline{P}(B)$ (так как $\underline{M}[A(x)+B(x)] \geqslant \underline{M}A(x)+\underline{M}B(x)$).

Вероятности, удовлетворяющие этим свойствам, называются *согласованными*. Смысл слова «согласованность» в том, что если бы какое-нибудь свойство не выполнялось, то хотя бы одна какая-то граница вероятности могла бы быть уточнена за счет других, т. е. $\underline{P}(A)$ увеличена или $\bar{P}(A)$ уменьшена.

Из свойств согласованности вероятностей непосредственно вытекают следующие свойства:

5. $P(\varnothing)=0$ — вероятность пустого события нуль.

6. $\underline{P}(A) \leqslant \bar{P}(A)$ — нижняя вероятность всегда не больше верхней.

7. $AB=\varnothing \Rightarrow \underline{P}(A+B) \leqslant \underline{P}(A)+\bar{P}(B) \leqslant \bar{P}(A+B)$ (частный случай соответствующего свойства средних, который может быть доказан и на основе **1**—**4**).

8. Для конечного числа попарно непересекающихся событий $\bar{P}(\sum_1^k A_i) \leqslant \sum_1^k \bar{P}(A_i)$ (следует по индукции из **3**).

9. Для конечного или счетного числа попарно непересекающихся событий $\underline{P}(\sum A_i) \geqslant \sum \underline{P}(A_i)$. Для конечного числа A_i это свойство непосредственно следует из **4**, а для счетного — из

$$\underline{P}\left(\sum_1^\infty A_i\right)=\underline{P}\left(\sum_1^k A_i+\sum_{k+1}^\infty A_i\right) \geqslant \underline{P}\left(\sum_1^k A_i\right)+\underline{P}\left(\sum_{k+1}^\infty A_i\right) \geqslant$$

$$\geqslant \sum_1^k \underline{P}(A_i)$$

предельным переходом в правой части при $k \to \infty$.

Свойство **8** развивает **3** и называется *конечной верхней полуаддитивностью* вероятностей. Свойство же **9** продолжает **4** не только на конечные, но и на счетные суммы и называется *нижней счетной полуаддитивностью*. Это более сильное свойство, поэтому нижняя вероятность по природе «более непрерывна», чем верхняя. Для верхней же свойство *счетной верхней полуаддитивности* верно лишь при дополнительном условии, составляющем левую часть следующего утверждения:

10. $\lim\limits_{k\to\infty} \bar{P}\left(\sum_{k+1}^\infty A_i\right)=0 \Rightarrow \bar{P}\left(\sum_1^\infty A_i\right) \leqslant \sum_1^\infty \bar{P}(A_i)$.

Для доказательства надо перейти к пределу при $k \to \infty$ в неравенстве

$$\bar{P}\left(\sum_1^k A_i+\sum_{k+1}^\infty A_i\right) \geqslant \sum_1^k \bar{P}(A_i)+\bar{P}\left(\sum_{k+1}^\infty A_i\right).$$

Пусть B_n, монотонно возрастая при $n \to \infty$, сходятся к B. Это записывается $B_n \uparrow B$ и означает, что $B_n \subset B_{n+1}$, $\forall n$, и для каждой

точки $x \in B$ с ростом n найдется такое B_n, которое все же x накроет. И тем не менее даже столь жестких требований сходимости событий недостаточно, чтобы гарантировать сходимость их вероятностей, что провозглашается следующим тезисом.

11. Вероятности в общем не являются непрерывными по отношению к монотонной сходимости событий. Это негативное свойство означает, что, имея $\underline{P}(B_n)$ и $\bar{P}(B_n)$ и зная, что $B_n \uparrow B$, тем не менее можно $\underline{P}(B)$ и $\bar{P}(B)$ брать отличными от пределов $\lim\limits_{n\to\infty} \underline{P}(B_n)$ и $\lim\limits_{n\to\infty} \bar{P}(B_n)$, не нарушив при этом свойства согласованности вероятностей. Покажем это на примере.

Пример 1.9. Пусть $B_n \uparrow B$, $B_n \neq B$ и первичными являются $\bar{P}(B_n)=p_1$, $\bar{P}(B)=p_2>p_1$. Тогда $\lim\limits_{n\to\infty} \bar{P}(B_n)=p_1<p_2$. Здесь согласованность вероятностей не мешает задать p_2 отличным от p_1. Если же дополнить указанный набор еще одной первичной нижней вероятностью $\underline{P}(B)=p$, $0<p \leqslant p_2$, то так как ни в одно из B_n событие B не вкладывается, то $\underline{P}(B_n)=0$, и следовательно, $\lim\limits_{n\to\infty} \underline{P}(B_n)=0$, тогда как $\underline{P}(B)=p>0$.

Продолжение первичных вероятностей. *Интервальным распределением вероятностей* (сокращенно ИРВ) называется ИМ, первичными для которой являются вероятности (точные или интервальные) набора событий. Первичными признаками ИРВ являются индикаторные функции $A(x)$ событий A первичного набора $\mathscr{A}$, и на них заданы вероятности в виде точных значений $P(A)$, интервалов $\underline{P}(A)$, $\bar{P}(A)$ или одной из границ, чаще верхней. Если нет нижней, то всегда можно положить ее равной 0, а не заданную $\bar{P}(A)$ считать равной 1. Это ничего не изменит, кроме того, что на всех событиях из $\mathscr{A}$ будут определены интервальные первичные вероятности, позволяя обозначение $\langle \underset{\sim}{P}(\mathscr{A}), \bar{P}(\mathscr{A}) \rangle$. Если же все первичные вероятности приведены к верхним, ИРВ обозначается $\langle \bar{P}(\mathscr{A}) \rangle$.

Рассмотрим, как продолжить вероятности, перенеся их на средние любых признаков. Поскольку первичными признаками ИРВ являются события из $\mathscr{A}$, то вторичными будут всевозможные их конечные линейные комбинации: $\mathscr{L}\mathscr{A}=\{g(x)=c+\sum c_i A_i(x), A_i \in \mathscr{A}\}$, (вторичные, они же $\mathscr{A}$-измеримые, функции), где c и c_i — произвольные коэффициенты, и первичные значения переносятся на вторичные признаки по (1.1):

$$\tilde{M}g=c+\sum \tilde{M}c_i A_i(x), \text{ где } \tilde{M}c_i A_i=\begin{cases} c_i \bar{P}(A_i) & \text{при } c_i>0, \\ c_i \underset{\sim}{P}(A_i) & \text{при } c_i<0. \end{cases}$$

Это первый шаг, хотя и не однозначный, так как одно g может по-разному записываться через A_i (тогда берется минимальное из $\tilde{M}g$ среди всевозможных записей).

Следующий шаг состоит в продолжении этих средних с их согласованием по (1.2) на любые ограниченные функции $f(x)$. Но

это будет возможно только, если первичные вероятности непротиворечивы: $g\geqslant 0\Rightarrow \tilde{M}g\geqslant 0$, $\forall g\in\mathscr{L}\mathscr{A}$. Тогда формула

$$\overline{M}f=\inf_{f(x)\leqslant g(x)\in\mathscr{L}\mathscr{A}}\tilde{M}g,\quad \forall f\in\mathscr{F}_0$$

дает согласованные значения средних на любых ограниченных сверху признаках. Класс $\mathscr{F}_0$ последних $\mathscr{A}$-мажорируем и потому будет естественной областью продолжения верхних средних всех ИРВ: $\mathscr{F}_{\mathscr{A}}=\mathscr{F}_0$. В частности, будут определены верхние вероятности $\overline{P}(B)$ (и через них нижние по свойству **2**) для всех событий $\forall B\subset\mathscr{X}$, другое дело, что они могут оказаться для многих событий тривиальными, т. е. равными 1.

Желание распространить средние на неограниченные признаки (такие как x, x^k, $\operatorname{tg} x$ и т. д.) по подобию математических ожиданий дает оправдание третьему шагу, к которому и перейдем.

Предельное продолжение средних. Действуем строго по аналогии с интегрированием, помня, что интегрирование по вероятностной мере ведет к математическим ожиданиям. В теории интегрирования: а) первичными даются меры множеств, б) их значения присваиваются интегралам от индикаторных функций, в) далее эти интегралы распространяются по аддитивности на интегралы от всех простых функций (сумм индикаторов), г) затем продолжаются на интегралы от измеримых ограниченных функций, д) последние, наконец, переносятся на неограниченные. Последний шаг применительно к ИМ и составляет предмет нашего рассмотрения.

Усечем неограниченную функцию f снизу уровнем $-H_1$ и сверху уровнем H_2, обозначив

$$f^{(-H_1,\,H_2)}(x)=\begin{cases}-H_1, & f(x)<-H_1,\\ f(x), & -H_1\leqslant f(x)\leqslant H_2,\\ H_2 & f(x)>H_2.\end{cases}$$

Функции $f^{-H_1,\,H_2}(x)$ ограничены и поэтому средние для них всегда определены. Положим для неограниченной функции

$$\overline{M}f=\lim_{H_1\to\infty}\lim_{H_2\to\infty}\overline{M}f^{(-H_1,\,H_2)}. \tag{1.4}$$

Собственно, так интуитивно и понимается всегда неограниченная функция как предел ее усеченного варианта при устремлении к бесконечности уровней усечения; естественно, в этом ключе следует понимать и средние. Важно, что сначала H_2 устремляется к ∞, так как в силу монотонности это дает наибольшее значение правой части, а затем уже H_1.

Для пределов (1.4) выполняются все аксиомы. Из них **А1—А3** доказываются с помощью неравенств:

1) $f\geqslant g\Rightarrow f^{(-H_1,\,H_2)}\geqslant g^{(-H_1,\,H_2)}$.

2) $b+f^{(-H_1,\,H_2)}+c=(b+f+c)^{(-b+H_1+c;\ b+H_2+c)}$;

3) $(f+g)^{(-H_1,\,H_2)}\leqslant f^{(-H_1/2,\ H_2+H_1/2)}+g^{(-H_1/2,\ H_2+H_1/2)}$,

в которых нужно взять $\overline{M}$ и перейти к пределам, сначала $H_2\to\infty$, а затем $H_1\to\infty$. Аксиома **А4** будет верна по определению. Причем, так как $(-f)^{(-H_1,\,H_2)}=-f^{(-H_2,\,H_1)}$, то $\underline{M}f=\lim\limits_{H_2\to\infty}\lim\limits_{H_1\to\infty}\underline{M}f^{(-H_1,\,H_2)}$.

В дальнейшем будем в основном иметь дело с верхними средними, оставляя нижнее «за кулисами» формулы обращения (аксиомы **А4**). Для заданной $\mathscr{M}$ обозначим $\mathscr{F}_\infty$ класс всех признаков, для которых существует и не равен ∞ предел (1.4): $\mathscr{F}_\infty=\{f:\overline{M}f<\infty\}$, назовем этот класс *предельной областью существования верхних средних* ИРВ, а соответствующую совокупности средних $\langle\overline{M}\mathscr{F}_\infty\rangle$ модель — *предельной* $\mathscr{M}_\infty$. Смысл продолжения (1.4) очень естествен: неограниченные признаки мыслятся как имеющие некоторый потолок, который безмерно высоко расположен (точно также понимается нами космическая бескрайность). И в обозримом диапазоне $(-H_1,\ H_2)$ вычисляется среднее, при увеличении H_i все более приближающее предельное значение.

Этой же точки зрения можно было придерживаться для любых ИМ, считая все неограниченные признаки, входящие в $\mathscr{F}$, имеющими некоторый общий потолок, столь недосягаемо высокий, что его в наших действиях просто удобно не замечать, оперируя «нижними» частями признаков. При такой интерпретации предельная ИМ $\mathscr{M}_\infty$ «уравнивается в правах» с построенной по первичным признакам согласно формуле продолжения.

Замечания.

1. Переход к пределу в (1.4) математически подразумевает непрерывность правой части при $H_1,\ H_2\to\infty$, а это есть дополнительное свойство средних, ниоткуда из доказанного нами ранее не следующее. Если не принимать этого свойства, то можно было бы брать $\overline{M}f$ отличным от предела правой части (1.4), поэтому с формальных позиций предельная модель $\mathscr{M}_\infty$ есть форма сужения заданной $\mathscr{M}\supset\mathscr{M}_\infty$ по правилу, соответствующему (1.4).

2. Принцип предельного расширения области существования применим и к общему классу ИМ $\langle\tilde{M}\mathscr{G}\rangle$. Сначала средние с $\mathscr{G}$ распространяются по формуле продолжения (1.2) на мажорируемую набором $\mathscr{G}$ область $\mathscr{F}_{\mathscr{G}}$, в частности, на абсолютно ограниченные признаки $\mathscr{F}_{00}\subset\mathscr{F}_{\mathscr{G}}$ (если все $g\in\mathscr{G}$ ограничены, то $\mathscr{F}_{\mathscr{G}}=\mathscr{F}_0$). Затем предельным переходом (1.4) распространяются с $\mathscr{F}_{00}$ на $\mathscr{F}_\infty$. Линейная оболочка $\mathscr{L}^+(\mathscr{F}_{\mathscr{G}}\cup\mathscr{F}_\infty)$ и станет расширенной областью существования средних, переход к которой эквивалентен сужению модели $\langle\tilde{M}\mathscr{G}\rangle$ к предельной форме $\langle\tilde{M}\mathscr{G}\rangle_\infty$.

Иллюстрация ИРВ. Наглядно для дискретных пространств $\mathscr{X}$ ИРВ представляются многогранниками векторов **P** вероятностей, грани которых параллельны осям $P(x_i)=0$, как это видно из рис. 1.7, где треугольник $\mathscr{S}$, соответствующий голой ИМ, срисован с рис. 1.4. Здесь трапеция $\mathscr{M}_2$ определяется всего двумя первичными вероятностями: $P_2(x_1)$ — левая ее грань и $\overline{P}_2(x_2)$ — верхняя, т. е. имеет порядок $\widetilde{2}$, тогда как $\mathscr{M}_1$ имеет шесть первич-

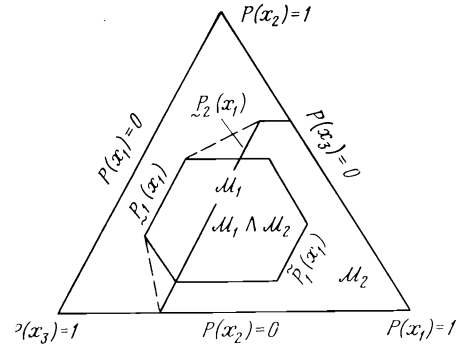


Рис. 1.7. Операция над ИРВ

ных граней, следовательно, шестого порядка. Пересечение двух ИРВ — снова ИРВ $\mathscr{M}_1 \bigwedge \bigwedge \mathscr{M}_2$ и первичными будут вероятности, полученные сложением воедино между собой первичных вероятностей $\mathscr{M}_1$ и $\mathscr{M}_2$, часть которых при согласовании окажется избыточной, так как проходят вне граней $\mathscr{M}_1 \bigwedge \mathscr{M}_2$, как, например, $\tilde{P}_2(x_2)$ или $\tilde{P}_1(x_1)$ на рис. 1.7.

Объединение двух ИРВ, обведенное на рис. 1.7 штриховкой, формирует новые контурные грани, которые не параллельны сторонам $\mathscr{T}$, т. е. уже не соответствуют вероятностям, что переводит в рамки более общих конструкций ИМ. Таким образом, класс всех ИРВ не замкнут относительно операции объединения. Голое ИРВ, соответствующее полному отсутствию нетривиальных первичных вероятностей, есть то же самое, что голая ИМ.

Перейдем к рассмотрению некоторых частных случаев ИРВ.

Конечно-аддитивные ИРВ. Интервальная модель, для которой первичной является система непересекающихся событий, называется *конечно-аддитивным интервальным распределением вероятностей* (короче, Σ-ИРВ). Обозначим $\mathscr{A}_\Sigma = \{A_1, A_2, \ldots\}$, $A_iA_j = \varnothing$, $i \neq j$ — набор попарно непересекающихся событий и пусть заданы $\underset{\sim}{P}(A_j)$, $\tilde{P}(A_j)$, $j=1, 2, \ldots$ Хотя не требуется, чтобы объединение $\underset{\sim}{A}_j$ охватывало все пространство $\mathscr{X}$, но удобно это считать, дополнив при необходимости исходный набор остаточным событием $A_0 = (\Sigma A_j)^c$ (если оно не пустое) и придав ему в качестве первичного тривиального вероятностный интервал $\underset{\sim}{P}(A_0) = 0$; $\tilde{P}(A_0) = 1$. Тогда $\mathscr{A}_\Sigma$ становится дроблением пространства $\mathscr{X}$ на непересекающиеся события, что далее и предполагается.

Непротиворечивость первичных вероятностей эквивалентна выполнению неравенств: $0 \leqslant \underset{\sim}{P}(A_j) \leqslant \tilde{P}(A_j)$, $\underset{\sim}{P}(\mathscr{X}) = \Sigma \underset{\sim}{P}(A_j) \leqslant 1$, $\tilde{P}(\mathscr{X}) = \Sigma \tilde{P}(A_j) \geqslant 1$. Последнее условие требуется только, если дробление $\mathscr{A}_\Sigma$ конечно. В случае счетного дробления в нем нет надобности, так как несмотря на расширение $\sum_1^k A_j$ при увеличении k всегда будет оставаться место для остаточного события A_0, для которого полагаем $\tilde{P}(A_0) = 1$, отсюда $\tilde{P}(\mathscr{X}) \geqslant 1$, где тильдой обозначена формально перенесенная с первичных вероятность $\mathscr{X}$.

Вторичными признаками будут всевозможные конечные линейные комбинации $g(x) = c + \Sigma c_j A_j(x)$, образующие линейный класс $\mathscr{L}\mathscr{A}_\Sigma$ функций. В него, в частности, входят так называемые *вторичные события* A_{J_k} — это те, которые набираются как объединения A_j: $A_{J_k} = \sum_{j \in J_k} A_j$, где J_k — конечный набор индексов. Для них $\underset{\sim}{P}(A_{J_k}) = \sum_{j \in J_k} \underset{\sim}{P}(A_j)$, $\tilde{P}(A_{J_k}) = \sum_{j \in J_k} \tilde{P}(A_j)$ — это перенесенные по аддитивности первичные вероятности (отсюда название: аддитивные ИРВ).

Уже говорилось о различии свойств 8 — верхней и **9** — нижней вероятностей. Это различие отражается и на перенесенных вероятностях, где также «лучшими» свойствами обладает нижняя вероятность. Проявляется это в том, что $\underset{\sim}{P}(A_J)$ могут быть по аддитивности распространены на суммы счетных множеств индексов J и это не повлияет на ИРВ. В самом деле, A_J при конечных поднаборах $J_k \subset J$ образуют внутренность, «фундамент» для A_J, так что $A_{J_k} \subset A_J$ и $\underset{\sim}{P}(A_{J_k}) \leqslant \underset{\sim}{P}(A_J)$, $\forall k$, поэтому в качестве $\underset{\sim}{P}(A_J)$ можно взять максимальное значение левой части последнего неравенства, получаемое переходом к пределу $J_k \uparrow J$, что ведет к формуле счетного суммирования: $\underset{\sim}{P}(A_J) = \sum_{j \in J} \underset{\sim}{P}(A_j)$, где J счетно. В частности, если число событий в наборе $\mathscr{A}_\Sigma$ счетно, то дополнение $J^c{}_k$ к конечному J_k будет счетным множеством индексов и

$$\underset{\sim}{P}(A^c_{J_k}) = \sum_{j \in J^c_k} \underset{\sim}{P}(A_j).$$

Формула счетного суммирования не распространяется на верхние вероятности, поскольку на счетном J для события A_J уже невозможно создать мажорирующую его «крышу» из конечного числа первичных событий.

Теорема 1.2. *Если первичные интервальные вероятности заданы на дроблении $\mathscr{A}_\Sigma$ пространства $\mathscr{X}$ на непересекающиеся события, то выражением*

$$\overline{P}(A_{J_k}) = \min\{\tilde{P}(A_{J_k}),\ 1 - \underset{\sim}{P}(A^c_{J_k})\} \tag{1.5}$$

они продолжаются, делаясь согласованными, на все вторичные события. Продолжение на все вторичные признаки осуществляется формулами

$$\overline{M}(\Sigma c_j A_j) = \min_c [c + \sum_{c_j > c} (c_j - c)\tilde{P}(A_j) - \sum_{c_j < c} (c - c_j)\underset{\sim}{P}(A_j)], \tag{1.6}$$

причем минимум достигается при $c = c^*$, удовлетворяющем уравнению

$$\sum_{c_j > c^*} \tilde{P}(A_j) + \sum_{c_j < c^*} \underset{\sim}{P}(A_j) + \sum_{c_j = c^*} [\varkappa \tilde{P}(A_j) + (1-\varkappa)\underset{\sim}{P}(A_j)] = 1$$

при однозначно существующем выборе $0 \leqslant \varkappa \leqslant 1$.

Доказательство вынесено в конец параграфа в дополнение 2.

Поясним формулу (1.5). Вероятность $\overline{P}(A_J)$ получается, с одной стороны, как сумма верхних вероятностей первичных событий, составляющих A_{J_k}, а с другой — данные об этой вероят-

ности черпаются из оценки нижней вероятности противоположного события $A^c_{J_k}$. Берется то из двух значений, которое более точное, т. е. меньшее по величине.

Из формулы (1.6) среднее для всего класса $\mathscr{L}\mathscr{A}_\Sigma$ вторичных признаков получается сдвигом: $\overline{M}(c+\sum c_jA_j)=c+\overline{M}(\sum c_jA_j)$. Понятно также, что (1.5) выводится из (1.6), и отсюда же сумма по $\underset{\sim}{P}(A_j)$ в конце правой части (1.6) допускается счетной.

Формула (1.6) становится нагляднее, если записать ее в виде

$$\overline{M}\left(\sum c_jA_j\right)=\sum_{c_j>c^*}c_j\bar{P}(A_j)+c^*P^*+\sum_{c_j<c^*}c_j\underset{\sim}{P}(A_j),$$

где $P^*=1-\sum_{c_j>c^*}\bar{P}(A_j)-\sum_{c_j<c^*}\underset{\sim}{P}(A_j)$. Тогда $\overline{M}\left(\sum c_jA_j\right)=\max_{\mathbf{P}}\sum c_jP_j$. Максимум ищется по векторам $\mathbf{P}$ вероятностей с компонентами $P_j\in[\underset{\sim}{P}(A_j),\ \bar{P}(A_j)]$ и достигается на векторе $\mathbf{P}^*$, компоненты P^*_j которого принимают минимально возможные значения $P^*_j=\underset{\sim}{P}(A_j)$ при малых c_j, т. е. $c_j<c^*$, максимальные $P^*_j=\bar{P}(A_j)$ при $c_j>c^*$, где c^* подбирается таким, чтобы $\mathbf{P}^*$ был вектором вероятностей, а именно $\sum P^*_j=1$. Этой цели служит и выбор P^* — компоненты вектора $\mathbf{P}$ индекса i, соответствующего равенству $c_i=c^*$.

Дальнейшее продолжение средних теперь уже на все ограниченные сверху признаки, которые в совокупности своей и составят естественную область существования средних для Σ-ИРВ, производится по известной формуле

$$\overline{M}f=\inf_{g:\,f(x)\leqslant g(x)\in\mathscr{L}\mathscr{A}_\Sigma}\overline{M}g,\quad f\in\mathscr{F}_0.$$

Счетно-аддитивные ИРВ. Мы сознательно долго воздерживались от безоговорочного распространения $\bar{P}(A_J)$, $A_J=\sum_{j\in J}A_j$, на счетные J по свойству суммируемости рядов, ибо этот шаг сразу вывел бы нас из привычных рамок принятой аксиоматики. Этот шаг — компетенция первичного набора, к чему мы теперь и обратимся.

Пусть $\mathscr{A}_\Sigma=\{A_1, A_2, \ldots\}$ — счетный набор непересекающихся событий, образующих разбиение $\mathscr{X}$. Мы видели, что если первичными являются значения $\underset{\sim}{P}(A_j)$, $\bar{P}(A_j)$, то нижние без ущерба переходят по аддитивности на $\underset{\sim}{P}(A_J)$ для любых, в том числе счетных J, а верхние $\bar{P}(A_{J_k})$ — только для конечных J_k, образуя костяк расчетов вероятностей (по (1.5)) для Σ-ИРВ.

Будем теперь считать, что первичными являются не только A_j (а отсюда и A_{J_k}), но и всевозможные счетные объединения A_J, в совокупности образующие систему $\mathscr{A}_\sigma$ событий, причем первичные вероятности задаются сразу счетно-аддитивными в том смысле, что $\bar{P}(A_J)=\sum_{j\in J}\bar{P}(A_j)$, $\underset{\sim}{P}(A_J)=\sum_{j\in J}\underset{\sim}{P}(A_j)$, $\forall J$, конечных и счетных J (ниже, в дополнении 3 показывается, что это требование

вовсе не является обязательным: первичной может быть $\mathscr{A}_\sigma$, а вероятности на ней заданы не счетно-аддитивными).

Интервальное распределение вероятностей, первичной для которого является система $\mathscr{A}_\sigma$ счетных сумм непересекающихся событий, а первичные вероятности (нижняя и верхняя) счетно-аддитивны, называется *счетно-аддитивным* (короче, σ-ИРВ) и обозначается $\langle\underset{\sim}{P}(A_\sigma),\ \bar{P}(A_\sigma)\rangle$.

При одних и тех же интервалах $\underset{\sim}{P}(A_j)$, $\bar{P}(A_j)$, $j=1, 2, \ldots$, за счет расширения первичного набора счетно-аддитивные ИРВ оказываются более узкими, чем Σ-ИРВ:

$$\langle\underset{\sim}{P}(\mathscr{A}_\sigma),\ \bar{P}(\mathscr{A}_\sigma)\rangle\subset\langle\underset{\sim}{P}(\mathscr{A}_\Sigma),\ \bar{P}(\mathscr{A}_\Sigma)\rangle.$$

Более того, σ-ИРВ, в общем, не относятся к классу Σ-ИРВ, потому что $\mathscr{A}_\sigma$ содержат счетные суммы событий, тогда как Σ-ИРВ задаются отдельными экземплярами событий и их вероятностями.

Таким образом, свойство счетной аддитивности эквивалентно фактическому расширению первичного набора событий и наложению дополнительных требований на первичные вероятности.

Для счетно-аддитивных ИРВ формула (1.5) согласования вероятностей на $\mathscr{A}_\sigma$ (они все будут теперь уже первичными) верна без каких-либо оговорок для всех $A_J\in\mathscr{A}_\sigma$. Равно как и формула (1.6) становится справедливой для вторичных признаков, каковыми являются теперь уже любые счетные суммы $\sum_1^\infty c_jA_j$ событий A_j из $\mathscr{A}_\sigma$. Для сравнения прямого и счетно-аддитивного ИРВ приведем пример.

Пример 1.10. Пусть $\mathscr{X}=\{0, 1, 2, \ldots\}$ — натуральный ряд чисел и заданы первичные вероятности $\underset{\sim}{P}_0$, $\bar{P}_0$, $\underset{\sim}{P}_1$, $\bar{P}_1$, ... так, что $\sum_0^\infty\underset{\sim}{P}_j\leqslant 1$. Получается Σ-ИРВ (собственно, для произвольных пространств $\mathscr{X}$ и дробления $\mathscr{A}_\Sigma=\{A_1, A_2, \ldots\}$ мы придем к такому же ИРВ, если отобразим $A_j\to j$, $j=0, 1, \ldots$). Для конечного множества J_k точек вероятности согласуются по (1.5). Рассмотрим счетное событие $A=\{0, 2, 4, \ldots\}$, состоящее из четных чисел, и A^c — из нечетных. Тогда для Σ-ИРВ

$$\underset{\sim}{P}(A)=\sum_{j=0}^\infty\underset{\sim}{P}_{2j},\quad\underset{\sim}{P}(A^c)=\sum_{j=0}^\infty\underset{\sim}{P}_{2j+1},$$

откуда получаются согласованные значения $\underset{\sim}{P}(A)=\underset{\sim}{P}(A)$, $\bar{P}(A)=1-\underset{\sim}{P}(A^c)$. Отметим, что $\bar{P}_j$ не участвуют в определении верхней вероятности счетных событий A по отсутствию конечных накрывающих A систем первичных событий.

Для счетно-аддитивного ИРВ дополнительным условием непротиворечивости будет $\sum_1^\infty\bar{P}_j\geqslant 1$. При конечных J_k вероятности те же. Для введенного выше

A имеем: $\bar{P}(A)=\sum_{j=0}^{\infty}\bar{P}_{2j}$, $\bar{P}(A^c)=\sum_{j=0}^{\infty}\bar{P}_{2j+1}$, и согласованными будут уже другие значения $\underset{\sim}{P}(A)=\max\{\underset{\sim}{P}(A),\ 1-\bar{P}(A^c)\}$; $\bar{P}(A)=\min\{\bar{P}(A),\ 1-\underset{\sim}{P}(A^c)\}$, что соответствует более точным вероятностям. Это увеличение точности достигается за счет того, что первичными исходными считаются не только натуральные числа, но и их всевозможные счетные суммы.

При точных первичных вероятностях $\underset{\sim}{P}_j=\bar{P}_j=P_j$, $j=0, 1, \ldots$, таких, что $\sum_{0}^{\infty}P_j=1$ (условие непротиворечивости), как нетрудно видеть, вероятность любого события A будет точной: $\underset{\sim}{P}(A)=\bar{P}(A)=P(A)$ как для конечно, так и для счетно-аддитивных ИРВ, т. е. оба типа ИРВ совпадут между собой и вероятность любого счетного подмножества будет точной, равной сумме вероятностей. Это есть закон счетной аддитивности точных вероятностей, верный для дискретных (конечных или счетных) дроблений $\mathscr{A}_\Sigma$ произвольного пространства $\mathscr{X}$.

Обобщения. Здесь будет рассмотрен случай, когда $\mathscr{A}_\Sigma$ не является дискретным, т. е. конечным или счетным. Удобно для наглядности воображать в качестве $\mathscr{X}$ числовую прямую $\mathscr{R}$.

Система $\mathscr{A}$ подмножеств называется *кольцом* и обозначается $\mathscr{K}$, если она замкнута по отношению к операциям пересечения и симметричной разности (Δ): $A, B\in\mathscr{K}\Rightarrow AB\in\mathscr{K}$, $A\Delta B=AB^c\cup A^cB\in\mathscr{K}$. А если к тому же она замкнута по отношению к счетным объединениям, то называется σ-*кольцом* и обозначается $\mathscr{K}_\sigma$. На числовой прямой $\mathscr{R}$ кольцо множеств образуется всевозможными конечными объединениями (суммами) непересекающихся отрезков, это будет *кольцо отрезков* $\mathscr{K}_\Sigma$ (подробнее см. [20]).

Мерой на кольце $\mathscr{K}$ называется неотрицательная конечно-аддитивная функция множеств, а на $\mathscr{K}_\sigma$ — счетно-аддитивная. Заметим, что суммы $A_{J_k}=\sum_{j\in J_k}A_j$ непересекающихся A_j в предыдущем изложении образовывали кольцо (а дополненные счетными суммами — σ-кольцо) и $\underset{\sim}{P}(A_J)$, $\bar{P}(A_J)$ — меры на нем (причем $\underset{\sim}{P}(A_J)$ без ущерба продолжаются до счетно-аддитивной меры).

Вернемся к общему случаю и будем считать, что первичные вероятности преподносятся двумя мерами $\underset{\sim}{P}(A)$ и $\bar{P}(A)$, $A\in\mathscr{K}$, на кольце $\mathscr{K}$. Конечно-аддитивная мера формирует Σ-ИРВ, а счетно-аддитивная на σ-кольце — соответственно σ-ИРВ.

Например, на $\mathscr{R}$ определены вероятности $\underset{\sim}{P}[a, b)$, $\bar{P}[a, b)$ любых отрезков и мы их переносим по аддитивности на суммы отрезков. При этом вероятности $\bar{P}$ могут быть или оказаться больше 1, что не вносит затруднений, так как откорректируется при согласовании. В общем, $\bar{P}(A)$, задаваемые как первичные вероятности на мелких множествах, оказываются уже не вероятностями, а мерами на широких множествах.

Требование непротиворечивости задания первичных мер состоит в следующем:

а) $0\leqslant\underset{\sim}{P}(A)\leqslant\bar{P}(A)$, $\forall A\in\mathscr{K}$;

б) $\underset{\sim}{P}(A)\leqslant 1$, $\forall A\in\mathscr{K}$;

в) $\bar{P}(\mathscr{X})\geqslant 1$.

Последнее условие имеет смысл, когда $\mathscr{X}$ входит в систему первичных событий. Иначе оно не нужно, так как по аксиоме **A1** все равно будет $P(\mathscr{X})=1$.

Формула продолжения и согласования вероятностей тождественна (1.5). Для произвольных событий она запишется:

$$\bar{P}(B)=\min\Big\{\inf_{B\subset A\in\mathscr{K}}\bar{P}(A),\ 1-\sup_{B^c\supset A\in\mathscr{K}}\underset{\sim}{P}(A)\Big\}.$$

Отметим, что если верхние вероятности не заданы, т. е. можно считать $\bar{P}(A)=1$, $\forall A\in\mathscr{K}$, то $\underset{\sim}{P}(A)=\underset{\sim}{P}(A)$, $\bar{P}(A)=1-\underset{\sim}{P}(A^c)$, $A\in\mathscr{K}$. Тогда нижняя вероятность аддитивна на суммах: $\underset{\sim}{P}(\sum A_i)=\sum\underset{\sim}{P}(A_i)$, а верхняя — в общем, нет.

Признак $f(x)$ называется $\mathscr{K}$-*измеримым*, если он представляется как равномерно сходящийся предел конечных линейных комбинаций событий из $\mathscr{K}$. Строго говоря, класс всех $\mathscr{K}$-измеримых признаков составляется замыканием класса $\mathscr{L}\mathscr{K}$ относительно равномерной сходимости. Слово «измеримый f» означает, что, умея измерять меры событий $A\in\mathscr{K}$, можно сколь угодно точно вычислить за конечное число шагов интеграл от f.

Если $f\in\mathscr{L}\mathscr{K}$, т. е. f есть конечная линейная комбинация событий из $\mathscr{K}$, то $\bar{M}f$ находится по (1.6). При $\mathscr{X}=\mathscr{R}$ для измеримых относительно кольца отрезков f (это все непрерывные функции x и с разрывами первого рода) суммы в (1.6) превращаются в интегралы:

$$\overline{M}f=\min_c\Big[c+\int(f(x)-c)^+d\bar{P}-\int(c-f(x))^+d\underset{\sim}{P}\Big],\qquad(1.7)$$

где плюс означает, что берется неотрицательная часть функции. Для Σ-ИРВ (первичным является кольцо $\mathscr{K}_\Sigma$ отрезков) здесь интеграл понимается в смысле Римана — Стилтьеса, а для σ-ИРВ — в смысле Лебега — Стилтьеса и для последнего класс признаков расширяется до измеримых по Лебегу.

Точные распределения вероятностей. Пусть на произвольной системе $\mathscr{A}$ первичных событий даются точные вероятности: $\underset{\sim}{P}(A)=\bar{P}(A)=P(A)$, $A\in\mathscr{A}$. Их согласованность эквивалентна непротиворечивости и состоит в том, что: 1) $A, B\in\mathscr{A}$, $A\subset B\Rightarrow P(A)\leqslant P(B)$; 2) если $A_j\in\mathscr{A}$ — попарно непересекающиеся события и их сумма также входит в систему $\mathscr{A}$, то выполняется закон аддитивных вероятностей:

$$\sum A_j\in\mathscr{A}\Rightarrow P(\sum A_j)=\sum P(A_j).$$

Вероятность суммы непересекающихся $A_j\in\mathscr{A}$ будет всегда точной, если сумма конечна (что сразу следует из свойств 8, 9), поэтому требование конечной аддитивности точных вероятностей

обязательно уже при их задании и равносильно их непротиворечивости. Но не требование счетной аддитивности, тем не менее которое должно выполняться, когда по заданию заведомо известно, что $\sum_1^\infty A_j \in \mathcal{A}$, т. е. что счетная сумма имеет точную вероятность.

Общие свойства точных вероятностей непосредственно вытекают из свойств интервальных вероятностей (см. начало настоящего параграфа). Продолжим начатую там нумерацию свойств, считая $A_i \in \mathcal{A}$.

12. $P(\mathscr{X}) = 1$, $P(\varnothing) = 0$.

13. $P(A) = 1 - P(A^c)$.

14. $P\left(\sum_1^k A_i\right) = \sum_1^k P(A_i)$.

15. $\lim\limits_{k\to\infty} \overline{P}\left(\sum_{k+1}^\infty A_i\right) = 0 \Rightarrow P\left(\sum_1^\infty A_i\right) = \sum_1^\infty P(A_i)$.

Обозначим $\mathcal{A}_*$ — набор событий, на котором при продолжении с $\mathcal{A}$ вероятности остаются точными. Очевидно, $\mathcal{A}_* \supset \mathcal{A}$. Из свойств **12** и **13** следует, что $\mathscr{X}$, $\varnothing \subset \mathcal{A}_*$ и что $A \in \mathcal{A}_* \Rightarrow A^c \in \mathcal{A}_*$. Свойство **14** означает замкнутость $\mathcal{A}_*$ относительно конечных сложений непересекающихся событий. Чем шире $\mathcal{A}$ в смысле количества событий и замкнутости операций, тем богаче будет $\mathcal{A}_*$, откуда могут появиться дополнительные свойства точных вероятностей.

Перейдем к случаю, когда исходный набор $\mathcal{A}$ замкнут относительно пересечений и разностей, т. е. образует собой кольцо множеств $\mathcal{A} = \mathcal{K}$ (можно и полукольцо, скажем, все отрезки числовой прямой). Тогда согласно свойствам **12**, **13**, **14** набор $\mathcal{A}_*$ событий, на которые продолжаются точные вероятности, образует *алгебру* множеств (алгебра есть кольцо с включенным в него $\mathscr{X}$, т. е. замкнутое относительно дополнений).

Конечно-аддитивными распределениями вероятностей (обозначаются $\mathscr{P}_\Sigma$) называются ИРВ, заданные первичными точными вероятностями на алгебре (полукольце, кольце) событий. Это частный случай Σ-ИРВ, когда первичные интервалы вероятностей превращаются в точные значения. Для этого случая, подставляя $\underline{P}(A_i) = P(A_i)$, $A_i \in \mathcal{A}_*$, имеем по свойству **9**: $\underline{P}\left(\sum_1^\infty A_i\right) = \sum_1^\infty P(A_i)$, т. е. точные исходные вероятности переходят по аддитивности в нижние вероятности счетных сумм. А вот верхние вероятности — в общем, нет. За исключением случая, когда остатки $\sum_{k+1}^\infty A_i$ сумм могут быть покрыты событиями B_k, вероятности которых при $k\to\infty$ делаются сколь угодно малыми. Тогда действует свойство **15**. В целях иллюстрации ниже приводится пример 1.11.

Свойство счетной аддитивности, механически перенесенное на любые счетные суммы, эквивалентно расширению первичного набора до *сигма-алгебры* $\mathcal{A}_\sigma$, включающей вместе с событиями любые их счетные суммы и дополнения к ним, и заданию на $\mathcal{A}_\sigma$ *счетно-аддитивного распределения* вероятностей $\mathscr{P}_\sigma$ (частный случай σ-ИРВ):

$$A_i \in \mathcal{A}_\sigma \Rightarrow \sum_1^\infty A_i \in \mathcal{A}_\sigma, \quad P_\sigma\left(\sum_1^\infty A_i\right) = \sum_1^\infty P_\sigma(A_i)$$

(первичной может быть $\mathcal{A}_\sigma$, а вероятности на ней точные, но не счетно-аддитивные, как это показывается в дополнении 2).

Для $\mathscr{P}_\sigma$ выполняется свойство монотонной сходимости: $B_n \uparrow B \Rightarrow P_\sigma(B_n) \to P_\sigma(B)$, в определенном смысле эквивалентное счетной аддитивности [19] точных вероятностей.

Средние от $\mathcal{A}_*$-измеримых признаков будут точными, получаемыми согласно (1.7) интегрированием $\underline{M}f = \int f(x)\,dP$, причем для $\mathscr{P}_\Sigma$ это будут интегралы Римана-Стилтьеса, а для $\mathscr{P}_\sigma$ — Лебега-Стилтьеса (тогда класс f с точными средними расширяется до измеримых по Лебегу). Продолжение средних на неограниченные признаки производится согласно (1.4):

$$\underline{M}f = \lim_{H_2\to\infty} \lim_{H_1\to\infty} Mf^{(-H_2, H_1)}, \quad \overline{M}f = \lim_{H_1\to\infty} \lim_{H_2\to\infty} Mf^{(-H_1, H_2)}.$$

При этом все усеченные признаки $f^{(-H_1, H_2)}$ должны быть измеримыми. Очевидно, среднее будет точным $\underline{M}f = \overline{M}f$, если пределы справа не зависят от порядка устремления H_1 и H_2 к бесконечности — это будет интеграл от неограниченной функции.

Для точных средних справедливо свойство $M\sum f_i = \sum Mf_i$, согласно которому символ M можно проносить за знаки конечных сумм. Это общее свойство. А в каких свойствах $\mathscr{P}_\Sigma$ будет отличаться от $\mathscr{P}_\sigma$? Ответ на поставленный вопрос дается примером.

Пример 1.11. Равномерное распределение (*мера-длина*). Пусть на $\mathscr{X} = [0; 1)$ первичными являются вероятности $P[a, b) = b - a$, $0 \leqslant a < b \leqslant 1$, равные длинам отрезков. При продолжении они определяют Σ-ИРВ $\mathscr{P}_\Sigma$ с точными вероятностями на алгебре отрезков $\mathcal{A}_\Sigma$ (куда входят суммы конечного числа отрезков и дополнения к ним). Точными нулевыми будут вероятности отдельных точек $P(x_i) = 0$ (получается при $a \to b = x_i$) и их конечных наборов: $P(\bigcup_1^k x_i) = \sum_1^k P(x_i) = 0$, поэтому считается $\bigcup_1^k x_i \in \mathcal{A}_\Sigma$. Для счетного числа точек все зависит от их расположения в смысле приемлемости свойства **15**. Например, событие $A = \{1, 1/2, 1/3, 1/4, \ldots\}$ будет иметь нулевую вероятность, так как остаток $\{1/k, 1/(k+1), \ldots\}$ покрывается отрезком $[0; 1/k)$ вероятности $1/k$, которая при увеличении k делается сколь угодно малой. А счетное событие D_∞ (множество (Дирихле), состоящее из всех рациональных точек отрезка $[0, 1)$, нельзя аналогичным образом покрыть, поэтому его вероятность вовсе неизвестна:

$$\underline{P}_\Sigma(D_\infty) = 0; \quad \overline{P}_\Sigma(D_\infty) = 1.$$

Результат сводится к фразе: «*умея измерять сколь угодно точно длины отрез-*

ков, нельзя никакими средствами замерить длину множества всех рациональных чисел»; практический смысл ее очевиден.

Счетно-аддитивное распределение $\mathscr{P}_\sigma$ при тех же начальных данных отрицает выделенный нами тезис. Оно имеет те же вероятности отрезков и их конечных сумм. Отличие в абсолютизации закона счетной аддитивности вероятностей. Так, для упомянутого множества Дирихле

$$\mathscr{P}_\sigma(D_\infty) = \sum_{x\in D_\infty} P(x) = \sum_1^\infty 0 = 0.$$

В нашем случае $\mathscr{P}_\sigma$ совпадает с мерой Лебега на отрезке [0; 1). Очевидно, $\mathscr{P}_\sigma \subset \mathscr{P}_\Sigma$ и первичными для $\mathscr{P}_\sigma$ фактически являются всевозможные конечные и счетные суммы отрезков и дополнений к ним, образующих борелевскую сигма-алгебру $\mathscr{A}_\sigma$. А точной эта мера после продолжения оказывается на несколько более широкой системе событий — так называемых лебеговых множествах, которые отличаются от борелевских лишь нулевыми событиями.

Средние $M\!f$ есть соответственно интегралы Римана (для $\mathscr{P}_\Sigma$) и Лебега (для $\mathscr{P}_\sigma$) на отрезке [0; 1) от интегрируемой функции $f(x)$.

Рассмотрим промежуточный случай между $\mathscr{P}_\Sigma$ и $\mathscr{P}_\sigma$. Пусть $\mathscr{P}_\Sigma$ дополняется еще одной первичной вероятностью $P(D_\infty)=1/2$ и обозначим полученное распределение $\mathscr{P}_\rho$. Имеем $\mathscr{P}_\rho = \mathscr{P}_\Sigma \wedge \langle P(D_\infty)=1/2\rangle$ и $\mathscr{P}_\sigma \subset \mathscr{P}_\rho \subset \mathscr{P}_\Sigma$, что являет пример, когда точными имеются вероятности отдельных исходов $x\in\mathscr{R}$, их счетной суммы D_∞, а в то же время закон счетной аддитивности не выполняется: $1/2 = P(D_\infty) > \sum\limits_{x\in D_\infty} P(x) = 0$. Иллюстрация необязательности счетной аддитивности.

Интервальные функции распределения. Пусть $\mathscr{X}=\mathscr{R}$ — числовая прямая, и первичными являются вложенные друг в друга неограниченные слева полуинтервалы $(-\infty, y)$, $y\in\mathscr{Y}$, где $\mathscr{Y}$ — произвольное подмножество $\mathscr{R}$.

Первичные вероятности

$$\underset{\sim}{P}(-\infty, y) = \underset{\sim}{F}(y), \quad \tilde{P}(-\infty, y) = \tilde{F}(y), \quad y\in\mathscr{Y},$$

как функции переменной y называются нижней и верхней первичными функциями распределения, а задаваемое ими в результате продолжения ИРВ называется *интервальной функцией распределения.*

Условие непротиворечивости первичных вероятностей выливается в требование

$$0 \leqslant \underset{\sim}{F}(x) \leqslant \tilde{F}(y) \leqslant 1, \quad \forall x \leqslant y, \quad x, y\in\mathscr{Y},$$

состоящее в том, что нижняя функция распределения нигде слева от y не должна возвышаться над верхней $\tilde{F}(y)$ (что будет обязательно выполнено, если, как это обычно делается, задавать $\underset{\sim}{F}(y)$ и $\tilde{F}(y)$ неубывающим и нижнюю не больше верхней: $\underset{\sim}{F}(y) \leqslant \tilde{F}(y)$, $\forall y$).

Согласование первичных вероятностей и продолжение их на любые полуинтервалы $(-\infty, x)$, $x\in\mathscr{R}$, проводится по формуле

$$\underline{F}(x) = \sup_{x\geqslant y\in\mathscr{Y}} \underset{\sim}{F}(y), \quad \overline{F}(x) = \inf_{x\leqslant y\in\mathscr{Y}} \tilde{F}(y). \tag{1.8}$$

Вероятности полуинтервалов продолжаются на вероятности одиночных отрезков:

$$\underline{P}(y, z) = [\underline{F}(z) - \overline{F}(y)]^+; \quad \overline{P}[y, z) = \overline{F}(z) - \underline{F}(y),$$

где плюс указывает на неотрицательную часть функции. Распространение этих формул на конечные суммы непересекающихся отрезков производится согласно выражениям:

$$\underline{P}\left(\sum_1^k [y_i, z_i)\right) = \sum_1^k \underline{P}[y_i, z_i);$$

$$\overline{P}\left(\sum_1^k [y_i, z_i)\right) = \overline{F}(z_k) - \underline{F}(y_1) - \sum_1^{k-1} [\overline{F}(y_{i+1}) - \underline{F}(z_i)]^+,$$

в которых полагается $y_1 \leqslant z_1 \leqslant y_2 \leqslant z_2 \leqslant \ldots \leqslant y_k \leqslant z_k$.

При точных функциях распределения $\underline{F}(x) = \overline{F}(x) = F(x)$, $\forall x$, (для этого таковыми они должны быть заданы) вероятности отрезков будут точными, равными приращениям: $P[y, z) = F(z) - F(y)$, как и для их конечных сумм, а продолжение на признаки будет соответствовать интегрированию по Риману — Стилтьесу: $Mg = \int g(x)\,dF(x)$.

Подобие ИРВ. Начнем с частных случаев. Пусть на произвольном $\mathscr{X}$ задан расширяющийся набор событий $\mathscr{A}_\uparrow = \{A_y,\ y\in\mathscr{Y}\subset\mathscr{R}\}$, $A_y \subset A_{y'}$ при $y<y'$. И пусть первичными являются вероятности этих событий $\underset{\sim}{F}(y) = P(A_y)$, $\tilde{F}(y) = \bar{P}(A_y)$. Тогда совершенно естественно отобразить $\mathscr{X}\to\mathscr{Y}$ так, чтобы $A_y \to (-\infty, y)$, и получить интервальную функцию распределения. Таким образом, ИРВ$\langle \underset{\sim}{P}(\mathscr{A}_\uparrow), \bar{P}(\mathscr{A}_\uparrow)\rangle$ по своей структуре и свойствам оказывается подобной интервальной функции распределения.

Аналогично, если задано Σ-ИРВ $\langle \underset{\sim}{P}(\mathscr{A}_\Sigma), \bar{P}(\mathscr{A}_\Sigma)\rangle$ на произвольном $\mathscr{X}$, то, отобразив элементы разбиений A_j в точки y_j числовой прямой, получим подобное ИРВ на подмножестве $\mathscr{Y} = \{y_1, y_2, \ldots\}$ числовой прямой. Если это окажется удобным, то можно считать $y_j = j$, и тогда $\mathscr{Y}$ — натуральный ряд чисел.

Основу подобия составляет взаимно-однозначное соответствие первичного набора множеств одного пространства, т. е. $\mathscr{X}$ и другого — $\mathscr{Y}$ с сохранением множественных операций объединения и пересечения (что называется гомеоморфизмом алгебр [20]). Очевидно, любое отображение $y = Sx$ пространства $\mathscr{X}$ и $\mathscr{Y}$ обеспечивает указанное соответствие.

Интервальное распределение вероятностей $\langle \tilde{P}^y(\mathscr{B})\rangle$ на $\mathscr{Y}$ называется *подобным* ИРВ$\langle \tilde{P}^x(\mathscr{A})\rangle$ на $\mathscr{X}$, если существует отобра-

жение S пространства $\mathscr{X}$ на $\mathscr{Y}$, при котором события $A\in\mathscr{A}$ переходят в события $B=SA\in\mathscr{B}$ с сохранением первичных вероятностей.

Конечно же, из $\mathscr{A}$ желательно изъять все события с несогласованными значениями $\bar{P}(A)$, тогда $\bar{P}^x(A)=\bar{P}^y(SA)$, $A\in\mathscr{A}$, $B\in\mathscr{B}$.

Смысл подобия состоит в упрощении структуры пространства $\mathscr{X}$ редукцией его в пространство $\mathscr{Y}$ как можно меньшей размерности. В то же время и согласование вероятностей, и их продолжение в обоих пространствах одинаково.

Семейства распределений. Пусть $\mathscr{P}_\theta$, $\theta\in\Theta$ — семейство точных распределений вероятностей на $\mathscr{X}$. Неизвестно, какое из $\mathscr{P}_\theta$ дает правильное описание явления, но какое-то из них это обязательно делает. Сказанное эквивалентно объединению $\mathscr{P}_\theta$, что ведет к ИМ:

$$\mathscr{M}=\bigvee_{\theta\in\Theta}\mathscr{P}_\theta.$$

Следовательно, любые семейства распределений вероятностей (аддитивных или счетно-аддитивных) дают ИМ, определяемую средними

$$\underline{M}g=\inf_{\theta\in\Theta}M_\theta g,\quad \overline{M}g=\sup_{\theta\in\Theta}M_\theta g,$$

где g — признаки, для которых средние по всем $\mathscr{P}_\theta$ являются точными.

Результат объединения, подчеркнем, есть в общем ИМ (а не ИРВ), за некоторым исключением, когда $\mathscr{P}_\theta$ является точками «тела» (можно, крайними) ИРВ. Например Σ-ИРВ $\langle\underset{\sim}{P}(\mathscr{A}_\Sigma),\ \bar{P}(\mathscr{A}_\Sigma)\rangle$ можно представить себе как объединение конечно-аддитивных распределений $\mathscr{P}_\Sigma$ таких, для которых $P(A_j)$ являются неизвестными и подчиняются условиям: $\underset{\sim}{P}(A_j)\leqslant P(A_j)\leqslant\bar{P}(A_j)$, $A_j\in\mathscr{A}_\Sigma$, (собственно, исходя из такого представления, в дополнении 2 доказываются формулы (1.5) и (1.6)).

Интервальную функцию распределения также можно мыслить себе как совокупность точных функций распределения, что позволяет выводить выражения для средних по формуле:

$$\overline{M}g=\sup_{\underline{F}\leqslant F\leqslant\overline{F}}\int g(x)\,dF(x).$$

Например, для признака $g(x)$, имеющего в точке x_m единственный глобальный максимум и не имеющего никаких локальных, получается:

$$\overline{M}g=g(x_m)\,[\overline{F}(x_m)-\underline{F}(x_m)]+\int_{-\infty}^{x_m}g(x)\,d\underline{F}(x)+\int_{x_m}^{\infty}g(x)\,d\overline{F}(x),$$

так как максимум достигается на $F(x)$, равной $\underline{F}(x)$ при $x<x_m$, имеющей в точке x_m скачок до $\bar{F}(x)$ и равной $\overline{F}(x)$ при $x>x_m$.

Относительные вероятности и средние. Истина постигается в сравнении. Это философское изречение применимо и к случайным явлениям. Подчас исходными данными являются сравнительные сведения о вероятностях и средних. Это суждения типа: «Событие A более вероятно, чем событие B», записываемое кратко «$PA\geqslant PB$»; или такое: «Признак f в среднем принимает большее значение, чем g», записываемое «$Mf\geqslant Mg$». Эти данные называются *относительными* (субъективными [14]) вероятностями и средними и характерны для экспертных оценок. Покажем, как они формально представляются в виде семейств распределений вероятностей и как преобразуются в средние значения соответствующих признаков.

Представим на миг, что вероятности событий A и B точные (что само собой подразумевает статистическую устойчивость явления) и соответствуют какому-либо точному распределению вероятностей $\mathscr{P}$ такому, что $P(A)\geqslant P(B)$. Для него по свойствам средних $M_{\mathscr{P}}(A-B)\geqslant M_{\mathscr{P}}A-M_{\mathscr{P}}B=P(A)-P(B)\geqslant 0$. Фраза «$PA\geqslant PB$» соответствует семейству всех таких распределений: $\mathscr{M}=\bigvee\mathscr{P}$, и по правилам вычисления средних для семейств имеем $\underline{M}(A-B)=\inf\limits_{\mathscr{P}}M_{\mathscr{P}}(A-B)=0$. Результат в терминах верхних средних запишется $\overline{M}(B-A)=0$ и может рассматриваться как соответствующее фразе «$PA\geqslant PB$» первичное среднее (см. табл. 1.1).

Из «$PA\geqslant PB$» продолжением первичного среднего вытекает $\underline{P}(B)\leqslant\underline{P}(A)\leqslant\bar{P}(B)$, $\bar{P}(B)\leqslant\bar{P}(A)$ (так как $\underline{M}B-\underline{M}A\leqslant 0=\overline{M}(B-A)\leqslant\overline{M}B-\underline{M}A$ и $0=\overline{M}(B-A)\geqslant\overline{M}B-\overline{M}A$) и интерпретируется как тот факт, что интервальная вероятность события A, перекрываясь, в общем, с интервальной вероятностью B, смещена в сторону бо́льших значений.

Разным фразам о старшинстве вероятностей соответствуют первичные средние вида $\overline{M}(B(x)-A(x))=0$, которые при разных A и B согласуются по правилам средних и продолжаются на любые признаки. Интересно, что и здесь будут верны правила логического вывода: «$PA\geqslant PB$» и «$PB\geqslant PC$»$\Rightarrow$«$PA\geqslant PC$» (так как $\overline{M}(C-B+B-A)\leqslant\overline{M}(C-B)+\overline{M}(B-A)\leqslant 0$).

Перейдем к средним. Совершенно аналогично показывается, что предложение «$Mf\geqslant Mg$» в устойчивых условиях эквивалентно $\underline{M}(f-g)=0$. Условием согласования этого предложения со средними от признаков f и g являются: $\underline{M}g\leqslant\underline{M}f\leqslant\overline{M}g$, $\overline{M}g\leqslant\overline{M}f$.

Замечание. Соотношение $\underline{M}f\geqslant\overline{M}g$, при котором интервалы средних не перекрываются между собой, отражает более жесткое предложение: «f в среднем больше g в любых (неустойчивых) условиях», когда точные средние Mf и Mg не существуют и допускаются лишь в интервальном понимании.

Дополнения. 1. Аксиоматизация ИРВ. Интервальные распределения вероятностей сами по себе могут быть определены как наборы интервальных

ных вероятностей $\underline{P}(A)$, $\overline{P}(A)$, $\forall A \subset \mathscr{X}$, связанных между собой аксиомами [21]: 1) $\underline{P}(\mathscr{X}) = \overline{P}(\mathscr{X}) = 1$; 2) $\overline{P}(A+B) \leqslant \overline{P}(A) + \overline{P}(B)$, $AB = \varnothing$; 3) $\underline{P}(A+B) \geqslant \underline{P}(A) + \underline{P}(B)$, $AB = \varnothing$; 4) $\underline{P}(A) = 1 - \overline{P}(A^c)$ (возможны и другие варианты эквивалентного выбора аксиом). Согласованные в смысле этих аксиом вероятности могут быть результатом продолжения первичных вероятностей, а дальнейшее продолжение вероятностей на средние признаков f по правилам ИМ приводит в итоге к интервальным средним $\underline{M}f$, $\overline{M}f$.

Этот путь традиционно вторит современному вероятностному подходу, где средние для измеримых f называются обычно *математическими ожиданиями* как результат математического расчета Mf по точным распределениям вероятностей $\mathscr{P}_\sigma$.

Аксиоматизация теории на базе вероятностей приводит к незамкнутой конструкции, поскольку очерчивает класс моделей только интервальными распределениями вероятностей, образующими все вместе лишь весьма узкий класс ИМ.

2. Доказательство теоремы 1.2. Докажем формулу (1.5). Событие A_J мажорируется либо самим собой, т. е. суммой $g_1(x) = \sum\limits_{j \in J} A_j(x)$, тогда $\overline{M}g_1 = \overline{P}(A_J)$, либо признаком $g_2(x) = 1 - \sum\limits_{j \notin J} A_j(x)$, тогда $\overline{M}g_2 = 1 - \underset{\sim}{P}(A^c{}_J)$. Минимальное (наиболее точное) из $\overline{M}g_1$ и $\overline{M}g_2$ и даст вероятность $\overline{P}(A_J)$ согласно (1.5). Легко убедиться, что никакой из других вторичных признаков, мажорирующих $A_J(x)$, не приведет к более точному значению этой вероятности.

Для доказательства (1.6) нужно представить $\langle \underset{\sim}{P}(\mathscr{A}_\Sigma), \overline{P}(\mathscr{A}_\Sigma) \rangle$ как семейство $\mathscr{A}_\Sigma$-точных распределений $\mathscr{P}$ таких, что $\underset{\sim}{P}(A_j) \leqslant P(A_j) \leqslant \overline{P}(A_j)$, и убедиться, что максимум по $\mathscr{P}$ при этих ограничениях

$$\overline{M} \sum c_j A_j(x) = \max_{\mathscr{P}} \sum c_j P(A_j) = \sum c_j P^*(A_j)$$

достигается на распределении вероятностей

$$P^*(A_j) = \begin{cases} \widehat{P}(A_j) & \text{при } c_j > c^*, \\ \varkappa \widehat{P}(A_j) + (1-\varkappa) \underset{\sim}{P}(A_j) & \text{при } c_j = c^*, \\ \underset{\sim}{P}(A_j) & \text{при } c_j < c^*, \end{cases}$$

где c^* и $\varkappa$ выбираются так, чтобы выполнялась нормировка $\sum P^*(A_j) = 1$, причем этот выбор однозначен. Распределение $\mathscr{P}^*$ принимает максимальные значения $\overline{P}(A_j)$ в той части $\mathscr{X}$, где функция $g(x) = \sum c_j A_j(x)$ велика, т. е. $c_j > c^*$, а для соблюдения нормировки вероятностей вынужденно оставляются самые минимальные значения $\underset{\sim}{P}(A_j)$ тем A_j, на которых $g(x)$ мала, т. е. $c_j < c^*$.

3. Примеры конечно- и счетно-аддитивных распределений. Пусть $\mathscr{X}$ — произвольное пространство, состоящее из несчетного числа точек x, и пусть вероятность каждой точки задана нулевой: $\overline{P}(x) \equiv 0$, $\forall x \in \mathscr{X}$. Это и будут первичные вероятности. Нулевые их значения передаются по аддитивности конечным суммам точек, составляющим здесь класс невозможных событий. Соответственно достоверными событиями, имеющими единичную нижнюю вероятность, будут противоположные события, т. е. $\mathscr{X}$ без любого конечного числа точек. Остальные события будут иметь тривиальные интервальные вероятности [0; 1]. Подчеркнем, что описанная модель — это не голое

ИРВ: как-никак данные о нулевых вероятностях точек все же присутствуют, что ведет к $\sum$-ИРВ.

Счетно-аддитивное распределение получится, если $\mathscr{X}$ несчетно и нулевые вероятности $\overline{P}(\bigcup\limits_1^\infty x_i) = 0$ исходно придаются любым конечным или счетным наборам точек. Здесь уже интервальные вероятности [0; 1] будут иметь только такие несчетные множества точек, что и дополнение к ним несчетно. Первичными для полученного σ-ИРВ являются счетные наборы точек и дополнения к ним, образующие вместе сигма-алгебру $\mathscr{A}_\sigma$, на которой вероятности счетно-аддитивны.

На $\mathscr{A}_\sigma$ будут заданы неаддитивные вероятности, если исходными иметь $\underline{P}(\bigcup\limits_1^k x_i) = 0$, $\overline{P}(\bigcup\limits_1^\infty x_i) = 1/2$ (вместо 1/2 может быть любая вероятность). Тогда исходное счетное множество точек будет иметь интервальную вероятность [0; 1/2].

Вернемся снова к началу примера и зададим для каждой точки x интервальную вероятность $0 \leqslant \underset{\sim}{P}(x) \leqslant \overline{P}(x) \leqslant 1$. Условие непротиворечивости сводится к тому, чтобы любые суммы нижних вероятностей не превышали единицу: $\sum \underset{\sim}{P}(x_i) \leqslant 1$, а это фактически означает, что $\underset{\sim}{P}(x_i)$ могут быть заданы не равными нулю лишь на дискретном множестве точек, причем так, что их сумма не больше 1. На $\overline{P}(x)$ никаких ограничений не накладывается. В результате:

$$\underline{P}(B) = \sum_{x_i \in B} \underset{\sim}{P}(x_i), \quad \overline{P}(B) = \min \left\{ \sum_1^k \widehat{P}(x),\ 1 \right\},$$

причем последнее равенство действует, если $B = \{x_1, \ldots, x_k\}$, а если B не является конечным, то $\overline{P}(B) = 1 - \underline{P}(B^c)$.

4. Гибридные распределения вероятностей. Они получаются добавлением к распределениям вероятностей не существующих для них средних (математических ожиданий). Поясним на примерах. Пусть дано распределение Коши $\mathscr{P}_C$ вероятностями отрезков $P_C(a, b) = \dfrac{1}{\pi}(\operatorname{arctg} b - \operatorname{arctg} a)$. Для него не существует ни математического ожидания, ни дисперсии, т. е. x, $x^2 \notin \mathscr{T}_\infty$. Присвоим извне недостающие значения MX, MX^2, сделав это в согласии с вероятностями, т. е. не меняя их в пересечении $\mathscr{P}_C \wedge \langle MX \rangle \wedge \langle MX^2 \rangle$. Это будет уже гибридная модель, не входящая в класс ИРВ, причем точные значения MX, MX^2 здесь могут заменяться на интервальные.

Другой пример, пусть $\mathscr{P}_л$ — равномерное распределение вероятностей на отрезке [0, 1]. Вероятности подотрезков равны их длинам, а математические ожидания — интегралам. Не существует математических ожиданий от неинтегрируемых по отрезку [0, 1] функций, скажем $f_k(x) = \dfrac{1}{(x-0{,}5)^k}$, $k \geqslant 1$. Заполняя «пустоты» любыми согласованными между собой значениями Mf_k, приходим к гибридным моделям: $\bigwedge\limits_k \langle Mf_k \rangle \wedge \mathscr{P}_л$.

Собственно, идея гибридных моделей близка к счетно-аддитивным распределениям вероятностей, а также к предельным формам моделей. В первом случае вероятностные интервалы для счетных сумм событий заменялись точными значениями согласно формуле счетной аддитивности (как видно из

предыдущего анализа, этого не обязательно делать именно таким образом), а во втором перенос средних на неограниченные признаки («пустоты», не вошедшие в $\mathcal{F}_{\mathcal{G}}$) производится по формуле интегрирования. Во всех этих случаях сужение ИМ не вызывало изменения исходных вероятностей.

1.5. ПРЕДСТАВЛЕНИЯ МОДЕЛЕЙ

Предисловие. Один итог, который уместно подвести, это то, что семейства моделей, а по сути их объединения, дают новые модели. Сказанное есть способ задания моделей совокупностями удобных, простых моделей (не обязательно семействами точных распределений вероятностей). Тогда тело модели наглядно представляется как выпуклая оболочка простейших ее частей, атомов; но может быть и не совсем простейших, а укрупненных по структуре, зато простых по описанию. Ясно, что представлений такого рода необычайно много, основные из них здесь рассматриваются.

Способы представлений делятся на универсальные и специальные. К *универсальным* относятся рассматриваемые здесь сечения тела ИМ параллельными гиперплоскостями, совокупности которых, как увидим, полностью определяются моделью и определяют ее. Сечения могут браться выборочно (задающие сечения), а также сами по себе могут описываться своими подсечениями.

К *специальным* способам относятся задания с помощью семейств, образованных сдвигами некоторой простой (стандартной) ИМ, введением неизвестного параметра в функциональное преобразование некоторого стандартного «шума» — процесса не столь уж сложной конфигурации, наконец заданием плотности (для нас она формальна) или семейств плотностей.

Сечения модели. *$\mathcal{G}$-простой* называется ИМ $\langle M\mathcal{G}\rangle$, определенная на первичном наборе $\mathcal{G}$ точными значениями Mg, $g\in\mathcal{G}$, первичных средних. Очевидно, $\langle M\mathcal{G}\rangle=\bigwedge\limits_{g\in\mathcal{G}}\langle Mg\rangle$, и результат будет непустым только в случае непротиворечивости набора Mg, $g\in\mathcal{G}$, состоящего в следующем:

$$\sum_1^k c_i g_i(x)\geqslant 0\Rightarrow\sum_1^k c_i Mg_i\geqslant 0,\quad \forall c_i,\ g_i\in\mathcal{G}.$$

$M_*\mathcal{G}$-сечением ИМ $\mathcal{M}$ называется пересечение: $\mathcal{M}_{M_*\mathcal{G}}=\mathcal{M}\wedge\langle M_*\mathcal{G}\rangle$. Таким образом, $\mathcal{M}_{M_*\mathcal{G}}$ — это ИМ, полученная прибавлением к средним $\mathcal{M}$ набора точных первичных средних M_*g, $g\in\mathcal{G}$. Совокупность $M_*\mathcal{G}=\{M_*g: g\in\mathcal{G}\}$ есть набор числовых характеристик секущей ИМ; звездочки означают, что это вносимые извне (по желанию меняемые) величины.

Если $\mathcal{F}$ есть область существования для верхних средних $\mathcal{M}$, то областью существования для $M_*\mathcal{G}$-сечения $\mathcal{M}_{M_*\mathcal{G}}$ будет $\mathcal{L}(\mathcal{F}\cup\mathcal{G})$, состоящая из всевозможных признаков вида $f+\sum c_i g_i$, $f\in\mathcal{F}$, $g_i\in\mathcal{G}$. Область существования сечения шире, чем у исходной

ИМ, если $\mathcal{G}\cup\mathcal{F}\neq\mathcal{F}$, т. е. в $\mathcal{G}$ включаются признаки, не входящие в $\mathcal{F}$.

В случае одного секущего признака получается M_*g-сечение $\mathcal{M}_{M_*g}$. Это сечение будет, очевидно, пустым, если $M_*g<\underline{M}g$ или $M_*g>\overline{M}g$. Геометрически секущую ИМ $\langle M_*g\rangle$ (при конечном числе элементов пространства $\mathcal{X}$) можно представить себе как гиперплоскость в пространстве векторов вероятностей. Для случая $\mathcal{X}=\{x_1, x_2, x_3\}$ сечение $\mathcal{M}_{M^*g}$ изображено на рис. 1.8 сплошной линией.

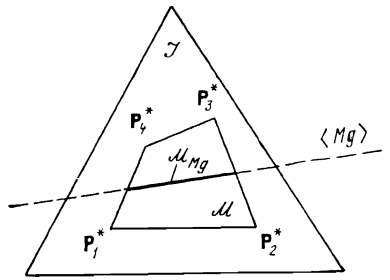


Рис. 1.8. Сечения модели

Рассмотрим, как нужно искать средние, соответствующие сечениям. Пусть $\mathcal{M}_{M_*g}=\mathcal{M}\wedge\langle M_*g\rangle$. Первичными средними для $\mathcal{M}_{M_*g}$ будут $\overline{M}h$, $h\in\mathcal{F}$ и M_*g, поэтому

$$\overline{M}_{M_*g}f=\inf_{h(x)+cg(x)\geqslant f(x),\ h\in\mathcal{F}}(\overline{M}h+cM_*g),$$

где нижняя грань ищется по $c\in\mathcal{R}$ и по $h\in\mathcal{F}$. При заданном c нижняя грань по h достигается, когда $h=f-cg$, в результате

$$\overline{M}_{M_*g}f=\min_c[\overline{M}(f-cg)+cM_*g]. \tag{1.9}$$

Замечание. Если g не принадлежит области существования обеих границ $\mathcal{M}$, т. е. $g\notin\mathcal{F}\cap(-\mathcal{F})$, то для любого признака $f\in\mathcal{F}\cap(-\mathcal{F})$ справедливо $\overline{M}_{M_*g}f=\overline{M}f$ (так как $\overline{M}(f-cg)=\infty$ при $c\neq 0$).

Аналогичным образом, если $\mathcal{G}_k=\{g_1, \ldots, g_k\}$ есть конечный набор признаков, то

$$\overline{M}_{M_*\mathcal{G}_k}f=\min_{\forall c_i}\left[\overline{M}\left(f-\sum_1^k c_i g_i\right)+\sum_1^k c_i M_* g_i\right]. \tag{1.10}$$

Для произвольного набора $\mathcal{G}$ на основании равенств $\mathcal{M}_{M_*\mathcal{G}}=\mathcal{M}\wedge\langle\mathcal{M}_*\mathcal{G}\rangle=\mathcal{M}\wedge\left(\bigwedge\limits_{\mathcal{G}_k\subset\mathcal{G}}\langle M_*\mathcal{G}_k\rangle\right)=\bigwedge\limits_{\mathcal{G}_k\subset\mathcal{G}}(\mathcal{M}\wedge\langle M_*\mathcal{G}_k\rangle)=\bigwedge\limits_{\mathcal{G}_k\subset\mathcal{G}}\mathcal{M}_{M_*\mathcal{G}_k}$ имеем:

$$\overline{M}_{M_*\mathcal{G}}f=\inf_{\mathcal{G}_k\subset\mathcal{G}}\overline{M}_{M_*\mathcal{G}_k}f,$$

где инфимум берется по всевозможным конечным поднаборам $\mathcal{G}_k$ признаков из $\mathcal{G}$.

Свойства сечений. Сечение обладает привычными свойствами ИМ. Помимо этого верны следующие свойства:

1. $\mathcal{M}_{M_*\mathcal{G}}\subset\mathcal{M}$.
2. $\mathcal{M}_{M_*\mathcal{X}}=\mathcal{M}$ при $M_*\mathcal{X}=1$, в противном случае $\mathcal{M}_{M_*\mathcal{X}}=\varnothing$.

3. $(\mathcal{M}_{M_*\mathcal{G}_1})_{M_*\mathcal{G}_2} = \mathcal{M}_{M_*(\mathcal{G}_1 \cup \mathcal{G}_2)}$, т. е. $M_*\mathcal{G}_2$-сечение $\mathcal{M}_{M_*\mathcal{G}_1}$ есть то же самое, что и $M_*(\mathcal{G}_1 \cup \mathcal{G}_2)$-сечение $\mathcal{M}$, т. е. $\mathcal{M} \wedge \langle M_*\mathcal{G}_1\rangle \wedge \langle M_*\mathcal{G}_2\rangle$.

4. Если $\tilde{M}h$, $h \in \mathcal{H}$ определяет модель $\langle \tilde{M}\mathcal{H}\rangle$, то $M_*\mathcal{G}$-сечение будет определяться средними $\tilde{M}\mathcal{H} \cup M_*\mathcal{G}$, т. е. $\langle \tilde{M}\mathcal{H}\rangle_{M_*\mathcal{G}} = \langle \tilde{M}\mathcal{H} \cup M_*\mathcal{G}\rangle$.

5. $\mathcal{M} = \bigwedge_\theta \mathcal{M}_\theta \Rightarrow \mathcal{M}_{M_*\mathcal{G}} = \bigwedge_\theta (\mathcal{M}_\theta)_{M_*\mathcal{G}}$ — пересечения перестановочны с сечениями.

6. Границы $\overline{M}_{M_*\mathcal{G}} f$ аддитивны относительно прибавления конечных линейных комбинаций признаков из $\mathcal{G}$: $\overline{M}_{M_*\mathcal{G}}(f + \sum c_i g_i) = \overline{M}_{M_*\mathcal{G}} f + \sum c_i M_* g_i$.

Доказательство этих свойств элементарно.

Из свойства 6 следует, что сечение будет одним и тем же для всех ИМ, получаемых «сдвигом» всех первичных признаков на $\sum c_i g_i$, $g_i \in \mathcal{G}$, и соответствующим сдвигом их средних на $\sum c_i M_* g_i$.

Теорема о представлении ИМ.

Теорема 1.3. *Любая ИМ $\mathcal{M}$ представляется как объединение ее $M_*\mathcal{G}$-сечений*:

$$\mathcal{M} = \bigvee_{M_*\mathcal{G}} \mathcal{M}_{M_*\mathcal{G}}, \tag{1.11}$$

где $\mathcal{G}$ есть любой взятый набор признаков, а объединение производится по значениям M_*g в интервалах $\underline{M}g \leqslant M_*g \leqslant \overline{M}g$, $g \in \mathcal{G}$.

Доказательство теоремы вынесено в дополнение 1 в конец параграфа.

Замечание. В объединении (1.11) вполне можно допустить M_*g пробегающими значения от $-\infty$ до ∞, так как при $M_*g < \underline{M}g$ или $M_*g > \overline{M}g$ сечение пусто: $\mathcal{M}_{M_*\mathcal{G}} = \varnothing$.

Если в качестве $\mathcal{G}$ взять произвольный набор $\mathcal{B}$ событий, то из теоремы следует, что каждая ИМ может быть представлена объединением ИМ с точными на наборе $\mathcal{B}$ вероятностями событий:

$$\mathcal{M} = \bigvee_{\underline{P}\mathcal{B} \leqslant P_*\mathcal{B} \leqslant \overline{P}\mathcal{B}} \mathcal{M}_{P_*\mathcal{B}},$$

где $P_*\mathcal{B} = \{P_*(B),\ B \in \mathcal{B}\}$ — обозначение совокупности вероятностей, причем $\underline{P}\mathcal{B}$ — нижних и $\overline{P}\mathcal{B}$ — верхних вероятностей.

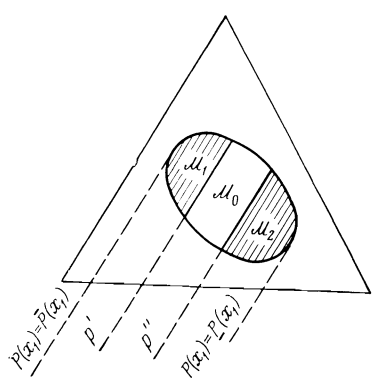


Рис. 1.9. Представление модели сечениями

Геометрическое толкование теоремы о представлении видно из рис. 1.8 и 1.9: $\mathcal{M}$ описывается последовательностью параллельных отрезков, полученных пересечением прямых $\langle M_*g\rangle$ с телом модели $\mathcal{M}$. Это при одной g. Если их несколько, то сечение сечений будет давать все более мелкие элементы (вплоть до атомов модели, а для дискретных пространств — векторов вероятностей, и тогда $\mathcal{M}$ представляется как семейство атомарных моделей — векторов вероятностей). Сказанное формулируется в виде следствия.

Следствие 1. $\langle \tilde{M}\mathcal{G}\rangle = \bigvee_{\langle M_*\mathcal{G}\rangle \subset \langle \tilde{M}\mathcal{G}\rangle} \langle M_*\mathcal{G}\rangle$.

Здесь $\langle M_*\mathcal{G}\rangle$ — простые модели, определенные точными значениями M_*g, $g \in \mathcal{G}$. Таким образом, модели с первичными средними $\tilde{M}g$, $g \in \mathcal{G}$, представляются как семейства моделей с точными значениями M_*g, $g \in \mathcal{G}$, такими, что $M_*g \leqslant \tilde{M}g$.

Следствие 1, очевидно, останется в силе, если вместо $\mathcal{G}$ взять любой включающий $\mathcal{G}$ набор признаков. А также любой набор признаков и событий, из которого линейными преобразованиями (или их замыканиями) может быть получен каждый признак из $\mathcal{G}$. В частности, если это система событий, то имеем утверждение.

Следствие 2. *Если все признаки набора $\mathcal{G}$ измеримы относительно системы $\mathcal{A}$ множеств* (т. е. представляются как конечные линейные комбинации индикаторов событий $\mathcal{A}$ или их замыкания относительно равномерной сходимости), *то*

$$\langle \tilde{M}\mathcal{A}\rangle = \bigvee_{\langle P_*\mathcal{A}\rangle \subset \langle \tilde{M}\mathcal{G}\rangle} \langle P_*\mathcal{A}\rangle.$$

Согласно этому следствию ИМ представляется как семейство точных на $\mathcal{A}$ распределений вероятностей. Если $\mathcal{A}$ есть алгебра или кольцо событий, то это будут конечно-аддитивные распределения вероятностей. А интересно, что будет, если $\mathcal{A}$ — сигма-алгебра, т. е. алгебра, замкнутая относительно счетных объединений? Тогда все равно ИМ представляется как семейство $\mathcal{A}$-точных распределений вероятностей, но это будут опять же в основном конечно-аддитивные распределения (!).

Можно сделать вывод, что счетно-аддитивные распределения сами по себе являются слишком редким исключением в «семье» распределений вероятностей, чтобы ими можно было описать многие ИМ (в частности, конечной размерности). Причем расширение системы $\mathcal{A}$ с целью дробления ею пространства $\mathcal{X}$ на все более мелкие части, а отсюда логический переход к борелевским сигма-алгебрам, и к более мелким лебеговским, хотя и несколько увеличивает описательные возможности счетно-аддитивных распределений, в принципиальной своей основе вывода не меняет.

Изюмина, скрытая в следствиях 1 и 2, состоит в том, что неустойчивые в статистическом смысле явления описываются в виде семейств точных моделей, соответствующих устойчивым явлениям, в частности, с помощью семейств точных распределений вероятностей. Неустойчивость статистическая взаимно «перекачи-

вается» в неустойчивость информационную, в наше незнание точных законов, неизвестность выбора. На первый взгляд, парадоксальный вывод, но на самом деле вполне естественный, так как в обоих случаях при независимых повторах в пределе будем получать разные средние арифметические, а это же и средние в интервальном их понимании.

Определение ИМ задающими сечениями. Только что говорилось о том, что можно мыслить себе ИМ в виде объединения или семейства более мелких ее частей — моделей. Но ведь это есть и способ задания $\mathcal{M}$, если ее исходно определять не через свои средние $\bar{M}f$, а как объединение более простых по описанию задающих ее моделей $\mathcal{M}^*_\theta$: $\mathcal{M}=\bigvee\limits_\theta \mathcal{M}^*_\theta$. Простых в том смысле, что для них легко находятся средние $\bar{M}_\theta f$. Тогда $\bar{M}f=\sup\limits_\theta \bar{M}^*_\theta f$. Это один из способов непрямого задания ИМ, различные аспекты которого здесь и обсуждаются.

Сначала рассмотрим тот случай, когда $\mathcal{M}^*_\theta$ являются $M^*\mathcal{G}$-точными, а роль многомерного параметра θ выполняет сам набор средних M^*g, $g\in\mathcal{G}$. Запишем

$$\mathcal{M}=\bigvee_{M^*\mathcal{G}} \mathcal{M}^*_{(M^*\mathcal{G})}. \qquad (1.12)$$

Справа символ $M^*\mathcal{G}$ сознательно заключен в круглые скобки, чтобы указать на тот факт, что для $\mathcal{M}^*_{(M^*\mathcal{G})}$ средние от признаков $g\in\mathcal{G}$ являются точными, равными значению соответствующего параметра: $M^*_{(M^*\mathcal{G})}g=M^*g$, так и на то, что в отличие от (1.11) «параметры» не обязаны пробегать все значения из $[\underline{M}g, \bar{M}g]$, $g\in\mathcal{G}$, а $\mathcal{M}^*_{(M^*\mathcal{G})}$ в свою очередь не обязательно должны быть $M\mathcal{G}$-сечениями модели $\mathcal{M}$. Как это хорошо видно из рис. 1.9, где $\mathcal{M}=\mathcal{M}_0\bigvee\mathcal{M}_1\bigvee\mathcal{M}_2$ может быть задана сечениями лишь ее частей $\mathcal{M}_1$ и $\mathcal{M}_2$ и здесь, если даже сечения $\mathcal{M}_0$ задавать пустыми, то все равно $\mathcal{M}=\mathcal{M}_1\bigvee\mathcal{M}_2$, т. е. выпуклая оболочка $\mathcal{M}_1$ и $\mathcal{M}_2$ (или сечений) определит $\mathcal{M}$.

В формуле (1.12) $\mathcal{M}^*_{(M^*\mathcal{G})}$ будем называть *задающими модель сечениями*. Рассмотрим пример.

Пример 1.12. Пусть $\mathcal{X}=\{x_1, x_2, x_3\}$ — три элементарных исхода и $\mathcal{M}$ интерпретируется как выпуклое семейство векторов вероятностей $\mathbf{P}$. Насколько видно из рис. 1.9, описание $\mathcal{M}$ первичными средними, эквивалентное описанию контура $\mathcal{M}$ касательными линиями (которых бесконечно много), является неудобным. В то же время каждое ее $P^*(x_1)$-сечение $\mathcal{M}^*_{P^*(x_1)}$ есть довольно простая по структуре ИМ, задаваемая точной вероятностью $P^*(x_1)$ и пределами изменения $P^*(x_2)$: $\underline{P}_{P^*(x_1)}(x_2)\leqslant P^*(x_2)\leqslant \bar{P}_{P^*(x_1)}(x_2)$, зависящими, в общем, от $P^*(x_1)$. Сечения $\mathcal{M}^*_{P^*(x_1)}$ будут задающими для $\mathcal{M}=\bigvee\limits_{P^*(x_1)}\mathcal{M}^*_{P^*(x_1)}$, нужно указать лишь пределы изменения параметра $P^*(x_1)$ либо в диапазоне от $\underline{P}(x_1)$ до $\bar{P}(x_1)$, соответствующем $\mathcal{M}$, либо в более узких двух диапазонах

$[\underline{P}(x_1), p']$ и $[p'', \bar{P}(x_1)]$, соответствующих отдельно $\mathcal{M}_1$ и $\mathcal{M}_2$, обозначенным на рис. 1.9.

Вернемся к формуле (1.12). Пусть модели $\mathcal{M}^*_{(M^*\mathcal{G})}$ описываются помимо значений M^*g, $g\in\mathcal{G}$, (разных для разных моделей) все одними и теми же верхними первичными средними $\bar{M}h$, $h\in\mathcal{H}$:

$$\mathcal{M}^*_{(M^*\mathcal{G})}=\langle \bar{M}\mathcal{H}\rangle\wedge\langle M^*{}_{\mathcal{G}}\rangle.$$

Тогда их объединение по изменениям M^*g, $g\in\mathcal{G}$, ограниченным сверху числами $\bar{M}g$, $g\in\mathcal{G}$, определит ИМ $\mathcal{M}$, первичными для которой будут те же самые $\bar{M}\mathcal{H}$, и плюс к этому, $\bar{M}\mathcal{G}$, что формально записывается:

$$\mathcal{M}=\bigvee_{M^*g\leqslant\bar{M}g,\ g\in\mathcal{G}}(\langle\bar{M}\mathcal{H}\rangle\wedge\langle M^*{}_{\mathcal{G}}\rangle)=\langle\bar{M}\mathcal{H}\rangle\wedge\langle\bar{M}{}_{\mathcal{G}}\rangle.$$

Поясним сказанное. Пусть $\mathcal{G}=g$ и изобразим на рис. 1.10 M^*g-сечение как плоский многогранник $\mathcal{M}^*_{(M^*g)}$ на гиперплоскости точного значения M^*g. Если его грани $\bar{M}h_j$ не меняются при M^*g-«сдвигах», то они сохраняются и для фигуры $\mathcal{M}$, полученной в результате параллельного перемещения этого плоского многогранника при изменении M^*g от $\underset{\sim}{M}g$ до $\bar{M}g$. Плюс к этому две грани будут соответствовать «крайним» значениям $M^*g=\underset{\sim}{M}g$ и $M^*g=\bar{M}g$. Кстати, расположенные на них многогранники полностью описывают $\mathcal{M}$: $\mathcal{M}=\mathcal{M}^*_{(\underset{\sim}{M}g)}\bigvee\mathcal{M}^*_{(\bar{M}g)}$. Подобная редукция описания возможна при любом наборе $\mathcal{G}$.

Несколько более общий случай по сравнению с предыдущим будет иметь место, если (при прежних остальных условиях) $\bar{M}_{(M^*\mathcal{G})}h$ зависят от M^*g_i, $g_i\in\mathcal{G}$, и линейно меняются при их изменениях:

$$\bar{M}_{(M^*\mathcal{G})}h=\tilde{m}_h+\sum c_i(h)\,M^*g_i,\quad h\in\mathcal{H},$$

где коэффициенты $c_i(h)$ зависят от h. На рис. 1.10 это будет выглядеть как изменение направления движения при смещении плоского многогранника, что вызовет и изменение итогового положения граней $\mathcal{M}$, соответствующих признакам h сечений, а значит, и самих h, которые переходят в $h-\sum c_i(h)\times g_i$ со значениями $M[h-\sum c_i(h)g_i]=\tilde{m}_h$, $h\in\mathcal{H}$. Эти первичные средние вместе с $\underset{\sim}{M}g$, $\bar{M}g$, $g\in\mathcal{G}$, и определят $\mathcal{M}$. Рассмотрим пример такого представления.

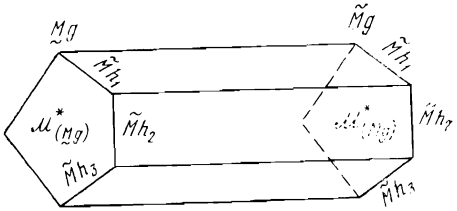


Рис. 1.10. Формирование граней модели

Пример 1.13. Пусть на $\mathcal{R}$ задающие сечения $\mathcal{M}^*_{(M^*X)}$ при каждом фиксированном M^*X заданы средними $\bar{M}^*_{(M^*X)}X^2=\tilde{m}_2+M^*X$. Здесь в принятых выше обозначениях $g(x)=x$, $h(x)=x^2$. Тогда объединения сечений по параметру M^*X, меняющемуся в пределах $\underset{\sim}{MX}\leqslant M^*X\leqslant\bar{M}X$, образует ИМ с первичными средними: $\bar{M}(X^2-X)=m_2$, $\underset{\sim}{MX}$, $\bar{M}\tilde{X}$.

Мы рассмотрели тот случай, когда первичные средние $M^*\mathcal{G}$-сечений зависят от $M^*\mathcal{G}$. Другой способ — сделать зависящим от $M^*\mathcal{G}$ сам вид первичных признаков $h_{M^*\mathcal{G}}$ задающих сечений.

Пример 1.14. Задание случайной величины средним и дисперсией. *Дисперсия* есть мощность центрированной к точно нулевому среднему с. в. $\bar{\sigma}^2=\bar{M}(X-MX)^2$. В общем случае нельзя определить дисперсию. Можно сделать это только при допущении о точных MX, т. е. для M^*X-сечений $\mathcal{M}^*_{(M^*X)}$ модели, задавая для них $\bar{M}^*_{(M^*X)}$ $(X-M^*X)^2=\hat{\sigma}^2_{(M^*X)}$. Объединение сечений по M^*X и даст модель случайной величины, заданной пределами $\underline{MX}$, $\bar{M}X$ изменения среднего и (при каждом M^*X) верхней дисперсией $\hat{\sigma}^2_{(M^*X)}$.

Преобразуя в каждом сечении выражение для дисперсий с учетом того, что M^*X является точным, имеем: $\hat{\sigma}^2_{(M^*X)}=\bar{M}^*_{(M^*X)}(X-M^*X)^2=\hat{M}^*_{(M^*X)}X^2-(M^*X)^2$, или $\bar{M}^*_{(M^*X)}X^2=\hat{\sigma}^2_{(M^*X)}-(M^*X)^2$. Отсюда следует, что первичный признак $h(x)=(x-M^*X)^2$ для M^*X-сечений эквивалентным образом заменяется на x^2 с первичным значением $\check{\sigma}^2_{(M^*X)}-(M^*X)^2$, зависящим от M^*X нелинейно.

Данная методика может быть продолжена на задание ИМ интервальными центральными моментами и куммулянтами.

Представление через стандартную ИМ. Вернемся к общему представлению: $\mathcal{M}=\bigvee\mathcal{M}_\theta$, частными случаями которого являются как запись ИМ в виде выпуклой оболочки простых распределений (вершин), так и в виде объединения сечений. Хорошо бы, чтобы все задающие $\mathcal{M}_\theta$ имели одинаковую, достаточно простую структуру.

Остановимся на том случае, когда все $\mathcal{M}_\theta$ получаются несложным образом из одной $\mathcal{M}_0$, называемой *стандартной*. Для этого наводится зависящее от θ соответствие между признаками $f\leftrightarrow f_\theta$ так, чтобы их средние для стандартной и задающей ИМ совпали: $\bar{M}_0 f=\bar{M}_\theta f_\theta$, при этом не нарушив согласованности. Тогда для нахождения $\bar{M}_\theta\varphi$ достаточно выявить соответствующий φ признак f (при заданном θ) и взять от него среднее $\bar{M}_0 f$.

Рассмотрим более строго вопрос, каким в этой схеме должно быть соответствие между признаками, как они связываются между собой? Пусть стандартная $\mathcal{M}_0$ определяется своими средними $\bar{M}_0 f$, $f\in\mathcal{F}$, и зададим $\mathcal{M}_\theta$ значениями

$$\bar{M}_\theta f=\bar{M}_0(\mathrm{L}_0 f), \qquad (1.13)$$

где L_θ — оператор, отображающий область $\mathcal{F}_\theta$ существования $\mathcal{M}_\theta$ в область $\mathcal{F}_0$ существования $\mathcal{M}_0$: $\mathcal{F}_\theta\xrightarrow{\mathrm{L}_\theta}\mathcal{F}_0$, $\theta\in\Theta$.

Утверждение 1.4. *Средние в формуле (1.13) будут согласованными, если оператор* L_θ *обладает следующими двумя свойствами: а) линейностью* $\mathrm{L}_\theta(c_1f_1+c_2f_2+c)=c_1\mathrm{L}_\theta f_1+c_2\mathrm{L}_\theta f_2+c$; *б) сохранением порядка* $f_1\geqslant f_2\Rightarrow\mathrm{L}_\theta f_1\geqslant\mathrm{L}_\theta f_2$.

Нужно доказать, что для границ, выраженных (1.13), выполняются аксиомы ИМ (см. стр. 15). Очевидны **А1** и **А4**. Далее, $b\geqslant 0\Rightarrow\bar{M}_\theta(bf+c)=\bar{M}_0\mathrm{L}_\theta(bf+c)=\bar{M}_0(b\mathrm{L}_\theta f+c)=b\bar{M}_0\mathrm{L}_\theta f+c=b\bar{M}_\theta f+c$, и доказана **А2**. Наконец, **А3** следует из соотношений $\bar{M}_\theta(f+g)=\bar{M}_0\mathrm{L}_\theta(f+g)\leqslant\bar{M}_0\mathrm{L}_\theta f+\bar{M}_0\mathrm{L}_\theta g=\bar{M}_\theta f+\bar{M}_\theta g$, что и доказывает утверждение.

Из свойств линейности и сохранения порядка оператора L_θ следует, что если первичным набором стандартной ИМ $\mathcal{M}_0=\langle\bar{M}_0\mathcal{G}\rangle$ является $\mathcal{G}$, то первичными наборами для $\mathcal{M}_\theta$ будут $\mathcal{G}_\theta=\{g_\theta:\mathrm{L}_\theta g_\theta=g,\ g\in\mathcal{G}\}$ со средними $\bar{M}_\theta g_\theta=\bar{M}_0 g$, где $\mathrm{L}_\theta g_\theta=g$.

Функциональные представления. Рассмотрим один частный случай предыдущего представления. Для этого обратимся к записи x в виде отображения $x=\mathrm{V}_\theta\xi$, где V_θ — известный оператор, зависящий от неизвестного параметра θ, принимающего значения из множества Θ. Такие записи обычны в задачах обнаружения и выделения сигналов, в которых шум ξ (вектор или процесс) действует в канале связи, описываемом оператором V_θ, где θ — неизвестный одномерный или многомерный параметр канала (или параметры сигнала и шума), а x — получаемые в результате векторные или в виде процесса наблюдения. Задана ИМ «шума» $\mathcal{M}_0^\xi$, играющая роль стандартной. Требуется составить ИМ $\mathcal{M}^x$ наблюдений.

Если оператор V_θ при каждом θ отображает «реализации» ξ в «реализации» x взаимно-однозначным образом, тогда L_θ, заданный так: $\mathrm{L}_\theta f(x)=f(\mathrm{V}_\theta\xi)$, будет удовлетворять посылам утверждения 1.4. Равенства $\bar{M}^x_\theta f(x)=\bar{M}^\xi_0 f(\mathrm{V}_\theta\xi)$ согласно (1.13) породят серию ИМ $\mathcal{M}_\theta{}^x$, объединение которых по θ даст модель $\mathcal{M}^x$ наблюдений, определяемую средними

$$\bar{M}f(x)=\sup_\theta\bar{M}^x_\theta f(x)=\sup_\theta\bar{M}^\xi_0 f(\mathrm{V}_\theta\xi).$$

Пример 1.15. В задачах радиолокации и связи смесь сигнала ω_t, где t есть время, с шумом ξ_t часто записывается в виде: $X_t=\theta\omega_t+\xi_t$, в котором $\theta\geqslant 0$ — неизвестная амплитуда сигнала. Пусть шум ξ_t описывается ИМ $\mathcal{M}_0$. Тогда модель $\mathcal{M}_\theta$ наблюдений X_t при каждом заданном θ определяется следующими значениями: $\bar{M}_\theta f(X_t)=\bar{M}_0 f(X_t-\theta\omega_t)$, а модель $\mathcal{M}$, равная объединению $\mathcal{M}_\theta$ по θ, — значениями $\bar{M}f=\sup_{\theta\geqslant 0}\bar{M}_0 f(X_t-\theta\omega_t)$. Они и дадут модель $\mathcal{M}$. Здесь, если $\mathcal{M}_0=\langle\bar{M}_0\mathcal{G}\rangle$ имеет первичным набором множество $\mathcal{G}$ функционалов, то первичными признаками $\mathcal{M}_\theta$ будут $g_\theta=g(X_t-\theta\omega_t)$ с теми же первичными средними, что и заданы на $\mathcal{G}$: $\bar{M}_\theta g(X_t-\theta\omega_t)=\bar{M}_0 g$.

Плотность. Рассматривается способ выразить одну ИМ через другую, стандартную, с помощью функции $p(x)$ переменной x.

Сначала рассмотрим наиболее простой случай, когда $\mathcal{X}=\mathcal{R}$, и на первичном кольце $\mathcal{K}_\Sigma$ всех отрезков задано конечно-адди-

тивное распределение $\mathscr{P}_1$ точными вероятностями $P_1[x, y)$, $x<y\in\mathscr{R}$. Плотность вероятностей по отношению к другому такому же $\mathscr{K}_\Sigma$-точному распределению $\mathscr{P}_0$ с вероятностями $P_0[x, y)$ задается формулой

$$p(x)=\lim_{y\downarrow x}\frac{P_1[x, y)}{P_0[x, y)}.$$

Плотность определена, если этот предел существует для всех x, кроме, может быть, множества точек $\mathscr{P}_0$-вероятности 0 (здесь $\mathscr{P}_0$ может вполне быть заменено на любую $\mathscr{K}_\Sigma$-точную меру, но удобней всего для этих целей мера длины, определенная как $P_0[a, b)=b-a$).

Смысл плотности вероятностей прозрачен сквозь призму аналогии с плотностью массы физического вещества, только слово «масса» заменяется на «меру» или вероятность. Точно такую же аналогию имеет плотность, когда $\mathscr{X}=\mathscr{R}^n$ (т. е. рассматривается случайный вектор). Тогда $p(x)$ есть предел отношения вероятностей непосредственных окрестностей точки x, в которых место отрезков занимают n-мерные брусы при неустанном уменьшении сторон брусов.

Если плотность $p(x)$ существует, то среднее по распределению $\mathscr{P}_1$ будет

$$M_1 g=\int g(x)\,p(x)\,dP_0(x)=M_0 gp$$

для всех интегрируемых по Риману — Стилтьесу функций $g(x)$, которые и образуют класс $\mathscr{L}\mathscr{K}_0$ вторичных признаков. Для любых же признаков f среднее $\overline{M}f$ определяется как верхняя грань Mg для всех $g\leqslant f$, $g\in\mathscr{L}\mathscr{K}_0$, откуда и следует формула

$$\overline{M}_1 f=\overline{M}_0 fp,\quad fp\in\mathscr{F},\tag{1.14}$$

где $\mathscr{F}$ —область существования средних для $\mathscr{P}_0$.

Из (1.14), если взять в качестве $f(x)$ индикаторную функцию отрезка $[x, y)$ (бруса) и устремить $y\downarrow x$, очевидно, будет следовать исходное определение плотности.

Возьмем (1.14) за основу формального определения плотности одной ИМ $\mathscr{M}_1$ (не обязательно точной) по отношению к другой $\mathscr{M}_0$. Считаем, что $\mathscr{M}_0$ имеет областью существования верхних средних класс $\mathscr{F}$ функций, а $\mathscr{M}_1$ — класс $\mathscr{F}_1=\{f/p: f\in\mathscr{F}\}\cup\mathscr{F}_0$, где $\mathscr{F}_0$ — все ограниченные сверху функции (добавление их к $\mathscr{F}_1$ нужно тогда, когда $p=\infty$ в некоторой области, так как в ней $f/p=0$).

Чтобы формула (1.14) задавала согласованные средние на признаках из $\mathscr{F}_1$, необходимо и достаточно, выполнения двух условий:

а) функция $p(x)$ неотрицательна: $p(x)\geqslant 0$;

б) среднее от функции $p(x)$ точное и равно 1:

$$\underline{M}_0 p=\overline{M}_0 p=1.$$

Функция $p(x)$, определенная уравнением (1.14), называется *формальной плотностью* $\mathscr{M}_1$ по $\mathscr{M}_0$ и обозначается $p(x)=\mathscr{M}_1/\mathscr{M}_0$.

Формальная плотность, если она существует, определена почти однозначно в том смысле, что если p_1 и p_2 — два варианта плотности, то событие, на котором они не равны друг другу, является $\mathscr{M}_0$-нулевым: $\bar{P}_0(p_1\neq p_2)=0$.

В самом деле, пусть $N=\{x: p_1\leqslant p_2\}$. Тогда из неравенств $\overline{M}_0(p_1-p_2)N\leqslant 0$ и $\overline{M}_0(p_1-p_2)N\geqslant\overline{M}_0p_1N-\overline{M}_0p_2N=0$ следует $M_0(p_1-p_2)N=0$, откуда $\bar{P}_0(N)=0$. Поменяв p_1 и p_2 местами, находим $\bar{P}_0(N')=0$, где $N'=\{x: p_1>p_2\}$. Объединяя N и N', получаем $\bar{P}_0(N\cup N')=0$, что и доказывает утверждение.

Событие N_0, нулевое для $\mathscr{M}$, обязано быть нулевым для $\mathscr{M}_1$: $\bar{P}_0(N_0)=0\Rightarrow\bar{P}_1(N_0)=\overline{M}_0N_0p=0$. При этом $p(x)$ на N_0 может в принципе быть любым, даже принимать значение $+\infty$ (так как $0\cdot\infty=0$). Более того, нулевым обязано быть событие на котором $p(x)=\infty$.

Если p есть формальная плотность $\mathscr{M}_1$ по $\mathscr{M}_0$ и $p>0$, то $1/p$ будет формальной плотностью $\mathscr{M}_0$ по $\mathscr{M}_1$:

$$p=\mathscr{M}_1/\mathscr{M}_0>0\Rightarrow 1/p=\mathscr{M}_0/\mathscr{M}_1.$$

Доказательство следует из равенства $\overline{M}_1(f/p)=\overline{M}_0(fp/p)=\overline{M}_0f$, и если $p=\infty$ при $x\in N$, то отношение p/p в этой области можно считать любым от 0 до ∞.

Теорема 1.5. *Пусть первичными для $\langle\overline{M}_0\mathscr{G}_0\rangle$ являются признаки набора $\mathscr{G}_0$ и $M_0p(x)=1$. Тогда первичными признаками для $\langle\overline{M}_1\mathscr{G}_1\rangle$, формальная плотность которой по отношению к $\langle\overline{M}_0\mathscr{G}_0\rangle$ существует и равна $p(x)$, будут*

$$\mathscr{G}_1=\{g_1(x): g_1(x)=g(x)/p(x),\quad g(x)\in\mathscr{G}_0\}\cup\{\pm 1/p(x)\}$$

со средними на них:

$$\tilde{M}_1(g/p)=\tilde{M}_0 g,\quad g\in\mathscr{G}_0;\quad \tilde{M}_1(1/p)=-\tilde{M}_1(-1/p)=1.$$

В самом деле, так как без ущерба $p(x)$ можно считать первичным признаком для $\langle\tilde{M}_0G_0\rangle$ со значением $M_0p=1$, то

$$\overline{M}_1 f=\overline{M}_0 fp=\inf\{[c+c_0M_0p+\sum c_i^+\tilde{M}_0g_i]: c+c_0p+\sum c_i^+g_i\geqslant fp\}=$$

$$=\inf\{[cM_1(1/p)+c_0+\sum c_i^+\tilde{M}_0g_i]: c/p+c_0+\sum c_i^+g_i/p\geqslant f\}.$$

Эта формула доказывает утверждение теоремы.

Таким образом, смысл формальной плотности $\mathscr{M}_1$ по отношению $\mathscr{M}_0$ состоит в пересчете вида первичных признаков $g_i\to g_i/p$, $p\to 1/p$, при одинаковых первичных средних, что может рассматриваться как взаимно-однозначное соответствие первичных признаков моделей. При этом достаточно рассматривать только те из них, первичные средние которых являются согласованными.

Рассмотрим следствия теоремы.

Следствие 1. *Область существования $\mathscr{F}_1$ верхних средних ИМ $\mathscr{M}_1$, формальная плотность которой по отношению к $\langle M_0\mathscr{G}_0\rangle$*

равна $p(x)$, составляет множество признаков, мажорируемых конечными линейными комбинациями вида $c+\sum c^+_i g_i/p$, $g_i\in\mathcal{G}_0$. При знаки из $\mathcal{F}_1$ представимы в виде $f=\sum c^+_i g_i/p+f_0$, где $g_i\in\mathcal{G}_0$, а f_0 — ограниченный сверху признак.

Следствие 2.

$$p=\mathcal{M}_1/\mathcal{M}_0=\mathcal{M}'_1/\mathcal{M}'_0\Rightarrow p=(\mathcal{M}_1\wedge\mathcal{M}'_1)/(\mathcal{M}_0\wedge\mathcal{M}'_0).$$

Можно сделать вывод, что для существования формальной плотности $p=\mathcal{M}_1/\mathcal{M}_0$ необходимо, чтобы $\mathcal{M}_0$ имела более богатый первичный набор, чем $\mathcal{M}_1$. Частное $\mathcal{G}/p$ приводит к потере данных о признаках набора $\mathcal{G}$ по крайней мере в той области, где $p(x)=0$. Отсюда размерность модели $\mathcal{M}_1$ должна быть никак не выше размерности $\mathcal{M}_0$. При $p(x)>0$, $\forall x$, размерности должны быть одинаковыми, а первичные наборы в известном смысле эквивалентными.

Рассмотрим примеры определения формальной плотности для ИМ, не являющейся точным распределением вероятностей.

Пример 1.15. Пусть $\mathcal{M}_0$ — моментная ИМ, определенная начальными моментами $\bar{M}_0X^i$, $i=1, \ldots, k$, верхними и потому неточными, кроме одного, $\bar{M}_0X^2=M_0X^2=\underset{\sim}{M_0X^2}$, с обязательно точным значением. Функция $p(x)=x^2$ будет формальной плотностью $x^2=\mathcal{M}_1/\mathcal{M}_0$ для такой $\mathcal{M}_1$, первичными средними которой являются $M_1(1/X^2)=1$, $\bar{M}_1(1/X)=\bar{M}_0X$, $\bar{M}_1X=\bar{M}_0X^3$, ..., $\bar{M}_1X^{k-2}=\bar{M}_0X^k$. Областью существования $\mathcal{F}_1$ будет множество признаков вида $c+\sum_1^k c^+_i X^{i-2}+f_0$, $f_0\in\mathcal{F}_0$. В этом примере, так как $p(x)>0$ при $x\neq 0$ функция $1/p(x)=1/x^2$ будет формальной плотностью $\mathcal{M}_0$ по отношению к $\mathcal{M}_1$, если исключить из числовой прямой, на которой они заданы, точку 0.

Отметим, что плотность одного ИРВ относительно другого может существовать лишь для точных распределений, так как требование $M_0p=M_1(1/p)=1$ исключает какие-либо отклонения.

Дополнения. 1. Доказательство теоремы о представлении. Докажем сначала теорему для сечения $\mathcal{M}$ одномерным параметром Mg. На основании определения операции объединения достаточно доказать равенство

$$\max_{\underline{Mg}\leqslant Mg\leqslant\overline{Mg}}\overline{M}_{Mg}f=\overline{\overline{M}}f.$$

Обозначим левую часть $\overline{\overline{M}}f$. Подставляя в нее формулу (1.9), получаем:

$$\overline{\overline{M}}f=\max_{\underline{Mg}\leqslant Mg\leqslant\overline{Mg}}\min_c[\overline{M}(f-cg)+cMg]=\max_{Mg}\min_c W(c,\ Mg).$$

Функция $W(c, Mg)$ линейна (и следовательно, вогнута) по Mg и выпукла по параметру c, так как для $0\leqslant\gamma\leqslant 1$:

$$W(\gamma c_1+(1-\gamma)c_2,\ \ Mg)=\overline{M}[\gamma f-\gamma c_1 g+(1-\gamma)f+(1-\gamma)c_2 g]+$$
$$+[\gamma c_1+(1-\gamma)c_2]Mg\leqslant\gamma\overline{M}(f-c_1 g)+\gamma c_1 Mg+$$
$$+(1-\gamma)\overline{M}(f-c_2 g)+(1-\gamma)c_2 Mg=\gamma W(c_1,\ Mg)+(1-\gamma)W(c_2,\ Mg).$$

Поэтому, используя известную теорему о минимаксе [22], без изменения результата максимум и минимум можно менять местами. Получаем

$$\overline{\overline{M}}f=\min_c\max_{Mg}W(c,\ Mg)=\min_c[\overline{M}(f-cg)+\max\{c\underline{M}g,\ c\overline{M}g\}],$$

и это равно $\bar{M}f$, поскольку минимум достигается при $c=0$ (в самом деле, при $c\geqslant 0$: $\bar{M}(f-cg)+\max\{c\underline{M}g,\ c\overline{M}g\}=\bar{M}(f-cg)+c\overline{M}g\geqslant\bar{M}f-c\overline{M}g+c\overline{M}g=\bar{M}f$ и то же самое неравенство справедливо при $c\leqslant 0$). Теорема доказана для случая, когда $\mathcal{G}$ состоит всего из одного признака.

По индукции теорема распространяется на случай, когда $\mathcal{G}=\mathcal{G}_k$ есть конечный набор. Наконец, для произвольного набора $\mathcal{G}$, если обозначить

$$\overline{\overline{M}}f=\sup_{M\mathcal{G}}\overline{M}_{M\mathcal{G}}f=\sup_{M\mathcal{G}}\inf_{\mathcal{G}_k\subset\mathcal{G}}\overline{M}_{M\mathcal{G}_k}f,$$

то нужно доказать равенство $\overline{\overline{M}}f=\bar{M}f$. Признаку f такому, что $|\bar{M}f|<\infty$ (откуда $|\bar{M}_{M\mathcal{G}}f|<\infty$), и фиксированному $\varepsilon>0$ всегда найдутся такие k' и конечный набор $\mathcal{G}_{k'}$, что

$$\inf_{\mathcal{G}_k\subset\mathcal{G}}\overline{M}_{M\mathcal{G}_k}f\geqslant\overline{M}_{M\mathcal{G}_{k'}}f-\varepsilon.$$

Отсюда

$$\overline{\overline{M}}f\geqslant\sup_{M\mathcal{G}_{k'}}\overline{M}_{M\mathcal{G}_k}f-\varepsilon=\overline{M}f-\varepsilon.$$

В то же время $\mathcal{M}_{M\mathcal{G}}\subset\mathcal{M}$, откуда получается $\bigvee_{M\mathcal{G}}\mathcal{M}_{M\mathcal{G}}\subset\mathcal{M}$ и $\overline{\overline{M}}f\leqslant\bar{M}f$. Окончательно, для любого признака f из области существования средних $\mathcal{F}$ имеем $|\overline{\overline{M}}f-\bar{M}f|\leqslant\varepsilon$, и утверждение теоремы следует из произвольности ε. Доказательство закончено.

2. Пример пересчета по формуле (1.13). Пусть $y=V_\theta(x)$ есть взаимно-однозначное отображение $\mathcal{X}$ в $\mathcal{X}$: $\mathcal{X}\xrightarrow{V_\theta}\mathcal{X}$. Тогда оператор $L_\theta f(x)=f(V_\theta(x))$, очевидно, будет обладать требуемыми свойствами а) и б) утверждения 1.4. В линейных пространствах $\mathcal{X}$ (например, если $\mathcal{X}=\mathcal{R}^n$) такого типа операторами являются преобразования сдвига $L_\theta f(x)=f(x-b_\theta x_0)$, $b_\theta\in\mathcal{R}$, где элемент x_0 характеризует направление сдвига. Подставляя в (1.13), получаем $\bar{M}_\theta f(x)=\bar{M}_0 f(x-b_\theta x_0)$. Здесь $\mathcal{M}_\theta$ выводится из $\mathcal{M}_0$ сдвигом всех $x\in\mathcal{X}$ на величину $b_\theta x_0$.

1.6. УСЛОВНЫЕ ИНТЕРВАЛЬНЫЕ МОДЕЛИ

Постановка проблемы. Интервальные модели дают описания явлений, еще не происшедших. Допустим, что такое безусловное описание $\mathcal{M}$ составлено в виде совокупности согласованных средних $\bar{M}f$, $f\in\mathcal{F}$. А потом вдруг дополнительно стало известно, что произошло событие B. Оно частично, но не полностью отражает результат явления, если только не является элементарным событием. Тогда после B неопределенность останется, а явление будет описываться новой моделью $\mathcal{M}_B$, называемой *условной* при

случившемся B. Условной модели соответствуют свои средние $\overline{M}_B f$.

Наличие достоверно происшедшего события B для условной модели соответствует вероятности $\underline{M}_B B(x) = \underline{P}_B(B) = 1$. Но не только это. Производится пересчет абсолютно всех средних $\overline{M} f$ в $\overline{M}_B f$, своего рода преобразование одних в другие, и как их часть — вероятностей $\overline{P}(A)$ в $\overline{P}_B(A)$. Вопрос, как?

Если событие B имеет точные вероятности $P(AB)$ и $P(B)$, то пересчет в условные хорошо известен и производится элементарно делением

$$P_B(A) = P(AB)/P(B).$$

А если неточные? Если неточной является только вероятность в числителе, то для расчета $\overline{P}_B(A)$ нужно подставить вместо нее $\overline{P}(AB)$, что соответствует супремуму отношения по $P(AB)$. По такому же принципу рассчитываются средние любых признаков.

Но как определить условные средние и вероятности, когда в знаменателе вероятность $P(B)$ является интервальной $\underline{P}(B)$, $\overline{P}(B)$? Основная трудность здесь в том, что если брать супремум правой части, то $P(AB)$ и $P(B)$ оказываются связанными: нельзя положить одновременно $P(B) = \underline{P}(B)$ и $P(AB) = \overline{P}(AB)$ (равно, как абсурдно желание, заполняя «вероятностной массой» до краев «отсек» AB, обеспечить при этом минимальное суммарное заполнение обоих отсеков AB и A^cB, составляющих B).

Определение условной интервальной модели. Начнем с рассуждений, которые и приведут нас к определению. Пусть $\mathcal{M}_{P_*(B)}$ есть $P_*(B)$-сечение $\mathcal{M}$. Вероятность события B для $\mathcal{M}_{P_*(B)}$ является точной, равной $P_*(B)$, и пусть $P_*(B) > 0$. Тогда соответствующие сечениям $\mathcal{M}_{P_*(B)}$ условные ИМ $\mathcal{M}_{P_*(B),B}$ получаются сужением области существования к признакам $f(x)B(x)$, $f \in \mathcal{F}$, где $\mathcal{F}$ — область существования $\overline{M} f$, и нормировкой всех средних на число $1/P_*(B)$, что дает

$$\overline{M}_{P_*(B),B} f = \overline{M} fB/P_*(B), \quad f \in \mathcal{F}.$$

Очевидно, определенные так средние удовлетворяют аксиомам ИМ.

Любая ИМ согласно теореме 1.3 о представлении записывается как объединение ее $P_*(B)$-сечений. Для каждого из сечений по указанной формуле находится условная ИМ, а их объединение и даст искомую условную ИМ.

Условной при случившемся событии B называется ИМ $\mathcal{M}_B$ на $\mathscr{X}$, определяемая из $\mathcal{M}$ средними:

$$\overline{M}_B f = \max_{\underline{P}(B) \leqslant P_*(B) \leqslant \overline{P}(B)} [\overline{M}_{P(B)} fB/P_*(B)], \quad f \in \mathcal{F}, \tag{1.15}$$

где $\overline{M}_{P_*(B)} fB$ есть средние для $P(B)$-сечений $\mathcal{M}_{P_*(B)} = \mathcal{M} \wedge \langle P_*(B) \rangle$ от признаков $f(x)B(x)$.

Средние по формуле (1.15) называются условными. Без труда проверяется, что они удовлетворяют аксиомам ИМ. Областью существования верхних средних условной ИМ будет класс признаков, совпадающих с функциями $\mathcal{F}$ на событии B, и произвольных (возможно, принимающих значения $\pm\infty$) вне этого события. Этот класс обозначаем $\mathcal{F}B$.

Нижние средние условной ИМ определяются по обычной формуле $\underline{M}_B f = -\overline{M}_B(-f)$.

Если $\underline{P}(B) = 0$, то правая часть (1.15) при максимуме, когда $P_*(B) = 0$, понимается как предел $P_*(B) \to 0$. Так же понимается эта формула, если событие B является нулевым, т. е. $\overline{P}(B) = 0$. Тогда

$$\overline{M}_B f = \lim_{P_*(B) \to 0} \frac{\overline{M}_{P_*(B)} fB}{P_*(B)} = \lim_{P_*(B) \to 0} \frac{P_*(B) \max\limits_x (fB)}{P_*(B)} = \max_x fB,$$

где при любых f (и неограниченных также) функция fB считается равной нулю в области B^c. Отсюда следует, что если событие B является невозможным (в частности, $B = \varnothing$), то условная ИМ $\mathcal{M}_B$ будет B-индикаторной $\mathcal{M}_B = \mathcal{I}_B$, а именно определяемой на $\mathscr{X}$ единственным первичным средним $\overline{P}_B(B^c) = 0$.

Подстановка в (1.15) формулы (1.9) § 1.5 для средних сечения раскрывает выражение для условных средних:

$$\overline{M}_B f = \max_{\underline{P}(B) \leqslant P_*(B) \leqslant \overline{P}(B)} \frac{\min\limits_c [\overline{M}(f - c)B + cP_*(B)]}{P_*(B)}. \tag{1.16}$$

Дадим примеры расчета по этой формуле.

Пример 1.16. Расчет условных вероятностей ИРВ. Пусть f в (1.16) есть $A(x)$. Тогда минимум по c в квадратных скобках выражения (1.16) будет достигаться либо при $c=0$, либо при $c=1$, в результате чего $\overline{M}_{P_*(B)} AB = \min\{\overline{P}(AB), P_*(B) - \underline{P}(A^cB)\}$. Подстановкой данной формулы в (1.15) находятся верхние условные вероятности: $\overline{P}_B(A) = \overline{P}_B(AB) = \overline{M}_B AB =$

$$= \max_{\underline{P}(B) \leqslant P_*(B) \leqslant \overline{P}(B)} \min \left\{ \frac{\overline{P}(AB)}{P_*(B)}, \; 1 - \frac{\underline{P}(A^cB)}{P_*(B)} \right\} = \frac{\overline{P}(AB)}{\overline{P}(AB) + \underline{P}(A^cB)},$$

где использован тот факт, что максимум достигается при равенстве членов под знаком минимума.

Пример 1.17. Пусть $\mathcal{M}$ на $\mathscr{R}$ определена двумя первичными средними $\underline{M}|X|$, $\overline{M}|X|$ и пусть $B = [a, b]$, $0 < a < b$. Без труда находятся $\underline{P}(B) = 0$ $\overline{P}(B) = \min\{1, \overline{M}|X|/a\}$. При заданном $P_*(B)$ для сечения имеем

$$\overline{M}_{P_*(B)} fB = \inf_{c + (c_1^+ - c_2^+)|x| + c_3 B(x) \geqslant f(x)} [c + c_1^+ \overline{M}|X| - c_2^+ \underline{M}|X| + c_3 P_*(B)].$$

Например, для события $D = [d_1, d_2] \subset B$ по этой формуле получаем

$$\overline{M}_{P_*(B)} D = \min\{\overline{M}|X|/c_1, \; P_*(B)\}, \quad \underline{M}_{P_*(B)} D = 0,$$

откуда, если $D \neq B$ и $D \neq \varnothing$, находим $\underline{M}_B D = 0$ и

$$\overline{M}_B D = \overline{P}_B(D) = \max_{P_*(B)} \min\left\{\frac{\overline{M}|X|}{d_1 P_*(B)},\ 1\right\} = \min\left\{\frac{\overline{M}|X|}{d_1 \underline{P}(B)},\ 1\right\} = 1.$$

Таким образом, границы для всех условных вероятностей событий в данном примере тривиальны. В случае $f(x) = |x|$ имеем:

$$\underline{M}_B|X| = \min_{0 \leqslant P_*(B) \leqslant \overline{P}(B)} \underline{M}|X|/P_*(B) = \underline{M}|X|/\overline{P}(B) = \frac{\underline{M}|X|}{\min\{\overline{M}|X|/a,\ 1\}};$$

$$\overline{M}_B|X| = \max_{0 \leqslant P_*(B) \leqslant \overline{P}(B)} \overline{M}|X|/P_*(B) = \infty,$$

и, в частности, при $\overline{M}|X| > a$ получаем $\underline{M}_B|X| = a\underline{M}|X|/\overline{M}|X|$.

Для другого события $B_1 = [-a,\ a]$ при тех же первичных средних имеем $\underline{P}(B_1) = [1 - \overline{M}|X|/a]^+$, $\overline{P}(B_1) = 1$, и если это событие произошло, то условные вероятности уже не будут, как выше, тривиальными, так как, например, для события $D_1 = [-d,\ d]$, $d < a$:

$$\underline{P}_{B_1}(D_1) = [1 - \overline{M}|X|/d]^+/[1 - \overline{M}|X|/a]^+,\ \overline{P}_{B_1}(D_1) = 1.$$

Без труда находятся средние от функции $|X|$:

$$\underline{M}_{B_1}|X| = \underline{M}|X|,\quad \overline{M}_{B_1}|X| = 1/[1/\overline{M}|X| - 1/a]^+.$$

Расчет условных моделей через вершины. Здесь рассматривается частный случай, но он позволяет нарисовать ясную наглядную картину условных моделей.

Пусть пространство $\mathscr{X} = \{x_1, \ldots, x_r\}$ конечно и $\mathscr{M}$ есть ИМ на нем. Обозначим $\mathbf{P}_\theta = (P_\theta(x_1), \ldots, P_\theta(x_r))$ — все векторы вероятностей, являющиеся вершинами (крайними точками) $\mathscr{M}$, так что $\mathscr{M} = \bigvee \mathbf{P}_\theta$. Соответствующие каждому $\mathbf{P}_\theta$ вероятности $P_\theta(B) = \sum_{x_i \in B} P_\theta(x_i)$ являются точными, потому объединение перегруппируется как запись $\mathscr{M}$ через $P(B)$-сечения, если каждое сечение представить набором вершин — векторов $\mathbf{P}_\theta$ с одинаковыми $P_\theta(B) = P(B)$.

Условная ИМ $\mathscr{M}_B$, полученная из $\mathscr{M}$, если стало известно, что событие B произошло, есть выпуклая оболочка (объединение) преобразованных к условным вершин, полученных делением компонент $\mathbf{P}_\theta$ на $P_\theta(B)$ с обнулением компонент $P_\theta(x_i)$, для которых $x_i \notin B$:

$$\mathscr{M}_B = \bigvee_{\theta:\ P_\theta(B) > 0} \mathbf{P}^B_\theta / P_\theta(B),$$

где $\mathbf{P}^B{}_\theta$ есть вектор с компонентами $P^B_\theta(x_i) = \begin{cases} P_\theta(x_i), & x_i \in B, \\ 0, & x_i \notin B, \end{cases}$

а объединение производится по тем вершинам, относительно которых вероятность B является ненулевой.

Геометрически получение условной ИМ выглядит следующим образом. Сначала каждая вершина $\mathbf{P}_\theta$ редуцируется к вектору

$\mathbf{P}^B{}_\theta$ обнулением всех компонент $P_\theta(x_i)$, для которых $x_i \notin B$. Эту редукцию можно интерпретировать как проекцию векторов $\mathbf{P}_\theta$ в соответствующее подпространство $\mathscr{R}^r{}_B$ пространства $\mathscr{R}^r$. Далее из начала координат в направлении векторов $\mathbf{P}^B{}_\theta$ проводятся лучи $\lambda P^B{}_\theta$, $\lambda \geqslant 0$, до пересечения с гиперплоскостью $\sum_{x_i \in B} P(x_i) = 1$ подпространства $\mathscr{R}^r{}_B$. Пересечение достигается при $\lambda = 1/P_\theta(B)$, и точки пересечения $P^B{}_\theta/P_\theta(B)$ дадут векторы условных вероятностей. Иллюстрация сказанного для $\mathscr{X} = \{x_1, x_2, x_3\}$ и $B = \{x_1, x_2\}$ приведена на рис. 1.11. Условной ИМ $\mathscr{M}_B$ здесь соответствует отрезок $[P_{1B},\ P_{3B}]$. Заметим, что число вершин $\mathscr{M}_B$ меньше, чем у $\mathscr{M}$: точка $\mathbf{P}_2$ является вершиной $\mathscr{M}$, но $\mathbf{P}_{2B}$ не есть вершина $\mathscr{M}_B$.

Мы показали, что условные ИМ записываются через вершины, если такую запись допускают безусловные. Если же $\mathscr{M}$ определена своими первичными средними $\overline{M}g$, $g \in \mathscr{G}$, то теоретически можно найти условную ИМ $\mathscr{M}_B$, сводя $\mathscr{M}$ к ее вершинам. Для этого нужно проделать следующую последовательность действий: первичные средние $\overline{M}\mathscr{G} \to$ согласованные грани $\mathscr{M} \to$ вершины $\mathscr{M} \to$ $\to$ вершины условной ИМ $\to$ первичные средние $\mathscr{M}_B$. Этот путь весьма долог, поскольку каждая из операций достаточно трудоемка. Но дело даже не в этом, а в том, что для бесконечных пространств $\mathscr{X}$ понятие вершины как вектора вероятностей не имеет смысла, что существенно ограничивает универсальность такой процедуры.

Некоторые свойства условных интервальных моделей. Для условных ИМ $\mathscr{M}_B$ событие B является достоверным $P_B(B) = 1$, а B^c — нулевым $\overline{P}_B(B^c) = 0$. С учетом этого условные ИМ обладают всеми свойствами ИМ, заданной на $\mathscr{X}$.

Рассмотрим, как будет меняться условная ИМ, если сужать или расширять событие B. Очевидно, в крайних случаях, когда событие B достоверно, т. е. $\underline{P}(B) = 1$, условная ИМ совпадает с исходной $\mathscr{M}_B = \mathscr{M}$, а если событие B является невозможным, т. е. $\overline{P}(B) = 0$, то условная ИМ будет B-индикаторной: $\mathscr{M}_B = \mathscr{I}_B$. Оказывается, последнее равенство имеет место не обязательно только для невозможных событий.

Свойство 1. *Если каждая из первичных функций $g \in \mathscr{G}$ безусловной модели $\mathscr{M} = \langle \overline{M}\mathscr{G} \rangle$ принимает на B постоянные значения, то условная к ней $\mathscr{M}_B$ будет B-индикаторной: $\mathscr{M}_B = \mathscr{I}_B$.*

В самом деле, в этом случае $\overline{M}_{P_*(B)}fB = P_*(B) \max fB$ и свойство непосредственно следует из определения условной ИМ.

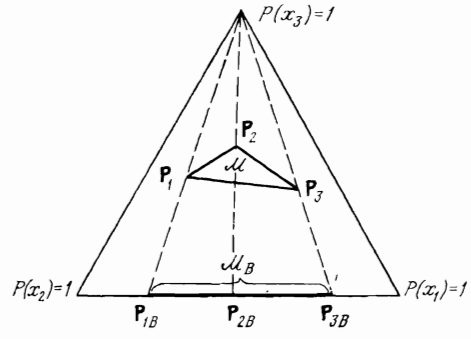


Рис. 1.11. Геометрическая иллюстрация условных моделей

Свойство 2. *Условная ИМ, соответствующая голой мо-*

дели $\mathscr{I}$, будет B-индикаторной $\mathscr{I}_B$, каковым бы условие B ни было.

Свойство 3. *Условная к объединению $\mathscr{M}=\mathscr{M}_1\vee\mathscr{M}_2$ модель равна объединению условных ИМ: $\mathscr{M}_B=\mathscr{M}_{1B}\vee\mathscr{M}_{2B}$.*

Свойство 4. $\mathscr{M}_1\subset\mathscr{M}_2\Rightarrow\mathscr{M}_{1B}\subset\mathscr{M}_{2B}$ — *при переходе к условным сохраняется операция включения.*

Свойство 3 распространяется на объединение произвольного числа моделей. Собственно, это свойство и было положено в основу определения условных моделей, когда от объединения сечений был сделан шаг к объединению условных к ним моделей. Свойство 4 есть прямое следствие 3, так как $\mathscr{M}_1\subset\mathscr{M}_2\Leftrightarrow\mathscr{M}_1\vee\vee\mathscr{M}_2=\mathscr{M}_2$.

Для иллюстрации восстановим наглядную картину, нарисованную на рис. 1.11. Считая $\mathscr{X}$ дискретным, представим себе $\mathscr{M}$ как выпуклое тело в пространство векторов вероятностей, и поместим точечный источник в начало координат. Тело $\mathscr{M}$ при освещении его источником бросает тень на гиперплоскость $\sum\limits_{x\in B}P(x)=P(B)=1$. Эта тень и даст условную ИМ. На основании такого представления получают наглядную интерпретацию свойств 3 и 4. Тень от выпуклого объединения тел будет равна объединению их теней, что иллюстрирует свойство 3. Но вот тень от пересечения тел не будет равна пересечению их теней. Два тела могут не пересекаться, но при освещении точечным источником бросать одинаковую тень на гиперплоскость. Отсюда — следующие два предложения.

Свойство 5. *Пересечению ИМ не будет, в общем, соответствовать пересечение условных ИМ.*

Свойство 6. *Двум различным безусловным ИМ может соответствовать одна и та же условная.*

Например, пусть $B(x)$ есть единственный первичный признак как для модели $\mathscr{M}_1$, так и для $\mathscr{M}_2$, заданных своими точными вероятностями $P_1(B)$, $P_2(B)\neq P_1(B)$. Тогда $\mathscr{M}_1$ и $\mathscr{M}_2$ не пересекаются между собой, но им соответствует одна и та же B-индикаторная условная ИМ $\mathscr{I}_B$.

О восстановлении безусловной модели по условным. Здесь будет обсуждаться вопрос о возможности восстановления $\mathscr{M}$ по набору соответствующих ей условных $\mathscr{M}_{B_i}$, где $\mathscr{B}_\Sigma=\{B_1, \ldots, B_k\}$ есть дробление пространства $\mathscr{X}$ на непересекающиеся события и $\mathscr{X}=\sum B_i$.

Пусть сначала вероятности событий B_i считаются точными для $\mathscr{M}$: $\underline{P}(B_i)=\overline{P}(B_i)=P(B_i)$, $i=1, \ldots, k$. Тогда $\overline{M}_{P(B_i)}fB_i=\overline{M}fB_i$ и на основании этого устанавливается справедливость неравенства

$$\overline{M}f\leqslant\sum\overline{M}fB_i=\sum P(B_i)\,\overline{M}_{B_i}f.$$

Правая часть есть усреднение условных средних $\overline{M}_{B_i}f$ по точному распределению вероятностей $P(B_i)$ событий B_i. Равенство левой и правой частей (соответствующее известной формуле полной вероятности) достигается только тогда, когда верхнее среднее $\overline{M}$ аддитивно на событиях B_i: $\overline{M}\sum fB_i=\sum\overline{M}fB_i$.

Перейдем теперь к более общему случаю, когда вероятности $P(B_i)$ не являются точными. Запишем $\sum P(B_i)\overline{M}_{B_i}f=M[\sum B_i(x)\overline{M}_{B_i}f]$, и заменим символ точного среднего на верхнее по $\mathscr{M}$. Это даст значения

$$\overline{\overline{M}}f=\overline{M}[\sum B_i\,\overline{M}_{B_i}f],\quad f\in\mathscr{F},$$

определяющие новую ИМ, обозначаемую $\overline{\overline{M}}$. Таким образом, *имеет место включение $\overline{\overline{M}}\supset\mathscr{M}$. А если $\mathscr{M}=\langle\overline{M}\mathscr{G}\rangle$ и все функции набора $\mathscr{G}$ $\mathscr{B}_\Sigma$-измеримы* (т. е. являются конечными линейными комбинациями $B_i(x)$ или их замыканиями), *то $\overline{\overline{M}}=\mathscr{M}$.*

Выделенный курсивом тезис станет понятным, если отметить, что для вычисления $\overline{M}f$ нужно знать как $\overline{M}_{B_i}f$, т. е. условные модели $\mathscr{M}_{B_i}$, так и средние всевозможных линейных комбинаций B_i: $\overline{M}\sum c_iB_i$, что (если взять первичным класс $\mathscr{L}\mathscr{B}_\Sigma$ этих комбинаций) составит $\mathscr{L}\mathscr{B}_\Sigma$-расширение $\mathscr{M}$. Если по условию второго предложения тезиса все $g\in\mathscr{G}$ являются $\mathscr{B}_\Sigma$-измеримыми, то $\mathscr{G}\subset\mathscr{L}\mathscr{B}_\Sigma$ и $\mathscr{L}\mathscr{B}_\Sigma$-расширение совпадает с исходной ИМ $\mathscr{M}$, при этом условные ИМ вырождаются $\mathscr{M}_{B_i}=\mathscr{I}_{B_i}$ в B_i-индикаторные.

Таким образом, сфера действия тезиса о восстановлении сводится к вырожденным условным моделям $\mathscr{M}_{B_i}=\mathscr{I}_{B_i}$.

В общем же случае при переходе к условным моделям происходит утеря данных об $\mathscr{M}$, так что восстановить можно будет не ее, а только более широкую (и следовательно, менее точную) модель $\overline{\overline{M}}$. Это ограничивает область приложения условных моделей в основном точными распределениями вероятностей.

Абстрактно-условные модели. В предыдущем изложении считалось, что произошло некоторое событие B и исследовалась условная модель $\mathscr{M}_B$. Совершенно формально пока мы заменим индикаторный $B(x)$ на произвольный признак $q(x)$ и будем по тем же формулам определять условную ИМ $\mathscr{M}_q$. Формула (1.15) позволяет это сделать. Как говорилось в начале § 1.1, $q(x)$ при $0\leqslant q(x)\leqslant 1$ может трактоваться как признак нечеткого события q, и тогда $\mathscr{M}_q$ будет условной ИМ, если такое событие произошло.

Пусть $q(x)$ есть функция на $\mathscr{X}$ такая, что $Mq>0$. *Абстрактно-условной*, соответствующей тому, что q произошло, называется модель $\mathscr{M}_q$, определяемая средними

$$\overline{M}_qf=\max_{\underline{M}q\leqslant M_*q\leqslant\overline{M}q}\frac{\overline{M}_{M_*q}(fq)}{M_*q}. \tag{1.17}$$

В числителе правой части (1.17) стоят средние M_*q-сечения $\mathscr{M}$.

Очевидно такое свойство: $\mathscr{M}_{c^+q}=\mathscr{M}_q$ — умножение $q(x)$ на неотрицательный коэффициент c^+ не меняет абстрактно-условной ИМ.

Пусть $p(x)$ есть плотность $\mathcal{M}$ по отношению к $\mathcal{M}^\circ$: $p=\mathcal{M}/\mathcal{M}^\circ$. Тогда соответствующая $\mathcal{M}$ абстрактно-условная модель $\mathcal{M}_q$ при случившемся $q(x)$ будет равна абстрактно-условной модели $\mathcal{M}^\circ_{qp}$ при случившемся $q(x)p(x)$:

$$p=\mathcal{M}/\mathcal{M}^\circ,\ \underline{M}q>0\Rightarrow\mathcal{M}_q=\mathcal{M}^\circ_{qp}.$$

В самом деле по определению плотности $\overline{M}f=\overline{M}^\circ fp$. Записывая (1.17) в развернутой форме, подставим формулу для M_*q-сечения и после несложных преобразований получаем

$$\overline{M}_q f=\max_{\underline{M}q\leqslant M_*q\leqslant\overline{M}q}\frac{\min\limits_c[\overline{M}(f-c)q+cM_*q]}{M_*q}=$$

$$=\max_{\underline{M}^\circ qp\leqslant M_*qp\leqslant\overline{M}^\circ qp}\frac{\min\limits_c[\overline{M}^\circ(f-c)qp+cM_*qp]}{M^*qp}=\overline{M}^\circ_{qp}f,$$

что и требовалось.

Из доказанного утверждения вытекает. *Если $p(x)$ есть плотность $\mathcal{M}$ по отношению к $\mathcal{M}^\circ$, то условная модель $\mathcal{M}_B$, соответствующая случившемуся событию $B\subset\mathscr{X}$, равна абстрактно-условной $\mathcal{M}^\circ_{Bp}$ при случившемся $B(x)p(x)=q(x)$:*

$$p=\mathcal{M}/\mathcal{M}^\circ\Rightarrow\mathcal{M}_B=\mathcal{M}^\circ_{Bp},\quad B\subset\mathscr{X}.$$

При $B=\mathscr{X}$ просто получается $\mathcal{M}=\mathcal{M}^\circ_p$.

Оба последних тезиса, любопытных в математическом аспекте, ищут наглядной интерпретации.

Таким образом, вводимое в настоящем параграфе понятие условной модели без труда распространяется на нечеткие события. Дадим иллюстрацию.

Формула условной вероятности (Байеса) для нечетких событий. Пусть для ИРВ требуется рассчитать вероятность A, если B случилось не достоверно, а с некоторым сомнением, произошло ли оно вообще. Имеем вместо B признак $q=\gamma B+1(1-\gamma)B^c$ (нечеткое событие), где γ есть коэффициент, интерпретируемый как вероятность того, что B произошло. Расчеты по формуле (1.17) дают после ряда вычислений следующее выражение, пригодное при $1/2\leqslant\gamma\leqslant1$, для вероятности A при свершившемся q:

$$\overline{P}_q(A)=\frac{(1-\gamma)\overline{P}(A)+(2\gamma-1)\overline{P}(AB)}{(1-\gamma)+(2\gamma-1)[\overline{P}(AB)+\underline{P}(A^cB)]}.$$

Как видно, при $\gamma=1$ результат тот же, что и в примере 1.16, а при $\gamma=1/2$ условной вероятностью станет априорная вероятность (безусловная) $\overline{P}_q(A)=\overline{P}(A)$ события A. Это и понятно, так как значение $\gamma=1/2$ эквивалентно $q\equiv1/2$, что в свою очередь заменимо на $q\equiv1$ (поскольку умножение q на константу не меняет абстрактно-условной вероятности) и ведет к достоверному событию $B=\mathscr{X}$. Формула для условной вероятности при $0\leqslant\gamma\leqslant1/2$

(соответствующей тому, что с преобладающей верой произошло даже B^c, а не B) получается из приведенной заменой B на B^c и γ — на $1-\gamma$, а для нижней условной вероятности — переменой нижних вероятностей с верхними и наоборот.

1.7. ЗАКЛЮЧЕНИЕ

Рассматриваются случайные явления, в описании которых прямо или косвенно может быть указано множество взаимно исключающих друг друга элементарных исходов, образующих пространство элементарных событий. Функции на этом пространстве называются признаками. Средние статистические значения признаков есть пределы средних арифметических результатов независимых повторов явления в одинаковых условиях и могут быть точными (для устойчивых явлений) и интервальными (для неустойчивых, неопределенных). Средние в интервальном понимании существуют в очень широкой области признаков, куда входят обязательно все ограниченные.

Интервальная модель есть совокупность нижних и верхних средних в области их существования, связанных между собой аксиомами § 1.1. Аксиомы согласуют средние между собой.

Любая ИМ (§ 1.2) формируется первичным набором $\mathscr{I}$ признаков, элементы которого предопределят тип модели, и непротиворечивым (корректным) заданием на $\mathscr{I}$ первичных средних, конкретизирующих ее вид. Ключевой является теорема 1.1, согласно которой средние с первичных признаков однозначно продолжаются с согласованием на все признаки, мажорируемые первичными, образуя область существования ИМ. Если среди первичных признаков имеются неограниченные, то они породят неограниченные признаки в области существования. Дополнительное расширение области существования может производиться предельным переходом (1.4) от усеченных сверху и снизу функций подобно тому, как понимается интеграл от неограниченных функций.

Интервальные модели отличаются друг от друга разными наборами первичных признаков и разными первичными средними, а в итоге — разными значениями средних в области их существования. По включениям этих значений можно судить, какая из ИМ более широкая, а какая менее (§ 1.3). Чем шире ИМ, тем меньше полезных данных о явлении в ней содержатся. Самой широкой среди всех является голая ИМ, соответствующая абсолютному отсутствию данных (или полнейшей неустойчивости явления). Расширение ИМ служит рабочим инструментом ее упрощений.

Через средние в § 1.3 определяются операции пересечения ИМ как добавление данных к уже имеющимся и объединения как рассеяние данных, рост неопределенности. Геометрически ИМ есть многогранники, в которых первичные признаки и их средние определяют соответственно направления граней и их положения. Пересечение многогранников есть их общая часть, поэтому будет снова многогранником, а объединение полагается расширить до его выпуклой оболочки, чтобы получилась ИМ.

Частным случаем ИМ, когда первичными взяты вероятности событий, являются интервальные распределения вероятностей, описанные в § 1.4. Первичный набор событий ИРВ произволен как по количеству, так и топологии. Если это полукольцо (например, отрезки числовой прямой) и на нем заданы точные вероятности, то они однозначно продолжаются, оставаясь точными, на

алгебру событий (суммы отрезков), а их согласованность равносильна аддитивности, что ведет к конечно-аддитивным распределениям вероятностей. Расширение первичного набора до счетной алгебры и задание вероятностей сознательно счетно-аддитивными ведет к сужению распределений вероятностей до счетно-аддитивных.

Обобщением указанных типов распределений вероятностей являются конечно- и счетно-аддитивные ИРВ, у которых при той же первичной системе событий вероятности заданы интервальными. Еще одним типом ИРВ является интервальная функция распределения вероятностей, первичной для которой является набор вкладывающихся событий; представима как семейство точных функций распределения, причем непрерывные из них есть конечно-аддитивные распределения вероятностей, а разрывные соответствуют их группам.

Любое семейство распределений вероятностей (конечно-аддитивных или счетно) суть некоторая ИМ. С другой стороны, любая ИМ с интервальными средними представима как объединение ее простых составляющих с точными средними (теорема 1.3 § 1.5). В частности, это может быть семейство конечно-аддитивных распределений вероятностей. Но не счетно-аддитивных как слишком специальных, чтобы стать универсальным «строительным материалом» для всех ИМ (исключая дискретные пространства элементарных событий).

Иная ИМ может приобрести наглядность, а подчас и физическую осмысленность правильным подбором ее представления через некоторые стандартные модели. Это может быть сделано удачным выбором вида соответствующего функционального преобразования. Еще один специальный способ состоит в записи средних одной ИМ через другую (стандартную) с помощью формальной плотности, понимаемой шире классической плотности вероятностей (§ 1.5).

При наличии достоверно свершившегося события ИМ трансформируется в условную пересчетом ее средних (§ 1.6). Формула (1.15) пересчета весьма сложна (кроме случая точных вероятностей). Она приложима к нечетким событиям. Переход к условным моделям сопровождается обычно расширением. Вернуться от условных к исходной модели, ничего не потеряв, можно лишь в редких исключениях, поэтому условные модели не занимают сколь-либо значительного места в интервальных методах.

Глава 2.

СОВМЕСТНЫЙ АНАЛИЗ

2.1. ДЕТЕРМИНИРОВАННЫЕ ПРЕОБРАЗОВАНИЯ ИСХОДОВ

Отображения. Определив интервальные модели и их свойства, пойдем дальше. Посмотрим, как они видоизменяются в смысле деформации признаков и их средних при преобразованиях пространства $\mathscr{X}$. Для этого классифицируем сначала сами преобразования.

Преобразование s пространства $\mathscr{X}$ в $\mathscr{Y}$

$$\mathscr{X} \xrightarrow{s} \mathscr{Y}$$

есть математическая запись реальных отношений между исходами на $\mathscr{X}$ и на $\mathscr{Y}$. Если $y=sx$ — функция, однозначно отображающая каждый исход $x\in\mathscr{X}$ в исход $y\in\mathscr{Y}$, то преобразование называется *детерминированным*.

Указав, как детерминированное s преобразует исходы, рассмотрим, как при этом преобразуются признаки, но не все, а пока только их частный класс — события. Здесь нет трудностей: $A\subset\mathscr{X}$ преобразуется в $B=\{y: y=sx,\ x\in A\}$ — это множество значений, которые принимает $y=sx$, когда x пробегает значения из A. Пишем формально $B=sA$ и будем называть B *образом* события A. Каждое событие A на $\mathscr{X}$ имеет свой образ B на $\mathscr{Y}$.

Свойства образов. Образ объединения двух и более событий из $\mathscr{X}$ равен соответственно объединению их образов на $\mathscr{Y}$: $s(A_1\cup A_2)=sA_1\cup sA_2$.

Образом пустого множества будет пустое множество: $s\varnothing=\varnothing$. Образом $\mathscr{X}$ будет, в общем, часть $\mathscr{Y}$, тогда говорим, что s есть отображение $\mathscr{X}$ *внутрь, или в* $\mathscr{Y}$. Если же образом $\mathscr{X}$ является все $\mathscr{Y}$ (это нетрудно сделать, исключив из $\mathscr{Y}$ те элементы, которые не входят в область значений sx), то s есть отображение $\mathscr{X}$ *на* $\mathscr{Y}$, что ниже и считается.

Включению событий соответствует включение их образов: $A_1\subset A_2\Rightarrow sA_1\subset sA_2$. То же верно для дополнения (если s есть отображение $\mathscr{X}$ на $\mathscr{Y}$): $sA^c=(sA)^c$. И чего нельзя сказать для пересечения: два непересекающихся события, скажем x_1 и x_2, могут отображаться в одно, как это видно из рис. 2.1.

Обозначим все элементы x, отображающиеся в точку y, через $A_y=s^{-1}y=\{x: sx=y\}$ и назовем прообразом или *изображением точки* y. *Изображением* $s^{-1}B$ *события* $B\in\mathscr{Y}$ будет объединение изображений точек, входящих в B: $A_B=\bigcup\limits_{y\in B} A_y=\{x: sx\subset B\}$. Непересекающимся B_1 и B_2 соответствуют непересекающиеся изображения $s^{-1}B_1$ и $s^{-1}B_2$, а алгебраическим операциям над ними — такие же операции над их изображениями. Множества $s^{-1}y\subset\mathscr{X}$ в результате не пересекаются и s можно рассматривать как взаимно-однозначное соответствие точек y пространства $\mathscr{Y}$ и подмножеств A_y пространства $\mathscr{X}$: $A_y\leftrightarrow y$, или соответствие $\bigcup\limits_{y\in B} A_y\leftrightarrow B$, приводящее к идентичности (изоморфизму) двух алгебр событий:

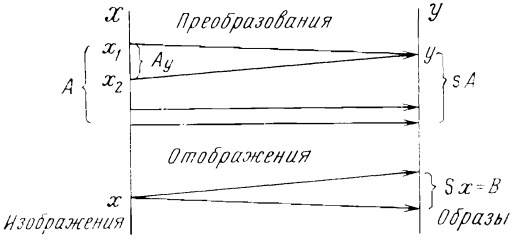


Рис. 2.1. Преобразования и отображения

алгебры $\mathscr{A}$ на $\mathscr{X}$, порожденной событиями A_y, и алгебры всех событий на $\mathscr{Y}$, обозначаемой $2^{\mathscr{Y}}$. Эти алгебры $\mathscr{A} \leftrightarrow 2^{\mathscr{Y}}$ изоморфны в том смысле, что алгебраические отношения и действия над множествами одной из них зеркально отображаются на отношения и такие же действия над другой.

Обобщим понятие преобразования, подводя его к изоморфизму двух произвольных алгебр событий. Для этого будем понимать S (обозначается в отличие от преобразования заглавной буквой) как отображение, при котором каждая точка x переходит, в общем, в подмножество $Sx = B_x$ пространства $\mathscr{Y}$, причем: а) для различных $x_1 \neq x_2$ множества $Sx_1 = B_1$ и $Sx_2 = B_2$, в которые отображаются эти точки, либо совпадают между собой, либо не пересекаются; б) $S\mathscr{X} = \bigcup Sx = \mathscr{Y}$. Равенство $Sx = B_x$ понимается как неясность, сомнение, куда из множества B_x попадет точка x, и в этом смысле S расплывчато.

Отображение S любые события из $\mathscr{X}$ переводит в события из $\mathscr{Y}$: $SA = \bigcup\limits_{x \in A} Sx$, и обратно: $S^{-1}B = \{x : Sx \subset B\}$. В нашем широком понимании для любого S обратное S^{-1} всегда определено. При повторных прямом и обратном отображениях $S^{-1}SA$ события, в общем, смазываются, становятся шире A, кроме некоторых из событий, представляющих наибольший интерес, так как именно они полностью характеризуют S. Это класс $\mathscr{A}_S$ всех тех событий на $\mathscr{X}$ (и его образ $\mathscr{B}_S$ на $\mathscr{Y}$), которое при прямом и затем обратном отображении «остаются четко на месте»:

$$\mathscr{A}_S = \{A : S^{-1}SA = A\} \overset{S}{\leftrightarrow} \mathscr{B}_S = \{B : B = SS^{-1}B\}.$$

Характерна однозначная связь: $A \leftrightarrow SA = B$, $S^{-1}B = A \leftrightarrow B$, индуцирующая две изоморфные алгебры $\mathscr{A}_S \leftrightarrow \mathscr{B}_S$, где $\mathscr{B}_S$ есть образ $\mathscr{A}_S$. Атомами алгебр являются $B_x = Sx$, $x \in \mathscr{X}$, для $\mathscr{B}_S$ и $A_x = S^{-1}Sx$ для $\mathscr{A}_S$. Если переобозначить их B_z, A_z, объединив совпадающие между собой B_x (и A_x) под одним индексом z, то S можно интерпретировать как взаимно-однозначное соответствие $A_z \leftrightarrow B_z$.

Итак, любое отображение S индуцирует изоморфные алгебры $\mathscr{A}_S$ и $\mathscr{B}_S$ и полностью определяется ими. В то же время любому изоморфизму можно подыскать соответствующее отображение S, понимая его как отображение событий из $\mathscr{X}$ (атомов A_z) в события из $\mathscr{Y}$ (атомы B_z).

Алгебры $\mathscr{A}_S$ и $\mathscr{B}_S$ охватывают те события, которые остаются четкими при преобразованиях. Через них могут быть определены образы (и изображения) всех остальных событий по формулам:

$$SA = \bigcap_{\substack{A' \in \mathscr{A}_S \\ A' \supset A}} SA'; \quad S^{-1}B = \bigcap_{\substack{B' \in \mathscr{B}_S \\ B' \supset B}} S^{-1}B'.$$

Преобразования признаков. Мы рассмотрели, как отображениями преобразуются события. Теперь обратимся к преобразованию признаков; установим, как одни из них при отображении переходят в другие.

Числовая функция $f(x)$, измеримая относительно индуцированной отображением S алгебры $\mathscr{A}_S$ событий, называется *S-представимым признаком*. Если понимать функцию f широко как отображение событий A в подмножества $f(A) = \{f(x) : x \in A\}$ числовой прямой, то для S-представимых признаков $f(A) \equiv f(S^{-1}SA)$, $\forall A$, что эквивалентно может быть взято за их определение; эти признаки принимают постоянные значения на «атомах» A_z алгебры $\mathscr{A}_S$. Класс всех S-представимых признаков обозначим $\mathscr{F}_S$. В него, очевидно, входят все индикаторные функции событий $\mathscr{A}_S$.

Класс $\mathscr{F}_S$ линеен и замкнут относительно любых арифметических действий, т. е. преобразование $F(f_1, f_2, \ldots)$ признаков из этого класса приводит к признакам из него же (можно допускать бесконечные значения признаков).

Каждый S-представимый признак имеет четко свой *образ* $\varphi(y) = f(S^{-1}y)$ (или $\varphi(B) = f(S^{-1}B)$), который является в свою очередь $\mathscr{B}_S$-измеримой функцией, или же, что то же самое, S^{-1}-представимым признаком $\varphi(B) \equiv \varphi(SS^{-1}B)$. Класс всех таких признаков обозначим $\Phi_{S^{-1}}$.

Класс $\mathscr{F}_S$ и его образ $\Phi_{S^{-1}}$ взаимно-однозначно связываются между собой: $\mathscr{F}_S \leftrightarrow \Phi_{S^{-1}}$ сохраняя упорядоченность функций: $f_1 \geqslant f_2 \Leftrightarrow \varphi_1 \geqslant \varphi_2$, где $f_1 \leftrightarrow \varphi_1$, $f_2 \leftrightarrow \varphi_2$; и идентичность арифметических действий: $F(f_1, f_2, \ldots) \leftrightarrow F(\varphi_1, \varphi_2, \ldots)$. Например, образом линейной комбинации $c + \sum c_i f_i(x)$, $f_i \in \mathscr{F}_S$, будет линейная комбинация образов $c + \sum c_i \varphi_i(y)$, $\varphi_i \in \Phi_{S^{-1}}$, $f_i \Leftrightarrow \varphi_i$.

Заметим, что если $y = Sx$ есть детерминированные преобразование, то $\Phi_{S^{-1}}$ — это вообще любые функции переменной y, а $\mathscr{F}_S$ — это множество функций вида $f(x) = \varphi(Sx)$.

Все S-представимые признаки и их образы дадут те, грубо говоря, «направления», по которым будет производиться расчет средних ИМ при отображениях S.

Расчет средних. Пусть на $\mathscr{X}$ задана ИМ $\mathscr{M}^x$ с областью $\mathscr{F}$ существования верхних средних $\overline{M}f$, $f \in \mathscr{F}$, и S есть отображение $\mathscr{X}$ на $\mathscr{Y}$. Требуется рассчитать соответствующую ИМ $\mathscr{M}^y$ на $\mathscr{Y}$, которую записываем $\mathscr{M}^y = S\mathscr{M}^x$. Какие данные о средних $\overline{M}\varphi$ на $\mathscr{Y}$ будут иметься? Что при этом теряется?

Для S-представимых признаков f ровно ничего:

$$\overline{M}\varphi = \overline{M}f, \quad \varphi \leftrightarrow f \in \mathscr{F}_S \cap \mathscr{F}, \tag{2.1}$$

и средние однозначно переносятся на их образы. Это очевидно, так как $\varphi(y)$ полностью повторяют значения $f(x) = \varphi(Sx)$.

Первичными для преобразованной ИМ $\mathscr{M}^y = S\mathscr{M}^x$ будут средние (2.1). Они согласованы на $\mathscr{B}_S$ в силу их согласованности на $\mathscr{A}_S$ и изоморфизма этих алгебр.

Последовательность этапов для определения $\mathscr{M}^y$ следующая:

1) сначала об $\mathcal{M}^x$ оставляются только сведения о средних подкласса S-представимых признаков $\mathcal{F}_S$, что соответствует $\mathcal{F}_S$-расширению $\mathcal{M}^x$;

2) средние с $\mathcal{F}_S$ переносятся на их образы по формуле (2.1), образуя первичные значения для $\mathcal{M}^y$;

3) наконец, первичные значения продолжаются на произвольные признаки.

Из сказанного следует, что будь на $\mathcal{X}$ задана $\mathcal{M}^x$, или ее $\mathcal{F}_S$-расширение $\langle \bar{M}^x \mathcal{F}_S \rangle$, итоговая $\mathcal{M}^y = S\mathcal{M}^x$ будет одной и той же. Средние только S-представимых признаков участвуют в расчете $\mathcal{M}^y$; остальные же «смазываются», теряют свои собственные средние и переходят в «подчинение» к набору $\bar{M}\mathcal{F}_S = \{\bar{M}f_S : f_S \in \mathcal{F}_S\}$.

Потерь, очевидно, не будет в том случае, если $\bar{M}\mathcal{F}_S$ определяют однозначно модель $\mathcal{M}^x$, т. е. ее первичные признаки $g \in \mathcal{G}$ все S-представимы: $\mathcal{G} \subset \mathcal{F}_S$. Тогда $\mathcal{F}_S$-расширение $\mathcal{M}^x$ совпадет с $\mathcal{M}^x$, а первичными для $\mathcal{M}^y$ будут образы $g \in \mathcal{G}$ с теми же, как у $\mathcal{M}^x$, средними, так что между первичными признаками и средними $\mathcal{M}^x$ и $\mathcal{M}^y$ устанавливается тождественная связь. Расчет средних производится уже не в три этапа, как выше указывалось, а в два, так как здесь $\mathcal{F}_S$ можно заменить на $\mathcal{G}$ и первый этап вырождается.

Пример 2.1. Пусть $\mathcal{X} = \mathcal{R}$ — числовая ось, и пусть $y = x^2$ — преобразование. Тогда $\mathcal{Y} = \mathcal{R}^+$ — полуось и средние $\bar{M}\varphi(Y)$ (для удобства случайные величины обозначаются заглавными буквами), определяющие $\mathcal{M}^y$, выражаются через средние $\mathcal{M}^x$ по формуле (2.1): $\bar{M}\varphi(Y) = \bar{M}q(X^2)$. Например, $\bar{M}Y = \bar{M}X^2$, $\bar{M}Y^2 = \bar{M}X^4$, $\bar{M}\cos Y = \bar{M}\cos X^2$ и т. д. Правые части рассчитываются на основании первичных данных о $\mathcal{M}^x$. Потерь при преобразовании не будет, если только первичные признаки $\mathcal{M}^x$ записываются сами через x^2. Скажем, $\mathcal{M}^x$ определена значениями $\bar{M}X^2$, $\bar{M}X^4$. Тогда $\mathcal{M}^y$ будет определена первичными значениями $\bar{M}Y = \bar{M}X^2$ и $\bar{M}Y^2 = \bar{M}X_4$, а все остальные $\bar{M}\varphi(Y)$, например, $\bar{M}\cos Y$, рассчитываются по ним или непосредственно переносятся как средние изображений $\bar{M}\varphi(X^2)$, например $\bar{M}\cos X^2$.

Пример 2.2. Детерминированные преобразования случайного процесса. Пусть X_t — процесс, определенный своими средними $\bar{M}f\{X_t\}$, где $f\{X_t\}$ — функционалы от X_t, и пусть S — его преобразование в $Y_t = SX_t$. Это может быть нелинейное преобразование $Y_t = v(X_t)$, например возведение в квадрат $Y_t = X^2_t$; ограничение; линейное преобразование; их комбинации, например $Y_t = \int v(X_\tau)\, h_{t,\tau} d\tau$ — нелинейность v с последующим фильтром. Собственно, вид преобразования не принципиален для расчета $M_q\{Y_t\}$, который крайне прост: эти средние в точности совпадают со средними $\bar{M}\varphi\{SX_t\}$ от $f\{X_t\} = \varphi\{SX_t\}$ — S-представимых функционалов: $M_q\{Y_t\} = \bar{M}\varphi\{SX_t\}$. Так, при преобразовании типа «нелинейность — фильтр», введенного чуть выше, имеем:

$$\bar{M}Y_t = \bar{M}\int v(X_\tau)\, h_{t,\tau}\, d\tau, \quad \bar{M}Y_t^2 = \bar{M}\left[\int v(X_\tau)\, h_{t,\tau} d\tau\right]^2;$$

$$\bar{M}\int Y_t\, dt = \bar{M}\int\int v(X_\tau)\, h_{t,\tau}\, dt d\tau,$$

и т. д. По ним ведется расчет соответствующих средних от остальных признаков процесса Y_t как по первичным значениям согласно теореме продолжения.

Подобие моделей. Речь пойдет об отображениях, которые не приводят к потере данных относительно модели. Совершенно ясно, что потерь не должно быть, если S взаимно однозначно отображает $\mathcal{X}$ на $\mathcal{Y}$, так как обратным отображением S^{-1} всегда можно успеть вернуться назад к исходному пространству и исходной модели, какой бы она ни была. Вопрос становится не таким тривиальным, если S является редукцией пространства, переводящей $\mathcal{X}$ в пространство $\mathcal{Y}$ меньшей размерности.

Отображение S называется *преобразованием подобия* (или просто подобием) для $\mathcal{M}^x$, если $\mathcal{M}^x = S^{-1}S\mathcal{M}^x$, т. е., отображая пространство $\mathcal{X}$ с помощью S в $\mathcal{Y}$, а затем обратно с помощью S^{-1}, приходим к той же самой отнюдь не более широкой модели. При этом S^{-1} будет также преобразованием подобия для $\mathcal{M}^y = S\mathcal{M}^x$, так как $SS^{-1}\mathcal{M}^y = \mathcal{M}^y$.

Модели $\mathcal{M}^x$ и $\mathcal{M}^y$, связанные между собой подобиями, называются *подобными* и обозначаются $\mathcal{M}^x \sim \mathcal{M}^y$.

Отображение S *будет подобием для* $\mathcal{M}^x = \langle \bar{M}^x \mathcal{G} \rangle$, *если* S-*представимы все первичные признаки*: $g(x) = \psi_g(Sx)$, $\forall g \in \mathcal{G}$. *При этом* $\mathcal{M}^x$ *преобразуется в подобную ей* $\mathcal{M}^y = \langle \bar{M}^x \psi \rangle$, *определенную первичными средними*: $\bar{M}^y \psi_g = \bar{M}^x g$, *и* $\Psi = \{\psi_g : g \in \mathcal{G}\}$.

Данное утверждение есть повторение рассуждений конца предыдущего раздела.

Наоборот, если S — подобие для $\mathcal{M}^x = \langle \bar{M}^x \mathcal{G} \rangle$, то, оставляя вместо $\mathcal{G}$ поднабор $\mathcal{G}_S$ всех S-представимых первичных признаков $\mathcal{G}_S \subset \mathcal{G}$, мы не изменим исходную модель: $\mathcal{M}^x = \langle \bar{M}^x \mathcal{G}_S \rangle$. Тогда класс $\mathcal{L}^+\mathcal{G}_S$ вторичных признаков будет S-представим и взаимно однозначно с сохранением отношений порядка и значений средних будет отображаться в класс $\mathcal{L}^+\Psi_S$, $\Psi_S = \{\psi_g : g \in \mathcal{G}_S\}$, вторичных признаков подобной $\mathcal{M}^y = S\mathcal{M}^x$, а $\mathcal{L}^+\mathcal{G}_S$ и $\mathcal{L}^+\Psi_S$, в свою очередь, будут подклассами $\mathcal{F}_S$ и соответственно $\Phi_{S^{-1}}$ всех представимых функций, достаточными для расчетов модели $\mathcal{M}^y$ по $\mathcal{M}^x$ при преобразовании S. Отсюда, если $\mathcal{M}^x$ и $\mathcal{M}^y$ подобны, то классы $\mathcal{F}_S$ и $\Phi_{S^{-1}}$ (как и $\mathcal{L}^+\mathcal{G}_S$, $\mathcal{L}^+\Psi_S$) взаимно однозначно отображаются друг в друга с зеркальностью арифметических действий, сохранением порядка и средних: $f \leftrightarrow \varphi$, $\bar{M}^x f = \bar{M}^y \varphi$, $f \in \mathcal{F}_S$, $\varphi \in \Phi_{S^{-1}}$.

Отношение подобия моделей рефлексивно: $\mathcal{M}^x \sim \mathcal{M}^x$, симметрично: $\mathcal{M}^x \sim \mathcal{M}^y \Rightarrow \mathcal{M}^y \sim \mathcal{M}^x$ и транзитивно: $\mathcal{M}^x \sim \mathcal{M}^y$, $\mathcal{M}^y \sim \mathcal{M}^z \Rightarrow \mathcal{M}^x \sim \mathcal{M}^z$. Подобные модели имеют одинаковые размерности, строго соответствующие друг другу грани и изображаются одинаковыми геометрическими телами (но в разных пространствах).

Пример 2.3. Пусть на разбиении $\mathcal{X}$ на непересекающиеся события: $\mathcal{X} = A_1 + A_2 + \ldots + A_k$, задано интервальное распределение вероятностей ИРВ границами вероятностей $\underset{\sim}{p}_i = \underset{\sim}{P}(A_i)$, $\tilde{p}_i = \tilde{P}(A_i)$. Эта модель подобна ИРВ на k-точечном пространстве $\mathcal{Y}_k = \{y_1, \ldots, y_k\}$ с теми же вероятностями

$\underset{\sim}{P}(y_i)=\underset{\sim}{p}_i$, $\bar{P}(y_i)=\bar{p}_i$. Отображением, наводящим подобие, будет $A_i \xrightarrow{S} y_i$, а обратным к нему будет изоморфизм $y_i \xrightarrow{S^{-1}} A_i$. Класс $\mathcal{F}_S$ образуют функции вида $c+\sum c_i A_i(x)$, а $\Phi_{S^{-1}}$ — вида $c+\sum c_i \delta_{y_i}(y)$.

Подобие сохраняют алгебраические действия над моделями:

$$\mathcal{M}^x_\theta \sim \mathcal{M}^y_\theta, \quad \forall\theta \Rightarrow \bigwedge_\theta \mathcal{M}^x_\theta \sim \bigwedge_\theta \mathcal{M}^y_\theta, \quad \bigvee_\theta \mathcal{M}^x_\theta \sim \bigvee_\theta \mathcal{M}^y_\theta.$$

2.2. СЛУЧАЙНЫЕ ПРЕОБРАЗОВАНИЯ

Переходные модели. В предыдущем параграфе рассмотрены отображения $\mathcal{X}$ в $\mathcal{Y}$ и их более общие формы — изоморфизмы, при которых каждая точка x с достоверностью 1 переходит в множество $B_x \subset \mathcal{Y}$, причем B_x при разных x совпадают между собой или не пересекаются.

Рассмотрим общий случай, когда x в принципе может перейти в любую точку $y \in \mathcal{Y}$, а знания об этом среднестатистические. Такие преобразования называются случайными: $\mathcal{X} \xrightarrow{Q} \mathcal{Y}$. Для случайных преобразований Q каждой точке x указывается некоторая ИМ $\mathcal{M}^y_x$ на $\mathcal{Y}$, называемая *переходной* из $\mathcal{X}$ на $\mathcal{Y}$ и определяемая своими средними значениями $\bar{M}^y_x\varphi(y)$. Способы ее задания точно такие же, как любой ИМ, а именно, первичными средними $\bar{M}^y_x\psi$, $\psi \in \Psi$, где ψ — первичные признаки на $\mathcal{Y}$, средние которых, да и вид их самих, в общем, зависят от x.

В частном случае, когда $\mathcal{M}^y_x$ есть интервальное распределение вероятностей, то первичными признаками будут события на $\mathcal{Y}$ и переходная ИМ будет полностью определяться верхними переходными вероятностями

$$\bar{q}(x, B)=\bar{P}^y_x(B), \quad B \subset \mathcal{Y},$$

указывающими, с какой наибольшей вероятностью точка x перейдет в событие B при преобразовании Q. Очевидно, $\underline{q}(x, B)=1-\bar{P}^y_x(B^c)$, т. е. нижние переходные вероятности сразу вычисляются по верхним.

Преобразования моделей. Пусть теперь заданы $\mathcal{M}^x$ на $\mathcal{X}$ и переходная модель $\mathcal{M}^y_x$ на $\mathcal{Y}$, описывающая случайное преобразование Q из $\mathcal{X}$ на $\mathcal{Y}$. Тогда соответствующая им $\mathcal{M}^y$ на $\mathcal{Y}$ будет определяться средними:

$$\bar{M}^y \varphi(y)=\bar{M}^x(\bar{M}^y_x \varphi(y)). \tag{2.2}$$

Здесь $\bar{M}^y_x\varphi(y)$ есть средние по переходной модели $\mathcal{M}^y_x$, а как функции x — это в свою очередь признаки на $\mathcal{X}$, средние которых по $\mathcal{M}^x$, обозначаемые $\bar{M}^x$, и дадут согласованные значения $\bar{M}^y\varphi$, определяющие $\mathcal{M}^y$. Результат преобразования записывается: $\mathcal{M}^y=Q\mathcal{M}^x$.

Если $\mathcal{F}^x$ и $\mathcal{F}^y_x$ есть соответственно области существования средних для $\mathcal{M}^x$ и $\mathcal{M}^y_x$, то для $\mathcal{M}^y$ эту область составляют такие функции $\varphi(y) \in \mathcal{F}^y_x$, что $\bar{M}^y_x\varphi(y) \in \mathcal{F}^x$. Сюда, конечно же, входят любые ограниченные функции.

Рассмотрим на примере последовательность вычислений по формуле (2.2).

Пример 2.5. Пусть как $\mathcal{M}^x=\langle \bar{M}^x g\rangle$ задана единственным первичным средним $\bar{M}^x g$, так и переходная модель $\mathcal{M}^y_x=\langle \bar{M}^y_x\psi\rangle$ — одним средним $\bar{M}^y_x\psi(y)=\bar{h}(x)$. Тогда согласно формуле (2.2) для признака $\varphi(y)$ имеем

$$\bar{M}^y\varphi=\bar{M}^x\bar{M}^y_x\varphi=\min_{c+c_1^+ g(x)\geqslant \bar{M}^y_x\varphi}[c+c_1^+\bar{M}^x g],$$

где $\bar{M}^y_x\varphi=\min_{d+d_1^+\psi(y)\geqslant\varphi(y)}[d+d_1^+\bar{h}(x)]$.

Среднее $\bar{M}^y\varphi$ будет определено, если определено $\bar{M}^y_x\varphi$ и мажорируемо линейными комбинациями $c+c^+_1 g(x)$. Причем для любых φ переходное среднее $\bar{M}^y_x\varphi$ как функция переменной x будет пропорционально $\bar{h}(x)$. Тогда и от $\mathcal{M}^x$ требуется только знания $\bar{M}^x\bar{h}(x)$, т. е. $\bar{h}(x)$ — его расширения. Этот факт не распространяется на несколько первичных значений $\bar{M}^y_x\psi_j=\bar{h}_j(x)$, так как тогда $\bar{M}^y_x\varphi$ не будут, в общем, линейными комбинациями $\bar{h}_j(x)$.

Объясним, почему так. Пусть первичных признаков переходной модели не один, а два, скажем ψ_1 и ψ_2 с соответствующими им верхними средними $\bar{h}_1(x)$ и $\bar{h}_2(x)$, тогда

$$\bar{M}^y_x\varphi=\min_{d+d_1^+\psi_1(y)+d_2^+\psi_2(y)\geqslant\varphi(y)}[d+d_1^+\bar{h}_1(x)+d_2^+\bar{h}_1(x)]$$

и правая часть не будет выражаться в виде линейной комбинации $h_1(x)$ и $\bar{h}_2(x)$, так как значения коэффициентов d и d^+_j, при которых достигается минимум, могут зависеть от x.

В примере был введен тот случай, когда $\bar{M}^y_x\varphi$ при любых φ образуют весьма ограниченный класс $\mathcal{F}^x{}_*$ признаков, причем это всегда замыкание $[\mathcal{L}^+\mathcal{H}]$ полулинейной оболочки $\mathcal{L}^+\mathcal{H}$ некоторого набора $\mathcal{H}$. Тогда для расчета $\mathcal{M}^y=Q\mathcal{M}^x$ и от $\mathcal{M}^x$ потребуется только знания $\bar{M}^x f$, $f \in \mathcal{F}^x{}_*$, а $\mathcal{F}^x{}_*$-расширение $\mathcal{M}^x$ не приведет к изменению $\mathcal{M}^y: Q\mathcal{M}^x=Q\langle \bar{M}^x\mathcal{F}^x{}_*\rangle$.

Остановимся на этом и рассмотрим для начала самый крайний случай. Пусть переходные ИМ одни и те же, независимо от x: $\mathcal{M}^y_x=\mathcal{M}_0$. Тогда $\bar{M}^y_x\varphi$ будут константами, не зависящими от x, и класс $\mathcal{F}^x{}_*$ вырождается в постоянную c. Получается, что о $\mathcal{M}^x$ вообще ничего не надо знать, чтобы рассчитать $\mathcal{M}^y: \mathcal{M}^y=\mathcal{M}_0$. Это тот случай, когда данные о статистических свойствах на $\mathcal{Y}$ не меняются, если становится известным «вход» x.

Рассмотрим другой случай. Пусть переходная модель $\mathcal{M}^y_x$ одна и та же на непересекающихся событиях A_j разбиения $\mathcal{A}_\Sigma=\{A_1, \ldots, A_k\}$ пространства $\mathcal{X}$. Достаточно, чтобы первичные средние, определяющие переходные ИМ, были постоянны на A_j, и тогда $\bar{M}^y_x\varphi$ будут постоянными при $x \in A_j$, т. е. $\mathcal{A}_\Sigma$-измеримыми, — что и даст класс $\mathcal{F}^x{}_*$. Модель $\mathcal{M}^x$ без ущерба для $\mathcal{M}^y=$

$=Q\mathcal{M}^x$ может быть расширена до ИМ, первичными которой являются средние $\overline{M}\sum c_i A_i(x)$ всех $\mathcal{A}_\Sigma$-измеримых функций (но никак не до ИРВ с первичными событиями из $\mathcal{A}_\Sigma$ или их объединениями).

Детерминированные преобразования есть частный случай случайных, когда переходные модели $\mathcal{M}^y{}_x=\langle P(B_x)=1\rangle$ являются голыми на взаимно непересекающиеся B_x, т. е. достоверно $x\to B_x$. Эти преобразования относятся к предыдущему типу, что будет видно, если для некоторого B_z обозначить $A_z=\{x: x\to B_z\}$, и тогда $A_z\leftrightarrow B_z$.

Свойства преобразований модели. Рассмотрим, какими свойствами обладает операция преобразования своего рода статистического пересчета моделей: $\mathcal{M}^y=Q\mathcal{M}^x$. Пусть на $\mathcal{X}$ данных нет — голая ИМ $\mathcal{J}^x$. Тогда $\mathcal{M}^y=Q\mathcal{J}^x$ будет определяться средними:

$$\overline{M}^y\varphi(y)=\sup_x \overline{M}^y_x\varphi(y)$$

и результат записывается:

$$Q\mathcal{J}^x=\bigvee_{x\in\mathcal{X}}\mathcal{M}^y_x.$$

Значит, преобразование голой ИМ эквивалентно объединению переходных ИМ по входу x, рассматриваемому как параметр.

Зафиксируем этот результат, дав ему расширенную формулировку: *объединение ИМ $\mathcal{M}_\theta$ по параметру $\theta\in\Theta$ может эквивалентно рассматриваться как результат случайного преобразования множества Θ, априорных данных о котором нет, в пространство исходов, а $\mathcal{M}_\theta$ — как переходная модель этого преобразования.*

Отсутствие априорных данных о x даст самую широкую модель на выходе, поэтому в общем случае $Q\mathcal{M}^x\subset Q\mathcal{J}^x$.

А что, если переходная $\mathcal{M}^y{}_x$ при всех x является голой, т. е. никаких данных о преобразовании нет? Тогда преобразование любой $\mathcal{M}^x$ ведет едино к голой модели, не несущей никаких статистических данных относительно исходов на $\mathcal{Y}$:

$$\mathcal{M}^y_x=\mathcal{J}^y\Rightarrow Q\mathcal{M}^x\equiv\mathcal{J}^y,\quad\forall\mathcal{M}^x.$$

Данный факт следует из формулы (2.2): $\overline{M}^y\varphi(y)=\overline{M}^x\sup\varphi(y)=\sup\varphi(y)$.

Рассмотрим теперь, как транформируются отношения включения и операции объединения, пересечения моделей на $\mathcal{X}$ после случайных преобразований Q.

1. Отношение включения сохраняется при преобразованиях:

$$\mathcal{M}^x_1\subset\mathcal{M}^x_2\Rightarrow Q\mathcal{M}^x_1\subset Q\mathcal{M}^x_2.$$

2. Преобразование объединения моделей на $\mathcal{X}$ равно объединению их преобразовний:

$$Q\Big(\bigvee_v\mathcal{M}^x_v\Big)=\bigvee_v Q\mathcal{M}^x_v.$$

12

В частности, если $\mathcal{M}^x=\bigvee_v\mathcal{P}_v$ представляется как объединение вершин, то преобразование $\mathcal{M}^x$ — как объединение преобразованных вершин:

$$Q\mathcal{M}^x=\bigvee_v Q\mathcal{P}_v.$$

3. Преобразование пересечения моделей включается в пересечение их преобразований:

$$Q\Big(\bigwedge_v\mathcal{M}^x_v\Big)\subset\bigwedge_v Q\mathcal{M}^x_v.$$

Причина тому, что в последней части стоит включение, а не равенство, состоит в ослаблении порядка между признаками при случайных преобразованиях. В частности (когда v — индекс, обозначающий номер первичного признака), получаем

$$Q\langle\overline{M}^x_{\mathcal{G}}\rangle=Q\bigwedge_{g\in\mathcal{G}}\langle\overline{M}^x g\rangle\subset\bigwedge_{g\in\mathcal{G}}Q\langle\overline{M}^x g\rangle.$$

4. Справедливо включение

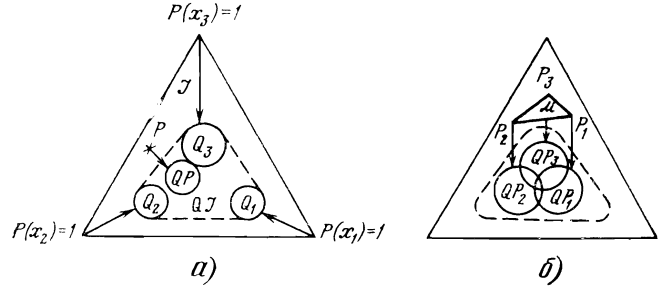
$$\Big(\bigvee_\theta Q_\theta\Big)\mathcal{M}^x\supset\bigvee_\theta Q_\theta\mathcal{M}^x,$$

где $Q=\bigvee_\theta Q_\theta$ — преобразование, соответствующее объединению переходных ИМ: $\mathcal{M}^y{}_x=\bigvee_\theta\mathcal{M}^y{}_{x,\theta}$.

В самом деле, $\overline{M}^y\varphi=\overline{M}^x\sup_\theta\overline{M}^y{}_{x,\theta}\varphi\geqslant\sup_\theta\overline{M}^x\overline{M}^y{}_{x,\theta}\varphi=\sup_\theta\overline{M}^y{}_\theta\varphi$.

Дадим на примере геометрическую иллюстрацию сказанного.

Пример 2.6. Пусть троеточие $\mathcal{X}=\{x_1, x_2, x_3\}$ случайным преобразованием Q отображается на себя. Переходная модель $\mathcal{M}^y{}_x$ при каждом значении x_i представляется как выпуклое множество Q_i векторов вероятностей $\mathbf{q}^\mathrm{T}=(q(x_i, y_1), q(x_i, y_2), q(x_i, y_3))$, так что $\overline{\mathcal{M}}^y{}_{x_i}\varphi=\max_{\mathbf{q}\in Q_i}\mathbf{q}^\mathrm{T}\boldsymbol{\varphi}$, где функция $\boldsymbol{\varphi}$ есть вектор. Множества Q_i, как представлено на рис. 2.2,*а*, — это результат преобразований индикаторных моделей $\langle P(x_i)=1\rangle$, $i=1, 2, 3$, являющихся



а) *б)*

Рис. 2.2. Образы моделей при случайных преобразованиях

вершинами $\mathcal{J}$. Преобразование вероятностей $\mathbf{P}=(P(x_1),\ P(x_2),\ P(x_3))$ ведет к семейству векторов:

$$Q\mathbf{P}=\left\{\mathbf{P}^y:P^y(y_j)=\sum_{i=1}^{3}P(x_i)\,q(x_i,\ y_j),\quad \mathbf{q}\in Q_i\right\}.$$

Как это видно из рис. 2.2,*б*, $Q\mathcal{M}^x$ составляется из преобразованных векторов вероятностей, лежащих в вершинах $\mathcal{M}^x$, и «загнанных» внутрь преобразования $Q\mathcal{J}^x$ голой ИМ. Суть третьего свойства раскрывается включением: $(Q\mathbf{P}_2\vee\vee Q\mathbf{P}_3)\wedge(Q\mathbf{P}_3\vee Q\mathbf{P}_1)\supset Q\mathbf{P}_3$.

Индикаторные преобразования, интервальная арифметика. Рассмотрим особый класс $\mathcal{M}^x$ и $\mathcal{M}^y{}_x$. Пусть $\mathcal{M}^x$ задана единственным предложением: «событие A из $\mathcal{X}$ достоверно». Это даст индикаторную модель: $\mathcal{M}^x=\langle P_x(A)=1\rangle$, определяемую согласованными средними $\overline{M}^x f(x)=\sup\limits_{x\in A}f(x)=\overline{f}(A)$, $\forall f\in\mathcal{F}_0$. Пусть случайный оператор осуществляет достоверно перевод каждой точки x в событие B_x на $\mathcal{Y}$, что описывается *индикаторными переходными моделями* $\mathcal{M}^y{}_x=\langle P^y(B_x)=1\rangle$, определяемыми средними: $\overline{M}^y{}_x\varphi(y)=\overline{\varphi}(B_x)$, $\forall\varphi\in\mathcal{F}_0$ (здесь не требуется, как для детерминированных изоморфных отображений, чтобы B_x взаимно не пересекались). Тогда согласно (2.2)

$$\overline{M}^y\varphi(y)=\sup_{x\in A}\left(\sup_{y\in B_x}\varphi(y)\right)=\overline{\varphi}\left(\bigcup_{x\in A}B_x\right),$$

т. е. результирующая $\mathcal{M}^y$ будет также B-индикаторной с событием $B=\bigcup\limits_{x\in A}B_x$: $\mathcal{M}^y=\langle P^y(B)=1\rangle$.

Таким образом, при индикаторных преобразованиях индикаторные модели переходят сами в себя, оставаясь в рамках этого очень простого класса. А по сути дела, производятся прямые преобразования одних событий в другие. В частных случаях, когда $\mathcal{X}=\mathcal{Y}=\mathcal{R}$ — числовая ось, A и B_x — интервалы на ней, а переходные операторы отражают простейшие арифметические действия, получаем правила преобразования интервалов — интервальную арифметику.

Сложение. Пусть $\mathcal{M}^x$ индикаторная на отрезке $A=[a,\ b]$ модель, а переходной оператор прибавляет к числу x отрезок $[c,\ d]$, т. е. $\mathcal{M}^y{}_x$ — индикаторные модели на отрезках $B_x=x+[c,\ d]=[x+c,\ x+d]$. Тогда индикаторным для $\mathcal{M}^y$ будет отрезок $B=[a,\ b]+[c,\ d]=[a+c,\ b+d]$, что и дает нам правило интервального сложения.

Вычитание. Аналогично предыдущему $[a,\ b]-[c,\ d]=[a-d,\ b-c]$. При сложении и вычитании ширина результирующего отрезка равна сумме составляющих.

Умножение. Модель $\mathcal{M}^x$ та же, а переходный оператор умножает каждое число x на отрезок, что ведет к отрезкам, равным $B_x=x[c,\ d]=[xc,\ xd]$ при $x\geqslant 0$ и $B_x=[xd,\ xc]$ при $x<0$. Объединяя B_x по $x\in[a,\ b]$, получаем результат умножения отрезков в

виде одного отрезка $B=[a,\ b]\times[c,\ d]=[\min\{ac,\ ad,\ bc,\ bd\},\ \max\{ac,\ ad,\ bc,\ bd\}]$, задающего индикаторную модель $\mathcal{M}^y$.

Деление. Модель $\mathcal{M}^x$ та же, а оператор осуществляет деление: при $x\geqslant 0$ имеем

$$B_x=x/[c,\ d]=\begin{cases}\left[\dfrac{x}{d},\ \dfrac{x}{c}\right] & \text{при } c\leqslant d<0,\\ \left(-\infty,\ \dfrac{x}{c}\right]\cup\left[\dfrac{x}{d},\ \infty\right) & \text{при } c\leqslant 0\leqslant d,\\ \left[\dfrac{x}{d},\ \dfrac{x}{c}\right] & \text{при } 0<c\leqslant d,\end{cases}$$

а при $x<0$ в правой части равенств переставляются d с c.

Объединение по $x\in[a,\ b]$ дает результат интервального деления:

$$[a,\ b]/[c,\ d]=\begin{cases}[a,\ b]\times[1/d,\ 1/c], & 0\notin[c,\ d],\\ (-\infty,\max\{a/c,\ b/d\}]\cup[\min\{a/d,b/c\},\ \infty), & c\leqslant 0\leqslant d.\end{cases}$$

В первом случае $0\notin[c,\ d]$ результат выражается через интервальное умножение, а во втором получается в виде объединения отрезков, и тогда при делении результат уже не будет одним отрезком, а дробится на два полуотрезка.

Как при сложении, так и при вычитании отрезок расширяется. Это ясно, так как размытые преобразования могут внести только дополнительную неопределенность, поэтому операции интервального сложения и вычитания не являются взаимно обратимыми, например $[a,\ b]+[c,\ d]-[c,\ d]=[a+c-d,\ b+d-c]$ (кроме $c=d$), равно, как операций умножения и деления.

Повторные действия приводят к дальнейшему расширению результирующих отрезков, причем при дроблении отрезка, вызванного операцией деления, последующие действия нужно совершать с каждой частью в отдельности, объединяя их потом между собой.

Простые преобразования. При рассмотрении в предыдущем параграфе детерминированных отображений S основой нахождения $\mathcal{M}^y=S\mathcal{M}^x$ служила взаимно-однозначная связь между классами S и S^{-1} представимых признаков: $\mathcal{F}_S\leftrightarrow\Phi_{S^{-1}}$, один класс на $\mathcal{X}$, другой —на $\mathcal{Y}$, и средние переносились с $\mathcal{F}_S$ на $\Phi_{S^{-1}}$, определяя $\mathcal{M}^y$. Покажем, что аналогичная связь существует и для некоторого типа случайных преобразований (обобщающих изоморфизмы). Определим их.

Пусть Φ — набор признаков на $\mathcal{Y}$. Назовем Φ-*простым* такое случайное преобразование из $\mathcal{X}$ на $\mathcal{Y}$, переходная модель $\mathcal{M}^y{}_x$ которого задается точными на Φ средними: $M^y{}_x\varphi=\overline{M}^y{}_x\varphi=M^y{}_x\varphi$, $\forall\varphi\in\Phi$; они же первичные, так что $\mathcal{M}^y{}_x=\langle M^y{}_x\overline{\Phi}\rangle$.

Ввиду того, что точные средние распространяются на линейную оболочку, оставаясь на ней точными, Φ-простые и $\mathcal{L}\Phi$-простые преобразования суть одно и то же, поэтому сразу удобно считать Φ линейным подклассом признаков на $\mathcal{Y}$.

При Φ-простых преобразованиях признаку $\varphi\in\Phi$ на $\mathcal{Y}$ ставится в соответствие признак f_φ на $\mathcal{X}$ вида

$$f_\varphi(x)=M^y_x\varphi(y),$$

называемый *изображением* φ, так что $\varphi\to f_\varphi$. Средние у признаков $\varphi\in\Phi$ и их изображений f_φ одни и те же:

$$\overline{M}^y\varphi(y)=\overline{M}^x M^y_x\varphi(y)=\overline{M}^x f_\varphi(x). \tag{2.3}$$

Обозначим линейный класс

$$\mathcal{F}_\Phi=\{f_\varphi(x),\quad \varphi\in\Phi\}$$

и назовем его *изображением* Φ. Таким образом, Φ-простое преобразование Q порождает отображение $\Phi\to\mathcal{F}_\Phi$ признаков на $\mathcal{Y}$ в их изображения на $\mathcal{X}$. Наоборот, каждому признаку $f\in\mathcal{F}_\Phi$ будет соответствовать подмножество Φ_f признаков $\varphi\in\Phi$, таких, что их изображение есть f:

$$f\to\Phi_f=\{\varphi:\varphi\in\Phi,\quad M^y_x\varphi(y)=f(x)\}.$$

Множество Φ_f называется *нечетким образом признака* f относительно Q. Каждому признаку f из $\mathcal{F}_\Phi$ соответствует свой образ, и объединение всех образов дает Φ. Сказанное иллюстрируется рис. 2.3.

Если $\varphi_1\geqslant\varphi_2$, то $f_{\varphi_1}=M^y{}_x\varphi_1\geqslant M^y{}_x\varphi_2=f_{\varphi_2}$. Следовательно, соотношение упорядоченности между признаками $\varphi\in\Phi$ влекут за собой такие же отношения между изображениями. Для образов это не всегда верно: может быть $f_1\geqslant f_2$, f_1, $f_2\in\mathcal{F}_\Phi$, но отнюдь не $\varphi_1\geqslant\varphi_2$ для $\varphi_1\in\Phi_{f_1}$, $\varphi_2\in\Phi_{f_2}$. Вывод тот, что отношения большеменьше внутри Φ беднее, чем внутри $\mathcal{F}_\Phi$, и следовательно, *случайные преобразования, даже простые, нарушают порядок между признаками.*

Рассмотрим, каким получается $\mathcal{M}^y=Q\mathcal{M}^x$ при простых преобразованиях. Во-первых, если $\mathcal{M}^x$ является $\mathcal{G}$-точной моделью ($\underline{M}^x g=\overline{M}^x g$, $g\in\mathcal{G}$) и $\mathcal{G}\in\mathcal{F}_\Phi$, то $\mathcal{M}^y$ будет $\Phi_{\mathcal{G}}=\bigcup\limits_{f\in\mathcal{G}}\Phi_f$-точной.

Пусть теперь $\mathcal{M}^x=\langle\overline{M}^x\mathcal{G}\rangle$ не является точной и задана на первичном наборе $\mathcal{G}$ признаков, и пусть $\mathcal{G}\subset\mathcal{F}_\Phi$. Так как $\mathcal{F}_\Phi$ есть линейный класс, то $\mathcal{L}\mathcal{G}\subset\mathcal{F}_\Phi$. Какой здесь будет $\mathcal{M}^y=Q\mathcal{M}^x$? Согласно формуле (2.3) $\overline{M}^y\varphi$ на признаках из Φ получаются перенесением средних с изображений этих признаков, т. е. с признаков из класса $\mathcal{F}_\Phi$. Дальнейшее продолжение на произвольные признаки на $\mathcal{Y}$ будет следовать из формулы (2.2)

$$\overline{M}^y h(y)=\overline{M}^x[\inf_{h\leqslant\varphi\in\Phi} M^y_x\varphi]=\overline{M}^x[\inf_{h\leqslant\varphi\in\Phi} f_\varphi(x)],$$

где инфимум берется по $\varphi\in\Phi$. Видно, что в общем случае

$$\overline{M}^y h(y)\leqslant\inf_{h\leqslant\varphi\in\Phi}\overline{M}^x f_\varphi=\inf_{h\leqslant\varphi\in\Phi}\overline{M}^y\varphi.$$

При строгом неравенстве класс Φ не будет первичным для $\mathcal{M}^y$, а именно, верно включение $\mathcal{M}^y\subset\langle\overline{M}^y\Phi\rangle$, где $\langle\overline{M}^y\Phi\rangle$ есть Φ-расширение $\mathcal{M}^y$. Тождественное же равенство будет иметь место тогда, когда инфимум проносится за символ M^x. Это как раз тот случай, когда $\overline{M}^y{}_x h\in\mathcal{F}_\Phi$ при всех h, для которых левая часть определена. Отсюда следует:

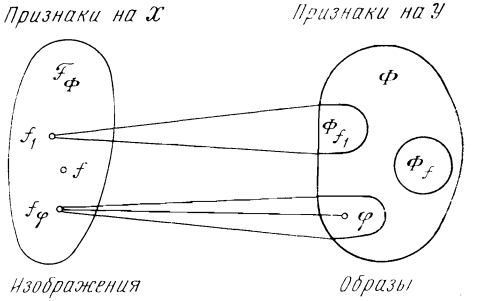


Рис. 2.3. Образы и изображения

Если случайное преобразование Q; *а) задано точными на классе* Φ *переходными моделями* $\mathcal{M}^y{}_x=\langle M^y{}_x\Phi\rangle$, *б) если* $\overline{M}^y{}_x h$, $\forall h\in\mathcal{F}^y$, *оказываются изображениями* $\varphi\in\Phi$, *то* $\mathcal{M}^y=Q\mathcal{M}^x$ *будет полностью определяться средними* (2.3) *на классе* Φ *признаков.*

Рассмотрим пример, когда условия этого утверждения выполняются.

Пример 2.7. Преобразование задано переходной плотностью.

Пусть $\mathcal{X}=\mathcal{R}^n$, $\mathcal{Y}=\mathcal{R}^m$ — векторные пространства, и пусть преобразование $\mathcal{R}^n\overset{Q}{\to}\mathcal{R}^m$ задается точной переходной плотностью вероятностей $p_x(y)$ на $\mathcal{R}^m$ при каждом заданном векторе $x\in\mathcal{R}^n$. Если это плотность по мере-длине, то класс Φ — это всевозможные интегрируемые функции $\varphi(y)$ и их изображениями будут

$$f_\varphi(x)=\int\limits_{y\in\mathcal{R}^m}\varphi(y)P_x(y)\,dy,\quad \varphi\in\Phi.$$

Каким же будет класс $\mathcal{F}_\Phi$ изображений? Чтобы ответить на этот вопрос, в плотности $p_x(y)$ «окрестим» y как параметр со значениями из $\mathcal{Y}$ и взглянем на переходную плотность как на набор функций переменной x, индексированный параметром y, для чего переобозначим ее: $q(x, y)=p_x(y)$. Если интеграл интерпретировать как линейную комбинацию функции $q(x, y)$ с весами $\varphi(y)$, то будет видно, что класс $\mathcal{F}_\Phi$ есть замыкание линейной оболочки $\sum c_i q(x, y_i)$. Этот класс линеен. Его размерность определяется числом линейно-независимых функций $q(x, y_i)$, т. е. размерностью их базиса. Он может быть невелик, если $q(x, y)$ при всевозможных $y\in\mathcal{Y}$ принимает как функция x лишь некоторые конкретные очертания.

При расчете $\mathcal{M}^y$ можно отбросить все знания о $\mathcal{M}^x$, оставив лишь $M^x f_\varphi(x)$, $f_\varphi\in\mathcal{F}_\Phi$, так как только они дадут средние $\overline{M}^y\varphi(y)$ для любых интегрируемых φ.

Дополнения. 1. Связь между преобразованиями. Класс индикаторных преобразований шире класса изоморфных отображений § 2.1, в которых тоже $x\to B_x$, но B_x должны либо совпадать, либо не пересекаться между собой.

Индикаторные преобразования $x\to B_x$, хотя $\mathcal{M}^x$ для них определяются точными первичными средними $M^y{}_x B_x=1$, не входят в класс простых преобра-

зований. Действительно, последние задаются точными значениями $M^{y}{}_{x}\varphi(y)$, $\varphi\in\Phi$, в которых первичные признаки $\varphi(y)$ переходных моделей не зависят от x (зависят лишь сами первичные значения), а B_x соответствует индикаторному признаку $B_x(y)$, зависящему от x.

Изоморфизмы § 2.1 есть частный случай простых преобразований, когда первичными переходной модели являются индикаторы $\varphi(y)=B_j(y)$, где B_j есть элементы разбиения $\mathcal{Y}$, а их вероятности $M^{y}{}_{x}\varphi(y)=P^{y}{}_{x}(B_j)$ равны 1, если $x\to B_j$, иначе — 0.

2. Случайное подобие. Дадим обобщение понятия подобия, показав, что подобными могут быть модели, связанные простыми случайными преобразованиями, наделенными свойствами подобных отображений § 2.1.

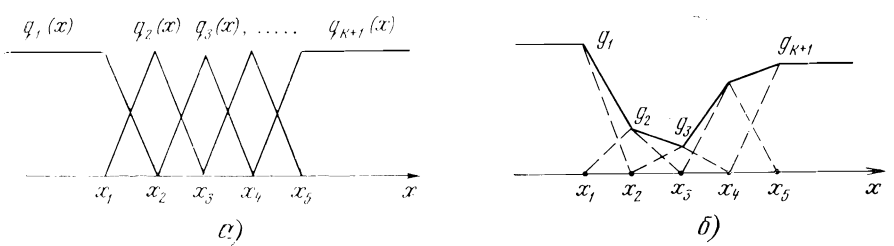
Модели $\mathcal{M}^x$ и $\mathcal{M}^y$ называются *подобными* между собой, что пишется $\mathcal{M}^x\sim\mathcal{M}^y$, если, во-первых, $\mathcal{M}^y=Q\mathcal{M}^x$ для некоторого преобразования Q (в общем, случайного) из $\mathscr{X}$ на $\mathcal{Y}$, и во-вторых, для Q существует «возвратное» (не всегда обратное) преобразование Q^- из $\mathcal{Y}$ на $\mathscr{X}$ (также, в общем, случайное) такое, что $\mathcal{M}^x=Q^-\mathcal{M}^y$, т. е. возвращающее к исходной модели на $\mathscr{X}$. Можно доказать, что случайное преобразование Q *будет преобразованием подобия, если выполняются три условия: а)* Q *задано точными на классе* Φ *переходными моделями* $\mathcal{M}^{y}{}_{x}=\langle M^{y}{}_{x}\Phi\rangle$; *б) признаки из* Φ *взаимно-однозначно с сохранением порядка связываются* $\varphi\leftrightarrow f_\varphi$ *с их изображениями* $f_\varphi=M^{y}{}_{x}\varphi$, $f_\varphi\in\mathcal{F}_\Phi$ *(наводя изоморфизм* Φ *и* $\mathcal{F}_\Phi$*); в) класс* $\mathcal{F}_\Phi$ *является определяющим для* $\mathcal{M}^x$ *(а* Φ *— определяющим для* $\mathcal{M}^y$*).*

Рассмотрим пример, когда условия утверждения выполняются.

Пусть каждый первичный признак g модели $\mathcal{M}^x$ допускает разложение $g(x)=\sum_1^{k+1} g_i q_i(x)$ по базису q_1, ..., q_{k+1}, такому, что: а) $q_i(x)\geqslant 0$; б) $\sum_1^{k+1} q_i(x)=1$; в) $g(x)\geqslant 0\Leftrightarrow g_i\geqslant 0$, $i=1, \ldots, k+1$. Тогда преобразование Q из $\mathscr{X}$ на $\mathcal{Y}=\{y_1, \ldots, y_{k+1}\}$, задаваемые вероятностями перехода $P_x(y_i)=q_i(x)$, является подобием для $\mathcal{M}^x$ и ведет к подобной ей $\mathcal{M}^y$, определенной средними

$$\overline{M}^y\varphi(y)=\overline{M}^y\sum_1^{k+1}\varphi_i\,\delta_{y_i}(y)=\overline{M}^x\sum_1^{k+1}\varphi_i\,q_i(x),\quad \varphi_i=\varphi(y_i),\ \forall\varphi_i.$$

Сказанное иллюстрируется рис. 2.4.



а) *б)*

Рис. 2.4. Случайное подобие: *а)* базис; *б)* вид первичных признаков $\mathcal{M}^x$

2.3. НЕЧЕТКИЕ СОБЫТИЯ И РАЗМЫТЫЕ ВЕРОЯТНОСТИ

Наблюдения и их изображения. Наблюдения есть результаты явления, фиксируемого через канал, измерительные устройства, органы чувств и пр. Удобно представлять себе два пространства: $\mathscr{X}$ — исходов явления, называемое *предметным* (иногда, универсальным [18]) пространством, и $\mathcal{Y}$ — пространство для описания результатов *наблюдений* (в виде чисел, суждений, словесных высказываний и т. д.) как событий B на $\mathcal{Y}$. Пространства $\mathscr{X}$ и $\mathcal{Y}$ связываются между собой случайным оператором, описываемым переходной моделью $\mathcal{M}^{y}{}_{x}$. Через призму этого оператора, по сути, мы и следим со стороны $\mathcal{Y}$ за тем, что происходит на $\mathscr{X}$. Каждое наблюдение $B\subset\mathcal{Y}$ будет иметь в предметном пространстве $\mathscr{X}$ свое интервальное изображение:

$$[\underline{q}(x,\ B),\ \overline{q}(x,\ B)]=[\underline{P}^{y}_{x}(B),\ \overline{P}^{y}_{x}(B)],\quad B\subset\mathcal{Y}-$$

это есть границы вероятности появления события B при исходе $x\in\mathscr{X}$, вычисленные относительно переходной модели $\mathcal{M}^{y}{}_{x}$. Размытость кривой изображения как функции переменной x характеризует нечеткость наблюдения B.

Изображения разных событий $B_i\subset\mathcal{Y}$ есть составная часть средних переходной модели и потому обязаны согласовываться между собой. Отсюда логическим связям на $\mathcal{Y}$ между событиями ставятся в соответствие отношения на $\mathscr{X}$ между изображениями:

$$1.\ B_1\subset B_2\Leftrightarrow\begin{cases}\underline{q}(x,\ B_1)\leqslant\underline{q}(x,\ B_2),\\ \overline{q}(x,\ B_1)\leqslant\overline{q}(x,\ B_2);\end{cases}$$

$$2.\ B_1B_2=\varnothing\Leftrightarrow\begin{cases}\underline{q}(x,\ B_1+B_2)\geqslant\underline{q}(x,\ B_1)+\underline{q}(x,\ B_2),\\ \overline{q}(x,\ B_1+B_2)\leqslant\overline{q}(x,\ B_1)+\overline{q}(x,\ B_2);\end{cases}$$

$$3.\ B_1=B_2^{c}\Leftrightarrow\begin{cases}\underline{q}(x,\ B_1)=1-\overline{q}(x,\ B_2),\\ \overline{q}(x,\ B_1)=1-\underline{q}(x,\ B_2);\end{cases}$$

4. Все $\mathcal{Y}$ и пустое $\varnothing$ есть наблюдения с тривиальными (тождественная единица и ноль) изображениями:

$$q(x,\ \mathcal{Y})\equiv 1,\quad q(x,\ \varnothing)\equiv 0.$$

Для определения изображений $\underline{P}_x(B)$, $\overline{P}_x(B)$ всех $B\subset\mathcal{Y}$ достаточно задать переходные вероятности на первичных событиях $B_i\in\mathcal{B}$ в виде $\underline{P}_x(B_i)$, $\tilde{P}_x(B_i)$, $B_i\in\mathcal{B}$, и затем перенести на любые события по известной формуле продолжения и согласования:

$$\overline{P}_x(B)=\inf_{c+\sum c_i B_i(y)\geqslant B(y)}[c+\sum(c_i^{+}\hat{P}_x(B_i)-(-c_i)^{+}\underset{\sim}{P}_x(B_i))],$$

$$\underline{P}_x(B)=1-\overline{P}_x(B^{c}),$$

где $c^+_i = c_i$ при $c_i \geqslant 0$ и $c^+_i = 0$ при $c_i < 0$. Это и приведет к согласованным изображениям любых $B \subset \mathcal{Y}$.

Наблюдения как события на $\mathcal{Y}$ разделяются на *первичные* B_i, для которых исходно известны изображения $\underline{q}(x, B_i) = \underline{P}_x(B_i)$, $\bar{q}(x, B_i) = \bar{P}_x(B_i)$ и все остальные, $B \subset \mathcal{Y}$, которые логически следуют (логика отношений суждений и используется, по существу, в неравенстве под знаком инфимума при нахождении $\bar{P}_x(B)$).

Из наблюдений B с помощью «предположительных высказываний» формируются *нечеткие суждения*, согласно которым с вероятностями γ_i, $i = 1, 2, \ldots$, производится случайный выбор одного из нескольких $B_i \subset \mathcal{Y}$. Иначе говоря, по смыслу неясно, какое из наблюдений B_i имело место и γ_i есть степень уверенности (вероятность) того, что произошло именно B_i. Изображением такого суждения на предметном пространстве будет $[\sum \gamma_i \underline{q}(x, B_i), \sum \gamma_i \bar{q}(x, B_i)]$. Так, если B является правильным с вероятностью p и ложным (противоположным B, т. е. B^c) с вероятностью $1-p$, то изображением на $\mathcal{X}$ будет

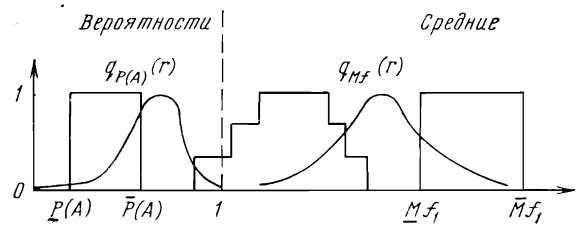
$$[p\underline{q}(x, B) + (1-p)(1-\bar{q}(x, B)),\ p\bar{q}(x, B) + (1-p)(1-\underline{q}(x, B))].$$

Каждое суждение, четкое в виде события B или нечеткое, объективно определяется своим изображением. Два суждения, имеющие одинаковые изображения, будут тождественными, так как они описывают одну и ту же ситуацию, только по-разному. Вообще любая пара границ $[\underline{q}(x), \bar{q}(x)]$, $0 \leqslant \underline{q}(x) \leqslant \bar{q}(x) \leqslant 1$, на предметном пространстве $\mathcal{X}$ есть *нечеткое событие*. Разумеется, это не обязательно изображения каких-нибудь наблюдений. Изображения как первичных наблюдений, да и всех остальных наблюдений, так и суждений составляют лишь частный класс нечетких событий. Таким образом, нечетких событий обычно существует значительно больше, чем наблюдений. Логика нечетких событий и действия между ними определяются через отношения и действия между границами по аналогии с отношениями 1—4 между изображениями.

Размытые вероятности и средние. Само по себе утверждение, что среднее Mf принимает такое-то конкретное значение, есть форма представления нашего знания о среднем, так сказать, наблюдения за ним. При точном среднем имеем четкое наблюдение за ним. Интервальное среднее $[\underline{M}f, \bar{M}f]$ есть одна из форм нечеткого наблюдения за средним или же размытых знаний. Правильнее — это изображение наших знаний, имеющихся данных о среднем на предметном пространстве его значений — числовой оси. При интервальном среднем изображение индикаторное, как это демонстрируется на рис. 2.5. Но ведь существуют и нами изучены другие, более общие формы изображений, в виде неких очертаний, они и будут здесь применены к средним, в этом наша цель.

Прежде нужно сделать одну оговорку. По своей внутренней сути средние, как и вероятности, выпуклы в том смысле, что

Рис. 2.5. Размытые вероятности и средние



области их значений — выпуклые множества. Нельзя иметь интервальное среднее в виде объединения двух непересекающихся отрезков $[\underline{M}f, \bar{M}f]$, $[\underline{M}f, \bar{M}f]$, $\bar{M}f < \underline{M}f$, он обязательно сольется в один: $[\underline{M}f, \bar{M}f]$. Это накладывает вполне конкретный отпечаток на общий вид кривых изображений средних, вводимых следующим определением.

Функция $q_{Mf}(r)$ на числовой прямой $r \in \mathcal{R}$ называется изображением среднего Mf, или *размытым средним признака* f, если она:

1) неотрицательна и не больше 1: $0 \leqslant q_{Mf}(r) \leqslant 1$;
2) равна 0 вне диапазона $[\inf f, \sup f]$ возможных значений Mf;
3) унимодальна (не имеет локальных максимумов);
4) полунепрерывна снизу;
5) хотя бы при одном r принимает значение 1.

Виды размытых средних изображены на рис. 2.5. При $f = A$ это будут размытые вероятности.

Нам понадобится понятие горизонтального *среза* от размытого среднего на высоте γ, $0 \leqslant \gamma \leqslant 1$; это интервал, внутри которого изображение среднего больше или равно γ:

$$[\underline{M}_{(\gamma)} f,\ \bar{M}_{(\gamma)} f] = \{r : q_{Mf}(r) \geqslant \gamma\}.$$

Указанные условия, ограничивающие вид размытых средних, по существу, требуют, чтобы срезом при любом $0 \leqslant \gamma \leqslant 1$ был интервал и только он. Этот интервал и создает тот фундамент, который связывает размытые средние с интервальными и далее с интервальными моделями. Величина γ есть степень доверия, или своего рода предпочтения, отводимого данному интервалу.

Любое размытое среднее эквивалентно определяется зависимостью от γ, пробегающей значения от 0 до 1, соответствующих срезам интервалов $[\underline{M}_{(\gamma)} f, \bar{M}_{(\gamma)} f]$, сужающихся при увеличении γ. Допускаются бесконечные значения границ этих интервалов.

Перейдем к понятию размытой модели средних. Представим на миг, что размытые средние $q_{Mf}(r)$ приданы всем признакам $\forall f$, ограниченным и неограниченным. Срезы $[\underline{M}_{(\gamma)} f, \bar{M}_{(\gamma)} f]$ на одной высоте γ есть интервальные средние. Если они согласованы $\forall f$, то определяют интервальную модель $\mathcal{M}_{(\gamma)}$ с областью существования $\mathcal{F}_{(\gamma)} = \{f : \bar{M}_{(\gamma)} f < \infty\}$. При γ, пробегающем значения от 0

до 1, образуется сужающаяся последовательность моделей: $\mathscr{M}_{(\gamma)} \subset \mathscr{M}_{(\gamma')}$, $\gamma \geqslant \gamma'$, которые, как бы лежащие друг на друге на разных уровнях γ, создают своего рода пирамиду, т. е. размытую модель.

Размытой моделью называется сужающаяся при увеличении γ от 0 до 1 последовательность $\mathscr{M}_{(\gamma)}$ не вырождающихся в пустую (при $\gamma=1$) интервальных моделей. Определяется она совокупностью $q_{Mf}(r)$, $\forall f$, размытых средних, согласованных между собой в каждом срезе.

Исходя из указанной интерпретации размытой модели вытекает и основной способ ее задания: сначала задаются размытые средние $\tilde{q}_{Mg}(r)$ на наборе $g \subset \mathscr{G}$ первичных признаков. Берутся γ-срезы $\underset{\sim}{M}_{(\gamma)}g$, $\tilde{M}_{(\gamma)}g$, которые как первичные средние дадут при увеличении γ сужающиеся модели $\mathscr{M}_{(\gamma)} = \langle \underset{\sim}{M}_{(\gamma)}\mathscr{G}, \tilde{M}_{(\gamma)}\mathscr{G} \rangle$. При $\gamma=1$ первичные средние должны быть непротиворечивыми, чтобы самый верхний срез был не пустым: $\mathscr{M}_{(1)} \neq \varnothing$ (тогда тем более они будут непротиворечивыми при всех $0 \leqslant \gamma \leqslant 1$). По $\mathscr{M}_{(\gamma)}$ находятся интервалы $\underline{M}_{(\gamma)}f$, $\bar{M}_{(\gamma)}f$ для $\forall f$, они и дадут γ-срезы, определяющие размытые средние $q_{Mf}(r)$, $\forall f$. Нетрудно убедиться в том, что для них выполняются все требуемые условия 1)—5), входящие в определение размытых средних. Как частный случай, получаются размытые вероятности $q_{P(B)}(r)$, $\forall B$, особенность которых в том, что они располагаются, как это видно из рис. 2.5, на интервале 0—1 значений r.

В процессе продолжения средних на все признаки параллельно происходит согласование первичных значений $\tilde{q}_{Mg}(r)$; получающиеся в результате новые кривые $q_{Mg}(r)$, в общем, делаются более узкими: $q_{Mg}(r) \leqslant \tilde{q}_{Mg}(r)$, и следовательно, более точными.

В качестве небольшого отступления разграничим уровни описаний. Первый уровень занимает градация событий на элементарные $x \in \mathscr{X}$, сложные $A \subset \mathscr{X}$, далее, определяемые точными изображениями $q(x, B)$, наконец, интервальными $[\underline{q}(x, B), \bar{q}(x, B)]$. Следующий уровень составляют статистические описания: это точное среднее Mf, интервальное $[\underline{M}f, \bar{M}f]$, размытое $q_{Mf}(r)$.

Наконец, можно было бы говорить еще о более высшем уровне, а именно, расплывчатости самих описаний, вводя вместо $q_{Mf}(r)$ интервал $\underline{q}_{Mf}(r)$, $\bar{q}_{Mf}(r)$ и далее, обобщая его до некоторой кривой принадлежности. Вопрос только в том, оправдано ли будет такое усложнение. Для ответа рассмотрим крайний случай, когда имеет место полное незнание среднего, что эквивалентно интервалам голой модели $[\underline{M}f, \bar{M}f] = [\inf f, \sup f]$, и оно же эквивалентно тривиальным изображениям вида $\underline{q}_{Mf}(r) \equiv 0$, $\bar{q}_{Mf}(r) \equiv 1$. А так как последнее, несомненно, менее удобная форма, переход к метаописаниям (описаниям описаний) и еще дальше вряд ли имеет какой-либо содержательный смысл.

Размытые действия. Интервальная арифметика § 2.2 пригодна для расчета ошибок при вычислениях, вызванных округлением чисел. Но в некоторых задачах возникает необходимость и даже имеется возможность указывать положения неизвестных чисел не в виде интервалов, т. е. категорически: да — значит, принадлежит интервалу, нет — не принадлежит, а более плавно в виде кривых предпочтений, отводимых тем или иным числовым значениям (называемым также кривыми принадлежности [15]). И с ними нужно производить действия арифметики или анализа.

Обозначим $a(r)$ — размытое изображение числа; это есть функция на $\mathscr{R}$, удовлетворяющая свойствам размытых средних (иллюстрированных рис. 2.5). Изображение $a(r)$ нужно интерпретировать как набор интервалов $A(\gamma) = \{r : a(r) \geqslant \gamma\}$, получаемых горизонтальными γ-срезами $a(r)$, причем каждый срез определяет интервальное число в виде индикаторной его модели $\mathscr{M}_{(\gamma)} = \langle P(A(\gamma)) = 1 \rangle$, а все вместе при $0 \leqslant \gamma \leqslant 1$ — размытую модель числа в том плане, как это говорилось в предыдущем разделе.

Любые действия над числами рассчитываются по правилам интервальной арифметики (т. е. по правилам преобразований индикаторных моделей) для каждого γ-среза отдельно и объединяются затем по γ в изображение результата. В более детальном изложении, если $a_j(r)$ есть изображения чисел, то $A_j(\gamma)$ будут фигурами, зеркально к $a_j(r)$ расположенными относительно главной диагонали, т. е. это те же самые $a_j(r)$, но основаниями положенные на ось ординат. При каждом γ значениями $A_j(\gamma)$ будут интервалы, поэтому преобразование $f(a_1(r), \ldots, a_J(r))$ рассчитывается по правилам таких же действий над интервальными числами, итогом которых станут интервалы $F(\gamma)$, $0 \leqslant \gamma \leqslant 1$, и их осталось переложить основаниями с оси ординат на ось абсцисс, получая размытый результат $f(r)$.

Мы имеем, таким образом, модельную интерпретацию размытых чисел и арифметических действий, за подробностями которых отсылаем к обзорной книге [18]. К указанной в ней теории ведет и следующий шаг. Он состоит в определении размытых функций как отображений $z \to a_z(r)$, а интервалов от них — как интегралов от границ интервальных функций $A_z(\gamma)$ (γ-срезов), дающих интервальный результат с переводом затем его к изображению. Хотя и пришли к известному результату, но указанная нами аргументация, по-видимому, полезна в развитие концепции нечетких описаний и действий, потому что вовлекает для этих целей содержащийся в настоящей книге общий аппарат и делает концепцию нечеткости математически строгой с точки зрения этого аппарата.

2.4. СОВМЕСТНЫЕ ИНТЕРВАЛЬНЫЕ МОДЕЛИ

Совместные и частные интервальные модели. Рассматриваются совместные модели, описывающие результаты двух произвольных случайных явлений с исходами $\mathscr{X}$ и $\mathscr{Y}$.

Пусть $\mathscr{X}\times\mathscr{Y}$ — прямое произведение двух пространств исходов, каждый элемент которого есть пара (x, y), $x\in\mathscr{X}$, $y\in\mathscr{Y}$. Модель $\mathscr{M}^{xy}$ на этом произведении называется *совместной*. Она определяется согласованными средними $\bar{M}^{xy}f(x, y)$, $\forall f\in\mathscr{F}^{xy}$, где $\mathscr{F}^{xy}$ — область существования верхних средних (включающая, по крайней мере, все ограниченные сверху функции двух переменных). А задается — первичными средними $\tilde{M}g(x, y)$, $g\in\mathscr{G}$, и тогда $\mathscr{M}^{xy}=\langle\tilde{M}\mathscr{G}\rangle$.

Признаки $f(x, y)$ двух переменных называются *совместными*, а отдельно каждой переменной $f(x)$, $\varphi(y)$ — *частными*; частные образуют подклассы $\mathscr{F}^x$ и соответственно $\mathscr{F}^y$ совместных $\mathscr{F}^{xy}$.

Средние на подклассах частных признаков $\bar{M}f(x)$, $f(x)\in\mathscr{F}^x$, $\bar{M}\varphi(y)$, $\varphi\in\mathscr{F}^y$, очевидно, согласованы и определяют *частные модели* $\mathscr{M}^x$ *и* $\mathscr{M}^y$. Итак, *частные модели получаются как часть средних совместной.*

Следующая теорема позволяет находить первичные средние частных моделей по совместной.

Теорема 2.1. О первичных признаках частных моделей. *Пусть* $\mathscr{M}^{xy}=\langle\tilde{M}\mathscr{G}\rangle$ *— совместная модель и* $\tilde{M}g(x, y)$, $g\in\mathscr{G}$, *суть ее первичные средние. Тогда соответствующая ей частная* $\mathscr{M}^x$ *на* $\mathscr{X}$ *будет определяться своими признаками вида* $\inf\limits_y\sum\limits_i c^+_i g_i(x, y)$ *с первичными средними на них, равными:*

$$\tilde{M}^x[\inf_y\sum_i c_i^+ g_i(x, y)]=\sum_i c_i^+\tilde{M}g_i(x, y),\ g_i\in\mathscr{G},$$

при всевозможном выборе неотрицательных коэффициентов c^+_i, $i=1, \ldots, k<\infty$.

В самом деле, на основании общей формулы продолжения, центрируя признаки, имеем:

$$\bar{M}^{xy}f(x)=\inf_{c+\sum c_i^+ g_i(x, y)\geqslant f(x)}[c+\sum c_i^+\tilde{M}g_i]=$$

$$=\inf_{c,\ c_i^+}\{c: c+\sum c_i^+[g_i(x, y)-\tilde{M}g_i]\geqslant f(x)\},$$

причем в условии на c, c^+_i неравенство должно соблюдаться при всех $y\in\mathscr{Y}$, что равносильно подстановке в его левую часть функции

$$c+\inf_y[\sum c_i^+(g_i(x, y)-\tilde{M}g_i)]=c+h_{\mathbf{c}}(x),\quad \mathbf{c}=(c_1, \ldots, c_k).$$

Функцию $h_{\mathbf{c}}(x)$ можно рассматривать как первичный признак для $\mathscr{M}^x$ с нулевым первичным средним. Для линейной комбинации таких признаков $\sum d^+_i h_{\mathbf{c}_i}(x)$ имеется мажорирующий признак $h_{\mathbf{c}^*}(x)$, $\mathbf{c}^*=\sum d^+_i\mathbf{c}_i$, с нулевым средним, поэтому

$$\bar{M}^x f(x)=\inf_{c,\ \mathbf{c}_i}\{c: c+\sum d_i^+ h_{\mathbf{c}_i}(x)\geqslant f(x)\}=$$

$$=\inf_{c,\ h_{\mathbf{c}^*}}\{c: c+h_{\mathbf{c}^*}(x)\geqslant f(x)\},$$

откуда и следует доказательство теоремы.

Таким образом, первичными для частной ИМ будут нижние грани $\inf\limits_y g(x, y)$, $g\in\mathscr{L}^+\mathscr{G}$ (минимум берется по исключаемой переменной) вторичных признаков совместной модели с сохранением средних.

К примеру, если $\mathscr{M}^{xy}=\langle\tilde{M}g\rangle$ определяется всего одним первичным средним $\tilde{M}g(x, y)$, то первичным для $\mathscr{M}^x$ будет $\tilde{M}[\inf\limits_y g(x, y)]=\tilde{M}g$. Если первичных средних два $\mathscr{M}^{xy}=\langle\tilde{M}g_1\rangle\wedge\langle\tilde{M}g_2\rangle$, то для частной ИМ первичные средние выглядят так:

$$\tilde{M}h_{\mathbf{c}}(x)=\tilde{M}\inf_y[g_1(x, y)+c^+g_2(x, y)]=\tilde{M}g_1+c^+\tilde{M}g_2.$$

Их уже не два, а много по причине произвольности c^+. В общем, даже при конечном наборе первичных средних $\tilde{M}g_i(x, y)$, $i=1, \ldots, k$, задающих $\mathscr{M}^{xy}$, нет гарантии, что частная $\mathscr{M}^x$ будет определяться конечным числом первичных значений, кроме ряда исключений, о которых и пойдет сейчас речь.

Пример 2.8. Пусть $\mathscr{X}=\{x_1, \ldots, x_k\}$, $\mathscr{Y}=\{y_1, \ldots, y_l\}$ и на произведении этих пространств заданы первичные вероятности, называемые совместными:

$$0\leqslant\bar{P}(x_i, y_j)\leqslant 1,\quad \sum_i\sum_j\bar{P}(x_i, y_j)\geqslant 1.$$

Эти вероятности согласованы и задают $\mathscr{M}^{xy}$. Первичными для частной $\mathscr{M}^x$ будут вероятности $\bar{P}(x_i)=\min\{1, \sum\limits_j\bar{P}(x_i, y_i)\}$. Любые же другие признаки, согласно теореме 2.1 имеющие вид

$$h_{\mathbf{c}}(x)=\inf_y\sum_i\sum_j c_{ij}^+\delta_{x_i}(x)\,\delta_{y_i}(y)=\sum\underline{c}_i^+\delta_{x_i}(x),$$

где $\underline{c}^+_i=\min c^+_{ij}$ вследствие неравенства $\tilde{M}h_{\mathbf{c}}(x)=\sum\sum c^+_{ij}\bar{P}(x_i, y_j)\geqslant\sum\sum\underline{c}^+_i\bar{P}(x_i, y_j)=\sum\underline{c}^+_i\sum\bar{P}(x_i, y_j)\geqslant\sum\underline{c}^+_i\bar{P}(x_i)$ поглощаются вероятностями $\bar{P}(x_i)$. В случае точных вероятностей $P(x_i, y_j)$, образующих совместное распределение: $\sum\sum P(x_i, y_j)=1$, частное распределение также будет точным, равным сумме $P(x_i)=\sum\limits_j P(x_i, y_j)$, что элементарно доказывается.

Следствие. *Частные признаки* $g_i(x)$, *имея согласованные средние и будучи первичными для совместной* $\mathscr{M}^{xy}$, *остаются точно такими же с теми же средними для частной* $\mathscr{M}^x$.

В самом деле, при нахождении по теореме 2.1 первичных признаков для $\mathscr{M}^x$ те из признаков $g_i\in\mathscr{G}$ совместной $\mathscr{M}^{xy}$, которые зависят только от переменной x, выносятся за знак инфимума

$$\inf_y[\sum_i d_i^+ g_i(x)+\sum_j c_j^+ g_j(x, y)]=\sum_i d_i^+ g_i(x)+\inf_y\sum_j c_j^+ g_j(x, y),$$

откуда видно, что сами эти $g_i(x)$ (а не их линейные комбинации) будут первичными для $\mathscr{M}^x$. Очевидно, согласованность их средних от $\mathscr{M}^{xy}$ передается к $\mathscr{M}^x$.

Пример 2.9. Задание совместной модели частными первичными признаками. Здесь будет рассмотрен случай, когда первичные

признаки $g\in\mathcal{G}$ совместной модели $\mathcal{M}^{xy}$ разделяются на зависящие только от переменной x либо только от переменной y:

$$g(x,\ y)=\begin{cases}h(x), & h\in\mathcal{H},\\ \psi(y), & \psi\in\Psi,\end{cases}$$

так что $\mathcal{G}=\mathcal{H}\cup\Psi$. И пусть их средние $\overline{M}h(x)$, $\overline{M}\psi(y)$ являются согласованными. Они же будут первичными для частных моделей (согласно теореме 2.1), причем, как нетрудно установить, на них распространяется свойство аддитивности средних:

$$\overline{M}^{xy}[f(x)+\varphi(y)]=\overline{M}f(x)+\overline{M}\varphi(y).$$

Разделение первичных признаков на функции только от x и только от y эквивалентно заданию отдельно $\mathcal{M}^x=\langle\overline{M}\mathcal{H}\rangle$ и $\mathcal{M}^y=\langle\overline{M}\Psi\rangle$ при полном отсутствии данных о причинной связи (зависимости) между исходами явлений $x\in\mathcal{X}$ и $y\in\mathcal{Y}$. Совместная модель равна пересечению $\langle\overline{M}\mathcal{H}\rangle\wedge\langle\overline{M}\Psi\rangle$ частных, но заданных каждая уже на $\mathcal{X}\times\mathcal{Y}$ наборами признаков только одной переменной.

Отношения между совместными и частными моделями.

1. Для голой совместной ИМ частная модель будет голой:

$$\mathcal{M}^{xy}=\mathcal{J}^{xy}\Rightarrow\mathcal{M}^x=\mathcal{J}^x.$$

Это ясно, поскольку, если ничего не известно об $\mathcal{X}\times\mathcal{Y}$, то также не будет никаких данных от $\mathcal{X}$.

2. При переходе от совместных моделей к частным сохраняется иерархия в смысле отношений включения:

$$\mathcal{M}_1^{xy}\subset\mathcal{M}_2^{xy}\Rightarrow\mathcal{M}_1^x\subset\mathcal{M}_2^x$$

так как $\overline{M}_1^x f(x)=\overline{M}_1^{xy}f(x)\leqslant\overline{M}_2^{xy}f(x)=\overline{M}_2^x f(x))$.

Возникает вопрос, сохраняются ли алгебраические операции объединения и пересечения моделей? Для операции объединения — сохраняются:

3. $\mathcal{M}^{xy}=\bigvee\mathcal{M}_\theta^{xy}\Rightarrow\mathcal{M}^x=\bigvee\mathcal{M}_\theta^x$

(так как $\overline{M}^x f(x)=\overline{M}^{xy}f(x)=\sup\limits_\theta\overline{M}_\theta^{xy}f(x)=\sup\limits_\theta\overline{M}_\theta^x f(x)$).

А операция пересечения, в общем, не сохраняется:

4. $\mathcal{M}^{xy}=\bigwedge\mathcal{M}_\theta^{xy}\not\Leftrightarrow\mathcal{M}^x=\bigwedge\mathcal{M}_\theta^x$.

В самом деле, при $\mathcal{M}^{xy}{}_\theta=\langle\overline{M}\mathcal{G}_\theta\rangle$ первичным для их пересечения будет набор $\mathcal{G}=\bigvee\mathcal{G}_\theta$ со средними $\overline{M}g(x,y)=\inf\limits_\theta\overline{M}_\theta g(x,\ y)$, $g\in\mathcal{G}$.

Теперь согласно теореме 2.1 первичными для $\mathcal{M}^x$ будут

$$\overline{M}^x[\inf_y\sum c_i^+ g_i(x,\ y)]=\sum c_i^+\overline{M}^{xy}g_i(x,\ y)=$$

$$=\sum c_i^+\inf_\theta\overline{M}_\theta g_i(x,\ y)\leqslant\inf_\theta[\sum c_i^+\overline{M}_\theta g_i(x,\ y)],$$

где квадратные скобки и есть первичные средние моделей $\mathcal{M}^x{}_\theta$, а инфимум соответствует их пересечению.

Представление совместных моделей случайными преобразованиями. В начале настоящей главы изучались преобразования $\mathcal{X}\to\mathcal{Y}$. Это некоторые операторы, преобразующие «вход» x в «выход» y. Там нас больше интересовали, во-первых, способы описания самих преобразований: детерминированных $y=sx$ (см. § 2.1) и случайных (см. § 2.2), задаваемых переходными моделями $\mathcal{M}^y{}_x$, а во-вторых, расчет «выходной» модели $\mathcal{M}^y$ по виду преобразования и «входной» $\mathcal{M}^x$. Здесь нас интересует другой вопрос: как с помощью случайных преобразований (и особого их случая — детерминированных) можно задавать совместные интервальные модели $\mathcal{M}^{xy}$ на $\mathcal{X}\times\mathcal{Y}$?

Пусть имеется модель $\mathcal{M}^x$ входа, т. е. известным путем определены согласованные $\overline{M}^x f(x)$, $f\in\mathcal{F}^x$. И пусть имеется случайное преобразование из $\mathcal{X}$ на $\mathcal{Y}$, задаваемое переходными $\mathcal{M}^y{}_x$, $\forall x\in\mathcal{X}$, т. е. при каждом x определены переходные средние $\overline{M}^y{}_x\varphi(y)$, $\varphi(y)\in\mathcal{F}^y{}_x$, где класс $\mathcal{F}^y{}_x$ при каждом x может быть, в общем, разным.

Произведением $\mathcal{M}^x$ на $\mathcal{M}^y{}_x$ назовем совместную модель на $\mathcal{M}^{xy}$, обозначаемую:

$$\mathcal{M}^{xy}=\mathcal{M}^x\mathcal{M}^y_x,$$

которая определяется средними:

$$\overline{M}^{xy}f(x,\ y)=\overline{M}^x\overline{M}^y_x f(x,\ y).\qquad(2.4)$$

Правая часть (2.4) есть последовательное вычисление на первом шаге при каждом $x\in\mathcal{X}$ переходных средних $\overline{M}^y{}_x f(x,\ y)$ по $\mathcal{M}^y{}_x$ от $f(x,\ y)$ как функций переменной y, а поскольку переходные средние будут функциями x, т. е. признаками на $\mathcal{X}$, то на втором шаге уже от них берется среднее $\overline{M}^x$. Область существования $\mathcal{F}^{xy}$ произведения моделей составляют признаки $f(x,\ y)$, при каждом x принадлежащие $\mathcal{F}^y{}_x$, причем только та их часть, для которой $\overline{M}^y{}_x f\in\mathcal{F}^x$. Это, по крайней мере, все ограниченные функции двух переменных.

Формула (2.4) для нижнего среднего (нужно заменить f на $-f$) записывается $\underline{M}^{xy}f=\underline{M}^x\underline{M}^y{}_x f$.

Прокомментируем понятие произведения. По сути дела, если интерпретировать как $\mathcal{M}^y{}_x$, так и $\mathcal{M}^x$ в виде семейств точных моделей (с точными средними), то и на первом шаге, состоящем в вычислении переходных средних $\overline{M}^y{}_x f$, $\underline{M}^y{}_x f$, и на втором, когда по ним окончательно находятся $\overline{M}f$, $\underline{M}f$, рассматриваются каждый раз наихудшие возможные варианты внутри семейств, причем раздельно для $\mathcal{M}^x{}_y$ и $\mathcal{M}^x$ и раздельно для нижнего и верхнего среднего. Это то, какие данные имелись бы о среднем Mf в наименее благоприятном случае при наличии данных (в интервальном виде или в виде семейств) о модели входа и о случайном преобразовании.

Ниже нам понадобятся следующие достаточно очевидные свойства проноса функции переменной x за знак среднего $\overline{M}^{y}_{x}$ переходной модели:

$$\overline{M}^{xy} c^{+}(x) f(x, y) = \overline{M}^{x} [c^{+}(x) \overline{M}^{y}_{x} f(x, y)],$$

$$\overline{M}^{xy} [d(x) + f(x, y)] = \overline{M}^{x} [d(x) + \overline{M}^{y}_{x} f(x, y)],$$

где $c^{+}(x)$ — произвольная неотрицательная функция, а $d(x)$ — любая функция переменной x (не выводящая из класса $\mathcal{F}^{xy}$).

Восстановление сомножителей разложимой модели. Совместная модель $\mathcal{M}^{xy}$, записываемая в виде произведения $\mathcal{M}^{x}\mathcal{M}^{y}_{x}$, называется *разложимой*. Покажем, как по совместной разложимой восстанавливаются модели-сомножители. Относительно первого из них $\mathcal{M}^{x}$ проблем не возникает: это есть частная модель, средние которой $\overline{M}^{x} f(x)$ составляют часть средних $\overline{M}^{xy} f(x, y)$ совместной модели.

Проблема восстановления второго сомножителя $\mathcal{M}^{y}_{x}$ несколько сложнее. Для этого нужно выделить «характерные» для переходных моделей $\mathcal{M}^{y}_{x}$ классы признаков $f(x, y)$, средние $\overline{M}^{xy} f$ на которых ее и определят. Характерность их должна проявляться в том, что при разных x это совершенно разные, непересекающиеся классы, «остро откликающиеся» на изменения x. Отсюда догадка, что это должны быть дельта-образные по x функции.

Как и ранее, будем обозначать $\delta_{x_1}(x)$ индикаторную функцию элементарного события $x_1 \in \mathcal{X}$. Символ $\overline{M}^{xy}$ для краткости заменяем на $\overline{M}$. Введем функцию $f(x, y) = \delta_{x_1}(x) \cdot \varphi(y)$, где $\varphi \in \mathcal{F}^{xy}$. Согласно (2.4)

$$\overline{M}\, \delta_{x_1}(x) \varphi(y) = \overline{M}^{x} \overline{M}^{y}_{x} \delta_{x_1}(x) \varphi(y) = \overline{M}^{x} \delta_{x_1}(x) \overline{M}^{y}_{x_1} \varphi(y) =$$

$$= \begin{cases} \overline{P}(x_1) \overline{M}^{y}_{x_1} \varphi(y), & \overline{M}^{y}_{x_1} \varphi(y) \geqslant 0, \\ \underline{P}(x_1) \overline{M}^{y}_{x_1} \varphi(y), & \overline{M}^{y}_{x_1} \varphi(y) < 0, \end{cases}$$

где $\underline{P}(x_1)$, $\overline{P}(x_1)$ — границы вероятностей элементарного события $\overline{x_1}$. Пусть $\overline{P}(x_1) > 0$ и $\overline{M}\delta_{x_1}(x)\varphi(y) \geqslant 0$. Из второго неравенства следует $\overline{M}^{y}_{x_1} \varphi(y) \geqslant 0$, в результате чего

$$\overline{M}^{y}_{x_1} \varphi(y) = \frac{\overline{M}\, \delta_{x_1}(x) \varphi(y)}{\overline{P}(x_1)} \quad \text{при} \quad \overline{M}\, \delta_{x_1}(x) \varphi(y) \geqslant 0. \tag{2.5}$$

Эта формула и определяет средние переходной $\mathcal{M}^{y}_{x_1}$ для тех $x_1 \in \mathcal{X}$, для которых $\overline{P}(x_1) > 0$.

Здесь требование $\overline{M}\delta_{x_1}(x)\varphi(y) \geqslant 0$ не является излишне обременительным. Действительно, если это неравенство не выполняется и $\varphi(y)$ ограничена, то оно будет выполнено для функции $\varphi_c(y) = \varphi(y) + c$ при $c \geqslant -\inf \varphi(y)$ в силу того, что $\varphi_c(y) \geqslant 0$. Определяя $\overline{M}^{y}_{x_1} \varphi_c(y)$ по (2.5), тем самым находим

$$\overline{M}^{y}_{x_1} \varphi(y) = \overline{M}^{y}_{x_1} [\varphi(y) + c] - c, \quad c \geqslant -\inf \varphi.$$

Таким образом, достаточно, чтобы (2.5) выполнялось для неотрицательных ограниченных функций $\varphi(y)$, так как для неограниченных оно получается предельным переходом от их усечений. Это своего рода формула продолжения границ. Итак, имеем.

Если $\mathcal{M}^{xy}$ разложима, причем частная $\mathcal{M}^{x}$ такова, что $\overline{P}(x_1) > 0$ для всех $x_1 \in \mathcal{X}$, то формула (2.5) позволяет восстановить переходные ИМ по совместной. При этом переходные и условные ИМ (при случившихся x) совпадают между собой.

Последняя часть следует из дополнения 1 к параграфу и смысл ее в том, что переходную модель можно восстановить, определяя условные согласно § 1.6 при истекших элементарных событиях $x_1 \in \mathcal{X}$.

Разложимость совместной модели. А всякую ли совместную модель $\mathcal{M}^{xy}$ можно разложить на произведение частной и переходной (условной), т. е. интерпретировать связь между исходами $x \in \mathcal{X}$ и $y \in \mathcal{Y}$ действием случайного оператора? Увы, далеко нет! И тогда подстановка условных моделей в (2.4) приведет к расширенной по сравнению с $\mathcal{M}^{xy}$ модели: $\overline{M}^{x}\overline{M}^{y}_{x} f(x, y) \geqslant \overline{M} f(x, y)$.

Пусть задана совместная модель $\mathcal{M}^{xy}$. Каким свойствам должны удовлетворять ее согласованные средние, чтобы она была разложима? Для решения этого вопроса поступаем так, как если бы $\mathcal{M}^{xy}$ была разложима, т. е. вычисляем переходные средние по (2.5), заменив $\varphi(y)$ на $f(x, y)$:

$$\overline{M}^{y}_{x_1} f(x, y) = \frac{\overline{M}\, \delta_{x_1}(x) f(x, y)}{\overline{P}(x_1)} \quad \text{при} \quad \overline{M}\, \delta_{x_1} f \geqslant 0, \quad \overline{P}(x_1) > 0.$$

А далее смотрим, получится ли при подстановке этого выражения в правую часть (2.4) значение $\overline{M} f(x, y)$, и если это так для всех $f \in \mathcal{F}^{xy}$, то это дает совершенно веские основания считать, что $\mathcal{M}^{xy}$ разложима.

Теорема 2.2. О разложимости совместных моделей. *Если $\overline{P}(x) > 0$, $\forall x \in \mathcal{X}$, то для разложимости совместной модели $\mathcal{M}^{xy}$ на произведение $\mathcal{M}^{x}\mathcal{M}^{y}_{x}$ необходимо и достаточно выполнения при всех $x_1 \in \mathcal{X}$ и любых неотрицательных $f^{+}(x, y)$ из $\mathcal{F}^{xy}$ тождества*:

$$\overline{M}^{x_1} \left[\frac{\overline{M}\, \delta_{x_1}(x) f^{+}(x, y)}{\overline{P}(x_1)} \right] \equiv \overline{M} f^{+}(x, y),$$

где $\overline{M}^{x_1}$ есть среднее по частной $\mathcal{M}^{x}$. При этом переходная модель совпадает с условной.

Доказательство вынесено в дополнение 2 к параграфу.

Прокомментируем требование ненулевых верхних вероятностей $\overline{P}(x) > 0$ теоремы. В большинстве реальных задач число данных о явлении конечно, что соответствует моделям $\mathcal{M}^{xy}$ конечной размерности. Для них (если сразу исключить невозможные исходы) обязательно верхние вероятности отдельных исходов ненулевые. Нулевые же вероятности $\overline{P}(x) \equiv 0$ есть предельный случай при неограниченном увеличении точности модели. Исходя из

этого должна интерпретироваться теорема 2.2 и ее основное тождество.

Посылки теоремы 2.2 весьма серьезны и труднопроверяемы. Отметим одно простое свойство, необходимое для разложимости совместной модели. Оно состоит в том, что для всех $\varphi(y)\in\mathcal{F}^{xy}$ и c должно быть справедливо равенство

$$\overline{M}\delta_{x_1}(x)[\varphi(y)-c]=\begin{cases}\overline{M}\,\delta_{x_1}(x)\,\varphi(y)-c\overline{P}(x_1) & \text{при}\geqslant 0,\\ \overline{M}\,\delta_{x_1}(x)\,\varphi(y)-c\,\underline{P}(x_1) & \text{при}<0,\end{cases}$$

где условие $\geqslant 0$ и <0 относится к значению среднего слева. Рассматриваемое свойство разложимых моделей выводится точно так же, как это сделано при $c=0$ при выводе (2.5). Оно иллюстрируется рис. 2.6, где представлен график значений среднего как функции параметра сдвига c. В области положительных значений среднего, а конкретнее, при c таких, что $\overline{M}^{y}_{x_1}\varphi(y)\geqslant c$, это есть линейная функция c. Так же, как при отрицательных, соответствующих $\overline{M}^{y}_{x}\varphi(y)<c$. Между ними функция терпит излом.

Рассуждение останется верным, если $\varphi(y)$ заменить на $f(x, y)$. Тогда $\delta_{x_1}(x)[f(x, y)-c]$ есть, по сути, вертикальный дельта-вырез функции $f(x, y)-c$ по координате $x=x_1$, а рассматриваемое нами свойство на срезе — как линейность оператора $\overline{M}$ к параметру сдвига c, преломленная согласно рис. 2.6 при пересечении оси.

Первичные средние разложимых интервальных моделей. Пусть $\mathcal{M}^{xy}=\mathcal{M}^{x}\mathcal{M}^{y}_{x}$ есть разложимая совместная модель и пусть $\mathcal{M}^{x}=\langle\overline{M}^{x}\mathcal{H}\rangle$, $\mathcal{M}^{y}_{x}=\langle\overline{M}^{y}_{x}\Psi\rangle$ заданы верхними первичными значениями $\overline{M}^{x}h(x)$, $h\in\mathcal{H}$; $\overline{M}^{y}_{x}\psi(x, y)$, $\psi\in\Psi$, где $h(x)$ — частные признаки, а $\psi(x, y)$ — первичные признаки на $\mathcal{Y}$ при заданных x, называемые переходными.

Здесь изучается связь между первичными средними совместной модели и моделями-сомножителями.

Теорема 2.3. *Интервальная модель в виде произведения* $\mathcal{M}^{xy}=\langle\overline{M}^{x}\mathcal{H}\rangle\langle\overline{M}^{y}_{x}\Psi\rangle$ *определяется центрированными признаками*

$$\overset{\circ}{\mathcal{Y}}=\{h(x)-\overline{M}h,\ h\in\mathcal{H}\}\cup$$

$$\cup\{c^{+}(x)[\psi(x, y)-\overline{M}^{y}_{x}\psi],\psi\in\Psi,\quad\forall c^{+}(x)\}\qquad(2.5)$$

все с нулевыми первичными средними: $\overline{M}\overset{\circ}{g}=0$, $\forall\overset{\circ}{g}\in\overset{\circ}{\mathcal{G}}$, *причем*

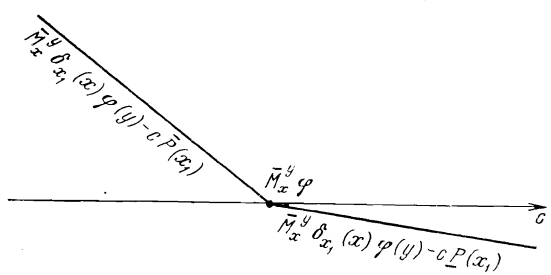


Рис. 2.6. Свойство разложимости модели

согласованным средним $\overline{M}^{y}_{x}\psi=\overline{M}^{y}_{x}\psi$ *соответствуют согласованные значения* $\overline{M}\overset{\circ}{g}=\overline{M}\overset{\circ}{g}$, $\overset{\circ}{g}=c^{+}(x)[\psi-\overline{M}^{y}_{x}\psi]$.

Прежде чем доказать теорему, дадим ее толкование. Обозначим $\overset{\circ}{h}(x)=h(x)-\overline{M}h$, $\overset{\circ}{\psi}(x, y)=\psi(x, y)-\overline{M}^{y}_{x}\psi$ — центрированные, т. е. приведенные к нулевым верхним средним $\overline{M}\overset{\circ}{h}=\overline{M}^{y}_{x}\overset{\circ}{\psi}=0$, первичные признаки как частной, так и переходной модели. Очевидно, $\langle\overline{M}^{x}\mathcal{H}\rangle=\langle\overline{M}^{x}\overset{\circ}{\mathcal{H}}\rangle$, $\langle\overline{M}^{y}_{x}\Psi\rangle=\langle\overline{M}^{y}_{x}\overset{\circ}{\Psi}\rangle$. Тогда теорема утверждает, что центрированными (с нулевыми средними) первичными признаками произведения останутся частные $\overset{\circ}{h}(x)\in\overset{\circ}{\mathcal{H}}$ признаки, дополненные совместными вида $\overset{\circ}{\psi}(x, y)c^{+}(x)$, $\overset{\circ}{\psi}\in\overset{\circ}{\Psi}$, равными центрированным переходным признакам, умноженным на произвольные неотрицательные функции $c^{+}(x)$ переменной x.

Для доказательства выпишем общее выражение среднего $\overline{M}f(x, y)$ разложимой модели через центрированные первичные значения сомножителей, реализующее согласно (2.4) двухшаговую процедуру вычисления:

$$\overline{M}f(x, y)=\overline{M}^{x}\overline{M}^{y}_{x}f(x, y)=$$

$$=\overline{M}^{x}C(x)=\inf\{d: d+\sum d^{+}_{j}\overset{\circ}{h}_{j}(x)\geqslant C(x)\},$$

где $C(x)=\inf\{c(x): c(x)+\sum c^{+}_{j}(x)\psi_j(x, y)\geqslant f(x, y)\}$. Сводя вместе два ограничения: одно на выбор d, а другое — на выбор $c(x)$, запишем их вместе, тогда окажется в наших руках заменить $C(x)$ на $c(x)$, сведя вычисление среднего к нахождению $\overline{M}f(x, y)=\inf\{d: d+\sum d^{+}_{j}\overset{\circ}{h}_{j}(x)\geqslant c(x),\ c(x)+\sum c^{+}_{j}(x)\times\psi_j(x, y)\geqslant f(x, y)\}$. В силу произвольности $c(x)$ первое ограничение вполне может быть заменено равенством, подставляя из которого $c(x)$ во второе ограничение, получаем $d+\sum d^{+}_{j}\overset{\circ}{h}_{j}(x)+\sum c^{+}_{j}(x)\overset{\circ}{\psi}(x, y)\geqslant f(x, y)$, что соответствует утверждаемым теоремой признакам, определяющим модель.

Замечание. В теореме 2.3 центрированные признаки $c^{+}(x)\overset{\circ}{\psi}(x, y)$ не при всех $c^{+}(x)$ будут обязательно согласованными и не все обязательно нужно считать первичными. Например, любая функция $c^{+}_{1}(x)$ есть первичный признак совместной модели, а форма $b^{+}c^{+}_{1}(x)+c$ дает хотя внешне другой, но фактически тот же самый признак. Сказанное относится и к $c^{+}_{j}(x)\overset{\circ}{\psi}_j(x, y)$, $b^{+}c^{+}_{j}(x)\overset{\circ}{\psi}_j(x, y)$, поэтому коэффициенты $c^{+}(x)$ можно каким-либо образом нормировать, например, полагая их принимающими значения от 0 до 1.

Из теоремы 2.3 следует, что за счет произвольности коэффициентов $c^{+}(x)$ как функций переменной x произведение моделей будет определяться значительно большим числом первичных значений, иметь большую размерность, нежели составляющие модели вместе взятые. В частности, размерность произведения ИМ несравненно выше размерностей сомножителей и может, в принципе, быть бесконечной.

Пример 2.10. Пусть $\underline{P}(A_1)$, $\bar{P}(A_1)$, $A_1 \subset \mathscr{X}$ — две первичные вероятности, определяющие $\mathscr{M}^x$, и пусть $\underline{P}_x(B_1)$, $\bar{P}_x(B_1)$, $B_1 \subset \mathscr{Y}$, — также две вероятности, задающие переходную ИМ размерности два $\mathscr{M}^{y}{}_x$. Нетрудно видеть, что и $\mathscr{M}^x$, и $\mathscr{M}^{y}{}_x$ есть ИРВ. Первичными признаками $\overset{\circ}{g}$ произведения $\mathscr{M}^x\mathscr{M}^{y}{}_x$, центрированными к нулевым средним, будут $A_1(x)-\bar{P}(A_1)$, $-A_1(x)+\underline{P}(A_1)$, $c^+(x)[B_1(y)-\bar{P}_x(B_1)]$, $c^+(x)[-B_1(y)+\underline{P}_x(B_1)]$ и для каждого из них $\overline{M}\overset{\circ}{g}=0$. Будем считать $0 \leqslant c^+(x) \leqslant 1$. Как видим, хотя сомножителями являются ИРВ, их произведение есть ИМ на $\mathscr{X}\times\mathscr{Y}$ с увеличенным числом центрированных первичных признаков, к которым, в частности, относятся

$$\overline{M}A(x)[B_1(y)-\bar{P}_x(B_1)]=0,\quad \overline{M}A(x)[-B_1(y)+\underline{P}_x(B_1)]=0$$

при любом событии $A \subset \mathscr{X}$. Чем шире пространство $\mathscr{X}$, т. е. чем больше элементарных исходов оно содержит, тем богаче делается состав набора первичных признаков, большей становится размерность совместной ИМ. Минимальной будет размерность при двух исходах $\mathscr{X}=\{x_1, x_2\}$. Тогда $A_1=x_1$ и первичных средних совместной ИМ всего шесть: четыре указаны выше двумя равенствами с подстановкой туда $A(x)=\delta_{x_1}(x)$ и $A(x)=\delta_{x_2}(x)$, и две исходные $\underline{P}(x_1)$ и $\bar{P}(x_1)$. При этом $\underline{P}_x(B_1)$ и $\bar{P}_x(B_1)$, в общем, могут быть разными при $x=x_1$ и $x=x_2$. При возрастании числа элементарных исходов пространства $\mathscr{X}$ размерность произведения неограниченно увеличивается, несмотря на то, что размерности сомножителей остаются равными двум.

Теорема 2.3 дает ответ на вопрос, какими групповыми свойствами должны обладать первичные признаки совместной ИМ для ее разложимости. Из этой теоремы вытекает.

Следствие. *Необходимым и достаточным условием разложимости $\mathscr{M}^{xy}$ является представимость ее центрированного набора первичных признаков в виде* (2.6).

Но трудности как раз состоят в представлении набора первичных признаков в виде (2.6). Очевидно, набор $\mathscr{G}$ должен быть достаточно богат. Одной из необходимых предпосылок разложимости является то, что наряду с первичным признаком g набору $\mathscr{G}$ должно принадлежать произведение $c^+(x)[g(x, y)-\bar{M}^{y}{}_x g]$, $\forall c^+(x) \geqslant 0$, где при $\bar{P}(x_1)>0$: $\bar{M}^{y}{}_{x_1} g=\bar{M}g(x, y)\delta_{x_1}(x)/\bar{P}(x_1)$ согласно (2.5).

Однако если $\mathscr{M}^{xy}$ не разложима, то всегда можно подыскать более широкую разложимую модель: $\mathscr{M}^x\mathscr{M}^{y}{}_x = \mathscr{M}^{xy}{}_* \supset \mathscr{M}^{xy}$. В худшем случае это будет голая совместная модель $\mathscr{J}^{xy}$, которая всегда разложима: $\mathscr{J}^{xy}=\mathscr{J}^x\mathscr{J}^{y}{}_x$, что легко проверяется.

Так как пересечение двух разложимых моделей не будет, в общем, разложимой моделью, то нельзя говорить о минимальной содержащей $\mathscr{M}^{xy}$ разложимой модели. Выбор разложимого расширения оказывается неоднозначным.

Вообще, напрашивается вывод, что раз разложимыми являются совместные модели очень ограниченного класса, то произведение не есть единый способ представления, а всего лишь удобный прием задания моделей совместных явлений, отражающий физическую природу перехода одного в другое.

Замечания. 1. Сказанное вступает в диссонанс с общеизвестными свойствами точных распределений вероятностей, всегда разложимых, поскольку для них определены точные условные (они же переходные) распределения вероятностей. Так, для точных вероятностей $P(x_i, y_i)$ на дискретных пространствах $\mathscr{X}$ и $\mathscr{Y}$ условное распределение получается по хорошо известной формуле

$$P_{x_i}(y_j)=P(x_i, y_j)/P(x_i),$$

в знаменателе которой стоит частное распределение (то же и для плотностей вероятностей на непрерывных пространствах). Внутри этого (узкого) класса моделей переход к условным (апостериорным) распределениям прост и универсален.

2. Описания моделей семействами распределений вероятностей расширяет возможности как самих распределений, так и их разложений, но и здесь имеется барьер в виде громоздкости такого типа описаний. В самом деле, в описаниях ИМ будут обязательно присутствовать фразы типа: «Все те распределения вероятностей, для которых $\bar{M}g \in [\underline{M}g, \bar{M}g]$, $g \in \mathscr{G}$, или же для которых $P(B) \in [\underline{P}(B), \bar{P}(B)]$, $B \in \mathscr{B}$, ...», и слово «все» во многом делает неконкретным, неконструктивным поэлементный состав модели. В частных случаях облегчение достигается сокращением семейств до обозримых размеров, но возможности такого упрощения охватывают, в основном, узкий класс моделей в виде параметрических семейств распределений вероятностей.

Подчиненные произведения. Предыдущими двумя замечаниями подготовлена почва для более широкого использования произведений моделей в совокупности с множественным их описанием как объединений семейств. Пусть совместная модель задана в следующем виде:

$$\mathscr{M}^{xy}=\bigvee_\theta \mathscr{M}^{xy}_\theta=\bigvee_\theta \mathscr{M}^x_\theta \mathscr{M}^{y}_{\theta,x},$$

где правая часть называется *подчиненным параметру* θ *произведением моделей*. Здесь заведомо семейство $\mathscr{M}^{xy}{}_\theta$, $\theta \in \Theta$, выбирается из разложимого класса.

Подчиненное произведение есть сокращенная запись следующего представления средних:

$$\overline{M}f(x, y)=\sup_\theta \overline{M}^x_\theta \overline{M}^{y}_{\theta, x} f(x, y).$$

Замечание. Произведение моделей $\mathscr{M}^x\mathscr{M}^{y}{}_x$, если каждая есть объединение: $\mathscr{M}^x=\bigvee_\theta \mathscr{M}^x{}_\theta$, $\mathscr{M}^{y}{}_x=\bigvee_\vartheta \mathscr{M}^{y}{}_{\vartheta,x}$, может рассматриваться как двойное объединение $\bigvee_\theta\bigvee_\vartheta \mathscr{M}^x{}_\theta \mathscr{M}^{y}{}_{\vartheta x}$, в котором θ и ϑ пробегают свои значения (возможно, из одного и того же множества Θ) несвязанно друг с другом. В этом состоит отличие от подчиненного произведения, для которого $\theta=\vartheta$ и, следовательно, параметры θ и ϑ сомножителей в значениях синхронны друг другу.

Введение подчиняющего параметра θ обретает наглядность и даже естественность, если он имеет физическую интерпретацию как параметр связи средних частной и переходных моделей, как параметр влияния на оба явления $\mathscr{X}$ и $\mathscr{Y}$ какого-либо одного постороннего фактора, описываемого неизвестным значением θ.

Например, пусть $\mathscr{X}$ описывает количество перегноя в почве, а $\mathscr{Y}$ — урожайность, скажем, травы. Эти два фактора будут зависеть от погодных условий, например от количества осадков. Это количество и может служить подчиняющим параметром θ.

Избавиться от влияния подчиняющего параметра можно, переходя к более широкой модели на основании следующего включения:

$$\bigvee_{\theta} \mathscr{M}^x_\theta \mathscr{M}^y_{\theta,\,x} \subset \Big(\bigvee_{\theta} \mathscr{M}^x_\theta\Big)\Big(\bigvee_{\theta} \mathscr{M}^y_{\theta,\,x}\Big).$$

Точно так же расширением можно избавиться от влияния x в переходной модели произведения $\mathscr{M}^x\mathscr{M}^y{}_x$, заменив $\mathscr{M}^y{}_x$ объединением по x. К чему мы таким образом придем, видно из следующего заголовка.

Свободные произведения. Пусть $\mathscr{X}$ и $\mathscr{Y}$ дают исходы двух различных явлений и на $\mathscr{X}\times\mathscr{Y}$ задана совместная $\mathscr{M}^{xy}$. Относительно нее вводится определение.

Явление $\mathscr{Y}$ называется свободным от $\mathscr{X}$, если совместная $\mathscr{M}^{xy}$ разлагается на произведение

$$\mathscr{M}^{xy} = \mathscr{M}^x\,\mathscr{M}^y,\quad \overline{M}f(x,\,y) = \overline{M}^x\,\overline{M}^y f(x,\,y),\quad \forall f \in \mathscr{F}^{xy}, \tag{2.7}$$

в котором переходная $\mathscr{M}^y{}_x = \mathscr{M}^y$ не зависит от x. Такое *произведение моделей называется свободным*. Частными для $\mathscr{M}^x\mathscr{M}^y$, очевидно, будут $\mathscr{M}^x$ и $\mathscr{M}^y$.

При $\mathscr{Y}$ свободном от $\mathscr{X}$ совместное среднее $\overline{M}^{xy}$ обладает свойствами:

$$\overline{M}\,f^+(x)\,\varphi^+(y) = \overline{M}f^+(x)\,\overline{M}\,\varphi^+(y);$$

$$\overline{M}\,[g(x) + f^+(x)\,\varphi(y)] = \overline{M}\,[g(x) + f^+(x)\,\overline{M}\,\varphi(y)];$$

$$\overline{M}\,[f(x) + \varphi(y)] = \overline{M}\,f(x) + \overline{M}\,\varphi(y).$$

Свобода $\mathscr{Y}$ от $\mathscr{X}$ имеет тот смысл, что, зная исход x явления на $\mathscr{X}$, ничего нового нельзя сказать относительно статистической модели на исходах последующего за ним явления $\mathscr{Y}$. Тем не менее из самой структуры формулы (2.7) следует, что каждый исход x может влиять на процедуру выбора исходов явления $\mathscr{Y}$ в рамках $\mathscr{M}^y$. Чем шире $\mathscr{M}^y$, тем большая степень такого влияния может иметь место. При точных $\mathscr{M}^y$ на дискретных $\mathscr{Y}$ свобода, как нетрудно видеть, эквивалентна независимости в классическом понимании.

Таким образом, понятие свободы обретает свой смысл лишь для неточных моделей, т. е. если есть выбор, неопределенность.

Рассмотрим, как будут «звучать фразы о свободе» в ракурсе первичных признаков. Из теоремы 2.3 и ее следствия имеем.

Утверждение 2.4. *Для того чтобы явление на $\mathscr{Y}$ было свободным от $\mathscr{X}$, необходимо и достаточно, чтобы набор первичных признаков совместной $\mathscr{M}^{xy}$ мог быть приведен к виду*

$$\mathring{\mathscr{g}} = \{h(x) - \bar{M}h,\ h \in \mathscr{H}\} \cup \{c^+(x)\,[\psi(y) - \bar{M}\psi],\quad \psi \in \Psi,$$
$$\forall c^+(x) \geqslant 0\},$$

все с нулевыми средними: $\bar{M}\mathring{g} \equiv 0$, $\mathring{g} \in \mathring{\mathscr{G}}$. *Тогда*

$$\mathscr{M}^{xy} = \langle \bar{M}\ \mathring{\mathscr{g}}\rangle = \langle \bar{M}^x\,\mathscr{H}\rangle\ \langle \bar{M}^y\,\Psi\rangle.$$

Понятие свободы не симметрично. А именно, произведение $\mathscr{M}^x\mathscr{M}^y$ есть не то же самое, что $\mathscr{M}^y\mathscr{M}^x$. Например, пусть $\mathscr{M}^x = \langle \bar{M}h\rangle$, $\mathscr{M}^y = \langle \bar{M}\psi\rangle$ — две частные модели, определенные каждая одним своим первичным средним $\bar{M}h$ и $\bar{M}\psi$. Тогда первичными признаками свободного произведения $\mathscr{M}^x\mathscr{M}^y$, приведенными к нулевым верхним средним, будут

$$\mathring{\mathscr{g}}_1 = \{h(x) - \bar{M}h\} \cup \{c^+(x)\,[\psi(y) - \bar{M}\,\psi],\quad \forall c^+(x) \geqslant 0\},$$

а перевернутого произведения $\mathscr{M}^y\mathscr{M}^x$ — будут уже другие признаки

$$\mathring{\mathscr{g}}_2 = \{d^+(y)\,[h(x) - \bar{M}h],\quad \forall d^+(y) \geqslant 0\} \cup \{\psi(y) - \bar{M}\,\psi\}.$$

Первая модель соответствует свободе $\mathscr{Y}$ от $\mathscr{X}$, а вторая — $\mathscr{X}$ от $\mathscr{Y}$. Разницу между ними поясним примером.

Пример 2.11. Пусть $\mathscr{X} = \{x_1, x_2\}$, $\mathscr{Y} = \{y_1, y_2\}$, при этом $\mathscr{M}^x$ определяется первичной вероятностью $\bar{P}(x_1)$, а на $\mathscr{Y}$ ИМ голая $\mathscr{M}^y = \mathscr{I}^y$. Тогда для произведения $\mathscr{M}^x\mathscr{I}^y$, относительно которого $\mathscr{Y}$ свободен от $\mathscr{X}$, первичный признак будет всего один $\delta_{x_1}(x)$ и для конкретного произведения признаков имеем

$$\overline{M}^x\,\overline{M}^y\,[\delta_{x_1}(x) - a]\,\delta_{y_1}(y) = \max_{x,\,y}\,[\delta_{x_1}(x) - a]\,\delta_{y_1}(y) = 1 - a.$$

При вычислениях здесь, как видно, первичная вероятность $\bar{P}(x_1)$ не участвовала. Для произведения $\mathscr{I}^y\mathscr{M}^x$ с переставленными сомножителями уже $\mathscr{X}$ будет свободен от $\mathscr{Y}$, а первичными будут $[d_1 + \delta_{y_1}(y) + d_2\delta_{y_2}(y)]\,[\delta_{x_1}(x) - \bar{P}(x_1)]$, поэтому $\overline{M}^y\overline{M}^x[\delta_{x_1}(x) - \bar{P}(x_1)]\,\delta_{y_1}(y) = \overline{M}^y\delta_{y_1}(y)\,\overline{M}^x[\delta_{x_1}(x) - \bar{P}(x_1)] = 0$. Это отличается от предыдущего при $a = \bar{P}(x_1)$.

Таким образом, свобода $\mathscr{Y}$ от $\mathscr{X}$ не тождественна свободе $\mathscr{X}$ от $\mathscr{Y}$. Хотя у свободных произведений обоих видов одинаковыми будут средние на признаках вида $f^+(x)\varphi^+(y)$:

$$\overline{M}^x\,\overline{M}^y\,f^+(x)\,\varphi^+(y) = \overline{M}^y\,\overline{M}^x\,f^+(x)\,\varphi^+(y),$$

но, в общем, они будут разными на совместных признаках $f(x, y)$.

Если эксперимент $\mathscr{Y}$ связать с поведением человека, то вкладываемый в слово «свобода» $\mathscr{Y}$ «житейский» смысл подобен фразам: «как хочет, так и поступает», «что волен, то и делает», как это видно из следующего примера.

Пример 2.12. Рассмотрим такую реальную ситуацию. Пусть первый раз монета подбрасывается так, что вероятность исхода равна 1/2, образуя экспе-

римент $\mathscr{X}$ с двумя исходами. Другой раз монета не подбрасывается, а показывается так или иначе некоторым лицом. Показ производится осмысленным образом, поэтому вероятность герба может быть любой от 0 до 1, в результате $\mathscr{M}^y = \mathscr{J}^y$. Ясно, что эксперимент $\mathscr{Y}$ свободен от $\mathscr{X}$, так что можно записать $\mathscr{M}^{xy} = \mathscr{M}^x \mathscr{J}^y$. Тем не менее свобода здесь не означает независимости: решение относительно того, какую монету показать, может созреть на основании результата подбрасывания. Свобода лишь означает, что о намерениях лица, показывающего монету, ничего не известно: они могут быть любыми.

Если поменять в этом примере последовательность действий: сначала производить эксперимент $\mathscr{Y}$ (показ монеты), а затем $\mathscr{X}$ (случайное подбрасывание), то получим совершенно другое произведение $\mathscr{J}^y \mathscr{M}^x$, относительно которого $\mathscr{X}$ не только свободен от $\mathscr{Y}$, но и более того, не зависит от $\mathscr{Y}$ в том смысле, который будет дан в следующем параграфе.

Любой совместной модели, разложимой на произведение $\mathscr{M}^x \mathscr{M}^y{}_x$, всегда можно подыскать минимальную более широкую $\mathscr{M}^{xy}{}_*$, относительно которой $\mathscr{Y}$ был бы свободен от $\mathscr{X}$. Для этого надо взять объединение переходных ИМ $\mathscr{M}^y{}_* = \bigvee_{\forall x} \mathscr{M}^y{}_x$ и образовать произведение $\mathscr{M}^{xy}{}_* = \mathscr{M}^x \mathscr{M}^y{}_*$. Наконец, для голой совместной модели имеет место равенство $\mathscr{J}^{xy} = \mathscr{J}^x \mathscr{J}^y = \mathscr{J}^y \mathscr{J}^x$, так что $\mathscr{Y}$ всегда будет свободен от $\mathscr{X}$ и наоборот.

Дополнения.

1. Теорема. *Если $\bar{P}(x) > 0$, $\forall x \in \mathscr{X}$, то для разложения совместной $\mathscr{M}^{xy}$ на произведение $\mathscr{M}^x \mathscr{M}^y{}_x$ необходимо и достаточно выполнения тождества*

$$\overline{M}\,\overline{\overline{M}}_x f(x, y) \equiv \overline{M} f(x, y), \quad \forall f \in \mathscr{F}^{xy},$$

где $\overline{\overline{M}}_x$ есть условные средние при заданных $x \in \mathscr{X}$.

Докажем это. Достаточность очевидна из равенства (2.4). Для доказательства необходимости требуется доказать, что условные средние $\overline{\overline{M}}_x$ совпадают с переходными $\overline{M}^y{}_x$, определенными формулой (2.5), т. е. требуется доказать равенство

$$\max_{\underline{P}(x_1) \leqslant P(x_1) \leqslant \overline{P}(x_1)} \frac{\overline{M}_{P(x_1)}\, \delta_{x_1} \varphi}{P(x_1)} = \frac{\overline{M}\, \delta_{x_1} \varphi}{\overline{P}(x_1)}, \qquad (*)$$

где $\bar{P}(x_1) > 0$ и $\overline{M}\delta_{x_1}\varphi \geqslant 0$. В левой части стоит условное среднее $\overline{\overline{M}}_x \varphi$, расписанное по формуле (1.15), и $\overline{M}_{P(x_1)}$ есть среднее по сечению $\mathscr{M}^{xy} \wedge \langle P(x_1) \rangle$, согласно формуле (1.9) равное

$$\overline{M}_{P(x_1)}\, \delta_{x_1} \varphi = \min_c \{ \overline{M} [\varphi(y) - c]\, \delta_{x_1}(x) + cP(x_1) \}. \qquad (**)$$

На основании записи (2.4) и свойств средних переходной ИМ имеет место равенство

$$\begin{aligned} \overline{M}[\varphi(y) - c]\, \delta_{x_1}(x) &= \overline{M}^x \overline{M}^y_x [\varphi(y) - c]\, \delta_{x_1}(x) = \\ &= \overline{P}(x_1)\, \overline{M}^y_{x_1} [\varphi(y) - c] = \overline{P}(x_1)\, \overline{M}^y_{x_1} \varphi(y) - c\overline{P}(x_1) = \\ &= \overline{M}\, \delta_{x_1}(x)\, \varphi(y) - c\overline{P}(x_1), \qquad (***) \end{aligned}$$

при $c \leqslant \overline{M}\, \delta_{x_1}(x)\, \varphi(y) / \overline{P}(x_1)$,

где последнее неравенство гарантирует $\overline{M}[\varphi(y) - c]\delta_{x_1}(x) \geqslant 0$, $\overline{M}^y{}_x [\varphi(y) - c] \geqslant 0$. Подстановка равенства $(***)$ в $(**)$ дает

$$\begin{aligned} \overline{M}_{P(x_1)}\, \delta_{x_1}(x)\, \varphi(y) &= \min_{c \leqslant \overline{M}\delta_{x_1}\varphi / \overline{P}(x_1)} [\overline{M}\, \delta_{x_1} \varphi - c\overline{P}(x_1) + cP(x_1)] = \\ &= \overline{M}\, \delta_{x_1} \varphi [P(x_1) / \overline{P}(x_1)]. \end{aligned}$$

Отсюда очевидно становится $(*)$, что и требовалось.

2. Доказательство теоремы 2.2. Необходимость следует из формулы (2.5), которая не изменится, если заменить в ней $\varphi(y)$ на $f(x, y)$. Для доказательства достаточности нужно формально определить переходные средние для $f^+(x, y)$ по указанной перед теоремой модификации формулы (2.5) и продолжить по свойству переноса на все ограниченные функции и далее устремлением уровней усечений к бесконечности — на неограниченные из $\mathscr{F}^{xy}$. Последняя часть теоремы 2.2 о совпадении переходной модели с условной вытекает из теоремы дополнения 1.

3. Как это видно из рис. 2.6, при $\underline{P}(x_1) > 0$ границы переходной $\overline{M}^y{}_{x_1} \varphi(y)$ являются решениями относительно m уравнения

$$\overline{M}[\varphi(y) - m]\, \delta_{x_1}(x) = 0.$$

Если же $\bar{P}(x_1) > 0$, но не исключается $\underline{P}(x_1) = 0$, тогда

$$\overline{M}^y_{x_1} \varphi(y) = \inf \{ c : \overline{M}[\varphi(y) - c]\, \delta_{x_1}(x) \leqslant 0 \}.$$

Наконец, учитывая, что $\overline{M}[\varphi(y) - c]\delta_{x_1}(x)$ в области положительных ее значений есть прямая по c, находим $\overline{M}^y{}_{x_1} \varphi$ как пересечение этой прямой (определяемой ее значениями при любых двух c_1 и c_2) с осью, так что для $\forall c_2 > c_1$

$$\overline{M}^y_{x_1} \varphi = \frac{c_1 + \overline{M}(\varphi - c_2)\, \delta_{x_1} - c_2\, \overline{M}(\varphi - c_1)\, \delta_{x_1}}{\overline{M}(\varphi - c_2)\, \delta_{x_1} - \overline{M}(\varphi - c_1)\, \delta_{x_1}}.$$

4. При $\bar{P}(x) > 0$, $\forall x$, для разложения совместной модели на произведение необходимым является выполнение следующего тождества:

$$\overline{M}\left[f(x, y) - \frac{\overline{M} f(x, y)\, \delta_{x_1}(x)}{\overline{P}(x_1)} \right] c^+(x) \equiv 0, \quad \forall c^+(x).$$

В самом деле, используя формулу, предшествующую теореме 2.2, имеем

$$\begin{aligned} \overline{M}[f - \overline{M} f\, \delta_x / \overline{P}(x)]\, c^+(x) &= \overline{M}^x\, \overline{M}^y_x \left(f - \overline{M}^y_x f \right) c^+(x) = \\ &= \overline{M}^x\, c^+(x) \left[\overline{M}^y_x f - \overline{M}^y_x f \right] \equiv 0. \end{aligned}$$

2.5. НЕЗАВИСИМОСТЬ

Определение независимости. Независимость — это отсутствие как и незнание взаимных связей исходов разных явлений, отраженное в определенных свойствах совместных моделей. Мы дадим формальное определение независимости как свойств средних, а обсуждать его адекватность нашим представлениям будем в процессе изложения. Замечательно то, что вводимое нами по-

нятие независимости действует и для неустойчивых в статистическом смысле явлений.

Два явления, одно с исходами $\mathscr{X}$ и другое — $\mathscr{Y}$ называются независимыми (или сокращенно, $\mathscr{X}$ и $\mathscr{Y}$ — независимы), если интервальное среднее произведений частных (разделенных по переменным) признаков равно произведениям интервальных средних, т. е. верна формула интервальной мультипликативности средних:

$$\overline{M}f(x)\varphi(y)=\overline{Mf(x)\,M\varphi(y)},\quad \forall f(x),\ \varphi(y)\subset\mathcal{F}^{xy}. \tag{2.8}$$

Здесь произведение средних справа раскрывается согласно правилам интервальной арифметики (§ 2.2), а именно:

$$\overline{Mf}\overline{M\varphi}=\max\{\overline{M}f\overline{M}\varphi,\ \overline{M}f\underline{M}\varphi,\ \underline{M}f\overline{M}\varphi,\ \underline{M}f\underline{M}\varphi\}- \tag{2.9}$$

— это максимальное число, которое может быть получено умножением чисел двух интервалов $\underline{M}f$, $\overline{M}f$ и $\underline{M}\varphi$, $\overline{M}\varphi$, причем допускаются бесконечные значения (если f принадлежит, а $-f$ не принадлежит области существования $\mathcal{F}^{xy}$). Аналогично нижнее среднее $\underline{Mf}(x)\varphi(y)$ будет равно минимуму в правой части (2.9), обозначаемому $\underline{Mf}\underline{M}\varphi$.

При некоторых положениях интервалов на числовой оси произведения средних упрощаются (аргументы для краткости опускаются и ниже всюду f — признак на $\mathscr{X}$, φ — на $\mathscr{Y}$):

$$\underline{M}f\geqslant 0,\quad \underline{M}\varphi\geqslant 0\Rightarrow \underline{Mf}\underline{M}\varphi=\underline{M}f\underline{M}\varphi,$$

$$\underline{M}f\geqslant 0,\quad \overline{M}\varphi\geqslant 0\Rightarrow \overline{Mf}\overline{M}\varphi=\overline{M}f\overline{M}\varphi,$$

$$\underline{M}f\geqslant 0,\quad \overline{M}\varphi\leqslant 0\Rightarrow \overline{Mf}\overline{M}\varphi=\underline{M}f\overline{M}\varphi,$$

$$\overline{M}f\geqslant 0,\quad \overline{M}\varphi\leqslant 0\Rightarrow \underline{Mf}\underline{M}\varphi=\overline{M}f\underline{M}\varphi.$$

Так, для неотрицательных функций f^+ и φ^+ всегда имеет место самый первый случай и $\underline{M}f^+\underline{M}\varphi^+=\underline{M}f^+\underline{M}\varphi^+$, $\overline{M}f^+\varphi^+=\overline{M}f^+\overline{M}\varphi^+$. В частности, для событий $A\subset\mathscr{X}$, $B\subset\mathscr{Y}$, подставляя их индикаторные функции $A(x)$, $B(y)$ в (2.9), получаем $\underline{P}(A,B)=\underline{P}(A)\underline{P}(B)$, $\overline{P}(A,B)=\overline{P}(A)\overline{P}(B)$.

Подчеркнем, что сами по себе эти пропорции между вероятностями не могут служить определением независимости (кроме случая точных вероятностей на дискретных пространствах), так как события — это всего лишь узкий подкласс из огромного разнообразия признаков. А независимость — емкое понятие, охватывающее все вместе признаки.

Если $\mathscr{I}^{xy}$ — голая совместная ИМ на $\mathscr{X}\times\mathscr{Y}$, то для всех признаков: $\underline{M}f=\inf f$, $\overline{M}f=\sup f$, и здесь нетрудно убедиться в следующем:

$$\underline{M}f\varphi\equiv\inf f\varphi=\underline{Mf}\underline{M}\varphi,\quad \overline{M}f\varphi=\sup f\varphi=\overline{Mf}\overline{M}\varphi,$$

где инфимум и супремум берутся по переменным x и y. Таким образом, *если никаких статистических данных об исходах $\mathscr{X}$ и $\mathscr{Y}$ нет, то явления независимы.* На первый взгляд несколько странный, но совершенно естественный, как выяснится дальше, вывод, вытекающий из общего правила: чем меньше мы знаем о явлении, чем более размытыми являются средние, тем меньше ценности несет в себе понятие независимости.

Если Mf и $M\varphi$ являются точными, то интервалы средних превращаются в точки, а интервальное произведение (2.8) в обычное: $Mf\varphi=MfM\varphi$.

Свойства независимости. Выведем свойства, позволяющие с разных сторон взглянуть на понятие независимости.

Свойство равноправия. *Явления $\mathscr{X}$ и $\mathscr{Y}$ равноправны по отношению к независимости: если $\mathscr{X}$ не зависит от $\mathscr{Y}$, то и $\mathscr{Y}$ не будет зависеть от $\mathscr{X}$.*

Свойство мультипликативности на сечениях. *Если $\mathscr{X}$ и $\mathscr{Y}$ независимы, то для Mf, $M\varphi$-сечения $\mathscr{M}^{xy}_{Mf\,M\varphi}$ совместной модели $\mathscr{M}^{xy}$ среднее произведения $f\varphi$ частных признаков является точным и равняется произведению средних:*

$$M_{Mf,\,M\varphi}f\varphi=MfM\varphi.$$

В самом деле, согласно формуле (1.10) для сечений имеем

$$\overline{M}_{Mf,\,M\varphi}f\varphi=\min_{c_1,c_2}[\overline{M}(f\varphi-c_1f-c_2\varphi)+c_1Mf+c_2M\varphi].$$

Перегруппируем выражение в квадратных скобках, записав его

$$[\overline{M}(f-c_1)(\varphi-c_2)-(Mf-c_1)(M\varphi-c_2)+MfM\varphi]=$$
$$=[\overline{M(f-c_1)\,M(\varphi-c_2)}-(Mf-c_1)(M\varphi-c_2)]+Mf\,M\varphi,$$

где использовано (2.8) и общая черта раскрывается согласно (2.9). Нетрудно теперь убедиться, что минимум квадратных скобок последнего выражения по c_1, c_2 равен 0.

Смысл доказанного свойства в том, что если бы мы вдруг узнали точно средние Mf, $M\varphi$ и добавили бы их в модель $\mathscr{M}^{xy}$ независимых $\mathscr{X}$ и $\mathscr{Y}$, то получили бы привычное свойство независимых явлений: среднее произведения $f\varphi$ равно произведению средних.

Свойство неизменности условных моделей. *Для независимых $\mathscr{X}$ и $\mathscr{Y}$ условная модель $\mathscr{M}^y_A$ на $\mathscr{Y}$ при условии, что случилось событие $A\subset\mathscr{X}$, не зависит от A и равна частной:*

$$\mathscr{M}^y_A=\mathscr{M}^y,\quad \forall A\subset\mathscr{X}.$$

Здесь $\mathscr{M}^y{}_A$ — частная к условной совместной модели $\mathscr{M}^{xy}{}_A$.

В самом деле, так же как для предыдущего свойства (полагая $c_2=0$) доказывается для $\underline{P}(A)=MA(x)$-сечения совместной модели равенство $\overline{M}_{P(A)}A(x)\varphi(y)=P(A)\overline{M}\varphi(y)$ и далее нужно применить формулу (1.15) для условного среднего, убеждаясь в $\overline{M}_A\varphi(y)=\overline{M}\varphi(y)$.

Таким образом, если $\mathscr{X}$ и $\mathscr{Y}$ независимы, то какие бы наблюдения ни велись за $\mathscr{X}$ (в виде свершившихся конкретных x или нечетких событий) о частной модели на $\mathscr{Y}$, все равно ничего нового не будет известно. Аналогично будет, если пространства $\mathscr{X}$ и $\mathscr{Y}$ переставить местами (в силу свойства равноправия).

Замечание. Само по себе свойство неизменности условных моделей $\mathscr{M}^y{}_A$ (как их одинаковость по $A\subset\mathscr{X}$) выполняется и для случая, когда $\mathscr{Y}$ свободен от $\mathscr{X}$, т. е. $\mathscr{M}^{xy}=\mathscr{M}^x\mathscr{M}^y$, поэтому нельзя положить это свойство в основу эквивалентного определения независимости.

Следующим является свойство аддитивности. *Среднее суммы разделенных по переменным частных признаков равно сумме средних*: $\overline{M}[\hat{f}(x)+\varphi(y)]=\overline{M}\hat{f}(x)+\overline{M}\varphi(y)$.

В самом деле, так как прибавление постоянных равенства не меняет, то, считая $\underline{M}f>0$, $\underline{M}\varphi>0$ и используя дважды (2.8), получаем

$$\overline{M}f\,\overline{M}\varphi+\overline{M}f+\overline{M}\varphi+1=\overline{M}(f+1)(\varphi+1)=$$

$$=\overline{M}(f\varphi+f+\varphi+1)\leqslant\overline{M}f\,\varphi+\overline{M}(f+\varphi)+1=\overline{M}f\,\overline{M}\varphi+\overline{M}(f+\varphi)+1,$$

откуда $\overline{M}f+\overline{M}\varphi\leqslant\overline{M}(f+\varphi)$, а так как обратное неравенство есть свойство средних, то это доказывает равенство.

Характеризационным для независимости, как выяснится из дальнейшего, является свойство взаимной нековариированности частных признаков: *если $\mathscr{X}$ и $\mathscr{Y}$ независимы, то нулевыми всегда будут верхние средние следующих произведений частных признаков $\hat{f}(x)$, $\varphi(y)\in\mathscr{F}^{xy}$*:

$$\overline{M}(f-\underline{M}f)(\varphi-\overline{M}\varphi)=\overline{M}(f-\overline{M}f)(\varphi-\underline{M}\varphi)=$$

$$=\overline{M}(\overline{M}f-f)(\varphi-\overline{M}\varphi)=\overline{M}(\underline{M}f-f)(\varphi-\underline{M}\varphi)=0,\quad\forall f,\ \varphi. \qquad (2.10)$$

Равенства (2.10) проверяются непосредственным использованием (2.8). Например,

$$\overline{M}(f-\underline{M}f)(\varphi-\overline{M}\varphi)=\overline{M(f-\underline{M}f)\,M(\varphi-\overline{M}\varphi)}=$$

$$=\max\{(\overline{M}f-\underline{M}f)(\overline{M}\varphi-\overline{M}\varphi),\quad(\overline{M}f-\underline{M}f)(\underline{M}\varphi-\overline{M}\varphi),$$

$$(\underline{M}f-\underline{M}f)(\overline{M}\varphi-\overline{M}\varphi),\quad(\underline{M}f-\underline{M}f)(\underline{M}\varphi-\overline{M}\varphi)\}=0.$$

Замечание. В левых частях равенств (2.10) участвуют произведения в известном смысле центрированных признаков. Средние этих произведений называются *ковариациями* $\hat{f}$ и φ. Их четыре, поэтому (2.10) есть обобщение известного для точных средних свойства равенства нулю ковариации независимых случайных величин: $M(f-Mf)(\varphi-M\varphi)=0$.

Независимое произведение. В тождестве (2.8), составляющем свойство независимости, справа стоит произведение средних, определяемых по частным моделям $\mathscr{M}^x$ и $\mathscr{M}^y$. Однако это отнюдь не означает, что частные модели $\mathscr{M}^x$ и $\mathscr{M}^y$ полностью должны определять совместную $\mathscr{M}^{xy}$. Нет, они обязаны согласно (2.8) определять только часть ее средних, а именно, на произведениях разделенных по переменным признаков. Ничто при этом не мешает еще знать средние $\overline{M}g(x,\ y)$ на неразделяемых признаках более точно, чем это можно было бы сделать продолжением разделенных.

Текущий раздел посвящается тому случаю, когда совместная модель $\mathscr{M}^{xy}$ однозначно определяется частными моделями независимых $\mathscr{X}$ и $\mathscr{Y}$, что даст наиболее широкую совместную модель независимых $\mathscr{X}$ и $\mathscr{Y}$ при фиксированных частных. Ее можно интерпретировать как отсутствие каких-либо «посторонних» взаимных средних: исследуются отдельно $\mathscr{X}$, $\mathscr{Y}$, и при условии, что они независимы, ставится вопрос, какая совместная модель получится? Определим ее.

Совместная модель $\mathscr{M}^{xy}$, первичными для которой являются средние (2.8) всевозможных произведений частных (разделенных по переменным) признаков, называется *независимым произведением* $\mathscr{M}^x$ и $\mathscr{M}^y$, и обозначается

$$\mathscr{M}^{xy}=\mathscr{M}^x\times\mathscr{M}^y,$$

$$\overline{M}^{xy}g(x,\ y)=(\overline{M}^x\times\overline{M}^y)\,g(x,\ y),\quad g(x,\ y)\in\mathscr{F}^{xy}.$$

Пусть частные модели заданы своими первичными средними: $\mathscr{M}^x=\langle\overline{M}^x\mathscr{H}\rangle$, $\mathscr{M}^y=\langle\overline{M}^y\Psi\rangle$, где $\mathscr{H}$ — набор частных признаков $h(x)$ на $\mathscr{X}$, а Ψ — признаков $\psi(y)$ на $\mathscr{Y}$. Процедура вычисления совместных средних $\overline{M}g$ для независимого произведения $\langle\overline{M}^x\mathscr{H}\rangle\times\langle\overline{M}^y\Psi\rangle$ разбивается на три этапа. Первый — это продолжение первичных средних на все частные признаки, мажорируемые конечными линейными комбинациями первичных, что дает $\overline{M}^xf(x)$, $f(x)\in\mathscr{L}^+\mathscr{H}$, и $\overline{M}^y\varphi(y)$, $\varphi(y)\in\mathscr{L}^+\Psi$. Второй — это вычисление по формуле (2.8) средних $\overline{M}f\varphi$ их произведений с учетом формулы обращения $\underline{M}^xf=-\overline{M}^x(-f)$ (если $-f\notin\mathscr{L}^+\mathscr{H}$, то считается $\underline{M}^xf=-\infty$). Средние $\overline{M}f\varphi$ будут конечными при $\pm f\in\mathscr{L}^+\mathscr{H}$ и $\pm\varphi\in\mathscr{L}^+\Psi$. Наконец, третий шаг — это по набору из всех $\overline{M}f\varphi$, первичному для совместной модели, вычисление $\overline{M}g(x,\ y)$ уже для любых g по формуле

$$\overline{M}g(x,\ y)=\inf_{c+\Sigma c_if_i(x)\varphi_i(y)\geqslant g(x,\ y)}[c+\Sigma c_i^+\overline{Mf_iM\varphi_i}].$$

Вытекающий из указанной многошаговой процедуры вывод состоит в том, что размерность модели, т. е. минимальное число определяющих ее первичных признаков, для независимого произведения $\mathscr{M}^x\times\mathscr{M}^y$ определяется не столько сомножителями, сколько размерами пространств $\mathscr{X}$ и $\mathscr{Y}$. Если число элементов этих пространств бесконечно, то независимо от размерностей $\mathscr{M}^x$ и $\mathscr{M}^y$ размерность их независимого произведения будет, в общем, бесконечной (за некоторым исключением, например, $\mathscr{T}^x\times\mathscr{T}^y$).

Пример 2.13. Пусть $A\subset\mathscr{X}$ и $B\subset\mathscr{Y}$ есть два события и $\bar{P}(A)$, $\bar{P}(B)$ — вероятности, являющиеся первичными для своих частных моделей размерно-

ти 1 соответственно $\mathscr{M}^x=\langle \bar{P}(A)\rangle$ и $\mathscr{M}^y=\langle \bar{P}(B)\rangle$. Покажем, что независимое произведение $\mathscr{M}^x\times\mathscr{M}^y$ будет иметь, в общем, бесконечную размерность (исключаем случаи, когда $\bar{P}(A)$ или $\bar{P}(B)$ равны 0 или 1, и когда пространства дискретны). Для ограниченных f и φ (составляющих области существования частных моделей), вводя обозначения $\bar{f}(A)=\sup\limits_{x\in A} f(x)$, $\underline{f}(A)=\inf\limits_{x\in A} f(x)$ и такие же для $\bar{\varphi}(B)$, $\underline{\varphi}(B)$, имеем

$$\bar{M}f=[\bar{f}(A)-\underline{f}(A^c)]^+\bar{P}(A)+\bar{f}(A^c),\quad \bar{M}\varphi=[\bar{\varphi}(B)-\underline{\varphi}(B^c)]^+\bar{P}(B)+\bar{\varphi}(B^c)$$

и $\underline{M}f=-\bar{M}(-f)$, то же $\underline{M}\varphi$. В частности, при $\bar{f}(A)\geqslant\bar{f}(A^c)$, $\bar{\varphi}(B)\geqslant\bar{\varphi}(B^c)$:

$$\underline{M}f=\underline{f}(A),\quad \underline{M}\varphi=\underline{\varphi}(B),\quad \bar{M}f\varphi=\max\{\bar{M}f\,\bar{M}\varphi,\ \underline{\varphi}(B)\,\bar{M}f,$$

$$\underline{f}(A)\,\bar{M}\varphi,\ \underline{f}(A)\,\underline{\varphi}(B)\}.$$

При $\bar{f}(A)>\underline{f}(A)$ и $\bar{\varphi}(B)>\underline{\varphi}(B)$ эти первичные средние совместной модели, как можно убедиться, не поглощают друг друга. Число их бесконечно. Здесь мы видим, что предположение независимости $\mathscr{X}$ и $\mathscr{Y}$ дает дополнительную пищу относительно средних произведения $f\varphi$ любых ограниченных признаков $f(x)$, $\varphi(y)$, а не только складываемых из индикаторных признаков событий A, B и их дополнений, т. е. принимающих на них постоянные значения.

Как указывалось в начале раздела, нужно отличать независимость как свойство совместных моделей, от независимого произведения частных моделей как конкретного вида совместной модели. Для последнего справедливы дополнительные свойства.

Свойства независимого произведения:

1. *Равноправие*: $\mathscr{M}^x\times\mathscr{M}^y=\mathscr{M}^y\times\mathscr{M}^x$.

2. *Сохранение порядка*:

$$\mathscr{M}_1^x\subset\mathscr{M}_2^x,\quad \mathscr{M}_1^y\subset\mathscr{M}_2^y\Rightarrow\mathscr{M}_1^x\times\mathscr{M}_1^y\subset\mathscr{M}_2^x\times\mathscr{M}_2^y.$$

3. *Дистрибутивность относительно объединения*:

$$\Big(\bigvee_v\mathscr{M}_v^x\Big)\times\Big(\bigvee_w\mathscr{M}_w^y\Big)=\bigvee_v\bigvee_w\mathscr{M}_v^x\times\mathscr{M}_w^y.$$

В самом деле, первое свойство следует из определения независимости. Второе — из очевидного соотношения:

$$\bar{M}_1 f\leqslant\bar{M}_2 f,\quad \underline{M}_1 f\geqslant\underline{M}_2 f,\quad \bar{M}_1\varphi\leqslant\bar{M}_2\varphi,$$

$$\underline{M}_1\varphi\geqslant\underline{M}_2\varphi\Rightarrow\overline{M_1 f M_1\varphi}\leqslant\overline{M_2 f M_2\varphi}.$$

Наконец, третье свойство расписывается в виде элементарного равенства

$$\max\{\max_v\bar{M}_v f\max_w\bar{M}_w\varphi,\quad \max_v\bar{M}_v f\min_w\underline{M}_w\varphi,$$

$$\min_v\underline{M}_v f\max_w\bar{M}_w\varphi,\quad \min_v\underline{M}_v f\min_w\underline{M}_w\}=\max_{v,w}\overline{M_v f M_w}\varphi.$$

Если в третьем свойстве в качестве v и w взять средние $Mf(x)$, $M\varphi(y)$ и использовать представление частных моделей в виде объединения сечений, то получим

$$\mathscr{M}^x\mathscr{M}^y=\Big(\bigvee_{M_*f}\mathscr{M}^x_{M_*f}\Big)\times\Big(\bigvee_{M_*\varphi}\mathscr{M}^y_{M_*\varphi}\Big)=\bigvee_{M_*f}\bigvee_{M_*\varphi}\mathscr{M}^x_{M_*f}\times\mathscr{M}^y_{M_*\varphi}.$$

А так как $\mathscr{M}^x_{M_*f}\times\mathscr{M}^y_{M_*\varphi}=(\mathscr{M}^x\times\mathscr{M}^y)_{M_*f,\,M_*\varphi}$, то отсюда вывод: *независимость сохраняется при переходе к* $(M_*f(x),\ M_*\varphi(y))$*-сечениям модели, и наоборот, из независимости в сечениях* (по разделенным признакам) *следует независимость* $\mathscr{X}$ *и* $\mathscr{Y}$.

4. *Независимые произведения эквивалентным образом задаются свойством* (2.10) *нековариированности частных признаков.*

Доказательство содержится в дополнении 1. Итак, нековариированность (2.10) не только есть следствие независимости, но и сама, если нековариированными являются любые пары $f(x)$, $\varphi(y)$ признаков, служит характеризацией независимого произведения, полностью его определяя.

Независимые произведения на дискретных пространствах исходов. Для дискретных явлений здесь будет дана интерпретация введенной нами независимости с позиций представлений моделей в виде семейства распределений вероятностей.

Пусть $\mathscr{X}=\{x_1,\ x_2,\ldots\}$ и $\mathscr{Y}=\{y_1,\ y_2,\ldots\}$ дискретны и независимы. Если вероятности элементарных исходов являются точными, то понятие независимости исходов $\mathscr{X}$ и $\mathscr{Y}$ превращается в тождественное равенство совместной вероятности (также точной) произведению частных:

$$P(x_i,\ y_j)\equiv P(x_i)P(y_j),\quad \forall i,\ j,$$

где $P(x_i)=\sum\limits_j P(x_i,\ y_j)$, $P(y_j)=\sum\limits_i P(x_i\ y_j)$. Это хорошо известно, и мы им не раз уже пользовались, в векторном виде записывая: $\mathbf{P}^{xy}=\mathbf{P}^x\times\mathbf{P}^y$. А если вероятности не точные, а интервальные? Тогда соответствующие модели $\mathscr{M}^{xy}$ изображаются выпуклыми семействами векторов совместных вероятностей $\mathbf{P}^{xy}$, или, что более экономично, крайними точками, а еще лучше не всеми, а только являющимися вершинами $\mathbf{P}^{xy}_\theta$ модели: $\mathscr{M}^{xy}=\bigvee\limits_\theta\mathbf{P}^{xy}_\theta$, где θ — индекс множества этих вершин, своего рода общий для $\mathscr{X}$ и $\mathscr{Y}$ фактор.

Принято считать, что независимость $\mathscr{X}$ и $\mathscr{Y}$ будет достигнута, когда на произведения разбиваются если не все векторы $\mathbf{P}^{xy}\subset\mathscr{M}^{xy}$, то хотя бы вершины $\mathbf{P}^{xy}_\theta$, т. е. $\mathscr{M}^{xy}=\bigvee\limits_\theta\mathbf{P}^x_\theta\times\mathbf{P}^y_\theta$. Но это не так, и вот пример, показывающий, к какому заблуждению это может привести!

Пример 2.14. Пусть $\mathscr{X}=\mathscr{Y}$ и вероятности $\mathbf{P}^{xy}_\theta$, где фактор θ принимает значения $\theta_1,\ \theta_2,\ \ldots$, таковы, что $\mathbf{P}_{\theta_i}(x_i,\ y_i)=\delta_{ij}$ (0 при $i\neq j$ и 1 при $i=j$), т. е. при заданном θ_i с вероятностью 1 появляется исход x_i и обязательно тот же самый исход $y_i=x_i$ на $\mathscr{Y}$. Для каждого $\mathbf{P}^{xy}_{\theta_i}$, как нетрудно видеть, $\mathscr{X}$ и $\mathscr{Y}$ независимы: $\mathbf{P}^{xy}_{\theta_i}=\mathbf{P}^x_{\theta_i}\times\mathbf{P}^y_{\theta_i}$, потому что детерминированные исходы формально независимы. Объединение $\mathscr{M}^{xy}=\bigvee\limits_{\theta_i}\mathbf{P}^x_{\theta_i}\times\mathbf{P}^y_{\theta_i}$ соответствует совместному явлению, в котором с неизвестной вероятностью выбирается какой-то исход $x_i\in\mathscr{X}$ и затем детерминированно — совпадающий с ним исход $y_i=x_i$

на $\mathcal{Y}$. Несомненно, раз y всегда совпадает с x, о независимости исходов не может быть и речи.

Что-то похожее на приведенный пример будет иметь место, если немного отклониться от точно нулевой и единичной частных вероятностей, так как и тогда представляется возможность по исходу x более или менее точно предсказать фактор θ, а через него уже исход y. Таким образом, причина дефицита независимости для модели в виде объединения $\bigvee_{\theta} \mathbf{P}^x{}_\theta \times \mathbf{P}^y{}_\theta$ кроется в наличии общего для x и y подчиняющего их свойства фактора θ. Для независимости нужно развязать факторы в соответствии со следующим утверждением.

На дискретных $\mathcal{X}$ и $\mathcal{Y}$ совместная $\mathcal{M}^{xy}$ образует независимое произведение тогда и только тогда, когда она представляется в виде объединения произведений точных векторов вероятностей

$$\mathcal{M}^{xy} = \bigvee_v \bigvee_w \mathbf{P}^x_v \times \mathbf{P}^y_w \tag{2.11}$$

с индексами v и w, несвязанно друг от друга пробегающими свои множества V и W значений.

Условиями утверждения гарантируется $\mathcal{M}^{xy} = \mathcal{M}^x \times \mathcal{M}^y$, где $\mathcal{M}^x = \bigvee \mathbf{P}^x{}_v$, $\mathcal{M}^y = \bigvee \mathbf{P}^y{}_w$, а совместные средние вычисляются из выражения:

$$\overline{M} g(x, y) = \sup_{v \in V} \sup_{w \in W} \sum_i \sum_j g(x_i, y_j) P_v(x_i) P_w(y_j), \quad \forall g \in \mathcal{F}_0^{xy}.$$

Необходимость следует из указанного представления частных моделей в виде семейств и свойства дистрибутивности независимого умножения относительно объединений; а для доказательства достаточности нужно подставить в выражение $\overline{M} g(x, y)$ произведение $g(x, y) = f(x)\varphi(y)$ и убедиться в справедливости тождества (2.11).

Формула (2.14) может быть взята за эквивалентное определение независимого произведения, но только для дискретных пространств $\mathcal{X}$ и $\mathcal{Y}$. Ее невозможно распространить на непрерывные пространства в связи с отсутствием там подобных векторам вероятностей атомов модели.

Представление (2.11) полезно потому, что придает некоторую наглядность доказательствам указанных выше свойств независимых произведений, как это демонстрируется в дополнении 2 к параграфу. А теперь — наглядный пример к данному разделу.

Пример 2.15. Если независимо два раза подбрасывается одна и та же гнутая монета, то имеем два независимых эксперимента с одинаковыми вероятностями герба $P(\Gamma_1) = P(\Gamma_2) = p$ при обоих подбрасываниях. Из-за неизвестности p: $\underline{p} \leqslant p \leqslant \overline{p}$, имеем совместную интервальную модель в виде объединения $\mathcal{M}^{xy}_* = \bigvee_{\underline{p} \leqslant p \leqslant \overline{p}} \mathbf{P}^x \times \mathbf{P}^y$, где вероятности $\mathbf{P} = (p, 1-p)$ одинаковы на $\mathcal{X}$ и $\mathcal{Y}$. Это есть стационарный эксперимент. Здесь вероятности существуют (хотя не известны) и подчинены друг другу, что так или иначе связывает исходы между собой, потому независимости, как мы ее определили, нет.

Другой случай, когда каждый раз совершенно неизвестно, какая монета, гнутая или нет, подбрасывается. Это уже неустойчивый эксперимент с совершенно независимыми исходами в нашем определении, а также и понимании независимого произведения, так как по результату первого подбрасывания уже совершенно ничего нового не скажешь о втором и

$$\mathcal{M}_2^{xy} = \bigvee_{\underline{p} \leqslant p_1 \leqslant \overline{p}} \bigvee_{\underline{p} \leqslant p_2 \leqslant \overline{p}} \mathbf{P}_1^x \times \mathbf{P}_2^y,$$

где $\mathbf{P}_1^x = (p_1, \ 1-p_1)$, $\mathbf{P}_2^y = (p_2, \ 1-p_2)$. $\underline{p} = \min\{\underline{p}, \ 1/2\}$, $\overline{p} = \max\{\overline{p}, \ 1/2\}$.

Геометрическая иллюстрация независимости. Для случая двухточечных экспериментов $\mathcal{X} = \{x_1, x_2\}$, $\mathcal{Y} = \{y_1, y_2\}$ совместная модель описывается семействами векторов вероятностей $\mathbf{P} = (P(x_1, y_1), P(x_1, y_2), P(x_2, y_1))$ в трехмерном пространстве (компонента $P(x_2, y_2)$ опущена, так как дополняет до 1 сумму остальных). Обратимся к рис. 2.7, где поверхность π есть множество векторов $\mathbf{P} = \mathbf{P}^x \times \mathbf{P}^y$ независимого произведения, т. е. с $P(x_i, y_j) = P(x_i) P(y_j)$. Через эти равенства по границам интервальных вероятностей $\underline{P}(x_1)$, $\overline{P}(x_1)$ и $\underline{P}(y_1)$, $\overline{P}(y_1)$ (первая пара задает $\mathcal{M}^x$, вторая — $\mathcal{M}^y$) определяются четыре точки $\mathbf{P}_{ij}$, $i, j = 1, 2$, на π; они и будут четырьмя вершинами тетраэдра независимого произведения $\mathcal{M}^x \times \mathcal{M}^y$.

Тетраэдр эквивалентно представляется: 1) объединением вершин $\mathbf{P}_{ij}$, 2) как фигура с четырьмя располагающимися на π ребрами, соответствующими границам вероятностей, 3) как множество, окаймленное гранями, соответствующими четырем уравнениям:

$$\begin{aligned}
&\overline{M}[\delta_{x_1}(x) - \underline{P}(x_1)][\delta_{y_1}(y) - \overline{P}(y_1)] = 0;\\
&\overline{M}[\delta_{x_1}(x) - \overline{P}(x_1)][\delta_{y_1}(y) - \underline{P}(y_1)] = 0;\\
&\overline{M}[\overline{P}(x_1) - \delta_{x_1}(x)][\delta_{y_1}(y) - \overline{P}(y_1)] = 0;\\
&\overline{M}[\underline{P}(x_1) - \delta_{x_1}(x)][\delta_{y_1}(y) - \underline{P}(y_1)] = 0.
\end{aligned} \tag{2.12}$$

Это есть уравнения (2.10) для $f(x) = \delta_{x_1}(x)$, $\varphi(y) = \delta_{y_1}(y)$. Так как при $\mathcal{X} = \{x_1, x_2\}$ любая функция $f(x)$ линейным образом приводится к $\delta_x(x)$ (например, при $f(x_2) > f(x_1)$, $\delta_{x_1}(x) = f(x) - f(x_1))/(f(x_2) - f(x_1))$, то отсюда следует: *система (2.12) нулевых средних, соответствующая нековариированности элементарных исходов, дает необходимые и достаточные условия независимости двухточечных экспериментов $\mathcal{X} = \{x_1, x_2\}$ и $\mathcal{Y} = \{y_1, y_2\}$.*

Условия нековариированности (2.12) достаточны для независимости, но не для независимого произведения. В самом деле, если на рис. 2.7 введением дополнительного первичного значения, скажем $\bar{p}(x_1, y_1)$ «отрезать» угол тетраэдра, не исказив первичных граней, то хотя равенства (2.12) выполняются, но независимого произведения уже не будет.

Сделаем выводы. Первый: любые «урезания» тетраэдра независимого произведения $\mathcal{M}^{xy}$ на рис. 2.7, при которых остаются

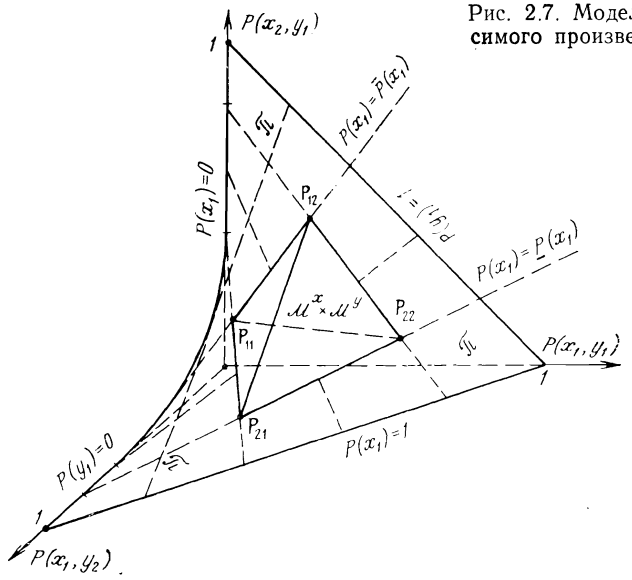


Рис. 2.7. Модель независимого произведения

хотя бы какие-нибудь части каждой из его граней, сохраняют независимость $\mathscr{X}$ и $\mathscr{Y}$, но нарушают независимое произведение. Таким «урезанием» можно учитывать дополнительные свойства экспериментов. Например, их стационарность как результат одинаковости условий проистекания (см. первую часть примера 2.15), которая в самом общем случае отражается дополнительными ограничивающими модель равенствами $M[f(x)-f(y)]\equiv 0$, $\forall f\in\mathscr{F}$ (об этом пойдет речь в следующих двух главах).

Второй: независимость есть некоторая пропорциональность размеров и особенность положения тела совместной модели по отношению к координатным осям, а независимое произведение — вместе с этим полагает и минимальное число граней модели.

Третий: любую совместную модель $\mathscr{M}^{xy}$ можно расширить до такой, при которой $\mathscr{X}$ и $\mathscr{Y}$ становятся независимыми, или же расширить до содержащего ее независимого произведения: $\mathscr{M}^{xy}\subset \subset \mathscr{M}_*{}^x\times\mathscr{M}_*{}^y$. Отметим, что справа $\mathscr{M}_*{}^x$ и $\mathscr{M}_*{}^y$ оказываются, в общем, шире частных $\mathscr{M}^x$ и $\mathscr{M}^y$ для $\mathscr{M}^{xy}$, и расширение не однозначно.

И, наконец, четвертый: наличие общего для $\mathscr{X}$ и $\mathscr{Y}$ подчиняющего параметра θ угрожает потерей независимости и соответствует модели $\bigvee\limits_{\theta}\mathbf{P}^{x_\theta}\times\mathbf{P}^{y_\theta}$, все вершины которой лежат на поверхности π (рис. 2.7).

Нековариированность случайных величин. Следствием независимости было выполнение для всех частных, разделенных по переменным признаков $f(x)$ и $\varphi(y)$ равенств (2.10), определяющих нековариированность f и φ. Изучим это понятие подробнее, считая f и φ произвольными совместными признаками, для чего в

(2.10) подставляются: $f=f_1(x, y)$, $\varphi=f_2(x, y)$ — на $\mathscr{X}\times\mathscr{Y}$ это случайные величины. Случайные величины f_1 и f_2 нековариированны, если их ковариации, определяемые левыми частями равенств (2.10), равны 0. Нековариированность — это свойство совместной модели. Для него верно следующее.

1. Понятие нековариированности равноправно относительно случайных величин.

2. Если f_1 и f_2 нековариированы, то такими же останутся вторичные поэлементные признаки $f'_1=b_1f_1+c_1$ и $f'_2=b_2f_2+c_2$, $\forall b_1$, b_2, c_1, $c_2\in\mathscr{R}$, (обозначаем их множества через $\mathscr{L}f_1$ и $\mathscr{L}f_2$).

3. Признаки f_1 и f_2 будут нековариированными, если и только если выполняется любое из следующих двух условий:

Условие А

$$\left.\begin{aligned}&\overline{M}(f_1-c)(f_2-\overline{M}f_2)=0\\&\overline{M}(c-f_1)(f_2-\underline{M}f_2)=0\end{aligned}\right\}\quad\text{при}\quad\forall c\leqslant\underline{M}f_1;$$

$$\left.\begin{aligned}&\overline{M}(f_1-\overline{M}f_1)(f_2-d)=0\\&\overline{M}(f_1-\underline{M}f_1)(d-f_2)=0\end{aligned}\right\}\quad\text{при}\quad\forall d\leqslant\underline{M}f_2.$$

Условие Б

$\overline{M}f'_1(f'_2-\overline{M}f'_2)=0$ при всех $f'_2\in\mathscr{L}f_2$ и $f'_1\in\mathscr{L}f_1$ таких, что $\underline{M}f'_1\geqslant 0$;

$\overline{M}(f'_1-\overline{M}f'_1)f'_2=0$ при всех $f'_1\in\mathscr{L}f_1$ и $f'_2\in\mathscr{L}f_2$ таких, что $\underline{M}f'_2\geqslant 0$.

4. Для голой совместной ИМ любые ограниченные признаки попарно нековариированы.

5. При точных средних Mf_1 и Mf_2 понятие нековариированности трансформируется в равенство $M(f_1-Mf_1)(f_2-Mf_2)=0$, а если к тому же $Mf_1=0$ или $Mf_2=0$, то совпадает с понятием *некоррелированности*: $Mf_1f_2=0$.

Классы $\mathscr{H}$ и Ψ признаков называются *нековариированными*, если любой признак $h\in\mathscr{H}$ нековариирован с любым $\psi\in\Psi$. Из свойства 2 совершенно ясно, что нековариированными будут и любые $h'\in\mathscr{L}h$ с $\psi'\in\mathscr{L}\psi$.

Независимость, свобода, нековариированность. Пусть $\mathscr{X}$ и $\mathscr{Y}$ — два явления. Если считать частные модели $\mathscr{M}^x=\langle\overline{M}^x\mathscr{H}\rangle$, $\mathscr{M}^y=\langle\overline{M}^y\Psi\rangle$ заданными их согласованными средними на первичных признаках $h(x)\in\mathscr{H}$, $\psi(y)\in\Psi$, то каждое из составляющих заголовок понятий соответствует определенной статистической связи как между первичными $h(x)$ и $\psi(y)$, так и между произвольными всеми остальными признаками вида $f(x)$ и $\varphi(y)$. Эта связь плюс частные модели вместе формируют совместную модель.

Систематизируя изложенные выше результаты, укажем, какие в итоге совместные модели получаются и какими наборами первичных средних они определяются.

Независимость $\mathscr{X}$ и $\mathscr{Y}$: $\mathscr{M}^{xy}_{\text{нз}}=\mathscr{M}^x\times\mathscr{M}^y$,

$$\overline{M}_{\text{нз}} f(x)\varphi(y)=\overline{Mf M\varphi},\quad \forall f,\ \varphi.$$

Свобода $\mathscr{Y}$ от $\mathscr{X}$: $\mathscr{M}^{xy}_{\text{св}}=\mathscr{M}^x\mathscr{M}^y$, $\overline{M}_{\text{св}} f(x,y)=\overline{M}^x(\overline{M}^y f(x,y))$,

$$\overline{M}\,\mathscr{H}\cup\{\overline{M}c^+(x)[\psi(y)-\overline{M}\psi]\},\ \psi\in\Psi\}.$$

Нековариированность $\mathscr{H}$ и Ψ: $\mathscr{M}^{xy}{}_{\text{нк}}$,

$$\overline{M}_{\text{нк}}(h-\underline{M}h)(\psi-\overline{M}\psi)=\overline{M}_{\text{нк}}(h-\overline{M}h)(\psi-\underline{M}\psi)=$$

$$=\overline{M}_{\text{цк}}(\overline{M}h-h)(\psi-\overline{M}\psi)=\overline{M}_{\text{нк}}(\underline{M}h-h)(\psi-\underline{M}\psi)=0,$$

$$\forall h\in\mathscr{H},\quad \psi\in\Psi.$$

Отсюда видно, что при одних и тех же частных моделях для независимого произведения количество первичных признаков существенно богаче, чем для свободного и чем при нековариированности, а так как средние на всех те же самые, то

$$\mathscr{M}_{\text{нз}}\subset\mathscr{M}_{\text{св}},\quad \mathscr{M}_{\text{нз}}\subset\mathscr{M}_{\text{нк}}.$$

Равенство будет лишь при точных векторах вероятностей в дискретных пространствах.

Что касается понятий свободы и нековариированности, то они, в общем, не сравнимы по включению между собой, так как $\mathscr{M}_{\text{св}}$ и $\mathscr{M}_{\text{нк}}$ задаются разными первичными признаками. В определенных классах моделей такая связь все же обнаруживается, как в следующем примере точных (на алгебрах) распределений вероятностей, важном также для осмысливания статуса принятого в современной литературе понятия независимости, которое в нашей терминологии оказывается ничем иным, как нековариированностью алгебр.

Пример 2.16. Алгебры $\mathscr{A}^x{}_0$ на $\mathscr{B}^y{}_0$ и $\mathscr{X}$ на $\mathscr{Y}$ называются нековариированными относительно точной на их произведении совместной модели $\mathscr{M}^{xy}=\mathscr{P}^x\times\mathscr{P}^y$, если $P(AB)=P(A)P(B)$ для любых $A\in\mathscr{A}^x{}_0$ и $B\in\mathscr{B}^y{}_0$. Совместную модель, для которой эти вероятности являются первичными, обозначим $\mathscr{M}_{\text{нк}}$. Для нее нековариированными будут любые интегрируемые (по $\mathscr{P}^x$ и $\mathscr{P}^y$) частные признаки $h(x)$ и $\psi(y)$: $M_{\text{нк}}h\psi=MhM\psi$, к которым относятся любые измеримые ограниченные функции, и эти значения можно считать первичными (что не меняет $\mathscr{M}_{\text{нк}}$).

Для произведений неизмеримых признаков $f(x)\varphi(y)$ продолжением с точных средних значений измеримых признаков $Mh\psi=MhM\psi$ находим:

$$\overline{M}_{\text{цк}} f(x)\varphi(y)=\inf_{\Sigma h_i(x)\psi_i(y)\geqslant f(x)\varphi(y)}\sum Mh_i(x)M\psi_i(y),$$

где инфимум берется по всем измеримым h_i и ψ_i.

Пусть теперь при тех же частных $\mathscr{P}^x$ и $\mathscr{P}^y$, определенных точными на своих алгебрах вероятностями, явления считаются независимыми. Тогда для $\mathscr{M}_{\text{нз}}=\mathscr{P}^x\times\mathscr{P}^y$ имеем для неизмеримых признаков

$$\overline{M}_{\text{нз}} f(x)\varphi(y)=\overline{Mf(x)M\varphi(y)}.$$

Разлагая $f=f^+-f^-$, $\varphi=\varphi^+-\varphi^-$ и считая $\{f\geqslant 0\}\in\mathscr{A}^x{}_0$, $\{\varphi\geqslant 0\}\in\mathscr{B}^y{}_0$, получим

$$\overline{M}_{\text{нк}} f\varphi=\overline{M}f^+\overline{M}\varphi^+-\underline{M}f^-\underline{M}\varphi^+-\underline{M}f^+\underline{M}\varphi^-+\overline{M}f^-\overline{M}\varphi^-,$$

$$\overline{M}_{\text{нз}} f\varphi=\max\{(\overline{M}f^+-\underline{M}f^-)(\overline{M}\varphi^+-\underline{M}\varphi^-),\ (\overline{M}f^+-\underline{M}f^-)\times$$

$$\times(\underline{M}\varphi^+-\overline{M}\varphi^-),\ (\underline{M}f^+-\overline{M}f^-)(\overline{M}\varphi^+-\underline{M}\varphi^-),\ (\underline{M}f^+-\overline{M}f^-)(\underline{M}\varphi^+-\overline{M}\varphi^-)\}.$$

Нетрудно видеть, что $\overline{M}_{\text{нз}}f\varphi\leqslant\overline{M}_{\text{нк}}f\varphi$. Это неравенство справедливо для произведений любых частных ограниченных признаков $f(x)$, $\varphi(y)$. Равенство будет в том случае, если оба признака неотрицательны (в том числе, для событий):

$$\overline{M}_{\text{нз}} f^+\varphi^+=\overline{M}_{\text{нк}} f^+\varphi^+=\overline{M}f^+\overline{M}\varphi^+,$$

где равенства справедливы отдельно для верхних и для нижних средних. Таким образом, и при точных вероятностях независимость в нашем определении является более строгим понятием, чем нековариированность алгебр.

Рассмотрим при тех же частных $\mathscr{P}^x$ и $\mathscr{P}^y$ понятие свободы $\mathscr{Y}$ от $\mathscr{X}$. Обозначим $\mathscr{M}_{\text{св}}=\mathscr{P}^x\mathscr{P}^y$. Для произведений измеримых (суммируемых) частных признаков имеем те же точные средние, что при нековариированности и независимости: $M_{\text{св}}h(x)\psi(y)=Mh(x)M\psi(y)$. При неизмеримых функциях и том же условии $\{f\geqslant 0\}\in\mathscr{A}^x{}_0$, $\{\varphi\geqslant 0\}\in\mathscr{B}^y{}_0$ получим

$$\overline{M}_{\text{св}} f\varphi=\overline{M}f^+(\overline{M}\varphi^+-\underline{M}\varphi^-)-\underline{M}f^-(\underline{M}\varphi^+-\overline{M}\varphi^-)=$$

$$=\overline{M}f^+\overline{M}\varphi^+-\overline{M}f^+\underline{M}\varphi^--\underline{M}f^-\underline{M}\varphi^++\underline{M}f^-\overline{M}\varphi^+.$$

Сравнение с предыдущими формулами ведет к неравенствам

$$\overline{M}_{\text{нз}} f\varphi\leqslant\overline{M}_{\text{св}} f\varphi\leqslant\overline{M}_{\text{нк}} f\varphi.$$

Так что в данном примере $\mathscr{M}_{\text{нз}}\subset\mathscr{M}_{\text{св}}\subset\mathscr{M}_{\text{нк}}$. Таким образом, при нековариированных алгебрах и точных вероятностях на них категория свободы занимает промежуточное положение между независимостью и нековариированностью.

Дополнения. 1. Теорема. *Для того чтобы модель $\mathscr{M}^{xy}$ представляла собой независимое произведение, необходимо и достаточно, чтобы первичными для нее были совместные произведения всевозможных частных нековариированных между собой признаков $f(x)$ и $\varphi(y)$.*

Необходимость эквивалентна свойству (2.10).

Достаточность. Обозначим $\mathscr{M}'$ ИМ с первичным набором (2.10) и заданными частными $\mathscr{M}^x$ и $\mathscr{M}^y$. Покажем, что $\mathscr{M}'=\mathscr{M}^x\times\mathscr{M}^y$. Согласно формуле (1.3)

$$\overline{M}' f\varphi=\inf_{c_i^+}\sup_{x,\,y}[f(x)\varphi(y)-\sum c^+\overset{\circ}{g}_i(x,\ y)],$$

где $\overset{\circ}{g}_i(x,\ y)$ есть всевозможные первичные признаки вида

$$(f-\underline{M}f)(\varphi-\overline{M}\varphi),\ (f-\overline{M}f)(\varphi-\underline{M}\psi),\ (\overline{M}f-f)(\varphi-\overline{M}\varphi),$$

$$(\underline{M}f-f)(\varphi-\underline{M}\varphi).$$

Так как этот набор есть часть набора (2.8), то $\mathscr{M}'\supset\mathscr{M}^x\times\mathscr{M}^y$ и потому $\overline{M}'f\varphi\geqslant\leqslant\overline{Mf}M\varphi$. Осталось доказать противоположное неравенство. Возьмем два частных признака $f(x)$ и $\varphi(y)$ с конечными $\underline{f}=\inf f$, $\overline{f}=\sup f$ и то же $\underline{\varphi}$, $\overline{\varphi}$. Нор-

мируем их, положив $f_0=(f-\underline{f})/(\overline{f}-\underline{f})$, $\varphi_0=(\varphi-\underline{\varphi})/(\overline{\varphi}-\underline{\varphi})$. Обозначаем $\mathscr{M}''$ — ИМ размерности четыре, определяемую нековариированными f_0 и φ_0, т. е. первичными средними (2.10) с подставленными $f=f_0$ и $\varphi=\varphi_0$. Так как $\mathscr{M}''\supset\mathscr{M}'$, то

$$\overline{M}'f\varphi\leqslant\overline{M}''f\varphi=\inf_{c_i^+}\max_{x,\,y}[f\varphi-c_1^+(f_0-\underline{M}f_0)(\varphi_0-\overline{M}\varphi_0)-$$

$$-c_2^+(f_0-\overline{M}f_0)(\varphi_0-\underline{M}\varphi_0)+c_3^+(f_0-\overline{M}f_0)(\varphi_0-\overline{M}\varphi_0)+$$

$$+c_4^+(f_0-\underline{M}f_0)(\varphi_0-\underline{M}\varphi_0)].$$

Нетрудно убедиться в том, что в правой части этого неравенства максимум по x и y квадратных скобок будет достигаться при тех x^* и y^*, при которых f и φ (f_0 и φ_0) принимают экстремальные значения. Пусть f_0, φ_0 максимальны (и равны 1) при $x=x^*_1$, $y=y^*_1$ и минимальны (равны 0) при $x=x^*_2$, $y=y^*_2$. Заменяя $\mathscr{X}$ и $\mathscr{Y}$ на двухточечные пространства $\mathscr{X}^*=\{x^*_1, x^*_2\}$, $\mathscr{Y}^*=\{y^*_1,y^*_2\}$ при тех же границах $\overline{M}^{x^*}f_0=\overline{M}f_0$, $\underline{M}^{x^*}f_0=\underline{M}f_0$, $\overline{M}^{y^*}\varphi_0=\overline{M}\varphi_0$, $\underline{M}^{y^*}\varphi_0=\underline{M}\varphi_0$, мы не изменим $\overline{M}''f\varphi=\overline{M}^{x^*y^*}f\varphi$. А для двухточечных пространств согласно следующему за (2.12) выделенному курсивом утверждению имеем $\overline{M}^{x^*y^*}f\varphi=\overline{M^{x^*}fM^{y^*}\varphi}=\overline{Mf}\overline{M}\varphi$, поэтому $\overline{M}'f\varphi\leqslant\overline{Mf}\overline{M}\varphi$, что и требовалось доказать.

2. Пример использования представления (2.11) в случае дискретных пространств $\mathscr{X}$ и $\mathscr{Y}$ для доказательства свойств аддитивности и нековариированности частных признаков $f(x)$, $\varphi(y)$:

$$\overline{M}(f+\varphi)=\sup_v\sup_w(M_vf+M_w\varphi)=\sup_vM_vf+\sup_wM_w\varphi=\overline{M}f+\overline{M}\varphi;$$

$$\overline{M}(f-\underline{M}f)(\varphi-\overline{M}\varphi)=\sup_v\sup_w(M_vf-\underline{M}f)(M_w\varphi-\overline{M}\varphi)=$$

$$=(\overline{M}f-\underline{M}f)(\overline{M}\varphi-\overline{M}\varphi)=0.$$

3. Пример соотношения понятий независимости и свободы. Вернемся к примеру 2.12. Сначала подбрасывается симметричная монета (эксперимент $\mathscr{X}$), затем «некто», зная результат подбрасывания, показывает одну из сторон своей монеты, делая это, как он хочет (эксперимент $\mathscr{Y}$). Игра идет на деньги, причем «некто» ничего с этого не получает. Если «некто» нейтрален, т. е. он может и не учитывать результат первого бросания, то эксперименты $\mathscr{X}$ и $\mathscr{Y}$ будут независимыми. Если же о намерениях этого «некто» ничего не известно, то эксперимент $\mathscr{Y}$ свободен от $\mathscr{X}$. Например, «некто» может захотеть обеспечить выигрыш одному из игроков или же выравнять выигрыши. Обозначим A — выпадение первого герба и B — показ второго. По условию задачи $\underline{P}(A)=\overline{P}(A)=1/2$, а так как «некто» может избрать любую стратегию, например, показывая только гербы или только решки, то $\underline{P}(B)=0$, $\overline{P}(B)=1$. Для свободного произведения первичными совместной ИМ будут $\underline{P}(A)=\overline{P}(A)=1/2$ (средние $\overline{M}(c^+_1A+c^+A^c)(B-\overline{P}(B)=0$ избыточны, так как $B-\overline{P}(B)=B-1\leqslant0$), а для независимого произведения —

$$\overline{M}(1/2-A)B=\overline{M}(1/2-A)B^c=\underline{M}(1/2-A)B=\overline{M}(1/2-A)B^c=0.$$

Равенства ведут к точным значениям средних $M(1/2-A)B=M(1/2-A)B^c=0$ или равенствам $P(AB)=P(AB^c)$, $P(AB^c)=P(A^cB^c)$, которые вместе с первичными $P(AB)+P(AB^c)=1/2=P(A^cB)+P(A^cB^c)$ определяют все включенные в независимое произведение векторы вероятностей. Отметим, что данный эксперимент отличается от двухкратного независимого подбрасывания монеты, при котором все четыре компоненты вектора вероятностей равны 1/4.

2.6. ЗАКЛЮЧЕНИЕ

В § 2.1 изучается воздействие преобразований пространств исходов одного в другое на вид интервальной модели. Всем известны формулы преобразований плотностей и насколько сложными они становятся для нелинейных и инерционных преобразований (достаточно вспомнить расчеты системы фильтр-ограничитель-фильтр). Для интервальных моделей принципы расчета гораздо проще и универсальнее. Среди огромного разнообразия признаков всегда найдутся согласованные с преобразованием, называемые представимыми признаками, которые, во-первых, совершенно элементарно преобразуются сами, и во-вторых, для них производится непосредственный перенос средних со входа на выход, определяя, таким образом, первичные данные выходной модели. Нужно только рассчитать средние представимых признаков на входе.

Процедура расчета упростится, если первичные признаки на входе все представимы и согласованы, тогда действиями будет прямой перенос средних со входа на выход, а модели оказываются подобными. Подобие означает схожесть структур и отсутствие невозвратимого ущерба при преобразовании. В частности, подобны между собой все плотности распределений вероятностей, так как переходят одна в другую при преобразованиях случайных величин.

Случайные преобразования § 2.2 как математическая запись расплывчатых, неопределенных действий даются интервальными моделями выходного явления при каждом значении входа и называются переходными моделями. Случайность преобразований добавляет неопределенности на выходе, приводя к расслоению признаков и расширению средних. Их частный случай ведет к интервальным действиям арифметики, заложенным в интервальном анализе.

Случайное преобразование наглядно сравнивается с взволнованным аквариумом полупрозрачной неоднородной жидкости, сквозь которую мы смотрим из комнаты на улицу. Предметы становятся искаженными, нечеткими, трудноразличимыми. Уже видим не дерево и машину, а их смутные случайные очертания, по которым рисуются усредненные изображения. Это и будут нечеткие наблюдения § 2.3 как результат расплывчатых случайных[1] преобразований, где вход составляет недоступная нам некоторая предметная область (улица), а выход — наблюдения, связываемые в суждения о том, что происходит. Кстати, в человеческом языке слова и фразы также имеют предметный смысл и относятся к предметной области по принципу: мы так привыкли понимать, т. е. среднестатистически. В этом ракурсе человек являет собой разновидность случайного преобразования входа (о чем мы говорим) в выход (что мы говорим).

На интервальные средние можно смотреть как на преобразования чисел в интервалы, т. е. как своеобразный итог «видения» точных средних (если таковые существуют) через призму ограниченного эксперимента, вынуждающего прибегать к осторожным оценкам в виде интервалов. Если посмотреть

[1] Именно случайных с заложенными в них статистическим закономерностями в отличие от теории нечетких множеств Заде [15].

чуть пристальнее, то это уже будут не интервалы, а их расплывчатые аналоги, дающие размытые изображения средних. Определенные для всех признаков, они составляют размытую модель (как пример размытых преобразований в конце § 2.3 строится размытая арифметика). Этим сделан четвертый шаг на пути от детерминизма к случайности в их следующей цепи: 1) детерминированные явления, 2) случайные явления, заданные распределениями вероятностей, 3) заданные интервальными моделями, 4) заданные размытыми моделями. Дальше дороги, вроде бы, не видно.

Не только преобразования способны отразить совместное поведение двух случайных явлений. Более широко и полно это делают совместные модели § 2.4, которые задаются первичными средними совместных признаков обоих явлений и продолжаются на все остальные. Средние частных признаков вместе образуют частные модели со своей структурой первичных данных (теорема 2.1).

Не всякие модели, увы, могут рассматриваться как результаты каких-нибудь преобразований, а лишь подкласс моделей, названных разложимыми. Разложимые модели записываются как произведения частных моделей на переходные. Особый случай, когда переходные модели от входа не зависят, ведет к понятию свободы выхода от входа, как будто кто-то своевольно распоряжается выходом в рамках модели, зная вход, учитывать который или нет — его дело.

Независимость есть неведение ничего дополнительного об одном явлении, если стал известен результат другого, и наоборот. Свойство симметричное. Независимость как объективная и субъективная реальность охватывает явления целиком со всеми их признаками и через средние признаков определяется в § 2.5. Широта этого понятия оказывается зависящей от данных о явлениях. Если даны точные значения вероятностей (средних), то независимость сводится к свойству мультипликативности, а для интервальных значений — к интервальному аналогу этого свойства. А если совсем не известны, то... независимость имеет место всегда ! Разве это не следует из смысла понятия? И интересен отсюда следующий вывод, что независимость может достигаться расширением совместной модели (по сути, забыванием связей). Чтобы вывод не показался чересчур странным, напомним, что модель — это всего лишь зеркало явления, отражатель его сторон-свойств в своих образах и на своем языке, поэтому и независимость проявляется в определенной заорганизованности совместной модели, особенностях ее структуры, которых можно достичь расширением модели.

Сказанное относится к понятиям некоррелированности и нековариированности в их интервальных определениях. Это более слабые свойства по сравнению с независимостью, т. е. соответствуют более широкой совместной модели. Связь между ними разбирается в последнем разделе главы. Нужно сказать, что классическое определение независимости как мультипликативности точных вероятностей на алгебрах событий есть в наших терминах всего лишь нековариированность алгебр, а это несколько слабее истинной независимости.

Глава 3.

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ, ПОСЛЕДОВАТЕЛЬНОСТИ, СУММЫ

3.1. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ, ПОСЛЕДОВАТЕЛЬНОСТИ

Определения. Случайной величиной (с. в.) называется случайное явление, пространством элементарных исходов которого является числовая прямая $\mathcal{R}$ или ее часть.

Случайные величины обозначаются заглавными буквами X, Y, а соответствующие им совместные и частные модели — обычным образом: $\mathcal{M}^{XY}$, $\mathcal{M}^X$, $\mathcal{M}^Y$. Оператор $\bar{M}$ без индексов означает взятие верхнего среднего от следующих за ним с. в. (или их преобразований) по совместной их модели.

Арифметические действия между с. в., например $X+Y$, X/Y и т. д., означают соответствующие преобразования (сложение, деление и т. д.) на прямом произведении пространств их значений, т. е. $\mathcal{R}\times\mathcal{R}\to\mathcal{R}$, а отношения $X<Y$ или $X\in[a, b]$ есть соответствующие события на этом произведении. Вероятности событий есть средние от индикаторных функций, например $P(X<Y)$, $\bar{P}(a\leqslant X\leqslant b)$, или просто $\bar{P}[a, b]$, $\bar{P}(a, b)$ в зависимости от характера замкнутости отрезка.

В принципе, определение с. в. не исключает бесконечных ее значений $X=\infty$ (или $-\infty$).

Случайная величина называется *ограниченной*, если можно указать H такое, что $\bar{P}(|X|>H)=0$. Случайная величина называется *дискретной*, если на прямой можно выделить конечное или счетное множество $\mathcal{X}_0=\{a_1, a_2, \ldots\}$ чисел, образующих достоверное событие: $P(X\in\mathcal{X}_0)=\lim\limits_{k\to\infty} P(\bigcup\limits_1^k a_i)=1$. Оно и будет множеством значений с. в. X.

Случайная величина называется *непрерывной в точке* a, если вероятность любого отрезка, содержащего точку a, стремится к 0 при устремлении длины этого отрезка к 0, т. е. если независимо от порядка устремления ε_1 и ε_2 к 0 имеет место равенство

$$\lim_{\varepsilon_1,\,\varepsilon_2\downarrow 0} \bar{P}[a-\varepsilon_1,\ \ a+\varepsilon_2]=0.$$

Случайная величина непрерывна в бесконечно удаленных точках $\pm\infty$, если $\bar{P}(X=\pm\infty)=0$. Случайная величина непрерывная в каждой точке, включая бесконечные $\pm\infty$, называется *непрерывной*.

Задание с. в. X производится обычным образом первичными средними $\bar{M}g(X)$, $g\in\mathcal{G}$, которые продолжаются далее на любые функции, мажорируемые линейными комбинациями первичных и составляющими область существования $\mathcal{F}^X=\{f:\bar{M}|f|<\infty\}$. Минимальное число определяющих X элементов $\mathcal{G}$ создаст размерность модели.

Границы $\underline{M}X$, $\overline{M}X$, если они существуют (т. е. $\pm x \in \mathcal{F}^X$), называются верхним и нижним средними значениями самой с. в. X; $\underline{M}X^2$, $\overline{M}X^2$ — среднеквадратическими значениями, или нижней и верхней мощностью с. в. X.

Разберем различные отношения между с. в.

Включение $\mathcal{M}^X \supset \mathcal{M}^Y$ эквивалентно $\overline{M}f(X) \geqslant \overline{M}f(Y)$, $\forall f \in \mathcal{F}^X$, и $\mathcal{F}^X \subset \mathcal{F}^Y$; включение означает, что первично среднестатистических данных в X заложено меньше, чем в Y, или просто они менее точные. Для краткости иногда записываем $X \supset Y$ и говорим, что с. в. X шире (в среднестатистическом смысле), чем Y, или же X включает Y. Самой широкой среди всех является голая с. в. (она же голый параметр), о которой нет никаких данных. Это выход «черного ящика» полностью неизвестной структуры. Иногда выгодно так считать в целях упрощения, даже если кое-какие данные об X имеются.

Включение нужно отличать от неравенства $X \geqslant Y$, означающего, что X всегда будет принимать значение, не меньшее Y: $\underline{P}(X \geqslant Y) = 1$. Неравенство, в частности, будет иметь место для двух признаков $X = f_1(\xi)$, $Y = f_2(\xi)$ одной и той же ξ, если один из них мажорирует другой: $f_1 \geqslant f_2 \Rightarrow X \geqslant Y$. Из $X \geqslant Y$ следует $\overline{M}f(Y) \geqslant \overline{M}f(X)$, но только для монотонно неубывающих функций f (в отличие от включения).

Для полноты картины укажем еще на отношение X больше Y в вероятностном смысле как тождественное неравенство: $\underline{P}(X > x) \geqslant \overline{P}(Y > x)$, $\forall x \in \mathcal{R}$, означающее, что вероятности превышений любых уровней x для с. в. X больше, чем для Y. Это самое слабое отношение упорядоченности с. в. среди введенных нами, называемое в литературе так: X стохастически больше Y.

Еще понадобится далее понятие симметрии. Случайная величина X называется *симметричной*, если среднее любой нечетной функции есть точный ноль: $Mf(X) = 0$ при $f(-x) = -f(x)$, $f \in \mathcal{F}^X$. В частности, если X симметрична, то $M \sin uX = 0$, $\forall u \in \mathcal{R}$, $MX^{2k+1} = 0$ для тех k, для которых $x^{2k+1} \in \mathcal{F}^X$.

Детерминированные преобразования. Резюмируем применительно к с. в. результаты § 2.1, где изучались детерминированные преобразования явлений.

Пусть X задана своими средними $\overline{M}f(X)$, $f \in \mathcal{F}^X$. Преобразование $Y = s(X)$ одной с. в. в другую (а при $s \in \mathcal{F}^X$ с. в. Y будет одним из признаков с. в. X) ведет к модели $\mathcal{M}^Y$, средние которой, как это следует из записи: $\overline{M}\varphi(Y) = \overline{M}\varphi(s(X))$, составляют часть средних модели $\mathcal{M}^X$. А именно, из всего многообразия $\mathcal{F}^X$ выбираются только средние s-представимых признаков: $f(x) = \varphi(s(x))$. Их согласованность между собой очевидна и $\mathcal{F}^Y = \{\varphi : \varphi s \in \mathcal{F}^X\}$ — область существования $\mathcal{M}^Y$.

Если X задается набором $\mathcal{G}$ первичных признаков, то после ее преобразования в Y этот набор, в общем, распадается на все $\varphi(y) \in \mathcal{F}^Y$ — все они потенциально будут первичными для $\mathcal{M}^Y$, что равносильно росту размерности $\mathcal{M}^Y$ по сравнению с $\mathcal{M}^X$.

Кроме одного случая, когда все признаки набора $\mathcal{G}$ s-представимы, т. е. $g(x) = \psi_g(s(x))$, $\forall g \in \mathcal{G}$, и тогда признаки $\psi_g(y)$, $g \in \mathcal{G}$ будут первичными для $\mathcal{M}^Y$, размерности $\mathcal{M}^X$ и $\mathcal{M}^Y$ будут одинаковы, а X и Y будут подобными между собой ($s(x)$ — преобразование подобия). В этом случае по средним s-представимых признаков $\mathcal{M}^X$ восстанавливаются остальные.

Преобразование s будет ограниченным, если область $s\mathcal{X}$ значений $s(x)$ — ограниченное множество на $\mathcal{R}$. К ним относятся: гармонические преобразования $s_u(x) = \cos(ux)$ и $\sin(ux)$ (где u — индекс признака) с областью значений $s\mathcal{X} = [-1, 1]$, индикаторные $s(x) = A(x)$ со значениями 0 и 1 и т. д. Многие нужные преобразования не являются ограниченными. Наиболее распространенными из них считаются: тождественное преобразование x, для которого $s\mathcal{X} = \mathcal{R}$; квадратическое x^2, $s\mathcal{X} = \mathcal{R}^+$; логарифмическое $\ln x$, $x \in \mathcal{R}^+$, $s\mathcal{X} = \mathcal{R}$; показательное $\exp x$, $s\mathcal{X} = \mathcal{R}^+$; гиперболическое $1/x$, $s\mathcal{X} = \mathcal{R}$. В последнем случае в точке $x = 0$ преобразование не определено, и если эта точка не является первичным событием для X, то ее полезно исключить из исходов этой с. в., что не приведет к видоизменению модели X, но зато позволит математически строго пользоваться этим преобразованием.

Нормальная случайная величина. Нормальная с. в. занимает особое положение, обязанное предельным теоремам. Мы рассмотрим ее с наших общих концепций, дав различную интерпретацию, а связанные с ней предельные теоремы будут рассмотрены через параграф уже после введения понятий сходимости.

Нормальной со средним m и дисперсией σ^2 называется с. в. Y, определенная продолжением (посредством интегрирования и (1.4)) нормальной плотности вероятностей (по отношению к мере-длине):

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-(x-m)^2/(2\sigma^2)\right]. \tag{3.1}$$

Первичным для нормальной с. в. является следующий набор вероятностей отрезков:

$$P(a,\ b) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right), \quad a < b, \tag{3.2}$$

где $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z \exp(-x^2/2)\,dx$ — функция Лапласа (она табулируется в учебниках и задачниках по теории вероятностей). Вероятности отрезков получаются интегрированием плотности в пределах от a до b, тогда как плотность получается из вероятностей как пределы их отношений к длинам отрезков при устремлении последних к 0.

Соответствующая нормальной с. в. модель обозначается $\mathcal{N}_{m,\sigma}$. Если свести ее к нулевому среднему и единичной дисперсии $\mathring{Y} = (Y-m)/\sigma$, то получим *стандартную нормальную с. в.*, соответствующую $\mathcal{N}_{0,1}$. Достаточно ее и рассматривать, так как любая другая к ней приводится согласно формуле $Y = m + \sigma\mathring{Y}$.

Область $\mathcal{F}_{\mathcal{N}}$ существования точных средних для $\mathcal{N}_{0,1}$ составляют все интегрируемые с весом $\exp(-y^2/2)$ функции, для них

$$M f(\mathring{Y}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y) \exp(-y^2/2)\, dy,$$

где интеграл понимается в смысле Римана. В частности, для гармонических средних

$$M \cos u\mathring{Y} = \exp(-u^2/2), \quad M \sin u\mathring{Y} \equiv 0, \quad \forall u, \tag{3.3}$$

а для моментов

$$M(\mathring{Y})^{2k+1} = 0, \quad M(\mathring{Y})^{2k} = (2k)!/(k!2^k), \quad k = 1, 2, \ldots \tag{3.4}$$

Теорема 3.1 характеризации нормальной с.в. *Эквивалентными являются следующие способы задания* $\mathcal{N}_{0,1}$:
1°. *плотностью* (3.1) *при* $m=0$, $\sigma=1$;
2°. *первичными вероятностями* (3.2) *интервалов*;
3°. *гармоническими средними* (3.3), *принятыми за первичные и дополненными первичными вероятностями* $\tilde{P}(|\mathring{Y}|>H)$, *заданными достаточно произвольно, но в рамках следующего требования непротиворечивости*: $1-2\Phi(H) \leqslant \tilde{P}(|\mathring{Y}|>H) \underset{H\to\infty}{\to} 0$;
4°. *моментами* (3.4), *взятыми в качестве первичного набора.*

Действительно, первичным для $\mathcal{N}_{0,1}$ можно было бы считать любой плотный в множестве $\mathcal{F}_{\mathcal{N}}$ интегрируемых с весом $\exp(-y^2/2)$ функций класс, или его базис. Такими являются индикаторы интервалов в 2°, степенные функции в 4° и гармонические в 3°. Необходимость введения дополнительных первичных вероятностей в 3° вызвана тем, что гармонические функции образуют базис лишь при ограниченной области значений аргумента.

Случайная величина, описываемая в виде следующего семейства нормальных моделей:

$$\bigvee_{\underline{m}\leqslant m\leqslant \overline{m}} \bigvee_{\underline{\sigma}\leqslant\sigma\leqslant\overline{\sigma}} \mathcal{N}_{m,\sigma} = \mathcal{N}_{\underline{m},\overline{m},\underline{\sigma},\overline{\sigma}},$$

называется *нормальной с интервальными средним* $\underline{m}$, $\overline{m}$ *и дисперсией* $\underline{\sigma}^2$, $\overline{\sigma}^2$. Объединение эквивалентно записи $Y=\sigma\mathring{Y}+m$, где $\mathring{Y}$ — стандартная нормальная с.в., свободная от m и σ, а m и σ принимают произвольные значения в отведенных им интервалах, что ведет к средним

$$\overline{M}f = \max_{\substack{\underline{m}\leqslant m\leqslant\overline{m}\\ \underline{\sigma}\leqslant\sigma\leqslant\overline{\sigma}}} \overline{M} f(\sigma\mathring{Y}+m), \quad \forall f \in \mathcal{F}_{\mathcal{N}}.$$

При бесконечных значениях $\underline{m}=-\infty$, $\overline{m}=\infty$ обозначаем $\mathcal{N}^{\circ}_{\underline{\sigma},\overline{\sigma}}$ и называем нормальной при неизвестном m и интервальном σ. Для

нее, как это находится из последней формулы, имеем

$$\overline{P}(a, b) = 2\Phi\left(\frac{b-a}{2\underline{\sigma}}\right), \quad \underline{P}(a, b) = 0.$$

Теми же будут вероятности, если $\overline{\sigma}$ не известна, тогда обозначаем $\mathcal{N}_{\underline{\sigma}}$, а если к тому же и $\underline{\sigma}$ не известна, т. е. $\underline{\sigma}=0$, то нормальная с.в. вырождается в голую.

Случайные последовательности. Способы описания с. в. непосредственно переносятся на случайные векторы $\mathbf{X}=(X_1, \ldots, X_n)$, составленные из последовательности с.в. Не столь по форме, сколь по содержанию описания при этом усложняются. Нас будут интересовать здесь разные упрощения: такие способы, которые позволили бы по частным моделям $\mathcal{M}^{X_i}$ с.в. X_i составить совместную $\mathcal{M}_{\mathbf{X}}$. Для этого-то и нужны понятия независимости, свободы, нековариированности предыдущей главы.

Последовательность $X_1, X_2, \ldots, X_n$ называется *последовательностью независимых с.в.*, если она полностью определена первичными средними вида

$$\overline{M} \prod_1^n f_i(X_i) = \overline{\prod_1^n M f_i(X_i)} = \max_{M^{(i)}=\overline{M} \text{ или } \underline{M}} \prod_1^n M^{(i)} f_i(X_i), \tag{3.5}$$

где справа стоит максимум по всевозможным сочетаниям произведений $\overline{M}f_i(X_i)$ и $\underline{M}f_i(X_i)$, причем все сомножители должны быть конечны (т. е. $\pm f_i \in \mathcal{F}^{X_i}$), чтобы правая часть была конечной. Это есть закон интервальной мультипликативности средних, положенный в основу независимости и независимого произведения с.в.

Для последовательности независимых с.в. справедливы следующие соотношения:

1. $\overline{M} \sum_1^n f_i(X_i) = \sum_1^n \overline{M} f_i(X_i)$.

2. $\overline{M} \prod_1^n f_i^+(X_i) = \prod_1^n \overline{M} f_i^+(X_i)$, $\underline{M} \prod_1^n f_i^+(X_i) = \prod_1^n \underline{M} f_i^+(X_i)$.

3. $\overline{M} \prod_1^k (X_i - \overline{M}X_i) \prod_{k+1}^n (X_i - \underline{M}X_i) =$

$$= \begin{cases} 0, & \text{если } k \text{ нечетно,} \\ \prod_1^n (\overline{M}X_i - \underline{M}X_i), & \text{если } k \text{ четно.} \end{cases}$$

$$\underline{M} \prod_1^k (X_i - \overline{M}X_i) \prod_{k+1}^n (X_i - \underline{M}X_i) =$$

$$= \begin{cases} 0, & \text{если } k \text{ четно,} \\ -\prod_1^n (\overline{M}X_i - \underline{M}X_i), & \text{если } k \text{ нечетно.} \end{cases}$$

4. Независимость сохраняется при преобразованиях $Y_1=f_1(X_1)$, $Y_2=f_2(X_2)$, ..., $Y_n=f_n(X_n)$.

5. Независимыми будут функции от любых непересекающихся поднаборов последовательности $Y_1=f_1(X_1, \ldots, X_k)$, $Y_2=f_2(X_{k+1}, \ldots, X_n)$.

6 При $i\neq j$

$$\overline{M}(X_i-\underline{M}X_i)(X_j-\overline{M}X_j)=\overline{M}(X_i-\overline{M}X_i)(X_j-\underline{M}X_j)=$$

$$=\underline{M}(X_i-\overline{M}X_i)(X_j-\overline{M}X_j)=\underline{M}(X_i-\underline{M}X_i)(X_j-\underline{M}X_j)=0. \quad (3.6)$$

Последовательность, удовлетворяющая свойству 6, называется *нековариированной.*

Последовательность называется *некоррелированной*, если $MX_iX_j=0$, $i\neq j$. При нулевых средних $MX_i=0$ понятия нековариированности и некоррелированности совпадают. Отметим, что если $X_i=m+\xi_i$ и ξ_i некоррелированы и имеют нулевые средние $M\xi_i=0$, то при неизвестном параметре m с.в. X_i не будут нековариированными, а будут подчиненно (при каждом заданном m) нековариированными.

Последовательность называется *свободной*, если она определена произведением частных $\mathcal{M}^{X_1}\mathcal{M}^{X_2}\ldots\mathcal{M}^{X_n}$, что соответствует следующему порядку вычислений средних: $\overline{M}f(X_1, \ldots, X_n)=\overline{M}^{X_1}(\overline{M}^{X_2}(\ldots(\overline{M}^{X_n}f(X_1, \ldots, X_n))\ldots))$.

Для свободных последовательностей каждое последующее значение X_i представляется как случайное преобразование предыдущих X_1, X_2, ..., X_{i-1}, причем структура этого преобразования не известна, а известна лишь частная $\mathcal{M}^{X_i}$. Для свободных последовательностей остаются справедливыми свойства 1 и 2.

Свойство свободы теряется при перестановках последовательности, в отличие от свойств независимости и нековариированности. Подчеркнем еще раз, что все указанные нами способы задания последовательности объединены тем, что требуют лишь знания частных ИМ элементов X_i с указанием характера взаимодействия X_i.

Еще одним способом будет, когда это взаимодействие не указано, а заданы лишь частные $\mathcal{M}^{X_i}$ (см. пример 2.9). Тогда средние разделенных по переменным признаков $\overline{M}g(X_i)$, $g\in\mathcal{G}_i$, первичные для отдельных с.в., образуют первичный набор совместной модели (отсюда следует соответствующий формуле продолжения способ вычисления совместных средних). На суммах разделенных признаков выполняется свойство 1 аддитивности.

При одинаковых частных моделях независимость X_i приводит к наиболее узкой среди остальных совместной модели, а последний случай полностью неизвестных связей между X_i — к самой широкой.

Однородность и стационарность последовательности. Последовательность называется однородной, если ее совместная ИМ не меняется при циклических сдвигах элементов, т. е. для однородной последовательности вектора $\mathbf{X}=(X_1, \ldots, X_n)$ и $S_k\mathbf{X}=(X_k, \ldots, X_n,$

$X_1, X_2, \ldots, X_{k-1})$, отличающиеся циклической перестановкой элементов, имеют одинаковые совместные ИМ: $\mathcal{M}^{\mathbf{X}}=\mathcal{M}^{S_k\mathbf{X}}$. Тем более одинаковыми должны быть частные модели $\mathcal{M}^{X_1}=\ldots=\mathcal{M}^{X_n}$. Однородность эквивалентна равенству средних при сдвигах:

$$\overline{M}f(\mathbf{X})=\overline{M}f(S_k\mathbf{X}), \quad \forall f\in\mathcal{F}.$$

Однородной будет независимая последовательность, элементы которой заданы одинаковыми частными ИМ. Или нековариированная, если только определяющие ее равенства (3.6) являются первичными средними совместной ИМ. Свободная последовательность даже при совпадении частных моделей не может быть однородной, так как свойство свободы не является равноправным к перестановкам.

Однородность отражает внешнюю симметрию статистических данных к циклическим сдвигам, наделяя ею ИМ. Более тонким является понятие стационарности. Признак $f(\mathbf{X})$ называется *стационарным* к циклическим сдвигам S_k последовательности $\mathbf{X}$, если

$$M[f(\mathbf{X})-f(S_k\mathbf{X})]=0, \quad \forall S_k.$$

Последовательность $\mathbf{X}$ называется *стационарной,* если стационарны любые признаки $f\in\mathcal{F}^{\mathbf{X}}$ из области существования средних. Важно, что в определении стационарности M — точное среднее.

Следствия стационарности последовательности. 1. Для всех частных признаков из области существования $M[f(X_i)-f(X_j)]=0$, $\forall i, j$ (очевидно, так как частные признаки принадлежат $\mathcal{F}^{X}$). 2. Из стационарности следует однородность. В самом деле $\overline{M}f(S_k\mathbf{X})=\overline{M}[f(S_k\mathbf{X})-f(\mathbf{X})+f(\mathbf{X})]=M[f(S_k\mathbf{X})-f(\mathbf{X})]+\overline{M}f(\mathbf{X})=\overline{M}f(\mathbf{X})$. 3. Если бы среднее $\overline{M}f(\mathbf{X})$ какого-то признака стало бы вдруг точно известным, то таким же оно оказалось бы при любых циклических сдвигах последовательности: $\overline{M}_{Mf(\mathbf{X})}f(S_k\mathbf{X})=Mf(\mathbf{X})$, где слева — среднее по сечению модели.

Стационарность последовательности как род статистической ее устойчивости (отсюда и точные M) вкладывается во внутреннюю симметрию ИМ и проявляется в абсолютной взаимной подчиненности средних (следствие 3). Стационарность обязана на практике неизменности во времени условий генерации элементов X_i. При стационарности даже голые частные модели не делают совместную голой. Например, пусть последовательность независима, т. е. определяется равенствами (3.5), и стационарна. С классических позиций имеем независимую одинаково распределенную выборку с неизвестными распределениями вероятностей элементов. Это отнюдь не голая модель, так как стационарность есть уже весьма существенные знания. Мы увидим в последней главе, как стационарность облегчает оценку средних частных признаков по длинному ряду наблюдений.

Обобщения. 1. Если стационарными являются не все признаки, а только набора $\mathcal{Q}$, то последовательность называется $\mathcal{Q}$-стационарной. Стационарными будем называть и соответствую-

щие признакам $q \in \mathcal{Q}$ параметры Mq (это не средние модели, а направления их сечения). Так можно говорить о стационарности среднего MX (соответствует стационарному признаку X), среднеквадратического MX^2, набора вероятностей и т. д.

2. Обобщением будет определение стационарности не к сдвигам, а к какой-то другой группе операторов $\mathbf{S}_k$, например группе всех перестановок. Оба обобщения относятся и к однородности.

3. Свойство стационарности модели эквивалентно в некотором смысле свойству главной диагонали: сечение ее по одной координате определит точно такие же значения всех других.

Зависимые последовательности. Выше уже обсуждалась возможность задания последовательности частными моделями ее элементов. На том же принципе базируется задание последовательности упрощенными совместными моделями (например, только соседних элементов), отражающими их связь между собой. Так, задавая корреляции соседних элементов $\underline{M}X_iX_{i+1}=\underline{b}$, $\overline{M}X_iX_{i+1}=\overline{b}$, $i=1, 2, \ldots$, получаем, если больше ничего не известно, модель однородной последовательности, для которой корреляции и будут ее первичными значениями. А в общем, заданными могут быть средние $\overline{M}f(X_i, X_{i+1}, \ldots, X_{i+k})$ функций не одного и двух, а любого числа элементов, характеризуя тем самым зависимость не только соседних элементов, но и через один, через два и т. д. элементов.

Определенные упрощения дает здесь допущение об однородности, позволяющее средние для какого-либо одного фрагмента последовательности сразу переносить на любые их циклические сдвиги.

Для задания зависимых последовательностей одна из упрощающих возможностей состоит в выражении ее через более простую, в частности, независимую последовательность ξ. Это будет функциональным представлением вида $\mathbf{X}=\mathbf{v}(\xi)$, одной из форм которой является *рекуррентное представление* любого из двух видов:

$$X_i = v_i(\xi_i, \xi_{i-1}, \ldots, \xi_1),\ X_i = w_i(\xi_i, X_{i-1}, \ldots, X_1),$$

где v_i и w_i — детерминированные преобразования. Второе представление является частью первого, что будет ясно, если последовательно выразить из $X_i=v_1(\xi_i)$ значение $\xi_1=v_1^{-1}(X_1)$, подставить его в $X_2=v_2(\xi_2, \xi_1)$, из которого снова выразить ξ_2 через X_2 и X_1 и т. д.

Рекуррентное представление:

$$X_i = w_i(\xi_i, X_{i-1}, \ldots, X_{i-k}),\ i = 1, 2, \ldots,$$

где ξ_i независимы, называется *k-связным марковским*. Односвязное марковское представление $X_i=w_i(\xi_i, X_{i-1})$ есть способ отражения инерционности значений последовательности и может порождаться как самой физической природой, так и диктоваться удобством, экономностью, подчас привычностью. При интервальных моделях ξ_i марковское представление (впрочем как и другие) делается универсальным: оно достижимо расширением ИМ

ξ_i как средством добиться адекватности выбранного нами описания $\mathcal{M}_{\text{марк}}$ реальному $\mathcal{M}$, понимая под адекватностью включение: $\mathcal{M}_{\text{марк}} \supset \mathcal{M}$.

3.2. СХОДИМОСТИ

Неравенства для случайных величин. Пусть X и Y — две случайные величины. Неважно, как они связываются между собой; например, это могут быть два признака $X=f_1(\xi)$, $Y=f_2(\xi)$ некоторой одной с. в. ξ, или же совершенно разные с. в. последовательности. Справедливы следующие аналоги классических неравенств:

1) Гельдера. При $r>1$ и $1/r+1/s=1$:

$$\overline{M}XY \leqslant (\overline{M}|X|^r)^{1/r} (\overline{M}|Y|^s)^{1/s},\ \underline{M}XY \leqslant (\underline{M}|X|^r)^{1/r} (\overline{M}|Y|^s)^{1/s}.$$

2) Минковского. При $r \geqslant 1$:

$$(\overline{M}|X+Y|^r)^{1/r} \leqslant (\overline{M}|X|^r)^{1/r} + (\overline{M}|Y|^r)^{1/r},$$

а при $0<r<1$

$$\overline{M}|X+Y|^r \leqslant \overline{M}|X|^r + \overline{M}|Y|^r,\quad \underline{M}|X+Y|^r \leqslant \underline{M}|X|^r \mp M|Y|^r.$$

3) Шварца — Буняковского:

$$(\overline{M}|X+Y|^2)^{1/2} \leqslant (\overline{M}X^2)^{1/2} + (\overline{M}Y^2)^{1/2}.$$

4) Маркова. При $r>0$, $a>0$:

$$\underline{P}(|X| \geqslant a) \leqslant \underline{M}|X|^r/a^r,\quad \overline{P}(|X| > a) \leqslant \overline{M}|X|^r/a^r.$$

5) Чебышева:

$$\underline{P}(|X| \geqslant a) \leqslant \underline{M}X^2/a^2,\quad \overline{P}(|X| > a) \leqslant \overline{M}X^2/a^2.$$

6) Йенсена. Если $\psi(x)$ выпукла и имеет производную, то

$$\overline{M}\,\psi(X) \geqslant \psi(\overline{M}X),\ \overline{M}\,\psi(X) \geqslant \psi(\underline{M}X).$$

7) Элементарные неравенства

$$\overline{M}|X+Y|^r \leqslant c_r\,\overline{M}|X|^r + c_r\,\overline{M}|Y|^r,$$

$$\underline{M}|X+Y|^r \leqslant c_r\,\underline{M}|X|^r + c_r\,\overline{M}|Y|^r,$$

где $c_r=1$ при $r \leqslant 1$ и $c_r=2^{r-1}$ при $r \geqslant 1$.

При точных средних эти неравенства переходят в классические.

Доказательство неравенств. 1. Первая формула следует из элементарного неравенства $ab \leqslant |a|^r/r + |b|^s/s$, если заменить в нем сначала a на $X/(\overline{M}|X|^r)^{1/r}$, а b — на $Y'=Y/(\overline{M}|Y|^s)^{1/s}$ и взять верхнее среднее от обеих частей, используя свойство полуаддитивности. Вторая формула получается в отличие от первой заменой a на $X'=X/(\underline{M}|X|^r)^{1/r}$, после чего используется

неравенство: $\underline{M}(|X'|^r+|Y'|^s)\leqslant \underline{M}|X'|^r+\overline{M}|Y'|^s$. Случаи $\overline{M}|X|^r=0$, $\overline{M}|Y|^s=0$ исключаются, так как при этом неравенства становятся тривиальными.

2. Первое из неравенств Минковского доказывается по классической схеме [20, с. 50] из неравенства Гельдера, а вторые два являются следствиями элементарного неравенства: $|a+b|^r\leqslant|a|^r+|b|^r$, $0\leqslant r\leqslant 1$, при $a=X$, $b=Y$.

3. Есть частный случай первого неравенства Минковского при $r=2$.

4. Следует из того, что индикаторы полуотрезков $(-\infty, -a]$, $[a, \infty)$ суть два единичных уступа, простирающиеся от точек $-a$ и a в противоположные стороны, меньшие функции $|x|^r/a^r$.

5. Есть частный случай неравенства Маркова при $r=2$.

6. Доказательство стандартно [20].

7. Следует из неравенства $|a+b|^r\leqslant c_r|a|^r+c_r|b|^r$.

Сходимость моделей. Здесь рассматривается сходимость частных ИМ $\mathcal{M}^{X_n}$ последовательности X_n при $n\to\infty$ к частной ИМ $\mathcal{M}^X$ с.в. X, в смысле сходимости средних $\overline{M}f(X_n)$ к $\overline{M}f(X)$. Обозначим: $\mathcal{F}^{X_n}$ и $\mathcal{F}^X$ — области существования для $\mathcal{M}^{X_n}$ и $\mathcal{M}^X$.

Последовательность с. в. X_n называется:

а) *ИМ-сходящейся к X в направлении класса $\mathcal{H}$ признаков*, что обозначается $\overline{M}\mathcal{H}(X_n)\to\overline{M}\mathcal{H}(X)$, если $\lim\limits_{n\to\infty}\overline{M}h(X_n)=\overline{M}h(X)$, $\forall h\in\mathcal{H}$;

б) *ИМ-сходящейся к X*: $\lim\limits_{n\to\infty}\mathcal{M}^{X_n}=\mathcal{M}^X$, если $\mathcal{F}^{X_n}=\mathcal{F}^X$ и ИМ-сходимость имеет место в направлении всех признаков из $\mathcal{F}^X$;

в) *ИМ-сходящейся в X*: $\lim\mathcal{M}^{X_n}\subset\mathcal{M}^X$, если $\lim\overline{M}f(X_n)\leqslant\overline{M}f(X)$, $\forall f\in\mathcal{F}^X$.

Если пределы средних не существуют, то в определениях они заменяются на $\overline{\lim}=\limsup$.

Теорема 3.2. *Для ИМ-сходимости X_n в X достаточно ИМ-сходимости в направлении набора $\mathcal{G}^X$ первичных признаков с. в. X:*

$$\overline{M}g(X_n)\to\overline{M}g(X),\ \forall g\in G^X\Rightarrow\lim\mathcal{M}^{X_n}\subset\mathcal{M}^X.$$

Доказательство. Пусть $f\in\mathcal{F}^X$. Согласно следствию теоремы 1.1 каждому заданному $\varepsilon>0$ можно указать такую конечную линейную комбинацию $g_\varepsilon=c+\sum c^+_i g_i$, что $g_\varepsilon(x)\geqslant f(x)$ и $\overline{M}g_\varepsilon-\overline{M}f\leqslant\varepsilon/2$, где $\overline{M}g_\varepsilon=c+\sum c^+_i\overline{M}g_i$, $g_i\in\mathcal{G}^X$. В силу сходимости первичных значений будут сходиться $\overline{M}g_\varepsilon(X_n)\to\overline{M}g_\varepsilon(X)$, откуда можно указать такое n_ε, что $|\overline{M}g_\varepsilon(X_n)-\overline{M}g_\varepsilon(X)|\leqslant\varepsilon/2$ при $n>n_\varepsilon$. В результате объединения двух неравенств и $g_\varepsilon\geqslant f$ имеем $\overline{M}f(X_n)\leqslant\overline{M}g_\varepsilon(X_n)\leqslant|\overline{M}g_\varepsilon(X_n)-\overline{M}g_\varepsilon(X)|+\overline{M}g_\varepsilon(X)\leqslant\overline{M}f+\varepsilon$ при $n>n_\varepsilon$. Отсюда $\lim\overline{M}f(X_n)\leqslant\overline{M}f+\varepsilon$. Произвольность ε доказывает требуемое неравенство определения в). При $f\notin\mathcal{F}^X$ результат тривиален, что и требовалось доказать.

Следствие. *Если $\mathcal{M}^{X_n}\supset\mathcal{M}^X$ для всех n, то для ИМ-сходимости X_n к X (т. е. $\lim\mathcal{M}^{X_n}=\mathcal{M}^X$) достаточно ИМ-сходимости в направлении набора $\mathcal{G}^X$ первичных признаков с. в. X.*

Случайная величина X называется *дескриптивной*, если существует последовательность $X_{(k)}$, $k=1, 2, \ldots$, *случайных величин*, описываемых конечным числом k первичных значений, при $k\to\infty$ ИМ-сходящаяся к X. Дескриптивность эквивалентна существованию последовательности $\mathcal{G}^X{}_{(k)}=\{g_1, \ldots, g_k\}$ наборов, таких что $\lim\limits_{k\to\infty}\langle\overline{M}\mathcal{G}^X{}_{(k)}\rangle=\mathcal{M}^X$, где $\langle\overline{M}\mathcal{G}^X{}_{(k)}\rangle$ есть $\mathcal{G}^X{}_{(k)}$-расширение $\mathcal{M}^X$ (полунаблюдаемое, если первичными оставить $\overline{M}g_i=\overline{M}g_i(X)$, $g_i\in\mathcal{G}_{(k)}$).

Дескриптивность — это возможность аппроксимировать модель с.в. X сколь угодно точно конечным числом данных о ней в виде набора средних ее признаков или первичных средних, это гарантия того, что при увеличении k для любого признака f (из области существования $\mathcal{F}^X$) среднее аппроксимирующей модели конечного порядка k будет сходиться к аппроксимируемому: $\lim\limits_{k\to\infty}\overline{M}f(X_{(k)})=\overline{M}f(X)$, $\forall f\in\mathcal{F}^X$.

В частности, дескриптивной будет с. в., определенная точными первичными вероятностями отрезков $P(x, x+\Delta x)$, если плотность $p(x)$ существует и ограничена. Для такой с. в. $\mathcal{G}_{(k)}$ образуют деления ограниченного отрезка $[-H, H]$ на k частей с устремлением длины каждого деления к 0, а H — к ∞.

Последовательность X_n называется *дескриптивной*, если существуют конечные наборы $\mathcal{G}_{(k)}$, такие что равномерно по n:

$$\lim\langle\overline{M}\mathcal{G}^{X_n}_{(k)}\rangle=\langle\overline{M}\mathcal{F}^{X_n}\rangle.$$

Теорема 3.3. *Если последовательность $X_1, X_2, \ldots$ и с.в. X дескриптивны, то для ИМ-сходимости X_n к X достаточно ИМ-сходимости в направлении объединения наборов первичных средних с. в. X_n, $n=1, 2, \ldots$, и X.*

Доказательство. В силу дескриптивности, каждой $f\in\mathcal{F}^X$ и заданному $\varepsilon>0$ можно указать такое k_ε и такую $g_{\varepsilon,k}\in\mathcal{L}+\mathcal{G}_{(k)}$, что $g_{\varepsilon,k}(x)\geqslant f(x)$ и $\overline{M}g_{\varepsilon,k}(X)-\overline{M}f(X)\leqslant\varepsilon$, $\overline{M}g_{\varepsilon,k}(X_n)-\overline{M}f(X_n)\leqslant\varepsilon$ при $\forall k>k_\varepsilon$. Отсюда $|\overline{M}f(X_n)-\overline{M}g_{\varepsilon,k}(X_n)|+|\overline{M}g_{\varepsilon,k}(X_n)-Mg_{\varepsilon,k}(X)|+|Mg_{\varepsilon,k}(X)-\overline{M}f(X)|\leqslant|\overline{M}f(X_n)-\overline{M}g_{\varepsilon,k}(X_n)|+|\overline{M}g_{\varepsilon,k}(X_n)-\overline{M}g_{\varepsilon,k}(X)|+|\overline{M}g_{\varepsilon,k}(X)-\overline{M}f(X)|\leqslant 2\varepsilon+|\overline{M}g_{\varepsilon,k}(X_n)-\overline{M}g_{\varepsilon,k}(X)|$. Поднаборы $\mathcal{G}_{(k)}$ всегда можно считать подмножествами объединения $\mathcal{G}=\bigcup_n\mathcal{G}^{X_n}$ первичных наборов. Сходимость средних на $\mathcal{G}_{(k)}$ будет вызывать сходимость на $\mathcal{L}+\mathcal{G}_{(k)}$, поэтому последнее слагаемое правой части неравенства стремится при $n\to\infty$ к 0. Произвольность ε доказывает сходимость $\lim\overline{M}f(X_n)=\overline{M}f(X)$, что и требовалось.

Условия теоремы 3.3 будут выполнены, если выполняется любое из следующих условий: а) X_n заданы ограниченными плотностями $p_n(x)$, ненулевыми лишь на конечном отрезке и сходящимися к плотности $p(x)$ с.в. X; б) X_n определены интервальными плотностями $\underset{\sim}{p}_n(x)$, $\tilde{p}_n(x)$ (определяющими вероятности отрезков $\underset{\sim}{P}(x, y)$, $\tilde{P}(x, y)$, $x\leqslant y$), сходящимися соответственно к $\underset{\sim}{p}(x)$, $\tilde{p}(x)$; в) X_n определены функциями распределения $\underline{F}_n(z)=\underline{P}(X_n<z)$, $\overline{F}_n(z)=\overline{P}(X_n<z)$, сходящимися к $\underline{F}(z)=\underline{P}(X<z)$, $\overline{F}(z)=\overline{P}(X<z)$.

Сходимость случайных величин и сходимость их моделей. Будем говорить, что X_n сходится к X *в среднеквадратическом* (сквсходится) и писать $X_n\overset{2}{\to}X$, если $\lim\overline{M}(X_n-X)^2=0$. Смысл скв-схо-

димости в том, что при увеличении n значения элементов последовательности X_n все ближе повторяют значения с. в. X, в пределе равняясь им. Скв-сходимость: 1) определена относительно совместных ИМ $\mathcal{M}^{X_n X}$, 2) требует, чтобы признаки $(x_n - x)^2$ принадлежали областям существования средних совместных ИМ (в этом недостаток скв-сходимости), откуда следует $x^2{}_n \in \mathcal{F}^{X_n}$, $x^2 \in \mathcal{F}^X$; 3) вынуждает определенную сходимость частных $\mathcal{M}^{X_n}$ к $\mathcal{M}^X$ (ИМ-сходимость).

Заострим внимание на последнем факте. Обозначим $\mathcal{H}_m$ — класс непрерывных в точке m функций, таких, для которых

$$\sup_{x \in \mathcal{R}} f(x)/x^2 < \infty, \tag{3.7}$$

т. е. имеющих скорость роста при увеличении $|x|$ не быстрее x^2. Пусть $\mathcal{H} = \bigcap_{\forall m} \mathcal{H}_m$ — класс непрерывных на всей $\mathcal{R}$ функций со свойствами (3.7).

Теорема 3.4. *Из скв-сходимости X_n к X следует ИМ-сходимость на классе $\mathcal{H}$ признаков:*

$$X_n \xrightarrow{2} X \Rightarrow \overline{M} f(X_n) \to \overline{M} f(X), \ \forall f \in \mathcal{H}.$$

Смысл теоремы совершенно прозрачен и без формального доказательства (достаточно громоздкого). Если значения X_n приближаются к X, то по непрерывности и $f(X_n)$ будут приближаться к $f(X)$, что вынуждает сходимость средних. Если X принимает значения лишь в ограниченной области Ω, то достаточно непрерывности f лишь в Ω, поэтому сходимость средних сохраняется на расширенном классе $\mathcal{H}_\Omega = \bigcap_{m \in \Omega} \mathcal{H}_m$. В исключительном случае, когда $\Omega = m$ — число, имеем следующее утверждение.

Теорема 3.5. *Сходимость в среднеквадратическом к постоянному числу $X_n \xrightarrow{2} m$ эквивалентна ИМ-сходимости X_n к $X = m$ в направлении класса $\mathcal{H}_m$.*

Таким образом, скв-сходимость к постоянному числу является более слабой формой по сравнению с ИМ-сходимостью (на всей $\mathcal{F}^X$), так как гарантирует сходимость средних лишь на подклассе $\mathcal{H}_m \subset \mathcal{F}^X$. Интересно отметить, что скв-сходимость к постоянному числу эквивалентна сходимости средних всего на трех (входящих в $\mathcal{H}_m$) признаках и равносильна ИМ-сходимости на $\mathcal{H}_m$:

$$X_n \xrightarrow{2} m \Leftrightarrow \{\overline{M} X_n^2 \to m^2, \underline{M} X_n \to m,$$

$$\overline{M} X_n \to m\} \Leftrightarrow \{\overline{M} f(X_n) \to \overline{M} f(m), \ \forall f \in \mathcal{H}_m\}.$$

Сходимость среднего арифметического, закон больших чисел. В теории вероятностей и развиваемой нами интервальной теории моделей случайных явлений этот закон носит ключевой характер.

По своему внутреннему содержанию среднее согласно объяснению § 1.1 есть физическая величина, достижимая как предел среднеарифметического результатов наблюдений за признаком f в серии независимых одинаковых повторений. Для устойчивых явлений пределом будет число Mf, причем сколько бы раз мы ни возвращались к новой серии испытаний — одно и то же. А для неустойчивых — это будут в каждой серии разные числа, но располагающиеся на некотором одном и том же отрезке $[\underline{M}f, \overline{M}f]$, тем более широком, чем глубже «поражены» нестабильностью внутренние законы генерации явления.

Теперь задача состоит в проверке, подтверждает ли сама построенная нами теория тот изначальный смысл, который вкладывался в ее конструкцию? Это и будет основным критерием состоятельности теории (если относить к следующим критериям доступность теории, интерпретируемость параметров и простоту применений). Все данные для указанной проверки уже имеются: определена независимость и, как форма ее проявления, — нековариированность, введены понятия сходимости. Приступим к исследованию среднего арифметического.

Пусть X_i, $i = 1, 2, \ldots$, — последовательность с. в. Ее элементы можно понимать как результаты наблюдений за самой с. в. или же за некоторым признаком $X = f(\xi)$ случайного явления ξ. Будем сначала считать, что средние $MX_i = m_i$ точно известны и $M(x_i - m_i)(X_j - m_j) = 0$ при $i \neq j$ — это есть следствие независимости (см. замечание к (2.10)), названное нами нековариированностью с. в. Она эквивалентна (при точных средних) некоррелированности центрированных с. в. $\mathring{X}_i = X_i - MX_i$, отражаемой равенствами: $\underline{M} \mathring{X}_i \mathring{X}_j = \overline{M} \mathring{X}_i \mathring{X}_j = 0$, $i \neq j$. Для таких $\mathring{X}_i$ верны неравенства:

$$\overline{M} (\Sigma \mathring{X}_i)^2 \leqslant \Sigma \overline{M} \mathring{X}_i^2, \ \underline{M} (\Sigma \mathring{X}_i)^2 \geqslant \Sigma \underline{M} \mathring{X}_i^2. \tag{3.8}$$

Доказываются они элементарно следующим образом:

$$\overline{M} (\Sigma \mathring{X}_i)^2 = \overline{M} \sum_i \sum_j \mathring{X}_i \mathring{X}_j \leqslant \sum_i \sum_j \overline{M} \mathring{X}_i \mathring{X}_j = \Sigma \overline{M} (\mathring{X}_i)^2.$$

Теорема 3.6. Устойчивый вариант закона больших чисел. *Пусть X_i, $i = 1, 2, \ldots$, — последовательность нековариированных с. в. с точными средними $m_i = MX_i$, такими, что существует и конечен предел $m = \lim_{n \to \infty} \frac{1}{n} \sum_1^n m_i$, и ограниченными дисперсиями $\overline{M}(X_i - m_i)^2 = \overline{\sigma^2}_i \leqslant \overline{\sigma^2}$. Тогда при $n \to \infty$ среднее арифметическое $S_n = \frac{1}{n} \sum_{i=1}^n X_i$ этих с. в.:* (I) *будет скв-сходиться $S_n \xrightarrow{2} m$ к постоянному числу m;* (II) *ИМ-сходиться к числу m в направлении класса $\mathcal{H}_m$ непрерывных в точке m признаков, имеющих скорость роста не быстрее x^2 (условие (3.7)).*

Доказательство. Обозначим $\overset{\circ}{X}_i = X_i - m_i$ — центрированные с. в. Для них $\overline{M}(\overset{\circ}{X}_i)^2 \leqslant \overline{\sigma}^2$, $M\overset{\circ}{X}_i\overset{\circ}{X}_j = 0$, $i \neq j$, откуда

$$\overline{M}\left(S_n - \frac{1}{n}\sum_1^n m_i\right)^2 = \overline{M}\left(\frac{1}{n}\sum_1^n \overset{\circ}{X}_i\right)^2 \leqslant$$

$$\leqslant \frac{1}{n^2}\sum_i\sum_j \overline{M}\,\overset{\circ}{X}_i\overset{\circ}{X}_j = \frac{1}{n^2}\sum \overline{M}\,(\overset{\circ}{X}_i)^2 \leqslant \frac{\overline{\sigma}^2}{n}$$

и $\overset{2}{S}_n \to m$. Из последнего согласно теореме 3.5 следует сходимость в направлении $\mathscr{H}_m$, что доказывает вторую часть.

Замечания. 1. В условиях теоремы 3.6 ограниченность дисперсий может быть заменена ограниченностью $\overline{M}X^2_i \leqslant b$, $i = 1, 2, \ldots$ (так как $\overline{M}(X_i - m_i)^2 \leqslant \overline{M}X^2_i$).

2. Если $MS_n = \sum_1^n m_i/n$ при $n \to \infty$ не сходится ни к какому числу, то на основании неравенства

$$\overline{M}\,f(S_n) = \overline{M}\,f(S_n - MS_n + MS_n) \leqslant \sup_{MS_n} \overline{M}\,f(\overset{\circ}{S}_n + MS_n)$$

с учетом вытекающей из теоремы 3.6 сходимости $\overset{\circ}{S}_n \to 0$ доказывается неравенство

$$\overline{\lim_{n\to\infty}}\,\overline{M}\,f(S_n) \leqslant \sup_{\underline{m} \leqslant m \leqslant \overline{m}} f(m), \tag{3.9}$$

где $\underline{m} = \underline{\lim} MS_n$, $\overline{m} = \overline{\lim} MS_n$, справедливое для всех f, непрерывных на концах отрезка $[\underline{m}, \overline{m}]$.

Самый наглядный и самый распространенный вариант закона больших чисел получается, когда все $m_i = m$, $i = 1, 2, \ldots$, т. е. одинаковы. Тогда, очевидно, $MS_n = m$ и среднее арифметическое S_n будет указанным в теореме 3.6 образом сходиться к этому m. Поэтому даже если m было первоначально неизвестно, в пределе, взяв среднее арифметическое наблюдений X_i, получим его точное значение. Это и есть классический закон больших чисел, а правильнее, многих чисел, неоднократно подтвержденный экспериментально, в частности, сериями подбрасывания монеты.

Пусть теперь среднее $m = MX_i$ есть стационарный неизвестный параметр, описываемый частной ИМ $\mathscr{M}^m$, т. е. m по i одно и то же, но не известно какое, а совместная ИМ равна $\mathscr{M}^{\mathbf{X}\,m} = \mathscr{M}^m \times \mathscr{M}^{\mathbf{X}}{}_m$. Относительно переходной $\mathscr{M}^{\mathbf{X}}{}_m$ при каждом m с. в. X_i считаются некоррелированными и $M_m X_i = m$, $i = 1, 2, \ldots$ Тогда S_n будет скв-сходиться к с. в. m, определенной ИМ $\mathscr{M}^m$. Согласно теореме 3.4 отсюда будет следовать сходимость $\overline{M}f(S_n) \to \overline{M}^m f(m)$, $\forall f \in \mathscr{H}$.

В частности, пусть $\mathscr{M}^m$ — индикаторная на отрезке $[\underline{m}, \overline{m}]$ ИМ, т. е. известно лишь, что $\underline{m} \leqslant m \leqslant \overline{m}$ и $\mathscr{M}^{\mathbf{X}} = \bigvee_{\underline{m} \leqslant m \leqslant \overline{m}} \mathscr{M}^{\mathbf{X}}{}_m$. Тогда S_n будет скв-сходиться к с. в. m с индикаторной на $[\underline{m}, \overline{m}]$ ИМ, откуда будет следовать ИМ-сходимость $\overline{M}f(S_n) \to \sup_{\underline{m} \leqslant m \leqslant \overline{m}} f(m)$ в направлении класса $\mathscr{H}_{\underline{m},\,\overline{m}}$ признаков, определенного следующими двумя условиями: 1) для каждого признака из этого класса выполняется (3.7), 2) каждый признак непрерывен на границах отрезка $[\underline{m}, \overline{m}]$ и в точке достижения им максимума. Здесь ИМ-сходимость является следствием сходимости значений.

Закон больших чисел для неустойчивых последовательностей. Откажемся в предыдущих рассуждениях от предположения, что MX_i являются либо точными, либо это неизвестный стационарный параметр. Будем считать, что при каждом i с. в. X_i независимы и их среднее может быть любым внутри интервала $[\underline{M}X_i, \overline{M}X_i]$, и в этом смысле X_i статистически неустойчивы. Тогда S_n не будет скв-сходиться ни к какой с. в. Можно говорить лишь об ИМ-сходимости, т. е. сходимости $\mathscr{M}^{S_n}$.

Теорема 3.7. *Пусть X_i, $i = 1, 2, \ldots$, независимы и заданы $\underline{M}X_i$, $\overline{M}X_i$ и $\overline{M}X^2_i \leqslant b$. Тогда при $n \to \infty$ среднее арифметическое этих с. в. S_n ИМ-сходится в направлении $\mathscr{H}_{\underline{m},\,\overline{m}}$ признаков к индикаторной на $[\underline{m}, \overline{m}]$ ИМ, где*

$$\underline{m} = \underline{\lim}\,\frac{1}{n}\sum_1^n \underline{M}X_i,\quad \overline{m} = \overline{\lim}\,\frac{1}{n}\sum_1^n \overline{M}X_i.$$

Доказательство. Нужно показать, что $\overline{M}f(S_n) \to \max_{\underline{m} \leqslant m \leqslant \overline{m}} f(m)$. Представляя $\mathscr{M}^{X_i} = \bigvee_{\underline{M}X_i \leqslant MX_i \leqslant \overline{M}X_i} \mathscr{M}^{X_i}_{MX_i}$, записываем

$$\mathscr{M}^{\mathbf{X}} = \bigvee_{\underline{M}X_1 \leqslant MX_1 \leqslant \overline{M}X_1} \bigvee_{\underline{M}X_2 \leqslant MX_2 \leqslant \overline{M}X_2} \ldots \left(\mathscr{M}^{X_1}_{MX_1} \times \mathscr{M}^{X_2}_{MX_2} \times \ldots\right) = \bigvee_{M\mathbf{X}} \mathscr{M}^{\mathbf{X}}_{M\mathbf{X}},$$

где $M\mathbf{X} = (MX_1, \ldots, MX_n)$ — вектор средних. Пусть признак $f \in \mathscr{H}_{\underline{m},\,\overline{m}}$ и пусть x_{max} есть точка его максимума внутри $[\underline{m}, \overline{m}]$. Тогда $\overline{M}f(S_n) = \sup_{M\mathbf{X}} \overline{M}_{M\mathbf{X}} f(S_n)$. Взяв в качестве $M\mathbf{X}$ такую последовательность, что $\frac{1}{n}\sum_1^n MX_i \to x_{max}$, получим согласно теореме 3.6 $\overline{M}_{M\mathbf{X}} f(S_n) \to f(x_{max})$, откуда

$$\lim \overline{M}f(S_n) \geqslant f(x_{max}) = \max_{\underline{m} \leqslant m \leqslant \overline{m}} f(m).$$

Осталось доказать противоположное неравенство. Для любого

$$\varepsilon > 0 : \overline{M}\,f(S_n) \leqslant \overline{M}\,f(S_n)\{\underline{m} - \varepsilon \leqslant S_n \leqslant \overline{m} + \varepsilon\} +$$

$$+ \overline{M}\,f(S_n)\{S_n < \underline{m} - \varepsilon\} + \overline{M}f(S_n)\{S_n > \overline{m} + \varepsilon\}.$$

Покажем, что при $n\to\infty$ последние два слагаемые стремятся к 0. Имеем

$$\overline{M}f(S_n)\{S_n<\underline{m}-\varepsilon\}=\sup_{M\mathbf{X}}\overline{M}_{M\mathbf{X}}f(S_n)\{S_n<\underline{m}-\varepsilon\}.$$

Так как для каждого $M\mathbf{X}$ среднее арифметическое его компонент удовлетворяет неравенству $\underline{m}\leqslant\underline{\lim}\,MS_n\leqslant MS_n\leqslant\overline{\lim}MS_n\leqslant\overline{m}$, то, применяя неравенство (3.9), получаем

$$\lim_{n\to\infty}\overline{M}_{M\mathbf{X}}f(S_n)\{S_n<\underline{m}-\varepsilon\}\leqslant\sup_{\underline{m}\leqslant m\leqslant\overline{m}}f(m)\{m<\underline{m}-\varepsilon\}=0,$$

откуда $\overline{M}f(S_n)\{S_n<\underline{m}-\varepsilon\}\to 0$. Аналогично $\overline{M}f(S_n)\{S_n>m+\varepsilon\}\to 0$. В результате

$$\lim_{n\to\infty}\overline{M}f(S_n)\leqslant\lim_{n\to 0}\overline{M}f(S_n)\{\underline{m}-\varepsilon\leqslant S_n\leqslant\overline{m}+\varepsilon\}\leqslant\max_{\underline{m}-\varepsilon\leqslant m\leqslant\overline{m}+\varepsilon}f(m).$$

В силу произвольности $\varepsilon>0$ и непрерывности $f(x)$ в точках $\underline{m}$ и $\overline{m}$ имеем $\lim\limits_{n\to\infty}\overline{M}f(S_n)\leqslant\max\limits_{\underline{m}\leqslant m\leqslant\overline{m}}f(m)$, что и требовалось доказать.

Замечание. Независимость в формулировке теоремы 3.7 может быть заменена на нековариированность X_i для каждого сечения $\mathscr{M}_{M\mathbf{X}}$ совместной модели вектором средних $M\mathbf{X}$.

Смысл теоремы 3.7 наиболее легко раскрывается, когда все $\underline{M}X_i=\underline{m}$, $\overline{M}X_i=\overline{m}$, $i=1, 2, \ldots$, одинаковы, в частности, когда выборка однородна, т. е. средние элементов могут «прыгать» неконтролируемым образом внутри одного и того же отрезка $[\underline{m}, \overline{m}]$. Точно так же будет «прыгать» и среднее MS_n. Причем сами по себе значения S_n могут отклоняться за отрезок $[\underline{m}, \overline{m}]$, но в пределе при $n\to\infty$ эти отклонения становятся все менее и менее возможными.

С платформы эксперимента серия неограниченных повторений опыта ведет в пределе к некоторому числу $m=\lim S_n$. Так вот, если в устойчивом случае согласно теореме 3.6 это должно быть в любой серии одно и то же число, то в неустойчивом мы каждый раз будем получать новые числа, и теорема 3.7 утверждает, что они должны лежать в отрезке $[\underline{m}, \overline{m}]$. Это и будет наиболее точный их диапазон при крайне неточных данных, когда о выборке известно лишь, что ее среднестатистическая мощность ограничена: $\overline{M}X^2_i\leqslant b$, и даны диапазоны средних $\underline{M}X_i$, $\overline{M}X_i$.

Дополнения. 1. Сходимость среднего арифметического в сечениях. *Пусть X_i независимы и имеют ограниченную среднюю мощность: $\overline{M}X^2_i\leqslant b$, $i=1, 2, \ldots$ Тогда при $n\to\infty$ их среднее арифметическое $S_n=\sum X_i/n$ для каждого заданного $m=MS_n$ (т. е. в каждом своем $m=MS_n$-сечении) сквсходится к m.*

Утверждение следует непосредственно из закона больших чисел 3.6, примененного к m-сечениям. Утверждается, что если бы средние m_i элементов стали точно известны, то к их предельному среднему арифметическому $m=\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_1 m_i$ скв-сходилось бы S_n.

2. Разновидности сходимостей. Последовательность X_n, при $n\to\infty$, называется сходящейся (по своим значениям) к с. в. X:

а) *в среднем* ($X_n\xrightarrow{r}X$), если $\overline{M}|X-X_n|^r\to 0$, где $r>0$;

б) *по вероятности* ($X_n\xrightarrow{\text{в}}X$), если $\bar{P}(|X-X_n|>\varepsilon)\to 0$, $\forall\varepsilon>0$;

в) *почти всюду* ($X_n\xrightarrow{\text{пв}}X$), если $\bar{P}(\bigcap\limits_1^n\{|X-X_n|>\varepsilon\})\to 0$, $\forall\varepsilon>0$.

Верны утверждения, близкие по своей сути и по способу доказательства к классическим [1]:

1) $X_n\xrightarrow{r}X\Rightarrow X_n\xrightarrow{\text{в}}X\Leftarrow X_n\xrightarrow{\text{пв}}X$.

2) $X_n\xrightarrow{r}X\Rightarrow X_n\xrightarrow{r'}X$, $\forall r'\leqslant r$.

3) X_n равномерно ограничены ($\exists h:\bar{P}(X_n>h)=0$, $\forall n$) и $X_n\xrightarrow{\text{в}}X\Rightarrow X_n\xrightarrow{r}X$.

4) Если X_n сходится (r, в., пв) к X и X_n равномерно ограничены, то $\overline{M}X_n\to\overline{M}X$, $\underline{M}X_n\to\underline{M}X$.

5) При условии конечности моментов $\overline{M}|X_n|^r<h$, $\overline{M}|X|^r<h$ верно:

$$X_n\xrightarrow{\text{в}}X\Rightarrow X_n\xrightarrow{r'}X,\quad\forall r'\leqslant r\Rightarrow X_n\xrightarrow{1}X\Rightarrow\overline{M}X_n\to\overline{M}X,\quad\underline{M}X_n\to\underline{M}X.$$

6) Если $f(x)$ абсолютно непрерывна, то сходимость $X_n\to X$ в., r или пв. влечет точно такую же сходимость $f(X_n)\to f(X)$, причем если $f(x)$ ограничена, то $\underline{M}f(X_n)\to\underline{M}f(X)$, $\overline{M}f(X_n)\to\overline{M}f(X)$.

7) Пусть моменты $\overline{M}|X|^r$ и $\overline{M}|X_n|^r$, $\forall n$, конечны. Тогда из сходимости в среднем будет следовать сходимость моментов:

$$X_n\xrightarrow{r}X\Rightarrow\overline{M}|X_n|^r\to\overline{M}|X|^r,\quad\underline{M}|X_n|^r\to\underline{M}|X|^r.$$

3.3. ДОПРЕДЕЛЬНАЯ И ПРЕДЕЛЬНАЯ ПРОБЛЕМЫ

Аппроксимация модели суммы независимых с. в. Рассмотрим сумму $\sum X_i$ (для краткости пределы суммирования, где они не обязательны, будем опускать). Слагаемые считаются независимыми и по праву независимого произведения совместная ИМ вектора $\mathbf{X}=(X_1, \ldots, X_n)$ полностью определяется частными: $\mathscr{M}^{\mathbf{X}}=\mathscr{M}^{X_1}\times\ldots\times\mathscr{M}^{X_n}$.

Первичными для совместной модели $\mathscr{M}^{\mathbf{X}}$ будут всевозможные произведения частных признаков $\prod g_i(x_i)$ с перенесением на них средних по закону интервальной мультипликативности (3.5). Расчет модели суммы производится по формуле продолжения

$$\overline{M}f\left(\sum X_i\right)=\inf\left\{\sum_j\overline{\prod_i Mg_{ij}(X_i)}:\sum_j\prod_i g_{ij}(x_i)\geqslant f\left(\sum x_i\right)\right\},$$

где нижняя грань ищется выбором $g_{ij}(x_i)$ из области существования $\mathscr{F}_i{}^{X_i}$ частных моделей.

Вычисления по этой формуле весьма громоздки. Исключение составляют те признаки f, которые при подстановке в них на место аргумента суммы $\sum x_i$ сами разлагаются на суммы произведений $\sum\prod g_{ij}(x_i)$, тогда среднее на произведениях находится по

свойству мультипликативности, а для средних сумм приближенные значения получаются по свойству полуаддитивности. Рассмотрим примеры конкретных f, их разложений и средних от сумм.

Линейный признак:

$$f(x)=x;\ f\left(\sum x_i\right)=\sum x_i;\ \overline{M}\sum X_i=\sum \overline{M}X_i.$$

Здесь считается $x\in\mathscr{F}_i{}^{X_i}$, $\forall i$, а равенство в правой части (вместо полуаддитивности) имеет место на основании свойства аддитивности средних сумм независимых с. в.

Квадратичный признак:

$$f(x)=x^2;\ f\left(\sum x_i\right)=\sum x_i^2+\sum_{i\neq j}\sum x_i x_j;$$

$$\overline{M}\left(\sum X_i\right)^2\leqslant\sum \overline{M}X_i^2+\sum_{i\neq j}\sum \overline{MX_i\,MX_j},$$

где неравенство обязано свойству полуаддитивности и считается x, $x^2\in\mathscr{F}_i{}^{X_i}$, $\forall i$, (ниже это молча предполагается). Правая часть полученного неравенства и является оценкой верхнего среднего квадрата суммы.

Несколько более точную оценку можно получить, если взять $M\mathbf{X}=(MX_1,\ \ldots,\ MX_n)$-сечение модели; тогда для верхнего среднего в силу того, что в сечениях случайные величины X_i остаются независимыми, верно равенство

$$\overline{M}_{M\mathbf{X}}\left(\sum X_i\right)^2=\sum \overline{M}_{M\mathbf{X}}X_i^2+\sum_{i\neq j}\sum MX_i\,MX_j.$$

Теперь если взять максимум правой части по $MX_i\in[\underline{M}X_i,\ \overline{M}X_i]$, $i=1,\ \ldots,\ n$, то с учетом того, что $\overline{M}_{M\mathbf{X}}X^2{}_i\leqslant\overline{M}X^2{}_i$ и максимум достигается при MX_i, равном $\underline{M}X_i$ или $\overline{M}X_i$, получим окончательную оценку в виде правой части неравенства

$$\overline{M}\left(\sum X_i\right)^2\leqslant\sum \overline{M}X_i^2+\max_{M^{(i)}=\underline{M}\text{ или }\overline{M}}\sum M^{(i)}X_i\,M^{(j)}X_j.$$

Подобное выражение верно и для нижнего среднего, только в этом случае неравенство меняет знак, верхнее среднее заменяется на нижнее, а максимум на минимум. Такой же путь возможен для оценки нижних и верхних средних $\underline{M}(\sum X_i)^r$, $\overline{M}(\sum X_i)^r$ степенных признаков порядка $r>2$, называемых начальными моментами, а также для оценки средних их линейных комбинаций — полиномов: $\overline{M}(a_1\sum X_i+\ldots+a_k(\sum X_i)^k)$, так как они тоже разлагаются на суммы произведений X_i.

Оценки в виде правых частей неравенств для средних значений степенных функций (отсюда и полиномов) образуют набор, аппроксимирующий сверху модель суммы, т. е. позволяющий сформировать расширенную модель суммы, пользуясь лишь знаниями моментов слагаемых. Одно из достоинств именно такого расширения заключается в непосредственной физической интерпретируемости характеристик, на которых оно основывается: первый момент есть среднее с. в., второй — средняя статистическая

мощность и т. д. А другое важное достоинство обязано тому, что согласно первой теореме Вейерштрасса [23, стр. 39] любую непрерывную (и кусочно-непрерывную) функцию на ограниченном отрезке можно сколь угодно точно в равномерной метрике аппроксимировать полиномами, что делает класс степенных признаков весьма распространенным. Степенные признаки x^k, $k=1, 2, \ldots$, образуют *первый универсальный класс признаков*.

Некоторое стеснение при работе в этом классе вызывает необходимость полагать, что $x^k{}_i\in\mathscr{F}_i{}^{X_i}$, $\forall i$, где k — порядок старшего момента. А так как x^k — неограниченная функция и она не обязана принадлежать области существования модели, то становится вынужденной формальная фраза: «Пусть существуют моменты k-го порядка случайных величин X_i». Голословность этой фразы оправдывается, возможно, тем, что для практики преобладающее большинство с. в. являются ограниченными и моменты существуют.

В дополнение к степенным и полиномиальным признакам укажем еще на экспоненциальные признаки как возможное направление для аппроксимации моделей суммы:

$$f_э(ux)=\exp(ux),$$

$$\underline{M}\exp\left(\sum u X_i\right)=\prod \underline{M}\exp(u X_i),\qquad \overline{M}\exp\left(\sum u X_i\right)=$$

$$=\prod \overline{M}\exp(u X_i),$$

где в последних формулах использовано свойство мультипликативности верхнего и нижнего средних на произведениях частных неотрицательных признаков независимых с. в. Незначительность в применениях экспоненциальных признаков обусловлена не столько их неограниченностью, хотя и это тоже влияет, сколько неудобством разложения произвольных функций в ряды по ним, взятым в качестве базиса.

Гармоническая аппроксимация. *Второй универсальный класс признаков* для расчета средних сумм образуется набором гармонических функций $f_s(ux)=\sin ux$, $f_c(ux)=\cos ux$, где u — параметр, $-\infty<u<\infty$ — своего рода индекс гармоники. Гармоники ограничены и формируют плотный класс в том смысле, что (согласно второй теореме Вейерштрасса [23, с. 41]) их линейными комбинациями могут быть сколь угодно приближены любые ограниченные непрерывные на конечном отрезке функции (аппроксимацию дает разложение Фурье). Гармонические функции при подстановке на место аргумента сумм разлагаются на суммы произведений, что позволяет произвести подсчет их средних, которым сейчас и займемся.

Назовем $|\overline{M}|f=\max\{|\underline{M}f|,\ |\overline{M}f|\}$ — *абсолютным средним признака* f. Обозначим

$$\underline{\nu}_i(u)=\underline{M}\cos u X_i,\ \underline{\lambda}_i(u)=\underline{M}\sin u X_i,$$

$$\overline{\nu}_i(u)=\overline{M}\cos u X_i,\ \overline{\lambda}_i(u)=\overline{M}\sin u X_i$$

и назовем *гармоническими средними с. в.* X_i. Введем *абсолютные гармонические средние*: $|\overline{\nu_i}|(u)=|\overline{M}|\cos uX_i$, $\Lambda_i(u)=|\overline{M}|\sin uX_i$,

$$|\nu_i|(u)=\min_{\underline{\nu}_i\leqslant\nu_i\leqslant\overline{\nu}_i}|\nu_i(u)|=\begin{cases}\underline{\nu}_i(u) & \text{при } \underline{\nu}_i(u)>0,\\ 0 & \text{при } \underline{\nu}_i(u)\leqslant 0\leqslant\overline{\nu}_i(u),\\ -\overline{\nu}_i(u) & \text{при } \overline{\nu}_i(u)<0.\end{cases}$$

очевидно, являющиеся неотрицательными и $|\underline{\nu_i}|(0)=|\overline{\nu_i}|(0)=1$, $\Lambda_i(0)=0$. По существу, $|\underline{\nu_i}|(u)$ есть минимальное по $\underline{M}\leqslant M\leqslant\overline{M}$ абсолютное среднее косинуса $|M\cos uX_i|$ с. в. X_i, $|\overline{\nu_i}|(u)$ — ее максимальное по модулю значение, а $\Lambda_i(u)=|\overline{\lambda_i}|(u)$ — то же для синуса $M\sin uX_i$.

Теорема 3.8. Основные неравенства для гармонических средних сумм. *Пусть* $X_1, \ldots, X_n$ *независимы. Тогда для всех* u *и* φ:

(I) $$|\overline{M}|\cos\left(u\sum_{i=1}^{n}X_i+\varphi\right)\leqslant\overline{A}_\Sigma(u)=$$
$$=\prod_{i=1}^{n}\sqrt{|\overline{\nu_i}|^2(u)+\Lambda_i^2(u)}\,.$$

А для тех u, *для которых* $\min_i\underline{\nu}_i(u)>0$, *а также*

$$\varphi_\Sigma(u)=\sum_{i=1}^{n}\operatorname{arctg}(\Lambda_i(u)/\underline{\nu}_i(u))\leqslant\pi/2,$$

справедливы следующие три неравенства:

(II) $$\overline{M}\cos u\sum_1^n X_i\leqslant\prod_1^n\overline{\nu}_i(u);$$

(III) $$\underline{M}\cos u\sum_1^n X_i\geqslant\prod_1^n\sqrt{\underline{\nu}_i^2(u)+\Lambda_i^2(u)}\cos\varphi_\Sigma(u);$$

(IV) $$|\overline{M}|\sin u\sum_1^n X_i\leqslant\overline{A}_\Sigma(u)\sin\left[\sum_1^n\operatorname{arctg}(\Lambda_i(u)/\overline{\nu}_i(u))\right].$$

Доказательство. Используем при каждом фиксированном $u\in\mathscr{R}$ представление

$$\langle\overline{M}\mathscr{F}^X\rangle=\vee\ldots\vee\langle\overline{M}_{\nu_1,\lambda_1}\mathscr{F}_1^{X_1}\rangle\times\ldots\times\langle\overline{M}_{\nu_n,\lambda_n}\mathscr{F}_n^{X_n}\rangle=\bigvee_{\nu,\lambda}\langle\overline{M}_{\nu,\lambda}\mathscr{F}^X\rangle,$$

где $\langle\overline{M}_{\nu_i,\lambda_i}\mathscr{F}_i^{X_i}\rangle$ есть $\nu_i=M\cos uX_i$, $\lambda_i=M\sin uX_i$-сечения $\langle\overline{M}\mathscr{F}_i^{X_i}\rangle$, а $\nu=(\nu_1,\ldots,\nu_n)$, $\lambda=(\lambda_1,\ldots,\lambda_n)$. Обозначая Re — действительную часть комплексного числа, а Im — мнимую, используя равенство $\cos(u\sum x_i+\varphi)=\operatorname{Re}\prod_1^n e^{j(ux_i+\varphi)}$, где j — комплексная единица, с учетом того, что правая его часть есть сумма произведений $\cos ux_i$ и $\sin ux_i$, а потому при точных ν_i и λ_i

среднее $M_{\nu,\lambda}$ будет точным и проносится за знак суммы и произведений, получаем

$$\overline{M}\cos\left(u\sum_1^n X_i+\varphi\right)=\max_{\nu,\lambda}\operatorname{Re}\prod_1^n(\nu_i(u)+j\lambda_i(u))\,e^{j\varphi}=$$
$$=\max_{\nu,\lambda}\prod_1^n\sqrt{\nu_i^2(u)+\lambda_i^2(u)}\times\cos\left[\sum_1^n\operatorname{arctg}(\lambda_i(u)/\nu_i(u))+\varphi\right]. \qquad (3.10)$$

Так как косинус меньше 1, то неравенства (I) теоремы отсюда становятся очевидными. Докажем остальные.

Выделим какое-нибудь одно значение индекса $i=k$ и преобразуем фрагмент правой части (3.10), зависящий от k (опустив для краткости аргумент u и положив $\varphi=0$)

$$\sqrt{\nu_k^2+\lambda_k^2}\cos[\varphi_k+\operatorname{arctg}(\lambda_k/\nu_k)]=\nu_k\cos\varphi_k+\lambda_k\sin\varphi_k,$$

где φ_k обозначена сумма в аргументе косинуса (3.10) без k-го слагаемого. Теперь видно, что при $|\varphi_k|\leqslant\pi/2$ максимум достигается при $\nu_i=\overline{\nu_i}$ и $\lambda_i=0$ (это будет еще яснее, если максимизировать не по каждому λ_i, а по всему вектору λ), а минимум — при $\nu_i=\underline{\nu_i}$ и $\lambda_i=|\overline{\lambda_i}|$, что доказывает (II) и (III).

Неравенство (IV) следует из соотношений

$$\left|\sqrt{\nu_i^2+\lambda_i^2}\sin(\varphi_i+\operatorname{arctg}\lambda_i/\nu_i)\right|=|\lambda_i\cos\varphi_i+\nu_i\sin\varphi_i|\leqslant$$
$$\leqslant|\overline{\lambda_i}|\cos|\varphi_i|+\overline{\nu}_i\sin|\varphi_i|,$$

что и требовалось.

Теорема позволяет по гармоническим средним слагаемых получать оценки гармонических средних сумм, эти оценки даются правыми частями неравенств (I)—(IV). Перейдем к одному ее упрощенному случаю.

Допредельная проблема, однородный случай. Рассмотрим однородную последовательность. Для нее по определению гармонические средние для каждой из с. в. X_i будут одни и те же при $i=1,\ldots,n$.:

$$\underline{M}\cos u\,X_i=\underline{\nu}(u),\ \overline{M}\cos u\,X_i=\overline{\nu}(u),\ |\overline{M}|\sin u\,X_i=\Lambda(u).$$

Любая последовательность может быть сделана таковой, если расширить ее модель, доводя интервалы гармонических средних объединением до самого широкого: $\underline{\nu}(u)=\min_i\underline{M}\cos uX_i$; $\overline{\nu}(u)=\max_i\overline{M}\cos uX_i$, и то же самое для синуса, и оставляя их в качестве первичных. Из теоремы 3.8 вытекает:

Утверждение 3.9. *Для сумм однородной последовательности независимых с. в.* $X_1, \ldots, X_n$ *при всех* u *и* φ *верно неравенство*

(!) $$|\overline{M}|\cos\left(u\sum X_i+\varphi\right)\leqslant(\overline{\nu}^2(u)+\Lambda^2(u))^{n/2},$$

а при $\Lambda(u)/\underline{\nu}(u) \leqslant \operatorname{tg}(\pi/2n)$ *справедливы уточненные неравенства*

(II) $\overline{M}\cos\left(u\sum_1^n X_i\right) \leqslant \overline{\nu}^n(u)$;

(III) $\underline{M}\cos\left(u\sum_1^n X_i\right) \geqslant (\underline{\nu}^2(u)+\Lambda^2(u))^{n/2}\cos\left(n\operatorname{arctg}\frac{\Lambda(u)}{\underline{\nu}(u)}\right)$;

(IV) $|\overline{M}|\sin\left(u\sum_1^n X_i\right) \leqslant (\overline{\nu}^2(u)+\Lambda^2(u))^{n/2}\sin\left(n\operatorname{arctg}\frac{\Lambda(u)}{\underline{\nu}(u)}\right)$.

Суть допредельной проблемы применительно к гармоническим признакам заключается в упрощении, унификации правых частей введенных только что неравенств при достаточно большом числе n слагаемых. Наша цель — проследить, как по мере роста n и увеличения данных о слагаемых сужаются в направлении гармонических, а затем и степенных признаков модели сумм.

Для практики, это ответ на вопрос, что общего и что конкретного можно сказать о модели суммы, скажем, пяти, десяти, ста независимых с. в., если имеются какие-то данные о слагаемых и вариант прямого расчета модели суммы исключается из-за излишней трудоемкости.

Вернемся к правым частям неравенств утверждения 3.9. Во-первых, упрощения в них возможны для биномов, записываемых (для простоты аргументы у $\nu(u)$ и $\Lambda(u)$ далее опускаются):

$$(\nu^2+\Lambda^2)^{n/2} = [1-(1-\nu^2-\Lambda^2)]^{n/2},$$

где ν равняется соответственно верхнему либо нижнему значениям. При тех условиях, когда малы $1-\nu^2$ (напомним, что $|\nu|\leqslant 1$) и Λ, т. е. в результате мало $F(\nu,\Lambda)=1-\nu^2-\Lambda^2$ (причем $F\geqslant 0$), бином правой части аппроксимируется экспонентой

$$[1-F(\nu,\Lambda)]^{n/2} = \exp\left\{-\frac{F(\nu,\Lambda)}{2}-\delta_n\left(\frac{F(\nu,\Lambda)n}{2}\right)\right\} \text{ при } |F(\nu,\Lambda)| \leqslant 1,$$

где $\delta_n(y) = -y-\frac{n}{2}\ln\left(1-\frac{2y}{n}\right) = 1+\frac{y}{2}\left(\frac{2y}{n}\right)+\frac{y}{3}\left(\frac{2y}{n}\right)^2+\ldots$ (получается, если подставить $y=Fn/2$, взять логарифм и разложить в ряд Маклорена). Поправка $\delta_n(y)$ возрастает при увеличении $y>0$ (равна ∞ при $y=n/2$) и убывает при росте n.

И во-вторых, при малых $1-\nu^2$ и Λ^2 упрощаются аргументы при косинусе и синусе в (III)—(IV). А именно, так как $\operatorname{arctg} x \leqslant x$ при $x>0$ и косинус на первой четверти периода убывает при увеличении аргумента, а синус — возрастает, то имеем:

Утверждение 3.10. *Для сумм однородной последовательности независимых с. в.* $X_1, \ldots, X_n$ *при всех* u *и* φ *верно неравенство*:

(I°) $|\overline{M}|\cos\left(u\sum_1^n X_i+\varphi\right) \leqslant \exp\left\{-\frac{F(\overline{\nu},\lambda)n}{2}-\delta_n\left(\frac{F(\overline{\nu},\Lambda)n}{2}\right)\right\}$,

а при u, *таких, что* $n\Lambda(u)/\underline{\nu}(u) \leqslant \pi/2$, *справедливы уточненные неравенства*:

(II°) $\overline{M}\cos\left(u\sum_1^n X_i\right) \leqslant \exp\left\{-\frac{(1-\overline{\nu}^2)n}{2}-\delta_n\left(\frac{(1-\overline{\nu})^2 n}{2}\right)\right\}$;

(III°) $\underline{M}\cos\left(u\sum_1^n X_i\right) \geqslant \exp\left\{-\frac{F(\underline{\nu},\Lambda)n}{2}-\delta_n\left(\frac{F(\underline{\nu},\Lambda)n}{2}\right)\right\}\cos\left(n\cdot\frac{\Lambda}{\underline{\nu}}\right)$;

(IV°) $|\overline{M}|\sin\left(u\sum_1^n X_i\right) \leqslant \exp\left\{-\frac{F(\overline{\nu},\Lambda)n}{2}-\delta_n\left(\frac{F(\overline{\nu},\Lambda)n}{2}\right)\right\}\sin\left(n\cdot\frac{\Lambda}{\overline{\nu}}\right)$.

Здесь $F(\nu,\Lambda)=1-\nu^2-\Lambda^2$. Неравенства имеют содержательный смысл, когда F настолько мало, что $Fn/2$ — небольшая величина.

Чтобы проследить зависимость средних значений, даваемых правыми частями неравенств (I°)—(IV°), от n удобно *нормировать* сумму $\sum X_i/\sqrt{n}$. В новой сумме при росте n диапазон случайного разброса каждого слагаемого $X_{in}=X_i/\sqrt{n}$ стягивается к 0, отсюда к нулю устремляется $\Lambda_{(n)}(u)=|\overline{M}|\sin uX_{in}$, так как аргумент синуса и сам синус становятся ничтожно малыми, а к единице устремляется $\underline{\nu}_{(n)}(u)=\underline{M}\cos uX_{in}$ (тем более,, $\overline{\nu}_{(n)}(u)$), так как косинус в пределах малых изменений аргумента будет близок к 1. Отсюда $F_{(n)}=1-\nu^2_{(n)}-\Lambda^2_{(n)}$ уменьшается, причем при определенных условиях пропорционально n, и тогда произведение $F_{(n)}n$ будет стабилизировано по n. Соответственно стабилизируются и правые части неравенств утверждения 3.10, дающие допредельные оценки гармонических средних для сумм (см. дополнение 1).

Допредельная проблема была бы не полной, если не рассматривать степенные признаки и связанные с ними оценки и упрощения. Это можно сделать лишь для определенного типа слагаемых сумм. А именно, для симметричных X_i с ограниченными моментами (определение см. в начале § 3.1), а более широко, считая нулевыми все нечетные моменты вплоть до порядка $2k$: $MX_i^{2r-1}=0$, $r=1, \ldots, k$. Тогда моменты нечетных порядков суммы вплоть до $2k-1$ будут нулевыми: $M(\sum X_i/\sqrt{n})^{2r-1}=0$, так как при разложении $(\sum X_i/\sqrt{n})^{2r-1}$ на слагаемые в каждом из них будет обязательно присутствовать хотя бы одна с. в. X_i в нечетной степени, среднее которой — ноль.

Для четных моментов сумм симметричных с. в. прямое разложение на слагаемые с последующим взятием среднего и исполь-

зованием свойств полуаддитивности и независимости ведет к неравенству

$$\overline{M}\left(\sum_1^n X_i/\sqrt{n}\right)^{2k} \leqslant \frac{1}{n^k}\left[n\,\overline{\mu}_{2k} + n(n-1)^+ \sum_j \overline{\mu}_{2(k-j)}\,\overline{\mu}_{2j}\,C_{2k}^{2j}/2! + n(n-1)(n-2)^+ \sum_{j+l<k} \overline{\mu}_{2(k-j-l)}\,\overline{\mu}_{2j}\,\overline{\mu}_{2l}\,\frac{(2k)!}{(2(k-j-l))!\,(2j)!\,(2l)!3!} + \ldots + n(n-1)\ldots(n-k+1)^+ \overline{\mu}_2^k\,\frac{(2k)!}{2^k k!}\right], \tag{3.11}$$

где $\overline{\mu}_{2j} = \overline{M}X_i^{2j}$. Противоположное неравенство с заменой верхних моментов на нижние $\underline{\mu}_{2j}$ верно для нижних средних $\underline{M}(\sum X_i/\sqrt{n})^{2k}$.

В частности, при $k=1$ в силу независимости X^2_i, а как следствие, аддитивности средних от суммы, имеет место равенство:

$$\overline{M}\left(\sum_1^n X_i/\sqrt{n}\right)^2 = \overline{\sigma}^2, \quad \underline{M}\left(\sum_1^n X_i/\sqrt{n}\right)^2 = \underline{\sigma}^2, \tag{3.12}$$

где $\overline{\sigma}^2 = \overline{\mu}_2 = \overline{M}X_i^2$, $\underline{\sigma}^2 = \underline{\mu}_2 = \underline{M}X_i^2$.

Вместе с (3.12) правые части (3.11) и есть допредельные оценки моментов сумм симметричных с. в. Самым замечательным в (3.11) является то, что значимым при увеличении n становится лишь последнее слагаемое. Этот факт занимает ключевое положение в предельной проблеме, к предварительному освещению которой для однородных последовательностей и переходим, оставив на § 3.4 развитие допредельной и предельной проблем на суммы общего вида.

Введение в предельную проблему. Суть ее составляют те предельные при $n\to\infty$ упрощения, которые получаются с точными значениями или с приближенными в виде правых частей неравенств для средних значений ключевых признаков нормированных сумм. Предельные теоремы, во-первых, дают ответ на вопрос, для каких конкретно признаков и при каких условиях средние допускают нетривиальные приближения. И во-вторых, указывают на предельные при $n\to\infty$ значения приближений, формирующие расширенную предельную модель сумм.

Наша конечная цель — проследить, как по мере накопления и уточнения данных о слагаемых (сужения их моделей) сужается предельная модель нормированной суммы и, в конечном счете, сходится к нормальной. Сходимость к нормальному закону прослеживается в двух универсальных направлениях: на классе степенных признаков и на классе гармонических согласно теореме 3.1 характеризации нормальной с. в.

Для степенных признаков сходимость означает сходимость моментов к нормальным значениям $MY^{2k-1}=0$, $MY^{2k} = \sigma^{2k}(2k)!/(k!2^k)$ (см. 3.4)), что в случае симметричных с. в. X_i вытекает непосредственно из (3.11) и (3.12) и приводит к следующему:

Утверждение 3.11. *Пусть последовательность X_i независима, однородна, ограничена (существуют все моменты), симметрична (средние MX_i и моменты нечетных порядков все равны 0) и пусть $\underline{M}X^2_i = \underline{\sigma}^2$, $\overline{M}X^2_i = \overline{\sigma}^2$. Тогда при $k=1, 2, \ldots$, справедливы такие соотношения для моментов нормированных сумм:*

1°. $M\left(\sum_1^n X_i/\sqrt{n}\right)^{2k-1} \equiv 0$;

2°. $\lim\limits_{n\to\infty} \overline{M}\left(\sum_1^n X_i/\sqrt{n}\right)^{2k} \leqslant \overline{\sigma}^{2k}(2k)!/(k!\,2^k)$;

3°. $\lim\limits_{n\to\infty} \underline{M}\left(\sum_1^n X_i/\sqrt{n}\right)^{2k} \geqslant \underline{\sigma}^{2k}(2k)!/(k!\,2^k)$.

В самом деле, в выражении (3.11) все слагаемые, кроме последнего, исчезают при увеличении n, а последнее и дает правую часть неравенства 2°. Для нижнего среднего неравенство (3.11) заменяется на противоположное, откуда следует 3°.

Правые части неравенств утверждения 3.11 в качестве первичных определяют предельную ИМ, сужающуюся по мере уточнения интервала $\underline{\sigma}^2$, $\overline{\sigma}^2$ дисперсий к нормальной модели.

Следствие. *Если в условиях утверждения 3.11 дисперсия является точной $\underline{\sigma}^2 = \overline{\sigma}^2 = \sigma^2$, то имеет место ИМ-сходимость суммы к нормальной с. в.*

Замечание. Чем старше порядок k, тем, в общем-то, медленнее имеет место сходимость моментов к их нормальным значениям. Сказанное верно хотя бы потому, что при этом в правой части (3.11) будет большее число слагаемых помимо последнего, неисчезающего.

В направлении гармонических признаков согласно теоремам 3.1, 3.8 и формуле (3.3) нормальной считается сходимость к нулю синусоидальных средних сумм $M\sin(u\sum X_i/\sqrt{n})$ (для симметричных слагаемых они точно равны нулю) и к $\exp(-\sigma^2u^2/2)$ — косинусоидальных $M\cos(u\sum X_i/\sqrt{n})$ при контроле условия, что $\lim\limits_{n\to\infty} \bar{P}(\sum X_i/\sqrt{n} > H) \underset{H\to\infty}{\to} 0$. Последнее условие всегда выполняется для независимой последовательности с нулевыми средними $MX_i=0$ и ограниченной дисперсии $\overline{M}X^2_i \leqslant \overline{\sigma}^2$ слагаемых, так как из аналога неравенства Чебышева (§ 3.2) с учетом равенства (3.12) получается:

$$\bar{P}\left(\sum X_i/\sqrt{n} > H\right) \leqslant \overline{\sigma}^2/H^2 \underset{H\to\infty}{\to} 0.$$

При условиях следствия совершенно резонно полагать, что гармонические средние также должны сходиться к нормальным их

значениям. В самом деле, равенство нулю синусоидальных средних следует из симметрии слагаемых. А сходимость к нормальным значениям косинусоидальных средних будет следствием предельных теорем для сумм общего вида, о которых будет говориться о последующих разделах.

Таким образом, имеются два направления доказательства предельных законов нормальной сходимости: степенное и гармоническое. В предверии нормального закона, когда либо n конечно (допредельный случай), либо условия на слагаемые недостаточны для нормальной сходимости, оба направления не подменяют, а дополняют друг друга; каждое из них дает свои грани допредельной модели, каждое по-своему характеризует приближение к нормальной с. в.

Нами сформулирован простейший вариант предельного закона, демонстрирующий основную идею, генеральную мысль. При этом требования к слагаемым предъявлялись очень жесткие: все имеют нулевые средние и нулевые нечетные моменты, а также точные одинаковые дисперсии, что само собой подразумевает стационарность этих параметров и необходимую для этого статистическую устойчивость выборки, а в конечном счете — абсолютное знание.

Допустим хоть на миг, что последовательность «чуть-чуть» статистически неустойчива. Это вынуждает рассматривать вместо точных средних интервальные $\underline{M}X_i$, $\overline{M}X_i$. Причем как бы ни были они близки к нулю, скажем, $[\underline{M}X_i,\ \overline{M}X_i]=[-\varepsilon,\ +\varepsilon]$, $\varepsilon>0$, т. е. сколь бы малым мы не взяли ε, все равно границы интервала средних нормированной суммы, равные

$$\underline{M}\left(\sum X_i/\sqrt{n}\right)=-\sqrt{n}\varepsilon,\ \overline{M}\left(\sum X_i/\sqrt{n}\right)=+\sqrt{n}\varepsilon,$$

при $n\to\infty$ будут неограниченно «разбегаться» в разные стороны, делая бессмысленной нормальную сходимость. Это есть демонстрация крайней критичности классических предельных результатов к вариациям условий.

Мы избежим упомянутой критичности, если не будем исключать сходимости законов сумм к другим, более широким, чем нормальная, моделям. Это и будет предельная проблема в ее расширенном понимании. В следующем параграфе мы и начнем постепенное продвижение по ней шаг за шагом от самых слабых допущений на слагаемые, имея в виду неустойчивость и неоднородность последовательности, к более сильным, приходя в конечном счете к классическим условиям, при которых имеет место нормальная сходимость.

Дополнение. Допредельная теорема для однородной последовательности. *Пусть* $X_1, \ldots, X_n$, $n\geqslant 3$, *независимы, имеют нулевые средние* $MX_i=0$ *и конечную дисперсию* $\overline{M}X^2{}_i=\overline{\sigma^2}$. *Тогда их нормированная сумма* $s_n=\sum X_i/\sqrt{n}$ *включается в с. в.* Y *(в смысле* $\mathcal{M}^{s_n}\subset\mathcal{M}^Y$*), задаваемую первичными средними* $MY=0$, $\overline{M}Y^2=\overline{\sigma^2}$ *и*

$$\underline{M}\cos uY=\exp\left\{-\frac{\overline{\sigma^2}u^2}{2}-\delta_n\left(\frac{\overline{\sigma^2}u^2}{2}\right)\right\}\cos\frac{c_*\overline{\sigma^2}u^2}{1-\overline{\sigma^2}u^2/(2n)}$$

при и таких, что в правой части равенства аргумент косинуса меньше $\pi/2$, *а* $c_*=0{,}3184$.

Доказательство. Во-первых, в силу нулевых средних слагаемых таким же будет $Ms_n=0$, а вследствие (3.12) $\overline{M}s^2{}_n=\overline{\sigma^2}$. Они дают средние степенных признаков нормированной суммы в направлении линейной и квадратичной функций. Для гармонических «направлений» неравенства (3.13) и (3.14) следующего раздела ведут к таким двум неравенствам: $F(\underline{\nu},\ \Lambda)=1-\underline{\nu}^2-\Lambda^2\leqslant 1-\underline{\nu}^2\leqslant\overline{\sigma^2}u^2/n$, $n\Lambda/\underline{\nu}\leqslant c_*\overline{\sigma}u^2/[1-\overline{\sigma^2}u^2/(2n)]$ — и их подстановка в (III°) утверждения 3.10 доказывает теорему.

3.4. ПРЕДЕЛЬНЫЕ МОДЕЛИ СУММ ОБЩЕГО ВИДА

Центральные допредельные неравенства. Рассмотрим суммы общего вида $\sum X_{in}$, понимая X_{in}, $i=1,\ \ldots,\ n$, как последовательности серий независимых с. в.

Видоизменим неравенства утверждения 3.10 применительно к сумме общего вида, как и ранее обозначая $F(\nu,\ \Lambda)=1-\nu^2-\Lambda^2$, $\underline{\nu}_{in}=\underline{M}\cos uX_{in}$, $|\overline{\nu}_{in}|=|\overline{M}|\cos uX_{in}$, $\Lambda_{in}=|\overline{M}|\sin uX_{in}$, переменную u для краткости опуская, где можно.

Теорема 3.12. Допредельные неравенства для гармонических средних. *Пусть* X_{in}, $i=1,\ \ldots,\ n$, — *последовательность независимых с. в.. Тогда для всех* φ *и* u *верно неравенство*

$$\text{I. } |\overline{M}|\cos\left(u\sum_{i=1}^{n}X_{in}+\varphi\right)\leqslant$$

$$\leqslant\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}F\left(|\overline{\nu}_{in}|,\ \Lambda_{in}\right)\right\},\ \forall\varphi,u;$$

а при $\min\limits_i \underline{\nu}_{in}(u)>0$ *и* $\Psi_\Sigma(u)=\sum\limits_{i=1}^{n}\dfrac{\Lambda_{in}}{\underline{\nu}_{in}}\leqslant\dfrac{\pi}{2}$ *верны неравенства*

$$\text{II. } \overline{M}\cos\left(u\sum_{i=1}^{n}X_{in}\right)\leqslant\exp\left\{-\sum_{i=1}^{n}(1-\overline{\nu}_{in})\right\};$$

$$\text{III. } \underline{M}\cos\left(u\sum_{i=1}^{n}X_{in}\right)\geqslant\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}F\left(\underline{\nu}_{in},\right.\right.$$

$$\left.\left.\Lambda_{in}\right)\left[-\frac{\ln(1-F_n)}{F_n}\right]\right\}\cos\Psi_\Sigma(u),$$

где $F_n=\max\limits_{i=1,\ \ldots,\ n}F\left(\underline{\nu}_{in},\ \Lambda_{in}\right)$;

$$\text{IV. } |\overline{M}|\sin\left(u\sum_{i=1}^{n}X_{in}\right)\leqslant$$

$$\leqslant\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}F\left(\overline{\nu}_{in},\ \Lambda_{in}\right)\right\}\sin\left(\sum_{i=1}^{n}\Lambda_{in}/\overline{\nu}_{in}\right).$$

Доказательство. Следует из неравенства (I) теоремы 3.9, если произведение $\prod_{i=1}^{n}(\overline{\nu}^2_{in}+\Lambda^2_{in})^{1/2}$, которое в нашем случае будет стоять в правой части этого неравенства, заменить на

$$\exp\left\{\frac{1}{2}\sum_{i=1}^{n}\ln\left[1-F\left(|\overline{\nu}_{in}|,\quad \Lambda_{in}\right)\right]\right\}$$

и воспользоваться неравенством $\ln(1+z)\leqslant z$. Также вытекают II из (II) утверждения 3.9, а IV — из (IV); в последнем случае арктангенс заменяется на аргумент, так как $\operatorname{arctg}|x|\leqslant|x|$, а $\sin\psi$ при $|\psi|\leqslant\pi/2$ есть неубывающая функция аргумента. Наконец, III получается из (III) с помощью предыдущего неравенства для арктангенса и следующего нетрудно проверяемого неравенства:

$$\ln(1-z)\geqslant z\frac{\ln(1-\varepsilon)}{\varepsilon}\quad\text{при}\quad 0\leqslant z\leqslant\varepsilon,\quad\forall\varepsilon>0,$$

с учетом того, что косинус на первом полупериоде есть убывающая функция аргумента. Доказательство закончено.

Первая ослабленная предельная теорема. Разные предположения относительно слагаемых X_{in} приводят к оценкам в виде неравенств для их гармонических или степенных средних и к «срабатыванию» тех или иных допредельных неравенств предыдущего раздела, правые части которых при устремлении $n\to\infty$ и дадут предельный закон нормированных сумм.

Теорема 3.13. *Пусть X_{in}, $i=1,\ \ldots,\ n$, независимы в каждой серии и выполняются три условия*

А°. $\lim\limits_{n\to\infty}\max\limits_{i=1,\ \ldots,\ n}\overline{M}X^2_{in}=0$;

Б°. $\lim\limits_{n\to\infty}\sum\limits_{i=1}^{n}\overline{M}X^2_{in}=\overline{\sigma}^2$;

В°. $\lim\limits_{n\to\infty}\sum\limits_{i=1}^{n}|\overline{M}|\,X_{in}=m$.

Тогда при $n\to\infty$ их сумма $\sum X_{in}$ ИМ-сходится в с.в. ξ, определенную первичными средними:

$$|\overline{M}|\,\xi=m,\ \overline{M}\,\xi^2=m^2+\overline{\sigma}^2,$$

$$\underset{\sim}{M}\cos u\,\xi=\exp\left(-u^2\overline{\sigma}^2/2\right)\cos\Psi_{\Sigma}(u),$$

где $\Psi_{\Sigma}(u)=|u|m+c_*u^2\overline{\sigma}^2$, $c_*=0{,}3184$, а переменная u пробегает значения, заключенные неравенством $\Psi_{\Sigma}(u)\leqslant\pi/2$.

Доказательство. Степенные средние $|\overline{M}|\xi$ и $\overline{M}\xi^2$ получаются легко из соответствующих допредельных неравенств, доказанных выше. Обратимся к косинусоидальному направлению, отправляясь от основополагающей теоремы 3.12. Из неравенства $\cos y\geqslant 1-y^2/2$, согласно которому косинусоида мажорирует располагающуюся под ней параболу, подстановкой $y=uX_{in}$ и взятием среднего получаем

$$\underline{\nu}_{in}=\underline{M}\cos(uX_{in})\geqslant 1-u^2\overline{M}X^2_{in}/2. \tag{3.13}$$

Из него следует: $F(\underline{\nu}_{in},\ \Lambda_{in})=1-\underline{\nu}^2_{in}-\Lambda^2_{in}\leqslant 1-\underline{\nu}^2_{in}\leqslant u^2\overline{M}X^2_{in}$ и $F_n\leqslant u^2\max\limits_i\overline{M}X^2_{in}$.

Из другого неравенства $|\sin y-y|\leqslant c_*y^2$, где c_* 0,3184 и равенство достигается при $y=3{,}124$ (показывается проверкой), подстановкой $y=uX_{in}$ и взятием среднего, находим

$$\Lambda_{in}\leqslant|u|\,|\overline{M}|\,X_{in}+c_*u^2\overline{M}X^2_{in}. \tag{3.14}$$

Используя оба полученных неравенства, выводим

$$\frac{\Lambda_{in}}{\underline{\nu}_{in}}\leqslant\frac{|u|\cdot|\overline{M}|\,X_{in}+c_*u^2\overline{M}X^2_{in}}{1-u^2\overline{M}X^2_{in}/2}.$$

Подстановка найденных неравенств в формулу III теоремы 3.12 с учетом $F_n\to 0$ (на основании А°) и пренебрежение членами второго порядка малости доказывает результат.

Для однородных внутри серий с.в. условия теоремы будет выполнены, если $|\overline{M}|X_{in}=m/n$ и $\overline{M}X^2_{in}=\overline{\sigma}^2/n$. Как видно, если $X_{in}=X_i/\sqrt{n}$, то условия теоремы достижимы лишь при $MX_i=0$ (иначе $m=\infty$).

Следствие 1. *Пусть $MX_i=0$, $\overline{M}X^2_i\leqslant\overline{\sigma}^2$. Тогда нормированная сумма $\sum X_i/\sqrt{n}$ при $n\to\infty$ ИМ-сходится в с.в. ζ, определенную первичными средними: $M\zeta=0$, $\overline{M}\zeta^2=\overline{\sigma}^2$, $\underset{\sim}{M}\cos u\zeta=\exp(-u^2\overline{\sigma}/2)\cos(0{,}3184u^2\overline{\sigma}^2)$ при и таких, что аргумент косинуса меньше $\pi/2$.*

Этот результат, кстати, может быть получен и из теоремы дополнения предыдущего параграфа.

Если усилить формировку, взяв $X_{in}=X_i/n$, то при $\overline{M}X^2_i<\infty$ будет $\overline{\sigma}^2=0$ и тогда получим:

Следствие 2. *Пусть $|\overline{M}|X_i=m$, $\overline{M}X^2_i<\infty$. Тогда среднее арифметическое $\sum X_i/n$ этих с.в. сходится в с.в. ξ, определенную первичными средними: $\overline{M}\xi^2=m^2$, $\underset{\sim}{M}\cos u\xi=\cos|u|m$ при $|u|m\leqslant\pi/2$.*

Отметим, что среднее $|\overline{M}|\xi=m$, вроде бы обязанное «перекочевать» из теоремы 3.13 в ее следствие, на самом деле поглощается средним $\overline{M}\xi^2=m^2$ благодаря неравенству $|\overline{M}|X\leqslant\sqrt{\overline{M}X^2}$ (следующему из неравенства Гельдера при $Y\equiv 1$).

Условия теоремы 3.13 так слабы, что позволили «сработать» лишь одному неравенству III допредельной теоремы 3.12. Пойдем дальше, вовлекая другие неравенства.

Вторая ослабленная предельная теорема. Определим те условия, при которых срабатывает неравенство I теоремы 3.12. Это не-

равенство инвариантно к сдвигам как суммы $\sum X_{in}+c$ на постоянную c, так и отдельных слагаемых, поэтому следует ожидать, что в условиях не требуется ограниченности суммы средних $\sum|\bar{M}|X_{in}$, но возникают дополнительные требования.

Теорема 3.14. *Пусть X_{in} независимы и пусть выполняются условия*: **а°**. $\lim\limits_{n\to\infty}\max\limits_{i=1,\ldots,n}\bar{M}X_{in}^2=0$;

б°. $\lim\limits_{n\to\infty}\sum(\bar{M}X_{in}^2)^{3/2}=0$;

в°. $\lim\limits_{n\to\infty}\sum\underline{M}X_{in}^2\{|X_{in}|\leqslant\delta\}=\underline{\sigma}^2,\ \forall\delta>0.$

Тогда

$$|\bar{M}|\cos\left(u\sum X_{in}+\varphi\right)\leqslant\exp\{-(\underline{\sigma}^2-a^2)u^2/2\},\ \forall u,\varphi,$$

где $a^2=\lim\limits_{n\to\infty}\sum|\bar{M}|^2X_{in}$ (обозначено $|\bar{M}|^2X=(|\bar{M}|X)^2$).

Доказательство. Во-первых, из (3.13) видно, что неравенство $\min\limits_i\nu_{in}(u)>0$ будет верно для тех u, при которых $u^2\max\limits_i\bar{M}X^2_{in}\leqslant 2$, т. е. $u^2\leqslant 2/\max\limits_i\bar{M}X^2_{in}$, а так как знаменатель первой части по условию а° стремится к 0 при $n\to\infty$, то в пределе оно будет верно для всех u, и следовательно, для всех u верно равенство $|\bar{\nu}_{in}|(u)=\bar{\nu}_{in}(u)$.

Из неравенства $\cos y\leqslant 1-(1-\varepsilon^2/12)y^2\{|y|\leqslant\varepsilon\}/2$, справедливого при любом $\varepsilon>0$, подставляя $y=uX_{in}$, получаем

$$1-\bar{\nu}_{in}(u)\geqslant\frac{u^2}{2}\left(1-\frac{\varepsilon^2}{12}\right)\underline{M}X_{in}^2\left\{|X_{in}|\leqslant\frac{\varepsilon}{u}\right\}. \tag{3.15}$$

Теперь подстановка полученного неравенства для $1-\bar{\nu}_{in}(u)$, а также неравенства (3.14) для Λ_{in} в правую часть I теоремы 3.12 и замена $\delta=\varepsilon/u$ с использованием $|\bar{M}|X\leqslant\sqrt{\bar{M}X^2}$ дает после небольших упрощений

$$\sum F(|\bar{\nu}_{in}|,\ \Lambda_{in})\geqslant u^2\left(1-\frac{u^2\delta^2}{12}\right)\sum\underline{M}X_{in}^2\{|X_{in}|\leqslant\delta\}-$$

$$-\frac{u^4}{4}\sum\left(\underline{M}X_{in}^2\right)^2-u^2\sum|\bar{M}|^2X_{in}-2c_*|u|^3\sum\left(\bar{M}X_{in}^2\right)^{3/2}-$$

$$-c_*^2u^4\sum(\bar{M}X_{in}^2)^2.$$

В силу условия в° второй и два последних члена правой части последнего неравенства стремятся к 0. В результате произвольности δ имеем

$$\lim_{n\to\infty}\sum F(|\bar{\nu}_{in}|,\ \Lambda_{in})\geqslant u^2(\underline{\sigma}^2-a^2)$$

и утверждение теоремы теперь следует из I теоремы 3.12, что и требовалось.

Следствие 1. *При условиях теоремы 3.14 сумма $\sum X_{in}$ ИМ-сходится в с. в. η, определенную первичными средними вида*

$$|\bar{M}|\cos(u\eta+\varphi)=\exp\{-(\underline{\sigma}^2-a^2)u^2/2\},\ \forall u,\varphi.$$

Следствие 2. *Пусть X_i, $i=1, 2, \ldots$, независимы, пусть $|\bar{M}|X_i\leqslant m$ и $\bar{M}X_i^2=\bar{\sigma}^2<\infty$, $\underline{M}X_i^2=\underline{\sigma}^2$. Тогда при $n\to\infty$ нормированная сумма $\sum X_i/\sqrt{n}$ ИМ-сходится в с. в. η, определенную в следствии 1 при $a=m$.*

В самом деле, для $X_{in}=X_i/\sqrt{n}$, как нетрудно убедиться, выполняются посылки **а°** и **б°** теоремы 3.14, а посылка **в°** следует из определения среднего для квадратичных (неограниченных) признаков как предела: $\underline{\sigma}^2=\lim\limits_{H\to\infty}\underline{M}X_i^2\{|X_i|\leqslant H\}$ при $H=\delta\sqrt{n}$.

Последнее следствие наиболее полно раскрывает смысл второй предельной теоремы 3.14. Во-первых, не требуется, в общем, условия ограниченности элементов последовательности или аналога этого условия — непременного спутника классических предельных результатов. Кстати заметим, что среднеквадратическая ограниченность мощности $\bar{M}X_i^2=\bar{\sigma}^2$ — не эквивалент ограниченности самих с. в., пример тому — нормальная с. в. И во-вторых, нетривиальные оценки гармонических характеристик сумм могут быть получены и в том случае, когда средние с. в. являются интервальными, что имеет место для статистически неустойчивых последовательностей и интерпретируется как совершенно неконтролируемые и не связанные друг с другом скачки средних MX_i внутри интервала $-m$, m.

Нетривиальными результатами следствия будут лишь при $\underline{\sigma}>m$ (условие отрицательности показателя экспоненты в теореме 3.14), а именно, грубо говоря, когда случайный разброс слагаемых, т. е. их независимые колебания превышают неконтролируемые (неустойчивые) флуктуации средних. Собственно, в этом и в отсутствии ограничений на суммарное верхнее среднее и дисперсию состоит основное отличие второй предельной теоремы от первой.

Замечания. 1. Условия теоремы 3.14 не исключают неограниченности как абсолютного среднего $|\bar{M}|\sum X_{in}$ суммы, так и ее «мощности» $\bar{M}(\sum X_{in})^2$.

2. Условия **а°** и **б°** теоремы будут выполнены, если об X_{in} известно только, что они независимы и ограничены числами ε_n, т. е. $P(|X_{in}|<\varepsilon_n)=1$, причем $\varepsilon_n\to 0$ при $n\to\infty$.

3. Если считать фиксированными $\lim\underline{M}\sum X_{in}$, $\lim\bar{M}\sum X_{in}$, $\lim\underline{M}(\sum X_{in})^2$, $\lim\bar{M}(\sum X_{in})^2$ и т. д., то $\sum X_{in}$ будет ИМ-сходиться в с. в. η_1, определенную помимо гармонических первичных средних теоремы 3.14 еще и первичными значениями $\underline{M}\eta_1$, $\bar{M}\eta_1$, $\underline{M}\eta_1^2$, $\bar{M}\eta_1^2$ и т. д., соответственно равными указанным пределам.

4. При $m=0$, $\bar{M}X_i^2\leqslant\bar{\sigma}^2$, $X_{in}=X_i/\sqrt{n}$ результат совпадет со следствием 1 теоремы 3.13.

Исследуем с. в. η, определенную следствием 1 теоремы 3.14. Если $f(y)=\sum\limits_u A(u)\cos(uy+\varphi_u)$ есть разложение функции $f(y)$ в ряд Фурье, то

$$|\overline{M}|\, f(\eta) \leqslant \sum_u |A(u)| \exp\{-\underline{\sigma}^2 u^2/2\}$$

(при дискретах u, плотно заполняющих числовую ось, сумма заменяется на интеграл). Правая часть последнего неравенства дает оценки средних для произвольных признаков с. в. η.

Третья ослабленная предельная теорема. Посылки первой и второй предельных теорем во многом не перекрываются: нельзя сказать, какие из них более сильные, а какие более слабые. Объединение их ведет к следующей теореме (используются прежние обозначения).

Теорема 3.15. *Пусть выполняются условия А°, Б° и В° первой предельной теоремы 3.13 и условие* **в°** *второй — 3.14. Тогда сумма* $\sum X_{in}$ *ИМ-сходится в с. в. Y, определенную первичными средними на степенных признаках*

1. $|\overline{M}|\, Y = m$, 2. $\overline{M} Y^2 = m^2 + \overline{\sigma}^2$;

и на гармонических

I. $|\widetilde{M}| \cos(uY + \varphi) = \exp\{-(\underline{\sigma}^2 - a^2) u^2/2\}$, $\forall u, \varphi$;

II. $\widetilde{M} \cos uY = \exp(-\underline{\sigma}^2 u^2/2)$;

III. $\underset{\sim}{M} \cos uY = \exp\left(-\overline{\sigma}^2 u^2/2\right) \cos \Psi_\Sigma(u)$;

IV. $|\widetilde{M}| \sin uY = \exp\{-(\underline{\sigma}^2 - a^2) u^2/2\} \sin \Psi_\Sigma(u)$;

где последние три соотношения даются при u, заключенном неравенством: $\Psi_\Sigma(u) = |u| m + c_* u^2 \overline{\sigma}^2 \leqslant \pi/2$.

Доказательство. Требуется установить лишь II и IV, так как все остальные следуют из предыдущих двух предельных теорем 3.13 и 3.14 (при этом условие **б°** в 3.14 есть следствие А° и Б° теоремы 3.13, поэтому выполняются все посылки двух теорем). Доказательство соотношения II получается вставлением неравенства (3.15) в неравенство II теоремы 3.12 и учета условия **в°**. Доказательство же IV в первом множителе правой части есть повторение доказательства I, а во втором — той части III, которая относилась к аргументу под знаком косинуса, что и требовалось.

Заметим, что не все фигурирующие в теореме 3.15 средние являются согласованными между собой, и их согласование — весьма трудоемкое занятие. Хотя этого в настоящий момент и не требуется.

Центральная теорема нормальной сходимости. Ключевыми для определения сходимости сумм к нормальной модели являются теорема 3.1 характеризации нормальной с. в. и теорема 3.2 ИМ-сходимости. Следуя им, нужно доказать, что: 1) правые части соотношений II и III теоремы 3.15 сходятся к одной и той же функции $\exp(-\sigma^2 u^2/2)$ (для этого должно быть $\sigma^2 = \overline{\sigma}^2 = \underline{\sigma}^2$); 2) правая часть IV должна равняться 0 при всех u. Еще одно требование: $\Phi(H/\sigma) \leqslant \bar{P}(\sum X_{in} > H) \to 0$ при $H \to \infty$ автоматически верно на основании неравенства Чебышева и условия Б°.

Теорема 3.16. *Пусть* X_{in} *независимы, имеют нулевые средние* $MX_{in} = 0$, *точные дисперсии* $MX^2_{in} = \sigma^2_{in}$, *причем* $\sum \sigma^2_{in} = \sigma^2 < \infty$; *пусть дисперсии убывают* $\lim\limits_{n\to\infty} \max\limits_{i=1,\ldots,n} \sigma^2_{in} = 0$, *и пусть выполняется условие Линдеберга — Феллера (ЛФ)* [1, с. 397]

$$\text{ЛФ}: \lim_{n\to\infty} \sum \overline{M} X^2_{in} \{|X_{in}| > \delta\} = 0, \ \forall \delta > 0.$$

Тогда $\sum X_{in}$ *ИМ-сходится в нормальную с. в.* $\mathcal{N}_{0,\sigma}$.

Доказательство. Проводится на основании теоремы 3.15. Во-первых, покажем, что $\Psi_\Sigma(u) \equiv 0$. В самом деле, из двух неравенств, верных при любом ε от 0 до 1:

$$|y - \sin y| \leqslant y^2 \frac{\varepsilon - \sin\varepsilon}{\varepsilon^2} \text{ при } |y| \leqslant \varepsilon,$$

$$|y - \sin y| \leqslant c_* y^2, \ \forall y,$$

выводим

$$|y - \sin y| \leqslant y^2 \frac{\varepsilon - \sin\varepsilon}{\varepsilon^2} \{|y| \leqslant \varepsilon\} + c_* y^2 \{|y| > \varepsilon\}.$$

Раскрывая соответствующим образом модуль, подставляя $y = uX_{in}$ и беря среднее, получаем с учетом $MX_{in} = 0$:

$$\Lambda_{in} = |\overline{M}| \sin u X_{in} \leqslant u^2 \left[\frac{\varepsilon - \sin\varepsilon}{\varepsilon^2} \overline{M} X^2_{in} \left\{ |X^2_{in}| \leqslant \frac{\varepsilon}{u} \right\} + \right.$$

$$\left. + c_* \overline{M} X^2_{in} \left\{ |X_{in}| > \frac{\varepsilon}{u} \right\} \right].$$

Теперь, принимая во внимание, что $\max\limits_i \nu^2_{in} \to 1$ (вытекает из (3.13) и условия убывания дисперсии), имеем для $\Psi_\Sigma(u) = \overline{\sum \Lambda_{in}/\nu_{in}}$ в пределе при $n \to \infty$:

$$\Psi_\Sigma(u) \leqslant u^2 \left[\frac{\varepsilon - \sin\varepsilon}{\varepsilon^2} \sigma^2 + c_* \lim_{n\to\infty} \sum \overline{M} X^2_{in} \left\{ |X_{in}| > \frac{\varepsilon}{u} \right\} \right], \ \forall \varepsilon > 0.$$

При любом $\varepsilon > 0$ в силу условия ЛФ последняя сумма стремится к 0. А первое слагаемое правой части стремится к 0 при $\varepsilon \to 0$, поэтому и вся правая часть будет равна 0.

С учетом доказанного из неравенств II, III и IV теоремы 3.15 получаем нормальную сходимость гармонических средних, что и требовалось.

Следствие. *Пусть* X_i, $i = 1, \ldots, n$, *независимы, имеют нулевые средние* $MX_i = 0$ *и точные дисперсии* $MX^2_i = \sigma^2$ (понимаемые как пределы (1.4)) *и пусть выполняется условие* **У**1:

У1. $\lim\limits_{H\to\infty} H^2 \bar{P}(|X_i| > H) = 0$ *равномерно по* i.

Тогда нормированная сумма $\sum X_i/\sqrt{n}$ *при* $n \to \infty$ *ИМ-сходится в* $\mathcal{N}_{0,\sigma}$.

В самом деле, среднее неограниченного признака определяется как предел среднего от усеченного варианта этого признака, откуда для X^2 имеем:

$$\sigma^2 = \underline{M}X_i^2 = \lim_{H\to\infty} \underline{M}\left(X_i^2\right)^{(-H,\,H)}.$$

Далее, используя условие **У1**, получаем

$$\overline{M}X_i^2\{|X_i|>H\} = \overline{M}\left[X_i^2 - X_i^2\{|X_i|\leqslant H\}\right] \leqslant \overline{M}X_i^2 - \underline{M}X_i^2\{|X_i|\leqslant H\} \leqslant$$

$$\leqslant \sigma^2 - \underline{M}\left[\left(X_i^2\right)^{(-H,\,H)} - H^2\{|X_i|>H\}\right] \leqslant$$

$$\leqslant \sigma^2 - \underline{M}\left(X_i^2\right)^{(-H,\,H)} + H^2\overline{P}(|X_i|>H)$$

и при устремлении H к ∞ правая часть стремится к 0. А тогда выполняется условие ЛФ:

$$\sum \overline{M}(X_i/\sqrt{n})^2\{|X_i|/\sqrt{n}>\delta\} \leqslant \max_i \overline{M}X_i^2\{|X_i|>\delta\sqrt{n}\} \xrightarrow[n\to\infty]{} 0$$

и нормальная сходимость имеет место согласно теореме 3.16.

Смысл условия **У1** в том, что вероятность превышений при увеличении уровня H должна уменьшаться быстрее значения $1/H^2$, предписанного неравенством Чебышева, что, конечно же, выполняется, если X_i ограничены одним числом.

В следствии требуется обязательная стационарность параметров $m=0$ и σ. К случаю, когда это требование не выполняется, сейчас и перейдем.

Интервальная нормальная сходимость. Основное препятствие в применениях теоремы нормальной сходимости состоит в требовании точного нулевого среднего слагаемых и точной дисперсии. Постараемся отказаться от этого требования, считая эти параметры неизвестными и меняющимися. Для этого пусть с. в. X_{in} раскрываются через некоторые «стандартные» независимые с. в. ξ_i с нулевыми средними и единичными дисперсиями

$$X_{in} = \sigma_i\xi_i/\sqrt{n} + m_i/n, \quad M\xi_i = 0, \quad M\xi_i^2 = 1, \quad i = 1, \ldots, n.$$

Здесь ξ_i считаются свободными от параметров m_i, σ_i, тогда и X_{in} будут независимыми. Делением на $\sqrt{n}$ и n уже учтена потребная для нормальной сходимости скорость убывания среднего и дисперсии X_{in} при $n\to\infty$.

Использованное нами представление называется *аддитивно-мультипликативным*. Хотя, в общем, оно является весьма специфичным, но для предельных теорем оно как раз оказывается универсальным. Дело в том, что для нормальной сходимости основным является выполнение условия ЛФ, базирующегося на следующем наборе признаков $X^2_{in}\{|X_{in}|>\delta\}$, $\forall\delta>0$. А этот-то набор признаков в асимптотике с учетом малости m_i/n пересчитывается в такого же типа набор, но уже с. в. $\xi_i : \xi^2_i\{|\xi_i|>\delta\sqrt{n}/\sigma\}$. Поэтому, считая для ξ_i первичными именно средние от этих признаков, дополненные $M\xi_i=0$, $M\xi^2_i=1$, получаем расширение: $X_{in} \subseteq \sigma_i\xi_i/\sqrt{n} + m_i/n$. Применим к правой части предельные результаты, и мы придем к теореме.

16

Теорема 3.17. *Пусть* $X_{in} \subseteq \sigma_i\xi_i/\sqrt{n} + m_i/n$, $i=1, \ldots, n$, *где* ξ_i *независимы с нулевыми средними* $M\xi_i=0$ *и единичными дисперсиями* $M\xi^2_i=1$ *и свободны от параметров* m_i *и* σ_i, *меняющихся в отрезках* $[\underline{m}_i, \overline{m}_i]$ *и* $[\underline{\sigma}_i, \overline{\sigma}_i]$. *Пусть все* ξ_i *удовлетворяют условию* **У1** *следствия, и пусть* $\overline{\sigma}_i$ *ограничены сверху:* $\overline{\sigma^2}_i \leqslant b < \infty$. *Тогда* $\sum X_{in}$ *ИМ-сходится в нормальную модель* $\mathcal{N}_{\underline{m},\,\overline{m},\,\underline{\sigma},\,\overline{\sigma}}$ *с интервальными средним* $\underline{m}$, $\overline{m}$ *и дисперсией* $\underline{\sigma^2}$, $\overline{\sigma^2}$, *равными средним арифметическим от тех же параметров слагаемых.*

Доказательство. Покажем, что при фиксированных векторах $\mathbf{m}=(m_1, \ldots, m_n)$ и $\boldsymbol{\sigma}^2=(\sigma^2_1, \ldots, \sigma^2_n)$, компоненты которых лежат в своих интервалах, условие ЛФ выполняется. Используя элементарное неравенство $(a+c)^2 \leqslant 2a^2+2c^2$, имеем

$$\lim_{n\to\infty} \sum \overline{M}_{\mathbf{m},\,\boldsymbol{\sigma}} \left(\frac{\sigma_i}{\sqrt{n}}\xi_i + \frac{m_i}{n}\right)^2 \left\{\left|\frac{\sigma_i}{\sqrt{n}}\xi_i + \frac{m_i}{n}\right| > \delta\right\} \leqslant$$

$$\leqslant \lim_{n\to\infty} n \max_i \left(2\frac{\sigma_i^2}{n}\overline{M}\xi_i^2\left\{|\xi_i| > \frac{\delta\sqrt{n}}{\sigma_i}\right\} + 2\frac{m_i^2}{n^2}\right) \leqslant$$

$$\leqslant 2b \lim_{H\to\infty} \overline{M}\xi_i^2\{|\xi_i|>H\} = 0,$$

где мы заменили $H=\delta\sqrt{n/b}$ и пренебрегли членами, стремящимися к нулю. Индексы $\mathbf{m}$ и $\boldsymbol{\sigma}$ у $\overline{M}$ соответствуют последовательно сначала $\mathbf{m}$, а затем $\boldsymbol{\sigma}$-сечению модели $\mathcal{M}^{\mathbf{X}}$ вектора $\mathbf{X}=(X_1, X_2, \ldots, X_n)$, поэтому согласно теореме о представлении имеем $\mathcal{M}^{\mathbf{X}} \subset \bigvee_{\mathbf{m}} \bigvee_{\boldsymbol{\sigma}} M^{\mathbf{X}}_{\mathbf{m},\,\boldsymbol{\sigma}}$, и также для суммы $\mathcal{M}^{\Sigma} = \bigvee_{m} \bigvee_{\sigma} \mathcal{M}^{\Sigma}_{m,\,\boldsymbol{\sigma}}$ где $m=\Sigma m_i/n$, $\sigma^2=\Sigma\sigma^2_i/n$. Установив выше условие ЛФ, мы доказали $\lim_{n\to\infty} \mathcal{M}^{\Sigma}_{m,\,\sigma} \subset \mathcal{N}_{m,\,\sigma}$, поэтому и $\lim_{n\to\infty} \mathcal{M}^{\Sigma} \subset \mathcal{N}_{\underline{m},\,\overline{m},\,\underline{\sigma},\,\overline{\sigma}}$, что и требовалось.

В условиях теоремы ничего не надо было знать о связи элементов m_i, σ_i, $i=1, 2, \ldots$, между собой. Это могут быть стационарные параметры, тогда $m_i=m$, $\sigma_i=\sigma$, $\forall i$, что соответствует одинаковым средним и дисперсиям для X_{in} при $i=1, \ldots, n$, но не известным и принадлежащим своим интервалам. Это могут быть и полностью нестационарные параметры, когда m_i и σ_i несвязанно друг от друга и неконтролируемым образом «скачут» внутри своих интервалов. Случай, соответствующий неустойчивой выборке.

Наиболее жестким является требование теоремы 3.17, чтобы средние m_i/n слагаемых скоро стремились к 0 при росте n, причем в $\sqrt{n}$ раз быстрее, чем сходится к 0 параметр разброса $\sigma_i/\sqrt{n}$. Нетрудно убедиться, что это условие может быть ослаблено, и теорема останется верной, если абсолютные средние $|\overline{M}|X_{in}$ слагаемых убывают не со скоростью $1/n$, а гораздо медленнее, а именно, $|\overline{M}|X_{in} \sim 1/n^{\alpha}$, где α — любое число из $1/2<\alpha\leqslant 1$. Тогда при $\alpha<1$ средние арифметические $\underline{m}=\sum \underline{M}X_{in}/n$, $\overline{m}=\sum \overline{M}X_{in}/n$, одна, другая или обе вместе могут вырождаться в

бесконечность. Это не должно смущать, так как, например, при $\underline{m}=-\infty$, $\overline{m}=\infty$ предельной будет нормальная модель $\mathcal{N}_{\underline{\sigma},\overline{\sigma}}$, соответствующая полностью неизвестному среднему и рассмотренная в § 3.1, а она еще далеко не голая модель. Возможны также варианты $\underline{m}=\overline{m}=\infty$, $\underline{m}=\overline{m}=-\infty$ и любые промежуточные случаи, когда одна из границ конечна.

И все равно, несмотря на сказанное, так как неравенство $\alpha>1/2$ должно быть строгим, для нормальной сходимости необходимым является требование малости средних: средние должны стремиться к 0 быстрее, чем средние квадратические отклонения. Иначе нормальной сходимости не будет и допустимы лишь слабые предельные результаты, на связь с которыми мы и укажем.

В случае интервальной нормальной модели Y имеем:

$$\overline{M} f(Y) = \sup_m \sup_\sigma \int_{-\infty}^{\infty} f(y) \frac{1}{\sqrt{2\pi}\sigma} \exp\{-(y-m)^2/(2\sigma^2)\}\, dy$$

для всех интегрируемых (по Риману) функций, где m и σ пробегают свои интервальные значения. Обозначив $m_*=\max\{|\underline{m}|, |\overline{m}|\}$, имеем

$$|\overline{M}|\, Y = m_*,\ \overline{M}\, Y^2 = m_*^2+\overline{\sigma}^2,$$

$$\underline{M} \cos uY = \exp\left(-\overline{\sigma}^2 u^2/2\right)\cos(um_*) \text{ при } |u|\leqslant \pi/(2m_*).$$

Эти значения нужно сравнить с первой предельной теоремой 3.13, где получены почти такие же первичные признаки предельной с. в. ξ. Почти, но не совсем, так как условия теоремы 3.13 слабее: в них нет требования ЛФ и, следовательно, не будет интервального нормального закона. И как видно, нижнее среднее для косинуса там получается менее точным.

Аналогичное замечание можно сделать по отношению к другим слабым предельным теоремам. Таким образом, основное усиление теоремы интервальной нормальной сходимости в отличие от слабых предельных теорем обязано именно условию ЛФ и требованию малости средних.

Дополнения.

1. Предельная модель при ограниченности средних модулей слагаемых. Везде в предельных результатах требовалась ограниченность $\overline{M}X^2_{in}$. А что, если эти значения не даны? Теоретически они вправе быть бесконечными, если $x^2\notin\mathcal{F}^{X_i}$. Насколько предельные модели сумм будут зависеть от существования $\overline{M}X^2_{in}$? А если вместо них заданы $\overline{M}|X_{in}|$, то что будет? На эти вопросы проливают свет следующие две теоремы.

Теорема. *Пусть X_{in} независимы и при $n\to\infty$:* А) $\max\limits_i \overline{M}|X_{in}|\to 0$, Б) $\Sigma\overline{M}|X_{in}|\to a$. *Тогда при $u\leqslant\pi/(2a)$ справедливы два неравенства:*

$$\lim_{n\to\infty} \underline{M}\cos(u\Sigma X_{in}) \geqslant \exp\{-c_0|u|a\}\cos(ua);$$

$$\lim_{n\to\infty} |\overline{M}|\sin(u\Sigma X_{in}) \leqslant \sin(|u|a);$$

где $c_0=0{,}725$.

Доказательство. Из неравенств $1-\cos x\leqslant c_0|x|$ (где $c_0=0{,}725$ есть постоянная, которая находится, как решение относительно c_0 и x_0 системы уравнений: $\sin x_0=c_0$, $c_0x_0=1-\cos x_0$) и $|\sin x|\leqslant|x|$, подставляя $x=uX_{in}$ и беря среднее, получаем

$$1-\underline{\nu}_{in}(u)\leqslant c_0|u|\,\overline{M}|X_{in}|,\quad \Lambda_{in}(u)\leqslant|u|\,\overline{M}|X_{in}|.$$

Теперь на основании неравенств

$$\Psi_\Sigma(u)\leqslant\frac{|u|}{\min\limits_i \underline{\nu}_{in}(u)}\Sigma\overline{M}|X_{in}| = \frac{a|u|}{1-|u|\,c_0\max\limits_i \overline{M}|X_{in}|};$$

$$\Sigma F(\underline{\nu}_{in},\ \Lambda_{in})\leqslant 2\Sigma(1-\underline{\nu}_{in}(u))\leqslant 2c_0|u|\Sigma\overline{M}|X_{in}|;$$

$$F_n=\max_i F(\underline{\nu}_{in},\ \Lambda_{in})\leqslant c_0|u|\max_i \overline{M}|X_{in}|,$$

подставляя которые в неравенство III допредельной теоремы 3.12, получаем при переходе к пределу $n\to\infty$ первый результат теоремы. Второй результат получается из неравенства IV теоремы 3.12, если учесть в пределе $\Sigma\Lambda_{in}/\underline{\nu}_{in}\to\Sigma\Lambda_{in}\leqslant|u|\Sigma\overline{M}|X_{in}|\to|u|a$ и монотонность синуса при $|u|a\leqslant\pi/2$, что и требовалось.

Следствие. *Пусть X_i независимы и $\overline{M}|X_i|=a$. Тогда $\Sigma X_i/n$ при $n\to\infty$ ИМ-сходится в с. в. Z, определенную первичными средними:* $|M|Z=a$, $\underline{M}\exp(uZ)=\exp(-|u|a)$, $\underline{M}\cos uZ=\exp(-c_0|u|a)\cos ua$, $|\overline{M}|\sin uZ=\sin|u|a$ *при* $|u|\leqslant\pi/(2a)$.

Этот результат должен рассматриваться вместе с теоремой 3.7 о сходимости среднего арифметического. Основная особенность в том, что при столь слабых условиях, как ограниченность средних абсолютных значений слагаемых, для получения сходящейся суммы нормировать X_i нужно не множителем $1/\sqrt{n}$, как в предельных теоремах, а множителем $1/n$, и при этом сходимость будет не к постоянному числу, как в законе больших чисел, а ИМ-сходимость в с. в. Z.

2. Слабейший предельный результат. Пусть не только $\overline{M}X^2_{in}$, но и $\overline{M}|X_{in}|$ не являются ограниченными. Оказывается, и в этом случае предельная модель все еще не будет тривиально голой.

Теорема. *Пусть X_1, X_2, ..., есть последовательность независимых с. в. и пусть $\underline{M}X_i=\underline{m}$, $\overline{M}X_i=\overline{m}$. Тогда при $n\to\infty$ среднее арифметическое $\Sigma X_i/n$ ИМ-сходится в с. в. ζ (т. е. $\lim\mathcal{M}\subset\mathcal{M}^\zeta$), определенную первичными средними:*

$$\underline{M}\,\zeta=\underline{m},\quad \overline{M}\,\zeta=\overline{m},$$

$$\underline{M}\exp(u\zeta)=\exp(u\underline{m}),\quad \underline{M}\exp(-u\zeta)=\exp(-u\overline{m}),\quad \forall u\geqslant 0.$$

Доказательство. Из неравенства $\exp x\geqslant 1+x$ получаем $\underline{M}\exp(uX)\geqslant 1+\underline{M}uX$. Далее на основании независимости X_i имеем

$$\underline{M}\exp\Big(u\sum_i X_i/n\Big)=\underline{M}\prod_i\exp(uX_i/n)\geqslant\prod_i\underline{M}\exp(uX_i/n)\geqslant(1+\underline{M}uX_i/n)^n.$$

Переходя к пределу $n\to\infty$ и используя классическую формулу Эйлера, получаем

$$\lim_{n\to\infty}\underline{M}\exp\Big(u\sum X_i/n\Big)\geqslant\exp(\underline{M}uX_i)\geqslant\begin{cases}\exp(u\underline{m}), & u>0,\\ \exp(u\overline{m}), & u<0,\end{cases}$$

что и требовалось.

3. На прямое разложение моментов сумм опирается следующее утверждение. *Пусть* X_{in} *независимы, симметричны, ограничены и пусть существуют* ε *и* H *такие, что* $\varepsilon \leqslant n\overline{M}X^2_{in} \leqslant H$, $\forall in$. Тогда

$$\lim_{n\to\infty} \overline{M}\,(\Sigma X_{in})^{2k} \leqslant \lim_{n\to\infty} \left(\Sigma \overline{M} X^2_{in}\right)^k (2k)!/(k!2^k).$$

3.5. ЗАКЛЮЧЕНИЕ

Если в предыдущих двух главах рассматривались общие пространства исходов, то здесь и далее — числовые пространства. Результатами случайных явлений будут случайные величины (с. в.), последовательности, затем процессы. Исходы теперь уже будут связываться не только множественными отношениями, но и числовыми: их можно упорядочивать, складывать между собой, умножать на скаляр, преобразовывать по правилам действия с числами, векторами, функциями. Такие возможности реализуются как в новых способах задания с. в. и последовательностей (см. § 3.1), так и новых формах представлений.

Случайные величины (и последовательности) в общей конструкции задаются средними признаков, а признаками являются всевозможные преобразования на числовой прямой — функции одной (многих) переменных. В том числе те неограниченные, которые мажорируются первичными. Вот почему не всегда существует как среднее самой с. в., ибо признаки в виде тождественного преобразования прямой не являются ограниченными, так и моменты, в частности, среднеквадратическое значение (при точном среднем — дисперсия).

Самым распространенным представителем с. в. является нормальная, в нашей конструкции задаваемая тремя эквивалентными способами, тремя разными наборами средних: 1) с помощью плотности, следовательно, вероятностями отрезков; 2) моментами; 3) гармоническими средними в виде характеристической функции. В направлении этих наборов и удобно судить о степени приближения к нормальной с. в.

Нужда предельных результатов, составляющих наиболее весомую часть главы, потребовала определить понятия сходимости моделей (ИМ-сходимости), состоящей в приближении средних одной модели к другой. Главная причина введения ИМ-сходимости состоит в ее нацеленности на предельные результаты для сумм независимых с. в. В частности, с ее помощью формулируется закон больших чисел применительно к статистически неустойчивым последовательностям (теорема 3.7).

Все давно привыкли, что при сложении независимых с. в. вправе ожидать нормального предельного закона, полагая при этом выполненными известные условия Линдеберга — Феллера. А если эти условия не выполняются, что вполне естественно при их исходной жесткости, состоящей в точном знании средних слагаемых и их дисперсий, а также неограниченном росте их числа? Представьте на миг, что средние слагаемых не совсем точные, т. е. хоть да чуть-чуть, а интервальные, и тут же окажетесь в тупиковой ситуации, так как средним нормированных сумм станет разбегающийся по ширине интервал, и предел просто теряется. Потеряется в рамках классического подхода, но не интервального, способного охватить любые промежуточные случаи, причем даже для конечных сумм, т. е. допредельного случая.

Вообще допредельные и предельные результаты важны потому, что операция суммирования самопроизвольно участвует в практике рождения многих с. в. Так, погрешность изготовления детали есть результат наложения друг на друга разных факторов. Сложение лежит в корнях процедуры фильтрации и т. д. И полезно знать, что не только данные о признаках слагаемых (в частности, вероятностях, среднем с. в.) переносятся по соответствующим формулам на суммы, но и само суммирование полагает внутри себя уточнение средних по характерным направлениям, даваемое степенными и гармоническими признаками. Такие направления универсальны в силу сугубо арифметических свойств, проявляемых при подстановке на место аргумента сумм. Их средними будут границы моментов, гармонические средние, при точных распределениях вероятностей объединяемые в характеристическую функцию (для нас не нужную).

В зависимости от числа слагаемых и данных о них для сумм получаются то более, то менее широкие интервальные модели, определяемые своими средними по универсальным направлениям как степенным, так и гармоническим. Интересно проследить, как характер данных о слагаемых влияет на ширину допредельной и предельной моделей. И как в крайнем случае точных данных, удовлетворяющих классическим условиям, предельной станет нормальная модель (см. теорему 3.17).

Изложение результатов построено так, что сначала в § 3.3 рассматривается случай однородных слагаемых, позволяющий вникнуть в суть, а затем в § 3.4 переносится на неоднородные суммы общего вида, где закономерности более общие, но и более сложные. Новые допредельные и предельные утверждения позволяют в полном объеме в терминах ИМ выявить вытекающие из суммирования данные еще задолго до того, как суммы стали предельно нормальными, и даже если таковыми в пределе не смогут стать.

Глава 4.
СЛУЧАЙНЫЕ ПРОЦЕССЫ

4.1. ОПИСАНИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ

Принцип описаний. Время — неумолимый движитель, без устали бежит-бежит. В этом безостановочном беге и возникают события, названные случайными на том основании, что факт их появления или непоявления не прогнозируется абсолютно точно. Но время дает нам еще одно проявление случайности: можно достоверно знать, что событие произойдет, но не знать момента возникновения, и событие становится случайным по времени, т. е. случайным процессом. Вообще любые случайные или неслучайные события, если учесть их положение во времени, образуют процесс. А удобна ли такая абсолютизация случайных процессов?

Наша цель — построение математических моделей — обязывает не усложнять, а упрощать. Введение времени как самостоятельного параметра оправдано при следующих обстоятельствах. Во-первых, если важным представляется момент появления со-

бытия, например в радиолокации, где запаздывание отраженного импульса несет сведения о расстоянии до цели. Во-вторых, при описании физических явлений, связность и естественность хода развития которых без времени проследить немыслимо, таких как рост агрокультуры, технологические процессы, сигналы динамической системы, шумы, помехи и так далее.

Процессы в природе могут быть самыми разнообразными: дробовые и атмосферные шумы, транспортные и промышленные, импульсные и гармонические помехи, всевозможные потоки в системах массового обслуживания и др. Задачей исследователя ставится разработка как можно более экономных и простых описаний, достигаемых выявлением наиболее существенных, важных сторон, своего рода «анкетных данных» процессов с последующим облачением этих данных в «тогу» первичных средних.

Собственно, в самой уже модели за счет выбора первичных признаков заключена потенциальная возможность к упрощениям, направленным редукциям, и чем экономнее модель, меньшим числом данных она задается, тем проще процесс в нашем представлении, т. е. в том виде, как нам удобно с ним иметь дело. Это принципиальное положение теории. И переход к дискретному времени тогда естествен как одна из разновидностей редукции.

Реализации и признаки. Формально *случайным процессом* (или просто процессом) *называется система случайных величин* X_t, $t\in\mathrm{T}$, *индексированная числовым параметром* t *— текущим временем*. Здесь T — множество значений t, в частности, это отрезок $[0, T]$ числовой прямой, а в пределе — полупрямая $\mathscr{R}^+$ или вся числовая ось $\mathscr{R}$. Если t — векторный параметр, например $\mathbf{t}=(t, z_1, z_2, z_3)$ — время и три координаты пространства, то $X_{\mathbf{t}}$ называется *полем*, наше изложение распространяется и на него. Наконец, если T — дискретный набор временных отсчетов: $\mathrm{T}=\{t_1, \ldots, t_n\}$, то процесс вырождается в случайную последовательность, а при $n=1$ — в случайную величину. Переход к последовательности связывает процесс с исследованием предыдущей главы, хотя сейчас нас в основном будет интересовать непрерывное время t.

Пространством элементарных событий, соответствующим процессу X_t, $t\in\mathrm{T}$, в общем, является множество всех возможных реализаций x_t, $t\in\mathrm{T}$ (функций времени). Множество $\mathscr{X}$ реализаций, имеющее единичную вероятность $\bar{P}(\mathscr{X})=1$, называется *достоверным* для данного процесса. Все реализации, не принадлежащие $\mathscr{X}$, оказываются невозможными.

Если T есть интервал прямой (либо вся прямая), а достоверным является множество $\mathscr{X}$ непрерывных реализаций, то процесс будет непрерывным. Если это множество дифференцируемых реализаций, то и процесс будет дифференцируемым, если ограниченных (т. е. $|x_t|\leqslant a$), то — ограниченным и т. д. Таким образом, некоторое свойство всех реализаций будет достоверным свойством процесса, т. е. выполняющимся с вероятностью 1.

По нашему мнению, для реальных, физически, так сказать, осязаемых моделей каждая возможная реализация должна иметь ненулевую верхнюю вероятность $\bar{P}(X_t=x_t)>0$, $x_t\in\mathscr{X}$. Так и будет получаться, если первичных данных о процессе конечное число и они не абсолютно точны, т. е. в известном смысле размыты.

Для нужд теории, несмотря на сказанное, нельзя исключать и тот крайний вариант, когда вероятности всех отдельных реализаций нулевые, понимая его как предельный или идеальный, соответствующий неограниченному набору данных. В этом варианте достоверное множество реализаций $\mathscr{X}$ может не быть определенным однозначно (эквивалентно неоднозначности нулевого множества). В самом деле, если $\mathscr{X}_k$, $k=1, 2, \ldots$, — разные варианты $\mathscr{X}$, то их конечные пересечения обязательно будут достоверными, но никак не счетные, так как пересечение всех $\mathscr{X}_k$ не приводит к достоверному множеству, а значит, минимальное из них не существует (эквивалентно тому, что объединение счетного числа нулевых множеств не ведет к нулевому множеству).

Любое множество реализаций, включающее хотя бы один какой-нибудь вариант $\mathscr{X}$, будет достоверным. Не всегда $\mathscr{X}$ нужно стремиться сделать как можно у́же, но желательно, чтобы оно было как можно проще (даже за счет некоторого его расширения).

Прежде чем подойти к описанию процессов, напомним общую конструкцию интервальных моделей. Она остается единообразной для любых случайных объектов, будь то случайные величины, последовательности, процессы, наконец, поля, и состоит из трех шагов: 1) анализируется структура признаков с взаимной их полуупорядоченностью; 2) выделяются первичные, на которых задаются первичные средние; 3) первичные средние продолжаются на все остальные признаки, образуя модель. Сложность модели будет определяться числом первичных признаков и, конечно, их структурой, а «подводными камнями» будет размерность пространства $\mathscr{X}$ и связанные с ней трудности контроля упорядочения признаков, о которых пойдет речь ниже.

Обратимся к случайным процессам. Их признаками будут всевозможные функционалы $f\{X_t\}$, ставящие в соответствие каждой реализации x_t одно число. Примерами таковых для процесса с непрерывным временем являются линейные признаки, к которым относятся, во-первых, интегралы

$$f_h\{X_t\}=\int_{\mathrm{T}} X_t\, h_t\, dt,$$

где h_t — весовая функция; во-вторых, «выхватывание» из процесса одного отсчета $f_\tau\{X_t\}=X_\tau$, соответствующего моменту τ (может быть получен из интеграла при h_t в виде дельта-функции Дирака); в-третьих, взятие первой производной $f\{X_t\}=dX_t/dt$ (если реализации процесса дифференцируемы), второй производной и т. д.

Квадратичные признаки X^2_t, X_tX_τ, $\partial^2X_tX_\tau/dtd\tau$, $f_H\{X_t\}=\int\limits_{\mathrm{T}}\int\limits_{\mathrm{T}} X_tX_\tau H_{t,\tau}dtd\tau$.

Индикаторные признаки есть индикаторные функции событий $\{X_t \geqslant a\}$, $\{a_1 < X_t < a_2\}$, состоящих в превышении или непревышении процессом уровней; сюда же относятся произведения индикаторных функций: $\{X_t \geqslant a,\ \forall t \in \mathrm{T}_1\}$, указывающие на одновременное превышение уровня a всеми значениями процесса из подмножества T_1 временной оси.

Гибридные признаки вовлекают разные классы предыдущих, как, например, линейно-индикаторный $\{\int_{\mathrm{T}} X_t h_t dt > a\}$, состоящий в фиксации превышения линейным признаком уровня a с формулировкой результата в виде 0 (нет) и 1 (да).

Гармонические признаки даются произведениями

$$\prod_i \overset{\sin}{\cos}(u_i X_{t_i})$$

и связываются с предельными теоремами предыдущей главы.

Вообще, признаков неисчислимое множество, а то, что здесь указано, лишь узкие их подклассы. Разберем упорядочение признаков как внутри подклассов, так и между ними. Для линейных признаков имеем $f_h \geqslant f_{h^*} \leftrightarrow h_t \geqslant h^*_t$, $\forall t \subset \mathrm{T}$. Для квадратичных $f_H \geqslant f_{H^*} \leftrightarrow H_{t,\tau} - H^*_{t,\tau}$ — неотрицательно определенное ядро. Линейные превращаются в квадратичные посредством возведения f_h в квадрат: $f^2_h = \int_{\mathrm{T}}\int_{\mathrm{T}} X_t X_\tau \cdot h_t h_\tau \, dt d\tau$ с $H_{t,\tau} = h_t h_\tau$, в результате чего $f^2_h \geqslant f_{H^*}$ эквивалентно неотрицательной определенности $h_t h_\tau - H^*_{t,\tau}$. Связь линейных и квадратичных с индикаторными признаками осуществляется через мажорирование индикаторных функций параболами, и наоборот. Так как $\{|x| \leqslant a\} \geqslant c_0 - c^+_1 x^2$ при $c_0 \leqslant 1$, $c^+_1 \geqslant c_0/a^2$, то сдвигом по оси абсцисс получаем неравенство:

$$\{a_1 \leqslant X_t \leqslant a_2\} \geqslant c_0 - c^+_1 \left(X_t - \frac{a_2 + a_1}{2}\right)^2$$

при $c_0 \leqslant 1$, $c^+_1 \geqslant \dfrac{4 c_0}{(a_2 - a_1)^2}$,

где в левой части стоит индикаторный признак, а в правой — вложенная в индикаторный прямоугольник парабола. Помещая параболу сверху, получаем

$$\{X_t > a_2\} \leqslant c^+_1 \left(X_t - \frac{a_2 + a_1}{2}\right)^2,\ \forall a_1 < a_2,\ c^+_1 \geqslant 4/(a_2 - a_1)^2.$$

Кстати, взяв среднее от обеих частей, придем к неравенству Чебышева (см. § 3.1).

Модель процесса. В формальном определении модель процесса есть совокупность согласованных средних $\bar{M}f$ на классе $\mathscr{F}$ функционалов-признаков, составляющих область существования верхних средних. В $\mathscr{F}$ входят, по крайней мере, все ограниченные функционалы. Но могут включаться и неограниченные, скажем, линейные и квадратичные. Это будет совершенно законно, если процесс ограничен, а если нет, то будет некоторый волюнтаризм, но оправданный энергетической конечностью реальных физических процессов, а также практической невозможностью (даже в смысле измерить) сколь угодно больших его значений. К этому вернемся в конце раздела.

По способу задания модель согласно общей нашей конструкции порождается любым непротиворечивым набором первичных средних $\bar{M}g$, $g \in \mathscr{G}$, где $\mathscr{G}$ — выделенная совокупность первичных признаков-функционалов. В частности, первичными будут вероятности, если $\mathscr{G}$ составляют индикаторные признаки.

Конечно же, по первичным средним вовсе не обязательно искать все $\bar{M}f$, а нужно удовлетвориться мыслью, что они существуют и в случае нужды вычислимы конструктивным путем, даваемым формулой согласования и продолжения (1.2).

При движении к основной цели: научиться экономно описывать первичными средними знакомые и наиболее характерные типовые черты и свойства процессов, полезно придать этим средним прикладную интерпретацию. Вот некоторые из примеров:

текущее среднее значение процесса: $\underline{m}_t = \underline{M} X_t$;
среднее интегральное значение: $\underline{M} \int X_t dt$;
текущая средняя мощность $\underline{r}_t = \underline{M} X^2_t$;
интегральная средняя мощность: $\underline{M} \int X^2_t dt$;
текущие начальные моменты: $\underline{M} X^{h}_t$;
корреляционные функции: $\underline{r}(t, \tau) = \underline{M} X_t X_\tau$;
вероятности превышений: $\underline{P}(X_t > a)$;
совместные вероятности: $\underline{P}(a_1 \leqslant X_t \leqslant a_2,\ \forall t \in \mathrm{T}_1)$.

Здесь черта над и под буквами предохраняет от повтора, означая, что соответствующее выражение верно отдельно как для нижнего, так и для верхнего средних.

В отличие от $\underline{M} X_t$ и $\overline{M} X_t$, которые равноправны, между нижним и верхним средними других признаков имеется большая разница. Нижнее среднеквадратическое значение $\underline{r}_t$ указывает на ту границу, ниже которой ни при каких условиях не может упасть среднеквадратическая мощность процесса, а $\bar{r}_t$ — выше чего она никогда не сможет подняться. То же самое относится к вероятностям. Таким образом, $\underline{r}_t$, $\underline{M}$, $\underline{P}$ служат для описания необходимых, обязательно присутствующих статистических свойств процесса, а в свою очередь, $\bar{r}_t$ и $\bar{M}$, $\bar{P}$ — возможных, не исключенных, не обязательных.

Любое из приведенных нами средних, — а ими, конечно же, не исчерпывается необозримое богатство выбора, — может быть включено в набор первичных; все зависит от того, какие среднестатистические данные о процессе доступны или могут быть из многообразия доступных разумно выделены для формирования математической модели. Например, если измерения значений процесса производятся с помощью инерционных технических средств,

выдающих на выход не значения самого процесса, а лишь интегральные его данные на отрезках $[i\Delta t,\ (i+1)\Delta t]$, то все первичные признаки будут иметь вид $g(Y_i)$, где $Y_i=\int\limits_{i\Delta t}^{(i+1)\Delta t} X_t dt$. Если принципиальными для процесса мы считаем выбросы, то целесообразно в первичные включать вероятности превышений, тогда ими и будет характеризоваться выделенная нами черта процесса. Если доступным для наблюдений является не вся, а часть T_1 временной оси, то признаками будут функционалы от X_t, $t\in T_1$, инвариантные к поведению реализаций вне T_1. О характерных признаках подробнее поговорим в следующем разделе. А сейчас продемонстрируем процедуру продолжения первичных средних на остальные признаки.

Пример 4.1. Пусть X_t имеет достоверным некоторое множество реализаций $\mathscr{X}$ и задан своими точными первичными значениями текущего среднего $MX_t=m_t$, $t\in T$. Ничего, кроме этого, о нем не известно. Формулой продолжения средних будет

$$\overline{M}f\{X_t\}=\inf\{[c_0+\Sigma c_i m_{t_i}]: f\{x_t\}\leqslant c_0+\Sigma c_i x_{t_i}\},$$

где неравенство на поиск инфимума должно выполнять при всех ограниченных реализациях x_t, а инфимум определяется выбором дискретов t_i и коэффициентов c_i конечных сумм. Из данной формулы получаем $M(c_0+\Sigma c_i X_{t_i})=c_0+\Sigma c_i MX_{t_i}$, т. е. оператор среднего, будучи точным, проносится за знак конечных сумм. По формуле продолжения могут быть найдены средние лишь от функционалов, представляемых в виде $f\{x_t\}=f_0\{x_t\}+\Sigma c_i X_{t_i}$, где функционал $f_0\{x_t\}$ равномерно ограничен на $\mathscr{X}$; они-то и образуют область существования $\mathscr{F}$ средних и для них:

$$\overline{M}f\{X_t\}=\sup_{x_t\in\mathscr{X}} f_0\{x_t\}+\Sigma c_i m_{t_i}.$$

Если теперь текущее среднее не является точным, а задается интервалами $\underline{m}_t$, $\overline{m}_t$, то все формулы остаются в силе с той лишь разницей, что $c_i m_{t_i}$ заменяются на $c_i\overline{m}_{t_i}$ при $c_i\geqslant 0$ и на $c_i\underline{m}_{t_i}$ при $c_i\leqslant 0$.

Рассмотрим упрощения на пути построения модели. Какова бы ни была исходная модель, любой процесс можно «подвести основанием» под фиксированный набор признаков $\mathscr{H}$, вычислив для этого $\overline{M}h$, $h\in\mathscr{H}$ и взяв их за первичные для новой модели. Это соответствует $\mathscr{H}$-расширению модели $\mathscr{M}$ и позволяет соответствующим выбором $\mathscr{H}$ упростить описание процесса, представить его в типовом виде.

Уже отмечалось, что вид первичного набора открывает возможность редукции самого процесса. Так, если первичными являются функционалы $g\{X_t\}=g\{\mathbf{X}\}$, зависящие лишь от отсчетов $\mathbf{X}=(X_{t_1},\ \ldots,\ X_{t_n})$ процесса в дискретные моменты, то переход от X_t к вектору $\mathbf{X}$, называемый *дискретизацией процесса*, будет преобразованием подобия (см. § 2.1). Здесь этот переход, по сути

дела, не меняет данных о модели. А поскольку к указанному виду первичных функционалов могут расширением быть приведены любые модели, то дискретизация, в общем, есть не что иное в смысле теории, как упрощающее расширение модели процесса. То же может быть сказано относительно *квантования*, состоящего в приведении значений процесса к заранее выбранным уровням.

Итак, любые преобразования процесса влекут за собой расширения его модели, кроме преобразований подобия, так как они сохраняют все известные свойства процесса.

Характерные черты процессов. Первый и прямой путь задания процессов состоит в выделении характерных статистических свойств и облачения их в форму первичных средних. Для этого может быть выделено, в общем, любое их число. Причем удачным считается наш выбор только тогда, когда без серьезного расширения тела модели меньшим оказывается число первичных значений, так как в результате будет проще модель. Поймать в чертах процесса наиболее важное, отличительное и воплотить в модель — есть искусство обладания математическим языком на базе инженерной интуиции.

Рассмотрим некоторые типовые свойства. Сведения об ограниченности процесса по абсолютной величине числом a формулируются как $\overline{P}(|X_t|>a)=0$. При t, пробегающим T, это даст не одно, а целый набор первичных значений.

Инерционность процесса в смысле невозможности быстрых его изменений описывается ограничениями на производную процесса в обычном или среднеквадратическом смысле: $\overline{P}(|dX_t/dt|>a)=0$; $\overline{M}(dX_t/dt)^2=c$. Это определит локальные свойства реализаций. Глобальный же характер их изменений отражается видом границ корреляционных функций $\underline{r}(t,\ \tau)$, $\overline{r}(t,\ \tau)$. Отметим, что при несовпадении нижней и верхней границ корреляций свойства непрерывности ими не определяются. Тогда непрерывность процесса воплощается, например, в способность средних вида $\overline{M}(X_t-X_\tau)^2$ сходиться к 0 при $\tau\to t$. Возможны другие средства описания непрерывности, например ограниченностью производной.

Глобально (и достаточно грубо) изменчивость процесса определяется *интервалом корреляции* $\tau_{кор}$ — минимальным числом таким, что $|\overline{M}|X_tX_{t+\tau}\equiv 0$ при $\tau>\tau_{кор}$. Охарактеризовать максимально допустимую интенсивность выбросов процесса выше уровня a помогают первичные значения $\overline{P}(|X_t|>a)$, тогда как гарантированная доля вероятностей этих выбросов эквивалентна ненулевой нижней границе $\underline{P}(|X_t|>a)$. Причем на основании неравенства Чебышева верхняя вероятность превышений для ее согласованности со значениями $\overline{M}X^2_t$ не может быть выше $\overline{M}X^2_t/a^2$, иначе при согласовании она должна замениться на это более точное значение. Нижняя же вероятность превышения автономна, так что если вдруг она окажется больше $\overline{M}X^2_t/a^2$, то это никак на саму нее не повлияет, зато при согласовании приведет к уточ-

нению $\underset{\sim}{M}X^2_t$, а именно, к росту до значения $\underline{M}X^2_t = a^2\underline{P}(|X_t|>a)$. Рассмотрим, как можно задать конкретный процесс.

Пример 4.2. Задание импульсной помехи. Особенность рассматриваемого процесса — наличие на оси времени хаотических импульсов разной, в общем, амплитуды, формы и продолжительности. Отсюда основная черта — это редкость ненулевых значений процесса. Поэтому первичными нужно сделать вероятности $\underline{P}(X_t\neq 0)=\underline{q}$, $\bar{P}(X_t\neq 0)=\bar{q}$. Отрезок $[\underline{q}, \bar{q}]$ указывает на вероятности присутствия в момент t импульса (ненулевого значения) процесса, а конечная его ширина — на незнание этой вероятности или же на неустойчивость процесса. При $\underline{q}=0$ не исключается, что импульсов вовсе не будет, а $\underline{q}>0$ гарантирует обязательную долю вероятности их присутствия. Значение же $\bar{q}$ задает максимально допустимую их концентрацию.

Дополнительно к $\underline{q}$, $\bar{q}$ более детальный характер превышений мог бы быть отражен вероятностями $\underline{P}(|X_t|>a)$ и $\bar{P}(|X_t|>a)$.

Дробление процесса на составляющие. Исследуем одну полезную интерпретацию процессов, вытекающую из теоремы 1.3 о представлении.

Назовем процесс *$\mathcal{G}'$-простым*, если модель его есть $\langle M\mathcal{G}'\rangle$ и задается точными на наборе $\mathcal{G}'$ функционалов первичными средними Mg, $g\in\mathcal{G}'$. Согласно теореме 1.3 о представлении модель любого процесса с первичным набором $\mathcal{G}$ признаков записывается как объединение *$\mathcal{G}'$-простых составляющих* моделей при $\mathcal{G}'$ более подробном, чем $\mathcal{G}$:

$$\langle \tilde{M}\ \rangle = \mathcal{M} = \bigvee_{\langle M\mathcal{G}'\rangle\subset\mathcal{M}} \langle M\ \rangle', \quad \mathcal{L}+\mathcal{G}'\supset\mathcal{G}.$$

Чем шире $\mathcal{G}'$, тем из более «мелких частиц» складывается модель процесса. Наиболее экономно считать $\mathcal{G}'=\mathcal{G}$.

Иногда удобнее вместо «модель» говорить «процесс», а объединение интерпретировать как семейство простых процессов. Тогда рассматриваемое представление позволяет вообразить себе X_t так, как будто бы вместо него действует любой из этих $\mathcal{G}'$-простых составляющих процессов, но неизвестно какой.

В более общем представлении модели имеем: $\mathcal{M}=\bigvee_\theta \mathcal{M}_\theta$, где θ — неизвестный (задающий) параметр.

Функциональные представления. Даются записью $X_t=\mathbf{V}_\theta\{\xi_t\}$, определяющей X_t посредством преобразования «стандартного» процесса ξ_t и неизвестного произвольного параметра θ. Оператор $\mathbf{V}_\theta$ включает те действия, какие нужно проделать с ξ_t для получения X_t, а в θ вкладывается неизвестная часть этих действий.

Для последовательности (дискретного времени t) примеры функциональной записи были рассмотрены в конце § 3.1. Они обобщаются на процессы. Одним из них является линейное представление:

$$X_t = \int h(t,\tau)\,\xi_\tau\,d\tau.$$

При подробном описании ξ_t импульсный отклик $h(t,\tau)$ фильтра может играть роль носителя априорного дефицита θ, разрушителя подробности. К примерам линейных представлений процесса относится запись в виде решения дифференциального уравнения $c_k d^kX_t/dt^k+\ldots+c_1dX_t/dt+c_0X_t+c=\xi_t$.

Для процесса на выходе канала связи с замираниями характерно мультипликативное представление: $X_t=\theta^+_t\xi_t$, где ξ_t — процесс с установленными свойствами и, главное, с точным значением $M\xi^2_t=1$, $\forall t$, а $\theta^+_t\geqslant 0$ — параметр замираний. В зависимость от того, быстрыми или медленными являются замирания, ставятся корреляционные свойства θ^+_t. Размах между максимальным и минимальным значениями θ^+_t характеризует глубину замираний, а разность $\bar{M}(\theta^+_t)^2-\underline{M}(\theta^+_t)^2=\bar{M}X^2_t-\underline{M}X^2_t$ — неточное знание мощности процесса X_t. Что касается связи процессов θ^+_t с ξ_t, то часто они независимы. Но может θ^+_t быть свободным от ξ_t, последнее оправдано физической возможностью влияния значений ξ_t на θ^+_t, когда ξ_t есть передаваемый по каналу сигнал, а X_t — то, что получилось в канале с учетом замираний. В этом варианте θ_t может зависеть от ξ_t, а точнее, подстраиваться под него непредсказуемым образом. Обратное проблематично.

Аддитивное представление: $X_t=\theta_t+\xi_t$, $M\xi_t=0$, позволяет развязать медленно меняющуюся составляющую θ_t от быстро флуктуирующей ξ_t, и плюс к этому — явно выделить в виде ξ_t ту часть процесса, о которой статистические данные имеются, от той θ_t, о которой кроме множества Θ возможных реализаций ничего не дано.

Различные аддитивные представления. Функциональные представления, в общем, ограничены по своим возможностям ввиду детерминизма их связей. Не всегда удается подобрать такие $\mathbf{V}$, Θ и ξ, чтобы получить требуемые свойства X_t, да и итог может оказаться настолько сложным, что станет ненужным. Облегчения можно достичь заменой равенства на включение: $X_t\subset\mathbf{V}_\theta\{\xi_t\}$, означающее, что ИМ правой части включает ИМ левой, т. е. правая часть воссоздает расширенную модель исключением некоторых сторон процесса X_t (желательно наименее важных). Будем это включение рассматривать применительно к аддитивному представлению.

Поставим вопрос, всякий ли процесс X_t можно записать как аддитивную смесь $m_t+\xi_t$ среднего его значения m_t и добавки ξ_t с нулевым средним $M\xi_t=0$? Оказывается, и мы сейчас это покажем, что нет, если иметь в виду равенство, и да — если включение, т. е. расширенное представление.

Рассмотрим произвольный процесс X_t, заданный моделью $\mathcal{M}$. Любое $\langle m_t=MX_t\rangle$-сечение модели

$$\mathcal{M}_{m_t} = \mathcal{M}\bigwedge_t \langle MX_\cdot = m_t\rangle$$

соответствует допущению, что среднее процесса известно точно.

Каждому m_t-сечению соответствует то же достоверное множество реализаций, что и самому процессу. Если $\mathscr{M}_{m_t}$ непусто, то реализация m_t среднего называется *собственной*. Множество

$$\mathfrak{M} = \{m_t\colon \mathscr{M}_{m_t} \neq \varnothing\}$$

называется *собственным семейством средних.*

Поскольку модель любого процесса представляется объединением ее сечений, имеем:

$$\mathscr{M} = \bigvee_{m_t \in \mathfrak{M}} \mathscr{M}_{m_t},$$

$$\overline{M} f\{X_t\} = \sup_{m_t \in \mathfrak{M}} \overline{M}_{m_t} f\{X_{t,m}\},$$

где нижняя формула — суть расшифровка объединения в верхней, а процесс $X_{t,m}$ соответствует сечению $\mathscr{M}_{m_t}$. Собственное семейство средних задает так называемые *свойства первого порядка* процесса. На основании этого семейства могут быть рассчитаны верхние (и нижние) средние конечных линейных комбинаций отсчетов процесса:

$$\overline{M} \sum c_i X_{t_i} = \sup_{m_t \in \mathfrak{M}} \sum c_i m_{t_i}.$$

Рассмотрим теперь возможность представления процесса X_t в виде суммы среднего и остатка. Если среднее m_t является точным, то собственное семейство $\mathfrak{M} = \{m_t\}$ состоит всего из одной реализации m_t и очевидной становится запись: $X_t = m_t + \xi_t$, $M\xi_t = 0$, где $\xi_t = X_t - m_t$ есть процесс, определенный через X_t значениями $\overline{M} f\{\xi_t\} = \overline{M} f\{X_t - m_t\}$, $\forall f \in \mathscr{F}$. Если X_t задан своими первичными средними $\overline{M} g\{X_t\}$, $g \in \mathscr{G}$, то первичными средними для ξ_t будут $\overline{M} g\{\xi_t + m_t\}$, $g \in \mathscr{G}$, и они получаются смещением на m_t функционалов g. Символьно это записывается $\mathscr{M}^{\xi+\mathbf{m}} = \mathscr{M}^{\mathbf{x}}$ или $\mathscr{M}^{\xi} = \mathscr{M}^{\mathbf{x}-\mathbf{m}}$, $\mathbf{m} = \{m_t,\ t \in \mathrm{T}\}$.

Пусть теперь $MX_t = m_t$ не является точным. Аналогом предыдущего будет так называемое подчиненно-аддитивное представление, согласно которому аддитивная запись верна для каждого MX_t-сечения $\mathscr{M}_{\mathbf{m}}$ исходной модели $\mathscr{M}$:

$$X_{t,\mathbf{m}} = m_t + \xi_{t,\mathbf{m}},$$
$$M_{\mathbf{m}}\, \xi_{t,\mathbf{m}} = 0,\ m_t \in \mathfrak{M}, \qquad (4.1)$$

где процесс $\xi_{t\,\mathbf{m}}$ определяется моделью $\mathscr{M}_{\mathbf{m}}^{\xi_{t,\mathbf{m}}} = \mathscr{M}_{\mathbf{m}}^{\mathbf{x}-\mathbf{m}}$ и имеет при каждом m_t нулевое среднее. В этом представлении первое слагаемое m_t при неточном его значении «вбирает» в себя статистически неустойчивую составляющую среднего процесса, а все имеющиеся статистические данные о X_t переносятся в остаток. Процесс X_t же сам складывается из семейства составляющих его $X_{t,\mathbf{m}}$ при m_t, пробегающим множество $\mathfrak{M}$.

Некоторым неудобством подчиненно-аддитивного представления является то, что остаток $\xi_{t,\mathbf{m}}$, в общем, будет зависеть от значения m_t. Чтобы освободиться от этой зависимости, заменим $\xi_{t,\mathbf{m}}$ на более широкий (в смысле модели) процесс $\xi_t \supset \xi_{t,\mathbf{m}}$, определенный объединением:

$$\mathscr{M}^{\xi} = \bigvee_{\mathbf{m} \in \mathfrak{M}} \mathscr{M}_{\mathbf{m}}^{\xi_t,\mathbf{m}}.$$

Сказанное наводит на следующее расширенное аддитивное представление произвольного процесса:

$$X_t \subseteq m_t + \xi_t,\ m_t \in \mathfrak{M},\ M\xi_t = 0, \qquad (4.2)$$

где m_t принимает произвольные значения из семейства $\mathfrak{M}$, а процесс ξ_t свободен от m_t. Символ $\subseteq$ означает, что расширение, стоящее в правой части (4.2), однозначно определено и является минимальным расширением подобного рода, включающим X_t.

Смысл перехода от (4.1) к (4.2) состоит в том, что «забываются» связи между остатком $\xi_{t,\mathbf{m}} = X_{t,\mathbf{m}} - m_t$ и средним m_t, приводя к расширению ξ_t по сравнению с $\xi_{t,\mathbf{m}}$, как это записано в (4.2). Опасно только, не станет ли это расширение чрезмерным. Например, если ξ_t в результате расширения окажется голым процессом, то включение (4.2) становится тривиальным. Такая опасность иллюстрируется примером.

Пример 4.3. Сугубо ограниченный процесс. Пусть процесс X_t определен первичными вероятностями $\tilde{P}(|X_t| \geqslant a) = 0$, $\forall t$. Это есть процесс, о котором известно только, что в любой момент он ограничен по модулю значением a. Для такого процесса $\underline{M} X_t = -a$, $\overline{M} X_t = a$ и собственное семейство $\mathfrak{M}$ средних образуют всевозможные реализации, заключенные между значениями $-a$, a. В представлении (4.2) здесь будет $\xi_t = 0$, так как нельзя из такого процесса выделить случайную составляющую с $M\xi_t = 0$.

При замене включения (4.2) на равенство, что будет, когда все $\xi_{t,\mathbf{m}}$ от m_t не зависят, т. е. $\xi_{t,\mathbf{m}} = \xi_t$, $\forall \mathbf{m}$, приходим к свободно-аддитивному представлению процесса с неизвестным средним:

$$X_t = m_t + \xi_t,\ m_t \in \mathfrak{M},\ M\xi_t = 0,$$

где ξ_t считается свободным от m_t. Его чаще будем просто называть аддитивным. Не всякий процесс допускает такое представление, а лишь тот, для которого m_t-сечения модели одинаковы между собой за вычетом m_t.

Дополнения. 1. Точное распределение вероятностей процесса. Абсолютно точным распределение процесса X_t, $t \in \mathrm{T}$, может быть лишь в случае, если T дискретно и X_t при каждом $t \in \mathrm{T}$ принимает дискретные значения. Если это не так, то нужен остов в виде алгебры $\mathscr{A}_t$, на событиях которых задаются точные совместные вероятности $P(A_{t_1}, A_{t_2}, \ldots, A_{t_n})$, $A_{t_i} \in \mathscr{A}_{t_i}$ где $t_1, t_2, \ldots, t_n$ — произвольные выборки отсчетов из T. Корректность вероятностей и вместе с тем их согласованность эквивалентна конечной адди-

тивности. Тогда вероятности, взятые в качестве первичных, и определят процесс с точным на пересечении $\mathscr{A}=\bigcap_t \mathscr{A}_t$ алгебр распределением вероятностей.

Если $\mathscr{A}_t$ есть счетные алгебры, принцип построения модели процесса сохраняется, порождая конечно-аддитивную меру на $\mathscr{A}$ (для которой, однако, вероятности $P(A_{t_1}, A_{t_2}, \ldots)$ не будут, в общем, точными при счетных множествах отсчетов). Конечно-аддитивной мере соответствует единственная счетно-аддитивная мера согласно известной теореме Колмогорова о согласованных распределениях [1]. Меры может и не существовать, а задаваться на сопряженном пространстве, тогда процесс в классическом определении называется обобщенным [2]. Например, это процесс, у которого $Y_i=\int X_t\varphi_i(t)dt$, $i=1, 2, \ldots$, есть нормальные с.в. с нулевыми средними и $MY_iY_j=\int\varphi_i(t)\varphi_j(t)dt$. При нашем подходе в невыполнении счетной аддитивности нет «криминала», а обобщенный процесс естествен как задание средними на линейных преобразованиях.

2. Задание процесса переходными моделями. Этот способ состоит в последовательном задании переходных моделей $\mathscr{M}^{X_t}$, $\mathscr{M}^{X_{t_1}}_{X_t}$, $\mathscr{M}^{X_{t_2}}_{X_t, X_{t_1}}\ldots$, при произвольном выборе отсчетов $t<t_1<t_2<\ldots$ Если все последующие переходные модели зависят только от последнего момента: $\mathscr{M}^{X_{t_n}}_{X_{t_1},\ldots,X_{t_{n-1}}}=\mathscr{M}^{X_{t_n}}_{X_{t_{n-1}}}$, то способ задания называется *марковским*. Он характерен для точных вероятностей и оказывается малоэффективным при неточных. Причина неудобства лежит, во-первых, в трудностях согласования переходных моделей, если t непрерывно, и во-вторых, в «расползании» модели $\mathscr{M}^{X_{t_n}}$ при увеличении t_n в силу обоюдного влияния на нее как ширины модели $\mathscr{M}^{X_{t_{n-1}}}$, так и расплывчатости переходной модели $\mathscr{M}^{X_{t_n}}_{X_{t_{n-1}}}$. При точных и тех, и других, заданных вероятностями, ничего подобного не наблюдается, что и дает жизнь таким представлениям.

4.2. КОРРЕЛЯЦИОННЫЕ СВОЙСТВА

Процессы второго порядка. *Корреляционными свойствами* процесса X_t будем называть согласованное множество средних $\overline{M}(\sum e_iX_{t_i}+\sum\sum d_{ij}X_{t_i}X_{t_j})$, или в векторной форме

$$\overline{M}\,(\mathbf{e}^{\mathrm{T}}\mathbf{X}+\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{X}),\ \mathbf{e}^{\mathrm{T}}=(e_1, e_2, \ldots),$$
$$\mathbf{D}=\{d_{ij}\},\ \mathbf{X}^{\mathrm{T}}=(X_{t_1}, X_{t_2}, \ldots) \tag{4.3}$$

при всевозможных выборах отсчетов $t_i\in\mathrm{T}$ и коэффициентов c_i и d_{ij} конечных сумм. Нижние $\underline{M}$ также определены, но не записаны, так как сразу выражаются через верхние. Удобно считать матрицу $\mathbf{D}$ симметричной: $d_{ij}=d_{ji}$, что, как можно видеть, не скажется на корреляционных свойствах.

[1] Боровков А. А. Курс теории вероятностей. — М.: Наука, 1972. — С. 261.
[2] Гельфанд Н. М., Виленкин Н. Я. Некоторые применения гармонического анализа. — М.: ФМЛ, 1974. — Вып. 4. — С. 302.

Отметим, что корреляционные свойства есть более широкое понятие, чем просто границы $\bar{r}_{t,\tau}=\overline{M}X_tX_\tau$ корреляционной функции (если только не считать корреляции точными). Процесс, заданный своими корреляционными свойствами, называется *процессом второго порядка*. В его первичной основе обязательно лежит какой-то непротиворечивый набор средних из (4.3), скажем,

$$\overline{M}\,(\mathbf{e}^{\mathrm{T}}_{(k)}\mathbf{X}_{(k)}+\mathbf{X}^{\mathrm{T}}_{(k)}\mathbf{D}_{(k)}\mathbf{X}_{(k)})=\bar{m}_{(k)},\ k=1, 2, \ldots,$$

где $\mathbf{X}_{(k)}$ составляются, в общем, из разных последовательностей дискретных отсчетов $\mathbf{X}_{(k)}=(X_{t_1}\ \ldots, X_{t_k})$, а $\bar{m}_{(k)}$ — первичные значения. В частности, это могут быть границы средних $\overline{M}X_t$ и корреляций $\bar{r}_{t,\tau}$ при всех или некоторых t и τ. Любой процесс расширением приводится к процессу второго порядка, для чего за первичные берутся средние вида (4.3).

Первичный набор продолжается на остальные корреляционные свойства в соответствии с общей формулой (с одновременным согласованием $\tilde{m}_{(k)}$, если исходно они заданы несогласованными):

$$\overline{M}\,(\mathbf{e}^{\mathrm{T}}\mathbf{X}+\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{X})=\inf\{[c+\sum c^+_k\bar{m}_{(k)}]: c+\sum c^+_k(\mathbf{e}^{\mathrm{T}}_{(k)}\mathbf{X}_{(k)}+$$
$$+\mathbf{X}^{\mathrm{T}}_{(k)}\mathbf{D}_{(k)}\mathbf{X}_{(k)})\geqslant\mathbf{e}^{\mathrm{T}}\mathbf{X}+\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{X}\}, \tag{4.4}$$

где двоеточием отделено условие, при котором ищется инфимум в (4.4) выбором c, c^+_k. Из этого условия становится ясным, что нетривиальными могут быть средние только для квадратичных форм от таких векторов $\mathbf{X}^{\mathrm{T}}=(X_{t_1}, X_{t_2}, \ldots)$, компоненты которых встречаются хотя бы в одном $\mathbf{X}_{(k)}$.

Сказанное порождает неудобства, поэтому подчас разумным и наглядным следует признать существование определенной гладкости корреляционных свойств по времени t, что дает основание переходу к фиксированному набору отсчетов $t_1, t_2, \ldots, t_n$, а в результате — к единому вектору $\mathbf{X}$, считая, что между отсчетами будет иметь место нечто промежуточное. Тогда условие в (4.4) перепишется:

$$c+\sum c^+_k(\mathbf{e}^{\mathrm{T}}_{(k)}\mathbf{X}+\mathbf{X}^{\mathrm{T}}\mathbf{D}_{(k)}\mathbf{X})\geqslant\mathbf{e}^{\mathrm{T}}\mathbf{X}+\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{X}.$$

Если принять еще одно допущение, что процесс имеет нулевое среднее $MX_t\equiv 0$, $\forall t$, и что заданными являются $\bar{m}_{(k)}=\overline{M}\mathbf{X}^{\mathrm{T}}\mathbf{D}_{(k)}\mathbf{X}$, то возникают дополнительные упрощения:

$$\overline{M}\,(\mathbf{e}^{\mathrm{T}}\mathbf{X}+\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{X})=\overline{M}\,\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{X}=\inf\{\sum c^+_k\bar{m}_{(k)}:\sum c^+_k\mathbf{X}^{\mathrm{T}}\mathbf{D}_{(k)}\mathbf{X}\geqslant$$
$$\geqslant\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{X}\}. \tag{4.5}$$

Здесь в отличие от общей формулы (4.4) свободный коэффициент c положен равным 0, ибо таким он будет получаться при нахождении инфимума в (4.4) и принятых допущениях.

Неравенство в условии формулы (4.5) эквивалентно неотрицательной определенности матриц $\sum c^+_k\mathbf{D}_{(k)}-\mathbf{D}$, что символически

выглядит следующим образом: $\sum c^+_k \mathbf{D}_{(k)} - \mathbf{D} \geqslant 0$. Будем говорить, что $\sum c^+_k \mathbf{D}_{(k)}$ *мажорирует* $\mathbf{D}$. Итак, продолжение средних на квадратичные формы при симметричных матрицах $\mathbf{D}$ эквивалентно поиску среди матричных конечных линейных сумм $\sum c^+_k \mathbf{D}_{(k)}$, мажорирующих $\mathbf{D}$, такой, которой соответствует минимальное значение $\sum c^+_k \bar{m}_{(k)}$.

Пример 4.4. Пусть заданными являются $MX_{t_i} = 0$, $\overline{MX^2_{t_i}} = \overline{\sigma^2}$, $\underline{MX^2_{t_i}} = \underline{\sigma^2} = -\overline{M}(-X^2_{t_i})$, $i = 1, \ldots, n$. Они всегда будут непротиворечивыми, если $\underline{\sigma^2} \leqslant \overline{\sigma^2}$, и согласованными. Здесь первичными будут единичные матрицы $\pm\mathbf{I}$ со средними соответственно $\overline{\sigma^2}$, $-\underline{\sigma^2}$ (приведенными к верхнему). Неравенство после двоеточия в (4.5) запишется так: $c\mathbf{I} - \mathbf{D} \geqslant 0$, тогда инфимум в (4.5) достигается при c, равном максимальному λ_{max} собственному числу матрицы $\mathbf{D}$, и в зависимости от его знака $\overline{M}\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{X} = \max\{\lambda_{max}\underline{\sigma^2}, \lambda_{max}\overline{\sigma^2}\}$.

Представление процессов второго порядка семействами средних и ковариационных функций. Процесс называется *простым второго порядка*, если он задается точным средним $MX_t = m_t$ и точной ковариацией $b(t, \tau)$, определяемой формулой

$$b(t, \tau) = M(X_t - m_t)(X_\tau - m_\tau).$$

Обозначим модель простого процесса второго порядка как $\langle \mathbf{m}, \mathbf{b} \rangle$, где сокращенно $\mathbf{m} = \{m_t\}$, $\mathbf{b} = \{b(t, \tau)\}$.

Объединение простых процессов записывается:

$$\mathscr{M} = \bigvee_{\mathbf{m} \in \mathfrak{M}} \bigvee_{\mathbf{b} \in \mathfrak{B}_\mathbf{m}} \langle \mathbf{m}, \mathbf{b} \rangle, \tag{4.6}$$

где множество $\mathfrak{M}$ называется собственным семейством средних, а множества $\mathfrak{B}_\mathbf{m}$ — *собственными семействами ковариаций* (в общем, вид которых может быть разным для разных средних $\mathbf{m}$). Объединение (4.6) ведет к процессу второго порядка, задаваемому средними

$$\overline{M}\Big(\sum c_i X_{t_i} + \sum\sum d_{ij} X_{t_i} X_{t_j}\Big) = \sup_{\mathbf{m} \in \mathfrak{M}} \Big[\sum c_i m_{t_i} + \sum\sum d_{ij} m_{t_i} m_{t_j} + \sup_{\mathbf{b} \in \mathfrak{B}_\mathbf{m}} \sum\sum d_{ij} b(t_i, t_j)\Big]. \tag{4.7}$$

Причем все ковариации семейств $\mathfrak{B}_\mathbf{m}$ должны быть неотрицательно определенными в том смысле, что $\sum\sum c_i c_j b(t_i, t_j) \geqslant 0$ при любом выборе отсчетов t_i и коэффициентов c_i конечных сумм. Покажем, что такой способ задания процессов второго порядка является универсальным.

Теорема 4.1. *Каждый процесс второго порядка эквивалентным образом может быть задан в виде (4.6) собственными выпуклыми семействами $\mathfrak{M}$ средних и $\mathfrak{B}_\mathbf{m}$, $\mathbf{m} \in \mathfrak{M}$, ковариаций.*

Доказательство. Запишем $\mathscr{M} = \bigvee_{\mathbf{m} \in \mathfrak{M}} \mathscr{M}_\mathbf{m}$, где $\mathscr{M}_\mathbf{m} = \mathscr{M} \wedge \langle \mathbf{m} \rangle$ есть $\mathbf{m}$-сечение. Обозначим $\mathscr{M}_{\mathbf{m}, \mathbf{b}} = \mathscr{M}_\mathbf{m} \wedge \langle \mathbf{b} \rangle$ модель процесса с точным средним m_t и точной ковариацией $b(t, \tau)$, которая будет собственной, если пересечение непусто.

Заметим, что $\langle \mathbf{b} \rangle$ предполагает точное значение m_t и имеет смысл лишь для m_t-сечения $\mathscr{M}_\mathbf{m}$. Так как X_t есть процесс второго порядка, то его m_t, а затем $b(t, \tau)$-сечением будет $\mathscr{M}_{\mathbf{m}, \mathbf{b}} = (\mathscr{M} \wedge \langle \mathbf{m} \rangle) \wedge \langle \mathbf{b} \rangle = \langle \mathbf{m} \rangle \wedge \langle \mathbf{b} \rangle = \langle \mathbf{m}, \mathbf{b} \rangle$, что соответствует простому процессу второго порядка. Обозначая $\mathscr{B}_\mathbf{m} = \{\mathbf{b} : \mathscr{M}_{\mathbf{m}, \mathbf{b}} \neq \varnothing\}$ и выражая $\mathscr{M}$ объединением сечений, получаем представление (4.6), что и доказывает теорему.

Скрытое содержание этой теоремы, заслуживающее пристального внимания, заключается в том, что статистически неустойчивый процесс, у которого точных средних и ковариаций вовсе не существует, представляется семейством составляющих его процессов с точными значениями m_t и $b(t, \tau)$, т. е. статистически устойчивых. Это общее представление удобно, когда оно не чересчур громоздко.

Формулы (4.7) для расчета средних через семейственное представление можно рассматривать как результат применения к процессам второго порядка подчиненно-аддитивного представления (4.1): $X_{t, \mathbf{m}} = m_t + \xi_{t, \mathbf{m}}$, где $\xi_{t, \mathbf{m}}$ имеют нулевые средние $M_\mathbf{m} \xi_{t, \mathbf{m}} \equiv \mathbf{0}$ и описываются семействами $\mathfrak{B}_\mathbf{m}$ ковариаций. Расширенным будет представление вида (4.2): $X_{t, m} \subseteq m_t + \xi_t$, где «добавка» ξ_t имеет нулевое среднее $M\xi_t = 0$, свободна от m_t и определяется объединением $\bigcup_{\mathbf{m} \in \mathfrak{M}} \mathfrak{B}_\mathbf{m}$ семейств. Связь характеристик расширенного представления с исходными на одном частном случае будет исследована ниже.

Интервальные ковариации и корреляции. Корреляционные свойства в нашей интерпретации очень разнообразны. А нельзя ли упрощенно описать связь между X_t и X_τ с учетом неопределенности этой связи, вызванной как нашим незнанием, невозможностью ее точно проанализировать, так и статистической неустойчивостью процесса?

Уже говорилось, что эта связь характеризуется границами корреляций: $\bar{r}(t, \tau) = \overline{M} X_t X_\tau$. Но этого мало, и вот почему. Каждое X_t согласно аддитивному представлению складывается из двух составляющих: полностью неизвестной, статистически неустойчивой m_t — это любая функция из собственного семейства $\mathfrak{M}$, и случайной добавки и $\xi_{t, m}$ (возможно, подчиняющейся m_t). И чаще оказывается, что своему размаху границы $\bar{r}(t, \tau)$ обязаны именно влиянию m_t, а этого не хотелось бы. Выделим корреляционные свойства, которые были бы отделены от m_t и характеризуют остаток. При точных средних m_t и корреляциях $r(t, \tau)$ такие свойства дает *ковариационная функция* $b(t, \tau) = r(t, \tau) - m_t m_\tau$, а при неточных m_t — *границы ковариаций*:

$$b_1(t, \tau) = \underline{M}(X_t - \underline{m_t})(X_\tau - \underline{m_\tau}), \quad b_2(t, \tau) = \underline{M}(X_t - \overline{m_t})(X_\tau - \overline{m_\tau}),$$

$$b_3(t, \tau) = \overline{M}(X_t - \underline{m_t})(X_\tau - \overline{m_\tau}), \quad b_4(t, \tau) = \overline{M}(X_t - \overline{m_t})(X_\tau - \underline{m_\tau}).$$

Границ четыре: b_1 и b_2 — нижние, а b_3 и b_4 — верхние.

Каждую из ковариаций можно расписать как некоторую сторону собственных семейств $\mathfrak{M}$ и $\mathfrak{B}_{\mathfrak{m}}$, используя для этого (4.7). Например,

$$b_1(t,\tau) = \inf_{\mathfrak{m}\in\mathfrak{M}} \inf_{b\in\mathfrak{B}_{\mathfrak{m}}} [b(t,\tau) + (m_t - \underline{m}_t)(m_\tau - \underline{m}_\tau)] = \inf_{\mathfrak{m}\in\mathfrak{M}} [\underline{b}_{\mathfrak{m}}(t,\tau) + (m_t - \underline{m}_t)(m_\tau - \underline{m}_\tau)],$$

где $\underline{b}_{\mathfrak{m}}(t,\tau)$ — нижняя грань ковариаций в $\mathfrak{B}_{\mathfrak{m}}$, и так для остальных (нетрудно вывести их в качестве упражнения).

Рассмотрим пример, из которого можно извлечь содержание каждой из ковариаций.

Пример 4.5. Пусть процесс X_t второго порядка имеет постоянное среднее и может находиться в одном из двух состояний, в которых среднее и ковариация равны соответственно либо m_1, $K_1(t,\tau)$, либо m_2, $K_2(t,\tau)$, где m_1 и m_2 — два числа, причем для определенности считаем $m_1 > m_2$. Это дает описания двух простых процессов, составляющих X_t, а вместе задает модель X_t в виде объединения: $\mathscr{M} = \mathscr{M}_1 \vee \mathscr{M}_2$, где $\mathscr{M}_1 = \langle m_1, K_1\rangle$, $\mathscr{M}_2 = \langle m_2, K_2\rangle$. Тогда

$$b_1(t,\tau) = \min\{K_1(t,\tau);\ (m_2 - m_1)^2 + K_2(t,\tau)\},$$
$$b_2(t,\tau) = \min\{K_1(t,\tau) + (m_2 - m_1)^2;\ K_2(t,\tau)\},$$
$$b_3(t,\tau) = b_4(t,\tau) = \max\{K_1(t,\tau);\ K_2(t,\tau)\}.$$

Здесь $b_1(t,\tau)$ и $b_2(t,\tau)$ могут оказаться чуть завышенными по сравнению с минимальным из двух значений $K_1(t,\tau)$ и $K_2(t,\tau)$, тем не менее $\min\{b_1(t,\tau), b_2(t,\tau)\} = \min\{K_1(t,\tau), K_2(t,\tau)\}$.

Если все четыре ковариации равны 0, то X_t и X_τ называются нековариированными (см. (2.10)), что может быть следствием не только того, что в подчиненно-аддитивном представлении (4.1) случайные составляющие $\xi_{t,\mathfrak{m}}$ при каждом m_t некоррелированы между собой, т. е. $M_{\mathfrak{m}}\xi_{t,\mathfrak{m}}\xi_{\tau,\mathfrak{m}} = 0$ (тогда $b_1(t,\tau) = \inf_{\mathfrak{m}}(m_t - \underline{m}_t)\times$ $\times(m_\tau - \underline{m}_\tau) = 0$ и аналогично для остальных ковариаций), но и того, что просто случайная составляющая отсутствует, т. е. $X_t \equiv m_t$, как в примере 4.3 сугубо-ограниченного процесса, о котором известно только, что его реализации не могут по модулю превышать некоторый уровень.

Используем расширенное аддитивное представление (4.2): $X_t \sqsubseteq m_t + \xi_t$. Так как ξ_t описывается объединенным по m семейством $\bigcup \mathfrak{B}_m$, то границами ковариаций будут

$$\underline{b}(t,\tau) = \inf_{\mathfrak{m}\in\mathfrak{M}} \inf_{b\in\mathfrak{B}_{\mathfrak{m}}} b(t,\tau),$$
$$\overline{b}(t,\tau) = \sup_{\mathfrak{m}\in\mathfrak{M}} \sup_{b\in\mathfrak{B}_{\mathfrak{m}}} b(t,\tau).$$

Это будут расширенные границы, так как, в общем,

$$\underline{b}(t,\tau) \leqslant \min\{b_1(t,\tau), b_2(t,\tau)\}, \qquad \overline{b}(t,\tau) \geqslant \max\{b_3(t,\tau), b_4(t,\tau)\}.$$

Равенства имеют место, очевидно, при $\mathfrak{B}_m \equiv \mathfrak{B}$, $\forall m$, но не только (см. пример 4.5).

Рассмотрим связь ковариаций с корреляционными функциями: нижней $\underline{r}(t,\tau) = \underline{M}X_tX_\tau$ и верхней $\bar{r}(t,\tau) = \overline{M}X_tX_\tau$. На основании расширенного аддитивного представления имеем:

$$\underline{r}(t,\tau) = \inf_{m_t\in\mathfrak{M}} [m_t\, m_\tau + \underline{b}_{\mathfrak{m}}(t,\tau)] \geqslant \inf_{m_t\in\mathfrak{M}} m_t\, m_\tau + \underline{b}(t,\tau);$$
$$\overline{r}(t,\tau) = \sup_{m_t\in\mathfrak{M}} [m_t\, m_\tau + \overline{b}_{\mathfrak{m}}(t,\tau)] \leqslant \sup_{m_t\in\mathfrak{M}} m_t\, m_\tau + \overline{b}(t,\tau).$$

При $t=\tau$ получаем соответственно

$$\underline{M}X_t^2 \geqslant (|\underline{M}|\, X_t)^2 + \underline{b}(t,t), \qquad \overline{M}X_t^2 \leqslant (|\overline{M}|\, X_t)^2 + \overline{b}(t,t).$$

Все эти неравенства заменяются на равенства при свободно-аддитивном представлении. В этом случае точные значения корреляций $r(t,\tau)$ эквивалентны точным средним $MX_t = m_t$ и точной ковариации $b(t,\tau)$. Вообще при точных средних m_t границы корреляции с поправкой на слагаемое $m_t m_\tau$ совпадают с границами ковариаций:

$$\underline{r}(t,\tau) = m_t\, m_\tau + \underline{b}(t,\tau), \qquad \overline{r}(t,\tau) = m_t\, m_\tau + \overline{b}(t,\tau).$$

Разложение процесса по базису. Смысл любых разложений процесса сводится к упрощающей его замене дискретным набором коэффициентов.

Пусть $e_j(t)$, $j=1, 2, \ldots$, есть система ортонормированных функций, заданных на отрезке $[0, T]$:

$$\int e_i(t)\, e_j(t)\, dt = \delta_{ij} = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases}$$

Здесь и далее интеграл от 0 до T — по мере-длине. Функции $e_j(t)$ являются координатными «осями» Гильбертова пространства $\mathscr{L}_2$ всех интегрируемых с квадратом функций: $\mathscr{L}_2 = \{c_t : \int c_t^2 dt < \infty\}$. Скалярным произведением в этом пространстве будет $(c_t; a_t) = \int c_t a_t dt$. В этих обозначениях ортонормированность записывается: $(e_i(t);\ e_j(t)) = \delta_{ij}$.

Система $e_j(t)$, $j=1, 2, \ldots$, ортонормированных функций называется полной, если для любой функции $c_t \in \mathscr{L}_2$ имеет место равенство

$$\lim_{k\to\infty} \int \left[c_t - \sum_{j=1}^{k} \left(\int c_t\, e_j(t)\, dt\right) e_j(t) \right]^2 dt = 0.$$

В этой формуле сумма внутри квадратных скобок есть разложение функции c_t в ряд по $e_j(t)$.

Полная ортонормированная система функций $e_j(t)$ называется *базисом* в $\mathscr{L}_2$. Например, базисы в $\mathscr{L}_2$ на $[0, T]$ образуют: 1) нормированные полиномы Лежандра, 2) гармонические функции

$\sqrt{2/T}\sin_{\cos}(2\pi jt/T)$, $j=1, 2, \ldots$, дополненные постоянной $e_0(t)=1/\sqrt{T}$.

Рассмотрим разложение процесса по базису

$$X_t=\sum_{j=1}^{\infty} X(j)\,e_j(t), \qquad X(j)=\int X_t\,e_j(t)\,dt,\ t\in[0,T]. \tag{4.8}$$

Считается, что множество $\mathscr{X}=\mathscr{L}_2$ достоверных реализаций процесса X_t должно состоять из интегрируемых с квадратом функций.

Свойства коэффициентов разложения будут полностью определяться свойствами исходного процесса. В частности, корреляционные свойства процесса однозначно определяют корреляционные свойства коэффициентов разложения, что видно из цепочки равенств, справедливых для конечных сумм:

$$\overline{M}\left[\sum c_i X(i)+\sum\sum d_{ij}X(i)X(j)\right]=$$
$$=\overline{M}\left[\sum c_j\int X_t e_j(t)\,dt+\sum\sum d_{ij}\int X_t e_i(t)\,dt\times\right.$$
$$\left.\times\int X_t e_j(t)\,dt\right]=\overline{M}\left[\int X_t g(t)\,dt+\iint X_t X_\tau h(t,\tau)\,dt\,d\tau\right]=$$
$$=\lim_{|\Delta t|\to 0}\overline{M}\left[\sum_m X_{t_m} g(t_m)(t_{m+1}-t_m)+\right.$$
$$\left.+\sum_m\sum_n X_{t_m}X_{t_n}h(t_m,t_n)(t_{m+1}-t_m)(t_{n+1}-t_n)\right],$$

где $g(t)=\sum c_i e_i(t)$, $h(t,\tau)=\sum\sum d_{ij}e_i(t)e_j(t)$.

Для каждой реализации из $\mathscr{L}_2$ справедливо неравенство $\int X^2_t dt\geqslant\sum_1^k X(j)^2$, доказываемое следующим образом:

$$0\leqslant\int\left[X_t-\sum_1^k X(j)\,e_j(t)\right]^2 dt=\int\left[X_t^2-2X_t\sum_1^k X(j)\,e_j(t)+\right.$$
$$\left.+\left(\sum_1^k X(j)\,e_j(t)\right)^2\right]dt=\int X_t^2\,dt-\sum_1^k X(j)^2.$$

Отсюда, если процесс имеет конечную среднюю энергию, т. е.

$$\int\overline{M}X_t^2\,dt<\infty,\ \text{то}\ \overline{M}\sum_1^k X(j)^2\leqslant\overline{M}\int X_t^2\,dt\leqslant\int\overline{M}X_t^2\,dt;$$

$$\underline{M}\sum_1^k X(j)^2\leqslant\underline{M}\int X_t^2\,dt.$$

Переходя к пределу $k\to\infty$, получаем

$$\overline{M}\sum_1^\infty X(j)^2\leqslant\overline{M}\int X_t^2\,dt,\qquad \underline{M}\sum_1^\infty X(j)^2\leqslant\underline{M}\int X_t^2\,dt. \tag{4.9}$$

Неравенства наводят на мысль, что между процессом X_t и его разложением (4.8) может иметь место «энергетический дисбаланс». В этом случае корреляционные свойства коэффициентов разложения не будут определять полных корреляционных свойств исходного процесса и переход от процесса к его разложению вызовет потери.

Рассмотрим тот случай, когда переход к разложению не связан с потерями в свойствах второго порядка. Будем понимать *разложение* (4.8) в *среднеквадратическом скв-смысле*, соответствующем равенству

$$\lim_{k\to\infty}\overline{M}\int\left[X_t-\sum_1^k X(j)\,e_j(t)\right]^2 dt=0.$$

Ниже считаем, что процесс имеет конечную среднюю энергию.

В силу неравенств (4.9) имеем

$$0\leqslant\overline{M}\int X_t^2\,dt-\overline{M}\sum_1^\infty X(j)^2\leqslant\overline{M}\left[\int X_t^2\,dt-\right.$$
$$\left.-\sum_1^\infty X(j)^2\right]=\lim_{k\to\infty}\overline{M}\int\left[X_t-\sum_1^k X(j)\,e_j(t)\right]^2 dt=0.$$

Отсюда видно, что неравенства в (4.9) заменяются на равенства

$$\overline{M}\sum_1^\infty X(j)^2=\overline{M}\int X_t^2\,dt,\qquad \underline{M}\sum_1^\infty X(j)^2=\underline{M}\int X_t^2\,dt.$$

В этом случае корреляционные свойства процесса будут эквивалентны корреляционным свойствам коэффициентов разложения. Сказанное вкладывается в теорему.

Теорема 4.2. *Пусть процесс непрерывен в скв-смысле на* $[0,T]:\lim_{\tau\to t}\overline{M}(X_\tau-X_t)^2=0$. *Тогда его разложение* (4.8) *в ряд справедливо в скв-смысле. При этом корреляционные свойства процесса* X_t *и его коэффициентов разложения* $X(j)$ *эквивалентны.*

Доказательство. Доказательство второй части теоремы, по сути, содержится в рассуждениях, предшествующих теореме, поэтому осталось доказать первую. При точных среднем и ковариации $b(t,\tau)$ эта теорема является известной [1, с. 500]. В общем случае свойства второго порядка эквивалентны заданию собственных семейств $\mathfrak{M}$ и $\mathfrak{B}_{\mathbf{m}}$, $\mathbf{m}\in\mathfrak{M}$. Из непрерывности процесса вытекает равномерная непрерывность всех функций из $\mathfrak{M}$ и всех процессов $\xi_{t,\,\mathbf{m}}$ в подчиненно-аддитивном представлении $X_t=m_t+\xi_{t,\,\mathbf{m}}$. Отсюда каждая реализация m_t среднего и каждый процесс $\xi_{t,\,\mathbf{m}}$ может быть разложен в ряд по базису, а в силу равномерной непрерывности скв-сходимость этого ряда будет равномерной относительно $\mathfrak{M}$ и $\mathfrak{B}_{\mathbf{m}}$. Поэтому, обозначая $m(j)$ и $\xi_{\mathbf{m}}(j)$ — коэффициенты разложения $m(t)$ и $\xi_{t,\,\mathbf{m}}$ в ряды по базису $e_j(t)$ и используя c_r-неравенство § 3.1, получаем

$$\overline{M}\int\left[X_t-\sum_1^k X(j)\,e_j(t)\right]^2 dt=$$

$$=\sup_{\mathbf{m}\in\mathfrak{M}} 2\overline{M}^{\xi}_{\mathbf{m}}\int\left[m_t-\sum_1^k m(j)\,e_j(t)+\xi_{t,\,\mathbf{m}}-\right.$$

$$\left.-\sum_1^k \xi_{\mathbf{m}}(j)\,e_j(t)\right]^2 dt\leqslant \sup_{\mathbf{m}\in\mathfrak{M}}\left\{2\int\left[m_t-\sum_1^k m(j)\,e_j(t)\right]^2 dt+\right.$$

$$\left.+\overline{M}^{\xi}_{\mathbf{m}}\,2\int\left[\xi_{t,\,\mathbf{m}}-\sum_1^k \xi_{\mathbf{m}}(j)\,e_j(t)\right]^2 dt\right\}=$$

$$=2\sup_{\mathbf{m}\in\mathfrak{M}}\left\{\int\left[m_t-\sum_1^k m(j)\,e_j(t)\right]^2 dt+\right.$$

$$\left.+\sup_{\mathbf{b}\in\mathfrak{B}_{\mathbf{m}}}\overline{M}^{\xi}_{\mathbf{m},\,\mathbf{b}}\int\left[\xi_{t,\,\mathbf{m},\,\mathbf{b}}-\sum_1^k \xi_{\mathbf{m},\mathbf{b}}(j)\,e_j(t)\right]^2 dt\right\}\leqslant$$

$$\leqslant 2\sup_{\mathbf{m}\in\mathfrak{M}}\int\left[m_t-\sum_1^k m(j)\,e_j(t)\right]^2 dt+$$

$$+\sup_{\mathbf{b}\in\mathfrak{B}}\left[b(t,\,\tau)-\sum_1^k\sum_1^k\int\int b(t,\,\tau)\,e_i(t)\,e_j(\tau)\,dt\,d\tau\right].$$

Оба слагаемых правой части стремятся к 0 при $k\to\infty$ в силу равномерной непрерывности семейств $\mathfrak{M}$ и $\mathfrak{B}=\bigcup\limits_{\mathbf{m}\in\mathfrak{M}}\mathfrak{B}_{\mathbf{m}}$, что доказывает теорему.

Отметим, что скв-непрерывный процесс второго порядка при его разложении преобразуется в последовательность также второго порядка.

4.3. ОДНОРОДНЫЕ И СТАЦИОНАРНЫЕ ПРОЦЕССЫ

Однородные процессы. Понятие однородности применительно к последовательностям вводилось в § 3.1. Здесь оно прикладывается к процессам.

Процесс X_t считается *однородным*, если его модель симметрична к сдвигам во времени: $\mathscr{M}^{X_t}\equiv\mathscr{M}^{X_{t+\tau}}$, что эквивалентно тождеству средних $\overline{M}f\{X_t\}\equiv\overline{M}f\{X_{t+\tau}\}$, $\forall f\in\mathscr{F}$, $\forall\tau$. Совершенно ясно, что понятие однородности имеет точный смысл лишь для процессов, определенных на всей длине временной оси $(-\infty,\ \infty)$. Но можно рассматривать участок однородного процесса любой конечной протяженности $[0,\ T]$. Однородность позволяет сэкономить на задании процесса, сделав это раз для одного положения признаков на временной оси и перенеся те же значения на любой сдвиг по времени, существенно преумножая тем самым первичный набор.

Процесс будет однородным, если выполняются два условия: 1) набор первичных функционалов, определяющих процесс, при

сдвиге во времени преобразуется сам в себя; 2) первичные средние при сдвиге во времени не меняются. Сказанное символьно записывается:

1) $g\{X_t\}\in\mathscr{g}\Rightarrow g\{X_{t+\tau}\}=g_\tau\{X_t\}\in\mathscr{g}$, $\forall\tau$;

2) $\overline{M}\,g\{X_t\}=\overline{M}\,g\{X_{t+\tau}\}$, $\forall g\in\mathscr{g}$, $\forall\tau$.

Примером однородного является процесс, определенный постоянными границами среднего $\underline{M}X_t=\underline{m}$, $\overline{M}X_t=\overline{m}$ и текущей мощности $\underline{M}X^2{}_t=\underline{r}$, $\overline{M}X^2{}_t=\overline{r}$. Первичный набор здесь создают всевозможные отсчеты процесса и их квадраты: $\mathscr{g}=\{\pm x_t,\ \pm x^2{}_t,\ -\infty<t<\infty\}$, и видно что этот набор не меняется при сдвигах во времени, как и средние на нем.

Объединение и пересечение (имея в виду модели) однородных процессов приводят к однородному процессу. Сужение же может привести к нарушению свойства однородности, поскольку внутри однородного процесса существуют неоднородные составляющие, которые могут «выскочить» при этом наружу.

Любой процесс X_t расширением («забыванием» его неоднородных особенностей) может быть приведен к однородному процессу Y_t. Для этого нужно положить $\overline{M}f\{Y_t\}=\sup\limits_{-\infty<\tau<\infty}\overline{M}f\{X_{t+\tau}\}$, $\forall f\in\mathscr{F}$, и эти значения зададут Y_t.

Голый процесс, соответствующий полному отсутствию каких-либо данных, всегда однороден: сдвиг во времени не меняет нулевых сведений о таком процессе.

Инвариантные во времени преобразования однородных процессов ведут к однородному процессу, например нелинейные безынерционные преобразования вида $Y_t=f\{X_t\}$, линейные преобразования вида $Y_t=\int h(t-\tau)X_\tau d\tau$. Складывание, вычитание и перемножение однородных процессов приводит снова к однородному процессу.

Расширяя понятие однородности, будем говорить не о неизменности всех свойств процесса по отношению к сдвигу во времени, а о неизменности только некоторых из них. Процесс называется $\mathscr{Q}$-однородным, если $\overline{M}q\{X_t\}=\overline{M}q\{X_{t+\tau}\}$, $\forall q\in\mathscr{Q}$, $\forall\tau$, т. е. при сдвиге во времени не меняются средние признаков набора $\mathscr{Q}$. Это *частично-однородный* процесс, но если $\mathscr{Q}$ является первичным набором, он будет однородным в общем смысле. То же самое будет, если $\mathscr{Q}\supset\mathscr{g}$, где $\mathscr{g}$ есть первичный набор. Если же $\mathscr{Q}\subset\mathscr{g}$, то $\mathscr{Q}$-однородный процесс сам по себе может не быть однородным, так как некоторые его первичные данные способны, в общем, зависеть от сдвига во времени.

Например, если первичными являются $\underline{M}X_t=\underline{m}$, $\overline{M}X_t=\overline{m}$, $\underline{M}X^2{}_t=\underline{r}$, $\overline{M}X^2{}_t=\overline{r}$, то процесс является однородным. Если же помимо этих имеются другие первичные данные, зависящие от времени, например $\overline{M}X^3{}_t=m_{(3)}(t)$, то процесс будет уже только частично $\{\pm x,\ \pm x^2\}$-однородным, а среднее и мощность — его однородные параметры.

Процесс, корреляционные свойства которого не меняются во времени, называется *однородным в широком смысле*. Для такого процесса тождественно по τ

$$\overline{M}\left(\sum c_i X_{t_i} + \sum\sum d_{ij} X_{t_i} X_{t_j}\right) \equiv$$

$$\equiv \overline{M}\left(\sum c_i X_{t_i+\tau} + \sum\sum d_{ij} X_{t_i+\tau} X_{t_j+\tau}\right).$$

Это частично-однородный процесс, но если других кроме корреляционных данных нет (т. е. процесс второго порядка), то он однороден в общем смысле.

Рассмотрим однородные процессы второго порядка. Собственное семейство $\mathfrak{M}$ средних для них не должно меняться при сдвигах во времени $m_t \in \mathfrak{M} \Rightarrow m_{t+\tau} \in \mathfrak{M}$, $\forall\tau$, и то же самое можно сказать о собственных семействах ковариаций: $b(t, t') \in \mathfrak{B}_{m_t} \Rightarrow$ $\Rightarrow b(t+\tau, t'+\tau) \in \mathfrak{B}_{m_{t+\tau}}$. Границы ковариаций должны зависеть только от разности аргументов: $\underline{b}(t, t') = \underline{b}(t-t')$, $\overline{b}(t, t') = \overline{b}(t-t')$, хотя, в общем, каждая из ковариаций собственного семейства, как будет видно, таким свойством не обязана обладать.

Детерминированная функция c_t, рассматриваемая как процесс с единственной возможной реализацией, будет однородна, если только эта реализация есть тождественная постоянная $c_t \equiv m$. Процесс второго порядка, определяемый точным средним m_t и ковариацией $b(t, t')$, будет однородным, если только среднее постоянно: $m_t = m$, а ковариация зависит от разности аргументов: $b(t, t') = b(t-t')$. Назовем такую *ковариацию однородной*.

Рассмотрим сечение однородного процесса X_t точным средним m_t и ковариацией $b(t, t')$, записав следующим образом:

$$\mathscr{M}_{\mathbf{m},\mathbf{b}} = \mathscr{M} \wedge \langle \mathbf{m}, \mathbf{b} \rangle.$$

Сечение может быть непустым, даже если m_t зависит от t и ковариация неоднородна. Следовательно, *для однородных процессов в собственные семейства* $\mathfrak{M}$ *и* $\mathfrak{B}_{\mathbf{m}}$ *входят, в общем, неоднородные функции* m_t *и* $b(t, t')$. Отсюда однородный процесс складывается и из неоднородных составляющих.

Пример 4.6. Однородный процесс первого порядка. Пусть однородный процесс X_t задан постоянными границами среднего: $M\underline{X}_t = \underline{m}$, $M\overline{X}_t = \overline{m}$. Имеем процесс первого порядка. Для него

$$\overline{M}\sum c_i X_{t_i} = \sum_{c_i \geqslant 0} c_i \overline{m} + \sum_{c_i < 0} c_i \underline{m}.$$

Под такое же среднее для заданных c_i можно подогнать другой процесс первого порядка X^*_t, задав его точным средним вида $MX^*_{t_i} = \overline{m}$ при t_i таких, что $c_i \geqslant 0$, и $MX^*_{t_i} = \underline{m}$ при t_i таких, что $c_i < 0$. Этот неоднородный процесс, поскольку среднее его меняется во времени, включается в предыдущий: $X^*_t \subset$ $\subset X_t$, более того, входит лишь как один составляющий элемент в представление X_t в виде объединения m_t-простых процессов: $\mathscr{M} = \bigvee_{\underline{m} \leqslant m_t \leqslant \overline{m}} \langle m_t \rangle$, что иллюстрирует нашу мысль о неоднородных составляющих однородного процесса.

Таким образом, однородность суть симметрия исходных данных о процессе к сдвигу во времени.

Стационарные процессы. В § 3.1 были введены понятия стационарных признаков и стационарной последовательности с. в. Переложим эти понятия на процессы.

Признак $g\{X_t\}$ (функционал от X_t) называется *стационарным*, если для любого сдвига τ

$$M[g\{X_t\} - g\{X_{t+\tau}\}] \equiv 0, \quad \forall\tau.$$

Эквивалентным написанному будет:

$$\mathscr{M} \wedge \langle Mg\{X_t\} = a \rangle \wedge \langle Mg\{X_{t+\tau}\} = a' \rangle = \varnothing \text{ при } a \neq a'.$$

Смысл в том, что если бы вдруг оказалось точно известным среднее значение $Mg\{X_t\} = a$ признака g, то оно осталось бы таким же при любых сдвигах процесса во времени. Иначе говоря, если обозначить Mg-сечения через $\mathscr{M}_{Mg} = \mathscr{M} \wedge \langle Mg\{X_t\} \rangle$, то для $\mathscr{M}_{Mg}$ все средние от функционалов $g\{X_{t+\tau}\}$ должны быть точными, равными значению $Mg\{X_t\}$ независимо от сдвига τ: $M_{Mg} g\{X_{t+\tau}\} \equiv Mg$.

Процесс называется *$\mathcal{Q}$-стационарным (частично-стационарным)*, если все признаки набора $\mathcal{Q}$ являются стационарными, и *стационарным*, если $\mathcal{Q}$ — любые признаки. Последнее определение требует комментариев. А существует ли вообще в природе стационарный процесс, т. е. такой, что каков бы признак f ни был взят, он оказывается стационарным? Если да, то это подразумевает, во-первых, возможность точного знания среднего любого признака, и во-вторых, полную идентичность работы внутреннего статистически устойчивого механизма процесса по времени. Но даже пусть такового процесса и нет, все равно абсолютизация стационарности как математической абстракции оказывается весьма удобной как уверенность, что какой бы набор $\mathcal{Q}$ признаков ни был взят, при сдвиге во времени их средние, если они вдруг станут точно известными, не меняются. Имея в виду, что практический выбор будет всегда ограничен нашими возможностями, так что фактически обращаться будем с частичной стационарностью.

Итак, стационарность — это статистическая устойчивость процесса в двух направлениях. Во-первых, по ансамблю, когда для процесса, если бы удалось его неоднократно «прокрутить» в одинаковых условиях, устанавливается (а иначе — декларируется) существование точных средних статистических Mq его признаков q. И во-вторых, по времени, когда утверждается неизменность этих средних статистических по течению процесса во времени.

$\mathcal{Q}$-стационарные процессы представляются как семейства $M\mathcal{Q}$-точных стационарных процессов: $\mathscr{M} = \bigvee_{M\mathcal{Q}} \mathscr{M}_{M\mathcal{Q}}$ и $\mathscr{M}_{M\mathcal{Q}}$ — это

сечение $\mathcal{M}$, т. е. складываются как ансамбли составляющих эти процессы стационарных «кусочков».

$\mathcal{Q}$-стационарные процессы, очевидно, будут $\mathcal{Q}$-однородными, а если $\mathcal{Q}=\mathcal{G}$ — есть первичный набор, то вообще однородными.

$\mathcal{G}$-стационарные с первичным набором $\mathcal{G}$ процессы могут интерпретироваться как семейства составляющих их простых стационарных процессов, что соответствует объединению:

$$\langle \overline{M}\ g\rangle = \bigvee_{M\mathcal{G}} \langle M\ g\rangle \qquad (4.10)$$

и эквивалентно следующей формуле расчета границ средних от вторичных признаков $\mathcal{L}\mathcal{G}$ (конечных линейных комбинаций первичных функционалов):

$$\overline{M}\ \Sigma\ c_i\ g_i\{X_t\} = \sup_{M\mathcal{G}}\ \Sigma\ c_i\ Mg_i,$$

где супремум берется по простым процессам и соответствующим им Mg_i, составляющим стационарные процессы.

Понятие стационарности существенно тоньше по сравнению с однородностью. Если однородность — это неизменность во времени внешних форм, а внутри может твориться все, что угодно, то стационарность — это неизменность внутренней структуры, тех простых частичек, на которые делится модель (см. (4.10)), и, как следствие, — внешних форм. При одинаковых первичных данных ИМ стационарного процесса всегда уже $\mathcal{G}$-однородного. Понятия $\mathcal{G}$-стационарности и $\mathcal{G}$-однородности совпадают, лишь когда модель сама есть одна «простая частичка», т. е. $\mathcal{G}$-точная.

В следующем утверждении устанавливаются алгебраические свойства моделей стационарных процессов.

Теорема 4.3. *Пересечение модели $\mathcal{Q}_1$-стационарного процесса с $\mathcal{Q}_2$-стационарным ведет к модели $\mathcal{Q}_1\cup\mathcal{Q}_2$-стационарного процесса. А модели $\mathcal{G}$-стационарных процессов, заданные на одном и том же первичном наборе $\mathcal{G}$ функционалов, при объединениях сохраняют свойство стационарности.*

Доказательство. Пусть X_t и Y_t — соответственно $\mathcal{Q}_1$- и $\mathcal{Q}_2$-стационарные процессы. Тогда для $q\in\mathcal{Q}_1$ согласно определению и на основании коммутативности пересечения имеем при $a\neq a'$:

$$\mathcal{M}^X \wedge \mathcal{M}^Y \wedge \langle Mq\{X_t\}=a\rangle \wedge \langle Mq\{X_{t+\tau}\}=a'\rangle = \varnothing \quad \forall\tau,$$

что доказывает стационарность параметра Mq, $q\in\mathcal{Q}_1$, для $\mathcal{M}^X\wedge\mathcal{M}^Y$. Это же, очевидно, верно и для $q\in\mathcal{Q}_2$, что доказывает первую часть теоремы.

Вторая часть вытекает из (4.10) и коммутативности операции объединения: $\langle\overline{M}^X\mathcal{G}\rangle\vee\langle\overline{M}^Y\mathcal{G}\rangle=(\vee\langle M^X\mathcal{G}\rangle)(\vee\langle M^Y\mathcal{G}\rangle)=\vee\langle M\mathcal{G}\rangle$, где в круглых скобках объединение производится соответственно по $\langle M^X\mathcal{G}\rangle\subset\langle\overline{M}^X\mathcal{G}\rangle$ и $\langle M^Y\mathcal{G}\rangle\subset\langle\overline{M}^Y\mathcal{G}\rangle$, а в конце — по $\langle M\mathcal{G}\rangle\subset\langle\overline{M}^X\mathcal{G}\rangle\vee\langle\overline{M}^Y\mathcal{G}\rangle$, что и требовалось доказать.

Смысл первой части теоремы 4.3 состоит в том, что если к известным имеющимся сведениям о частично-стационарном процессе добавляются дополнительные данные, обладающие свойством стационарности, то процесс останется стационарным. Например, если для стационарного процесса второго порядка указано дополнительно, что вероятности $P(a<X_t<b)$ существуют и не зависят от t, то процесс будет также стационарным, но уже на более широком ансамбле признаков. В развитие этой мысли абсолютно стационарный процесс может представляться как пересечение всех включающих его частично-стационарных процессов.

Рассмотрим пример частично-стационарного процесса.

Пример 4.7. Процесс первого порядка со стационарными свойствами. Определим объединение $\vee\langle MX_t=m\rangle$ как семейство процессов, о каждом из которых известно только, что среднее во времени не меняется и чему оно равно. В добавление к этому заданы границы $\underline{m}$, $\overline{m}$ среднего m. Для такого процесса (сравни с примером 4.6 однородного процесса) верно:

$$\overline{M}\ \Sigma\ c_i X_{t_i} = \max_{\underline{m}\leqslant m_i\leqslant\overline{m}}\ \Sigma\ c_i\ m_i = \max\left\{\underline{m}\ \Sigma\ c_i,\ \overline{m}\ \Sigma\ c_i\right\}.$$

Из этого равенства следует полезный вывод: получается одна и та же модель первого порядка, принимает ли m значения внутри интервала $\underline{m}$, $\overline{m}$, или только крайние $\underline{m}$ и $\overline{m}$, т. е. та же модель достигается как объединение двух составляющих $\langle MX_t=\underline{m}\rangle\vee\langle MX_t=\overline{m}\rangle$. Рассматриваемый нами процесс представляется $X_t=m+\xi_t$, где m принимает любое значение в отрезке $[\underline{m},\ \overline{m}]$, а ξ_t свободен от m и $M\xi_t=0$ — это все, что о ξ_t известно.

Отметим, что для $\mathcal{Q}$-стационарных процессов со стационарным средним MX_t (т. е. x_t как признаки входят в $\mathcal{Q}$) характерно подчиненно-аддитивное представление: $X_t=m+\xi_{t,m}$, где m — независящий от t параметр, а добавка свободна от m и при каждом его значении имеет нулевое среднее. Сказанное относится и к следующему процессу.

Процесс второго порядка называется стационарным, если стационарными являются все его признаки корреляций. Это есть разновидность частичной стационарности, где $\mathcal{Q}$ — линейно-квадратичные функционалы.

Процесс второго порядка X_t будет стационарным, если и только если выполняются два условия: 1) *собственное множество $\mathfrak{M}$ средних содержит лишь константы m*; 2) *собственные семейства $\mathfrak{B}_m$ ковариаций составляют однородные функции $b(t-t')$, зависящие только от разности аргументов.*

Модель стационарного процесса второго порядка представляется в виде

$$\mathcal{M} = \bigvee_{m\in\mathfrak{M}}\ \bigvee_{b\in\mathfrak{B}_m} \langle m, b\rangle,$$

где $\langle m, b\rangle$ определяет простой стационарный процесс, заданный своими постоянным средним m и однородной ковариацией $b(\tau)$.

Интересно отметить, что процесс второго порядка может быть стационарным, хотя средние и ковариации совершенно неизвестны. Для такого процесса $\mathfrak{M}$ составляют любые константы, а $\mathfrak{B}_m = \mathfrak{B}$ — класс всевозможных неотрицательно определенных функций $b(t—t')$, зависящих лишь от разности аргументов.

Замечание. По аналогии с предыдущим произвольный процесс (не обязательно второго порядка) со стационарными корреляционными свойствами будет называться *стационарным в широком смысле*. При этом некоторые другие его свойства (например, вероятности превышений), выходящие за рамки свойств второго порядка, могут зависеть от времени. Стационарный в широком смысле процесс расширением (при котором сохраняются только корреляционные свойства) приводится к стационарному процессу второго порядка.

Спектральные двойники процессов. Если рассматривать ограниченный интервал времени 0, T длительностью T, то гармонические функции $\sqrt{2/T}\sin(k2\pi t/T)$, $\sqrt{2/T}\cos(k2\pi t/T)$, $k=1, 2, \ldots$, образуют ортонормированный базис. Дополним его постоянной $e_0(t)=1/\sqrt{T}$. Согласно теореме 4.2 любой скв-непрерывный процесс в скв-смысле может быть разложен в ряд по этому базису

$$X_t = X_0 + \sum_{k=1}^{\infty}\left[X_k^s \sin\left(k\frac{2\pi}{T}t\right) + X_k^c \cos\left(k\frac{2\pi}{T}t\right)\right],\ 0\leqslant t\leqslant T, \tag{4.11}$$

где $X_0 = \frac{1}{\sqrt{T}}\int_0^T X_t\,dt$, $X_k^s = \sqrt{\frac{2}{T}}\int_0^T X_t \sin\left(k\frac{2\pi}{T}t\right)dt$,

$X_k^c = \sqrt{\frac{2}{T}}\int_0^T \cos\left(k\frac{2\pi}{T}t\right)dt$.

Ряд (4.11) называется *рядом Фурье*, а $X^s{}_k$ и $X^c{}_k$ — коэффициентами Фурье или *спектральными коэффициентами*.

Подчиненно-аддитивному представлению $X_t = m_t + \xi_{t,\mathfrak{m}}$ соответствуют аналогичные представления коэффициентов разложения Фурье левой и правой частей: $X^s{}_k = m^s{}_k + \xi^s{}_{k,\mathfrak{m}}$, $X^c{}_k = m^c{}_k + \xi^c{}_{k,\mathfrak{m}}$.

Смысл разложения процесса в ряд Фурье состоит в том, чтобы представить процесс в удобном виде, используя конечный набор спектральных коэффициентов. При этом $\mathscr{M}$ процесса преобразуется в модель $\mathscr{M}_\Phi$ коэффициентов Фурье, что формально записывается: $\mathscr{M}_\Phi = S_\Phi \mathscr{M}$, а S_Φ называется преобразованием Фурье.

Модель $\mathscr{M}_\Phi$ определена средними, рассчитываемыми по формуле:

$$\overline{M}_\Phi \varphi\{X_k^s, X_k^c, k=1, 2, \ldots\} = \overline{M}\varphi\left\{\sqrt{\frac{2}{T}}\int X_t \sin\left(k\frac{2\pi}{T}t\right)dt,\right.$$

$$\left.\sqrt{\frac{2}{T}}\int X_t \cos\left(k\frac{2\pi}{T}t\right)dt,\ k=1, 2, \ldots\right\} = \overline{M} f_\varphi\{X_t\},$$

где φ — функционал в пространстве значений коэффициентов

Фурье, а f_φ — его изображение в пространстве реализаций (см. § 2.1).

Если ограничиться рассмотрением только корреляционных свойств, то точным средним m_t и ковариациям $b(t, t')$ процесса X_t будут соответствовать точные средние $m^s{}_k$, $m^c{}_k$ коэффициентов Фурье и точные их ковариации, определяемые выражениями:

$$b_{kl}^{ss} = \frac{2}{T}\int_0^T\int_0^T b(t, t')\sin\left(k\frac{2\pi}{T}t\right)\sin\left(l\frac{2\pi}{T}t'\right)dtdt';$$

$$b_{kl}^{sc} = \frac{2}{T}\int_0^T\int_0^T b(t, t')\sin\left(k\frac{2\pi}{T}t\right)\cos\left(l\frac{2\pi}{T}t'\right)dtdt';$$

$$b_{kl}^{cc} = \frac{2}{T}\int_0^T\int_0^T b(t, t')\cos\left(k\frac{2\pi}{T}t\right)\cos\left(l\frac{2\pi}{T}t'\right)dtdt'.$$

Обозначим матрицу этих коэффициентов через B_Φ.

Собственным семействам средних $\mathfrak{M}$ и $\mathfrak{B}_\mathfrak{m}$ ковариаций процесса X_t будут соответствовать собственные семейства средних $\mathfrak{M}_\Phi = S_\Phi\mathfrak{M}$, где S_Φ — обозначение преобразования Фурье, и матриц $\mathfrak{B}_{\Phi,\mathfrak{m}}$ ковариаций в пространстве коэффициентов Фурье. Это соответствие взаимно однозначно.

Перейдем к тому случаю, когда X_t есть стационарный в широком смысле скв-непрерывный процесс, полагая для простоты $MX_t=0$. Тогда все функции собственного семейства $\mathfrak{B}$ будут непрерывными, зависящими от разности аргументов, а ковариации коэффициентов Фурье будут равны $b^{sc}{}_{kl} = 0$ — это для перекрестных синус-косинус, а для совпадающих — уже не будут нулевыми и для отдельной $b(t—t') \in \mathfrak{B}$ записываются:

$$\begin{matrix} b_{kl}^{ss} \\ b_{kl}^{cc} \end{matrix} = \frac{2}{\pi}\int_0^T b(\tau)\left[\frac{1}{k-l}\cos\frac{(k+l)\pi\tau}{T}\sin\frac{(k-l)\pi(T-\tau)}{T} \pm\right.$$

$$\left.\pm\frac{1}{k+l}\cos\frac{(k-l)\pi\tau}{T}\sin\frac{(k+l)\pi\tau}{T}\right]d\tau, \quad k\neq l;$$

$$b_{kk}^{ss} = b_{kk}^{cc} = 2\int_0^T b(\tau)\frac{T-\tau}{T}\cos\frac{2k\pi\tau}{T}d\tau + 0\left(\frac{1}{T}\right).$$

Из данных выражений видно, что при $T\to\infty$ и $k2\pi/T\to\omega$ имеет место сходимость:

$$b^{ss}{}_{kl}\to\delta_{kl}B(\omega),\ b^{cc}{}_{kl}\to\delta_{kl}B(\omega),$$

где $B(\omega) = 2\int_0^\infty b(\tau)\cos\omega\tau\,d\tau$

называется *энергетическим спектром процесса*, а $\delta_{kl}=1$ при $k=l$ и 0 при $k\neq l$. Каждой точной неотрицательно определенной кор-

реляционной функции $b(t-t')$ соответствует энергетический спектр $B(\omega)$, обладающий свойствами:

$$B(\omega) \geqslant 0;\ B(\omega) = B(-\omega);\ B(\omega) \underset{|\omega|\to\infty}{\to} 0;\ 2\int_0^\infty B(\omega)\, d\omega = b(0).$$

Собственному семейству $\mathfrak{B}$ ковариаций, таким образом, ставится в соответствие собственное семейство энергетических спектров.

Рассмотрим следующие спектральные преобразования процесса:

$$X^s_{\omega,T} = \sqrt{\frac{2}{T}} \int_0^T X_t \sin \omega t dt, \qquad X^c_{\omega,T} = \sqrt{\frac{2}{T}} \int_0^T X_t \cos \omega t \cdot dt.$$

Здесь ω называется частотой, а $X^s_{\omega,T}$ и $X^c_{\omega,T}$ — спектральными составляющими процесса. Очевидно, $X^s_{2\pi k/T,\,T} = X^s_k$, $X^c_{2\pi k/T,\,T} = X^c_k$, на основании чего можно сделать вывод, что спектральные составляющие $X^s_{\omega,T}$ и $X^c_{\omega,T}$, $\omega \in \mathcal{R}$, дают описание на интервале $(0, T)$ скв-непрерывного процесса.

Задать модель $\mathcal{M}$ процесса X_t — все равно, что задать соответствующую ей модель в частотной области при $-\infty < \omega < \infty$, где процесс X_t, $0 \leqslant t \leqslant T$, заменяется на пару процессов $X^s_{\omega,T}$, $X^c_{\omega,T}$, которую удобно записать в комплексном виде

$$\dot{X}_{\omega,T} = X^c_{\omega,T} + \mathrm{j} X^s_{\omega,T} = \sqrt{\frac{2}{T}} \int_0^T X_t \exp(\mathrm{j}\omega t)\, dt$$

и называть *спектральным двойником* X_t. Это комплексная функция ω. По формуле (4.11) можно, зная спектральный двойник, однозначно восстановить X_t, поэтому X_t и $\dot{X}_{\omega,T}$ эквивалентны.

Спектральные процессы. Понятие спектра процесса имеет под собой твердую физическую основу и защищено с практической стороны экспериментами, а с научной — теоремой Винера — Хинчина. Мы дадим отличающуюся от классической математическую интерпретацию спектра — ту, которая логично вытекает из наших построений для стационарного непрерывного процесса.

Понятие спектра так или иначе предполагает бесконечное время, поэтому пусть процесс задан на $[0, T]$ и $T \to \infty$. Считаем $MX_t = 0$. Спектральный двойник $\dot{X}_{\omega,T}$, рассматриваемый при каждом фиксированном ω как последовательность индексированных параметром T случайных величин, при $T \to \infty$ не является скв-сходящимся ни к какой предельной случайной величине; скв-предела не существует. Но это не столь важно. Важно то, что для стационарных процессов «перекрестные» корреляции между разными частотами при $T \to \infty$ стремятся к 0, а именно, верно следующее:

$$\lim_{T\to\infty} M(\dot{X}_{\omega,T}\, \overset{*}{\dot{X}}_{\omega',T}) = 2\underline{B}(\omega)\, \delta_0(\omega - \omega');$$

$$\lim_{T\to\infty} \overline{M}(\dot{X}_{\omega,T}\, \overset{*}{\dot{X}}_{\omega',T}) = 2\overline{B}(\omega)\, \delta_0(\omega - \omega'),$$

где $\underline{B}(\omega) = \inf\limits_{b\in\mathfrak{B}} 2\int_0^\infty b(\tau)\cos\omega\tau d\tau$, $\overline{B}(\omega) = \sup\limits_{b\in\mathfrak{B}} 2\int_0^\infty b(\tau)\cos\omega\tau d\tau$, а $\delta_0(\omega - \omega')$ есть 1 при $\omega = \omega'$, иначе 0. Отсюда видно, что к нулю стремится корреляция между действительной и мнимой частями спектрального двойника.

Обозначим $\dot{X}_\omega$ процесс, обладающий предельными при $T \to \infty$ корреляционными свойствами процесса $\dot{X}_{\omega,T}$, и назовем его *спектральным процессом*. Согласно написанным выше соотношениям для спектрального процесса имеем следующие определяющие его равенства:

$$MX^c_\omega X^c_{\omega'} = MX^s_\omega X^s_{\omega'} = MX^c_\omega X^s_{\omega'} = MX^c_\omega X^s_\omega = 0,\ \omega \neq \omega'. \qquad (4.12)$$

Корреляционные свойства $\dot{X}_\omega$, с учетом (4.12) приводятся к виду

$$\overline{M}\left[\sum c_i (X^c_{\omega_i})^2 + \sum d_j (X^s_{\omega_j})^2\right] =$$

$$= \sup_{B(\omega)} \left[\sum c_i B(\omega_i) + \sum d_j B(\omega_j)\right],$$

где $B(\omega)$ способствуют собственному семейству ковариаций. Отсюда

$$\overline{M}\left[(X^c_\omega)^2 - (X^s_\omega)^2\right] = \underline{M}\left[(X^c_\omega)^2 - (X^s_\omega)^2\right] = 0. \qquad (4.13)$$

Из полученных уравнений видно, что для задания модели спектрального процесса достаточно задать либо $\overline{M}\sum c_i (X^c_{\omega_i})^2$, либо $\overline{M}\sum c_i (X^s_{\omega_i})^2$, либо

$$\overline{M} \sum c_i |\dot{X}_{\omega_i}|^2 = \sup_{B(\omega)} \sum c_i B(\omega_i),\ \forall \omega_i. \qquad (4.14)$$

Доказано такое утверждение.

Спектральный двойник $\dot{X}_{\omega,T}$ скв-непрерывного стационарного в широком смысле процесса X_t с нулевым средним при $T \to \infty$ ИМ-сходится в направлении корреляционных свойств к спектральному процессу $\dot{X}_\omega$, заданному первичными значениями (4.12), (4.13), (4.14).

Констатируемая утверждением сходимость является необходимым условием стационарности в широком смысле скв-непрерывного процесса. Но не достаточным, так как спектральный двойник процесса, являющегося нестационарным лишь в некоторой локальной области времени (пример: $X_t = m_t + \xi_t$, где ξ_t стационарен, а m_t — любой урезанный вне фиксированного интервала процесс: $m_t = 0$ при $|t| > T_0$) и стационарным на остальной временной оси, также будет сходиться к $\dot{X}_\omega$.

Необходимо отметить, что спектральный процесс $\dot{X}_\omega$, определенный первичными значениями (4.12)—(4.14), в общем, не имеет аналога во временной оси, поэтому его нужно рассматривать как: 1) способ задания предельных (при $T \to \infty$) корреляционных

свойств исходного стационарного в широком смысле процесса X_t; 2) самостоятельный способ задания не существующих, но удобных в каком-то смысле процессов (как белого шума, для которого $B(\omega)=\text{const}$ и для которого аналогов во времени даже в предельном смысле нет).

Одним из основных достоинств перехода к спектральному процессу является простота и наглядность его пересчета при однородных линейных преобразованиях, о чем речь в следующем параграфе.

4.4. ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ ПРОЦЕССА

Гладкость преобразований и непрерывность процессов. *Линейными* называются преобразования вида

$$Y_t=\int_0^T h(t,\tau)\,X_\tau\,d\tau, \tag{4.15}$$

где $h(t,\tau)$ есть импульсный отклик фильтра, т. е. его реакция на дельта-скачок на входе в момент t; Y_t — процесс на выходе, если на вход поступает X_t. Считается, что реализации процесса интегрируемы с весом $h(t,\tau)$.

Многие процессы обязаны своему виду и свойствами линейным преобразованиям, вызванным характерной для систем наблюдений и измерений инерционностью среды, устройств и звеньев. И чем больше инерционность фильтрующей системы, чем медленней меняется $h(t,\tau)$ по τ, тем плавнее, глаже будут пропускаемые на ее выход реализации Y_t в силу способности такой системы нивелировать скачки и быстрые колебания входного процесса. Сказанное подтверждается неравенством

$$|Y_{t+\Delta t}-Y_t|=\left|\int_0^T [h(t+\Delta t,\tau)-h(t,\tau)]\,X_\tau\,d\tau\right|\leqslant$$

$$\leqslant(\max_\tau|X_\tau|)\int_0^T|h(t+\Delta t,\tau)-h(t,\tau)|\,d\tau,$$

из которого видно, что при ограниченном входном процессе $\max|X_\tau|\leqslant H$ скорость изменения выходного за промежуток Δt определяется тем, насколько отклик фильтра изменится, если дельта-скачок на входе сдвинуть по времени на Δt.

Поделив обе части на Δt и устремив к нулю, получим неравенство для производной

$$|dY_t/dt|\leqslant(\max|X_t|)\int_0^T|dh(t,\tau)/dt|\,d\tau.$$

Такое же неравенство верно и для старших производных.

Расчет выхода фильтра. Расчет модели процесса Y_t, определяемого в (4.15), производится строго по общей методике § 2.1. Сначала выявляются h-представимые признаки входного процесса, т. е. записываемые в виде функционалов $g\{\int h(t,\tau)X_\tau d\tau\}=gh\{X\}$. Для них рассчитываются средние $\bar{M}gh\{X\}$, и они же напрямую переносятся на аналогичные средние $\bar{M}g\{Y\}$ процесса Y_t, в совокупности своей и определяющие его модель.

Проследим этот путь на примере входного процесса второго порядка. Первичными для него являются средние линейно-квадратичных признаков, записанные в интегральном виде

$$\tilde{M}\left[\int c_t^{(k)}\,X_t\,dt+\int\int X_t\,D_{t,t'}^{(k)}\,X_{t'}dtdt'\right]=\tilde{m}_k,\quad k=1,2,\ldots,$$

или сокращенно для левой части: $\tilde{M}(\mathbf{c}^{(k)},\ \mathbf{D}^{(k)})=m_k$, k — номер первичного признака. Первичные значения и определят по формуле согласования и продолжения средние от любых пар $\bar{M}(\mathbf{c},\mathbf{D})$, составляющих все вместе корреляционные свойства. Из них h-представимые признаки, т. е. допускающие запись

$$c_t=\int h(t,\tau)\,e_\tau\,d\tau,\ D_{t,t'}=\int\int h(t,\tau)\,H_{\tau,\tau'}\,h(t',\tau')\,d\tau\,d\tau'$$

или сокращенно (по аналогии с векторной формой) $(\mathbf{c},\ \mathbf{D})=(\mathbf{he},\ \mathbf{hHh})$, определят модель выходного процесса своими средними

$$\bar{M}^Y(\mathbf{e},\mathbf{H})=\bar{M}^X(\mathbf{c},\mathbf{D}).$$

Напрашиваются следующие выводы.

1. При линейных преобразованиях процессы второго порядка переходят также в процессы второго порядка, т. е. корреляционные свойства в корреляционные свойства.

2. Если все первичные признаки входного процесса второго порядка $\mathbf{h}$-представимы, т. е. $(\mathbf{c}^{(k)},\ \mathbf{D}^{(k)})=(\mathbf{he}^{(k)},\ \mathbf{hH}^{(k)}\mathbf{h})$, $k=1,2,\ldots$, то первичными для выходного процесса будут $(\mathbf{e}^{(k)},\ \mathbf{H}^{(k)})$ с теми же средними. Тогда входной и выходной процессы будут подобными.

3. Если преобразование обратимо в смысле существования отклика $h^{-1}(t,\tau)$ обратного фильтра, определяемого уравнением

$$\int h(t,\tau)\,h^{-1}(t',\tau)\,d\tau=\delta(t-t'),$$

то преобразование наводит подобие между процессами X_t и Y_t, а первичными для Y_t будут признаки

$$e_t^{(k)}=\int h^{-1}(t,\tau)\,c_\tau^{(k)}\,d\tau,$$

$$H_{t,t'}^{(k)}=\int\int h^{-1}(t,\tau)\,D_{\tau,\tau'}^{(k)}\,h^{-1}(t',\tau')\,d\tau\,d\tau'.$$

Линейные преобразования можно изучать с помощью подчиненно-аддитивного представления: $X_t=m_t+\xi_{t,\mathbf{m}}$, где для процесса второго порядка слагаемые определяются собственными семействами средних $\mathfrak{M}$ и ковариаций $\mathfrak{B}_\mathbf{m}$ (см. (4.6)). Тогда такое же представление будет иметь выходной процесс, записываемый

$$Y_t=\int h(t,\tau)\,m_\tau\,d\tau+\int h(t,\tau)\,\xi_{\tau,\mathbf{m}}\,d\tau=n_t+\eta_{t,\mathbf{n}}$$

и рассматриваемый как отдельное прохождение через фильтр

среднего и случайной добавки, что и определит нам собственное семейство средних $\mathfrak{M}$ и ковариаций $\mathfrak{K}_n$ на выходе:

$$\mathfrak{M} = \{n_t : n_t = \int h(t, \tau)\, m_\tau\, d\tau,\ m_\tau \in \mathfrak{M}\},$$

$$\mathfrak{K}_n = \{K_n(t, t') = \int\int h(t, \tau)\, h(t', \tau')\, b(\tau, \tau')\, d\tau\, d\tau',\ b \in \mathfrak{B}_m\}.$$

Они и определят полностью выходной процесс второго порядка, по ним могут быть рассчитаны корреляционные свойства. Но расчет легче дается использованием собственных семейств входного процесса, что и продемонстрируем на примере.

Пример 4.8. Расчет границ ковариации. Пусть $X_t = m_t + \xi_t$, где $M\xi_t = 0$ и ξ_t свободен *от* m_t, в результате чего $\mathfrak{B}_m \equiv \mathfrak{B}$, $\forall m$. Пусть также отклик фильтра неотрицателен $h(t, \tau) \geqslant 0$. Тогда границы ковариаций выходного процесса рассчитываются как максимум по входным

$$\overline{K}(t, t') = \int\int h(t, \tau)\, h(t', \tau')\, \overline{b}(\tau, \tau')\, d\tau\, d\tau'.$$

Пусть однородный процесс второго порядка (для него $\overline{b}(t, t') = \overline{b}(t - t')$), имеющий конечный интервал $\tau_{кор}$ корреляции (т. е. $\overline{b}(t - t') = 0$ при $|t - t'| > \tau_{кор}$), пропускается через инерционный фильтр такой, что отклик неотрицателен $h(t, \tau) \geqslant 0$ и как функция переменной τ мало меняется за $\tau_{кор}$: $h(t, \tau + \tau_{кор}) \approx h(t, \tau)$. Тогда границы ковариаций выходного процесса будут полностью приобретать черты линейного звена, как это видно из следующего упрощения предыдущего выражения:

$$\overline{K}(t, t') = \int\int h(t, \tau)\, h(t', \tau + \Delta)\, \overline{b}(\Delta)\, d\tau\, d\Delta \approx$$

$$\approx \int \overline{b}(\Delta)\, d\Delta \int h(t, \tau)\, h(t', \tau)\, d\tau = \overline{\gamma}^2 K_0(t, t'),$$

где $K_0(t, t')$ определяется исключительно откликом фильтра. Можно было бы представить $Y_t = \gamma Y^0_t$, где множитель γ произволен в интервале $(\underline{\gamma}, \overline{\gamma})$, а Y^0_t имеет нулевое среднее и точную ковариацию $K_0(t, t')$. Однако это представление верно лишь для расчета границ ковариаций и отнюдь не означает, что собственное семейство сужается до $\gamma^2 K_0(t, \tau)$ (как это было бы, если бы входной процесс был стационарен).

Назовем *фильтр однородным*, если $h(t, \tau) = h(t - \tau)$, т. е. его реакция с точностью до сдвига одинакова вне зависимости от момента поступления входа. Нетрудно видеть, что *фильтрация однородного процесса второго порядка однородным фильтром снова ведет к однородному процессу.*

Линейное преобразование и представление стационарного процесса. Здесь считаем, что входной процесс стационарный второго порядка и $MX_t = 0$. Тогда собственные семейства составляют ковариации, зависящие только от разности аргументов: $b(t - t')$. Очевидно, что если фильтр однородный $h(t, \tau) = h(t - \tau)$, то и выходное множество ковариаций будет таким же, зависящим лишь от разности аргументов:

$$K(t - t') = \int h(t - \tau)\, h(t' - \tau')\, b(\tau - \tau')\, d\tau\, d\tau'.$$

Таким образом, *процесс остается стационарным, если его пропустить через однородный фильтр.*

Рассмотрим по образу и подобию примера 4.8 прохождение стационарного широкополосного процесса через узкополосный фильтр. Последнее подразумевает, что за интервал, на котором $b(\tau - \tau')$, принимают ненулевые значения, $h(t)$ меняется слабо, откуда получаем:

$$K(t - t') \simeq \int b(\Delta)\, d\Delta \int h(t - \tau)\, h(t' - \tau)\, d\tau = \gamma^2 K_0(t - t'). \quad (4.16)$$

Написанное позволяет следующим образом представить выходной процесс: $Y_t = \gamma Y^0_t$, где $MY^0_t = 0$, $MY^0_t Y^0_{t'} = K_0(t - t')$ определяется исключительно фильтром, а входной процесс влияет лишь на множитель γ. В итоге выходной процесс имеет точную ковариацию, определяемую фильтром.

Сказанное особенно станет наглядным, если перейти от X_t к его предельному спектральному двойнику $\dot{X}_\omega$, смысл введения которого во многом и состоял в упрощении расчетов спектральных и, следовательно, корреляционных свойств выходного процесса. В самом деле, предельный спектральный двойник на выходе находится простым умножением входного на частотную характеристику фильтра:

$$\dot{Y}_\omega = \dot{H}_\omega \dot{X}_\omega,\ \dot{H}_\omega = \int_{-\infty}^{\infty} h(\tau) \exp(j\omega\tau)\, d\tau.$$

Как и для всякого стационарного процесса, выходной спектральный процесс задается некоррелированностью его действительной и мнимой составляющих (4.12), равенством их мощностей (4.13), а его особенности (4.14) отражаются следующими корреляционными свойствами (записанными в интегральном виде):

$$\overline{M} \int_{-\infty}^{\infty} c_\omega |\dot{Y}_\omega|^2\, d\omega = \overline{M} \int_{-\infty}^{\infty} c_\omega |\dot{H}_\omega|^2\, |\dot{X}_\omega|^2\, d\omega.$$

С позиций собственных ковариаций правая часть равна

$$2 \sup_{B(\omega)} \int_{-\infty}^{\infty} c_\omega |\dot{H}_\omega|^2 B(\omega)\, d\omega,$$

где $B(\omega)$ — преобразование Фурье от $b(\tau)$ и супремум берется по соответствующим собственному семейству спектрам. Например, если X_t был определен границами $\overline{B}(\omega)$ энергетического спектра, то Y_t будет определен границами $|\overline{\dot{H}}_\omega|^2 \overline{B}(\omega)$.

Нетрудно видеть, что если $B(\omega)$ — широкополосный спектр, а $|\dot{H}_\omega|$ — узок, то $|\dot{H}_\omega|^2 B(\omega) \simeq B(\omega_0) |\dot{H}_\omega|^2 = \gamma^2 |\dot{H}_\omega|$, где ω_0 — средняя частота «настройки» фильтра, т. е. выходные свойства определяются только видом фильтра с точностью до энергетического множителя.

Любой стационарный процесс второго порядка (с нулевым средним) может быть представлен как результат линейного преобразования некоторого «стандартного» процесса $Z(t)$, имеющего

заданный ненулевой при каждом ω спектр $B_0(\omega)>0$, с помощью фильтра. В самом деле, выходной процесс будет определяться следующим семейством энергетических спектров:

$$\mathfrak{K}_\omega = \{K(\omega) = |\dot{H}_\omega|^2 B_0(\omega),\ \dot{H}_\omega \in \dot{\mathfrak{H}}\},$$

и всегда можно подобрать такое множество $\dot{\mathfrak{H}}$ частотных характеристик фильтра, чтобы получить заданное собственное семейство $\mathfrak{B}$.

Стандартным может быть любой формальный спектральный процесс $\dot{Z}_\omega$, заданный своим энергетическим спектром. В частном случае, если им является «*белый шум*»: $B_0(\omega)=b_0$, то $|H_\omega|^2 = K(\omega)/b_0$ и тогда с точностью до постоянного коэффициента семейство $\mathfrak{B}_\omega$ идентично $\dot{\mathfrak{H}}$. Отметим, что «белый шум» как спектральный процесс $\dot{Z}_\omega$ не только не имеет аналога во временной области, но не является в пределе спектральным двойником никакого реального процесса X_t. Однако он определен сам по себе в частотной области и дает удобную форму для представлений других процессов.

Узкополосные процессы. Получаются при фильтрации широкополосного стационарного процесса узкополосным фильтром, настроенным на среднюю частоту ω_0. Импульсный отклик фильтра с узкой полосой частот пропускания записывается

$$h(t-t') = H(t-t')\cos[\omega_0(t-t')+\varphi],$$

где $H(\tau)$ — огибающая, медленно меняющаяся по сравнению с периодом $2\pi/\omega_0$ частоты настройки. Подстановка отклика в (4.16) убеждает, что все собственные выходные ковариации обязательно приближаются при сужении полосы фильтра к одному виду

$$K(\tau) \simeq \frac{1}{2}\gamma^2 \int H(t)\,H(t+\tau)\,dt \cos\omega_0\tau = \gamma^2 K_0(\tau)\cos\omega_0\tau,$$

включающему колебательный множитель. А поскольку точно такими же ковариациями обладает представление

$$Y_t = \eta_t^s \sin\omega_0 t + \eta_t^c \cos\omega_0 t,$$

где η^c_t и η^s_t, называемые когерентной и квадратурной составляющими процесса, некоррелированы, стационарны, имеют нулевые средние и ковариацию $\gamma^2 K_0(\tau)$ приходим к выводу, что любой узкополосный стационарный процесс в комплексной форме записывается

$$\dot{Y}_t = \dot{\eta}_t \exp(-j\omega_0 t),\ \dot{\eta}_t = \eta_t^c + j\eta_t^s,$$

где $\dot{\eta}_t$ — комплексная огибающая процесса, а сам процесс Y_t равен действительной части $\dot{Y}_t$: $Y_t = \mathrm{Re}\,\dot{Y}_t$. Такое представление широко используется в статистической радиотехнике и при обработке узкополосных сигналов.

При переходе к спектральным процессам указанная запись запись превращается в $\dot{Y}_\omega = \dot{\eta}_{\omega-\omega_0}$.

4.5. ЗАКЛЮЧЕНИЕ

Процесс — это явление, исходами которого являются реализации непрерывного времени. Современная теория знает разные подходы к описаниям случайных процессов. Самый общий, состоящий в задании процесса согласованными между собой многомерными распределениями вероятностей, оказывается сложным, неэффективным и труднопримени́мым. Исключение, пожалуй, составляют нормальные процессы, и то из-за их близости к другому подходу — определению процесса его корреляционными свойствами. Корреляционный подход прост в понимании, полагается на физическую природу инерционности процессов, интерпретируется через спектр и находит широкое применение во всех областях. Еще один подход состоит в использовании функциональных преобразований стандартных процессов, обычно белого шума (так задаются диффузионные процессы), и занимает промежуточное положение между указанными двумя. Всем классическим подходам свойственны абсолютно завершенные по детализации конструкции: если известны вероятности, то вся совокупность, то же с корреляциями. Все же в самом существовании корреляционного подхода и его успехах усматривается тенденция вынужденного отхода от абсолюта в сторону упрощений, ибо корреляции суть лишь составная часть, толика всего необозримого арсенала вероятностных свойств.

Требование дальнейших упрощений вызывает необходимость открыть простор любым частичным, сокращенным описаниям, заданиям процесса его отдельными свойствами, и не обязательно в точном, а можно в размытом, интервальном виде. Незавершенные для классической теории, такие конструкции оказываются совершенно законченными и строгими, даже естественными, для интервальных моделей, где любые вероятности, корреляции, моменты (точные или интервальные) как фрагменты средних, если их принять за первичные, уже как-то определяют процесс, причем чем в меньшем числе, тем проще.

Сохранение обязательных, наиболее видимых, характерных черт и «забывание» всех второстепенных доводит сложность описаний процесса до уровня, мало отличающегося от моделей явлений с простыми исходами (типа дискретных и непрерывных случайных величин, последовательностей) и к ним нередко сводится. А для этого требуется направленный подбор признаков, которыми служат функционалы на пространстве реализаций, и задание их средних. Таковыми для импульсной помехи могут быть вероятности превышений одного или сетки уровней. Для процесса, рожденного инерционным устройством, характерными являются частично известные корреляции, интервал корреляции, свойства непрерывности реализаций и т. п. Любые желаемые черты при соответствующем навыке переводятся на язык первичных средних.

Новый подход требует пересмотра некоторых положений современной теории. Так, не всякий процесс записывается как сумма его среднего и остатка, а верны более общие аддитивные представления (конец § 4.1). Необычно определяется ковариация, если среднее интервально. Используются и известные приемы. Так, процесс, заданный корреляционными свойствами (второго порядка), представляется семействами точных собственных средних и ковариационных функций.

Упрощение структуры описаний может достигаться за счет неизменности свойств процесса во времени, что позволяет, задавая средние при одном начале отсчета, перенести их по сдвигу на все остальные, преумножая тем самым

число первичных данных. Сказанное охватывается понятием однородности процесса как инвариантности во времени внешнего облика в виде первичных средних, а отсюда и всех остальных. Значительно более тонким и сложным оказывается понятие стационарности как сохранность во времени внутренней, подчас незримой микроструктуры модели процесса.

Стационарность позволяет перенести процесс в спектральную область, и не столько перенести (так как это можно сделать и для других процессов), сколько выделить простые свойства спектра, особенно полезные в задачах стационарной фильтрации и при описании процессов через спектральные двойники. Спектральные описания в такой степени автономны, что позволяют задавать вырожденные процессы типа белого шума, не являющиеся ничьими двойниками, но крайне удобные для представлений свойств других, вполне реальных процессов.

Глава 5.
ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ

5.1. СТАТИСТИЧЕСКИЕ МОДЕЛИ

Что такое математическая статистика? В любой деятельности, будь то производство, социальная сфера или быт, приходится принимать решения. В последнее понятие вкладывается научный смысл, причем значительно более объемный, чем обиходный. Считается решением, во-первых, ответ на вопрос, каково значение интересующего нас параметра, определяющего положение объекта или состояние системы, тогда имеем задачу оценивания параметров. Например, определение дальности до цели. Проблемой оценивания охватываются решения, преподносимые в форме интервала (интервальные оценки) и вообще в виде нечеткого события (расплывчатые оценки). Такое оценивание называется доверительным, когда расплывчатость служит гарантом надежности.

Во-вторых, решение — это непосредственное выделение значений некоторой физической величины в ее течении во времени, тогда имеем задачу фильтрации.

В-третьих, решением будет выбор одной из двух гипотез типа есть цель или нет, параметр ноль или не ноль, быть или не быть и пр. и пр.

В-четвертых, это проверка одной из многих гипотез, например, в каком из нескольких рабочих (частотных) каналов присутствует сигнал, кто из представленных для опознания есть преступник, на какую отметку знает студент, сдающий экзамен, и т. д. Если это есть выбор одного из дискретного набора значений числового параметра, то при сближении дискретов и, соответственно, увеличении их числа задача проверки гипотез сближается с оцениванием, так как решение будет все больше сводиться к выбору конкретного значения параметра.

Таким образом, общая проблема принятия решений распадается на предметные области, которые между собой тесно соприкасаются.

Что же необходимо для принятия решений? Пищу для решений дают наблюдения, так или иначе связывающиеся с интересующими нас параметрами состояния системы. Если искомые параметры можно наблюдать либо измерять напрямую, то проблемы нет и мы получаем абсолютно точный результат. Лучшего не может быть. Труднее, если наблюдения косвенные, искажены погрешностями, помехами, измерения неточные и результат завуалирован шумами. Тут-то и возникают потребности в статистических методах.

Само по себе прилагательное «статистический» означает, что используются усредненные по многим наблюдениям данные, своего рода собирательный среднестатистический опыт. И новизна нашего подхода по сравнению с классическим в том, что оформляется этот опыт в виде интервальных моделей средних, что позволяет охватить практически самый разноликий статистический материал в его бедности и богатстве, с учетом формы, объема, неопределенности и степени доверия к нему.

Сами решения в статистических методах имеют статистическую окраску: они не обязаны быть совсем точными всегда, раз это невозможно сделать однажды, но в среднем должны приводить к наилучшему результату. Это и есть основная задача статистических методов — оптимальный синтез, которому посвящена вся вторая часть книги.

Математическая статистика — наука анализа решающих правил и их синтеза — имеет давние традиции, корнями уходящими в историю теории вероятностей. Росли эти две науки вместе, подтягивая друг друга. Математическая статистика давала пищу теории вероятностей требованием освоения новых моделей, развивая для них методы. И в результате бурного совместного роста, обязанного XX веку, мы оказались перед поразительным разнообразием моделей и методов. Доходило дело до абсурда, когда исследователи сначала придумывали на основе здравого смысла правила, а затем «наводили на них наукообразие», подыскивая модели, для которых эти правила оптимальны (если это и есть способ оправдания, то лишь своего существования). А потребителям ничего не оставалось, как верить или делать вид, что верят, следуя известной сказке про голого короля.

Корни подобных абсурдов лежат в том беспрекословном подчинении, незримом фатализме, с которым выбор модели обуславливает метод синтеза и в итоге вид оптимального решения. И если пользоваться арсеналом точных моделей, то их кажущееся многообразие, с одной стороны, и ненадежность с другой — порождают одинаковое сомнение в вариантах выбора, делают равноценными совершенно разные модели, тем самым обезличивая оптимальные процедуры решений.

Напрашивающийся выход состоит в расширении арсенала моделей, дополнения его простыми, грубыми, надежными моделями для удовлетворения спроса такими, которым вполне можно и нужно доверять. Не набирать каждый раз модели как семейства точных, поскольку это долгий путь, а иметь готовые образцы на все случаи жизни — вот наша цель! Принципиально то, что в реальных задачах данных всегда конечное число и они не могут быть абсолютно точными. Именно таковыми являются основные развиваемые нами модели, обретая потенциальную надежность в ущерб утерянной точности. И именно в этом заложен смысл подготовленных нами алгоритмических методов, ориентированных на конечное число данных, а при неограниченном увеличении

рассыпающихся в «фейерверк» современных аналитических методов (так или иначе находящих разумное свое обоснование в рамках предлагаемого общего подхода).

Новые модели пригодны для любых «климатических» условий: «переносят» как изобилие, так и дефицит исходных статистических данных, работают в условиях статистической неустойчивости (отраженной в интервальных средних), а также при частичном и полном отсутствии статистических данных. При этом в рамках индикаторных моделей интервальные средние могут подменяться указаниями интервалов, допусков на наблюдения, **приближая нас** к интервальному анализу, и в этом плане теория еще ждет своего развития.

Статистические интервальные модели. Приступим к строгой математической формулировке проблематики. Задача состоит в обработке наблюдений $y \in \mathcal{Y}$ с целью подготовки данных и вынесения решений относительно состояний $x \in \mathcal{X}$ объекта или явления. Примером состояний может быть наличие либо отсутствие сигнала, скрытого шума, направление или дальность до цели и т. д. Переменную x будем называть *параметром состояний*. Область $\mathcal{X}$ значений x, в общем, весьма произвольна: дискретное множество, векторное или функциональное пространство и т. п.

Предмет математической статистики возникает тогда, когда прямое наблюдение за состоянием x либо невозможно, либо затруднено наличием внутренних или внешних случайностей. Чтобы задача имела смысл, от состояний x должны зависеть свойства наблюдений $y \in \mathcal{Y}$, и тогда y будет описываться не одной, а семейством моделей $\mathcal{M}^y{}_x$, $x \in \mathcal{X}$. Это есть переходные модели, эквивалентные некоторому случайному оператору Q (каналу), в соответствии с которым $\mathcal{X} \xrightarrow{Q} \mathcal{Y}$.

Само состояние x также, в общем, априори описывается моделью $\mathcal{M}^x$, так что $\mathcal{M}^y = Q\mathcal{M}^x$. Произведение $\mathcal{M}^{xy} = \mathcal{M}^x \mathcal{M}^y{}_x$ дает совместное математическое описание исходов на $\mathcal{Y}$ и значений интересующего нас параметра x состояний. Совместная $\mathcal{M}^{xy}$ называется *статистической интервальной моделью* (СИМ).

Статистическая интервальная модель $\mathcal{M}^{xy}$ в зависимости от того, распадается она на произведение $\mathcal{M}^x \mathcal{M}^y{}_x$ или нет, называется соответственно *разложимой и неразложимой*. Различие между ними состоит в способе формирования совместной модели. Для неразложимых это делается с помощью первичных средних $\bar{M}g(x, y)$, $g \in \mathcal{G}$, содержащих данные о совместном поведении x и y. Эти средние могут быть найдены экспериментально, когда состояния x находятся вне нашего влияния, так что можно лишь пассивно следить за значениями x и сопровождающими их реализациями y. Разложимые модели являются результатом анализа поведения наблюдений y при каждом $x \in \mathcal{X}$ в отдельности. Они будут иметь место тогда, когда на этапе формирования модели можно управлять значениями x состояний. Поясним разницу между моделями на примерах.

Пример 5.1. Пусть синтез модели осуществляется на основании повторного совместного наблюдения реализаций $x^{(n)}$ и $y^{(n)}$, $n=1, \ldots, N$, при этом собираются сведения о средних значениях признаков $g(x, y)\in\mathcal{G}$. Далее усреднением и выставлением доверительных границ находятся первичные значения

$$\underset{\sim}{M}g(x,\ y)=\frac{1}{N}\sum_{n=1}^{N}g\left(x^{(n)},\ y^{(n)}\right)-\Delta;$$

$$\tilde{M}g(x,\ y)=\frac{1}{N}\sum_{n=1}^{N}g\left(x^{(n)},\ y^{(n)}\right)+\Delta.$$

В этом случае СИМ получается неразложимой. Если же имеется возможность управлять состояниями x и набирать статистику о среднем $M^{y}{}_{x}g(y)$ при каждом значении $x\in\mathcal{X}$, то будут получаться доверительные границы $\underset{\sim}{M}{}^{y}{}_{x}g(y)$, $\tilde{M}^{y}{}_{x}g(y)$, задающие переходные модели разложимой СИМ.

Пример 5.2. Пусть состояний всего два: x_0 и x_1, и каждому из них соответствует своя модель $\mathcal{M}^{y}{}_{0}$ и $\mathcal{M}^{y}{}_{1}$, интерпретируемая как семейство точных распределений вероятностей $\mathcal{P}^{y}:\mathcal{M}^{y}{}_{x}=\bigvee\limits_{\mathcal{P}^{y}\in\mathcal{M}^{y}_{x}}\mathcal{P}^{y}$, $x=x_0$ или x_1. Тогда если $\mathcal{P}^{y}$ могут выбираться произвольно внутри $\mathcal{M}^{y}{}_{x}$ вне связи с тем, равно $x=x_0$ или x_1, тогда

$$\overline{M}f(x,\ y)=\overline{M}^{x}\left[\delta_{x_0}(x)\cdot\sup_{\mathcal{P}\subset\mathcal{M}^{y}_{0}}\overline{M}^{y}_{\mathcal{P}}f(x,\ y)+\delta_{x_1}(x)\sup_{\mathcal{P}\subset\mathcal{M}^{y}_{1}}\overline{M}^{y}_{\mathcal{P}}f(x,\ y)\right]$$

и имеем разложимую СИМ, причем $\mathcal{M}^{y}{}_{x_0}=\mathcal{M}^{y_0}$, $\mathcal{M}^{y}{}_{x_1}=\mathcal{M}^{y_1}$. Если же $\mathcal{P}^{y_1}$ из $\mathcal{M}^{y_1}$ каким-то образом подчинены $\mathcal{P}^{y_0}$ из $\mathcal{M}^{y_0}$, так что выбор одного распределения вынуждает вид другого, то, обозначая связь $\mathcal{P}^{y_1}=S\mathcal{P}^{y_0}$, получаем

$$\overline{M}f(x,\ y)=\sup_{\mathcal{P}_0\in\mathcal{M}^{y}_{0}}\overline{M}^{x}\left[\delta_{x_0}(x)\,\overline{M}^{y}_{\mathcal{P}_0}f(x,\ y)+\delta_{x_1}(x)\,\overline{M}^{y}_{S\mathcal{P}_0}f(x,\ y)\right].$$

В этом случае СИМ сужается и становится неразложимой.

Параметрами состояний могут быть средние характеристики случайного объекта или явления, скажем $x=Mq(y)$, где q — некоторая функция. Тогда $\mathcal{M}^{y}{}_{x}$ суть Mq-сечения из представления:

$$\mathcal{M}^{y}=\bigvee_{x\in\mathcal{X}}\mathcal{M}^{y}_{x},\ \text{где}\ \mathcal{M}^{y}_{x}=\mathcal{M}^{y}\wedge\langle Mq=x\rangle$$

и $\mathcal{X}=\{x:\mathcal{M}^{y}{}_{x}\neq\varnothing\}$. Модель разлагается на произведение $\mathcal{M}^{xy}=\mathcal{J}^{x}\mathcal{M}^{y}{}_{x}$, где $\mathcal{J}^{x}$ есть голая на $\mathcal{X}$ модель. Это произведение эквивалентно следующему порядку вычисления средних: $\overline{M}f(x,\ y)=\sup\limits_{x}\overline{M}^{y}{}_{x}f(x,\ y)$, и соответствует полному отсутствию априорных данных о x.

Одним из способов возможных упрощений СИМ является ее расширение, так как этот способ видоизменения СИМ не ухудшает ее надежности (а скорее наоборот, в отличие от сужения). Например, расширением неразложимая СИМ может быть приведена к разложимой. Вообще расширением можно свести СИМ к любой из заранее выбранных упрощенных форм. Вопрос только в том, к каким потерям точности это приведет, так как при неудачно выбранной форме приведенной СИМ расширение может обратиться в «раздевание» вплоть до голой модели $\mathcal{J}^{xy}$, следовательно, к потере любых данных как относительно x, так и y. Приведение возможно и к другим формам, о которых пойдет речь, и задача инженера-исследователя — подобрать такую, какая ближе всего стоит к решаемой задаче с целью составить наиболее экономное описание.

Функциональные представления наблюдений. Одной из форм задания СИМ является функциональное представление вида

$$y=V_x\xi,\ x\in\mathcal{X},\ y\in\mathcal{Y},\ \xi\in\Xi,\qquad(5.1)$$

где ξ — некоторое случайное явление, называемое флуктуациями (либо шумом, помехой), а V есть оператор связи y с x и ξ, отображающий произведение пространств $\mathcal{X}\times\Xi$ в $\mathcal{Y}$. В этом случае достаточно знать совместную модель $\mathcal{M}^{x\xi}$, которая вместе с V однозначно определит СИМ $\mathcal{M}^{xy}$. Преимущества представления (5.1) будут ощутимы лишь тогда, когда простой по форме является $\mathcal{M}^{x\xi}$ (или без особых потерь в расширении ее удастся упростить). Например, когда $\mathcal{M}^{x\xi}=\mathcal{M}^{x}\mathcal{M}^{\xi}$, т. е. флуктуации ξ свободны от x.

Можно заранее выбрать оператор V и расширением свести СИМ к виду (5.1). Тогда возникает вопрос о правильном выборе V, извлекающим из ξ все отличительное, что только требуется для y.

Модели с мешающими параметрами. В определенных задачах выделяют так называемые мешающие параметры $\theta\in\Theta$, считая СИМ $\mathcal{M}_{\theta}{}^{xy}$ подчиненной θ. Если θ может быть любым внутри Θ, так что априори его модель является голой $\mathcal{M}^{\theta}=\mathcal{J}^{\theta}$, то введение θ эквивалентно представлению СИМ в виде объединения $\mathcal{M}^{xy}=\bigvee\limits_{\theta}\mathcal{M}_{\theta}{}^{xy}$. При этом СИМ будет также частной к произведению $\mathcal{M}^{\theta xy}=\mathcal{J}^{\theta}\,\mathcal{M}_{\theta}{}^{xy}$.

Мешающий параметр может входить в функциональное представление $y=V_{\theta,x}\xi$, где он отражает неизвестные данные об операторе V. Это представление имеет смысл только вместе с упрощающими предположениями о совместной модели θ, x и ξ, что рассматривается в примере.

Пример 5.3. Пусть имеет место представление $y_t=\theta^{+}{}_{t}(x_t+\xi_t)$, где x_t есть сигнал, передаваемый по каналу связи; ξ_t — аддитивный шум. Здесь мешающий параметр $\theta^{+}{}_{t}\geqslant 0$ отражает замирания в канале. В случае независимости совместная модель запишется $\mathcal{M}^{\theta x\xi}=\mathcal{M}^{\theta}\times\mathcal{M}^{x}\times\mathcal{M}^{\xi}$. Если сведения о независимости нет, но имеет место свобода ξ от θ и x (означающая, что шум способен в некоторой мере «подстраиваться» под значения θ и x в рамках заданной $\mathcal{M}^{\xi}$) и свобода θ от x, то: $\mathcal{M}^{\theta x\xi}=\mathcal{M}^{\theta}\mathcal{M}^{x}\mathcal{M}^{\xi}$. Наконец, возможен и другой вариант, когда $\mathcal{M}^{\theta x\xi}=\mathcal{M}^{x\xi}\mathcal{M}^{\theta}$ и здесь уже параметр θ может тактически «подстраиваться» под x и ξ в рамках $\mathcal{M}^{\theta}$.

Робастные модели. Получаются заданием или интерпретацией моделей семействами распределений вероятностей, оформленны-

ми как некоторые окрестности в пространстве распределений. Так может описываться и сама по себе СИМ, и модели флуктуаций (или параметра состояний) в функциональном представлении наблюдений. Задается семейство $\mathscr{M}=\bigvee\mathscr{P}$ в виде метрических ограничений типа

$$\{\mathscr{P}: d(\mathscr{P}, \mathscr{P}_0) \leqslant \varepsilon\},$$

где d — «расстояние» от распределения вероятностей $\mathscr{P}$ до «центра» $\mathscr{P}_0$. Читается, как семейство всех распределений вероятностей $\mathscr{P}$, отстоящих от $\mathscr{P}_0$ не более, чем на ε. Разные метрики d ведут к разным семействам, разным моделям. Одно из распространенных семейств дают интервальные плотности: $\underset{\sim}{p}(z)$, $\tilde{p}(z)$ — семейство всех плотностей, располагающихся между нижней и верхней границами (здесь z заменяет переменные x, y или обе вместе). Выделим «центр» $p_0(z)=[\underset{\sim}{p}(z)+\tilde{p}(z)]/2$, он формален (это может и не быть плотность); тогда семейство $\{p(z): \underset{\sim}{p}(z) \leqslant p(z) \leqslant \tilde{p}(z)\}$ выражается метрически так:

$$\left\{p(z): \max_z \frac{|p(z)-p_0(z)|}{\tilde{p}(z)-\underset{\sim}{p}(z)} \leqslant \frac{1}{2}\right\}.$$

Частные варианты интервальной плотности, соответствующие $\tilde{p}(z)\equiv\infty$, дает семейство «засоренных» [11] распределений, эквивалентно представляемое: $(1-\varepsilon)p_0(z)+\varepsilon p_\forall(z)$, где $p_0(z)$ — заданная, $p_\forall(z)$ — совершенно неизвестная плотность, и тогда $\underset{\sim}{p}(z)=(1-\varepsilon)p_0(z)$. Семейство записывается через функции распределения: $(1-\varepsilon)F_0(z)+\varepsilon F_\forall(z)$, где $F_\forall(z)$ — произвольна. Понимается такая модель так, как будто с вероятностью $(1-\varepsilon)$ случайный исход z подчиняется заданному «чистому» распределению вероятностей, соответствующему $F_0(z)$ (или $p_0(z)$), а с вероятностью ε может быть все, что угодно, что и вызывает «засорение» чистых знаний и появление семейства.

Робастный подход имеет поддержку в виде теоремы 1.3 о представлении, по которой модели интерпретируются как семейства распределений вероятностей. И этот подход мог бы быть универсальным, если бы не огромные трудности, встающие на пути описания метрическими ограничениями самых разнообразных семейств и свойств, особенно таких, как зависимость между случайными величинами внутри последовательности или процесса. Угроза получить громоздкий ком описаний вместе с вытекающими отсюда трудностями синтеза очень ограничивает сферу действия робастных моделей и методов.

5.2. ОПТИМАЛЬНЫЕ ПРАВИЛА

Расплывчатые решения и решающие правила. Целью статистических методов является вынесение решений относительно состояний $x\in\mathscr{X}$ по наблюдениям y. Каждое решение есть некоторое суждение о состояниях. Это суждение может выражаться в детерминированной форме в виде указания конкретного элемента $\hat{x}$ пространства $\mathscr{X}$, так и в нечеткой форме в виде подмножеств $\mathscr{X}$, или в виде нечетких событий $\underline{q}(x)$, $\bar{q}(x)$ (с. 88), где специфично для решений границы обязаны совпадать и обозначаются $d(x)$.

Решениями называются любые нечеткие события на пространстве состояний $\mathscr{X}$, описываемые функциями $d(x)$, $x\in\mathscr{X}$, такими, что $0\leqslant d(x)\leqslant 1$. Функция $d(x)$ — изображение мнения относительно возможных x, выраженного в виде некоторой кривой предпочтений разным значениям x по шкале $[0; 1]$. Это же при каждом x и степень уверенности, даже вероятности, с которой элемент x включается как возможный представитель решения.

Множество всех возможных решений (всех функций $0\leqslant d(x)\leqslant 1$ на $\mathscr{X}$) обозначим $D_\forall$. В $D_\forall$ входит решение $d(x)\equiv 1$, соответствующее фразе: «Какое-то состояние на $\mathscr{X}$ имеет место». Туда же входят *индикаторные решения* $d(x)=A(x)$, $A\subset\mathscr{X}$, соответствующие фразе: «Имеет место какое-то одно состояние из множества A на $\mathscr{X}$». В случае $\mathscr{X}=\mathscr{R}$, когда множество A есть интервал прямой: $A=[a, b]$, индикаторное решение называется *интервальным решением*. Решение в виде дельта-функции: $d(x)=\delta_{\hat{x}}(x)$, состоящее в указании одного конкретного состояния $\hat{x}$, называется *детерминированным*. Наконец, решение $d(x)\equiv 0$ соответствует тому, что никакое из состояний $\mathscr{X}$ не имеет места.

Удобно выдавать 0 за белое, 1 — за абсолютно черное, и воображать себе в общем $d(x)$, как размазанное пятно переменной контрастности на белом фоне, своего рода как нечеткое изображение цели на экране осциллографа.

Величину $\sup d(x)-\inf d(x)$, равную высоте $d(x)$, будем называть *контрастностью решения*, а решение, принимающее хотя бы раз как значение 0, так и значение 1 — *контрастным*. Контрастность, это в некотором смысле несомненность, уверенность решений. Множество контрастных решений обозначим D_{01}: $D_{01}=\{d(x): \inf d(x)=0, \sup d(x)=1\}$. В D_{01} входят как все детерминированные решения $D_{\text{дет}}$, так и все индикаторные (интервальные) $D_{\text{и}}$. На этом основании верно включение: $D_{\text{дет}}\subset D_{\text{и}}\subset D_{01}\subset D_\forall$.

Вернемся к множеству $D_\forall$ всех решений. Оно замкнуто относительно логических операций: «не $d(x)$» $\Leftrightarrow 1-d(x)$; «$d_1(x)$ или $d_2(x)$» $\Leftrightarrow d(x)=\max\{d_1(x), d_2(x)\}$; «$d_1(x)$ и $d_2(x)$» $\Leftrightarrow d(x)=\min\{d_1(x), d_2(x)\}$ (то же можно сказать относительно $D_{\text{и}}$). Множество $D_\forall$ замкнуто и относительно рандомизации: «выбор $d_i(x)$ с вероятностями p_i, $\sum p_i=1$» $\Leftrightarrow d(x)=\sum p_i d_i(x)$. Рандомизацией выражаются решения, высказанные в виде сомнения. Так, фраза: «Вероятно (с вероятностью p) имеет место решение $d_*(x)$» отражается абстрактным событием $d(x)=pd_*(x)$. Если, скажем, $d_*(x)=A(x)$ — индикаторное решение, то решение $pA(x)$ будет соответствовать предложению: «Имеет место одно из состояний множества A со степенью уверенности p, и никакое из других состояний (т. е. из A^c)». Таким образом, разнообразие нечетких событий позволяет выразить разные оттенки решений.

Перейдем теперь от решений к решающим правилам. Они каждому наблюдению y указывают, какое решение при этом следовало бы принимать; т. е. полностью задают схему, процедуру принятия решений, какое бы y ни случилось. Если решения нечеткие, то правила называются *расплывчатыми*.

Чисто формально решающее правило $\partial_y(x)$ есть отображение пространства наблюдений $\mathcal{Y}$ в множество всех решений $D_\forall$: $\mathcal{Y} \xrightarrow{\partial_y} D_\forall$.

Мы ограничим классы правил, если заменим $D_\forall$ на некоторое его подмножество D. Такие правила каждому y ставят в соответствие решение $d(x)$ из множества D, $D \subset D_\forall$.

Правила классифицируются по виду решений. Если $D = D_{\text{дет}}$ есть множество детерминированных решений, то правило называется детерминированным; если $D = D_{\text{и}}$ есть индикаторные (интервальные) решения, то правило называется индикаторным (интервальным). Наконец, если $D = D_{01}$ — множество контрастных решений, то правило называется контрастным. Обозначения классов правил сродни решениям: $\mathcal{D}_\forall$, $\mathcal{D}_{\text{дет}}$, $\mathcal{D}_{\text{и}}$, $\mathcal{D}_{01}$.

Потери. Решение $d(x')$ принимается в конечном счете относительно состояний $x' \in \mathcal{X}$, поэтому требуется охарактеризовать его правильность, насколько оно угадывает искомое x, охватывает что ли его. Будем характеризовать противоположную величину — неправильность — с помощью *потерь* $\pi(x, d(x'))$, определяющих плату за решение $d(x')$, если на самом деле имело место состояние x. Это, в общем, функционал от нечетких решений $d(x')$, зависящий от истинного состояния x. В частном случае детерминированных решений потери $\pi(x, \hat{x})$ будут функцией двух переменных: истинного состояния x и принимаемого решения $\hat{x}$. Ниже даются примеры потерь.

Пример 5.4. Дельта-потери. Используются при детерминированных решениях и имеют вид обратного дельта-выброса: $\pi(x, x') = 1 - \delta_{\hat{x}}(x)$. Потери равны 0 при $\hat{x} = x$ (правильном решении) и равны 1 при ошибочном. Такие потери означают, что нас не волнует, какую ошибку дает решение $\hat{x}$ по отношению к истинному состоянию x, а лишь интересует сам факт, имеется ли ошибка (тогда потери равны 1) или нет (тогда 0). Эта крайняя категоричность: либо все, либо ничего — сглаживается при обобщении с детерминированных на произвольные нечеткие решения в следующем примере.

Пример 5.5. Составные потери. Пусть $\pi(x, d(x')) = 1 - d(x) + \lambda\Omega\{d(x)\}$. Потери равны величине неуверенности $1 - d(x)$, с которой решение $d(x)$ судит об истинном состоянии x плюс ущерб $\lambda\Omega\{d\}$ за расплывчатость. Ущерб пропорционален ширине и обычно есть интеграл от $d(x)$. На прямой $\mathcal{X} = \mathcal{R}$ это будет площадь под функцией $d(x)$, определяющая интегральную ширину. Для детерминированных правил ширина нулевая: $\Omega\{\hat{x}\} = 0$ — и для них составные потери совпадут с дельта-потерями. Параметр λ есть весовой коэффициент, увеличение которого повышает акцент ущерба за расплывчатость.

Пример 5.6. Квадратичные потери. Эти потери используются для детерминированных правил и равны квадрату «расстояния» между принимаемыми решениями $\hat{x}$ и истинным состоянием x. В случае $\mathcal{X} = \mathcal{R}$ потери $\pi(x, \hat{x}) = (x - \hat{x})^2$, а при многомерном параметре состояний $\mathcal{X} = \mathcal{R}^n$ — это будет квадрат длины вектора ошибки: $\pi(x, \hat{x}) = \sum_1^n (x_i - \hat{x}_i)^2$. Наконец, для процессов квадратичные потери превратятся в интеграл: $\pi(x, \hat{x}) = \int (x_t - \hat{x}_t)^2 dt$. Обобщением являются потери, определяемые в виде метрики на $\mathcal{X}$, а для линейных пространств $\mathcal{X}$ — в виде нормы: $\pi(x, \hat{x}) = \|x - \hat{x}\|^2$. Распространить такие потери на интервальные решения $d(x') = [\check{x}, \hat{x}]$ можно было бы, скажем, положив $\pi(x, d(x')) = \min\{\|x - \check{x}\|^2, \|x - \hat{x}\|^2\}$.

Потери целенаправленно выбираются исходя из того, какие решения (решающие правила) мы можем реализовать на практике, на какие их стороны следует обратить особое внимание и что желательно получить. Немаловажна и простота. Так, для задач расплывчатого оценивания удобными оказываются составные потери, а для задач фильтрации — квадратичные.

Потери сами по себе могут быть неоднозначными, если их выбор вызывает сомнения. Тогда они определяются двумя границами: $\underline{\pi}(x, d(x'))$, $\overline{\pi}(x, d(x'))$, своего рода наилучшими и наихудшими потерями. В частности, может быть задана только граница потерь сверху $\overline{\pi}$ и тогда, учитывая, что $\underline{\pi}$ неотрицательна, остается считать $\underline{\pi} \equiv 0$. При равенстве $\underline{\pi}(x, d) = \overline{\pi}(x, d)$ потери называются точными и обозначаются $\pi(x, d)$.

Риск. Текущие свойства решений оцениваются потерями, а глобальные свойства правил ∂_y (переменную x удобно опускать) как процедур принятия решений, завися от потерь, определяются существенным образом совместным поведением наблюдений y и их связью с состояниями x, т. е. видом СИМ $\mathcal{M}^{xy}$. Так как ни конкретных значений x, ни каково ожидается y на стадии синтеза правил мы не знаем, то свойства нужно характеризовать в среднем, с учетом среднестатистических изменений x и y. Для этого нужно усреднить потери и получить *нижний и верхний средние риски*:

$$\underline{\Pi}(\partial) = \underline{M}^{xy}\,\underline{\pi}(x, \partial_y), \qquad \overline{\Pi}(\partial) = \overline{M}^{xy}\,\overline{\pi}(x, \partial_y).$$

Нижний — это риск, лучше (меньше) которого быть не может, а верхний — наихудший из возможных; с одной стороны, риск оптимистичный, а с другой — риск пессимистичный. Забегая вперед, отметим, что риски могут вычисляться и по-другому, не только как средние значения, а с добавками, ну скажем, за расплывчатость правил или другие их негативные черты. Это будет уже не средний риск, а составной, о котором пойдет речь в следующей главе. Такую возможность обобщения держим всегда в уме.

В дальнейшем удобно иметь в качестве риска не два, а одно число, взвешивая между собой нижний и верхний его значения:

$$\Pi^\varkappa(\partial) = (1 - \varkappa)\,\underline{\Pi}(\partial) + \varkappa\,\overline{\Pi}(\partial), \qquad \varkappa \geqslant 0, \tag{5.2}$$

где $\varkappa$ есть *коэффициент пессимизма*. При $\varkappa=0$ и $\varkappa=1$ риск равен соответственно нижней и верхней его границам. Случай $\varkappa=0$ ведет к $\underline{\Pi}(\partial)$ и соответствует крайнему оптимизму, т. е. расчету на наилучший расклад: «авось повезет», тогда как $\varkappa=1$ зовет к $\overline{\Pi}(\partial)$ и стратегии пессимизма. Допускается $\varkappa>1$ — это *сверхпессимизм*. Значение $\varkappa=1/2$ соответствует *полуоптимизму*. Два значения $\varkappa=1$ и $\varkappa=1/2$ занимают особое положение, о чем пойдет речь дальше.

Статистическая задача. Формализуется как совокупность исходных данных в виде $\Upsilon=(\mathcal{M}^{xy}, \mathcal{D}, [\Pi], \varkappa)$, где $\mathcal{M}^{xy}$ есть СИМ; $\mathcal{D}$ — класс решающих правил, которыми мы собираемся ограничиться; $\varkappa$ — коэффициент пессимизма; $[\Pi]$ — способ вычисления риска. Для среднего риска его заменяет функция потерь, заданная точно $[\pi]$ или интервально $[\underline{\pi}, \overline{\pi}]$.

Статистическая задача может быть более и менее широкой. Для двух задач Υ_1 и Υ_2 говорим, что Υ_1 не шире Υ_2, если $\varkappa_1=\varkappa_2$, $\mathcal{D}_1=\mathcal{D}_2=\mathcal{D}$ и выполняются неравенства: $\underline{\Pi}_1(\partial)\geqslant\underline{\Pi}_2(\partial)$, $\overline{\Pi}_1(\partial)\leqslant\overline{\Pi}_2(\partial)$, $\forall\partial\in\mathcal{D}$. Тогда Υ_1 является более узкой задачей, чем Υ_2, как в случае $\mathcal{M}_1{}^{xy}\subset\mathcal{M}_2{}^{xy}$ или $[\underline{\pi}_1, \overline{\pi}_1]\subset[\underline{\pi}_2, \overline{\pi}_2]$.

Статистическая задача поставлена, если для любого решающего правила $\partial_y\in\mathcal{D}$ могут быть вычислены риски $\Pi^{\varkappa}(\partial)$. Слишком широкая статистическая задача (например, $\mathcal{M}^{xy}=\mathcal{J}^{xy}$ или $\underline{\pi}\equiv 0$, $\overline{\pi}\equiv 1$) приводит к тривиальным рискам и ее нужно сужать. Кроме того, целевое расширение статистической задачи (расширение $\mathcal{M}^{xy}$ либо $[\underline{\pi}, \overline{\pi}]$) может служить средством ее упрощения.

Классификация статистических задач. Характер статистической задачи определяется многими факторами: структурой пространств $\mathcal{X}$ и $\mathcal{Y}$, видом СИМ $\mathcal{M}^{xy}$ и т. д. Но для классификации наиболее важен вид пространства состояний $\mathcal{X}$ и какие решения D относительно состояний могут приниматься.

Если D имеет две степени свободы, т. е. любое $d\in D$ представляется как линейная комбинация $d(x)=c_1A(x)+c_2(1-A(x))$, $A\subset\mathcal{X}$, то имем задачу проверки двух гипотез: одна из них состоит в том, что $x\in A$, и альтернатива $x\notin A$. Если $d(x)=\sum c_iA_i(x)$, где множества A_i образуют разбиение $\mathcal{X}$, то имеем задачу проверки нескольких гипотез, т. е. что $x\in A_i$. К проверке гипотез всегда можно свести задачу, если $\mathcal{X}$ дискретно и состоит из конечного числа элементов, хотя это же и задача оценивания значений x, четкой грани здесь нет.

Если $\mathcal{X}$ непрерывно, например $\mathcal{X}=\mathcal{R}$, и на решения никаких ограничений не накладывается, то имеем задачу оценивания параметра. Во всех случаях числа

$$\underline{\alpha}(\partial)=\underline{M}^{xy}[1-\partial_y(x)]=1-\overline{M}^{xy}\partial_y(x),$$

$$\overline{\alpha}(\partial)=1-\underline{M}^{xy}\partial_y(x)$$

есть *нижняя и верхняя вероятности ошибки* правил. Соответственно

$$\alpha^{\varkappa}(\partial)=(1-\varkappa)\,\underline{\alpha}(\partial)+\varkappa\,\overline{\alpha}(\partial)$$

будет *взвешенной вероятностью ошибки*.

Если на класс $\mathcal{D}$ накладывается только одно требование, чтобы для всех $\partial\in\mathcal{D}$ фиксированной была вероятность ошибки $\alpha^{\varkappa}(\partial)$, то имеем задачу доверительного решения. Доверительные решения обязательно должны быть расплывчаты, так как для детерминированных оценок числового параметра ошибка, если отбросить вырожденные случаи, будет равна 1.

Детерминированные правила $\partial\in\mathcal{D}_{\text{дет}}$ записываются: $\partial_y(x)=\delta_{\hat{x}_y}(x)$, или просто $\hat{x}_y$, где $\hat{x}\in\mathcal{X}$. Алгоритм $\hat{x}_y$ есть отображение $\mathcal{Y}\to\mathcal{X}$, указывающее каждому y оценку $\hat{x}$ состояния.

Оптимальность и пессимизм. *Оптимальными называются правила $\partial^{\varkappa}{}_y$, минимизирующие риск* $\Pi^{\varkappa}(\partial)$. Они дают решение поставленной статистической задачи Υ. Для оптимальных правил риск равен

$$v(\Upsilon)=\inf_{\partial\in\mathcal{D}}\Pi^{\varkappa}(\partial)$$

и называется *ценой задачи*.

Минимум риска на классе $\mathcal{D}$ может не достигаться, и в то же время $v(\Upsilon)<\infty$. Тогда будет существовать подоптимальная последовательность $\partial_{(n)}\in\mathcal{D}$ правил, качество которых сколь угодно приближается к наилучшему: $v(\Upsilon)=\lim\limits_{n\to\infty}\Pi^{\varkappa}(\partial_{(n)})$. При любом $\varepsilon>0$ в этой последовательности можно указать такое n_ε, что $\Pi^{\varkappa}(\partial_{(n)})-v(\Upsilon)<\varepsilon$ при $n>n_\varepsilon$ и, следовательно, ущерб не будет превышать ε, сколь угодное малое.

При $\varkappa=1$ оптимальные правила (подоптимальные последовательности) минимизируют максимальный риск и называются *минимаксными*.

При $\varkappa\geqslant 1$ расширение задачи (в смысле данного выше определения) увеличивает цену, равно как и дальнейшее сверх 1 увеличение коэффициента пессимизма $\varkappa$. Чтобы сказанное стало совсем прозрачно, нужно переписать риск (5.2) в виде

$$\Pi^{\varkappa}(\partial)=\underline{\Pi}(\partial)+\varkappa[\overline{\Pi}(\partial)-\underline{\Pi}(\partial)] \qquad (5.2')$$

и обратить внимание на то, что он состоит из суммы нижнего риска плюс разброса риска (от наилучшего к наихудшему случаю), взвешенного коэффициентом пессимизма. Коэффициент $\varkappa$, таким образом, играет роль регулятора изменчивости риска оптимального правила $\partial^{\varkappa}$, так как при росте $\varkappa$ увеличивается вес второй части (5.2), что вынудит в итоге синтеза уменьшение разности $\overline{\Pi}(\partial^{\varkappa})-\underline{\Pi}(\partial^{\varkappa})$, делая риск более устойчивым в смысле разброса, более гарантированным. Значение $\varkappa=1$ свойственно пессимизму, а $\varkappa>1$ — сверхпессимизму.

Уменьшение $\varkappa$ ниже 1 приводит к принципиальным изменениям в статистической задаче. Теперь уже расширение СИМ в целях упрощения может даже уменьшить цену задачи, так как приводя к увеличению верхнего риска $\overline{\Pi}(\partial)$, в то же самое время уменьшает нижний $\underline{\Pi}(\partial)$. Причем при $\varkappa>1/2$ преобладающим в весовой сумме будет верхний риск, а при $\varkappa<1/2$ — нижний. Случай чрезмерного оптимизма $\varkappa<1/2$ вызывает противоречие: чем меньше известно (чем шире СИМ), тем лучше. Что называется: «хорошо ловить рыбу в мутной воде».

А имеет ли вообще смысл оптимизм $\varkappa<1$? Наверное да, так как пессимистичный настрой, хотя и делает правила устойчивыми, но достигает этого, поступясь ухудшением качества, разменниваясь ценою задачи. Оптимизм же при удаче даст выброс качества; напротив, при неудаче приведет к дополнительному ухудшению, усугубив положение. И вопрос в том, что лучше иметь, либо заведомо что-то гарантированное, надежное, либо что-то неустойчивое, но временами то более хорошее, то более плохое. Вопрос решается в сторону оптимизма, когда ожидаемое как гарантированное заведомо худо и удовлетворить не может. Тогда добавок оптимизма за счет усугубления перепадов в качестве оставляет надежду (свойственную игрокам) на благоприятный исход, если, конечно, повезет. Лучше надежда, чем гарантированная обреченность (поэтому в азартные игры любят играть люди малообеспеченные).

Случай $\varkappa=1/2$ полуоптимизма является особым. В нем одинаково заложен расчет на «удачу» и на «неудачу». И особенность его, как это с первого взгляда ни покажется странным, в рекомендации пользоваться именно точными распределениями вероятностей при поиске оптимальных правил.

Теорема 5.1. *Пусть СИМ определено согласованными первичными вероятностями* $\underset{\sim}{P}(x, y)$, $\tilde{P}(x, y)$ *на дискретных* $\mathscr{X}$ *и* $\mathscr{Y}$, *причем* $\sum_x\sum_y[\underset{\sim}{P}(x, y)+\tilde{P}(x, y)]=2$. *Тогда при точных потерях* $\pi(x, d_y)$ *в режиме полуоптимизма* $\varkappa=1/2$ *средний риск любого правила* d_y *равен*

$$\Pi^{1/2}(\partial)=\sum \pi(x, \partial_y) P_0(x, y), \qquad P_0(x, y)=[\underset{\sim}{P}(x, y)+\tilde{P}(x, y)]/2.$$

В самом деле, при этих условиях, как это следует из формулы (1.6) для средних $\underline{M}f$, $\overline{M}f$ у аддитивных ИРВ, имеем $\underline{M}f+\overline{M}f=2\sum_x\sum_y f(x, y)P_0(x, y)$, откуда подстановкой вместо f потерь получаем результат.

Основной вывод тот, что раз риски выражаются через точное совместное распределение вероятностей $P_0(x, y)$ (согласно условию теоремы $\sum\sum P_0(x, y)=1$), то и *поиск оптимального правила следует производить по этому точному распределению* P_0.

Теорема 5.1 один к одному переносится на произвольные пространства $\mathscr{X}$ и $\mathscr{Y}$, на произведении которых заданы интервальные плотности $\underset{\sim}{p}(x, y)$, $\tilde{p}(x, y)$ такие, что центральной (среднеарифметической) между ними $p_0=(\underset{\sim}{p}+\tilde{p})/2$ будет также плотность. Например,

$$\underset{\sim}{p}(x, y)=p_0(x, y)-\Delta(x, y), \quad \tilde{p}(x, y)=p_0(x, y)+\Delta(x, y),$$

где p_0 есть некоторое предположительное значение плотности, а Δ — ошибка в ее знании. Тогда риск $\Pi^{(1/2)}(\partial)$ вычисляется по p_0 и не будет зависеть от ошибки Δ, и следовательно, от ширины СИМ. Все будет определяться центральной плотностью $p_0(x, y)$.

Конечно же, это частный результат, соответствующий СИМ, заданной аддитивным интервальным распределением вероятностей, но он определяет статус точных вероятностных моделей при решении статистических задач, когда предположение о том или ином виде плотности или распределении вероятностей выдается за уверенность в адекватном выборе модели.

Проблема достаточности. Хорошо, когда на стадии предварительного анализа статистической задачи по ее внешним чертам сразу же возникает возможность сузить класс решающих правил до размеров, облегчающих поиск оптимального. Так сказать, создать своего рода ограждение, достаточное в плане уверенности, что оптимальное правило находится внутри него. Тогда нахождение оптимального правила приобретает двуэтапность: сначала как можно большее сужение диапазона поиска, а затем уже окончательный выбор. Сейчас речь пойдет о первом этапе, получившем название достаточной редукции.

Достаточная редукция производится по внешнему облику задачи, поэтому будет однотипно определяемой для семейств статистических задач, объединенных одинаковыми чертами, такими как: 1) особенности СИМ, число и вид первичных признаков; 2) характер $\mathscr{X}$ и $\mathscr{D}$, задающих постановочную цель решающего правила; 3) вид функции потерь; 4) область значений коэффициента пессимизма $\varkappa$. Причем тем весомее будут те или иные приемы достаточности, чем для более широкого семейства задач они пригодны.

В свете сказанного проблему достаточности будем подразделять на *глобальную*, когда редукция производится с позиций среднего риска по виду СИМ, или же только по внешнему облику функции потерь, и *специальную*, когда достаточная редукция связывается с конкретными $\mathscr{X}$ и $\mathscr{D}$, привязывается к задаче проверки гипотез или оценивания и к выбранному типу риска. В этой главе мы касаемся только глобальной проблемы, оставив специальные на последующее предметное изложение по главам.

Дадим формальное определение. Подкласс $\mathscr{D}^*$ решающих правил называется *достаточным*, если произвольному правилу $\partial\in\mathscr{D}$ при любом заданном $\varepsilon>0$ можно указать правило $\partial^*\in\mathscr{D}^*$ такое, что $\Pi^{\varkappa}(\partial^*)\leqslant\Pi^{\varkappa}(\partial)+\varepsilon$, т. е. любому правилу из исходного (соответствующего Υ) класса $\mathscr{D}$ можно указать не худшее его в смысле

риска (если только на сколь угодно малую величину ε) правило из достаточного подкласса $\mathscr{D}^*\subset\mathscr{D}$.

Достаточность и функция потерь. Следующее утверждение о достаточности связывается с неограниченными потерями. Тогда риск $\overline{\Pi}(\partial)=\bar{M}\pi(x,\partial_y)$ будет конечным лишь для тех ∂^*, при которых $\overline{\pi}(x,\partial_y)$ принадлежит области существования верхних средних СИМ. Для остальных он бесконечен, как бесконечным будет $\Pi^{\varkappa}(\partial)$ при любом $\varkappa>0$ и такие правила могут быть исключены из $\mathscr{D}$, редуцированы, что приводит к следующему выводу.

Достаточным при $\varkappa>0$ является подкласс $\mathscr{D}^*$ решающих правил, для которых потери $\overline{\pi}(x,\partial_y)$ принадлежат области существования СИМ. Так, если $\mathscr{M}^{xy}=\langle\tilde{M}\mathscr{G}\rangle$, где $g_i(x,y)\in\mathscr{G}$ — первичные признаки СИМ с заданными $\tilde{M}g_i$, то в $\mathscr{D}^*$ включаются правила ∂^*, потери которых $\overline{\pi}(x,\partial^*)$ мажорируемы хотя бы одной из конечных сумм $g(x,y)=c_0+\sum c^+{}_i g_i(x,y)$, составляющих класс $\mathscr{L}^+\mathscr{G}$ вторичных признаков, с переносом на них $\tilde{M}g=c_0+\sum c^+{}_i\tilde{M}g_i$, как это следует из общей формулы продолжения средних:

$$\overline{\Pi}(\partial)=\inf\{\tilde{M}g:\overline{\pi}(x,\partial_y)\leqslant g(x,y)\in\mathscr{L}+\mathscr{G}\}.$$

Рассмотрим пример применения сделанного утверждения.

Пример 5.7. Пусть $\mathscr{X}=\mathscr{Y}=D=\mathscr{R}_T$ есть пространства реализаций на интервале $[0,T]$ и пусть правила детерминированные $\mathscr{D}_{\text{дет}}$, а потери для них $\overline{\pi}(x,\hat{x})=\int_0^T(x_t-\hat{x}_t)^2dt$ равны среднему квадрату ошибки. Пусть СИМ $\mathscr{M}^{xy}$ задана корреляционными свойствами $\bar{M}(c+\sum c_iX_{t_i}+\sum\sum d_{ij}X_{t_i}X_{t_j})$. Тогда достаточными будут правила $\hat{x}_{t,y}$, которые при $y_t\to\infty$ растут как функция y_t не быстрее линейной степени. К ним относятся линейные: $\hat{x}_{t,y}=\int_0^T h(t,\tau)y_\tau\,d\tau$. Квадратические $\hat{x}_{t,y}=\int_0^T h(t,\tau)y^2{}_\tau d\tau$ уже не будут входить в достаточный класс, так как потери для них не мажорируемы квадратичными формами, составляющими все вторичные признаки для заданной СИМ.

Целевое расширение СИМ управляет достаточным классом, может служить эффективным инструментом упрощений. Кроме того случая, когда расширение не затрагивает верхних средних следующего класса признаков (функций x, y):

$$\mathscr{G}_\pi=\{-\underline{\pi}(x,\partial^*),\ \overline{\pi}(x,\partial^*):\partial^*\in\mathscr{D}^*\},$$

по существу, определяющего риск для всех ∂^* из достаточного класса правил.

Чтобы указать еще на одну черту достаточного класса, центрируем первичные признаки СИМ $\langle\tilde{M}\mathscr{G}\rangle$ (приведем к $\tilde{M}\overset{\circ}{g}=0$), положив для этого $\overset{\circ}{g}=g-\tilde{M}g$, $g\in\mathscr{G}$. Обозначим $\overset{\circ}{\mathscr{G}}$ — центрированный набор. Тогда согласно модифицированной формуле продолжения (1.3) нижний и верхний средние риски запишутся:

$$\underline{\Pi}(\partial)=-\inf_{g\in\mathscr{L}+\overset{\circ}{\mathscr{G}}}\sup_{x,y}[-\underline{\pi}(x,\partial_y)-g(x,y)];$$

$$\overline{\Pi}(\partial)=\inf_{g\in\mathscr{L}+\overset{\circ}{\mathscr{G}}}\sup_{x,y}[\overline{\pi}(x,\partial_y)-g(x,y)].$$

Отсюда видно, что риски в определенном смысле равны величине наилучшего приближения функции потерь (для верхнего — сверху, для нижнего — снизу) центрированными (с нулевыми $\tilde{M}\overset{\circ}{g}$) вторичными признаками СИМ из $\mathscr{L}+\overset{\circ}{\mathscr{G}}$. Достаточными будут классы таких правил, для которых $-\underline{\pi}(x,\partial_y)$ и $\overline{\pi}(x,\partial_y)$ лучше других аппроксимируются сверху функциями класса $\mathscr{L}+\overset{\circ}{\mathscr{G}}$. Причем при расширении СИМ функции этого класса смещаются в область отрицательных значений, в результате аппроксимация ухудшается и это влечет за собой увеличение риска (по крайней мере, при $\varkappa\geqslant1$).

На основании только одного вида функции потерь можно иногда сократить множество D решений, и соответственно класс решающих правил. *Множество решений D^* будет достаточным* (на D^* основывается достаточный класс правил), *если каждому $d\in D$ можно указать такое $d^*\in D^*$, что при всех $x\in\mathscr{X}$ имеет место неравенство*: $\pi(x,d^*)\leqslant\pi(x,d)$. Например, пусть решения детерминированные $D=D_{\text{дет}}$, а потери $\pi(x,x')$ принимают постоянные значения на разбиении $A_1,\ldots,A_k$ пространства $\mathscr{X}$: $\pi(x,x')=\pi_i(x)$, $x'\in A_i$. Тогда достаточным будет конечное множество $D'=\{x'_1,\ldots,x'_k\}$ решений, где x'_i есть произвольно выбранные представители множеств A_i.

От функции потерь многое зависит как в плане глобальной, так и специальной достаточности. Сам ее выбор может стимулировать принятие того или иного вида решений, например только детерминированных или только расплывчатых. Овладение этим рычагом управления — особая проблема, остающаяся вне нашего рассмотрения. Мы же в дальнейшем будем иметь дело с наболее простыми, типовыми видами потерь, для которых разрешим и не сверхгромоздок путь нахождения оптимальных правил. Сказанное относится и к более общим типам риска. На первом плане стоит, конечно же, простота синтеза и практическая разумность риска. Регулировочными становятся те параметры, которые без усложнений удается вплести в структуру риска (как это было, например, с коэффициентом пессимизма).

5.3. ДОСТАТОЧНАЯ РЕДУКЦИЯ НАБЛЮДЕНИЙ

Теорема о представимости. Здесь и ниже с позиций среднего риска освещается глобальная проблема достаточности в плане таких преобразований $z=Qy$ наблюдений y, которые, сокра-

щая y, не выводят из достаточного класса $\mathscr{D}^*$, т. е. сразу на начальном этапе без ущерба для статистической задачи совокупность наблюдений редуцируется преобразованием Q до значений z меньшей размерности. Это возможно только тогда, когда решающие правила достаточного класса $\mathscr{D}^*$ могут быть все записаны через z, т. е. Q-представимы в смысле следующей записи $\partial^*{}_z(x)=\partial^*{}_{Qy}(x)$. Преобразование Q, от которого зависит достаточный класс правил, называется *достаточным*. Оно дает сокращенные сведения о наблюдениях, вполне достаточные для построения оптимального решающего правила.

Теорема 5.2. *Преобразование* Qy *будет достаточным при* $\varkappa\geqslant 1$, *если все первичные признаки СИМ* $\mathscr{M}^{xy}=\langle\widetilde{M}\mathscr{G}\rangle$ *являются* Q-*представимыми, т. е. записываются*

$$g_j(x, y)=h_j(x, Qy),\ \forall\, g_j\in\mathscr{G}. \tag{5.3}$$

Доказательство. В самом деле, тогда и все вторичные признаки класса $\mathscr{X}+\mathscr{G}$ будут $\mathscr{G}$-представимы в смысле (5.3). Они будут постоянными на подмножествах $Q^{-1}Qy=Q^{-1}z$ пространства $\mathscr{Y}$ таких, что $Qy=z$ при фиксированных z, поэтому верхние средние одинаковы для любого выбранного признака $f(x, y)$ и Q-представимой его мажоранты $\sup f(x, y')$, где супремум берется по $y'\in Q^{-1}Qy$. Произвольному правилу ∂_y укажем Q-представимое $\partial^*{}_y$, положив его на подмножествах $Q^{-1}z=\{y: Qy=z\}=Q^{-1}Qy$ пространства $\mathscr{Y}$ равным ∂_{y_z}, где y_z — представитель $Q^{-1}z$. Тогда на основании сказанного:

$$\underline{\Pi}(\partial)=\underline{M}^{xy}\inf_{y'\in Q^{-1}Qy}\underline{\pi}(x,\ \partial_{y'})\quad\leqslant\underline{M}^{xy}\,\underline{\pi}\left(x,\ \partial^*_y\right)=\underline{\Pi}(\partial^*);$$

$$\overline{\Pi}(\partial)=\overline{M}^{xy}\sup_{y'\in Q^{-1}Qy}\overline{\pi}(x,\ \partial_{y'})\quad\geqslant\overline{M}^{xy}\,\overline{\pi}\left(x,\ \partial^*_y\right)=\overline{\Pi}(\partial^*).$$

В результате при $\varkappa\geqslant 1$ согласно (5.2) $\Pi^\varkappa(\partial)\geqslant\Pi^\varkappa(\partial^*)$, что и требовалось.

Если широко рассматривать Q как детерминированное преобразование $\mathscr{X}\times\mathscr{Y}$ в $\mathscr{X}\times\mathscr{Z}$, оставляющее элемент x на месте: $Q(x, y)=(x, Qy)$, и ведущее к преобразованию СИМ; $Q\mathscr{M}^{xy}=\mathscr{M}^{xz}$, то условия (5.3) эквивалентны тому, что Q наводит подобие (см. § 2.1): $\mathscr{M}^{xy}\backsim\mathscr{M}^{xz}$, т. е. не влечет потерь данных о СИМ: $Q^{-1}Q\mathscr{M}^{xy}=\mathscr{M}^{xy}$. Таким образом, Qy, если это *преобразование подобия СИМ, будет достаточным при* $\varkappa\geqslant 1$, *и наоборот*.

Замечания. 1. Если вовлекать широкий класс преобразований $Q(x, y)$, меняющих не только y, но и x, то утверждения о достаточности будут существенно связываться с видом потерь и в этом смысле будут специальны. Причина: соблюдение или нет нужного (для теоремы 5.2) факта, что из Q-представимости $\partial^*{}_y$ следует Q-представимость потерь $\underline{\pi}(x, \partial^*{}_y)$, $\overline{\pi}(x, \partial^*{}_y)$.

2. Как мы увидели сейчас и увидим ниже, утверждения о достаточности работают обычно при пессимизме $\varkappa\geqslant 1$, корни чего в том, что редукция «сокращает благоприятные возможности» и ущемляет диапазон свободного выбора. Кроме того случая, когда все самые благоприятные возможности остаются внутри сокращенного класса, так же как и наименее благоприятные.

Первичные признаки и достаточность. Преобразование $z=Qy$, областью значений которого является числовая прямая $z\in\mathscr{R}$, вектор $z\in\mathscr{R}^k$ или любое числовое пространство $\mathscr{Z}$, называется *числовым*. Оно эквивалентно набору признаков: $z_i=Q_i(y)$, $i=1, \ldots, k$, как функций наблюдений на $\mathscr{Y}$. Если Q достаточно, то и соответствующий ему *набор признаков наблюдений* называется *достаточным*.

Свяжем достаточный набор с первичными признаками СИМ $\mathscr{M}^{xy}=\langle\widetilde{M}\mathscr{G}\rangle$. Порядок следующий: сначала по первичным признакам $g_i(x, y)$ СИМ выделяются достаточные признаки наблюдений, ровно столько, сколько останется после исключения повторов, а уже размерность пространства $\mathscr{Z}$ сама собой подстраивается под их число.

Теорема 5.3. *Достаточными при* $\varkappa\geqslant 1$ *признаками наблюдений является набор первичных признаков СИМ* $g(x, y)\in\mathscr{G}$ *(рассматриваемых при каждом* x *как функции переменной* y*), развернутой по* x *и* g:

$$Qy=\{g(x, y): x\in\mathscr{X},\ g\in\mathscr{G}\}.$$

Утверждение станет прозрачным, если $\mathscr{X}$ дискретно и состоит из элементов $x_1, \ldots, x_k$. Тогда достаточный набор образуют следующие вектор-признаки $(g_j(x_1, y), g_j(x_2, y), \ldots, g_j(x_k, y))$ наблюдений, расположенные друг за другом в цепочку при g_j, пробегающих $\mathscr{G}$. Получается вектор-цепочка Qy признаков, общая размерность которой равна произведению числа элементов $\mathscr{X}$ на число признаков набора $\mathscr{G}$. Для уменьшения размерности повторы в этой цепочке уместно сократить.

Доказательство утверждения эквивалентно доказательству представимости каждого $g_j(x, y)$ в виде (5.3) и следует из того, что $g_j(x, y)$ есть, по сути дела, выбор j-го вектор-признака из цепочки Qy, т. е. проекция Qy в некоторое подпространство пространства функций переменной y, составляющих «оси» Qy.

Возможность дополнительной редукции возникает, когда все элементы цепочки зависят полностью от небольшого числа одних и тех же функций $q_1(y)$, $q_2(y)$, ..., которые и составят на основании (5.3) достаточный набор. Например, все они зависят от одной и той же функции, что возвращает нас к теореме 5.2.

Перейдем к модификациям достаточности для разных конкретных способов задания СИМ, применяя теорему 5.3 к соответствующим наборам признаков. Сначала используем разложение СИМ.

Следствие 1. *Пусть СИМ разложима на произведение:* $\mathscr{M}^{xy}=\mathscr{M}^x\mathscr{M}^y{}_x$, *где переходная модель* $\mathscr{M}^y{}_x$ *задается первичными средними* $\widetilde{M}_x\psi_x(y)$, $\psi_x(y)\in\Psi_x$. *Тогда преобразование в виде вектора-цепочки*

$$Qy=\{\psi_x(y): \psi_x(y)\in\Psi_x,\ x\in\mathscr{X}\}$$

будет достаточным (*при* $\varkappa\geqslant 1$).

Доказательство вытекает из того, что первичными для произведения будут $c^+(x)[\psi_x(y)-\bar{M}_x\psi]$ вместе с первичными признаками $\mathcal{M}^x$ (теорема 2.3).

Учитывая указанную выше зависимость достаточного набора от преобразования подобия, можно переформулировать следствие 1 так: преобразования подобия Q_xy для переходных $\mathcal{M}^{y}{}_{x}$, выстроенные в вектор-цепочку по x, образуют тандем $Qy=\{Q_xy : x\in\mathcal{X}\}$, являющийся преобразованием подобия СИМ (следовательно, от него зависит достаточный набор признаков наблюдений).

Перейдем к функциональному представлению наблюдений.

Следствие 2. *Пусть* $y=V_x\xi$ *и считается заданной модель* $\mathcal{M}^{\xi}=\langle\bar{M}\Psi\rangle$, *причем выполняются условия*: а) ξ *свободен от* x; б) *каждый признак* $\psi(\xi)\in\Psi$ *представляется в виде*: $\psi(\xi)=r_\psi(Q_xV_x\xi)$, *где* Q_x — *некоторые преобразования* y, *зависящие от* x. *Тогда их вектор-цепочка* $Qy=\{Q_xy : x\in\mathcal{X}\}$ *образует достаточное* (*при* $\varkappa\geqslant 1$) *преобразование наблюдений.*

В самом деле, условие а) гарантирует разложимость СИМ, так как $\bar{M}^{xy}f(x, y)=\bar{M}^x\bar{M}^{\xi}f(x, V_x\xi)=\bar{M}^x\bar{M}^{y}{}_{x}f(x, y)$, а первичными для переходных моделей на основании условия б) будут признаки $\psi(Qy)$, $\psi\in\Psi$, образующие согласно следствию 1 достаточный набор и зависящие от преобразований Q_xy, которые и будут достаточными.

Пример 5.9. Пусть $y_t=x_t+\xi_t$, $t\in[0, T]$, где шум ξ_t свободен от сигнала x_t, интегрируем и задан первичными средними $\bar{M}\psi(\int\xi_t dt)$, $\psi\in\Psi$. Выражение в круглых скобках записывается: $\int[(x_t+\xi_t)-x_t]dt=\int(y_t-x_t)dt$, что и определяет Q_xy. При всех x_t преобразование Q_xy зависит только от $\int y_t dt$, поэтому интеграл и будет достаточным при $\varkappa\geqslant 1$ признаком наблюдений, определяющим достаточную редукцию пространства реализаций $\mathcal{Y}$ в числовую прямую $\mathcal{R}$.

Рассмотрим теперь случай задания СИМ посредством мешающего параметра θ. Два очередных следствия, не требующие доказательств, являются некоторыми обобщениями предыдущих.

Следствие 3. *Пусть СИМ записывается как семейство* $\mathcal{M}^{xy}=\bigvee\mathcal{M}_\theta{}^{xy}$ *и при каждом фиксированном* θ *существует преобразование* $Q_\theta y$ *подобия для* $\mathcal{M}_\theta{}^{xy}$. *Тогда цепочка* $Qy=\{Q_\theta y, \theta\in\Theta\}$ *при* $\varkappa\geqslant 1$ *даст достаточное преобразование.*

Следствие 4. *Пусть* $y=V_{\theta,x}\xi$, *где* ξ *свободен от* x *и* θ. *Пусть существует отображение* $S\xi$ *подобия для* $\mathcal{M}^{\xi}$, *записываемое* $S\xi=Q_{\theta,x}(V_{\theta,x}\xi)$. *Тогда вектор-отображение* $Qy=\{Q_{\theta,x}y : \theta\in\Theta, x\in\mathcal{X}\}$ *будет достаточным при* $\varkappa\geqslant 1$.

Достаточные преобразования и факторизация. Давно известна связь достаточности с факторизацией плотностей распределений вероятностей [3, 4]. Здесь, следуя по этому же пути, будут получены сначала общие касающиеся интервальных моделей утверждения, аналогичные факторизации, и далее указана связь с классическими.

Теорема 5.4. О факторизации. *Пусть в исходах преобразования* $z=Qy$, *отображающего* $\mathcal{Y}$ *в* $\mathcal{Z}$, *СИМ разлагается на произведение*

$$\mathcal{M}^{xy}=\mathcal{M}^{zx}\,\mathcal{M}^{y}_{z}. \tag{5.4}$$

Тогда Qy *будет достаточным для тех статистических задач, в которых*: а) $\varkappa=1$; б) *класс* $\mathcal{D}$ *решающих правил таков, что* $d_y\in\mathcal{D}\Rightarrow\bar{M}^{y}{}_{z=Qy}d_y\in\mathcal{D}$; в) *функция потерь* $\underline{\pi}(x, d)$ *при каждом* x *выпукла относительно* $d\in D$.

В самом деле, используя факторизацию (5.4) и аналог неравенства Йенсена из § 3.2, имеем: $\underline{\Pi}(\partial)=\bar{M}^{xy}\underline{\pi}(x, d_y)=\bar{M}^{xz}\bar{M}^{y}{}_{z}\underline{\pi}(x, d_y)\geqslant\bar{M}^{xz}\underline{\pi}(x, \bar{M}^{y}{}_{z}d_y)=\bar{M}^{xz}\underline{\pi}(x, \partial^*{}_z)$, где $\partial^*{}_z=\bar{M}^{y}{}_{z}d_y$; отсюда $\underline{\Pi}(\partial)\geqslant\underline{\Pi}(\partial^*)$.

Условие б) теоремы выполнится, если множество D решений есть выпуклое подмножество числового пространства, а $\mathcal{D}=\mathcal{D}_V$ составляют всевозможные отображения $\mathcal{Y}$ в D (например, все детерминированные правила).

В дальнейшем нам понадобится тот факт, что для $\mathcal{M}^{y}{}_{z}$ фактическим пространством возможных исходов будет множество $\mathcal{Y}_z=Q^{-1}z=\{y : Qy=z\}$ всех тех y, которые преобразуются в одно z, так что $Q\mathcal{Y}_z=z$ и $\underline{P}^{y}{}_{z}(\mathcal{Y}_z)=1$.

Теорему 5.4 можно распространить на широкий диапазон значений $\varkappa$, если потребовать от средних $M^{y}{}_{z}$ точности (т. е. $\underline{M}^{y}{}_{z}=\bar{M}^{y}{}_{z}$).

Теорема 5.4а. *Пусть стоит задача* $\Upsilon=(\mathcal{M}^{xz}, \mathcal{M}^{y}{}_{z}, \mathcal{D}, [\underline{\pi}, \bar{\pi}], \varkappa)$, *где* $z=Qy$, *и переходные средние* $M^{y}{}_{z}\underline{\pi}(x, d_y)$, $M^{y}{}_{z}\bar{\pi}(x, d_y)$ *являются точными для всех* $d_y\in\mathcal{D}$, *причем* $d_y\in\mathcal{D}\Rightarrow M^{y}{}_{z}d_y\in\mathcal{D}$. *Тогда преобразование* Qy *будет достаточным для задачи* Υ *при всех* $\varkappa\leqslant 1$, *когда нижние* $\underline{\pi}(x, d)$ *и верхние* $\bar{\pi}(x, d)$ *потери при каждом* x *выпуклы как функции* d; *и будет достаточным при* $\varkappa\geqslant 1$, *когда верхние потери выпуклы, а нижние — вогнуты.*

В самом деле, если нижние потери выпуклы, тогда на основании аналога неравенства Йенсена в § 3.2 имеем: $\underline{\Pi}(\partial)=\underline{M}^{xz}M^{y}{}_{z}\underline{\pi}(x, d_y)\geqslant\underline{M}^{xz}\underline{\pi}(x, M^{y}{}_{z}d_y)=\underline{M}^{xz}\underline{\pi}(x, \partial^*{}_z)=\underline{\Pi}(\partial^*)$, поэтому при $\varkappa\leqslant 1$ будет справедливо неравенство: $\underline{\Pi}^{\varkappa}(\partial)\geqslant\underline{\Pi}^{\varkappa}(\partial)$. Наоборот, если эти потери вогнуты, то $\underline{\Pi}(\partial)\leqslant\underline{\Pi}(\partial^*)$ и при $\varkappa\geqslant 1$ имеем $\underline{\Pi}^{\varkappa}(\partial)\geqslant\underline{\Pi}^{\varkappa}(\partial^*)$.

В частности, если $\underline{\pi}(x, d)\equiv 0$, то нижние потери будут как выпуклыми, так и вогнутыми, и тогда преобразование Qy будет достаточным при любых $\varkappa$. При точных потерях π (т. е. $\underline{\pi}=\bar{\pi}$) допустим лишь случай $\varkappa\leqslant 1$, для чего потребуется выпуклость π, что характерно для классических моделей.

Пример 5.10. Пусть $\mathcal{M}^{xy}=\mathcal{P}^{xy}$ — распределение точных (на алгебре $\mathcal{A}$ пространства $\mathcal{X}\times\mathcal{Y}$) вероятностей, заданное точной совместной плотностью $p(x, y)$, и пусть эта плотность факторизуется: $p(x, y)=r(x, Q(y))h(y)$ на произ-

ведение измеримых (относительно $\mathscr{A}$) функций. Пусть $\mathscr{D}$ — всевозможные измеримые решающие правила d, а потери $\pi(x, \partial_y)$ являются точными измеримыми. Тогда средний риск будет точным, равным $\Pi(\partial)=M^{xy}\pi(x, \partial_y)=M^{xz}M^{y}{}_{z}\pi(x, \partial_y)$, где $M^{y}{}_{z}$ соответствует распределению вероятностей на $\mathscr{Y}$ (а точнее, на $\mathscr{Y}_z$) при заданном z. Условие б) теоремы 5.4 выполнится, если D — выпуклое множество числового пространства. По теореме 5.4а при выпуклых потерях $\pi(x, d)$ и выпуклом множестве D решений преобразование $Q(y)$ будет достаточным. Сказанное составляет суть известной из литературы теоремы о факторизации плотности и связанными с нею достаточными статистиками [3, 4].

Рассмотрим подробнее условие факторизуемости (5.4) с точки зрения первичных признаков $g(x, y)$ СИМ. Согласно теореме 2.3 о первичных признаках произведения моделей условие (5.4) фактически означает, что признаки $g\in\mathscr{G}$, приведенные к нулевым средним $\overset{\circ}{g}=g-\bar{M}g$, имеют вид либо 1) $g(x, Qy)-\bar{M}g$, либо 2) $c^{+}(x, Qy)[\psi(y)-\bar{M}\psi]$, причем последние, если они не тривиальны, получаются перебором всех неотрицательных $c^{+}(x, z)$ (здесь $\psi(y)$ — первичны для $\mathscr{M}^{y}{}_{z}$ вместе с $\underline{P}^{y}{}_{z}(\mathscr{Y}_z)=1$). Если же признаков вида 2) нет, что соответствует голой в области $\mathscr{Y}_z$ переходной модели $\mathscr{M}^{y}{}_{z}$, то остаются признаки 1) и теорема 5.4 факторизации подпадает под теорему 5.2 о связи достаточности с Q-представимостью признаков, что распространяет ее сферу действия на любые потери при $\varkappa\geqslant 1$.

5.4. РЕДУКЦИЯ НАБЛЮДЕНИЙ И ИНВАРИАНТНОСТЬ

Инвариантные модели. Свойство инвариантности отражает некоторые стороны симметрии СИМ подобно симметрии шара с центром в начале координат к повороту осей. Но прежде изучим само понятие инвариантности, широко используемое в настоящее время в научно-статистической литературе, и свяжем с нашим понятием достаточности.

Пусть $sy\in\mathscr{S}$ есть множество обратимых преобразований $\mathscr{Y}$ на себя. Очевидно, это множество всегда можно считать алгебраической группой. В самом деле, произведение $s_1s_2y=s_1(s_2y)$ есть снова обратимое преобразование, а тождественное преобразование $y\to y$ дает единичный элемент группы, поэтому $\mathscr{S}$ называем *группой преобразований.*

Функция $\varphi(y)$ называется *инвариантной к группе преобразований* $\mathscr{S}$ пространства $\mathscr{Y}$, если $\varphi(sy)=\varphi(y)$, $\forall s\in\mathscr{S}$. При преобразованиях группы S каждая точка y совершает движение по траектории $O_y=\{sy,\ s\in\mathscr{S}\}$, называемой *орбитой* элемента y. Две разные орбиты либо совпадают, либо не пересекаются. На каждой орбите группа $\mathscr{S}$ транзитивна в том смысле, что любой элемент этой орбиты можно преобразованием из $\mathscr{S}$ перевести в любой другой. Пространство $\mathscr{Y}$ разбивается разноименными орбитами O_y на непересекающиеся подмножества: $\mathscr{Y}=\sum_{y}O_y$.

Максимальным инвариантом относительно группы $\mathscr{S}$ преобразований пространства $\mathscr{Y}$ называется отображение $I(y)$, при котором равенство $I(y)=I(y')$ эквивалентно принадлежности y и y' одной и той же орбите. Максимальный инвариант принимает на каждой орбите постоянные значения, причем разные для разноименных орбит. Область значений максимального инварианта может лежать в весьма произвольном пространстве.

Утверждение 5.5. *Любая инвариантная к группе* $\mathscr{S}$ *функция* $\varphi(y)$ *представима через максимальный инвариант*: $\varphi(y)=\psi(I(y))$.

В самом деле, если y и y' таковы, что $I(y)=I(y')$, то $y'=sy$ для некоторого $s\in\mathscr{S}$ и поэтому $\varphi(y)=\varphi(y')$.

Для нахождения $I(y)$ может быть использован следующий факт [9]: для любой функции $f(y)$ ее инфимум $\inf_{s} f(sy)=f(y)$ (как и sup) всегда инвариантен к группе $\mathscr{S}$, а если инфимум достигается и единствен: $\min_{s} f(sy)=f(s_y(y))$, то $I(y)=s_y(y)$ является максимальным инвариантом.

Статистическая модель $\langle\bar{M}\mathscr{G}\rangle$ *называется инвариантной к группе* $\mathscr{S}$ *преобразований*, если все ее первичные признаки инвариантны к $\mathscr{S}$: $g(x, sy)=g(x, y)$, $\forall g\in\mathscr{G}$. Согласно утверждению 5.5 все первичные признаки инвариантной СИМ представимы через максимальный инвариант $I(y)$, т. е. записываются $g(x, I(y))$. Но тогда на основании теоремы 5.2 имеем следующую теорему.

Теорема 5.6. *Максимальный инвариант является достаточным* (*при* $\varkappa\geqslant 1$) *преобразованием при инвариантной СИМ, а также преобразованием ее подобия.*

Пример 5.11. Пусть $y=(y_1, \ldots, y_n)$ и пусть первичные признаки СИМ инвариантны к перестановкам координат, например приводятся к виду $g(x, (\sum y^{q}{}_{i})^{m})$, тогда инвариантные правила достаточны, поэтому инвариантным должно быть оптимальное (при $\varkappa\geqslant 1$) решающее правило; оно будет функцией максимального инварианта, которым является здесь порядковая статистика — последовательность, расположенная в порядке неубывания: $I(y)=(y_{(1)}\leqslant y_{(2)}\leqslant\ldots\leqslant y_{(n)})$.

Симметрия, инвариантность и достаточность. Наше изложение здесь основывается на том соображении, что если СИМ при некоторых преобразованиях пространства $\mathscr{Y}$ на себя не меняется, то в общем-то не должны меняться и оптимальные решающие правила.

Модель $\mathscr{M}^{xy}$ называется *симметричной* к преобразованиям sy пространства $\mathscr{Y}$ на себя, если она не меняется при этих преобразованиях, т. е. $s\mathscr{M}^{xy}=\mathscr{M}^{xy}$, или что то же самое

$$\overline{M}f(x, sy)=\overline{M}f(x, y),\ \forall f\in\mathscr{F}.$$

Преобразования s считаем далее обратимыми.

Если СИМ симметрична к преобразованиям s_1 и s_2, то она

будет симметрична к их произведению (последовательному применению) $s_1 s_2$, а также к обратным преобразованиям s_1^{-1}, s_2^{-1}. Следовательно, она будет симметрична к всевозможным произведениям (последовательным применениям) s_i и s_j^{-1}, образующим алгебраическую группу $\mathscr{S}$ преобразований. Так как СИМ будет симметрична к группе $\mathscr{S}$ в том и только в том случае, если она симметрична к каждому преобразованию $s \in \mathscr{S}$, то имеет смысл говорить о симметрии ко всей группе $\mathscr{S}$.

Симметрия СИМ означает, что каждому первичному признаку $g \in \mathscr{G}$ и преобразованию $s \in \mathscr{S}$ может быть найден другой $g^* \in \mathscr{G}$, такой, что $g(x, sy) = g^*(x, y)$ и $\bar{M} g(x, y) = \bar{M} g^*(x, y)$, т. е. преобразования симметрии совершают перестановку первичных признаков внутри $\mathscr{G}$, не меняя их средних.

Инвариантные к $\mathscr{S}$ СИМ симметричны к $\mathscr{S}$, но не всегда наоборот, поэтому понятие симметрии более емкое. Любую СИМ $\mathscr{M}^{xy}$ расширением можно сделать симметричной $\mathscr{M}_*{}^{xy}$, $\mathscr{M}_*{}^{xy} \supset \mathscr{M}^{xy}$; для чего надо образовать средние

$$\overline{M}_* f(x, y) = \sup_s \overline{M} f(x, sy).$$

В самом деле, для $\mathscr{M}_*{}^{xy}$ и $s\mathscr{M}_*{}^{xy}$ имеем $s\overline{M}_* f(x, y) = \overline{M}_* f(x, sy) = \overline{M}_* f(x, y)$, поэтому $s\mathscr{M}_*{}^{xy} = \mathscr{M}_*{}^{xy}$. Если $\mathscr{G}$ есть набор первичных признаков для $\mathscr{M}^{xy}$, то симметризацию достаточно провести на $\mathscr{G}$, что приводит, в общем, к расширению набора, и тем не менее, упрощает СИМ, так как первичные средние выравниваются на признаках $g(x, sy)$, преобразующихся друг в друга при s, пробегающих $\mathscr{S}$.

Примеры симметричных СИМ дают однородные процессы, для которых группу $\mathscr{S}$ образуют сдвиги во времени.

Свойство постоянства риска. *Если СИМ симметрична к преобразованиям $\mathscr{S}$ пространства $\mathscr{Y}$, то риск будет неизменен при преобразованиях* $s \in \mathscr{S}$: $\Pi^{\varkappa}(\partial_{sy}) = \Pi^{\varkappa}(\partial_y)$, так как согласно определению симметрии

$$\underline{M}\,\underline{\pi}(x, \partial_{sy}) = \underline{M}\,\underline{\pi}(x, \partial_y), \quad \overline{M}\,\overline{\pi}(x, \partial_{sy}) = \overline{M}\,\overline{\pi}(x, \partial_y).$$

Сейчас мы свяжем понятия симметрии, инвариантности и достаточности в смысле среднего риска, считая класс $\mathscr{D}$ замкнутым относительно $\mathscr{S}$ в смысле $\partial_y \in \mathscr{D} \Rightarrow \partial_{sy} \in \mathscr{D}$, $\forall s \in \mathscr{S}$, и рандомизированным, т. е. выпуклым.

Теорема 5.7. *Пусть СИМ симметрична к группе $\mathscr{S}$ преобразований пространства $\mathscr{Y}$ и выполняется любое из условий*: 1) *группа $\mathscr{S}$ дискретна*; 2) *оптимальное правило ∂_y^* является единственным. Тогда максимальный инвариант к группе $\mathscr{S}$ есть достаточное преобразование, причем для всех $\varkappa$.*

Доказательство. Если $\partial^{\varkappa}{}_y$ единственно, то на основании свойства постоянства риска имеем $\inf \Pi^{\varkappa}(\partial_y) = \Pi^{\varkappa}(\partial^{\varkappa}{}_y) = \Pi^{\varkappa}(\partial^{\varkappa}{}_{sy})$ и в силу единственности оптимального правила верно $\partial^{\varkappa}{}_y = \partial^{\varkappa}{}_{sy}$, т. е. оно инвариантно, и следовательно, есть функция максимального инварианта. Если оптимальное правило

не является единственным, то совокупность оптимальных правил $\{\partial^{\varkappa}{}_y\}$ должна быть замкнута относительно группы $\mathscr{S}$ преобразований в том смысле, что для заданного $\partial^{\varkappa}{}_y$ все правила $\partial^{\varkappa}{}_{sy}$ при $s \in \mathscr{S}$ должны принадлежать этой совокупности. Причем, если группа $\mathscr{S}$ дискретна $\mathscr{S} = \{s_1, \ldots, s_k\}$, то, полагая, что $\partial^*{}_y$ получается равновероятным рандомизированным выбором $\partial^{\varkappa}{}_{sy}$, придем к инвариантному правилу: $\partial^*{}_y = \sum_{i=1}^{k} \partial^{\varkappa}{}_{s_i x}/k$.

Если СИМ разложима $\mathscr{M}^{xy} = \mathscr{M}^x \mathscr{M}^y{}_x$, то симметрия $\mathscr{M}^{xy}$ эквивалентна симметрии переходных моделей $\mathscr{M}^y{}_x$ в смысле $s\mathscr{M}^y{}_x = \mathscr{M}^y{}_x$, $\forall x$.

Пример 5.12. Пусть $\mathbf{y} = (y_1, \ldots, y_n)$ и СИМ разложима. Пусть при каждом $x \in \mathscr{X}$ последовательность y_i независима и однородна, т. е.

$$\mathscr{M}^{\mathbf{y}}_x = \mathscr{M}^{y_1}_x \times \ldots \times \mathscr{M}^{y_n}_x, \quad \mathscr{M}^{y_i}_x \equiv \mathscr{M}^{y_1}_x.$$

Тогда переходные модели $\mathscr{M}^y{}_x$ будут симметричны к перестановкам y_i между собой. Максимальным инвариантом к группе перестановок является порядковая статистика $(y_{(1)} \leqslant y_{(2)} \leqslant \ldots \leqslant y_{(n)})$. Следовательно, оптимальное правило должно быть инвариантным к перестановкам и являться функцией порядковой статистики.

Порядковая статистика будет максимальным инвариантом и в том случае, если y_i при каждом заданном x имеют одинаковые модели $\mathscr{M}^{y_1}_x = \ldots = \mathscr{M}^{y_n}_x$ но совершенно неизвестно, как y_i связаны друг с другом (первичный набор образуют $\{g_x(y_i)$, $g \in \mathscr{G}$, $i = 1, \ldots, n\}$, а первичные средние не зависят от i).

5.5. ДЕТЕРМИНИРОВАННЫЕ РЕШЕНИЯ И ФИЛЬТРАЦИЯ

Общие соображения. *Детерминированными* называются правила класса $\mathscr{D}_{\text{дет}}$, для которых решениями являются сами искомые состояния $\hat{x} \in \mathscr{X}$. Правила обозначаются $\hat{x}_y$ и представляют собой инструкцию, указывающую, какое $\hat{x}$ выбирать (предлагать как решение) при каждом возможном наблюдении y.

Когда состоянием становится параметр (т. е. $\mathscr{X} = \mathscr{R}$ — числовая прямая), детерминированные правила переходят в детерминированное оценивание. Оценивание будет, когда x — вектор ($\mathscr{X} = \mathscr{R}^h$). Последний случай иногда называют фильтрацией, понимая под вектором отсчеты процесса. А в общем, фильтрация охватывает случай, когда x — процесс ($\mathscr{X}$ — пространство реализаций). Четкой грани здесь нет.

Детерминированные — это решительные решения, когда не должно быть места нечетким, осторожным высказываниям, суждениям, а требуются конкретные действия. Например, нужно очистить речевой сигнал от шумов, чтобы получить конкретную реализацию отфильтрованного процесса. Или в теории управления по наблюдениям за объектом требуется сформировать сигнал: он тоже должен быть четким как управляющее воздействие на физическую систему.

Проблема детерминизма в оценивании имеет давние традиции и богатую историю. Мы не ставим задачей обзор. Наша

цель — выявить те особенности, которые вносятся в нее новыми моделями и регулировочной шкалой оптимизма-пессимизма. Последняя позволяет делать правила то более избирательными к оговоренным ситуациям и качественными к ним, то более грубыми и устойчивыми (в смысле среднего риска) к отклонениям от них.

Потери правил $\hat{x}_y$ меряются обычно как некоторое расстояние $\pi(x, \hat{x}_y)$ между истинным значением x и предлагаемым оценкой $\hat{x}_y$. Это обязательно нелинейная, подчас неограниченная функция, как в случае степенного ее типа $|x-\hat{x}_y|^k$. Будучи преобразованием переменных x и $\hat{x}$, функция потерь искажает вид первичных признаков СИМ, вторгаясь в связь между структурой признаков и видом правил, но сохраняет, что очень важно, Qy-представимость: Q-представимость правил эквивалентна такой же представимости потерь и следует в свою очередь из представимости первичных признаков. Последнее, как утверждается теоремой 5.2, дает основания достаточной редукции наблюдений. Приведем пример редукции.

Пример 5.13. Пусть наблюдаются $y_1, \ldots, y_n$, а x — искомый числовой параметр. Пусть первичным для СИМ является среднее $\bar{M}\sum_i (y_i-x)^2=\bar{b}$. Первичный признак представляется $\sum_i (y_i-x)^2=\sum_i y^2{}_i-2x\sum_i y_i+x^2$ и при любых x есть функция $\sum_i y_i$ и $\sum_i y^2{}_i$. Эти суммы согласно теореме 5.2 есть достаточные при $\varkappa\geqslant 1$ признаки, отсюда $\hat{q}_y$ должна быть функцией от них. Некоторое обобщение получается, когда суммы взвешены коэффициентами $\sum_i c_i(y_i-x)^2$, тогда достаточны $\sum_i c_i y_i$, $\sum_i c_i y^2{}_i$.

Совершенно теми же согласно следствию 2 к теореме 5.3 достаточные признаки будут, если $y_i=x+\xi_i$, $i=1, \ldots, n$, а флуктуации ξ_i заданы первичным значением $\bar{M}\sum_i c_i\xi^2{}_i$. Если y_t — процесс, то индекс i заменяется на время t, а суммы — на интегралы по t и достаточными признаками становятся интегралы $\int c_i y_t dt$, $\int c_t y^2{}_t dt$.

Оптимальные решения при дельта-потерях. Приступим к изложению принципов построения оптимальных детерминированных оценок в зависимости от вида потерь и значения коэффициента пессимизма. Пусть сначала берутся *дельта-потери* $\pi(x, \hat{x}) = 1-\delta_{\hat{x}}(x)$, равные 0 при правильном решении $\hat{x}=x$ и 1 при неправильном $\hat{x}\neq x$. Эти потери очень критичны (если не сказать капризны) к сколь угодно малым отклонениям $\hat{x}$ от x, так как сразу подскакивают от минимального до своего максимального значения. Зато работа с дельта потерями удобна вследствие простых результатов, ибо цена приобретает «лицеприятный» вид:

$$v(\Upsilon) = 1 - \sup_{\hat{x}_y} P^{1-\varkappa}(x=\hat{x}_y), \qquad (5.5)$$

где $P^{1-\varkappa}(x=\hat{x}_y)=\varkappa\underline{P}(x=\hat{x}_y)+(1-\varkappa)\bar{P}(x=\hat{x}_y)$ — взвешенная вероятность правильного решения. При $\varkappa=0$ оптимальное правило будет находиться максимизацией верхней границы вероятности $\bar{P}(x=\hat{x}_y)$, а при пессимизме $\varkappa=1$ — нижней $\underline{P}(x=\hat{x}_y)$.

Пример 5.14. Пусть СИМ задана первичными вероятностями:

$$0\leqslant \underset{\sim}{P}(x, y)\leqslant \tilde{P}(x, y)\leqslant 1, \quad \forall (x, y)\in \mathscr{X}\times\mathscr{Y}$$

(согласованными в смысле $\sum \underset{\sim}{P}\leqslant 1$, $\sum \tilde{P}\geqslant 1$, где суммы по x, y), задающими аддитивное ИРВ на дискретных $\mathscr{X}\times\mathscr{Y}$. По формуле (1.5) для аддитивных ИРВ имеем:

$$\underline{P}(x=\hat{x}_y) = \max\left\{\sum_{x=\hat{x}_y} \underset{\sim}{P}(x, y),\ 1-\sum_{x\neq\hat{x}_y} \tilde{P}(x, y)\right\},$$

$$\bar{P}(x=\hat{x}_y) = \min\left\{\sum_{x=\hat{x}_y} \tilde{P}(x, y),\ 1-\sum_{x\neq\hat{x}_y} \underset{\sim}{P}(x, y)\right\}.$$

Так как множество $\{x\neq\hat{x}_y\}$ точек пространства $\mathscr{X}\times\mathscr{Y}$ оказывается обычно существенно шире множества $\{x=\hat{x}_y\}$, то в первой формуле правая из двух частей имеет тенденцию стать отрицательной, а во второй — левая стать меньше правой (при $\sum \underset{\sim}{P}<1$), поэтому часто оказывается, что

$$\underline{P}(x=\hat{x}_y)=\sum_{x=\hat{x}_y} \underset{\sim}{P}(x, y),\quad \bar{P}(x=\hat{x}_y)=\sum_{x=\hat{x}_y} \tilde{P}(x, y),$$

и оптимальное правило будет определяться максимизацией по x при заданном y взвешенной совместной вероятности $P^{1-\varkappa}(x, y)=\varkappa\underset{\sim}{P}(x, y)+(1-\varkappa)\tilde{P}(x, y)$, К этим правилам скоро вернемся из-за их более общего назначения.

Замечание. Обратим внимание на одну характерную особенность оптимальных правил при дельта-потерях. Множество $\{x=\hat{x}_y\}$ представляет собой линию в пространстве $\mathscr{X}\times\mathscr{Y}$ и, как факт, невозможно найти ни одной линейной комбинации первичных признаков СИМ, которая мажорировалась бы индикаторной функцией множества $\{x=\hat{x}_y\}$ и имела неотрицательное нижнее среднее, поэтому для таких задач $\underline{P}(x=\hat{x}_y)=0$. Обращаясь теперь к формуле (5.5), видим, что при $\varkappa=1$ риск независимо от $\hat{x}_y$ всегда будет максимальным, равным 1. Оптимальное правило становится бессмысленным. К бессмыслице приводит и $\varkappa>1$. Лишь при $\varkappa<1$ оптимальные правила будут нетривиальными и находятся максимизацией $\bar{P}(x=\hat{x}_y)$.

Расширим СИМ до аддитивного ИРВ, определенного границами $\underline{P}(x, y)$, $\bar{P}(x, y)$ (вычисленными по исходной СИМ и принятыми за первичные). Правило, минимизирующее риск при указанном расширении СИМ, а потому квазиоптимальное, обозначается $\hat{x}^{\varkappa}_y$. Оно согласно (5.5) максимизирует при каждом $y\in\mathscr{Y}$ взвешенную вероятность: $\max_x P^{1-\varkappa}(x, y)$ и называется *пра-*

вилом взвешенного правдоподобия (в примере 5.14 оно оптимально). В частности, при $\varkappa=0$ из него вытекает правило $\hat{x}^0{}_y$, максимизирующее $\bar{P}(x,\ y)$: $\bar{P}(\hat{x}^0{}_y)=\max\limits_x \bar{P}(x,\ y)$, называемое *правилом максимального правдоподобия*. Таким образом, *метод максимального правдоподобия соответствует максимально возможному оптимизму* $\varkappa=0$.

При пессимизме $\varkappa=1$ квазиоптимальным будет правило $\hat{x}^1{}_y: \underline{P}(\hat{x}^1{}_y,\ y)=\max\limits_x \underline{P}(x,\ y)$, максимизирующее нижнюю границу вероятности.

Введем модификации рассмотренных правил, учитывающие мешающие параметры. Пусть $\mathscr{M}^{xy}=\bigvee\limits_\theta \mathscr{P}_\theta{}^{xy}$, где $\mathscr{P}_\theta{}^{xy}$ при каждом θ определяется точными вероятностями $P_\theta(x,\ y)$. Тогда $\underset{\sim}{P}(x,\ y)=\inf\limits_\theta P_\theta(x,\ y)$, $\bar{P}(x,\ y)=\sup\limits_\theta P_\theta(x,\ y)$ и оценка $\hat{x}^0{}_y$ максимизирует максимальную по θ вероятность, а $\hat{x}^1{}_y$ — минимальную.

Если соответственно замечанию выше $\underset{\sim}{P}(x=\hat{x}_y)=0$, и тем более $\underset{\sim}{P}(x,\ y)\leqslant\underset{\sim}{P}(x=\hat{x}_y)=0$, то $P^{1-\varkappa}(x,\ y)$ становится пропорциональной $\bar{P}(x,\ y)$, поэтому правило максимального правдоподобия $\hat{x}^0{}_y$, соответствующее $\varkappa=0$, будет квазиоптимальным при любых $\varkappa<1$. Указанная пропорциональность будет иметь место и когда $P(x,\ y)=\rho\bar{P}(x,\ y)$ для некоторого $\rho\leqslant 1$ (при $\rho=1$ вероятность точная).

Постановка задачи линейной фильтрации сигнала при квадратичных потерях. Пусть наблюдения записываются

$$y_i=x_i+\xi_i,\ i=1,\ldots,n,\qquad M\xi_i=0,$$

где x_i условно будем называть сигналом, а ξ_i — шумом. Или в векторах-столбцах $\mathbf{y}=\mathbf{x}+\xi$, $M\xi=0$. Ставится задача фильтрации $\mathbf{x}$ при квадратичных потерях

$$\pi(\mathbf{x},\hat{\mathbf{x}})=\|\mathbf{x}-\hat{\mathbf{x}}\|^2=(\mathbf{x}-\hat{\mathbf{x}})^{\mathrm{T}}(\mathbf{x}-\hat{\mathbf{x}}),$$

где $\|\mathbf{z}\|^2$ — квадрат нормы вектора. Векторы $\mathbf{x}$ и ξ считаются некоррелированными: $M\mathbf{x}\xi^{\mathrm{T}}=0$, заданными свойствами второго порядка в виде согласованного набора средних $\bar{M}\mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x}$, $\bar{M}\xi^{\mathrm{T}}\mathbf{C}\xi$ при всевозможных матрицах $\mathbf{C}$. Это делается путем задания первичных значений $\bar{M}\mathbf{x}^{\mathrm{T}}\mathbf{H}\mathbf{x}$, $\mathbf{H}\in\mathscr{H}$, $\bar{M}\xi^{\mathrm{T}}\mathbf{G}\xi$, $\mathbf{G}\in\mathscr{G}$, с их последующим продолжением согласованным образом на вторичные признаки (т. е. на квадратичные формы) по формуле

$$\overline{M}\,\mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x}=\inf\Big\{\sum c_i^+\,\bar{M}\,\mathbf{x}^{\mathrm{T}}\mathbf{H}_i\,\mathbf{x}:\ \sum c_i^+\,\mathbf{H}_i-\mathbf{C}\geqslant 0\Big\},\qquad(5.6)$$

где $\mathbf{H}_i\in\mathscr{H}$. В (5.6) из сумм выброшен свободный коэффициент, ибо он получается равным 0. Неравенство под знаком инфимума понимается как матричное, т. е. как неотрицательная

определенность матрицы $\sum c^+{}_i\mathbf{H}_i-\mathbf{C}$. Такие же формулы верны для ξ.

В классе правил, допускающих разложения в ряды Вольтерра, т. е. типа

$$(\hat{\mathbf{x}}_\mathbf{y})_i=\sum_j L_{ij}\,y_j+\sum_k\sum_j L^{(2)}_{ikj}\,y_k\,y_j+\sum_k\sum_j\sum_l L^{(3)}_{ikjl}\,y_k\,y_j\,y_l+\cdots,$$

на основании теоремы 5.1 достаточными будут линейные правила $\hat{\mathbf{x}}_\mathbf{y}=\mathbf{L}\mathbf{y}$, где $\mathbf{L}=\{L_{ij}\}$ — матрица размерности $n\times n$ (так как только для них потери мажорируемы квадратичным формами). Для линейных правил при квадратичных потерях и некоррелированных $\mathbf{x}$ и ξ нижний и верхний риски запишутся

$$\underline{\Pi}(\mathbf{L}\mathbf{y})=\underline{M}^{\mathbf{xy}}\,\|\mathbf{x}-\mathbf{L}\mathbf{y}\|^2=\underline{M}^{\mathbf{x}}\,\|\mathbf{x}-\mathbf{L}\mathbf{x}\|^2+\underline{M}^{\xi}\,\|\mathbf{L}\,\xi\|^2,$$

$$\overline{\Pi}(\mathbf{L}\,\mathbf{y})=\overline{M}^{\mathbf{xy}}\,\|\mathbf{x}-\mathbf{L}\mathbf{y}\|^2=\overline{M}^{\mathbf{x}}\,\|\mathbf{x}-\mathbf{L}\mathbf{x}\|^2+\overline{M}^{\xi}\,\|\mathbf{L}\,\xi\|^2$$

и вычисляются согласно (5.6). Риски складываются из двух слагаемых: первое определяет ошибку фильтрации за счет инерционности фильтра $\mathbf{L}$, который не успевает отслеживать изменения сигнала (в силу отличия от единичной матрицы $\mathbf{I}$), а второе — за счет проникновения шума на выход фильтра. Первое слагаемое растет при увеличении инерционности фильтра, а второе — уменьшается, и между ними есть оптимум.

Выбор матрицы $\mathbf{L}^*$, определяющей оптимальный фильтр, производится исходя из минимизации среднего риска

$$\Pi^{\varkappa}(\mathbf{L})=M^{\varkappa}\,\mathbf{x}^{\mathrm{T}}(\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\,\mathbf{x}+M^{\varkappa}\,\xi^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{L}\,\xi,\qquad(5.7)$$

где $M^{\varkappa}$ есть взвешенное среднее, равное $(1-\varkappa)\underline{M}+\varkappa\bar{M}$.

При точных матрицах корреляций $M\mathbf{x}\mathbf{x}^{\mathrm{T}}=\mathbf{K}$, $M\xi\xi^{\mathrm{T}}=\mathbf{B}$ риск от $\varkappa$ не будет зависеть: $\Pi(\mathbf{L})=\mathrm{tr}\,(\mathbf{I}-\mathbf{L})\,(\mathbf{I}-\mathbf{L})^{\mathrm{T}}\mathbf{K}+\mathrm{tr}\,\mathbf{L}\mathbf{L}^{\mathrm{T}}\mathbf{B}$, где tr — след матрицы, равный сумме диагональных элементов. Оптимальным в этом случае при обратимой матрице $(\mathbf{K}+\mathbf{B})$ будет фильтр

$$\mathbf{L}^*=(\mathbf{K}+\mathbf{B})^{-1}\,\mathbf{K},\qquad \Pi(\mathbf{L}^*)=\mathrm{tr}\,\mathbf{K}\,(\mathbf{K}+\mathbf{B})^{-1}\,\mathbf{B},\qquad(5.8)$$

где вторым указано выражение для ошибки фильтрации.

Фильтрация сигнала с известными корреляционными свойствами из шума ограниченной мощности. Пусть корреляционная матрица сигнала $\mathbf{x}$ является точно известной и равной $\mathbf{K}$, тогда как относительно шума известны сведения лишь об его средней «мощности»

$$\underline{M}\ \sum\xi_i^2/n=\underline{M}\,\|\xi\|^2/n=\underline{\sigma}^2,\ \overline{M}\,\|\xi\|^2/n=\overline{\sigma}^2.$$

Тогда аналогично (5.6): $\underline{M}\,\xi^{\mathrm{T}}\mathbf{L}\mathbf{L}\,\xi=\sup\limits_{c\mathbf{I}\leqslant\mathbf{L}^{\mathrm{T}}\mathbf{L}} cn\underline{\sigma}^2=\lambda^2_{min}\,n\sigma^2$,

$$\overline{M}\,\xi^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{L}\,\xi=\inf_{\{c\mathbf{I}\geqslant\mathbf{L}^{\mathrm{T}}\mathbf{L}} cn\,\overline{\sigma}^2=\lambda^2_{max}\,n\overline{\sigma}^2,$$

где неравенства матричные и $\lambda^2{}_{min}$ и $\lambda^2{}_{max}$ есть соответственно

минимальное и максимальное собственные числа матрицы $\mathbf{L}^\tau\mathbf{L}$. В результате средний риск

$$\Pi^\varkappa(\mathbf{L}) = \operatorname{tr}(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^\tau\mathbf{K} + (1-\varkappa)\lambda_{min}^2 n\underline{\sigma}^2 + \varkappa\lambda_{max}^2 n\overline{\sigma}^2.$$

Представим $\mathbf{K}=\mathbf{F}^\tau\mathbf{\Gamma}\mathbf{F}$, где $\mathbf{F}$ — матрица собственных векторов, а $\mathbf{\Gamma}$ — диагональная матрица собственных чисел γ_i матрицы $\mathbf{K}$. Тогда оптимальное правило запишется в виде $\mathbf{L}\mathbf{y}=\mathbf{F}\mathbf{\Lambda}$, где $\mathbf{\Lambda}$ — диагональная матрица корней квадратных λ_i из собственных элементов $\lambda^2{}_i$ матрицы $\mathbf{L}^\tau\mathbf{L}$. В результате средний риск станет

$$\Pi^\varkappa(\mathbf{L}) = \sum(1-\lambda_i)^2\gamma_i + (1-\varkappa)\lambda_{min}^2 n\underline{\sigma}^2 + \varkappa\lambda_{max}^2 n\overline{\sigma}^2.$$

Задача нахождения оптимального правила сводится к минимизации найденного выражения по λ_i. Дадим результат, тем более, что он тривиален, не приводя утомительных, но в то же время несложных выкладок. При $\varkappa\geqslant 1$ оптимальными будут $\lambda^*{}_i\equiv c$, что соответствует фильтрации $\hat{x}_\mathbf{y}=c\mathbf{y}$, сводящейся к прямому ослаблению входных наблюдений на величину $c=\operatorname{tr}\mathbf{K}/(\operatorname{tr}\mathbf{K}+\varkappa n\overline{\sigma}^2)$. Причем c уменьшается при сравнительном увеличении средней статистической энергии шума $n\overline{\sigma}^2$, приходящейся на n отсчетов, по отношению к энергии сигнала $\operatorname{tr}\mathbf{K}$. В самом деле, если шум велик, то он дает основной вклад в ошибку на выходе и его нужно ослаблять. Незнание структуры шума ведет к тому, что наблюдения фактически не фильтруются.

Пусть теперь $0\leqslant\varkappa<1$. Результат будет весьма интересным, показывающим на вырожденные стороны режима оптимизма. При условии $\varkappa n\overline{\sigma}^2/\operatorname{tr}\mathbf{K}<\sqrt{(1-\varkappa)n\underline{\sigma}^2/\gamma_{min}}$ оптимальными снова будут $\lambda^*{}_i=c$ для всех i, кроме одного $i0$: $\gamma_{i0}=\min\gamma_i=\gamma_{min}$, $\lambda^*{}_{i0}=0$, назначаемого нулевым. При синтезе фильтра наивно предполагается, что в наиболее благоприятном случае шум весь войдет именно в это наименее «сигнальное» направление γ_{min}, а мы «захлопнем» его путем приравнивания нулю: $\lambda^*{}_{i0}=0$. Представьте, что шумом с вероятностью $1-\varkappa$ управляет «свой» человек, наиблагоприятнейшим образом к нам расположенный, а с вероятностью $\varkappa$ — противник. Оба они прекрасно осведомлены о наших действиях. Нужно защититься по-возможности от противника, и оставить «отдушину для благоприятных» акций «своего», гарантировав себе таким образом минимальный ущерб.

Последнее не может служить утешением, а скорее является вырождением, поэтому общий вывод следующий: *незнание структуры шума не позволяет никак использовать знание, даже точное, корреляционных свойств сигнала.*

Фильтрация при некоррелированном шуме. Пусть снова $\mathbf{y}=\mathbf{x}+\boldsymbol{\xi}$, корреляционная матрица $\mathbf{K}$ вектора сигнала $\mathbf{x}$ известна точно, а шум некоррелирован и задан первичными средними:

$$\overline{M}\,\xi_i^2=\overline{\sigma}_i^2,\ \underline{M}\xi_i^2=\underline{\sigma}_i^2,\ M\xi_i\xi_j=0,\ i\neq j.$$

Тогда $M\|\mathbf{L}\boldsymbol{\xi}\|^2=\sup\{\sum c_i\sigma_i^2 : \sum c_i\xi_i^2+\sum\limits_{i\neq j}\sum\xi_i c_{ij}\xi_i\leqslant$

$$\leqslant\sum\sum\sum\xi_i L_{ij}L_{jk}\xi_k\}=\sup\{\sum c_i\underline{\sigma}_i^2 : c_i\leqslant\sum_i L_{ij}^2\}=\sum\sum L_{il}^2\underline{\sigma}_i^2,$$

и точно так же $\overline{M}\|\mathbf{L}\boldsymbol{\xi}\|^2=\sum\sum L^2{}_{ij}\overline{\sigma}^2{}_i$. В результате риск

$$\Pi^\varkappa(\mathbf{L}) = \operatorname{tr}(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^\tau\mathbf{K} + \operatorname{tr}\mathbf{L}\mathbf{L}^\tau\mathbf{D}_\varkappa,$$

где $\sigma^2{}_{\varkappa i}=(1-\varkappa)\underline{\sigma}^2{}_i+\varkappa\overline{\sigma}^2{}_i$ и $\mathbf{D}_\chi$ есть диагональная матрица из $\sigma^2{}_{\varkappa i}$. Минимизация риска по $\mathbf{L}$ ведет к значению

$$\mathbf{L}^*=(\mathbf{K}+\mathbf{D}_\varkappa)^{-1}\mathbf{K},\qquad \Pi^\varkappa(\mathbf{L}^*)=\operatorname{tr}\mathbf{K}(\mathbf{K}+\mathbf{D}_\varkappa)^{-1}\mathbf{D}_\varkappa.$$

Оптимальный фильтр, как это видно из сравнения с (5.8), получается точно такой же, как если бы $M\xi^2{}_i$ были точно известны и равнялись $\sigma^2{}_{\varkappa i}$.

Обобщение. Пусть $\mathbf{y}=\mathbf{x}+\mathbf{H}\boldsymbol{\xi}$, где $\mathbf{H}$ — известная матрица, вектор $\boldsymbol{\xi}$ — тот же, как и был, а шум $\mathbf{H}\boldsymbol{\xi}$ есть результат прохождения независимых отсчетов ξ_i через известный фильтр, описываемый матрицей $\mathbf{H}$. Шум в результате будет коррелированный с неточно известными корреляциями, обязанными различию $\underline{\sigma}^2{}_i$ и $\overline{\sigma}^2{}_i$. Тогда

$$\Pi^\varkappa(\mathbf{L}) = \operatorname{tr}(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^\tau\mathbf{K} + \operatorname{tr}\mathbf{H}\mathbf{L}\mathbf{L}^\tau\mathbf{H}^\tau\mathbf{D}_\varkappa,$$

и в результате минимизации по $\mathbf{L}$ получим

$$\mathbf{L}^*=(\mathbf{K}+\mathbf{H}\mathbf{D}_\varkappa\mathbf{H}^\tau)^{-1}\mathbf{K},\qquad \Pi^\varkappa(\mathbf{L}^*)=\operatorname{tr}\mathbf{K}(\mathbf{K}+\mathbf{H}\mathbf{D}_\varkappa\mathbf{H}^\tau)^{-1}\mathbf{H}\mathbf{D}_\varkappa\mathbf{H}^\tau.$$

Вытекающая из этих формул инструкция призывает заменить неточно известные корреляции шума на матрицу $B_{ij}=\sum\limits_k H_{ik}\sigma^2{}_{\varkappa k}H_{kj}$, считая ее про себя точной матрицей корреляций шума, и далее расчеты с ней производить по (5.8). Например, если $y_i=x_i+\sum\limits_l h_l\xi_{i+l}$, т. е. шум генерируется скользящим суммированием независимых ξ_i, то $B_{ij}=\sum\limits_l h_{i-l}\sigma^2{}_{\varkappa l}h_{l-j}$. Здесь мы использовали прямой метод, который свел задачу к точным корреляциям. В следующем разделе использован робастный подход.

Корреляции заданы с погрешностями. Пусть $\mathbf{y}=\mathbf{x}+\boldsymbol{\xi}$, сигнал $\mathbf{x}$ и шум $\boldsymbol{\xi}$ некоррелированы между собой и заданы каждый своими корреляционными свойствами. Последние, как известно (§ 4.2), эквивалентны собственным семействам, соответственно $\mathscr{K}$ (для $\mathbf{x}$) и $\mathscr{B}$ (для $\boldsymbol{\xi}$) ковариаций, что позволяет при вычислениях риска пользоваться формулой

$$\Pi^\varkappa(\mathbf{L}) = (1-\varkappa)\inf_{\mathbf{K}\in\mathscr{K}}\operatorname{tr}(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^\tau\mathbf{K} + \varkappa\sup_{\mathbf{K}\in\mathscr{K}}\operatorname{tr}(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^\tau\mathbf{K} +$$

$$+(1-\varkappa)\inf_{\mathbf{B}\in\mathscr{B}}\operatorname{tr}\mathbf{L}\mathbf{L}^\tau\mathbf{B} + \varkappa\sup_{\mathbf{B}\in\mathscr{B}}\operatorname{tr}\mathbf{L}\mathbf{L}^\tau\mathbf{B}.$$

Рассмотрим случай, когда $\mathscr{K}$ и $\mathscr{B}$ заданы метрическими ограничениями вида

$$\mathscr{K}=\{\mathbf{K}:\|\mathbf{K}-\mathbf{K}_0\|\leqslant\Delta_{\mathbf{K}}\},\ \mathscr{B}=\{\mathbf{B}:\|\mathbf{B}-\mathbf{B}_0\|\leqslant\Delta_{\mathbf{B}}\},$$

где норма матриц понимается как максимальное по модулю собственное число. Можем для определенности считать, что $\mathbf{K}_0$ и $\mathbf{B}_0$ суть оценки корреляционных матриц, известные с ошибками, не превышающими соответственно допусков $\Delta_{\mathbf{K}}$ и $\Delta_{\mathbf{B}}$.

Верно равенство: $\sup\{\mathrm{tr}\,\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{K}:\|\mathbf{K}-\mathbf{K}_0\|\leqslant\Delta_{\mathbf{K}}\}=\mathrm{tr}\,\mathbf{H}\mathbf{H}^{\mathrm{T}}(\mathbf{K}_0+\Delta_{\mathbf{K}}\mathbf{I})$, причем супремум достигается при $\mathbf{K}^*=\mathbf{K}_0+\Delta_{\mathbf{K}}\mathbf{I}$, а $\mathbf{I}$ — единичная матрица. Матрица $\mathbf{K}^*$ получается из $\mathbf{K}_0$ увеличением всех собственных чисел на $\Delta_{\mathbf{K}}$. Используя тот факт, что семейства $\mathscr{K}$ и $\mathscr{B}$ должны состоять из неотрицательно определенных матриц, аналогично предыдущему имеем

$$\inf\{\mathrm{tr}\,\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{K}:\|\mathbf{K}-\mathbf{K}_0\|\leqslant\Delta_{\mathbf{K}}\}=\mathrm{tr}\,\mathbf{H}\mathbf{H}^{\mathrm{T}}(\mathbf{K}_0-\Delta_{\mathbf{K}}\mathbf{I})^{+},$$

где $(\mathbf{K}_0-\Delta_{\mathbf{K}}\mathbf{I})^{+}$ получается из $\mathbf{K}_0-\Delta_{\mathbf{K}}\mathbf{I}$ приравниванием нулю всех отрицательных собственных чисел при сохранении собственных векторов. На основании этих двух равенств выводим

$$\Pi^{\varkappa}(\mathbf{L})=(1-\varkappa)\,\mathrm{tr}\,(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{K}_0-\Delta_{\mathbf{K}}\mathbf{I})^{+}+$$
$$+\varkappa\,\mathrm{tr}\,(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{K}_0+\Delta_{\mathbf{K}}\mathbf{I})+(1-\varkappa)\,\mathrm{tr}\,\mathbf{L}\mathbf{L}^{\mathrm{T}}(\mathbf{B}_0-\Delta_{\mathbf{B}}\mathbf{I})^{+}+$$
$$+\varkappa\,\mathrm{tr}\,\mathbf{L}\mathbf{L}^{\mathrm{T}}(\mathbf{B}_0+\Delta_{\mathbf{B}}\mathbf{I})=\mathrm{tr}\,(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^{\mathrm{T}}\mathbf{K}_{\varkappa}+\mathrm{tr}\,\mathbf{L}\mathbf{B}_{\varkappa}\mathbf{L}^{\mathrm{T}},$$

где $\mathbf{K}_{\varkappa}=(1-\varkappa)(\mathbf{K}_0-\Delta_{\mathbf{K}}\mathbf{I})^{+}+\varkappa(\mathbf{K}_0+\Delta_{\mathbf{K}}\mathbf{I})$, и аналогично $\mathbf{B}_{\varkappa}$.

Оптимальной, минимизирующей риск, будет матрица $\mathbf{L}^*=(\mathbf{K}_{\varkappa}+\mathbf{B}_{\varkappa})^{-1}\mathbf{K}_{\varkappa}$. Заметим, что в последнем выражении $\mathbf{K}_{\varkappa}$ получается из $\mathbf{K}_0$ пересчетом собственных чисел λ_{0i} матрицы $\mathbf{K}_0$ в соответствии с равенством

$$\lambda_i=(1-\varkappa)(\lambda_{0i}-\Delta_{\mathbf{K}})^{+}+\varkappa(\lambda_{0i}+\Delta_{\mathbf{K}})=\begin{cases}\varkappa\lambda_{0i}+\varkappa\Delta_{\mathbf{K}} & \text{при } \lambda_{0i}<\Delta_{\mathbf{K}},\\ \lambda_{0i}+(2\varkappa-1)\Delta_{\mathbf{K}}, & \text{при } \lambda_{0i}\geqslant\Delta_{\mathbf{K}},\end{cases}$$

а собственные векторы сохраняются неизменными. Так же получается и $\mathbf{B}_{\varkappa}$ из $\mathbf{B}_0$. Числа λ_i возрастают по сравнению с λ_{0i} при $\varkappa>1/2$, а при $\varkappa\leqslant 1/2$ растут лишь те из них, которые достаточно малы, а именно, $\lambda_{0i}<\Delta_{\mathbf{K}}\varkappa/(1-\varkappa)$. Но всегда $\lambda_i\geqslant\varkappa\Delta_{\mathbf{K}}$. Закономерности пересчета можно проследить на рис. 5.1, о котором пойдет речь в дополнении 2.

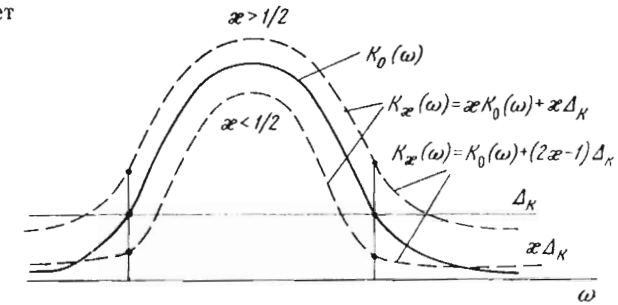
При $\mathbf{K}_0-\Delta_{\mathbf{K}}\mathbf{I}>0$ и $\mathbf{B}_0-\Delta_{\mathbf{B}}\mathbf{I}>0$ (что соответствует положительной определенности матриц) имеем

$$\mathbf{K}_{\varkappa}=\mathbf{K}_0+(2\varkappa-1)\Delta_{\mathbf{K}}\mathbf{I},\qquad \mathbf{B}_{\varkappa}=\mathbf{B}_0+(2\varkappa-1)\Delta_{\mathbf{B}}\mathbf{I}.$$

Таким образом, наличие ошибок $\Delta_{\mathbf{K}}$ и $\Delta_{\mathbf{B}}$ в корреляционных функциях компенсируется прибавлением к $\mathbf{x}$ и ξ добавок в виде некоррелированных векторов η и ζ: $M\eta^2_i=\Delta_{\mathbf{K}}$, $M\zeta^2_i=\Delta_{\mathbf{B}}$, $M\eta_i\eta_j=M\zeta_i\zeta_j=0$, $i\neq j$, и переходом к модели $\mathbf{y}=\mathbf{x}'+\xi'$, в которой $\mathbf{x}'=\mathbf{x}+\eta$, $\xi'=\xi+\zeta$ [1], с последующим синтезом для нее оптимального фильтра по (5.8), который и будет оптимальным для исходной задачи.

[1] Кузнецов В. П. Об устойчивой линейной фильтрации случайных сигналов// Радиотехника и электроника. — 1975. — № 1. — С. 2405—2408.

Рис. 5.1. Перерасчет энергетического спектра



Дополнения. 1. Результаты один к одному переносятся на процессы $y_t=x_t+\xi_t$: матрицы $\mathbf{K}$ и $\mathbf{B}$ заменяются на ядра $K(t,\tau)$, $B(t,\tau)$ операторов, а след $\mathrm{tr}\,\mathbf{H}$ — на интеграл $\int H(t,t)dt$. Ядром единичного оператора $\mathbf{I}$ будет дельта-функция, являющаяся корреляционной функцией «белого шума» спектральной интенсивности 1.

2. Пусть процессы x_t и ξ_t являются независимыми (или некоррелированными) стационарными и определены своими энергетическими спектрами $K(\omega)$ и $B(\omega)$, заданными метрическими ограничениями

$$\sup_{\omega}|K(\omega)-K_0(\omega)|\leqslant\Delta_K,\qquad \sup_{\omega}|B(\omega)-B_0(\omega)|\leqslant\Delta_B,$$

где $K_0(\omega)$ и $B_0(\omega)$ — некоторые предполагаемые значения (оценки). Поиск оптимального фильтра ориентируется на пересчитанный спектр (рис. 5.1) $K_{\varkappa}(\omega)$, вычисляемый по аналогии с λ_i.

5.6. ЗАКЛЮЧЕНИЕ

Предлагаемая здесь теория статистического синтеза по своему построению подобна классической. Можно аргументированно говорить об оптимальности, если определено это понятие. А для этого должен быть отработан аппарат анализа, выдвинуты критерии сравнения алгоритмов: какой из любых выбранных лучше, какой хуже. Аппарат будет действенен, если анализ каждого алгоритма принципиально возможен и не настолько трудоемок, что где-то грозит остановкой. Тогда возникает мысль, что раз алгоритмы упорядочиваются в колонну друг за другом по их качественным показателям, то среди них существует наилучший, к которому можно приблизиться, двигаясь в голову колонны. Причем для этого не обязателен перебор всех алгоритмов, имеются другие эффективные приемы и их разработка — наша цель.

Синтез не может обходиться без анализа, а анализ — без подготовительных работ следующего рода. Сначала, учитывая случайную природу среды, нужно определить характер случайности, т. е. математическую модель в виде СИМ (статистическая интервальная модель). Последняя есть совместное описание поведения наблюдений, т. е. входа алгоритма, и не наблюдаемых, но интересующих нас внутренних состояний среды. С состояниями связываются выходы алгоритмов — принимаемые ими решения.

Вид множества решений составляет лицо задачи. Если имеются всего два решения, то это будет проверка двух гипотез. Нескольким решениям соответствует многоальтернативная задача. Если нужно оценить параметр, то множеством решений будут точки числовой оси. Наконец, в задачах фильтрации решениями будут реализации обработанного согласно алгоритму сигнала.

Каждый алгоритм есть правило действия, инструкция, предписывающая, какой выход назначить каждому наблюдению, т. е. входу. Изюмина нашего подхода в том, что это не обязана быть совсем строгая инструкция (детерминированное правило), а может быть набор рекомендаций в виде списка сравнительных предпочтений, которыми наделяются разные решения. Алгоритмы — неодушевленные объекты, а окончательное решение пусть останется за человеком. Сказанное подводит нас к понятию нечетких решений и расплывчатых алгоритмов.

Практика применений и род математического аппарата синтеза могут потребовать алгоритмов закрепленной структуры (скажем, линейных). Тогда с самого начала вводится ограничение правил в виде класса допустимых алгоритмов, внутри которого и будет производиться затем выбор наилучшего.

Подготовительные работы еще не закончены. Нужно уметь сопоставить истинные значения состояний с принятыми решениями. Конечно же, полное совпадение — это очень хорошо; но не было бы статистической задачи, если бы алгоритм не имел «права на ошибку». Нужно назначить плату за ошибки в виде функции потерь.

Только теперь после введения всех атрибутов статистической задачи можно приступить к синтезу, т. е. нахождению оптимального правила принятия решений. Критерием сравнения и выбора лучшего будет риск, равный средним потерям. Здесь возникают две особенности. Одна — конструктивная, состоит в вычислении риска продолжением первичных средних СИМ на функцию потерь (частным случаем такого продолжения будет интегрирование по вероятностному распределению).

Другая — интервальность риска как среднего одного из признаков СИМ, причем нижний риск есть наименьшее, наиболее оптимистичное его значение, а верхний — пессимистичное. Идея брать некоторую промежуточную величину ведет к коэффициенту пессимизма, взвешивающему риски. Хотим иметь надежный гарантированный результат, берем коэффициент пессимизма равным 1, ориентируясь полностью на наихудший верхний риск. Берем ноль — рассчитываем на нижний риск, сверхоптимистично делая ставку на полную удачу (как в методе максимального правдоподобия § 5.5). Оптимальные правила при излишнем оптимизме лихорадочны по свойствам. Чуть лучше они делаются при полуоптимизме, когда нижний и верхний риски суммируются с одинаковыми весами. Режим полуоптимизма ослабляет требования к надежности модели (определяющей «здоровье» алгоритма) и одобряет нарочный переход к идеальным «погодным условиям» в виде точных моделей, оправдывая тем самым применимость распределений вероятностей. Полный иммунитет к «погоде» приобретают оптимальные правила, синтезированные в режиме пессимизма и сверхпессимизма. По свойствам они делаются устойчивыми, робастными.

Алгоритмы суть изделия научных лабораторий. Удобно их «производство» сделать двухэтапным: сначала — предварительная заготовка, а затем — окончательная оптимизация. Первый этап составляет смысл достаточной редукции. Цель его — сузить класс решающих правил, указав, на каком материале он должен основываться, т. е. какие предварительные сокращения наблюдений не приведут к потере свойств. Интересно и важно (теорема 5.3), что эта редукция всецело определяется структурой первичных признаков, исходно задающих СИМ. Чем проще первичный набор, тем глубже возможна редукция.

Достаточность — прерогатива исключительно режима пессимизма. В нашем изложении связывается с классическим понятием теоремой 5.4 о факторизации (расширенной применительно к интервальным моделям, а значит, и к семействам распределений вероятностей).

Инвариантность интервальных моделей и их симметрия к преобразованиям пространства наблюдений порождает такие же особенности оптимальных правил, а следовательно, предопределяет в какой-то мере вид достаточных преобразований (§ 5.4).

Последний параграф § 5.5 главы отдается иллюстрированному применению основных утверждений теории к задачам детерминированного (точечного) оценивания и фильтрации. Выясняется влияние коэффициента пессимизма на вид алгоритмов оптимальной фильтрации.

Глава 6.

РАСПЛЫВЧАТОЕ ОЦЕНИВАНИЕ

6.1. ОБЩИЕ ВОПРОСЫ

Ошибки правил. Детерминированные (точечные) оценки $\hat{x}_y$ по своему внутреннему содержанию таковы, что обладают мизерной вероятностью $\bar{P}(\hat{x}_y=x)$ «угадывания» правильного состояния, а чаще всего и вовсе нулевой (соответственно вероятностью ошибки $P(\hat{x}_y\neq x)$ — единичной). И эта природная их черта не есть следствие того, что оценки плохи (для определения целей они могут быть очень даже хороши), а просто вызвана тем, что хоть как-то попасть «дрожащей» точкой $\hat{x}_y$ в точку x на числовой прямой или числовом пространстве невозможно, учитывая количество точек и ничтожность точечных размеров. Это можно сделать только тогда, когда истинное x «накалывается» не точкой, а накрывается чем-то ощутимым, например интервалом, что и подводит нас к расплывчатым оценкам.

Потребность в расплывчатых оценках возникает прежде всего там, где помимо значения состояния нужна точность, с которой оно оценивается. Например, ставится задача — сопровождать указание расстояния до цели величиной погрешности, диапазоном разброса, с которым оно измеряется. Так порождается *доверительный интервал*. Чем шире он, тем меньше вероятность, что при указании произойдет ошибка. Эта ошибка может входить как составляющая риска, но может быть частью исходных технических требований на оценку, аппаратуру, тогда ошибку нужно фиксировать, поддерживать на определенном уровне. В этом случае говорим об оценке фиксированного уровня α ошибки, или просто уровня α.

В наших рассуждениях сейчас не столь важно, является ли $\mathscr{X}$ числовым или каким другим, и даже не число элементов x в $\mathscr{X}$, поэтому говорим о правилах как общих случаях оценок.

Пусть $\partial_y(x)$ — расплывчатое правило — функция как x, так и y. Рассмотрим сначала его надежность — величину, обратную ошибке. Для заданного правила $\partial_y(x)$ нижняя и верхняя надежности соответственно равны $\underline{M}\partial_y$, $\overline{M}\partial_y$, где средние берутся по СИМ $\mathscr{M}^{xy}$. В самом деле, величина $\partial_y(x)$ есть уверенность, с которой x указывается как возможное значение искомого состояния по результату наблюдения y. При разных y и x эта величина колеблется от 0 до 1 сообразно виду правила, причем 0 соответствует тому, что x отвергается как приемлемый вариант состояний, а 1 — наоборот, соответствует бесприкословному включению; величина $\partial_y(x)=1/2$ означает, что при заданном y некоторое x принимается как возможное состояние с вероятностью 1/2, т. е. с половинчатой уверенностью. Надежность же есть уверенность в среднем, т. е. определенная в среднестатистическом смысле, что и заложено в $\underline{M}\partial_y$.

Ненадежность меряется как нижняя и верхняя вероятности ошибки: $\underline{\alpha}(\partial)=1-\overline{M}\partial_y$, $\overline{\alpha}(\partial)=1-\underline{M}\partial_y$, а взвешенная вероятность ошибки как объединенный показатель

$$\alpha^{\varkappa}(\partial)=(1-\varkappa)\,\underline{\alpha}(\partial)+\varkappa\,\overline{\alpha}(\partial)=1-(1-\varkappa)\,\overline{M}\,\partial_y-\varkappa\,\underline{M}\,\partial_y, \qquad (6.1)$$

где $\varkappa$ — коэффициент пессимизма.

Нулевой вероятностью ошибки обладают тривиальные правила $\partial_y(x)\equiv 1$, тождественно равные 1, соответствующие при любом x фразе: «Какое-то x из $\mathscr{X}$ имеет место» (но какое?). Его надежность равна 1, так как ошибиться здесь невозможно (по определению пространства элементарных исходов). Другая крайность — единичная вероятность ошибки, свойственная точечным (детерминированным) оценкам, причем чаще единичной будет оказываться именно верхняя ошибка $\overline{\alpha}(\partial)=1$, а нижняя при этом может вполне быть нулевой (тогда взвешенная ошибка будет равна $\varkappa$ — вере в неблагопряитный исход). Предметный интерес для теории представляют не крайности, а «золотая середина», т. е. унимодальные как функции x при каждом y контрастные (достигающие 0 и 1) правила ограниченной по x ширины.

Не ограждая класс $\mathscr{D}$ расплывчатых правил какими-либо барьерами, т. е. $\mathscr{D}=\mathscr{D}_{\forall}$, выделим из него подкласс следующим условием.

Правила ∂_y, для которых $\alpha^{x}(\partial)=\alpha$, называются *правилами уровня* α, а при числовом $\mathscr{X}$ — *доверительными оценками* x. В дальнейшем мы будем заниматься синтезом именно таких оценок.

Расплывчатость, риск. Помимо ошибок у рассматриваемых правил имеется другая «теневая сторона» — их расплывчатость.

Желательно, чтобы она была как можно меньше, т. е. при каждом y конкретнее указывалось бы искомое состояние x, но это вступает в противодействие с ошибкой правил, которая при этом увеличивается.

Расплывчатость характеризуется шириной правил как функций переменной x при заданном y; может измеряться разными способами. Прежде чем приступить к их рассмотрению, подумаем, а зачем здесь какое-то разнообразие, нужно ли оно? Оказывается, да, так как это своего рода регулировка, инструмент направленного воздействия на оптимальные правила, придания им тех или иных желаемых качеств, черт, оттенков.

Приведем разные шкалы расплычатости, пригодные как для скалярного, так и векторного параметров $x\in\mathscr{R}^k$ (лишь для векторного $x=(x_1, \ldots, x_k)$ и интегралы по $dx=dx_1\cdot\ldots\cdot dx_k$ становятся кратными).

1. *Интегральная шкала*: $\Omega(\partial_y)=\int\partial_y(x)\,dx$ — наиболее проста по смыслу, так как придает всем иксам одни и те же веса. Для одномерного x это будет приведенная (к прямоугольнику единичной высоты той же площади) ширина, а многомерного — приведенный объем.

2. *Взвешенная шкала*: $\Omega_q(\partial_y)=\int q(x)\,\partial_y(x)\,dx$. Весовая функция $q(x)$ нужна для выделения более важных (скажем, в смысловом или стратегическом отношении) состояний, вынуждая их более точное оценивание в ущерб остальным.

3. *Обобщенная шкала*: $\Omega^{\gamma}(\partial_y)=\int q(x)\,\partial_y(x)^{\gamma}dx$. Возведение решений в степень $\gamma<1$ увеличивает $\partial^{\gamma}y$ по сравнению с ∂_y, если последнее <1, тем самым увеличивая плату за неуверенность решений и стимулируя их повышенную категоричность типа 0 и 1 (нет, да). При $\gamma>1$, наоборот, будут поощряться неуверенные решения.

4. *Эффективная ширина*: $\int\partial_y(x)\,dx/\sup\limits_x\partial_y(x)$. Это есть при каждом y отношение площади (объема) под $\partial_y(x)$ как функции переменной x к высоте $\overline{\partial}_y$. Таким способом создаются благоприятные условия для контрастных категоричных правил, у которых $\overline{\partial}_y=1$, и для них это будет просто интегральная шкала.

5. *Периметр* свойствен векторному x и определяется как максимальная по компонентам x_i интегральная ширина правил:

$$\max_x\max_i\,[q_i\int\partial_y(x)\,dx_i],$$

где q_i — веса, выделяющие относительную «важность» разных компонент вектора x.

Введенные шкалы пригодны и для дискретных x, если интегралы обменять на суммы по x.

Величина расплывчатости по введенным шкалам зависит от наблюдений y. Так как расплывчатость правил $\partial_y(x)$ может быть, в общем, разной при разных наблюдениях y, поэтому нужно говорить о средней расплывчатости (ширине), усредняя $\Omega(\partial_y)$ по y согласно СИМ $\mathscr{M}^{xy}$ (по частной к ней $\mathscr{M}^{y}$). Средняя ширина есть

штраф за расплывчатость и меряется, в общем, сверху и снизу: $\overline{M}\Omega(\partial_y)$, $\underline{M}\Omega(\partial_y)$, где $\Omega(\ )$ — любая из указанных выше шкал.

Если брать взвешенную сумму $(1-\varkappa)\underline{M}\Omega(\partial_y)+\varkappa\overline{M}\Omega(\partial_y)$, то при пессимизме $\varkappa\geqslant 1$ нижнее значение будет с отрицательным знаком, и это уже, извините, будет не штраф, а поощрение расплывчатости, что сделает невозможным основные результаты о достаточности (справедливые, как увидим, при $\varkappa\geqslant 1$), поэтому за штраф будем брать верхнее значение $\overline{M}\Omega(\partial_y)$ (по крайней мере, годное при пессимизме).

Назовем *составным риском расплывчатых правил* взвешенную сумму вероятности ошибки и верхнего штрафа за расплывчатость (для скалярного x это средняя ширина):

$$\Pi^{\varkappa}_{\lambda}(\partial)=\alpha^{\varkappa}(\partial)+\lambda\overline{M}\Omega(\partial_y). \tag{6.2}$$

Весовой коэффициент λ призван регулировать отношение между слагаемыми: при увеличении λ бо́льшая важность придается расплывчатости, нежели ошибке, в результате оптимальные правила, т. е. те, которые минимизируют $\Pi^{\varkappa}_{\lambda}(\partial)$, будут более точными и менее надежными. И наоборот.

Обратим внимание, что составной риск (6.2) не есть средний риск (среднее от потерь) в смысле предыдущей главы, а представляет собой некоторую очень простую форму характеризации негативных сторон правила. Тем не менее он сохраняет структуру риска как взвешенного коэффициентом пессимизма нижнего и верхнего его значений:

$$\Pi^{\varkappa}_{\lambda}(\partial)=(1-\varkappa)\underline{\Pi}(\partial)+\varkappa\overline{\Pi}(\partial)$$

при $\underline{\Pi}(\partial)=\underline{M}\partial_y$, $\overline{\Pi}(\partial)=\overline{M}\partial_y+\dfrac{\lambda}{\varkappa}\overline{M}\Omega(\partial_y)$.

Почему мы вдруг рассматриваем составной риск, а не средний? Да потому что составной риск делает простым синтез расплывчатых оценок и удовлетворяет (с помощью регулировочных шкал) всем практическим требованиям.

Регулируя λ, можно добиться любого фиксированного уровня α оптимального правила: если ошибка оказалась меньше α, то λ следует чуть уменьшить, стимулируя тем самым расплывчатость, а если больше — то увеличить, и далее снова посмотреть, что при этом получится с ошибкой оптимального правила.

Минимизация по $\partial\in\mathscr{D}_{\forall}$ составного риска с параллельной «подгонкой» значения λ есть способ синтеза оптимальных правил фиксированного уровня α, обладающих минимальной средней шириной в классе правил уровня α, т. е. решающих задачу:

$$\min_{\partial:\alpha^{\varkappa}(\partial)=\alpha}\overline{M}\Omega(\partial_y).$$

Оптимальные расплывчатые правила при заданных совместных плотностях вероятностей. Совместная плотность являет собой крайнее исключение, когда данные о практически всех вероятностях событий, связанных как с x, так и с y, имеются в виде точных значений. Выделение такого случая можно было бы

назвать отступлением от общей нашей тенденции иметь дело с грубыми моделями, описываемыми конечным числом данным. Тем не менее такой отход позволяет, во-первых, уяснить влияние шкалы расплывчатости на вид оптимального правила, и во-вторых, продемонстрировать простоту нового аппарата синтеза доверительных правил на классической «почве» распределений вероятностей.

Пусть $\mathscr{X}=\mathscr{R}^k$, т. е. $\mathbf{x}=(x_1, ..., x_k)$ и $\mathscr{Y}=\mathscr{R}^n$, т. е. $\mathbf{y}=(y_1, ..., y_n)$, и пусть СИМ задается точной совместной плотностью $p(\mathbf{x}, \mathbf{y})$ по мере-длине на $\mathscr{R}^{k+n}$ (обобщение на другие пространства и меры для нас не принципиально). Тогда частные плотности будут равны: $p(\mathbf{y})=\int p(\mathbf{x}, \mathbf{y})\,d\mathbf{x}$, $p(\mathbf{x})=\int p(\mathbf{x}, \mathbf{y})d\mathbf{y}$. Считаем автоматически выполненными все необходимые условия измеримости по отношению к алгебре отрезков на $\mathscr{R}^{k+n}$, обязанной сопровождать плотность (см. § 1.5).

Для (измеримых) оценок $\partial_y(x)$ ошибка будет точной величиной, равной $\alpha(\partial)=1-\int\int\partial_y(\mathbf{x})p(\mathbf{x}, \mathbf{y})d\mathbf{x}d\mathbf{y}$. Перейдем к составлению риска и нахождению оптимальных оценок для разных шкал расплывчатости.

Взвешенная шкала. Составной риск для нее записывается:

$$\Pi_{\lambda}(\partial)=1+\int\int\partial_{\mathbf{y}}(\mathbf{x})\,[-p(\mathbf{x}, \mathbf{y})+\lambda p(\mathbf{y})\,q(\mathbf{x})]\,d\mathbf{x}\,d\mathbf{y}.$$

Так как $\partial_{\mathbf{y}}(\mathbf{x})$ заключено в пределах от 0 до 1, то правилом, минимизирующим этот риск, будет:

$$\partial^*_{\mathbf{y}}(\mathbf{x})=\begin{cases}1 \text{ при } p(\mathbf{x}, \mathbf{y})>\lambda p(\mathbf{y})\,q(\mathbf{x}),\\ 0 \text{ при } p(\mathbf{x}, \mathbf{y})<\lambda p(\mathbf{y})\,q(\mathbf{x}).\end{cases} \tag{6.3}$$

Оптимальным уровня α будет найденное правило при таком выборе λ, чтобы

$$1-\alpha=M^{\mathbf{xy}}\partial^*_{\mathbf{y}}(\mathbf{x})=\iint\limits_{p(\mathbf{x},\mathbf{y})>\lambda p(\mathbf{y})q(\mathbf{x})}p(\mathbf{x}, \mathbf{y})\,d\mathbf{x}\,d\mathbf{y},$$

где неравенство снизу определяет совместную на $\mathscr{X}\times\mathscr{Y}$ область интегрирования. Обозначим $p_{\mathbf{y}}(\mathbf{x})=p(\mathbf{x}, \mathbf{y})/p(\mathbf{y})$ и назовем *апостериорной плотностью*. Оптимальное решающее правило будет индикаторным и состоит в сравнении при каждом $\mathbf{y}$ отношения $p_{\mathbf{y}}(\mathbf{x})/q(\mathbf{x})$ с порогом λ и присвоении $\partial^*_{\mathbf{y}}$ значения 1, если это отношение выше порога, и 0, если ниже.

Функция $q(\mathbf{x})$ определяет вес штрафов в зависимости от $\mathbf{x}$. Для интегральной шкалы $q(\mathbf{x})\equiv 1$ и тогда $\partial^*{}_{\mathbf{y}}$ определяется сравнением апостериорной плотности $p_{\mathbf{y}}(\mathbf{x})$ с порогом λ. При $q(\mathbf{x})=p(\mathbf{x})$ с порогом будет сравниваться отношение $p(\mathbf{x}, \mathbf{y})/[p(\mathbf{x})p(\mathbf{y})]$.

Возможна замена усредненного на максимальный по $\mathbf{y}$ штраф $\sup_{\mathbf{y}}\int\partial_{\mathbf{y}}(\mathbf{x})\ q(\mathbf{x})d\mathbf{x}$, имеющий смысл, если при любом $\mathbf{y}$ требуется контролиро-

вать ширину правила $\partial_y(x)$ по переменной x. Тогда составной риск приобретает вид:

$$\Pi_\lambda(\partial) = 1 - M^{xy}\partial_y(x) + \lambda \sup_y \int \partial_y(x)\, q(x)\, dx.$$

Записав $\inf_\partial \Pi_\lambda(\partial) = \sup_{p_0(y)} \inf_\partial [1 - M^{xy}\partial_y(x) + \lambda p_0(y) \int \partial_y(x)\, q(x)\, dx]$ (где, используя соответствующие утверждения теории игр [22], inf и sup поменяли местами), мы придем к правилу (6.3), в котором $p(y)$ заменено на $p_0(y)$, определяемое минимизацией выражения $\int\int [p(x, y) - \lambda p_0(y) q(x)]\, dx dy$. В некотором смысле $p_0(y)$ выбирается таким образом, чтобы произведение $p_0(x)q(x)$ было по возможности более «похожим» на $p(x, y)$.

Обобщенная шкала. Составной риск

$$\Pi_\lambda(\partial) = 1 - \int\int [-\partial_y(x)\, p(x, y) + \lambda\, \partial_y(x)^\gamma\, p(y)\, q(x)]\, dx\, dy.$$

Оптимальное решающее правило может быть найдено минимизацией при каждом x и y выражения в квадратных скобках. После несложных вычислений приходим к следующему его виду:

$$\partial_y^*(x) = \min\left\{1, \left[\frac{p(x, y)}{\lambda\gamma\, p(y)\, q(x)}\right]^{1/(\gamma-1)}\right\}, \quad \gamma \geqslant 1,$$

где λ находится из уравнения

$$\alpha(\partial) = \int\int p(x, y)\left[1 - \left(\frac{p(x, y)}{\lambda\gamma\, p(y)\, q(x)}\right)^{1/(\gamma-1)}\right]^+ dx\, dy = \alpha.$$

Параметр γ определяет наше отношение к сомнительным решениям $\partial_y(x) < 1$. При $\gamma = 1$ возвращаемся к (6.3). При $\gamma = 2$ получаем $\partial_y(x) = \min\{1, p_y(x)/[2\lambda q(x)]\}$. Интересно отметить, что при $q(x) \equiv 1$ данное решающее правило совпадает по форме с апостериорной плотностью, возможно, усеченной сверху величиной 2λ.

Замечания. 1. Выводы настоящего раздела согласно теореме 5.1 останутся в силе, если синтез производится в режиме полуоптимизма $\varkappa = 1/2$, а СИМ задана интервальной плотностью $\underset{\sim}{p}(x, y)$, $\tilde{p}(x, y)$, так что $\int\int [\underset{\sim}{p}(x, y) + \tilde{p}(x, y)]\, dx dy = 2$. Тогда в предыдущие выражения нужно подставлять $p(x, y) = [\underset{\sim}{p}(x, y) + \tilde{p}(x, y)]/2$.

2. При дискретных x и y плотности заменяются на вероятности, а интегралы — на суммы.

Достаточные классы расплывчатых правил. Нам неважно опять, каковыми являются $\mathscr{X}$ и $\mathscr{Y}$. Это могут быть числовые или дискретные пространства, поэтому вместо оценок в общем говорим о расплывчатых решениях. Неважно также, какой является при этом шкала расплывчатости; она может быть совсем другой, чем рассмотренные выше. Важно, что для шкал должно выполняться очевидное условие

$$d_1(x) > d_2(x) \Rightarrow \Omega(d_1) \geqslant \Omega(d_2), \tag{6.4}$$

т. е. более расплывчатым и более неопределенным решениям должно соответствовать большее (по крайней мере, не меньшее) число по шкале Ω. Как и выше, полагаем $\mathscr{D} = \mathscr{D}_\forall$ — класс всех оценок.

Выпишем выражение (6.2) для составного риска, раскрыв подробнее ошибку:

$$\Pi_\lambda^\varkappa(\partial) = 1 - \varkappa \underline{M}\, \partial_y + (\varkappa - 1)\, \overline{M}\, \partial_y + \lambda\, \overline{M}\, \Omega(\partial_y). \tag{6.5}$$

Следующая теорема о достаточности является ключевой для настоящей главы. Хотя достаточность здесь специальная, так как привязывается к конкретному риску (6.5), но зато напрямую обращается к структуре первичных признаков СИМ, а заодно и позволяет крайне просто найти верхнюю (наименее благоприятную) ошибку, наиболее важную для нас.

Теорема 6.1. *При СИМ $\langle \overline{M}\mathscr{G} \rangle$, пессимизме $\varkappa \geqslant 1$ и условии (6.4), достаточном в смысле составного риска (6.5) классом расплывчатых правил будет усеченный снизу осью абсцисс подкласс вторичных, не превышающих 1 признаков*

$$\mathscr{D}^* = \{\partial_y(x) = [c_0 - \sum c_i^+ g_i(x, y)]^+ : c_0 \in \mathscr{R},\ c_i^+ \in \mathscr{R}^+,$$
$$g_i \in \mathscr{G},\ \partial_y(x) \leqslant 1\}. \tag{6.6}$$

Причем в него могут быть включены только те $\partial^*{}_y(x)$, для которых нижняя надежность равна

$$\underline{M}\, \partial_y^* = c_0 - \sum c_i^+\, \overline{M}\, g_i. \tag{6.7}$$

Доказательство. Для любого ∂_y согласно следствию к теореме 1.1, переписанному для нижнего среднего, имеем $\underline{M}\partial = \sup_{-\partial \leqslant -g \in \mathscr{L}^+ \mathscr{G}} \underline{M}g$. Отсюда заданному $\varepsilon > 0$ можно всегда подыскать такую $-g_\varepsilon \in \mathscr{L}^+\mathscr{G}$, что $-g_\varepsilon \geqslant -\partial$ и $M\partial \leqslant Mg_\varepsilon + \varepsilon$. С учетом неравенств $g_\varepsilon(x, y) \leqslant g_\varepsilon(x, y)^+ \leqslant \partial_y(x)$, где плюс означает взятие неотрицательной части, имеем $\underline{M}\partial - \varepsilon \leqslant \underline{M}g_\varepsilon \leqslant \underline{M}g^+{}_\varepsilon$, $\overline{M}\partial \geqslant \overline{M}g^+{}_\varepsilon$. Функция $g_\varepsilon(x, y)^+$ есть решающее правило, так как $0 \leqslant \overline{g^+}_\varepsilon \leqslant \overline{\partial} \leqslant 1$. Теперь из (6.5) на основании условия (6.4) и найденных отношений получаем:

$$\Pi_\lambda^\varkappa(\partial) = 1 - \varkappa\, \underline{M}\partial + (\varkappa - 1)\, \overline{M}\partial + \lambda \overline{M}\Omega(\partial) \geqslant 1 - \varkappa\, \underline{M}g_\varepsilon^+ +$$
$$+ (\varkappa - 1)\, \overline{M}\, g_\varepsilon^+ + \lambda \overline{M}\Omega\left(g_\varepsilon^+\right) - \varkappa\varepsilon = \Pi_\lambda^\varkappa\left(g_\varepsilon^+\right) - \varkappa\varepsilon,$$

что и доказывает достаточность класса $\mathscr{D}^*$ правил.

Осталось доказать последнюю часть теоремы. Если в (6.7) вместо равенства стоит неравенство, то по следствию к теореме 1.1 существует другой вторичный признак $-g \in \mathscr{L}^+\mathscr{G}$, для которого $g(x, y)^+ \leqslant \partial_y(x)$ и $\underline{M}\partial = \underline{M}g + \varepsilon$, а так как в силу первого неравенства $\overline{M}g^+ \leqslant \overline{M}\partial$, то риск (6.5) у правила g^+ будет (при $\varepsilon \to 0$) не больше, чем у ∂, и последнее может быть исключено из достаточного класса.

Заметим, что обозначение $\overline{M}g_i$ в правой части формулы (6.7) вместо $\overline{M}g_i$ означает, что оставляются лишь полулинейные комбинации согласованных первичных признаков.

Следующие ниже утверждения получаются из теоремы 6.1 расшифровкой класса вторичных признаков при разных способах задания СИМ. Эти утверждения тут же сопровождаются поясняющими примерами.

Следствие 1. *Пусть СИМ разложима* $\mathcal{M}^{xy}=\mathcal{M}^x\mathcal{M}^{y}{}_x$, *причем* $\mathcal{M}^x=\langle\bar{M}\mathcal{H}\rangle$, $\mathcal{M}^{y}{}_x=\langle\bar{M}_x\Psi\rangle$ *определены своими первичными средними* $\bar{M}h_i(x)$, $h_i\in\mathcal{H}$; $\bar{M}_x\psi_j(y)$, $\psi_j\in\Psi$. *Тогда при* $\varkappa\geqslant 1$ *достаточным будет класс правил*

$$\mathcal{D}^*=\{\partial_y(x)=[c_0-\sum a_i^+ h_i(x)-\sum c_i^+(x)(\psi_i(x,y)-$$

$$-\bar{M}_x\psi_i)]^+ : h_i\in\mathcal{H},\ \psi_i\in\Psi\},$$

причем таких, что $\partial_y(x)\leqslant 1$, $\underline{M}^{y}{}_x\partial_y(x)=c_0-\sum a^+{}_i\bar{M}h_i$.

Пример 6.1. Пусть $\mathcal{M}^{y}{}_x$ определяется всего одним первичным признаком $\psi(y)$ и значением $\bar{M}_x\psi$, зависящим от x, а $\mathcal{M}^x=\mathcal{J}^x$ — голая модель. Тогда $\mathcal{D}^*$ составляют $\partial_y(x)=[c_0-c^+(x)(\psi(y)-\bar{M}_x\psi)]^+$ при выборе коэффициентов c_0 и $c^+(x)\geqslant 0$, удовлетворяющих неравенству $c_0+c^+(x)(\psi(y)-\bar{M}_x\psi)\leqslant 1$, причем $\underline{M}\partial=c_0$. Мы видим, что так как $c^+(x)$ есть произвольная неотрицательная функция, то правила $\partial_y(x)$ как функции переменной x при каждом заданном y могут иметь произвольный вид, но как функций переменной y их вид целиком определяется первичной функцией $\psi(y)$.

Следствие 2. *Пусть СИМ задана функциональным представлением* $y=V_x\xi$ *и моделью* $\mathcal{M}^{x\xi}=\langle\bar{M}^x\mathcal{H}\rangle\langle\bar{M}^{\xi}\Psi\rangle$ *(т. е. флуктуации* ξ *свободны от* x*). Пусть оператор* V_x *обратим:* $\xi=V_x^{-1}y$. *Тогда при* $\varkappa\geqslant 1$ *достаточными в смысле составного риска (6.5) будут решающие правила*

$$\mathcal{D}^*=\{\partial_y(x)=[c_0-\sum a_i^+ h_i(x)-\sum c_i^+(x)(\psi_i(V_x^{-1}y)-$$

$$-\bar{M}\psi_i)]^+ : h_i\in\mathcal{H},\ \psi_i\in\Psi\},$$

причем такие, что $\partial_y(x)\leqslant 1$, $\underline{M\partial}=c_0-\sum a^+{}_i\bar{M}h_i$.

Пример 6.2. Пусть наблюдения есть смесь сигнала x и шума: $y_i=x+\xi_i$, $i=1,\ \ldots,\ n$, $x\in\mathcal{R}$, $\xi_i\in\mathcal{R}$, или в векторном виде $\mathbf{y}=\mathbf{1}x+\xi$, где $\mathbf{1}=(1,\ \ldots,\ 1)^T$ — единичный вектор. Считаем, что шум ξ свободен от x и имеет модель $\mathcal{M}^{\xi}$, заданную первичными значениями $\bar{M}\psi_j(\xi)$, $\psi_j\in\Psi$, а модель сигнала x — значениями $\bar{M}h_i(x)$, $h_i\in\mathcal{H}$. Тогда класс $\mathcal{D}^*$ образуют правила вида $\partial_{\mathbf{y}}(x)=[c_0-\sum a^+{}_i h_i(x)-\sum c^+{}_j(x)(\psi_j(\mathbf{y}-\mathbf{1}x)-\bar{M}\psi_j)]^+$, при $\partial_y(x)\leqslant 1$, $\underline{M\partial}=c_0-\sum a^+{}_i\bar{M}h_i$.

Откажемся в следствии 2 от предположения свободы ξ от x.

Следствие 3. *Пусть* $y=V_x\xi$, *причем оператор* V_x *обратим, и пусть первичными средними заданы частные* $\mathcal{M}^x=\langle\bar{M}\mathcal{H}\rangle$, *и* $\mathcal{M}^{\xi}=\langle\bar{M}\Psi\rangle$. *Тогда при* $\varkappa\geqslant 1$ *и риске (6.4) достаточным будет следующий класс правил:*

$$\mathcal{D}^*=\{\partial_y(x)=[c_0-\sum a_i^+ h_i(x)-\sum c_i^+\psi_i(V_x^{-1}y)]^+ : h_i\in$$

$$\in\mathcal{H},\ \psi_i\in\Psi\},$$

причем таких, что $\partial_y(x)\leqslant 1$ *и* $\underline{M\partial}=c_0-\sum a^+{}_i\bar{M}h_i-\sum c^+{}_i\bar{M}\psi_i$.

Пример 6.3. Пусть $y_i=\sum_{j=1}^{k} w_{ij}x_j+\xi_i$, $i=1,\ \ldots,\ n$, или в векторных обозначениях $\mathbf{y}=\mathbf{W}\mathbf{x}+\xi$, где $\mathbf{W}$ — матрица w_{ij}. Пусть заданы только первичные средние $\bar{M}\psi_i$ флуктуаций ξ, а о связи ξ с x совершенно ничего не известно. Тогда достаточный (при $\varkappa\geqslant 1$) класс образуют правила вида $\partial_{\mathbf{y}}(\mathbf{x})=[c_0-\sum c^+{}_i\psi_i(\mathbf{y}-\mathbf{W}\mathbf{x})]^+$ при таком выборе коэффициентов, что $\partial_{\mathbf{y}}(\mathbf{x})\leqslant 1$ и $\underline{M\partial}=c_0-\sum c^+{}_i\bar{M}\psi_i$. Нетрудно видеть, если перейти к $k=1$, $\mathbf{W}=1$, что этот класс проще того, что получился в предыдущем примере.

Следствие 4. *Пусть* $y=V_x\xi$, *причем* V_x *обратим, и пусть об* x *ничего не известно* $\mathcal{M}^x=\mathcal{J}^x$, *как и о его связи с* ξ *(либо* x *свободен от* ξ*), а* $\mathcal{M}^{\xi}=\langle\bar{M}\Psi\rangle$ *основывается на первичных признаках* $\psi_i(\xi)$, $\psi_i\in\Psi$. *Тогда достаточный класс образуют правила вида* $[c_0-\sum c^+{}_i\psi_i(\xi)]^+\leqslant 1$, *куда подставляется* $\xi=V_x^{-1}y$, *причем* $\underline{M\partial}=c_0-\sum c^+{}_i\bar{M}\psi_i$.

Например, если $\mathbf{y}=\mathbf{W}\mathbf{x}+\xi$, то оптимальное правило есть функция $\mathbf{y}-\mathbf{W}\mathbf{x}$: $\partial_{\mathbf{y}}(\mathbf{x})=\partial(\mathbf{y}-\mathbf{W}\mathbf{x})$, в этом виде его и нужно искать.

Замечание. Если СИМ задается в виде семейства $\bigvee\mathcal{M}_\theta{}^{xy}$, или в виде пересечений $\bigwedge\mathcal{M}_\theta{}^{xy}$, а класс $\mathcal{D}^*$ достаточен при каждой $\mathcal{M}_\theta{}^{xy}$, то он будет достаточным для объединений и для пересечений. Это верно хотя бы потому, что классы вторичных признаков, если они одни и те же для $\mathcal{M}_\theta{}^{xy}$, при объединениях и пересечениях сохраняются.

Оптимизм и достаточность. Согласно предыдущему разделу режим пессимизма (т. е. $\varkappa\geqslant 1$) позволяет произвести предварительную обработку, состоящую в редукции к достаточному классу $\mathcal{D}^*$ расплывчатых правил, что облегчает решение статистической задачи в плане нахождения оптимального правила.

Возникают два вопроса. Один — возможна ли при оптимизме $\varkappa<1$ такая предварительная обработка? Общего ответа на этот вопрос дать нельзя. Причина в том, что любая редукция наблюдений ведет к потере данных и расширению интервальной модели. В свою очередь, это уменьшает $\underline{M\partial}$ и увеличивает $\bar{M}\partial$, т. е. эти величины как слагаемые риска (6.5) будут меняться в противоположных направлениях и при $\varkappa<1$ трудно уследить, в какую сторону поведет их взвешенную сумму, тем более делать общие утверждения на этот счет, составляющие смысл достаточности.

Отсюда возникает другой вопрос, насколько целесообразно пользоваться определенными при $\varkappa\geqslant 1$ достаточными классами $\mathcal{D}^*$, если $\varkappa<1$? Отчасти ответ освещается следующим легко проверяемым утверждением.

Утверждение 6.2. *Пусть для всех правил достаточного (при* $\varkappa\geqslant 1$*) класса* $\mathcal{D}^*$ *верно:* $\bar{M}\partial^*{}_y=\sup_{x,\,y}\partial^*{}_y(x)$. *Тогда класс* $\mathcal{D}^*$ *будет достаточным и при* $\varkappa<1$.

Итак, видим, что рассмотренные в предыдущем пункте достаточные классы правил $\mathcal{D}^*$ будут достаточными и при $\varkappa<1$, если

в $\alpha^{\varkappa}(\partial)$ «нейтрализовано» слагаемое с $\overline{M}\partial_y$. Иначе возникает чисто математическая возможность уменьшить ошибку $\alpha^{\varkappa}(\partial)$ за счет искусственного увеличения $\overline{M}\partial_y$ при сохранении неизменным $\underline{M}\partial_y$. Приведем пример реализации такой возможности, чтобы осмыслить всю абсурдность, к которой можно прийти, если пойти возникшим путем.

Пример 6.4. Допустим СИМ задана точными вероятностями $p_{ij}=P(x\in A_i,\ y\in B_j)$ на произведении разбиений $A_i\in\mathscr{A}^{\varkappa}_{\Sigma}$ и $B_j\in\mathscr{B}^{y}_{\Sigma}$ соответственно пространств $\mathscr{X}$ и $\mathscr{Y}$. Оптимальное правило по теореме 6.1 записывается в виде $\partial_y(x)=1-\sum\sum c_{ij}A_i(x)B_j(y)$, где $0\leqslant c_{ij}\leqslant 1$. Пусть теперь $\varkappa<1$, а $\mathscr{X}$ и $\mathscr{Y}$ есть линейные пространства (скажем, $\mathscr{X}\in\mathscr{R}$, $\mathscr{Y}\in\mathscr{R}^n$), причем множества A_i и B_j содержат каждое более, чем по одной точке этого пространства. Тогда, добавив к каждому слагаемому $c_{ij}A_i(x)B_j(y)$, у которого $c_{ij}<1$, произведение дельта-функций $(1-c_{ij})\delta_{x_{(i)}}(x)\delta_{y_{(j)}}(y)$, где $x_{(i)}$ и $y_{(j)}$ есть произвольные представители множеств A_i и соответственно B_j, приходим к правилу

$$\partial^*_y(x)=1-\sum\sum[c_{ij}A_i(x)B_j(y)+(1-c_{ij})\delta_{x_{(i)}}(x)\delta_{y_{(j)}}(y)]$$

с дельта-выбросами, у которого за счет нулевой площади выбросов штраф тот же: $\int\partial^*_y(x)dx=\int\partial_y(x)dx$, но $\overline{M}\partial^*_y(x)=1\geqslant\overline{M}\partial_y(x)=\underline{M}\partial_y(x)=1-\sum\sum c_{ij}p_{ij}$. Цель выбросов — обеспечить в наиблагоприятнейшем случае единичную надежность $\overline{M}\partial^*=1$ и тем самым уменьшить риск при оптимизме.

Таким образом, отход при $\varkappa<1$ от достаточного класса $\mathscr{D}^*$ (соответствующего $\varkappa\geqslant 1$) может привести к «нарушению гармонии» в правилах, вряд ли практически оправданному, поэтому использование класса $\mathscr{D}^*$ представляется разумным и при $\varkappa<1$, даже если условия утверждения 6.2 не выполняются.

Симметрия статистических моделей и эквивариантность расплывчатых правил. Часто в задаче обнаруживается симметрия следующего содержания: синхронные изменения наблюдений эквивалентны соответствующему сдвигу параметра состояний. Мысль подтверждается на примере, когда $y_i=x+\xi_i$, где подъем всех y_i на число a равносилен сдвигу x на то же самое число. Нужно ожидать переноса названного свойства на оптимальные оценки x, а в общем, на правила, о чем и пойдет речь. Но в широком плане, когда синхронные изменения y и x математически постулируются как связанные между собой преобразования пространств $\mathscr{X}$ и $\mathscr{Y}$, или просто как особый класс преобразований произведения пространств $\mathscr{X}\times\mathscr{Y}$. Причем важно не только, чтобы эти преобразования не перекрещивали между собой $\mathscr{X}$ и $\mathscr{Y}$, но и чтобы они образовывали группу, куда входили последовательные преобразования, равно как и обратные. Это нужно для последующего применения принципов инвариантности.

Перейдем к формальному изложению. Рассмотрим группу $\mathscr{S}$ преобразований пространства $\mathscr{X}\times\mathscr{Y}$, сохраняющую на месте по отдельности $\mathscr{X}$ и $\mathscr{Y}$, т. е. такую группу, что каждое $\mathbf{s}\in\mathscr{S}$ отображает $\mathscr{X}$ на $\mathscr{X}$ и $\mathscr{Y}$ на $\mathscr{Y}$: $\mathbf{s}(x,\ y)=(\mathrm{s}_1x,\ \mathrm{s}_2y)$. Любое преобразование $\mathbf{s}$ записывается как совместная пара $(\mathrm{s}_1,\ \mathrm{s}_2)$ преобразований, первое из них s_1 действует только на $\mathscr{X}$, а второе s_2 — только на $\mathscr{Y}$. Называются s_1 и s_2 частными к $\mathbf{s}$ преобразованиями. Дадим обобщение на совместные преобразования $\mathbf{s}=(\mathrm{s}_1,\ \mathrm{s}_2)$ результатов § 5.4 (где рассматривались преобразования лишь одного пространства наблюдений $\mathscr{Y}$), используя введенные там понятия и некоторые результаты, интерпретированные к нашему случаю.

Статистическая интервальная модель называется *симметричной к группе* $\mathscr{S}$, если для любого преобразования из этой группы

$$\overline{M}f(\mathrm{s}_1x,\ \mathrm{s}_2y)=\overline{M}f(x,\ y),\quad\forall f\in\mathscr{F}.$$

Симметрия достигается, если симметричен первичный набор $\mathscr{G}$ модели в том смысле, что каждой $g(x,\ y)\in\mathscr{G}$ и каждому преобразованию $\mathbf{s}\in\mathscr{S}$ может быть указан другой первичный признак $g_{\mathbf{s}}(x,\ y)=g(\mathrm{s}_1x,\ \mathrm{s}_2y)\in\mathscr{G}$ такой, что $\overline{M}g_{\mathbf{s}}(x,\ y)=\overline{M}g(x,\ y)$.

Утверждение 6.3. *Если* $\mathscr{M}^{xy}$ *симметрична к группе* $\mathscr{S}$, *то частные модели* $\mathscr{M}^x$ *и* $\mathscr{M}^y$ *будут симметричны к своим частным преобразованиям* s_1x *и* s_2y, $(\mathrm{s}_1,\ \mathrm{s}_2)\in\mathscr{S}$.

Утверждение следует из определения и соответствует равенствам $\overline{M}f(\mathrm{s}_1x)\equiv\overline{M}f(x)$, $\overline{M}\varphi(\mathrm{s}_2y)\equiv\overline{M}\varphi(y)$.

Функция $f(x,\ y)$ называется *эквивариантной* по отношению к группе преобразований $\mathscr{S}$ пространства $\mathscr{X}\times\mathscr{Y}$, если $f(\mathrm{s}_1x,\ \mathrm{s}_2y)\equiv f(x,\ y)$, $(\mathrm{s}_1,\ \mathrm{s}_2)\in\mathscr{S}$. Модель $\mathscr{M}^{xy}$ называется *эквивариантной к группе* $\mathscr{S}$, если все ее первичные признаки эквивариантны к $\mathscr{S}$. Смысл эквивариантности — компенсировать изменения y соответствующим изменением x. Инвариантность СИМ есть особый случай эквивариантности, когда $(\mathrm{s}_1x,\ \mathrm{s}_2y)=(x,\ \mathrm{s}_2y)$, т. е. каждое $\mathbf{s}\in\mathscr{S}$ преобразует только пространство $\mathscr{Y}$ (s_1 оставляет x четко на месте).

Если СИМ $\langle\overline{M}\mathscr{G}\rangle$ *эквивариантна к* $\mathscr{S}$, *то она будет симметрична к* $\mathscr{S}$. В самом деле все функции класса $\mathscr{L}^+\mathscr{G}$ будут эквивариантными и

$$\overline{M}f(\mathrm{s}_1x,\ \mathrm{s}_2y)=\inf\{\overline{M}g:f(\mathrm{s}_1x,\ \mathrm{s}_2y)\leqslant g(x,\ y)\in\mathscr{L}^+\mathscr{G}\}=$$

$$=\inf\{\overline{M}g:f(x,\ y)\leqslant g(x,\ y)\in\mathscr{L}^+\mathscr{G}\}=\overline{M}f(x,\ y).$$

В соответствии с определением *решающее правило* будет (как f) *эквивариантным к* $\mathscr{S}$, если $\partial_{\mathrm{s}_2y}(\mathrm{s}_1x)\equiv\partial_y(x)$. Иначе может быть записано $\partial_{\mathrm{s}_2y}(x)\equiv\partial_y(\mathrm{s}_1^{-1}x)$, $(\mathrm{s}_1,\ \mathrm{s}_2)\in\mathscr{S}$. Например, для детерминированных правил имеем $\hat{x}_{\mathrm{s}_2y}=\mathrm{s}_1\hat{x}_y$, отсюда преобразование s_2y выборки y ведет к «смещению» решения до решения $\mathrm{s}_1\hat{x}$ в пространстве состояний.

Класс эквивариантных функций замкнут относительно линейных операций и нелинейных преобразований, т. е. из эквивариантности всех $g(x,\ y)\in\mathscr{G}$ следует эквивариантность их линейных комбинаций (а в общем — эквивариантность любых функций от них $\varphi(g(x,\ y))$). Отсюда на основании теоремы 6.1 имеем:

Утверждение 6.4. *При $\varkappa \geqslant 1$ достаточными для эквивариантных СИМ в смысле составного риска (6.5) будут эквивариантные правила.*

Значит, и оптимальные правила будут эквивариантны. Сейчас покажем, что в ряде случаев эквивариантными оптимальные правила должны быть и при симметричных СИМ. Назовем *шкалу инвариантной к* $\mathscr{S}$, если $\Omega(\partial_y(s_1 x)) = \Omega(\partial_y(x))$, $\forall (s_1, s_2) \in \mathscr{S}$. В случае $\Omega(\partial_y(x)) = \int \partial_y(x)\,dx$ шкала будет инвариантной, пожалуй, только к сдвигам $sx = s + x$.

Теорема 6.5. *Пусть СИМ симметрична к группе преобразований $\mathscr{S}$, а риск составной, определенный формулой (6.5), причем шкала Ω инвариантна к $\mathscr{S}$. Допустим, также выполнено любое из двух условий:*

А. *Оптимальное правило единственно или число этих правил конечно;*

Б. *Группа $\mathscr{S}$ дискретна. Тогда при любых $\varkappa$ оптимальное правило (а если их несколько, то хотя бы одно из них) будет эквивариантным к $\mathscr{S}$.*

Доказательство. Пусть оптимальное правило является единственным. Тогда, используя сначала инвариантность шкалы, а затем следующую из утверждения 6.3 симметрию частной модели, получим $\overline{M}\Omega(\partial_y(x)) = \overline{M}\Omega(\partial_y(s_1 x)) = \overline{M}\Omega(\partial_{s_2 y}(s_1 x))$. Отсюда

$$\Pi^{\varkappa}_{\lambda}(\partial_y(x)) = 1 - M^{(1-\varkappa)}\,\partial_y(x) + \lambda \cdot \overline{M}\,\Omega(\partial_y(x)) =$$

$$= 1 - M^{(1-\varkappa)}\,\partial_{s_2 y}(s_1 x) + \lambda \cdot \overline{M}\,\Omega(\partial_{s_2 y}(s_1 x)) = \Pi^{\varkappa}_{\lambda}(\partial_{s_2 y}(s_1 x))$$

и в силу единственности оптимального правила $\partial_y(x) = \partial_{s_2 y}(s_1 x)$.

Если теперь оптимальных правил не одно, а несколько, скажем, $\partial_y(x)_i$, $i = 1, \ldots, l$, то преобразованиями $s \in \mathscr{S}$ одно переходит в другое, так что заданным $s \in \mathscr{S}$ и i найдется такое j, что $\partial_{s_2 y}(s_1 x)_i = \partial_y(x)_j$. Производя равномерную рандомизацию между $\partial_y(x)_i$ (т. е. выбирая $\partial_y(x)_i$ с вероятностями $1/l$), приходим к эквивариантному решающему правилу. То же самое необходимо проделать, если группа $\mathscr{S}$ дискретна: $\mathscr{S} = (s^{(1)}, \ldots, s^{(l)})$. Теорема доказана.

Теорема проста для случая, когда $\mathscr{S}$ дискретна. Если же нет, то трудности может вызвать проверка условия **А**.

Максимальный инвариант относительно группы $\mathscr{S}$, поскольку по определению $\mathscr{S}$ не перемешивает между собой $\mathscr{X}$ и $\mathscr{Y}$, записывается как пара $I_1(x)$, $I_2(y)$, и тогда эквивариантное правило записывается через максимальный инвариант: $\partial_y(x) = \partial'_{I_2(y)}(I_1(x))$. При условиях утверждения 6.4 и теоремы 6.5 оптимальные правила нужно искать именно в таком виде.

6.2. ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ ПРИ ЗАДАННЫХ РАСПРЕДЕЛЕНИЯХ ВЕРОЯТНОСТЕЙ ФЛУКТУАЦИЙ

Предисловие. Отсюда мы начинаем «осаду» проблемы доверительного оценивания числового параметра $x \in \mathscr{R}$, априорных данных о котором совсем нет, с целью получения конкретных расплывчатых оценок $\partial_y(x)$, минимизирующих штраф $\overline{M}\Omega(\partial)$ за расплывчатость при заданной ошибке (уровне) $\alpha^{\varkappa}(\partial) = \alpha$, или что то же самое, минимизирующих составной риск $\Pi^{\varkappa}_{\lambda}(\partial) = \alpha^{\varkappa}(\partial) + \lambda \overline{M}\Omega(\partial)$ при соответствующем подборе λ, обеспечивающем уровень α. Основная задача — научиться пользоваться аппаратом. Здесь в качестве первого шага не лишне выяснить, как работает предлагаемый аппарат на привычной и хорошо «утоптанной» многими исследователями почве заданных распределений вероятностей и их семейств, и к чему он приводит. Этому и посвящен данный параграф. Неожиданным оказалось то, что и на «утрамбованном грунте» наш подход дает пробиться свежим «росткам» в виде как общих выражений для доверительных оценок векторного параметра x, так и получения новых доверительных интервалов для скалярного параметра.

Оценка регрессии при известной плотности вероятностей. Пусть $y_i = \sum_{j=1}^{k} W_{ij} x_j + \xi_i$, $i = 1, \ldots, n$, или в векторных обозначениях $\mathbf{y} = \mathbf{W}\mathbf{x} + \boldsymbol{\xi}$. Считается, что относительно вектора $\mathbf{x}^{\mathrm{T}} = (x_1, \ldots, x_k)$ ничего не известно, а вектор $\boldsymbol{\xi}^{\mathrm{T}} = (\xi_1, \ldots, \xi_n)$, именуемый флуктуациями, определен n-мерной плотностью $p_{\xi}(\mathbf{z})$, $\mathbf{z}^{\mathrm{T}} = (z_1, \ldots, z_n)$, по отношению к мере-длине на $\mathscr{R}^n$. Требуется оценить вектор $\mathbf{x}$. Относительно связи $\mathbf{x}$ и $\boldsymbol{\xi}$ никаких предположений не делается, она не известна (или так нам удобно это считать).

Согласно следствию 4 к теореме 6.1 достаточный (при $\varkappa \geqslant 1$) класс образуют правила вида $\partial(\mathbf{y} - \mathbf{W}\mathbf{x})$, где $\partial(\mathbf{z})$ измеримы относительно алгебры отрезков. Считая за штраф верхнюю среднюю площадь под $\partial_y(\mathbf{x})$, что соответствует интегральной шкале расплывчатости, и учитывая отсутствие сведений о $\mathbf{x}$, преобразуем штраф:

$$\overline{M}\Omega(\partial) = \overline{M}^{\mathbf{y}} \int \partial(\mathbf{y} - \mathbf{W}\mathbf{u})\,d\mathbf{u} = \overline{M}^{\boldsymbol{\xi}} \sup_{\mathbf{x}} \int \partial(\mathbf{W}\mathbf{x} + \boldsymbol{\xi} - \mathbf{W}\mathbf{u})\,d\mathbf{u} =$$

$$= \overline{M}^{\boldsymbol{\xi}} \int \partial(\boldsymbol{\xi} - \mathbf{W}\mathbf{u})\,d\mathbf{u} = \int p_{\xi}(\mathbf{z})\,d\mathbf{z} \int \partial(\mathbf{z} - \mathbf{W}\mathbf{u})\,d\mathbf{u} =$$

$$= \int\int p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u})\,\partial(\mathbf{z})\,d\mathbf{z}\,d\mathbf{u}.$$

В силу измеримости $\partial(\mathbf{z})$ ошибка будет точной, равной $\alpha(\partial) = 1 - M\partial(\boldsymbol{\xi}) = 1 - \int \partial(\mathbf{z}) p_{\xi}(\mathbf{z})\,d\mathbf{z}$, и составной риск

$$\Pi^{\varkappa}_{\lambda}(\partial) = 1 - \int \partial(\mathbf{z})\,p_{\xi}(\mathbf{z})\,d\mathbf{z} + \lambda \int\int p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u})\,\partial(\mathbf{z})\,d\mathbf{z}\,d\mathbf{u} =$$

$$= 1 + \int \partial(\mathbf{z})\,[\lambda \int p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u})\,d\mathbf{u} - p_{\xi}(\mathbf{z})]\,d\mathbf{z}.$$

Минимизирует риск интервальная оценка $\partial^*(\mathbf{z})$, равная 1 при $\lambda \int p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u})\,d\mathbf{u} - p_{\xi}(\mathbf{z}) \leqslant 0$ и 0 в противном случае. Цена будет $v = 1 - \int [p_{\xi}(\mathbf{z}) - \lambda \int p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u})\,d\mathbf{u}]^{+} d\mathbf{z}$. После замены $\lambda_{\alpha} = 1/\lambda$ оценка записывается $\int p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u})\,d\mathbf{u} / p_{\xi}(\mathbf{z}) \leqslant \lambda_{\alpha}$ (указываются значения $\mathbf{z}$, при которых $\partial^*(\mathbf{z}) = 1$). Здесь λ_{α} находится по заданной ошибке из уравнения $\alpha(\partial^*) = 1 - \int \partial^*(\mathbf{z})\,p_{\xi}(\mathbf{z})\,d\mathbf{z} = \alpha$.

Выражение для доверительного интервала в явном виде получается заменой $\mathbf{z}=\mathbf{y}-\mathbf{W}\mathbf{u}$ и приобретает окончательный общий вид:

$$\partial_{\mathbf{y}}^{*}(\mathbf{x})=1 \text{ при } \frac{\int p_{\xi}(\mathbf{y}-\mathbf{W}\mathbf{x}+\mathbf{W}\mathbf{u})\,d\mathbf{u}}{p_{\xi}(\mathbf{y}-\mathbf{W}\mathbf{x})} \leqslant \lambda_{\alpha}, \tag{6.8}$$

иначе значение правила равно 0.

Нужно разрешить это неравенство относительно $\mathbf{x}$ и подобрать λ_α, что и ведет к доверительной индикаторной оценке заданного уровня α, а при одномерном $\mathbf{x}$ — к доверительному интервалу. Отметим, что, во-первых, постоянные множители левой части (6.8), не зависящие от $\mathbf{y}-\mathbf{W}\mathbf{x}$, удобно перенести в правую часть, и во-вторых, целесообразно искать упрощения с помощью монотонных нелинейных преобразований обеих частей, что сразу же и продемонстрируем.

Рассмотрим тот случай, когда ξ есть нормальный вектор с нулевым средним и матрицей корреляций $\mathbf{B}$, что соответствует плотности $p_{\xi}(\mathbf{z})=((2\pi)^{n}\det\mathbf{B})^{-1/2}\exp(-\mathbf{z}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{z}/2)$. Тогда по (6.8) после несложных вычислений находим

$$\frac{\int p_{\xi}(\mathbf{z}+\mathbf{W}\mathbf{u})\,d\mathbf{u}}{p_{\xi}(\mathbf{z})}=\frac{(\sqrt{2\pi})^{k}}{\sqrt{\det\mathbf{W}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{W}}}\times$$

$$\times\exp\left\{\frac{1}{2}\mathbf{z}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{W}\,(\mathbf{W}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{W})^{-1}\,\mathbf{W}^{\mathrm{T}}\mathbf{B}^{-1}\,\mathbf{z}\right\}.$$

Отсюда отбрасыванием постоянного множителя, логарифмированием и подстановкой $\mathbf{z}=\mathbf{y}-\mathbf{W}\mathbf{x}$ получаем индикаторную оценку, равную 1 в области значений $\mathbf{x}$, ограниченной неравенством

$$(\mathbf{y}-\mathbf{W}\mathbf{x})^{\mathrm{T}}\mathbf{B}^{-1}\,\mathbf{W}\,(\mathbf{W}^{\mathrm{T}}\mathbf{B}^{-1}\,\mathbf{W})^{-1}\,\mathbf{W}^{\mathrm{T}}\,\mathbf{B}^{-1}\,(\mathbf{y}-\mathbf{W}\mathbf{x})\leqslant\chi_{\alpha}^{2}(k), \tag{6.9}$$

где $\chi^2_\alpha(k)$ — критическая точка распределения хи-квадрата с k степенями свободы, и равную 0 вне этой области.

Пример 6.5. Рассмотрим простейший случай одного неизменного по i параметра x, который нужно оценить по выборке $y_i=x+\xi_i$, $i=1, \ldots, n$, где ξ_i независимы и нормально распределены с $M\xi_i=0$, $M\xi^2_i=1$. Тогда $k=1$, и из (6.9) получим доверительный интервал для x, записанный в форме неравенства $|\hat{y}-x|\leqslant\lambda_\alpha$, где и $\hat{y}=\sum_1^n y_i/n$ и $\lambda_\alpha=\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1-\alpha}{2}\right)$, а $\Phi(u)=$ $=\frac{1}{\sqrt{2\pi}}\int_0^u\exp(-u^2/2)\,du$ — функция Лапласа. В результате получается хорошо известный доверительный интервал для среднего при заданной дисперсии нормальной выборки

$$|\hat{y}-x|\leqslant\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1-\alpha}{2}\right).$$

Пусть теперь $y_i=xw_i+\xi_i$, $i=1, \ldots, n$, где ξ_i нормальны, зависимы и заданы корреляционной матрицей $\mathbf{B}=\{M\xi_i\xi_j,\ i,\ j=1, \ldots, n\}$,

а $M\xi_i=0$. Тогда оптимальной доверительной оценкой уже будет

$$\left|x-\frac{\mathbf{y}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{w}}\right|\leqslant(\mathbf{w}^{\mathrm{T}}\,\mathbf{B}^{-1}\,\mathbf{w})^{-1/2}\,\Phi^{-1}\left(\frac{1-\alpha}{2}\right). \tag{6.10}$$

Применительно к процессам $y_t=xw_t+\xi_t$, $M\xi_t=0$, $B(t,\tau)=$ $=M\xi_t\xi_\tau$ формула (6.10) переписывается:

$$\left|x-\frac{\int y_t\,h_t\,dt}{\int w_t\,h_t\,dt}\right|\leqslant\left(\int w_t\,h_t\,dt\right)^{-1/2}\Phi^{-1}\left(\frac{1-\alpha}{2}\right),$$

где h_t есть решение уравнения $\int B(t,\tau)h_\tau d\tau=w_t$.

При взвешенной шкале расплывчатости в знаменателе левой части неравенства в (6.8) появляется весовой множитель $q(\mathbf{y}-\mathbf{W}\mathbf{x})$, расширяющий интервал для тех x, которые имеют «повышенный приоритет» при сужении для остальных x.

В случае обобщенной шкалы $\Omega(\partial)=\int q(x)\partial_{\mathbf{y}}(x)^{\gamma}dx$, $\gamma>1$, оптимальная оценка уже не будет доверительным интервалом, а повторяет с некоторыми искажениями по форме плотность

$$\partial^{*}(\mathbf{z})=\min\left\{1;\left[\frac{p_{\xi}(\mathbf{z})}{\lambda_{\alpha}\,q(\mathbf{z})\int p_{\xi}(\mathbf{z}-\mathbf{W}\mathbf{u})\,d\mathbf{u}}\right]^{1/(\gamma-1)}\right\},$$

где λ_α ищется по заданному уровню α. Например, при одном параметре x, при $q(\mathbf{z})\equiv1$, нормальной плотности и независимых одродных флуктуациях с $M\xi_i=0$, $M\xi^2_i=\sigma^2$, имеем (рис. 6.1)

$$\partial_{y}^{*}(x)=\min\left\{1;\lambda\exp\left[-\frac{n}{2\sigma^{2}(\gamma-1)}(x-\hat{y})^{2}\right]\right\}.$$

Доверительное оценивание дисперсии. Пусть $y_i=x\xi_i$, $i=1, \ldots, n$, или в векторах $\mathbf{y}=x\xi$, где $x>0$ — оцениваемый параметр, свободный от ξ и описываемый голой моделью $\mathcal{J}^x$ (что соответствует полному отсутствию данных о x). Пусть известна плотность $p_\xi(\mathbf{z})$. Тогда достаточным при $\varkappa\geqslant1$ будет класс всех измеримых правил вида $\partial_{\mathbf{y}}(x)=\partial(\mathbf{y}/x)$. Ошибка равна $\alpha(\partial)=1-M\partial(\mathbf{y}/x)=$ $=1-\int p_\xi(\mathbf{z})\partial(\mathbf{z})d\mathbf{z}$, а штраф

$$\overline{M}^{\mathbf{y}}\int_0^\infty\partial(\mathbf{y}/u')\,du'=\overline{M}^{\xi}\sup_{x>0}\int_0^\infty\partial(x\,\xi/u')\,du'=$$

$$=\int p_{\xi}(\mathbf{z})\sup_{x>0}\int_0^\infty\partial(xz/u')\,du'\,d\mathbf{z}=\bar{\sigma}\int p_{\xi}(\mathbf{z})\int_0^\infty\partial(\mathbf{z}/u)\,du d\mathbf{z},$$

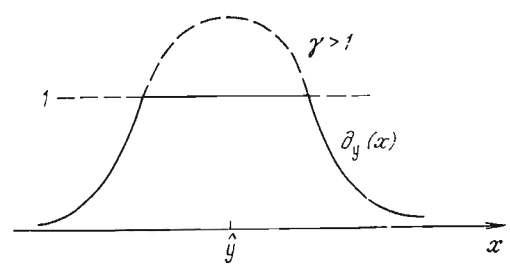


Рис. 6.1. Доверительная оценка среднего нормальной выборки при обобщенном штрафе

где произведена замена $u=u'/x$ и сделано допущение $\sup\limits_{x>0} x=\bar{\sigma}$. Необходимость такого допущения состоит в том, что ширина доверительной оценки растет с увеличением x, стремясь к ∞ при $x\to\infty$, вот и приходится вводить ограничение $x<\bar{\sigma}$ (причем $\bar{\sigma}$ можно считать любым сколь угодно большим числом). Тогда составной риск запишется

$$\Pi_\lambda(\partial)=1-\int \partial(\mathbf{z})\, p_\xi(\mathbf{z})\, d\mathbf{z}+\lambda\bar{\sigma}\int p_\xi(\mathbf{z})\int_0^\infty \partial(\mathbf{z}/u)\, du d\mathbf{z}.$$

Обозначая $\lambda\bar{\sigma}=1/\lambda_\alpha$ и производя во втором интеграле замену $\mathbf{z}'=\mathbf{z}/u$, учитывая при этом, что $d\mathbf{z}=u^n d\mathbf{z}'$, приводим риск к виду

$$\Pi_\lambda(\partial)=1-\int \partial(\mathbf{z})\left[p_\xi(\mathbf{z})-\lambda_\alpha^{-1}\int_0^\infty u^n\, p_\xi(u\,\mathbf{z})\, du\right] d\mathbf{z}.$$

Отсюда сразу же следует, что оптимальное доверительное решающее правило должно быть интервальным и иметь вид

$$\int_0^\infty u^n\, p_\xi(u\mathbf{y}/x)\, du/p_\xi(\mathbf{y}/x)\leqslant\lambda_\alpha, \tag{6.11}$$

где неравенство указывает значения x, при которых $\partial^*(\mathbf{y}/x)=1$ (иначе — 0); λ_α находится из уравнения $\int \partial^*(\mathbf{z})p_\xi(\mathbf{z})d\mathbf{z}=1-\alpha$, а при нахождении доверительного интервала совершена подстановка $\mathbf{z}=\mathbf{y}/x$.

В частном случае нормальной стандартной плотности флуктуаций оптимальный доверительный интервал в терминах переменной $\mathbf{z}=\mathbf{y}/x$ приобретает вид

$$\int_0^\infty u^n \exp\left(\|\mathbf{z}\|^2\frac{1-u^2}{2}\right)du=\frac{\exp(\|\mathbf{z}\|^2/2)}{\|\mathbf{z}\|^{n+1}}\int_0^\infty u^n\exp(-u^2/2)\,du\leqslant\lambda.$$

Отнесем интеграл к правой части, обозначив ее λ_n, и пусть $\underline{u}(\lambda_n)$ и $\bar{u}(\lambda_n)$ есть решения уравнения $\exp(u^2/2)=\lambda_n u^{n+1}$. Тогда искомый доверительный интервал запишется $\underline{u}^2(\lambda_n)\leqslant\|\mathbf{z}\|^2\leqslant\bar{u}^2(\lambda_n)$, если верно неравенство $\lambda_n\geqslant\exp[(n+1)^2/2]/(n+1)^{n+1}$. При равенстве доверительный интервал вырождается в точку, причем $\underline{u}^2(\lambda_n)=\bar{u}^2(\lambda_n)=n+1$, а при обратном неравенстве — бессмыслен.

Найдем λ_n в соответствии с заданной ошибкой α. Для этого обратим внимание на тот факт, что $\|\xi\|^2$ по определению есть случайная величина, распределенная по закону хи-квадрата с n степенями свободы. Подставляя плотность этого распределения, выводим

$$P[\underline{u}^2(\lambda_n)\leqslant\|\xi\|^2\leqslant\bar{u}^2(\lambda_n)]=\int_{\underline{u}^2(\lambda_n)}^{\bar{u}^2(\lambda_n)}\frac{v^{n/2}-1}{2^{n/2}\,\Gamma(n/2)}\,e^{-v/2}\,dv=1-\alpha,$$

где $\Gamma(\ )$ — гамма-функция. Отсюда находится значение $\lambda_{\alpha,n}$ как решение этого уравнения. Обозначая $\underline{u}_{\alpha,n}=\underline{u}(\lambda_{\alpha,n})$, $\bar{u}_{\alpha,n}=\bar{u}(\lambda_{\alpha,n})$ и подставляя $\|\mathbf{z}\|^2=\|\mathbf{y}\|^2/x^2$, получим окончательно следующий доверительный интервал уровня α:

$$\|\mathbf{y}\|^2/\bar{u}^2_{\alpha,n}\leqslant x^2\leqslant\|\mathbf{y}\|^2/\underline{u}^2_{\alpha,n}.$$

При $\alpha\to1$ имеем $\underline{u}^2_{\alpha,n}\uparrow n+1$, $\bar{u}^2_{\alpha,n}\downarrow n+1$ и доверительный интервал стягивается к точке $\|\mathbf{y}\|^2/(n+1)=\sum y_i^2/(n+1)$.

Таким образом, *получен доверительный интервал для дисперсии нормальной выборки, обладающий вообще минимальной расплывчатостью*, в том числе среди всех известных. Математическое ожидание наблюдений $\mathbf{y}$ считалось нулевым:

$$My_i=0,\ i=1,\ \dots,\ n.$$

Пусть теперь среднее $My_i=\theta$ совершенно неизвестно, т. е. $y_i=x\xi_i+\theta$. Считая θ свободным параметром, а ξ_i — независимыми стандартными нормальными с. в., путем вовлечения понятия инвариантности (к сдвигу) можно прийти к тому, что оптимальной оценкой дисперсии является доверительный интервал:

$$n\hat{\sigma}^2_{\mathbf{y}}/\bar{u}^2_{\alpha,n-1}\leqslant x^2\leqslant n\hat{\sigma}^2_{\mathbf{y}}/\underline{u}^2_{\alpha,n-1}, \tag{6.12}$$

где числа $\underline{u}_{\alpha,n-1}$ и $\bar{u}_{\alpha,n-1}$ были определены выше, а $\hat{\sigma^2}_{\mathbf{y}}=\sum_1^n(y_i-\hat{y})^2/n$.

6.3. ОЦЕНКА ПАРАМЕТРОВ РЕГРЕССИИ ПО ЭНЕРГЕТИЧЕСКИМ И КОРРЕЛЯЦИОННЫМ ДАННЫМ О ФЛУКТУАЦИЯХ

Обоснование. Наиболее обиходными, доступными сведениями о процессе помимо среднего являются его энергетические характеристики: средняя мощность, текущая мощность, а также корреляционные свойства, спектр. Причина их распространения в том, что верхняя граница средней мощности есть просто предел энергетических возможностей источника излучения. А корреляционные свойства обычно обязаны фильтру, будь он объект естественной природы как инерционность среды или искусственной в виде начальных каскадов приемника, который преодолевает процесс прежде, чем попасть на устройство обработки и принятия решений. Кстати, энергетические данные есть часть корреляционных, как и некоррелированность выборки, обычно достигаемая, если отсчеты процесса разнесены достаточно широко друг от друга.

Здесь строятся оптимальные доверительные оценки для регрессионных параметров, в частности, параметра сдвига, когда известны те или иные данные указанного сорта о флуктуациях. Оценки получаются расплывчатые, но никак не индикаторные, т. е. не в виде доверительных интервалов. Любопытно по ходу изложения проследить, как связывается расплывчатая форма этих оценок с видом первичных признаков, незримо присутствующих в энергетических, корреляционных и других исходных (первичных) данных, а также как по мере увеличения количества и улучше-

ния качества (точности) этих данных улучшаются оценки, становясь более точными, более надежными.

Оценка параметра сдвига при заданной мощности флуктуаций. Пусть $y_i = x + \xi_i$, $i = 1, \ldots, n$, или в векторных обозначениях $\mathbf{y} = \mathbf{1}x + \boldsymbol{\xi}$. Пусть единственное, что известно, так это $\overline{M}^{\xi} \|\boldsymbol{\xi}\|^2/n = \overline{\sigma^2}$ — верхняя средняя по отсчетам (и по ансамблю) мощность флуктуаций. Тогда СИМ $\mathscr{M}^{xy}$ будет определена единственным первичным средним, получаемым подстановкой $\xi_i = y_i - x$:

$$\overline{M}^{xy} \sum (y_i - x)^2/n = \overline{\sigma^2}. \tag{6.13}$$

Здесь сумма (далее все суммы по i от 1 до n) составляет первичный признак. Ищем при этих данных оптимальную оценку параметра x. Достаточным для поставленной задачи согласно следствию 4 к теореме 6.1 будет следующий класс оценок:

$$\partial_{\mathbf{y}}(x) = [c_0 - c \sum (y_i - x)^2/n]^+, \quad 0 \leqslant c_0 \leqslant 1, \quad c > 0, \tag{6.14}$$

причем для них $\underline{M}\partial_{\mathbf{y}}(x) = [c_0 - c\overline{\sigma^2}]^+$.

Чтобы искать оптимальную оценку, вычислим штраф. Обратим внимание, что оценка (6.14) как функция x есть парабола, что видно, если переписать входящий в (6.14) первичный признак следующим образом: $\sum (y_i - x)^2/n = \hat{\sigma}^2_{\mathbf{y}} + (x - \hat{y})^2$, где $\hat{y} = \sum y_i/n$ — выборочное среднее, $\hat{\sigma}^2_{\mathbf{y}} = \sum (y_i - \hat{y})^2/n$ — выборочная дисперсия. Парабола усечена снизу осью абсцисс. Интегрируя ее, получаем

$$\Omega(\partial) = \int \partial_{\mathbf{y}}(x)\, dx = \int_{\hat{y} - \sqrt{[c_0/c - \hat{\sigma}^2_{\mathbf{y}}]^+}}^{\hat{y} + \sqrt{[c_0/c - \hat{\sigma}^2_{\mathbf{y}}]^+}} [c_0 - c\hat{\sigma}^2_{\mathbf{y}} - c(x - \hat{y})^2]\, dx =$$

$$= \frac{4\left([c_0 - c\hat{\sigma}^2_{\mathbf{y}}]^+\right)^{3/2}}{3\sqrt{c}},$$

где пределы интегрирования соответствуют положительной части $\partial_{\mathbf{y}}(x)$. Осталось найти $\overline{M}^{\mathbf{y}} \Omega(\partial) = \frac{4}{3\sqrt{c}} \overline{M}^{\mathbf{y}} \left([c_0 - c\hat{\sigma}^2_{\mathbf{y}}]^+\right)^{3/2}$. Частная модель $\mathscr{M}^{\mathbf{y}}$ согласно теореме 2.1 основывается на первичном значении:

$$\overline{M}^{\mathbf{y}} \inf_x \sum (y_i - x)^2/n = \overline{M}^{\mathbf{y}}\, \hat{\sigma}^2_{\mathbf{y}} = \overline{\sigma^2}.$$

Здесь $\hat{\sigma}^2_{\mathbf{y}}$ есть первичный признак, а так как он всего один, то $\overline{M}^{\mathbf{y}}\, \hat{\sigma}^2_{\mathbf{y}} = \inf_{\mathbf{y}} \hat{\sigma}^2_{\mathbf{y}} = 0$, и поэтому

$$\overline{M}^{\mathbf{y}} \left([c_0 - c\hat{\sigma}^2_{\mathbf{y}}]^+\right)^{3/2} = \sup_{\mathbf{y}} \left([c_0 - c\hat{\sigma}^2_{\mathbf{y}}]^+\right)^{3/2} =$$

$$= \left(c_0 - c \inf_{\mathbf{y}} \hat{\sigma}^2_{\mathbf{y}}\right)^{3/2} = c_0^{3/2}.$$

Осталось найти $\overline{M}^{xy} \partial_{\mathbf{y}}(x)$. Правило $\partial_{\mathbf{y}}(x)$ не мажорируется первичным признаком (6.13), записанным $\hat{\sigma}^2_{\mathbf{y}} + (x - \hat{y})^2$, а он всего один, поэтому

$$\overline{M}^{xy} \partial_{\mathbf{y}}(x) = \sup_{x,\, \mathbf{y}} [c_0 - c \sum (y_i - x)^2/n]^+ = c_0.$$

В результате $\alpha^{\varkappa}(\partial) = 1 - \varkappa \overline{M}\partial_{\mathbf{y}}(x) + (\varkappa - 1)\overline{M}\partial_{\mathbf{y}}(x) = 1 - \varkappa[c_0 - c\overline{\sigma^2}]^+ + (\varkappa - 1)c_0$.

Теперь записываем составной риск:

$$\Pi_{\lambda}(\partial) = 1 - \varkappa [c_0 - c\overline{\sigma^2}]^+ + (\varkappa - 1)c_0 + \lambda c_0 \frac{4}{3} \sqrt{c_0/c}. \tag{6.15}$$

Здесь нужно рассматривать два случая: $c_0 - c\overline{\sigma^2} \geqslant 0$ и $c_0 - c\overline{\sigma^2} < 0$. Последний из них не представляет интереса, так как соответствует ошибке $\alpha^{\varkappa}(\partial) = (\varkappa - 1)c_0$, определяемой исключительно пессимизмом (а не знаниями о свойствах флуктуаций). Полагая $c_0 - c\overline{\sigma^2} \geqslant 0$, найдем сначала оптимальное значение c при заданном c_0. Для этого, дифференцируя правую часть составного риска по c и приравнивая 0, получаем

$$c = c_0 (2\lambda/(3\varkappa\overline{\sigma^2}))^{2/3} \quad \text{и}$$

$$\min_c \Pi_{\lambda}(\partial) = 1 + c_0 [(12\lambda^2 \varkappa \overline{\sigma^2})^{1/3} - 1].$$

Отсюда видно, что оптимальное значение c_0 должно быть равным либо 0, либо 1. Поскольку неравенство $c_0 - c\overline{\sigma^2} \geqslant 0$ исключает значение $c_0 = 0$, полагаем $c_0 = 1$. Коэффициент c находится по заданной ошибке α из уравнения $\alpha = 1 - \varkappa + c\varkappa\overline{\sigma^2} + \varkappa - 1 = c\varkappa\overline{\sigma^2}$ и равен $c^* = \alpha/(\varkappa\overline{\sigma^2})$. Причем условие $c_0 - c^*\overline{\sigma^2} > 0$ соответствует $\alpha/\varkappa < 1$.

Таким образом, *при* $\varkappa \geqslant 1$ *оптимальная оценка имеет вид*

$$\partial^*_{\mathbf{y}}(x) = [1 - c^* \sum (y_i - x)^2/n]^+ = [1 - c^* \hat{\sigma}^2_{\mathbf{y}} - c^* (x - \hat{y})^2]^+, \quad c^* = \frac{\alpha}{\varkappa\overline{\sigma^2}}. \tag{6.16}$$

Как функция переменной x, есть парабола, нанесенная на рис. 6.2 штриховой линией, максимальная при значении x, равном выборочному среднему $\hat{y}$, и усеченная снизу осью абсцисс. Максимальное значение параболы $1 - c^*\hat{\sigma}^2$, не равно 1, что говорит о неконтрастности оценки, а размах параболы у основания, равный $2\sqrt{1/c^* - \hat{\sigma}^2_{\mathbf{y}}}$, характеризует ее расплывчатость. Здесь неприятным является влияние $\hat{\sigma}^2_{\mathbf{y}}$, от которого, как оказывается, нетрудно избавиться, что сейчас и будет рассмотрено.

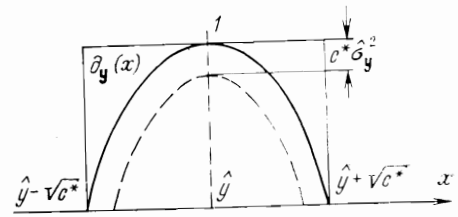


Рис. 6.2. Доверительная парабола при энергетических данных

ледем контрастную оценку (т. е. $\sup_x \partial_y \equiv 1$, $\inf_x \partial_y \equiv 0$) вида

$$\partial_y(x) = [1 - c^*(x - \hat{y})^2]^+, \quad c^* = \alpha/(\varkappa\overline{\sigma^2}). \tag{6.17}$$

Для нее $\underline{M}\,\partial_y(x) = \underline{M}\,[1 + c^*\hat{\sigma}_y^2 - c^*\sum(y_i - x)^2/n]^+ = 1 + c^*\underline{M}\hat{\sigma}_y^2 - c^*\overline{M}\sum(y_i - x)^2/n = 1 - c^*\overline{\sigma}^2 = 1 - \alpha/\varkappa$ и $\overline{M}\,\partial_y(x) = 1$, поэтому уровень $\alpha^{\varkappa}(\partial) = 1 - \varkappa\underline{M}\,\partial + (\varkappa - 1)\overline{M}\,\partial = \alpha$, а штраф $\overline{M}^{\mathbf{y}}\int \partial_y(x)\,dx = M^{\mathbf{y}}\,4/(3\sqrt{c^*}) = 4/(3\sqrt{c^*})$.

Таким образом, и штраф, и уровень оценок (6.16) и (6.17) совпадают, *контрастная оценка* (6.17) *также является оптимальной.* Это парабола по переменной x, нанесенная сплошной линией на рис. 6.2.

Можно сказать, что любая оценка, заключенная в пределах между оценками (6.16) и (6.17)

$$1 - c^*\hat{\sigma}_y^2 - c^*(x - \hat{y})^2]^+ \leqslant \partial_y(x) \leqslant [1 - c^*(x - \hat{y})^2]^+,$$

будет оптимальной. Причина такой неоднозначности оптимальной оценки объясняется тем, что как $\underline{M}\partial_y$, так и $\overline{M}\partial_y$, а следовательно, риск ориентирован на такой наименее благоприятный процесс, у которого $\hat{\sigma}^2{}_y = 0$, что означает постоянство реализаций флуктуаций $\xi_1 = \xi_2 = \ldots = \xi_n$, при этом оценки (6.16) и (6.17) совпадают. Это же и причина, по которой точность оценки, ее ширина не зависит от объема выборки n.

Оценки (6.16) и (6.17) называются оценками энергетического типа, так как используют только среднемощностные данные о флуктуациях. Сейчас мы выявим некоторые их дополнительные стороны и дадим обобщения.

Развитие энергетического типа оценивания. 1. Если помимо $\overline{\sigma^2}$ задана нижняя граница $\underline{\sigma^2} = \underline{M}^{\xi}\|\xi\|^2/n$ мощности флуктуаций, то оптимальная оценка не изменится, так как новое среднее, взятое за первичное, не меняет ни частной модели $\mathcal{M}^{\mathbf{y}}$, ни шкалы $\Omega(\partial)$, ни значений $\underline{M}\partial_y$, $\overline{M}\partial_y$.

2. Пусть помимо верхней мощности $\overline{\sigma^2}$ известно, что флуктуации имеют нулевые средние, а именно, либо $M\xi_i = 0$, $i = 1, \ldots, n$, либо $M\hat{\xi} = 0$. В обоих случаях в силу симметрии СИМ к перестановкам y_i оптимальный алгоритм будет принадлежать следующему достаточному классу:

$$\partial_y(x) = [c_0 + c_1\sum(y_i - x)/n - c\sum(y_i - x)^2/n]^+.$$

Слагаемое с коэффициентом c_1 даст лишь смещение в сторону по оси x оценки $\partial_y(x)$ как функции x, что, как нетрудно убедиться, приведет к увеличению риска, поэтому $c_1 = 0$. Частная модель $\mathcal{M}^{\mathbf{y}}$

как при нулевых данных $M\xi_i = 0$, так и когда этих данных о $M\xi_i$ нет, одна и та же, поэтому штраф рассматриваемой оценки, а следовательно, и риск будут такими же, как у оценок (6.16), (6.17). Таким образом, *данные о нулевом среднем значении флуктуаций не меняют решения исходной задачи*: оптимальная оценка будет иметь вид (6.16) или (6.17). Сказанное переносится и на случай, когда заданы границы: а) $\underline{M}\xi_i = -m$, $\overline{M}\xi_i = m$, $i = 1, \ldots, n$, либо б) $\underline{M}\hat{\xi} = -m$, $\overline{M}\hat{\xi} = m$.

3. Оценки (6.16) и (6.17) «настроены» на максимальный штраф $\overline{M}\Omega(\partial)$ и соответствуют $\varkappa \geqslant 1$. Тем не менее они остаются оптимальными и при $\varkappa < 1$, а также при взвешенном (коэффициентом пессимизма) штрафе вида

$$M^{\varkappa}\,\Omega(\partial) = \varkappa\overline{M}\int \partial_y(x)\,dx + (1 - \varkappa)\underline{M}\int \partial_y(x)\,dx.$$

В самом деле, $\overline{M}\partial_y(x) = \sup \partial_y(x)$ и согласно утверждению 6.2 достаточным при $\varkappa < 1$ будет тот же класс (6.14) оценок. Полагая $c_0 = 1$ (оптимальное значение), нетрудно убедиться, что переход к взвешенному штрафу не может повлиять на вид оптимальной оценки, так как значение c однозначно определяется ошибкой α.

4. Если искать оптимальную оценку в классе интервальных, то она примет вид $\hat{y} \pm \overline{\sigma}\sqrt{\varkappa/\alpha}$ и получается заменой в (6.17) усеченной параболы на прямоугольник единичной высоты с тем же основанием, как это выглядит на рис. 6.2. Ошибка этой оценки будет также равна α, однако расплывчатость ее по сравнению с (6.17) возрастет за счет увеличения площади под ней.

5. Мы считали, что о связи x и ξ ничего не известно. То же самое согласно следствию 4 теоремы 6.1 будет, если считать, что в представлении $\mathbf{y} = \mathbf{1}x + \xi$ о параметре x известно только, что он свободен от ξ, т. е. $\mathcal{M}^{x\xi} = \mathcal{M}^{\xi}\mathcal{J}^{x}$ и теперь уже x может произвольно подстраиваться под значения ξ. Можно показать, что допущение о свободе x от ξ не меняет оптимальной оценки.

6. Оценивание параметра регрессии. Пусть СИМ задана следующим образом: $y_i = w_i x + \xi_i$, $i = 1, \ldots, n$, $\overline{M}\|\xi\|^2/n = \overline{\sigma^2}$, где w_i — функция регрессии (скажем, форма детерминированного сигнала), x — параметр регрессии (амплитуда сигнала). Достаточный класс в этом случае образуют оценки вида

$$\partial_y(x) = [c_0 - c\,\|\mathbf{y} - \mathbf{w}\,x\|^2/n]^+, \quad \mathbf{w} = (w_1, \ldots, w_n)^{\mathrm{T}}.$$

Представим здесь квадрат нормы в виде $\|\mathbf{w}\|^2[\hat{\hat{\sigma}}^2{}_y - (x - \hat{y})^2]$ в обозначения $\hat{\hat{\mathbf{y}}} = \sum y_i w_i/\|\mathbf{w}\|^2$, $\hat{\hat{\sigma}}^2{}_y = \|\mathbf{y}\|^2/\|\mathbf{w}\|^2 - \hat{y}^2$ и запишем штраф (верхнюю среднюю ширину): $\lambda \sup_y (4c_0/3)\sqrt{c_0 n/(c\|\mathbf{w}\|^2) - \hat{\hat{\sigma}}^2{}_y}$. Супремум достигается при $\hat{\hat{\sigma}}^2{}_y = 0$, поэтому штраф равен $\lambda 4c_0\sqrt{c_0 n/c}/$

(3$\|\mathbf{w}\|$). Из сравнения с (6.15) видим, что оптимальной должна быть оценка (6.17) (или (6.16)), если заменить в ней $\hat{y}$ на $\hat{\hat{y}}$ ($\hat{\sigma}_{\mathbf{y}}$ на $\hat{\hat{\sigma}}_{\mathbf{y}}$), а c^* — на $\alpha\|\mathbf{w}\|^2/(\varkappa\bar{\sigma}^2 n)$.

7. Оценивание векторного параметра регрессии. Пусть $y_i=\sum_{j=1}^{k} w_{ij}x_j+\xi_i$, $i=1, \ldots, n$, или в векторном виде $\mathbf{y}=\mathbf{W}\mathbf{x}+\boldsymbol{\xi}$, и пусть $\bar{M}\|\boldsymbol{\xi}\|^2/n=\bar{\sigma}^2$ определяет $\mathscr{M}^{\xi}$. Оптимальной уровня α (при $\varkappa\geqslant 1$) будет оценка $[1-c^*\|\boldsymbol{\xi}\|^2]^+$, $c^*=\alpha/(\varkappa n\bar{\sigma}^2)$, куда подставляется $\boldsymbol{\xi}=\mathbf{y}-\mathbf{W}x$.

Записанная оценка неконтрастна, так как $\max_x\|\mathbf{y}-\mathbf{W}\mathbf{x}\|=\|\mathbf{y}_\perp\|\geqslant 0$, откуда $\max_x \partial_{\mathbf{y}}(\mathbf{x})\leqslant 1$, где $\mathbf{y}_\perp$ — проекция $\mathbf{y}$ в подпространство, ортогональное вектор-столбцам матрицы $\mathbf{W}$. Разложим y на две ортогональные составляющие: $\mathbf{y}=\mathbf{y}_W+\mathbf{y}_\perp$, где $\mathbf{y}_W=\mathbf{W}(\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{y}$ — проекция $\mathbf{y}$ в подпространство вектор-столбцов $\mathbf{W}$. Замена в вышезаписанной оценке $\mathbf{y}$ на $\mathbf{y}_W$ не изменяет $\bar{\alpha}$ (поскольку $\bar{M}\|\boldsymbol{\xi}\|^2=\bar{M}\|\boldsymbol{\xi}_W\|^2$, где $\boldsymbol{\xi}_W$ — аналогичная проекция $\boldsymbol{\xi}$), поэтому доверительная оценка многомерного параметра регрессии

$$\partial_{\mathbf{y}}(x)=[1-\alpha\,\|\mathbf{y}_W-\mathbf{W}x\|^2/(\varkappa n\bar{\sigma}^2)]^+$$

будет оптимальной уровня α и одновременно контрастной.

В частном случае, когда параметр регрессии скалярный, результат, как нетрудно убедиться, совпадает с рассмотренным в предыдущем пункте.

8. Неоднородные флуктуации. Пусть СИМ задана следующим образом:

$$\mathbf{y}=\mathbf{W}\mathbf{x}+\boldsymbol{\xi},\qquad \bar{M}\,\|\mathbf{A}\boldsymbol{\xi}\|^2=n\bar{\sigma}^2,$$

где $\mathbf{A}$ — матрица $n\times n$.

Совершая преобразование $\mathbf{z}=\mathbf{A}\mathbf{y}=\mathbf{A}\mathbf{W}\mathbf{x}+\mathbf{A}\boldsymbol{\xi}=\mathbf{A}\mathbf{W}\mathbf{x}+\boldsymbol{\eta}$, видим, что $\bar{M}\|\boldsymbol{\eta}\|^2=n\bar{\sigma}^2$, и задача в новых наблюдениях $\mathbf{z}$ сводится к рассмотренной в предыдущем пункте с заменой $\mathbf{W}$ на $\mathbf{A}\mathbf{W}$.

В частном случае, когда заданы $\bar{M}\sum a_i^2\xi_i^2=n\bar{\sigma}^2$, матрица $\mathbf{A}$ становится диагональной с элементами a_i по главной диагонали. Такой результат легко переносится на процессы:

$$y_t=w_t x+\xi_t,\ t\in[0,T],\qquad \bar{M}\,\frac{1}{T}\int_0^T a_t^2\,\xi_t^2\,dt=\bar{\sigma}^2,$$

где w_t и a_t — известные функции. Тогда для получения оптимальной оценки нужно в формулах (6.16), (6.17), $\hat{y}$, $\hat{\sigma}_{\mathbf{y}}$ и c^* заменить на

$$\hat{\hat{y}}_T=\frac{\int_0^T y_t w_t a_t^2\,dt}{\|\mathbf{w}\mathbf{a}\|^2},\quad \hat{\hat{\sigma}}_{\mathbf{y}}^2=\frac{\int_0^T y_t^2 a_t^2\,dt}{\|\mathbf{w}\mathbf{a}\|^2}-\hat{\hat{y}}_T^2,\quad c^*=\frac{\alpha\,\|\mathbf{w}\mathbf{a}\|^2}{\varkappa\bar{\sigma}^2 T},$$

где $\|\mathbf{w}\mathbf{a}\|^2=\int_0^T w_t^2 a_t^2 dt$. Так, оценка типа (6.17) принимает вид

$$\partial_{\mathbf{y}}^*(x)=[1-\alpha\,\|\mathbf{w}\mathbf{a}\|^2\,(x-\hat{\hat{y}}_T)^2/(\varkappa\bar{\sigma}^2 T)]^+.$$

9. Пусть теперь исходным является не задание «мощности» флуктуаций, а среднее значение некоторого неотрицательного функционала от реализаций флуктуаций: $\bar{M}F\{\boldsymbol{\xi}\}=\mu$. Тогда оценка будет иметь вид $\partial_{\mathbf{y}}(x)=[c_0-cF\{\mathbf{y}-\mathbf{W}\mathbf{x}\}]^+$, причем $\underline{M}\partial=[c_0-c\mu]$. Рассмотрим два примера таких оценок.

Пример 6.6. Пусть x — параметр сдвига и заданным является среднее эффективное значение реализации флуктуаций $\bar{M}|\xi|=\mu$, $|\boldsymbol{\xi}|=\Sigma|\xi_i|$. Тогда оптимальная оценка уровня α записывается: $\partial(\boldsymbol{\xi})=[c_0-c^*|\boldsymbol{\xi}|]^+$, $c_0=1$, $c^*=\alpha/(\varkappa\mu)$. Сюда нужно подставить $\boldsymbol{\xi}=\mathbf{y}-1x$. Оценка как функция переменной x обретает вид линейно-ломаной с узлами в точках y_i. Форма записи оценки не изменится при заданной любой другой норме $|\xi|$.

Пример 6.7. Пусть единственным первичным значением является

$$\bar{M}\Sigma\{|\xi_i|>h\}/n=p_h,$$

т. е. исходным является задание верхнего среднего относительной частоты превышений абсолютными значениями ξ_i порога h. Тогда оптимальная оценка имеет вид

$$\partial_{\mathbf{y}}(x)=[c_0-c\,\Sigma\{|\xi_i|>h\}/n]^+,\quad c_0=c=\min\{1,\ (1-\alpha)/(1-\varkappa p_h)\}.$$

Оптимальная оценка параметра сдвига при однородных некоррелированных флуктуациях. Дополним сведения относительно флуктуаций, обогащая первичный набор новыми признаками, и посмотрим, как видоизменится оценка, какие новые свойства у нее появятся. Пусть $y_i=x+\xi_i$, $i=1, \ldots, n$, причем известно только, что ξ_i некоррелированы $M\xi_i\xi_j=0$, $i\neq j$, и имеют одинаковые «мощности» $\bar{M}\xi^2{}_i=\bar{\sigma}^2$. Это соответствует СИМ с первичными средними, получаемыми подстановкой $\xi_i=y_i-x$ и приобретающими вид:

$$M\,(y_i-x)\,(y_j-x)=0,\ i\neq j;\quad \bar{M}\,(y_i-x)^2=\bar{\sigma}^2,$$

где i и j пробегают значения от 1 до n.

Поскольку СИМ симметрична к перестановкам y_i между собой, а группа перестановок дискретна, то достаточным будет класс инвариантных к перестановкам правил (см. утверждение 6.5). Вместе с теоремой 6.1 это позволяет получить следующий достаточный класс оценок:

$$\partial_{\mathbf{y}}(x)=[c_0-c_1\sum_{i\neq j}\sum(y_i-x)\,(y_j-x)/n^2-c_2\,\Sigma(y_i-x)^2/n]^+=$$

$$=[c_0-c_1(\hat{y}-x)^2+(c_1/n-c_2)\,(\hat{y}^2-2x\,\hat{y}+x^2)]^+,\quad \hat{y}^2=\Sigma\,y_i^2/n,$$

и для них $\underline{M}\partial_{\mathbf{y}}(x)=[c_0-c_2\bar{\sigma}^2]^+$. Осталось найти коэффициенты c_0, c_1 и c_2, минимизирующие риск и удовлетворяющие ограничению $\partial_{\mathbf{y}}(x)\leqslant 1$.

Будем искать $\Omega(\partial)$, для чего перепишем $\partial_y(x)$ в виде:

$$\partial_y(x) = [c_0 - c_1\hat{y}^2 + (c_1/n - c_2)\hat{y}^2 +$$

$$+ 2x(c_1(n-1)/n + c_2)\hat{y} - x^2(c_1(n-1)/n + c_2)]^+ = [\hat{a}_0 + \hat{a}_1 x - a_2 x^2]^+,$$

где $\hat{a}_0 = c_0 - a_2\hat{y}^2 + (c_1/n - c_2)\hat{\sigma}^2_y$, $\hat{a}_1 = 2a_2\hat{y}$, $a_2 = c_1(n-1)/n + c_2$.

Условие $\partial_y(x) \leqslant 1$ соответствует двум неравенствам $a_2 \geqslant 0$ и $\sup_x \partial_y(x) = \hat{a}_0 + \hat{a}^2_1/4a_2 \geqslant 1$. Так как $\partial_y(x) \geqslant 0$ при $\hat{a}_0 + \hat{a}^2_1/2a_2 \geqslant 0$ и $|x - \hat{a}_1/2a_2| \leqslant \sqrt{[\hat{a}^2_1 + 4\hat{a}_0 a_2]^+}/2a_2$, то, проинтегрировав $\partial_y(x)$ по x в пределах, соответствующих последнему неравенству, найдем

$$\Omega(\partial) = \int \partial_y(x)\,dx = (2/3a_2)\sqrt{[\hat{a}^2_1 + 4\hat{a}_0 a_2]^+}\,[\hat{a}_0 + \hat{a}^2_1/4a_2]^+.$$

Учитывая теперь, что $\overline{M}\partial(x) = \overline{M}^y[\hat{a}_0 + \hat{a}^2_1/4a_2]^+ = \sup_y [\hat{a}_0 + \hat{a}^2_1/4a_2]^+$, сформируем выражение для составного риска:

$$\Pi^{\varkappa}_{\lambda}(\partial) = 1 - \varkappa[c_0 - c_2\overline{\sigma^2}]^+ + (\varkappa - 1)S + 4\lambda S^{3/2}/3\sqrt{a_2}.$$

Найдем в этом выражении $S = \sup_y [\hat{a}_0 - \hat{a}^2_1/4a_2]^+ = c_0 + \inf_{\hat{y}} a_2\hat{y}^2 + \sup_{\hat{\sigma}^2_y}(c_1/n - c_2)\hat{\sigma}^2_y = c_0$. Отметим, что для справедливости последнего выражения должно выполняться $c_1/n \leqslant c_2$ (иначе супремум от $\hat{\sigma}^2_y$ равен ∞). В результате

$$\Pi^{\varkappa}_{\lambda}(\partial) = 1 - \varkappa[c_0 - c_2\overline{\sigma^2}]^+ + (\varkappa - 1)c_0 + 4\lambda c_0^{3/2}/3\sqrt{a_2}.$$

Оптимальным c_1, минимизирующим риск при заданных c_0 и c_2, будет такое значение, которое максимизирует $a_2 = c_1(n-1)/n + c_2$ при условии $c_1/n \leqslant c_2$. Так как чем больше c_1, тем больше a_2 и тем меньше риск, то оптимальным будет наибольшее возможное значение: $c_1 = c_2 n$, откуда сразу же $a_2 = c_1$, и после подстановки этих значений оценка приобретает вид $\partial_y(x) = [c_0 - c_1(\hat{y} - x)^2]^+$, а ее риск становится равным

$$\Pi^{\varkappa}_{\lambda}(\partial) = 1 - \varkappa[c_0 - c_1\overline{\sigma^2}/n]^+ + (\varkappa - 1)c_0 + \lambda c_0(4/3)\sqrt{c_0/c_1}.$$

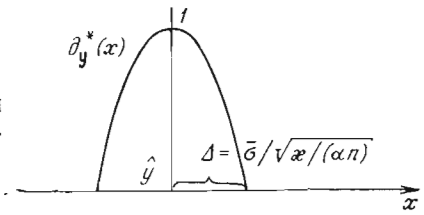
Полученное выражение отличается от (6.15) лишь тем, что вместо $\overline{\sigma^2}$ здесь стоит $\overline{\sigma^2}/n$, поэтому оптимальная оценка будет иметь вид (6.17) с заменой $\overline{\sigma^2}$ на $\overline{\sigma^2}/n$, что мы и сформулируем как результат.

При однородной некоррелированной выборке и $\varkappa \geqslant 1$ оптимальная расплывчатая оценка уровня α имеет вид

$$\partial^*_y(x) = [1 - c^*(x - \hat{y})^2]^+,\ c^* = \alpha n/\varkappa\overline{\sigma^2}. \qquad (6.18)$$

Иллюстрация ее дана на рис. 6.3. При увеличении n оценка становится все более узкой, менее расплывчатой, что очевидно должно быть, так как при получении $\hat{y}$ как суммы некоррелированные флуктуации складываются в беспорядке, стохастически, сглаживая в известном смысле друг друга, в результате разброс уменьшается, что позволяет для фиксации α сообразно этому

Рис. 6.3. Расплывчатая оценка при некоррелированных флуктуациях



сократить ширину. Причем такой же оценка будет, если дополнительно известно, что $M\xi_i = 0$, эти новые данные оказываются излишними.

Оценка (6.18) инвариантна к перестановкам наблюдений. Она же эквивариантна к преобразованию сдвига в том смысле, что одновременное прибавление ко всем y_i и к x любого числа a не меняет ее значения: $\partial_y(x) \equiv \partial_{y+1a}(x+a)$. Последнее и позволяет исследовать (6.18) при $x = 0$, исследуя тем самым $\partial(\xi)$.

Оценивание сдвига при неоднородных некоррелированных флуктуациях. Пусть выборка не является однородной, но остается некоррелированной, т. е. $y_i = x + \xi_i$, $M\xi_i\xi_j = 0$, $i \neq j$, $\overline{M}\xi^2_i = \overline{\sigma^2}_i$, $i = 1, \ldots, n$. Определенная таким образом СИМ уже не будет симметричной к перестановкам. Достаточным будет следующий класс оценок:

$$\partial_y(x) = [c_0 - \sum\sum c_{ij}(y_i - x)(y_j - x)]^+,$$

где $c_{ij} = c_{ji}$.

Перепишем: $\partial_y(x) = [\hat{a}_0 + \hat{a}_1 x - a_2 x^2]^+$, где $\hat{a}_0 = c_0 - \sum\sum c_{ij}y_i y_j$, $\hat{a}_1 = \sum\sum c_{ij}(y_i + y_j) = 2\sum c_i y_i$, $c_i = \sum_j c_{ij}$, $a_2 = \sum\sum c_{ij}$. При введенных обозначениях штраф

$$\overline{M}^y \int \partial_y(x)\,dx = \frac{4\lambda}{3a_2^2}[\sup_y(\hat{a}_0 a_2 + \hat{a}^2_1/4)^+]^{3/2}.$$

Ищем внутренний супремум в квадратных скобках:

$$\sup_y(\hat{a}_0 a_2 + \hat{a}^2_1/4) = c_0 a_2 + \inf_y[a_2\sum\sum c_{ij}y_i y_j - \sum\sum c_i c_j y_i y_j].$$

Чтобы инфимум не равнялся $-\infty$, матрица $\mathbf{Q}$ элементов $Q_{ij} = (\sum\sum c_{kl})c_{ij} - c_i c_j$, $i, j = 1, \ldots, n$, должна быть неотрицательно определенной, и тогда $\sup_y(\hat{a}_0 a_2 + a^2_1/4) = c_0 a_2$. Далее $\overline{M}\partial_y(x) = [c_0 - \sum_i c^+_{ii}\sigma^2_i]^+$, где c^+_{ii} есть положительные части c_{ii}. Наконец, $\overline{M}\partial_y(x) = \sup_{x,y}\partial_y(x) = (1/a_2)\sup_{x,y}(\hat{a}_0 a_2 + \hat{a}^2_1/4)^+ = c_0$, а $\partial_y(x) \leqslant 1$ эквивалентно условию $c_0 \leqslant 1$. Используя найденные значения и полагая $c_0 - \sum c^+_{ii}\overline{\sigma^2}_i > 0$, запишем риск

$$\Pi^{\varkappa}_{\lambda}(\partial) = 1 - c_0 + \varkappa\sum c^+_{ii}\overline{\sigma}^2_i + \lambda 4 \cdot c_0^{3/2}/3\sqrt{\sum\sum c_{ij}}.$$

Нужно минимизировать риск c_{ij} при условии неотрицательной определенности матрицы $\mathbf{Q}$. Разложим $\mathbf{Q}$ на сумму матриц ранга 1: $\mathbf{Q} = \sum \mathbf{b}_{(k)}\mathbf{b}^{\mathrm{T}}_{(k)}$, где $\mathbf{b}_{(k)} = (b_{k1}, \ldots, b_{kn})^{\mathrm{T}}$ — векторы. Тогда нужно будет минимизировать $\Pi_\lambda(\partial)$ независимым выбором векторов $\mathbf{b}_{(k)}$. Совершенно ясно, что все эти векторы све-

дутся к одному, так что $c_{ij}=b_ib_j$. Отсюда $\sum\sum c_{ij}=(\sum b_i)^2$ и $c_{ii}=b^2_i$. Оценка и риск запишутся так:

$$\partial_{\mathbf{y}}(x)=[c_0-(\sum b_i^2(y_i-x))^2]^+,$$

$$\Pi^{\varkappa}_{\lambda}(\partial)=1-c_0+\varkappa\sum b_i^2\bar{\sigma}_i^2+\lambda\,4\cdot c_0^{3/2}\,3\sum b_i.$$

Осталось минимизировать риск по b_i и c_0. Дифференцируя по b_i и приравнивая 0, получаем $b_i=2\lambda c_0^{3/2}/[3\varkappa\bar{\sigma}^2_i(\sum b_i)^2]$. Суммируя по i, выводим уравнение, из которого находим сумму $\sum b_i=[2\lambda(\sum(1/\bar{\sigma}^2_i))/(3\varkappa)]^{1/3}c_0^{1/2}$. Подставляя эту сумму в выражение для b_i и далее в $\Pi^{\varkappa}_{\lambda}(\partial)$, имеем

$$b_i=c_0^{1/2}\left(\frac{2\lambda}{3\varkappa}\right)^{1/3}\Big/\left[\bar{\sigma}_i^2\left(\Sigma\left(1/\bar{\sigma}_i^2\right)\right)^{2/3}\right]\text{ и}$$

$$\Pi_{\lambda}(\partial)=1-c_0\left[1-(12\lambda^2\varkappa)^{1/3}/\left(\Sigma\left(1/\bar{\sigma}_i^2\right)\right)^{1/3}\right].$$

Очевидно, c_0 будет равно 1 или 0 в зависимости от того, больше или меньше 0 выражение в квадратных скобках. Нас интересует лишь первый случай, поэтому $c_0=1$ и тогда $b_i=\lambda/\bar{\sigma}^2_i$. Определяя λ по заданной ошибке $\alpha=\varkappa\sum b^2_i\bar{\sigma}^2_i=\varkappa\lambda\sum(1/\bar{\sigma}^2_i)$, находим $\lambda^2=\alpha/[\varkappa\sum(1/\bar{\sigma}^2_i)]$. Сформулируем окончательный результат.

Оптимальная (при $\varkappa\geqslant 1$) уровня α оценка параметра сдвига неоднородной некоррелированной выборки имеет вид

$$\partial^*_{\mathbf{y}}(x)=\left[1-\frac{\alpha}{\varkappa}\left(\sum\frac{1}{\bar{\sigma}_i^2}\right)\left(x-\frac{\sum y_i/\bar{\sigma}_i^2}{\sum 1/\bar{\sigma}_i^2}\right)^2\right]^+.$$

В обозначениях $\bar{\sigma}^2_{\text{ср}}=n/(\sum 1/\bar{\sigma}^2_i)$ и $\hat{y}=\bar{\sigma}^2_{\text{ср}}\sum(y_i/(n\bar{\sigma}^2_i))$ оптимальная оценка записывается

$$\partial^*_{\mathbf{y}}(x)=\left[1-\frac{\alpha n}{\varkappa\bar{\sigma}^2_{\text{ср}}}(x-\hat{y})^2\right]^+. \tag{6.19}$$

Отметим, что при одинаковых дисперсиях $\bar{\sigma}_i=\bar{\sigma}$ имеем $\bar{\sigma}_{\text{ср}}=\bar{\sigma}$ и $\hat{\hat{y}}=\hat{y}$, поэтому из (6.19) частным случаем будет (6.17).

Обобщения оценок 1. Оценка параметра регрессии. Пусть $y_i=w_ix+\xi_i$, $M\xi_i\xi_j=0$, $i\neq j$, $M\xi^2_i=\bar{\sigma}^2_i$, $i=1,\ldots,n$. Переходя к новым наблюдениям $z_i=y_i/w_i=x+(\xi_i/w_i)=x+\eta_i$, $M\eta^2_i=\bar{\sigma}^2_i/w^2_i$, мы сводим задачу к (6.19), в которой y_i заменяются на z_i, а $\bar{\sigma}^2_i$ — на $\bar{\sigma}^2_i/w^2_i$. Получается следующая оптимальная оценка:

$$\partial^*_{\mathbf{y}}(x)=\left[1-\frac{\alpha}{\varkappa}\left(\sum\frac{w_i^2}{\bar{\sigma}_i^2}\right)\left(x-\frac{\sum y_i w_i/\bar{\sigma}_i^2}{\sum w_i/\bar{\sigma}_i^2}\right)^2\right]^+. \tag{6.20}$$

Отметим, во-первых, что отсчеты y_i индекса $i: w_i=0$ исключаются. Во-вторых, оценка эквивариантна к преобразованиям сдвига вида $\mathbf{s}(x,y)=(x+a,\ \mathbf{y}+\mathbf{w}a)$. В-третьих, оценка (6.20) базируется на $\hat{y}=(\sum y_iw_i/\bar{\sigma}^2_i)/(\sum w_i/\bar{\sigma}^2_i)$ — такой детерминированной оценке x, которая максимизирует в классе линейных оценок вида $\sum c_iy_i$ отношение $(\sum c_iw_i)^2/(\sum c^2_i\bar{\sigma}^2_i)$ (интерпретируемое в задачах радиотехники как выходное после линейной обработки отношение энергии полезного сигнала к мощности шума).

2. Рассмотрим более общую задачу. Пусть известны корреляции флуктуаций: $y_i=w_ix+\xi_i$, $M\xi_i\xi_j=B_{ij}$, $i,\ j=1,\ldots,n$. Здесь элементы B_{ij} образуют неотрицательно определенную симметричную матрицу $\mathbf{B}$. Разлагая ее по собственным функциям, получим $\mathbf{B}=\mathbf{F}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{F}$, где $\boldsymbol{\Sigma}$ — диагональная матрица с элементами σ^2_i по главной диагонали, а $\mathbf{F}$ — унитарная матрица F_{ij}: $\mathbf{F}^{\mathrm{T}}=\mathbf{F}^{-1}$.

Переходя к векторам-столбцам $\mathbf{w}$ и $\boldsymbol{\xi}$ элементов w_i и ξ_i, запишем $\mathbf{y}=\mathbf{w}x+\boldsymbol{\xi}$, $\mathbf{B}=M\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}$. Совершим преобразование $\mathbf{z}=\mathbf{F}\mathbf{y}=x\mathbf{F}\mathbf{w}+\mathbf{F}\boldsymbol{\xi}=x\check{\mathbf{w}}+\boldsymbol{\eta}$, где $\check{w}_i=\sum_j F_{ij}w_j$, $\eta_i=\sum_j F_{ij}\xi_j$. Нетрудно видеть, что $M\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}=M\mathbf{F}\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}=\mathbf{F}\mathbf{B}\mathbf{F}^{\mathrm{T}}=\mathbf{F}\mathbf{F}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{F}\mathbf{F}^{\mathrm{T}}=\boldsymbol{\Sigma}$, так что η_i некоррелированы между собой и $M\eta^2_i=\sigma^2_i$. Таким образом, в наблюдениях z_i задача сводится к уже рассмотренной, только в формуле (6.20) нужно заменить y_i на z_i, w_i на $\check{w}_i$, а $\bar{\sigma}^2_i$ на σ^2_i. Так как

$$\sum\check{w}_i^2/\sigma_i^2=\check{\mathbf{w}}^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\check{\mathbf{w}}=\mathbf{w}^{\mathrm{T}}\,\mathbf{F}^{\mathrm{T}}\,\mathbf{F}\,\mathbf{B}^{-1}\,\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathbf{w}=\mathbf{w}^{\mathrm{T}}\,\mathbf{B}^{-1}\mathbf{w},$$

$$\sum z_i\check{w}_i/\sigma_i^2=\mathbf{z}^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\check{\mathbf{w}}=\mathbf{y}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathbf{B}^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathbf{w}=\mathbf{y}^{\mathrm{T}}\mathbf{B}^{-1}\underline{w},$$

то после соответствующей подстановки в (6.20) получаем оптимальную оценку

$$\partial^*_{\mathbf{y}}(x)=\left[1-\frac{\alpha}{\varkappa}\,\mathbf{w}^{\mathrm{T}}\,\mathbf{B}^{-1}\mathbf{w}\,(x-\hat{\hat{x}}_{\mathbf{y}})^2\right]^+, \tag{6.21}$$

где $\hat{\hat{x}}_{\mathbf{y}}=\mathbf{y}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{w}/\mathbf{w}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{w}$ (полезно сравнить ее с (6.10)). Если учесть, что верно $M\hat{\hat{x}}^2_{\xi}=1/\mathbf{w}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{w}$, то оценка (6.21) записывается иначе:

$$\partial^*_{\mathbf{y}}(x)=[1-(\alpha/\varkappa)\,(x-\hat{\hat{x}}_{\mathbf{y}})^2/M\,\hat{\hat{x}}^2_{\xi}]^+.$$

Отметим, что $\hat{\hat{x}}_{\mathbf{y}}$ при $M\xi_i=0$ есть детерминированная оценка x, минимизирующая в классе линейных оценок $\mathbf{c}^{\mathrm{T}}\mathbf{y}$ отношение «сигнал-шум»: $(M\mathbf{c}^{\mathrm{T}}\mathbf{y})^2/M(\mathbf{c}^{\mathrm{T}}\mathbf{y})^2=(\mathbf{c}^{\mathrm{T}}\mathbf{w})^2/(\mathbf{c}^{\mathrm{T}}\mathbf{B}\mathbf{c})$.

3. Оценивание амплитуды детерминированного сигнала. Пусть требуется оценить x по данным

$$y_t=w_t\,x+\xi_t,\qquad M\,\xi_t\,\xi_{t'}=B\,(t,t'),\qquad t,t'\in[0,T],$$

где $B(t,t')$ — корреляционная функция шума, это есть ядро положительно определенного обратимого оператора. Оценка имеет точно приведенный в (6.21) вид с той лишь разницей, что квадратичные формы $\mathbf{w}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{w}$ и $\mathbf{y}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{w}$ заменяются на интегралы

$$\mathbf{w}^{\mathrm{T}}\,\mathbf{B}^{-1}\mathbf{w}=\int_0^T w_t\,h_t\,dt,\qquad \mathbf{y}^{\mathrm{T}}\,\mathbf{B}^{-1}\mathbf{w}=\int_0^T y_t\,h_t\,dt,$$

где h_t получается решением интегрального уравнения

$$\int_0^T B(t, t')\, h_t\, dt = w_{t'}.$$

Отметим, что точное знание корреляций есть некоторая идеализация реальных знаний, если ошибки невелики. В следующем разделе будет изучено влияние ошибок в знании корреляций на способы формирования оценок и оптимальный вид их.

Оценка амплитуды сигнала при колебаниях его формы и неточных корреляциях шума. Пусть $y_t = w_t x + \xi_t$, где $t = t_1, t_2, \ldots, t_n$ (или $t \in [0, T]$), w_t определяет форму сигнала, а x — его амплитуду. Пусть шум ξ_t имеет нулевое среднее $M\xi_t = 0$, а его корреляционная функция $B(t, \tau)$ неизвестна и лежит внутри заданного выпуклого ограниченного семейства $\mathfrak{B}$:

$$M\,\xi_t\,\xi_\tau = B(t, \tau) \in \mathfrak{B}, \text{ т.е. } \mathscr{M}^\xi = \bigvee_{\mathbf{B}\in\mathfrak{B}} \mathscr{M}^\xi_{\mathbf{B}}.$$

Будем искать оптимальную оценку x в достаточном для указанной задачи классе правил вида $\partial(y_t - w_t x)$. Запишем для этого риск

$$\Pi^{\varkappa}_{\lambda}(\partial) = 1 - \varkappa \underline{M}^\xi \partial(\xi) + (\varkappa - 1)\overline{M}^\xi \partial(\xi) + \lambda \overline{M}^\xi \sup_{x'} \int \partial(\xi + w(x' - x))\, dx =$$

$$= 1 - \varkappa \inf_{\mathbf{B}} \underline{M}^\xi_{\mathbf{B}} \partial(\xi) + (\varkappa - 1) \sup_{\mathbf{B}} \overline{M}^\xi_{\mathbf{B}} \partial(\xi) + \lambda \sup_{\mathbf{B}} \overline{M}^\xi_{\mathbf{B}} \int \partial(\xi - wx)\, dx =$$

$$= 1 - \varkappa \inf_{\mathbf{B}} \underline{M}^\xi_{\mathbf{B}} \partial(\xi) + (\varkappa - 1)\overline{\partial} + \lambda \sup_{\xi} \int \partial(\xi - wx)\, dx,$$

где использовано равенство $\overline{M}^\xi_{\mathbf{B}} f(\xi) = \sup f(\xi) = \overline{f}$ и индекс t у ξ_t и w_t для краткости опущен. Задача синтеза оптимального правила обретает вид

$$\inf_{\partial} \sup_{\mathbf{B}} \left[1 - \varkappa \underline{M}^\xi_{\mathbf{B}} \partial(\xi) + (\varkappa - 1)\overline{\partial} + \lambda \sup_{\xi} \int \partial(\xi - w\,x)\, dx\right].$$

Так как

$$\underline{M}^\xi_{\gamma\mathbf{B}_1 + (1-\gamma)\mathbf{B}_2} \partial(\xi) = \sup\Big\{ \sum_t \sum_\tau [\gamma B_1(t, \tau) +$$

$$+ (1 - \gamma) B_2(t, \tau)]\, c_{t,\tau} : \sum_t c_t \xi_t + \sum_t \sum_\tau c_{t,\tau} \xi_t \xi_\tau \leqslant \partial(\xi)\Big\} =$$

$$= \gamma \underline{M}^\xi_{\mathbf{B}_1} \partial(\xi) + (1 - \gamma) \underline{M}^\xi_{\mathbf{B}_2} \partial(\xi), \quad 0 \leqslant \gamma \leqslant 1,$$

то риск является вогнутой функцией $\mathbf{B}$. Учитывая выпуклость и ограниченность семейства $\mathfrak{B}$, мы можем инфимум и супремум поменять местами и сначала отыскать оценку при заданной корреляционной функции $B(t, \tau)$, обозначая риск $\Pi^{\varkappa}_{\lambda, \mathbf{B}}(\partial)$, а затем, максимизируя риск, определить наименее благоприятное $\mathbf{B}_{\text{н.б}}$ и подставить в оценку.

Оптимальная оценка, минимизирующая $\Pi^{\varkappa}_{\lambda, \mathbf{B}}(\partial)$, и риск для нее записываются в матричном виде (при непрерывном t легко переписывающимся в соответствующие интегральные аналоги):

$$\partial_{\mathbf{y}}(x) = \left[1 - \left(\frac{2\lambda}{3\varkappa}\, \mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{w}\right)^{2/3} \left(x - \frac{\mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{y}}{\mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{w}}\right)^2\right]^+,$$

$$\Pi^{\varkappa}_{\lambda, \mathbf{B}}(\partial) = (12\,\lambda^2\,\varkappa)^{1/3} / (\mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{w})^{1/3},$$

где при некотором λ это — формула (6.21), и было сделано допущение, что $\Pi^{\varkappa}_{\lambda, \mathbf{B}}(\partial) \leqslant 1$. Теперь из выражения для риска видно, что *наименее благоприятной* $\mathbf{B}_{\text{н.б}}$ *будет корреляционная функция, минимизирующая квадратичную форму*:

$$\mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1}_{\text{н.б}} \mathbf{w} = \min_{\mathbf{B}\in\mathfrak{B}} \mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{w}.$$

Окончательно *оптимальная оценка уровня* α *примет вид* (6.21), *куда вместо* $\mathbf{B}$ *подставляется* $\mathbf{B}_{\text{н.б}}$. Рассмотрим примеры получения таких оценок для различных семейств $\mathfrak{B}$.

Пример 6.8. Пусть заданы границы корреляционной функции $\underline{B}(t, \tau)$, $\overline{B}(t, \tau)$, являющиеся неотрицательно определенными ядрами. Тогда наименее благоприятным будет значение $B_{\text{н.б}}(t, \tau) = \overline{B}(t, \tau)$, соответствующее максимальной мощности $\overline{B}(t, t) = \overline{b}^2_t$ шума и максимальным корреляциям.

В другом случае пусть $\mathfrak{B} = \{\mathbf{B} : \|\mathbf{B} - \mathbf{B}_0\| \leqslant \nu\}$, где норма $\|\mathbf{B}\|$ есть максимальное собственное число оператора с ядром $B(t, \tau)$. Здесь $B_0(t, \tau)$ есть некоторое предполагаемое (оценочное) значение корреляционной функции, а ν — величина ошибки. В этом случае

$$\min_{\|\mathbf{B} - \mathbf{B}_0\| \leqslant \nu} \mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{w} = \mathbf{w}^{\mathrm{T}} (\mathbf{B}_0 + \nu \mathbf{I})^{-1} \mathbf{w},$$

где $\mathbf{I}$ — единичный оператор, соответствующий ядру в виде дельта-функции Дирака. Ядро $\nu\mathbf{I}$ интерпретируется как корреляционная функция «белого шума» спектральной интенсивности ν. Поэтому сумма $\mathbf{B}_0 + \nu\mathbf{I}$ равносильна добавлению к флуктуациям «белого шума». Таким образом, *ошибка в знании корреляционной функции флуктуаций компенсируется прибавлением добавки в виде «белого шума»* (подобный вывод, если вспомнить, мы уже получали в задаче фильтрации конца § 5.5).

Теперь обсудим тот случай, когда неточно известна функция w_t, определяющая форму полезного сигнала. Оценка (6.21) уже не будет в этом случае оптимальной, так как она призывает к линейной обработке наблюдений y_t, тогда как при неточно известном сигнале не исключено, что следует перейти к нелинейной обработке. По крайней мере, оценки вида $[c_0 - \sum\sum c_{ij}(y_i - x) \times (y_j - x)]^+$ уже не образуют достаточного класса. Тем не менее, если изменения сигнала малы по сравнению с некоторым гипотетическим значением w^0_t, то все же можно ограничиться линейными оценками, подбирая в $\hat{x}_{\mathbf{y}} = \mathbf{c}^{\mathrm{T}}\mathbf{y}$ значение $\mathbf{c}$ таким образом, чтобы максимизировать отношение сигнал-шум на выходе фильтра линейной обработки при наименее благоприятном отклике w_t, т. е. решая задачу

$$\max_{\mathbf{c}} \min_{\mathbf{w}\in\mathscr{W}} (\mathbf{c}^{\mathrm{T}} \mathbf{w})^2 / \mathbf{c}^{\mathrm{T}} \mathbf{B} \mathbf{c} \text{ при } \mathbf{c}^{\mathrm{T}} \mathbf{w} = 1.$$

Пример 6.9. Пусть w_t может колебаться около значения w°_t, так что $\|\mathbf{w} - \mathbf{w}^\circ\|^2 = \sum_t (w_t - w^\circ_t)^2 \leqslant \Delta^2$. Тогда

$$\min_{\mathbf{w}\in\mathscr{W}} (\mathbf{c}^{\mathrm{T}} \mathbf{w})^2 = \min_{\|\mathbf{w} - \mathbf{w}^\circ\| \leqslant \Delta} (\sum c_t w_t)^2 = \left(\sum c_t w^\circ_t - \|\mathbf{c}\| \Delta\right)^2 = (\mathbf{c}^{\mathrm{T}} \mathbf{w}^\circ - \|\mathbf{c}\| \Delta)^2,$$

где минимум достигается при $w_t = w^\circ_t - c_t\Delta/\|c\|$. Максимизация по c отношения сигнал-шум $(c^T w^\circ - \|c\|\Delta)^2/c^T B c$ ведет к значению

$$c^* = (B - \nu I)^{-1} w / w^T (B - \nu I)^{-1} w,$$

где ν находится как решение уравнения $\nu^2 w^T (B+\nu I)^{-2} w = \Delta^2$.

При непрерывном времени t суммы заменяются на интегралы, а матричные выражения — на интегральные аналоги.

Таким образом, *наличие колебаний сигнала в случае линейной обработки компенсируется введением дополнительного аддитивного «белого шума» спектральной интенсивности* ν[1]. Полезно провести сравнение с предыдущим примером, где такая же добавка являлась следствием неточного знания корреляционных свойств шума.

6.4. ОЦЕНИВАНИЕ ПАРАМЕТРА СДВИГА ПО МОМЕНТАМ И ГАРМОНИЧЕСКИМ СРЕДНИМ

Оценивание по моментам. Какова будет оценка параметра x по наблюдениям $y_i = x + \xi_i$, $i = 1, \ldots, n$, когда известны моменты более высоких, чем первого и второго порядков? Тогда вид оценки, в общем, будет отличаться от усеченной снизу параболы, каковой она являлась выше. Начнем с примера.

Пример 6.10. Допустим, что исходными для ξ являются: $M\xi^3_i \xi^3_j = 0$, $i \neq j$, $\overline{M}\xi^6_i = \bar{m}_6$. Тогда при $\varkappa \geqslant 1$ достаточным будет следующий класс оценок: $\partial(\xi) = [c_0 - c_1 \sum\sum_{i \neq j} \xi^3_i \xi^3_j - c_2 \sum \xi^6_i]^+$ (куда для получения $\partial_y(x)$ нужно подставить $\xi_i = y_i - x$). В частности, к этому классу относится оценка вида

$$\partial(\xi) = [1 - (\hat{\xi}^3/\Delta_n)^2]^+, \quad \hat{\xi}^3 = \sum \xi^3_i / n,$$

которая заменой $\xi^3_i = \eta_i$ выводится из (6.18), причем $M\eta_i = 0$, $\overline{M}\eta^2_i = \bar{m}_6$, поэтому для фиксации уровня должно быть $\Delta_n = \sqrt{\varkappa \bar{m}_6/(\alpha n)}$. Если в той же самой задаче дополнительно известны промежуточные моменты (вплоть до шестого), то, как можно будет видеть из сравнения со следующим примером, более хорошими могут оказаться оценки вида $[1 - (\hat{\xi}/\Delta)^6]^+$, так что форма оценок во многом определяется видом исходных (первичных) признаков.

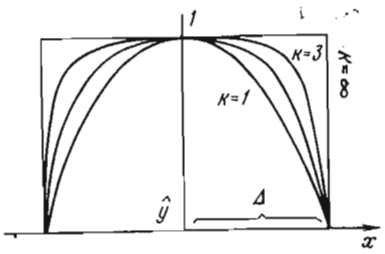
Перейдем от примера к общему случаю. Пусть ограниченными являются моменты вплоть до порядка $2k$. Зададимся оценками вида

$$\partial(\xi) = [1 - (\hat{\xi}/\Delta_n)^{2k}]^+, \quad k = 1, 2, \ldots, \tag{6.22}$$

и назовем их *оценками степенного типа*. Как функция переменной x оценка степенного вида $\partial(y - 1x)$ при $k > 1$ уже не перевернутая

[1] Впервые показано в статье: Кузнецов В. П. О синтезе линейных обнаружителей при неточно заданном сигнале и неполностью известных свойствах нормального шума//Радиотехника и электроника. — 1974. — № 12. — С. 2529—2538.

Рис. 6.4. Оценки степенного типа



усеченная снизу парабола, а кривая, более тупая, чем парабола у вершины, и более круто спадающая у основания, как это показано на рис. 6.4, трансформирующаяся при $k \to \infty$ к индикаторной форме, т. е. доверительному интервалу.

Величина $2\Delta_n$ — *размах* оценки у основания. Очевидно $\underline{M}\partial(\xi) \geqslant 1 - \overline{M}\hat{\xi}^{2k}/\Delta^{2k}_n$ (так как $\partial(z) \geqslant 1 - (\hat{z}/\Delta_n)^{2k}$), $\overline{M}\partial(\xi) \leqslant 1$ (так как $\partial(z) \leqslant 1$), откуда при $\varkappa \geqslant 1$ имеем

$$\alpha^\varkappa(\partial) = 1 - \varkappa \underline{M}\partial + (\varkappa - 1)\overline{M}\partial \leqslant \varkappa \overline{M}\hat{\xi}^{2k}/\Delta^{2k}_n.$$

Теперь, чтобы уровень $\alpha^\varkappa(\partial)$ оценки был не больше α, нужно брать $\Delta_n \leqslant (\varkappa \overline{M}\hat{\xi}^{2k}/\alpha)^{1/2k}$. Для определения размаха оценки Δ_n нужно знать $\overline{M}\hat{\xi}^{2k}$. В свою очередь, для нахождения $\overline{M}\hat{\xi}^{2k}$, как это следует из неравенства $\overline{M}\hat{\xi}^{2k} \leqslant \sum \ldots \sum |\overline{M}| \xi_{i_1} \ldots \xi_{i_k}| / n^{2k}$, достаточно знать смешанные моменты вплоть до порядка $2k$.

Чтобы получить содержательные результаты, нужно сделать предположения относительно существования моментов. Характер этих предположений и их влияния на размах оценки установим сначала на примере.

Пример 6.11. Пусть $k = 2$. Тогда для независимой симметричной последовательности с конечным четвертым моментом:

$$|\overline{M}|\, \xi_{i_1} \xi_{i_2} \xi_{i_3} \xi_{i_4} = \begin{cases} 0 & \text{при } i_1 \neq i_2, \quad i_1 \neq i_3, \quad i_1 \neq i_4, \\ \overline{\sigma}^4 & \text{при } i_1 = i_2 \neq i_3 = i_4, \\ \overline{m}_4 & \text{при } i_i = i_2 = i_3 = i_4, \end{cases}$$

где $\overline{\sigma}^2 = \overline{M}\xi^2_i$, $\overline{m}_4 = \overline{M}\xi^4_i$. Отсюда имеем

$$\overline{M}\hat{\xi}^4 \leqslant \frac{\overline{m}_4}{n} + \frac{3\overline{\sigma}^4(n-1)}{n^3}, \quad \Delta_n \leqslant \frac{1}{\sqrt{n}} \left[\frac{\varkappa}{\alpha} \left(\frac{\overline{m}_4}{n} + \frac{3\overline{\sigma}^4(n-1)}{n} \right) \right]^{1/4}.$$

Аналогично, если $k = 3$, то

$$\overline{M}\hat{\xi}^6 \leqslant \frac{1}{n^6} \left(n\bar{m}_6 + C^2_6 n(n-1)\overline{m}_4\overline{\sigma}^2 + n(n-1)(n-2)\overline{\sigma}^6 \frac{6!}{(2!)^3 3!} \right),$$

$$\Delta_n \leqslant \frac{1}{\sqrt{n}} \left[\frac{\varkappa}{\alpha} \left(\frac{\bar{m}_6}{n^2} + C^2_6 \frac{n-1}{n^2} \overline{m}_4 \overline{\sigma}^2 + 15 \frac{(n-1)(n-2)}{n^2} \overline{\sigma}^6 \right) \right]^{1/6}.$$

Здесь видно, что чем больше n, тем в меньшей мере влияют величины старших моментов на ширину оценки (6.22). При $n \to \infty$ существенным оказывается лишь значение второго момента $\overline{\sigma}^2$.

Сделаем предположения. Пусть k — некоторое целое число и обратимся к оценке степенного типа (6.22). Считаем ξ_i одно-

родной последовательностью симметричных с.в., у которой $\bar{m}_{2k}=\bar{M}\xi_i^{2k}<\infty$. Для нее $M\xi_i^{2j-1}=0$ при $j=1, \ldots, k$, и $\bar{m}_{2j}$ конечны при $j\leqslant k$. Нужны также предположения относительно некоррелированности между собой степеней отсчетов, а именно, считаем $|\bar{M}|\xi_1^{j_1}\xi_2^{j_2}\ldots\xi_n^{j_n}=0$, если нечетно хотя бы одно j_i, где j_i принимают целые значения от 0 до $\sum j_i\leqslant 2k$ (требуемая некоррелированность будет, если ξ_i независимы и симметричны). Тогда для расчета $\bar{M}\hat{\xi}^{2k}$, а отсюда и для расчета ошибки $\alpha^{\varkappa}(\partial)=\varkappa\bar{M}\hat{\xi}^{2k}/\Delta^{2k}{}_n$ оценки (6.22) уместно воспользоваться допредельной формулой (3.11), а в асимптотике при $n\to\infty$ — утверждением 3.11, согласно которому $\lim\limits_{n\to\infty}\bar{M}\hat{\xi}^{2k}n^k\leqslant\bar{\sigma}^{2k}(2k)!/(k!2^k)$.

Подставляя правую часть в ошибку и приравнивая последнюю α, находим асимптотическое значение размаха Δ_n оценки степенного типа, нормированное множителем $\sqrt{n}$:

$$\lim_{n\to\infty}\Delta_n\sqrt{n}=\frac{\bar{\sigma}}{\sqrt{2}}\left[\frac{\varkappa(2k)!}{\alpha k!}\right]^{1/2k}. \qquad (6.23)$$

Обратим внимание, что при $n\to\infty$ размах Δ_n уменьшается пропорционально $1/\sqrt{n}$, так как правая часть от n не зависит. Как функция числа k правая часть имеет явно выраженный минимум, дающий наименьший размах оценки при данном уровне α. Это рождает задачу асимптотически (при $n\to\infty$) оптимального выбора числа k, т. е. подстройке формы оценки по параметру k в рамках структуры степенного типа (6.22), к чему и перейдем.

Асимптотическая подстройка оценки степенного типа. Считаем ξ_i симметричными, ограниченными (поэтому существуют моменты любых порядков) и независимыми. В результате при любом целом k будут выполнены предпосылки формулы (6.23), дающей асимптотическое при $n\to\infty$ выражение для размаха оценки степенного типа.

Ширина оценки (6.22) получается интегрированием

$$\Omega_n(\partial)=\int\limits_{-\Delta_n}^{\Delta_n}[1-(x/\Delta_n)^{2k}]\,dx=2\Delta_n\frac{2k}{2k+1}.$$

Подстановка сюда значения Δ_n из (6.23), обеспечивающего по крайней мере асимптотически уровень α, приводит к выражению для *асимптотически нормированной ширины*

$$\Omega_*=\lim_{n\to\infty}\Omega_n(\partial)\sqrt{n}=\sqrt{2}\,\bar{\sigma}\left[\frac{\varkappa(2k)!}{\alpha k!}\right]^{1/2k}\frac{2k}{2k+1}.$$

Поставим задачу выбора такого k, при котором достигается минимум правой части. Ее решение — сфера численных методов. Но на качественном уровне вполне приемлемы и аналитические, воспользоваться которыми позволяет формула Стирлинга:

$k!\approx(k/e)^k\sqrt{2\pi k}$. Подстановка ее на место факториалов в Ω_* дает:

$$\Omega_*=2\bar{\sigma}(\sqrt{2}\,\varkappa/\alpha)^{1/2k}\sqrt{2k/e}\,\frac{2k}{2k+1},$$

а минимизация по k ведет к уравнению $l_\alpha=k(2k+3)/(2k+1)$, где $l_\alpha=\ln(\sqrt{2}\varkappa/\alpha)$, решение которого есть $k^*=l_\alpha/2-3/4+\sqrt{l^2{}_\alpha-3l_\alpha/2+9/2}$. Нужно взять ближайшее целое к нему значение $[k^*]$, подстановка которого в Ω_* даст асимптотически нормированную ширину оценки (6.22). Ее приведенное (к единичной дисперсии) значение $\Omega_*/\bar{\sigma}$ при $\varkappa=1$ сведено в таблицу для сравнения с приведенной шириной $2\Phi^{-1}(1/2-\alpha/2)$ доверительного интервала нормальной выборки примера 6.5. При составлении таблицы брались по очереди целые значения $k^*=1, 2, \ldots$, по ним находились l_α, затем α (оптимальные уровни, соответствующие заданному k^*) и вычислялись $\Omega_*/\bar{\sigma}$, а параллельно для сравнения находились $2\Phi^{-1}(1/2-\alpha/2)$.

Степень оценки	1	2	3	4	5	6
Оптимальный уровень	0,267	0,086	0,03	0,01	$3{,}84\cdot10^{-3}$	$1{,}4\cdot10^{-3}$
Расплывчатость $\Omega_*/\bar{\sigma}$	2,62	3,9	4,84	5,62	6,3	6,9
$2\Phi^{-1}(1/2-\alpha/2)$	2,22	3,44	4,36	5,08	5,76	6,4

Мысль, проводимая нами и подкрепленная таблицей, вот в чем. При увеличении n нормированная сумма $\sum\xi_i/\sqrt{n}$ ограниченной симметричной независимой последовательности с.в. в соответствии с центральной предельной теоремой 3.16 должна сходиться к нормальному закону, это факт. Но тогда следует ожидать, что ширина оценки (6.22) тоже должна приближаться к ширине доверительного интервала для нормальной последовательности ξ_i (сумма которых, как известно, тоже нормальна). Эти ожидания, как явствует из таблицы, полностью оправдываются. Оправдываются и доказывают принципиальную возможность получения качественных оценок при крайне скудных знаниях о флуктуациях ξ_i. Отсутствие же полного совпадения объясняется тем, что из предельных свойств суммы $\sum\xi_i$ «выхвачены» только моменты, да и сама оценка по постановке заключена внутри форм степенного типа. Еще важно, что наши оценки «впитывают» в себя все имеющиеся и следующие из них данные, даже побочные.

Использование допредельных и предельных результатов. Приступим к более методичному привлечению допредельных и предельных результатов гл. 3 для синтеза оценок. В предыдущих

двух разделах рассматривались только степенные признаки сумм и, как следствие, моменты. В то же время не использованы данные о гармонических признаках, возникающие при суммировании независимых с. в. согласно исследованиям гл. 3.

Прежде чем высказать соображения на этот счет, сделаем небольшую остановку, чтобы резюмировать характерные свойства полученных выше оценок параметра сдвига x. Во-первых, оценки $\partial_{\mathbf{y}}(x)$ являются функциями $\xi_i = y_i - x$, что равносильно их эквивариантности к преобразованиям сдвига: $\partial_{\mathbf{y}+1a}(x+a) = \partial_{\mathbf{y}}(x) = \partial(\mathbf{y}-\mathbf{1}x) = \partial(\boldsymbol{\xi})$. Первопричина такого свойства в том, что любые значения x находятся в одинаковом по привилегиям априорном положении, если данных об x нет (данные, скажем, о диапазоне значений x умышленно можно игнорировать с целью достичь упрощений) и если шкала расплывчатости не взвешена. Во-вторых, для однородной выборки $y_i = x + \xi_i$ наблюдений оценка является функцией суммы $\sum \xi_i$ как последствие симметрии задачи к перестановкам наблюдений, в результате все наблюдения оказываются одинаково равноценными участниками в поиске скрытого x.

И, наконец, самое важное в настоящий момент, что оценки являются функциями не просто суммы, а *нормированной суммы* ζ_n. Это либо среднее арифметическое $\hat{\xi}$ в оценке (6.17), когда данных о корреляциях ξ_i нет, либо нормированная сумма $\hat{\xi}\sqrt{n} = \sum \xi_i / \sqrt{n}$, если ξ_i некоррелированная или независимая однородная последовательность. Последнее легко вытекает из (6.22), (6.23), если учесть, что оценка зависит от отношения $\hat{\xi}/\Delta_n$, а размах Δ_n пропорционален $1/\sqrt{n}$.

В общем, оценки записываются: $\partial_{\mathbf{y}}(x) = v(\zeta_n)$, где $v(z)$ есть некоторая унимодальная симметричная функция z, равная 1 при $z=0$ и спадающая к 0 при отклонении z от 0. Интегральная ширина функции $v(z)$, а для независимых флуктуаций с поправкой на множитель $1/\sqrt{n}$ дает расплывчатость оценки. В упомянутом случае $\zeta_n = \hat{\xi}\sqrt{n} = (\hat{y} - x)\sqrt{n}$ и

$$\int \partial_{\mathbf{y}}(x)\, dx = \int v((\hat{y} - x)\sqrt{n})\, dx = \int v(z)\, dz/\sqrt{n}.$$

В аргументе функции $v(\zeta_n)$ стоит нормированная сумма ζ_n, в поведении которой выделяется то, что ζ_n при увеличении n стабилизируется как случайная величина, т. е. не вырождается ни к нулю, ни к бесконечности. Удобно представлять себе с. в. ζ_n, как будто совершающую ограниченные колебания в диапазоне, меняющемся при увеличении n несущественно, в то время, как среднестатистические данные об ζ_n претерпевают изменения, да еще какие! Базу исследованиям дает гл. 3, где показано, что нормированная сумма сообразно исходным данным стремится к разным по ширине моделям, первичными для которых являются

степенные и гармонические признаки (вплоть до нормальной и интервальной нормальной моделей). Здесь интересно напомнить, что сведения о степенных и гармонических признаках сумм ζ_n возникают и уточняются даже тогда, когда похожих данных о слагаемых вроде бы и не было.

Сказанное наводит на мысль для независимых флуктуаций ξ_i при расчете ошибки оценки $\partial(\boldsymbol{\xi}) = v(\zeta_n)$, $\zeta_n = \sum \xi_i / \sqrt{n}$, и синтезе оптимальной оценки как можно шире пользоваться данными о ζ_n, предоставляемыми допредельными и предельными теоремами в виде средних степенных $\overline{M}\xi^k{}_n$ и гармонических $\overline{M}\sin u\zeta_n$, $\overline{M}\cos u\zeta_n$ признаков, взятых за первичные для ζ_n. При этом вроде бы «забываются» исходные данные о флуктуациях после того, как они были использованы для получения необходимых сведений относительно ζ_n.

Теперь расчет ошибки будет осуществляться так: для вычисления $\overline{M}\partial_{\mathbf{y}}(x) = \overline{M}v(\zeta_n)$ функция $v(z)$ мажорируется полиномами, смесью гармоник или теми, и другими, а при расчете $\underline{M}\partial_{\mathbf{y}}(x)$ (соответствующей верхней ошибке), наоборот, мажорирует их в соответствии с общим принципом продолжения средних. А если вспомнить принцип достаточности, формулируемый теоремой 6.1, то $v(z)$ и вовсе сама должна записываться в виде степенного ряда, гармонического ряда Фурье или получаться смешением того и другого. Так, в общем-то, и делалось в предыдущем разделе, но только для степенных признаков. Нам же предстоит это проделать еще и для гармонических.

Замечания. 1. Нужно отметить, что оценивание по допредельным неравенствам и предельным результатам нельзя назвать совсем оптимальными хотя бы потому, что мы «забываем» данные о ξ_i и пользуемся только тем, что узнаем по ним о ζ_n. И далее, хотя по данным о ζ_n оценка строится наилучшим образом, но такие оценки все же точнее назвать квазиоптимальными, а при $n \to \infty$ — *асимптотически оптимальными.*

2. Нормироваться сумма $\sum \xi_i$ в оценке $v(\zeta_n)$ при независимых флуктуациях не обязательно должна коэффициентом $1/\sqrt{n}$, это может быть и $1/n$, если данных об отдельных слагаемых почти нет.

Синтез квазиоптимальных оценок по гармоническим средним. Пусть ξ_i, $i = 1, 2, \ldots, n$, — однородная последовательность независимых с. в. и пусть $\partial(\boldsymbol{\xi}) = v(\zeta_n)$ есть оценка параметра сдвига, где $\zeta_n = \sum \xi_i / \sqrt{n}$ — нормированная сумма. Функцию $v(z)$, считая ее симметричной по отношению к началу координат, запишем через преобразование Фурье $V(u)$:

$$v(z) = \int_{-\infty}^{\infty} V(u) \cos(uz)\, du, \qquad V(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} v(z) \cos(uz)\, dz.$$

Это прямая и обратная формулы преобразования Фурье.

Раскладывая $V(u)=V^+(u)-V^-(u)$ на положительную и отрицательную части, имеем

$$\underline{M}\,v(\zeta_n)\geqslant\int_U V^+(u)\,\underline{M}\cos(u\zeta_n)\,du-\int_U V^-(u)\,\overline{M}\cos(u\zeta_n)\,du.$$

Мы видим, что ошибка $\bar{\alpha}(\partial)=1-\underline{M}v(\zeta_n)$ оценки выражается через гармонические средние $\underline{M}\cos(u\zeta_n)$ и $\overline{M}\cos(u\zeta_n)$ нормированных сумм. Данные гармонических средних доставляются допредельными и предельными теоремами главы 3, в общем, не для всех u, а для некоторой области U, учтенной при интегрировании.

Для простоты конкретизируем задачу. Пусть ξ_i независимы, имеют нулевые средние $M\xi_i=0$ и конечные дисперсии $\overline{M}\xi_i^2=\bar{\sigma}^2$. При этих данных последовательность ξ_i, в общем, статистически неустойчива, сходимости ζ_n к нормальной модели не будет. Тем не менее значения $M\cos(u\zeta_n)$ могли бы быть получены из допредельной теоремы дополнения к § 3.3, которые и позволили бы в принципе рассчитать ошибку $\bar{\alpha}(\partial)$ при конечном n (и следовательно, взвешенную ошибку, так как $\underline{\alpha}(\partial)=0$). Мы не будем выписывать все громоздкие следующие этому пути формулы хотя бы потому, что они принципиально мало чем отличаются от более простых вытекающих из них при $n\to\infty$ асимптотических формул, тех, что даются следствием 1 теоремы 3.13, на которых и остановимся.

Последние предоопределяют такие асимптотические данные о ζ_n:

$$M\zeta_n=0,\ \overline{M}\zeta_n^2=\bar{\sigma}^2,\ \underline{M}\cos(u\zeta_n)=\exp(-u^2\bar{\sigma}^2/2)\cos(c^*u^2\bar{\sigma}^2),\quad(6.24)$$

где $c^*=0{,}3184$, а $|u|^2\leqslant\pi/(2c^*\bar{\sigma}^2)$. Это неравенство для u и определяет область U, так что если $v(z)$ такова, что $V(u)\geqslant 0$ для всех u, то имеем асимптотическое значение

$$\lim_{n\to\infty}\underline{M}\,v(\zeta_n)\geqslant 2\int_0^{\pi/(2c^*\bar{\sigma}^2)}V(u)\exp(-u^2\bar{\sigma}^2/2)\cos(c^*u^2\bar{\sigma}^2)\,du,$$

где при выставлении пределов использована симметрия подынтегрального выражения.

Замечание. Гармонические средние в (6.22) может быть не столь значительно, но все же уточняют модель суммы по сравнению с $\overline{M}\zeta_n^2=\bar{\sigma}^2$. По одному только последнему значению $\bar{\sigma}^2$ продолжением имели бы: $\underline{M}\cos(u\zeta_n)\geqslant 1-u^2\bar{\sigma}^2/2$, а то, что содержится о ζ_n в выражении (6.24), несколько лучшее, хотя при $u\to 0$ они и стремятся друг к другу.

Пример 6.12. Пусть $v(z)$ есть по форме «гауссов колокол»: $v_\delta(z)=\exp(-\pi z^2/\delta^2)$, где δ — его интегральная ширина: $\int v_\delta(z)dz=\delta$. Тогда $V(u)=(\delta/2\pi)\exp(-u^2\delta^2/4\pi)\geqslant 0$ и ошибка асимптотически будет не больше:

$$\bar{\alpha}(\partial)\leqslant 1-\frac{\delta}{\pi}\int_0^{\pi/(2c^*\bar{\sigma}^2)}\exp\left\{-\frac{u^2}{2}\left(\bar{\sigma}^2+\frac{\delta^2}{2\pi}\right)\right\}\cos(c^*u^2\bar{\sigma}^2)\,du.$$

Приравнивая правую часть значению $\alpha/\varkappa$, рассчитывается ширина δ, обеспечивающая нужный уровень α. При этом напомним, что ширина самой оценки $v(\zeta_n)$ будет равна $\delta/\sqrt{n}$, убывая с ростом n пропорционально $\sqrt{n}$ при фиксированном асимптотически α.

Вопрос, конечно, почему в примере в качестве исходной формы $v(\zeta_n)$ взят «гауссов колокол»? Именно потому, что преобразование Фурье от него есть всюду неотрицательная функция. Этому факту имеется разумное обоснование. В самом деле, при принятых данных имеем $\overline{M}\cos u\zeta_n=1$, поэтому наличие отрицательной части $V(u)$ внесет довольно неприятный вклад в ошибку оценки в виде слагаемого $\int V^-(u)du$. В результате те $v(z)$, преобразование Фурье которых неотрицательно, должны иметь определенное преимущество в плане достаточности.

Последнее суждение требует некоторой корректировки, так как не принимает в счет еще одного обстоятельства: согласно (6.24) помимо гармонических средних имеются данные о квадратичном признаке $\overline{M}\zeta_n^2=\bar{\sigma}^2$. Если только на нем основывать оценку, то это привело бы нас обратно к оценке (6.18) параметра сдвига по некоррелированной выборке (и значительно более простым путем, правда, при дополнительном условии $M\xi_i=0$). А целесообразно, по-видимому, использовать все данные о ζ_n, гармонические и степенные, приводящие к следующей смешанной форме оценок:

$$v(z)=[c_0+\sum c_i^+\cos(u_i z)-c^+z^2]^+,\ c_0+\sum c_i^+=1,$$

при последующем выборе u_i, c_0, $c^+{}_i$ и c^+. Оценки этой формы симметричны и при $z=0$ согласно условию справа достигают максимального, равного 1, значения. Для них

$$\underline{M}\,v(\zeta_n)\geqslant c_0+\sum c_i^+\,\underline{M}\cos(u_i\zeta_n)-c^+\bar{\sigma}^2=1-\alpha/\varkappa,$$

а ширина, нормированная множителем $1/\sqrt{n}$, найдется интегрированием $v(z)$.

Пример 6.13. Пусть форма оценки задана выражением

$$v(z)=[\cos(\pi z/2\Delta)-z^2/(16\Delta^2)]^+,$$

где делитель при z^2 подобран так, чтобы выражение в квадратных скобках не имело побочных выбросов выше 0, кроме одного основного при $z=0$. Тогда подстановка на место аргумента $\zeta_n=\xi\sqrt{n}$ и использование (6.24) приводит к уравнению

$$\underline{M}\,\partial(\zeta_n)=\exp(-\pi^2 b/2)\cos(c^*\pi^2 b)-b/4=1-\alpha/\varkappa,$$

где $b=\bar{\sigma}^2/(4\Delta^2)$. Решением уравнения находится b, а отсюда и Δ.

Применение других допредельных и предельных результатов принципиально мало отличается от только что проделанного. По этому поводу сейчас выскажем некоторые соображения применительно к неоднородной выборке.

Об оценивании параметра сдвига при неоднородных флуктуациях. Пусть $y_i = x + \xi_i$, $i = 1, \ldots, n$, где ξ_i — независимые с.в. с нулевыми средними $M\xi_i = 0$ и разными дисперсиями: $\overline{M}\xi^2_i = \overline{\sigma}^2_i$. Тогда нормированную сумму ζ_n нужно заменить на взвешенную $\zeta_n = \sum_{i=1}^{n} c_{in}\xi_i$, $c_{in} \geqslant 0$, записав оценку через $v(\zeta_n)$. Интегральной шириной оценки по x будет величина $\int v(\sum c_{in} y_i - x\sum c_{in})\,dx = (\sum c_{in})^{-1} \int v(z)\,dz$, обратно пропорциональная сумме $\sum c_{in}$. Ошибка монотонно зависит в конечном счете от дисперсии: $\overline{M}\zeta^2_n = \sum c^2_{in}\overline{\sigma}^2_i$, поэтому выбор коэффициентов c_{in} будет, по крайней мере, близким к оптимальному, если минимизирует $\sum c^2_{in}\overline{\sigma}^2_i$ при заданном $\sum c_{in}$. Это ведет к значениям $c_{in} = \lambda/\overline{\sigma}^2_i$, откуда $\zeta_n = \lambda\sum \xi_i/\overline{\sigma}^2_i$, т. е. весовые коэффициенты оказываются обратными дисперсиями. Становится ясным, что чем больше $\overline{\sigma}^2_i$, тем с меньшим весом должен участвовать i-й отсчет в сумме ζ_n, ограждая ее тем самым от влияния этого излишне интенсивного, излишне шумящего отсчета.

С теми же весами получалась оценка (6.19), но для некоррелированных ξ_i. У нас ξ_i независимы, что позволяет достичь дополнительного эффекта оценки применением к сумме $\sum \xi_i/\overline{\sigma}^2_i$ допредельных и предельных теорем § 3.4. Дальнейший путь строго следует «колее» предыдущего раздела, и мы подкрепим его примером.

Пример 6.14. Пусть ξ_i независимы, ограничены, заданы $\overline{\sigma}^2_i$ и симметричны. Рассмотрим оценку вида $\partial(\xi) = [1 - (\zeta_n/\delta)^{2k}]^+$, $\zeta_n = \lambda\sum \xi_i/\overline{\sigma}^2_i$. Пусть $\overline{\sigma}^2_i \geqslant \varepsilon > 0$. Применение дополнения 3 к § 3.4 (самый конец главы 3) дает

$$\lim_{n\to\infty} \overline{M}\,\zeta_n^{2k} \leqslant \lim_{n\to\infty} \frac{(2k)!}{k!\,2^k} [\lambda^2 \Sigma (1/\overline{\sigma}^2_i)]^k,$$

откуда $\delta = \lambda[\varkappa(2k)!/(k!\alpha)]^{1/2k}\sqrt{\Sigma(1/2\overline{\sigma}^2_i)}$. При $\lambda = [\Sigma(1/\overline{\sigma}^2_i)]^{-1}$ случайная величина ζ_n будет стабилизирована по n, а асимптотическая нормированная ширина оценки

$$\Omega_* = \lim_{n\to\infty} \Omega_n(\partial)\sqrt{n} = \sqrt{2}\left[\frac{\varkappa(2k)!}{\alpha k!}\right]^{1/2k} \overline{\sigma}_{\text{ср}},$$

где $\overline{\sigma}^2_{\text{ср}} = \left[\frac{1}{n}\Sigma\left(1/\overline{\sigma}^2_i\right)\right]^{-1}$. Ее минимум по k ищется точно так, как это проделывалось в начале настоящего параграфа при асимптотической подстройке оценки степенного типа.

6.5. ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРА МАСШТАБА

Общие соображения. Параметр масштаба определяет энергетические свойства наблюдений: мощность, эффективное значение. Задача его оценивания возникает, во-первых, когда уровень интенсивности процесса несет полезную информацию об интересующем нас положении объекта. Например, величина гидроакустических шумов может характеризовать расстояние до косяка рыб или до корабля. Сам шум в этом случае в некотором смысле обретает ранг полезности. А во-вторых, потребность измерения уровня шумов возникает при определении и текущем контроле показателей работающей аппаратуры, например, для фиксации вероятности ложных тревог в радиолокационных станциях. И в этом случае оценивание уровня шума нельзя считать «второсортным», второстепенным делом, так как оценки составным элементом входят в единый синтезируемый комплекс функционирования аппаратуры.

С позиций современной теории вероятностей оценивание масштабных параметров занимает особое положение. Обязано оно непрерывному времени и допущению о точном знании корреляционных свойств, так как тогда, разлагая процесс конечной временной протяженности от 0 до T в ряд по собственным функциям корреляционного ядра, получаем бесконечное число некоррелированных коэффициентов разложения, которые нормировкой (делением на корни из их дисперсий) все приводятся к одинаковым дисперсиям. В результате происходит превращение наблюдений в бесконечную последовательность преобразованных, новых значений, некоррелированных между собой и одинаково реагирующих на изменение масштаба, и таким образом, несущих неограниченный запас сведений об этом параметре; запас, позволяющий сколь угодно точно его выделить (причем по отрезку $(0, T)$ наблюдений сколь угодно малой протяженности!). Этот парадокс — эффект точного знания — получил название сингулярности [1].

Явление сингулярности оказывается серьезной преградой в реальном оценивании параметра масштаба, так как все оценки рефлекторно тянутся в ее сторону как сулящую «райские кущи» неограниченного улучшения их качественных показателей, перекрывая поиски других путей и напрочь заставляя забыть о том, что сокращение длины интервала $(0, T)$ наблюдений предъявляет все более жесткие требования к точности корреляционной функции, чтобы хоть как-то гарантировать некоррелированность даже небольшого числа компонентов разложения.

Наша цель — строить оценки параметров масштаба по реальным, конечным данным, что само собой исключит любую возможность таких чисто математических «фокусов», как сингулярность, поставит теорию оценивания на почву реальности. Причем, как всегда, за начало возьмем самые слабые исходные данные (о явлении) энергетического толка, а затем будем их постепенно усиливать в нужном направлении, не забывая следить за происходящими при этом видоизменениями оптимальной оценки. В конце концов, полезно знать, какие реальные сведения должны иметься,

[1] Гихман И. И., Скороход А. В. Теория случайных процессов. — М.: Наука, 1971. — Т. 1. — 664 с.

чтобы с нужной точностью оценить параметр масштаба, в частности, его разновидность — дисперсию.

Оценивание параметра масштаба по заданной мощности флуктуаций. Пусть $y_i = x^{l/2}\xi_i$, $i=1, \ldots, n$, $x>0$, $l\neq 0$. Требуется оценить параметр x, если известно, что ограничена средняя мощность флуктуаций: $\overline{M\xi^2}=\overline{\sigma}^2$, $\widehat{\xi^2}=\sum\xi_i^2/n$, и более ничего не дано.

Отметим, что при $\overline{\sigma}=1$ и $l=1$ параметр x будет *мощностью* (при $M\xi_i=0$ — *дисперсией*) наблюдений, а при $l=2$ — *средним эффективным* их значением. При $l=-2$ это будет *параметр нормировки*. В общем случае $l\neq 0$ будем называть x *параметром масштаба*.

Согласно следствию 4 теоремы 6.1 достаточный класс образуют оценки, зависящие от $\widehat{\xi^2}$ с подстановкой сюда $\xi_i=y_i/x^{l/2}$, что ведет к их виду $\partial_y(x)=[c_0-c^+\widehat{y^2}/x^l]^+$, и для них $\underline{M}\partial_y(x)=[c_0-c^+\overline{\sigma}^2]^+$, $\overline{M}\partial_y(x)=c_0$.

Найдем интегральную ширину и штраф, рассматривая сначала случай $l<0$,

$$\Omega(\partial)=\int_0^\infty [c_0-x^{-l}c^+\widehat{y^2}]^+dx=\frac{c_0 l}{l-1}\left(\frac{c_0}{c^+\widehat{y^2}}\right)^{-1/l},$$

$$\overline{M}\,\Omega(\partial)=\frac{c_0 l}{l-1}\sup_{\widehat{y^2}}\left(\frac{c_0}{c^+\widehat{y^2}}\right)^{-1/l}.$$

Правая часть равна ∞ при $\widehat{y^2}=0$, поэтому есть смысл предположить $\widehat{y^2}\geqslant\varepsilon>0$, считая ε сколь угодно малым числом. Тогда составной риск запишется

$$\Pi_\lambda^\varkappa(\partial)=1-\varkappa[c_0-c^+\overline{\sigma}^2]^+ +(\varkappa-1)c_0+\lambda c_0\frac{l}{l-1}\left(\frac{c_0}{c^+\varepsilon}\right)^{-1/l}.$$

Из этого выражения после некоторых выкладок следует, что оптимальным будет значение $c_0=1$. Значение же c^+ будет определяться уровнем α, что при $\alpha<\varkappa$ даст $c^+=\alpha/(\varkappa\overline{\sigma}^2)$. После подстановки найденных значений *оптимальная оценка параметра* x примет вид

$$\partial_y^*(x)=[1-x^{-l}\alpha\,\widehat{y^2}/(\varkappa\overline{\sigma}^2)]^+,\quad x\geqslant 0. \tag{6.25}$$

Так, при $l=-2$ (тогда $y_i=\xi_i/x$) оптимальная оценка параметра нормировки x получится равной: $\partial^*{}_y(x)=[1-x^2\alpha\widehat{y^2}/(\varkappa\overline{\sigma}^2)]^+$, $x>0$. Как видно из рис. 6.5,*а*, это есть полупарабола с максимумом в точке $x=0$, усеченная снизу осью абсцисс и слева — осью ординат.

Пусть теперь $l>0$. Оценка будет иметь вид рис. 6.5,*б* и имеет бесконечную ширину. Тогда, чтобы штраф не равнялся бесконечности, нужно рассматривать взвешенную шкалу интегрирования по x с весовым множителем $q(x)$, убывающим при $x\to\infty$, по крайней мере, быстрее $1/x$. Считая $q(x)=x^{-1-\gamma}$, $\gamma>0$, (т. е. чем больше

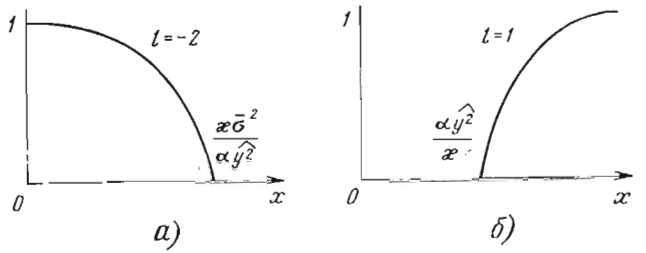


Рис. 6.5, *а*, *б*. Оцена параметра масштаба по энергетическим данным

x, тем меньше берется вес), найдем штраф как верхнее среднее взвешенной ширины:

$$\overline{M}^\gamma\,\Omega(\partial)=\overline{M}^\gamma\int_0^\infty\left[c_0-\frac{c^+\widehat{y^2}}{x^l}\right]^+\frac{dx}{x^{1+\gamma}}=\overline{M}^\gamma\int_{\left(\frac{c^+}{c_0}\widehat{y^2}\right)^{1/l}}^\infty\left(\frac{c_0}{x^{1+\gamma}}-\right.$$

$$\left.-\frac{c^+\widehat{y^2}}{x^{1+\gamma+l}}\right)dx=\frac{c_0 l}{\gamma+l}\max_{\widehat{y^2}}\left(\frac{c_0}{c^+\widehat{y^2}}\right)^{\gamma/l}=\frac{c_0 l}{\gamma+l}\left(\frac{c_0}{c^+\varepsilon}\right)^{\gamma/l}.$$

Составляя риск и производя его минимизацию (точно так же, как это делалось при $l<0$), получим оценку (6.25), которая оказывается оптимальной при любых l.

Так, при $l=1$ и $\overline{\sigma}^2=1$ (x есть мощность наблюдений) доверительная оценка x примет вид (см. рис. 6.5,*б*): $\partial_y(x)=[1-\alpha\widehat{y^2}/(x\varkappa)]^+$. Как функция переменной x оценка равна 0 при $x\leqslant\alpha\widehat{y^2}/\varkappa$ и стремится к 1 при $x\to\infty$.

Дополнения. 1. Тот же самый вид (6.25) будет иметь оптимальная оценка для непрерывных реализаций $y_t=x^{l/2}\xi_t$, $0\leqslant t\leqslant T$, с той лишь разницей, что $\widehat{y^2}=\frac{1}{T}\int_0^T y_t^2\,dt$.

2. Сведения о точной мощности (дисперсии) $\widehat{M\xi^2}=\sigma^2$, а также о нулевом среднем флуктуаций $M\xi_i=0$ (или же $\widehat{M\xi}=0$) не меняют вида оценки, т. е. эти сведения оказываются ненужными. То же самое, если x является свободным от ξ.

3. Столь большая расплывчатость получающихся оценок есть следствие крайне скудных данных о флуктуациях. Дело в том, что поставленная задача, когда известна только средняя мощность флуктуаций, в пессимистическом настрое эквивалентна оцениванию параметра масштаба по одному-единственному наблюдению, так как не исключается случай постоянных реализаций $\xi_1=\xi_2=\ldots=\xi_n$.

4. Если заданы точное значение $M\widehat{\xi^2}=\sigma^2$ и верхняя граница $\overline{M}(\widehat{\xi^2})^2=\overline{m}_4$, то доверительная оценка уровня α примет вид

$$\partial_y(x)=[1-c^*(1-x^{-l}\,\widehat{y^2})^2]^+,\quad c^*=\alpha/[\varkappa(1+\overline{m}_4-2\overline{\sigma^2})].$$

Эта оценка, как видно из рис. 6.6, равна 1 при $x=(\widehat{y^2})^{1/l}$ и убывает по мере отклонения x от этого значения. Некоторое улучшение оценки здесь достигается за счет данных о моментах четвертого порядка. Это в некотором роде общее свойство оценок моментов: для получения более удовлетворительных оценок момента k-го порядка нужно иметь сведения о моментах порядка $2k$. Сейчас мы лишний раз в этом убедимся.

Оценивание параметра масштаба по некоррелированной выборке. Пусть $y_i=x^{l/2}\xi_i$, $i=1,\ldots,n$, $x>0$, $l\neq 0$, и нужно оценить параметр масштаба x, если флуктуации некоррелированы и имеют одинаковые мощности:

$$M\xi_i^2=\sigma^2,\ M\xi_i\xi_j=0,\ i\neq j,\ i,j=j=1,\ldots,n.$$

Оптимальную оценку будем искать в следующем достаточном классе: $\partial_y(x)=[c_0-x^{-l}(c_1\hat{y}^2+c_2\hat{\sigma}^2{}_y)]^+$, $\hat{\sigma}^2{}_y=\widehat{y^2}-\hat{y}^2$. При нахождении этого класса учтены первичные данные и симметрия поставленной задачи к перестановкам наблюдений. Для этого класса $M\partial_y(x)=[c_0-c_1\sigma^2/n-c_2\sigma^2(n-1)/n]^+$, $\overline{M}\partial_y(x)=c_0$ при $c_1,\ c_2\geqslant 0$.

Точно так же, как это делалось в предыдущем разделе, только с заменой $c^+\hat{y}^2$ на $c_1\hat{y}^2+c_2\hat{\sigma}^2{}_y$, вычисляется штраф. При $l<0$ он будет равен

$$\overline{M}\,\Omega(\partial)=c_0^{(l-1)/l}\,l/[(l-1)\inf(c_1\,\hat{y}^2+c_2\,\hat{\sigma}^2_y)]=c_0^{(l-1)/l}\,l/(l-1)\,c_2\,\varepsilon,$$

где инфимум достигается при $\hat{y}=0$ и $\hat{\sigma}^2{}_y=\varepsilon^2$, причем допущено $\widehat{y^2}\geqslant\varepsilon>0$ (иначе ширина будет бесконечна), считая ε сколь можно малым. Оптимальным будет значение $c_0=1$, и тогда при $\alpha<\varkappa$ риск записывается двумя уравнениями: $\Pi^{\varkappa}{}_{\lambda}(\partial)=\alpha+\lambda/c_2$, $\alpha=\varkappa(c_1\sigma^2/n+c_2\sigma^2(n-1)/n)$. Чем больше c_2, тем меньше риск. Поэтому учитывая последнее равенство, оптимальными будут $c_1=0$, $c_2=\alpha n/(\varkappa(n-1)\sigma^2)$, в результате чего оптимальная оценка приобретает вид

$$\partial^*_y(x)=\left[1-x^{-l}\,\frac{\alpha n}{\varkappa(n-1)\sigma^2}\,\hat{\sigma}^2_y\right]^+.$$

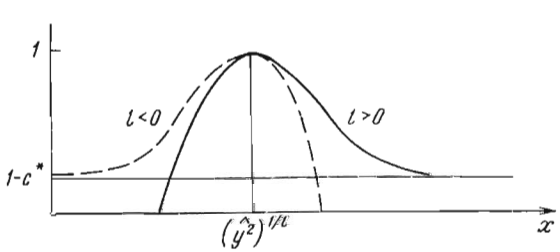


Рис. 6.6. Оценка параметра масштаба при заданном четвертом моменте

Точно такой же вид будет иметь оценка при $l>0$ (если брать взвешенную множителем $x^{1-\nu}$ шкалу интегрирования в штрафе).

Таким образом, *данные о некоррелированности наблюдений при отсутствии другой информации лишь незначительно меняют оценку и ее свойства по сравнению с рис. 6.5,а и б*, а именно, оценка сужается в $n/(n-1)$ раз, что при увеличении n становится практически не заметным. Причина этого неприятного эффекта кроется в том, что для получения хороших оценок требуются данные о четвертых моментах флуктуаций.

Допустим теперь, что помимо некоррелированности $M\xi_i\xi_j=0$, $i\neq j$, и $M\xi^2{}_i\xi^2{}_j=\sigma^4$, $i\neq j$, $\overline{M}\xi^4{}_i=\overline{m}_4$. Тогда достаточный класс образуют оценки вида $\partial_y(x)=[c_0+c_1x^{-l}\widehat{y^2}+c_2x^{-l}\sum\limits_{i\neq j}\sum y_iy_j-c_3x^{-2l}\widehat{y^4}-c_5x^{-2l}\sum\limits_{i\neq j}\sum y^2{}_iy^2{}_j]^+$, где $\widehat{y^4}=\sum y^4{}_i/n$. Постараемся более или менее «угадать» вид оптимальной оценки из этого класса, записав ее в виде

$$\partial_y(x)=[1-c\,(1-x^{-l}\,\widehat{y^2}/a)^2]^+. \tag{6.26}$$

Очевидно следующее: $\overline{M}\partial_y(x)=\overline{\partial}=1$, причем максимум оценки достигается при $x^l=\widehat{y^2}/a$. Если расписать оценку через ξ, раскрыв круглые скобки, то получим

$$\partial(\xi)=\left[1-c+\frac{2c}{a}\,\widehat{\xi^2}-\frac{c}{a^2n^2}\left(\sum\xi_i^4+\sum\limits_{i\neq j}\sum\xi_i\xi_j\right)\right]^+$$

и отсюда будет следовать, что

$$\underline{M}\partial_y(x)=[1-c+2c\sigma^2/a-c\overline{m}_4/na^2-c(n-1)\sigma^4/na^2]^+.$$

Считая выражение в квадратных скобках неотрицательным, ищем сначала оптимальное значение параметра a, тогда как c потом будет найдено исходя из заданного уровня α. Если вычислить штраф, полагая $\widehat{y^2}\geqslant\varepsilon>0$, то он будет равен $\overline{M}\Omega(\partial)=\varepsilon^{1/l}a^{-1/l}r(c)$, где $r(c)=(1-c)(c_+^{1/l}-c_-^{-1/l})+2c(c_+^{(l-1)/l}-c_-^{(l-1)/l})/(1-l)-c(c_+^{(2l-1)/l}-c_-^{(2l-1)/l})/(1-2l)$, а $c_+=1+1/\sqrt{c}$, $c_-=1-1/\sqrt{c}$. Подставляя найденные значения в составной риск, дифференцируя его по a и приравнивая 0, получим значение a как решение уравнения: $\lambda l\varepsilon^{1/l}r(c)\,a^{(2l-1)/l}+\sigma^2a-(\overline{m}_4-\sigma^4)/n-\sigma^4=0$. Чем меньше уровень α, тем меньше будет значение λ. Поэтому при малых α первым слагаемым левой части равенства можно пренебречь, положив $\lambda=0$. В результате получим значение параметра a в явном виде: $a=\sigma^2+(\overline{m}_4/\sigma^2-\sigma^2)/n$. Дальнейшие упрощения могут быть получены, если считать n большим, и тогда $a\approx\sigma^2$. Отметим, что

увеличение a по сравнению с σ^2 ведет к увеличению средней ширины при $l<0$ и уменьшению при $l>0$.

Полагая для простоты $a=\sigma^2$, получаем исходя из заданной ошибки α значение $c^*=\alpha n\sigma^4/\varkappa(\bar{m}_4-\sigma^4)$. В результате искомая оценка приобретает вид

$$\partial_y(x)=\left[1-\frac{\alpha n\sigma^4}{\varkappa(\bar{m}_4-\sigma^4)}(1-x^{-l}\hat{y}^2/\sigma^2)^2\right]^+. \qquad (6.27)$$

Оценка имеет максимум при $x^l=\hat{y}^2/\sigma^2$ и принимает ненулевые значения в интервале с границами $\left[\frac{\sigma^2}{\hat{y}^2}\left(1\pm\sqrt{\frac{\varkappa}{\alpha n}\left(\frac{\bar{m}_4}{\sigma^4}-1\right)}\right)\right]^{-1/l}$. Размах этого интервала, характеризующий расплывчатость оценки, будет тем меньше, чем больше α, а самое главное, эта ширина будет стремиться к 0 при увеличении n со скоростью $1/\sqrt{n}$. Это то самое, чего мы добивались.

Развитие проблемы. 1. Для расчета оценки (6.27) использовалось точное значение мощности σ^2. В случае неточного значения: $\underline{\sigma}^2=\underline{M}\xi^2_i$, $\bar{\sigma}^2=\bar{M}\xi^2_i$ ошибка в (6.26) равна $\alpha^x(\partial)=\varkappa(1-\underline{M}\partial_y)=\varkappa c[1-2\underline{\sigma}^2/a+\bar{m}_4/na^2+\bar{\sigma}^4(n-1)/na^2]$. Оптимальным значением a, минимизирующим ошибку, будет $a=\bar{\sigma}^4/\underline{\sigma}^2+(\bar{m}_4-\bar{\sigma}^4)/n\underline{\sigma}^2$.

2. Данные о $\bar{m}_4-\sigma^4$, необходимые для оценки (6.27), могут быть получены косвенным путем. Например, если ξ_i ограничены числом H, т. е. $\underline{P}(|\hat{\xi}_i|<H)=1$, то неравенство $\bar{M}(\xi^2-\sigma^2)^2\leqslant H^4/4$ позволяет $\bar{m}_4-\sigma^4$ заменить на $H^4/4$.

3. Рассмотрим вопрос об асимптотически (при $n\to\infty$) наилучших оценках, если известно, что $\bar{M}\xi^{2k}_j<\infty$, $\forall k>0$. Считая ξ_i независимыми и полагая для простоты $\underline{M}\xi^2_i=\sigma^2=1$, введем следующий класс степенных оценок

$$\partial(\xi)=[1-(1-\hat{\xi}^2)^{2k}/\Delta^{2k}]+,$$

куда для получения $\partial_y(x)$ вместо ξ_i нужно подставить $x^{-l/2}y_i$. Если обозначить $\eta_i=\xi^2_i-1$, то $M\eta_i=\underline{M}\xi^2_i-1=0$ и тогда $\partial(\eta)=[1-(\hat{\eta}/\Delta)^{2k}]+$. Мы видим, что в новых обозначениях оценка совпадает с (6.22). Тогда при $n\to\infty$ по аналогии с оценкой (6.22) имеем

$$1-\underline{M}\partial=[\bar{M}(\xi_i^2-1)^2/\Delta^2 n]^k(2k)!/k!\,2^k+0(1/n).$$

Отсюда, пренебрегая последним слагаемым правой части и используя то, что $\bar{M}\partial=1$, получаем ширину Δ оценки по заданному уровню α:

$$\Delta=\left(\frac{\varkappa(2k)!}{\alpha k!\,2^k}\right)^{1/2k}\sqrt{\frac{\bar{M}(\xi_i^2-1)^2}{n}}.$$

Если воспользоваться теперь в целях упрощений формулой Стирлинга, то получим $[(2k)!/k!2^k]^{1/2k}\approx\sqrt{2k/e}2^{1/4k}$, откуда

$$\Delta\approx\frac{1}{\sqrt{n}}\left(\sqrt{2}\,\frac{\varkappa}{\alpha}\right)^{1/2k}\sqrt{2k/e}\,H,\quad H=\sqrt{\bar{M}(\xi_i^2-1)^2}.$$

При оптимальном значении $k^*=\ln(\sqrt{2}\varkappa/\alpha)$, минимизирующем ширину Δ оценки, полагая k^* ближайшим целым числом, имеем окончательно $\Delta_*=(H/\sqrt{n})\sqrt{2\ln(\sqrt{2}\varkappa/\alpha)}$ и в результате оценка $\partial_y(x)$ приобретает вид:

$$\partial_y(x)=[1-(1-x^{-l}\hat{y}^2)^{2k^*}/\Delta_*^{2k^*}]+.$$

4. Данные о нулевом среднем $M\xi_i=0$ не меняют найденных оценок. Сказанное не означает, что средние $M\xi_i$ могут быть произвольными, так как некоррелированность $M\xi_i\xi_j=0$, $i\neq j$, сама по себе ограничивает возможные вариации среднего,

5. Пусть $y_i=x^{l/2}(\xi_i+\theta)$, $i=1,\ldots,n$, где $\theta\in\mathscr{R}$ — скалярный мешающий параметр. Тогда оценку следует искать в виде: $\partial_y(x)=[1-c(1-x^{-l}\hat{\sigma}^2_y/\sigma^2)^2]+$. Чтобы ошибка этой оценки стремилась к 0 при увеличении n, требуются следующие условия: $M\xi_i\xi_j=0$, $i\neq j$; $M\xi^2_i=\sigma^2$; $M\xi^2_i\xi^2_j=\sigma^4$, $i\neq j$; $\bar{M}\xi^4_i=\bar{m}_4$; $M\xi_i\xi_j\xi_k\xi_l=0$, $i\neq j$, $i\neq k$, $i\neq l$. При этих условиях $1-\underline{M}\partial_y(x)=c[\bar{m}_4(1/n-2/n^2+1/n^3)/\sigma^4-1/n+3/n^2-1/n^3]=\alpha/\varkappa$, откуда и находится коэффициент c. Такой же останется оценка для наблюдений $y_i=x^{l/2}\xi_i+m$, $m\in\mathscr{R}$.

6. Рассмотрим случай, когда мощности флуктуаций ξ_i по известному закону меняются: $M\xi^2_i=\sigma^2_i$, $\bar{M}\xi^4_i=\sigma^4_i\bar{m}_4$. Тогда переходя к новым наблюдениям $z=y_i/\sigma_i=x^{l/2}\xi_i/\sigma_i=x^{l/2}\eta_i$, мы сводим задачу к рассмотренной с $M\eta^2_i=1$. Оценка будет отличаться от найденных в этом пункте лишь тем, что вместо $\hat{y}^2$ подставляется $\sum y^2_i/(\sigma^2_i n)$. Аналогично, если задана корреляционная матрица $\mathbf{B}$ вектора ξ, то вектор $\eta=\mathbf{B}^{-1/2}\xi$ будет обладать только что рассмотренными свойствами, поэтому $\hat{y}^2$ заменяется на $\mathbf{y}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{y}/n$.

Мы не сможем в этой задаче перейти к процессам $y_t=x^{t/2}\xi_t$, пока не изучим влияние ошибок в знании σ^2_i на свойства оценок. Дело в том, что для процессов с точным значением корреляционной функции $B(t,\tau)=M\xi_t\xi_\tau$, разлагая ядро по собственным функциям $B(t,\tau)=\sum_i\sigma^2_i\varphi_i(t)\varphi_i(\tau)$ в новых наблюдениях $z_i=\int y_t\varphi_i(t)\,dt/\sigma_i=x^{l/2}\eta_i$ получаем бесконечно длинную некоррелированную последовательность случайных величин η_i, $M\eta^2_i=1$, что соответствует $n=\infty$ и приведет к абсолютно точной оценке x. Этот парадокс сингулярности, о котором говорилось во введении к параграфу, есть следствие допущения о точном знании $B(t,\tau)$, предположения, увы, практически не выполнимого. Выход содержится в следующем пункте.

7. Пусть мощности флуктуаций не являются точно известными, т. е. $y_i=x^{l/2}\xi_i$, $M\xi_i\xi_j=0$, $\underline{M}\xi^2_i=\underline{\sigma}^2_i$, $\bar{M}\xi^2_i=\bar{\sigma}^2_i$, $\bar{M}\xi^2_i\xi^2_j=\bar{\sigma}^2_i\bar{\sigma}^2_j$,

$\bar{M}\xi^4_i = m_i$, $i \neq j = 1, \ldots, n$. Ищем оценку в виде $\partial_y(x) = [1 - c(1 - \sum y^2_i/d_i)^2]^+$, $d_i \geqslant 0$. Для нее $\bar{M}\partial_y = 1$ и

$$\bar{\alpha}(\partial) = 1 - \underline{M}\,\partial_y = c\left[1 - 2\sum \frac{\underline{\sigma}^2_i}{d_i} + \sum \frac{\bar{m}_i - \bar{\sigma}^4_i}{d_i^2} + \left(\sum \frac{\bar{\sigma}^2_i}{d_i}\right)^2\right]^+.$$

Определим d_i исходя из минимизации ошибки $\bar{\alpha}(\partial)$. Для этого продифференцируем правую часть равенства по d_i и приравняем 0. Получим $d_i = (\bar{m}_i - \bar{\sigma}^4_i)/[\underline{\sigma}^2_i - \gamma\bar{\sigma}^2_i]^+$, где γ находится из уравнения $\gamma = \sum \bar{\sigma}^2_i[\underline{\sigma}^2_i - \gamma\bar{\sigma}^2_i]^+/(\bar{m}_i - \bar{\sigma}^4_i)$. Отметим, что $\gamma < 1$. Мы видим, что $d_i = \infty$ при $\underline{\sigma}^2_i/\bar{\sigma}^2_i < \gamma$. Таким образом, если ошибка в знании σ^2_i достаточно велика, то наблюдение y_i делится на ∞ и исключается. *Накапливаются лишь те y^2_i/d_i, для которых $\underline{\sigma}^2_i/\bar{\sigma}^2_i > \gamma$.*

Сказанное относится и к процессам, так как в разложении неточной корреляционной функции в ряд по собственным функциям ядра коэффициенты разложения, обозначенные так же σ^2_i, станут неизвестными. Причем чем больше индекс i, тем меньше их абсолютные значения и тем меньше будет отношение $\underline{\sigma}^2_i/\bar{\sigma}^2_i$. Поэтому, как и выше, придется ограничиться лишь конечным усеченным рядом, включающим только «главные компоненты» разложения.

6.6. ЗАКЛЮЧЕНИЕ

Логика такова, что на неясный вопрос не жди вразумительного ответа. Естественным ответом на размытость моделей, вызванную априорной бедностью данных, будет расплывчатость оценок.

Оценки искомых параметров должны подчиняться двум диаметрально противоположным требованиям. С одной стороны, быть надежными, и следовательно, в нужной мере расплывчатыми, а с другой — потребительский интерес вынуждает их к конкретности, т. е. по возможности к наибольшей точности. Разрешение этого противоречия ведет к оптимальным доверительным оценкам фиксированной надежности.

Введение составного риска как суммы штрафа за расплывчатость (в виде площади под кривой решений) плюс величина (вероятность) взвешенной ошибки достигает своей цели — существенного упрощения методов синтеза доверительных оценок. Даже при классических распределениях вероятностей минимизация составного риска ведет к новым по содержанию общим формулам для доверительных интервалов (§ 6.2) (для нормальных распределений превращающимся в известные), ставить и решать новые задачи, например, найти доверительную оценку дисперсии, обладающую минимальной расплывчатостью; получить оптимальные доверительные оценки параметров регрессии.

Самое важное, что составной риск, органично сочетаясь с конструкцией интервальных моделей в виде первичных средних, выходит на простые структуры оптимальных оценок, которыми будут усеченные снизу линейные комбинации первичных признаков, о чем гласит теорема 6.1. В зависимости от первичных данных получаются самые разнообразные по форме оценки, дающие

по шкале 0—1 картину предпочтений значениям параметра. В частности, при вероятностных моделях это будут доверительные интервалы (предпочтение полное 1 внутри и 0 вне интервала) ввиду того, что первичные признаки здесь индикаторные.

Меньше данных — легче поиск оптимальной оценки. Разве не это правильная пока непривычная концепция? Нет данных — никакого поиска не надо, оценки тривиальны (всем значениям будет приписано одинаковое предпочтение).

Однородность исходных данных, их симметрия откликается в структурах оценок уменьшением числа варьируемых коэффициентов, подлежащих оптимизации. Коэффициенты связываются линейным уравнением, фиксирующим надежность, а в остальном находятся из минимума интегральной расплывчатости оценки. Их поиск вкладывается в типовую параметрическую задачу оптимизации с ограничениями. Мы ее решаем аналитически, не исключая полезности численных способов в виде стандартных программ.

Оптимальные доверительные оценки параметра сдвига находятся в § 6.3, где считается известной только усредненная мощность флуктуаций. Более точными они становятся для некоррелированных флуктуаций, а затем обобщаются на случай заданных корреляций и на оценивание параметра (одномерного) регрессии. Примечательно, что параболический вид все эти оценки наследуют от квадратичных признаков, составляющих свойства второго порядка наблюдений. И характерно для нашей теории, что оценки по сложности одинаковы как для последовательности наблюдений, так и для процессов, если состав первичных данных одинаков. По мере роста числа данных оценки несколько усложняются, зато и уточняются, становятся по форме более узкими при той же надежности.

Предугадывается потребительский вопрос, лучше ли полученные оптимальные оценки, чем привычные доверительные интервалы для нормального распределения? На наивный вопрос ответ неутешителен: конечно же, хуже! И могут быть намного, так как разные оценки рассчитаны на разные условия. Одни условия — это режим полного априорного благополучия в виде нормального распределения вероятностей, по нему считается доверительный интервал. Полученные нами оценки оптимальны в условиях полного неблагополучия, свойственного бедным знаниями моментным задачам, поэтому более расплывчаты.

И все же, неблагополучие их относительное. При независимых отсчетах по мере увеличения числа наблюдений без каких бы то ни было дополнительных предположений наши оценки приближаются по свойствам к нормальным (§ 6.4). Можно лишь подозревать, что этому мы обязаны внутренним законам нормальной сходимости, формально при синтезе нигде не задействованным. Данный результат демонстрирует великолепное «пищеварение» разработанного аппарата, позволяющего при крайне бедном априорном пайке усваивать любые питательные вещества. Внедрение допредельных и предельных результатов в оценку на систематическую основу поставлено в конце § 6.4.

Конец главы посвящен задаче оценивания параметра масштаба (в частности, мощности) по данным о свойствах второго порядка наблюдений. Неожиданным оказывается здесь то, что отсутствие ограничений на моменты четвертого порядка не позволяет получить сколь-либо хорошие оценки даже при неограниченном росте длины наблюдений. Но исправить их при наличии нуж-

ных ограничений оказывается очень просто. И совершенно правильно ожидалось, что проблема сингулярности (состоящая в абсолютной точности оценки дисперсии, а в общем-то и сдвига) сама собой становится невозможной при отказе от идеальных вероятностных моделей и переходе к реальным интервальным. Сингулярность — побочный плод теоретизированного изобилия и недосягаемая мечта для привычной практическим задачам априорной бедности.

Глава 7.
ПРОВЕРКА ГИПОТЕЗ

7.1. ОБЩИЕ ПОЛОЖЕНИЯ

Введение. Отвлеченно представим себе некоторый ящик, который неважно, чем начинен, и что в нем происходит и как, но важно то, что процессы внутри могут находиться в некоторых двух состояниях. Одно состояние обычно, рабочее, условно именуемое нулевым, не вызывает у нас никаких реакций и не требует вмешательства, а другое, напротив, критическое, называемое альтернативным. Нужно по наблюдениям за выходом этого «ящика» решить, какое из двух состояний в настоящий момент присутствует, причем в условиях когда выход настолько засорен шумами или флуктуациями, что однозначно и с полной уверенностью решение принять невозможно, так как появление одних и тех же реализаций свойственно как нулевому, так и альтернативному состояниям, но с разными вероятностями.

Последнее очень существенно и подразумевает наличие среднестатистических данных о выходе, о наблюдениях. Это то, за что можно и нужно ухватиться для решения проблемы. Источник получения данных для нас сейчас не принципиален: либо это результат исследований внутренней физической природы процесса, либо последствие предварительных наблюдений за выходом, когда к этому говорится, какому состоянию он соответствует (обучение с учителем), или может быть обучение без учителя. Важна форма представления данных, которая должна соответствовать нашей генеральной концепции формирования интервальной модели средних: иметь вид размытых среднестатистических свойств признаков выходных наблюдений. Благодаря языку первичных данных это очень доступная и универсальная форма, в том мы не раз успели убедиться. Причем обязательно данные так или иначе должны связываться с состояниями, какое оно, нулевое или альтернативное(иначе решение станет тривиальным).

Важным здесь является анализ и получение данных не о любых доступных признаках с целью использования их всех, а только о наиболее характерных, выделяющих те направления, которые оказываются наиболее чувствительными к смене состояний (лучше меньше, да лучше). Тогда первичных данных будет меньше, а задача — проще. Именно этот случай нас наиболее интересует еще и потому, что при невозможности обучения или

предварительного анализа многие реальные задачи оказываются «сидящими на голодном пайке» априорного дефицита.

Поскольку действительные состояния неизвестны и предстоит это еще решить, то называются они *гипотезами*: соответственно *нулевая* гипотеза и *альтернативная* (проще, альтернатива). Они могут иметь разный вес, приоритет. Ведь отказ от нулевой гипотезы, т. е. принятие альтернативы, если оно неправильное, может привести в итоге к тяжелейшим последствиям. Например, принимается неверное решение о движении группы неопознанных объектов в направлении чьей-то территории, что может вызвать мгновенную ответную агрессию, которая не найдет оправданий. Или принимается решение о безвредности нового лекарства, тогда как это на самом деле не так. Или осуждается невиновный.

Сказанное подводит к принципу, так сказать, «презумпции гипотезы», согласно которому лучше несколько раз отвергнуть альтернативу, пусть даже ошибочно, когда она на самом деле верна, чем хотя бы раз неправильно отказаться от гипотезы. Последствием этого принципа чаще будет приниматься гипотеза, а альтернатива — лишь тогда, когда этому есть вполне веские основания, не вызывающие особых сомнений. Таким образом, гипотезе придается приоритет, вес, и тем самым в нее закладывается консервативное начало, а в альтернативу — прогрессивное (иногда, агрессивное). Это важно для понимания смысла правил проверки гипотез с фиксированным уровнем ошибок, о которых ниже пойдет речь.

Не остается в стороне и задача различения гипотез, соответствующая бесприоритетным гипотезам, например, тот случай, когда по каналу связи передаются посылки 0 (соответствует нулевой гипотезе) или 1 (альтернативе). На самом же деле, если вес заменить вероятностью гипотезы (средней ожидаемой частотой чередования нулевых посылок), то выбором вероятности можно всегда учесть желаемый приоритет, как и наоборот. Это убеждает нас (и будет математически подкреплено), что задачи проверки и различения гипотез, т. е. с приоритетной и бесприоритетной гипотезами являются смежными, подменивающими друг друга.

Математическое оформление задачи. Пусть по наблюдениям $y \in \mathcal{Y}$ требуется проверить *нулевую гипотезу* с соответствующей ей моделью $\mathcal{M}^y_0$, против *альтернативы* с соответствующей $\mathcal{M}^y_1$. Состояний x в этом случае два: формально $x=0$ при нулевой гипотезе и $x=1$ — при альтернативе. Решающее правило $\partial_y(x)$ по общему определению § 5.1 в результате трансформируется в пару $\partial_y=(\partial^0_y, \partial^1_y)$, где числа ∂^0_y и ∂^1_y (представляющие удобную запись $\partial_y(0)$ и $\partial_y(1)$) выражают степень предпочтения, отдаваемого соответственно нулевой гипотезе и альтернативе. В силу относительности предпочтений будем считать $\partial^0_y+\partial^1_y=1$, что позволяет в дальнейшем рассматривать либо ∂^0_y, либо ∂^1_y.

С учетом последнего равенства правило ∂_y можно интерпретировать как *рандомизацию*, при которой с вероятностью ∂^0_y при-

нимается нулевая гипотеза, а с вероятностью ∂^1_y — альтернатива. Рандомизация здесь имеет смысл процедуры, реализующей отношение предпочтения, если обязательно требуется принять одну из двух гипотез (в целом же понятие предпочтения более объемное, чем рандомизация).

Правило такое, что $\inf_y \partial^0_y = 0$, $\sup_y \partial^0_y = 1$, назовем *контрастным*, а если ∂^0_y (отсюда и ∂^1_y) принимают только значения 0 и 1 — то *детерминированным*.

Модели $\mathcal{M}^y_0$ и $\mathcal{M}^y_1$, класс $\mathcal{D}$ возможных правил, коэффициент пессимизма $\varkappa$ — все вместе они образуют *статистическую задачу проверки гипотез*. Свойства правил меряются ошибками, которые и определим сейчас.

Уровнем правила называется среднестатистическая величина ошибки, состоящей в принятии альтернативы при условии, что гипотеза верна. Меряется уровень относительно $\mathcal{M}^y_0$: $\alpha^\varkappa(\partial) = \varkappa\overline{M}_0\partial^1_y + (1-\varkappa)\underline{M}_0\partial^1_y = 1-\varkappa\underline{M}_0\partial^0_y - (1-\varkappa)\overline{M}_0\partial^0_y$, где $\underline{M}_0$ и $\overline{M}_0$ соответствуют $\mathcal{M}^y_0$. Та же самая $\alpha^\varkappa(\partial)$ — величина *ошибки первого рода*, задающая консервативность правила: чем меньше $\alpha^\varkappa(\partial)$, тем сохранней станет гипотеза. Балансом к ней является *ошибка второго рода*, определяемая формулой

$$\beta^\varkappa(\partial) = \varkappa\,\overline{M}_1\,\partial^0_y + (1-\varkappa)\,\underline{M}_1\,\partial^0_y = 1 - \varkappa\,\underline{M}_1\partial^1_y - (1-\varkappa)\,\overline{M}_1\,\partial^1_y .$$

Это есть вероятность неправильного принятия гипотезы при условии верной альтернативы (измеряется относительно $\mathcal{M}^y_1$).

Если интерпретировать значения правила ∂_y в интервале 0—1 как рандомизацию, то $\alpha^\varkappa(\partial)$ и $\beta^\varkappa(\partial)$ будут иметь смысл вероятностей ошибок, взвешенных коэффициентом пессимизма. А в общем, это будут количественные выражения среднего предпочтения, отдаваемого ошибочно одной гипотезе при справедливости другой. Вычисляются вероятности ошибок по методике продолжения средних с первичных данных (задающих исходно $\mathcal{M}_0$ и $\mathcal{M}_1$) на ∂^0_y, ∂^1_y, рассматриваемые как признаки пространства $\mathcal{Y}$.

Обе ошибки балансируют друг друга: желание уменьшить одну из них вызывает тут же увеличение другой и наоборот. Оптимизация на уровне синтеза состоит в связанном понижении обеих ошибок или только одной из них при заданном значении другой, что вложено в следующее определение.

Оптимальным при уровне α называется правило проверки гипотез, минимизирующее ошибку второго рода $\beta^\varkappa(\partial)$ при фиксированном значении величины ошибки первого рода: $\alpha^\varkappa(\partial) = \alpha$. Нахождение оптимальных правил и есть та основная задача, которая ставится в этой главе.

Оптимальное правило ищется в классе всех правил, т. е. всех функций ∂^0_y со значениями в отрезке $[0; 1]$. Для его нахождения, если воспользоваться методом множителей Лагранжа, достаточно сделать два шага. Первый — это решение *смежной задачи*: минимизации по ∂, *весовой суммы ошибок* $r^\varkappa_\lambda(\partial) = \lambda\alpha^\varkappa(\partial) + \beta^\varkappa(\partial)$, при этом минимизирующее правило будет зависеть от множителя λ,

придающего вес гипотезе. Второй шаг будет состоять в выборе такого λ, чтобы уровень минимизирующего правила равнялся α. Полученное правило и будет оптимальным уровня α правилом проверки гипотезы.

Правило, минимизирующее риск в виде весовой суммы ошибок, называется оптимальным правилом различения гипотез. Подмечена эквивалентность этих двух видов правил, где риск может рассматриваться как усредненная ошибка при заданных вероятностях гипотез.

Действительно, весовая сумма ошибок, если ее переписать $r^\varkappa_\lambda(\partial) = [p_0\alpha^\varkappa(\partial) + (1-p_0)\beta^\varkappa(\partial)]/(1-p_0)$, положив для этого $\lambda = p_0/(1-p_0)$, при интерпретации p_0 как вероятности гипотезы с точностью до множителя совпадает с полной вероятностью ошибки, заключенной в квадратные скобки. Таким образом, каждому λ соответствует задача с заданной вероятностью p_0, и наоборот.

Класс $\mathcal{D}^*$ правил называется достаточным, если любому правилу ∂ и числам $\lambda>0$, $\varepsilon>0$ можно указать такое $\partial^*\in\mathcal{D}^*$, что $r^\varkappa_\lambda(\partial^*) \leqslant r^\varkappa_\lambda(\partial) + \varepsilon$.

Понятие достаточности вытекает из общего определения § 5.2 и сформулировано применительно к смежной задаче. Если ошибка первого рода минимизирующего правила непрерывна по λ, то тот же класс $\mathcal{D}^*$ будет достаточным в основной задаче, причем при любом α. Если вдруг такой непрерывности нет, то ее можно добиться некоторым расширением $\mathcal{D}^*$.

Основная теорема о достаточности. Обозначим, как всегда, $\mathcal{L}+\mathcal{G}$ — класс вторичных признаков, основанных на $\mathcal{G}$, т. е. $\mathcal{L}+\mathcal{G} = \{g(y) = c_0 + \sum c^+_i g_i(y),\ g_i\in\mathcal{G}\}$. При проверке гипотез, если $\mathcal{G}^0$ — первичные признаки, задающие нулевую гипотезу $\mathcal{M}^y_0 = \langle\overline{M}_0\mathcal{G}^0\rangle$, а $\mathcal{G}^1$ — задают альтернативу, $\mathcal{M}^y_1 = \langle\overline{M}_1\mathcal{G}^1\rangle$, этих классов будет два: $\mathcal{L}+\mathcal{G}^0$ и $\mathcal{L}+\mathcal{G}^1$.

Следующая теорема связывает достаточные классы правил с видом вторичных (и следовательно, первичных) признаков, соответствующих гипотезе и альтернативе. В самом деле, на каком из классов $\mathcal{L}+\mathcal{G}^0$ или $\mathcal{L}+\mathcal{G}^1$ основывать правило? Оказывается, на том и на другом, а при $\varkappa = 1$ — либо на том, либо на другом, как сейчас будет доказано.

Теорема 7.1. *Пусть $\langle\overline{M}^y_0\mathcal{G}^0\rangle$ и $\langle\overline{M}^y_1\mathcal{G}^1\rangle$ определяют соответственно гипотезу и альтернативу, где $\mathcal{G}^0$ и $\mathcal{G}^1$ — приведенные к верхним первичные наборы. Пусть $0\leqslant g^0(y)\in\mathcal{L}+\mathcal{G}^0$, $0\leqslant g^1(y)\in\mathcal{L}+\mathcal{G}^1$ — неотрицательные из вторичных признаков гипотезы и альтернативы.*

1. *При $\varkappa\geqslant 1$ достаточными будут классы правил любого из двух видов* **а** *или* **б**:

а. $\partial^1_y = \min\{1, g^0(y), g^1(y)\}$, $\partial^0_y = \max\{0, 1-g^0(y), 1-g^1(y)\}$;

б. $\partial^1_y = \max\{0, 1-g^0(y), 1-g^1(y)\}$, $\partial^0_y = \min\{1, g^0(y), g^1(y)\}$.

2. *При* $\varkappa=1$ *достаточными будут правила любого следующего вида* **а** *или* **б**:

а. $\partial^1_y=\min\{1,\ g^0(y)\},\ \partial^0_y=[1-g^0(y)]^+$,

причем такие, что $\overline{\alpha}(\partial)=\overline{M}_0 g^0(y)=c_{00}+\sum c^+_{i0}\overline{M}_0 g^0_i$ *при* $g^0(y)=c_{00}+\sum c^+_{i0} g^0_i\ \ y)$;

б. $\partial^1_y=[1-g^1(y)]^+,\ \partial^0_y=\min\{1, g^1(y)\}$, *причем такие, что* $\overline{\beta}(\partial)=\overline{M}_1 g^1(y)=c_{01}+\sum c^+_{i1}\overline{M}_1\partial^1_i$ *при* $g^1(y)=c_{01}+\sum c^+_{i1} g^1_i(y)$.

Доказательство. **1а.** Пусть $\overset{\smile}{\partial}_y$ — какое-то правило. Покажем, что можно найти не худшее его правило из достаточного класса, т. е. с таким же риском или отличающимся на сколь угодно малое ε. Выпишем весовую сумму ошибок $r^\varkappa_\lambda(\overset{\smile}{\partial})=\lambda\varkappa\overline{M}_0\overset{\smile}{\partial}{}^1_y+(\varkappa-1)\overline{M}_1\overset{\smile}{\partial}{}^1_y-\lambda(\varkappa-1)\underline{M}_0\overset{\smile}{\partial}{}^1_y-\varkappa\underline{M}_1\overset{\smile}{\partial}{}^1_y$. Идея доказательства будет состоять в мажорировании произвольно выбранного $\overset{\smile}{\partial}{}^1_y$ по очереди вторичными признаками гипотезы и альтернативы, которые и передадут $\overset{\smile}{\partial}{}^1_y$ свои средние значения, определяя $\overline{M}_0\overset{\smile}{\partial}{}^1_y$ и $\overline{M}_1\overset{\smile}{\partial}{}^1_y$. Нижняя грань признаков и есть правило достаточного класса, не худшее $\overset{\smile}{\partial}{}^1_y$. А теперь — более строго и подробно.

По следствию к теореме 1.1 правилу $\overset{\smile}{\partial}{}^1_y$ и заданному $\varepsilon>0$ можно указать такой мажорирующий его вторичный признак нулевой гипотезы $g^0(y)=c_{00}+\sum c^+_{i0}g^0_i(y)\in\mathscr{L}+\mathscr{G}^0$, что $g^0(y)>\overset{\smile}{\partial}{}^1_y$ и $\overline{M}_0 g^0(y)-\overline{M}_0\overset{\smile}{\partial}{}^1_y\leqslant\varepsilon/\lambda$ (причем $\overline{M}_0 g^0(y)=c_{00}+\sum c^+_{i0}\overline{M}g^0_i$). Аналогично можно указать такой $g^1(y)\in\mathscr{L}+\mathscr{G}^1$, что $g^1(y)\geqslant\overset{\smile}{\partial}{}^1_y$ и $\overline{M}_1 g^1(y)-\overline{M}_1\overset{\smile}{\partial}{}^1_y\leqslant\varepsilon$ (причем $\overline{M}_1 g^1(y)=c_{01}+\sum c^+_{i1}\overline{M}g^1_i$). Теперь введем $\partial^1_y=\min\{1,\ g^0(y),\ g^1(y)\}$. Это есть решающее правило, так как $0\leqslant\partial^1_y\leqslant 1$, оно мажорируется функциями $g^0(y)$ и $g^1(y)$, поэтому с учетом $\varkappa\geqslant 1$ имеем $\lambda\varkappa\overline{M}_0\partial^1_y+(\varkappa-1)\overline{M}_1\partial^1_y\leqslant\lambda\varkappa\overline{M}_0 g^0(y)+(\varkappa-1)\overline{M}_1 g^1(y)\leqslant\lambda\varkappa\overline{M}_0\overset{\smile}{\partial}{}^1_y+(\varkappa-1)\overline{M}_1\overset{\smile}{\partial}{}^1_y+\varepsilon$. По мажорируемости $\partial^1_y\geqslant\overset{\smile}{\partial}{}^1_y$ верны неравенства $\underline{M}_0\partial^1_y\geqslant\underline{M}_0\overset{\smile}{\partial}{}^1_y$, $\underline{M}_1\partial^1_y\geqslant\underline{M}_1\overset{\smile}{\partial}{}^1_y$, в результате чего получаем $r^\varkappa(\partial)\leqslant r^\varkappa_\lambda(\overset{\smile}{\partial})+\varepsilon$, что доказывает достаточность правил вида, взятого для ∂^1_y, а формула для ∂^0_y эквивалентна $1-\partial^1_y$.

1б. Здесь доказательство отличается от предыдущего разве что выбором $g^0(y)\in\mathscr{L}+\mathscr{G}^0$ и $g^1(y)\in\mathscr{L}+\mathscr{G}^1$, удовлетворяющими неравенствам $g^0(y)\geqslant\overset{\smile}{\partial}{}^0_y$, $g^1(y)\geqslant\overset{\smile}{\partial}{}^0_y$, и таким, при котором $\overline{M}_0 g^0(y)-\overline{M}_0\overset{\smile}{\partial}{}^0_y\leqslant\varepsilon/\lambda$, $\overline{M}_1 g^1(y)-\overline{M}_1\overset{\smile}{\partial}{}^0\leqslant\varepsilon/\lambda$. Тогда, вводя правило $\partial^0_y=\min\{1,\ g^0(y),\ g^1(y)\}$, точно как выше доказывается $r^\varkappa_\lambda(\partial)\leqslant r^\varkappa_\lambda(\overset{\smile}{\partial})+\varepsilon$.

2а. При $\varkappa=1$ имеем $r^1_\lambda(\overset{\smile}{\partial})=\lambda\overline{M}_0\overset{\smile}{\partial}{}^1_y+1-\overline{M}_1\overset{\smile}{\partial}{}^1_y$. Выберем $g^0(y)$, как это делалось в **1а**, и положим $\partial^1_y=\min\{1;\ g^0(y)\}$. Тогда $\lambda\overline{M}_0\partial^1_y\leqslant\lambda\overline{M}_0 g^0(y)\leqslant\lambda\overline{M}_0\overset{\smile}{\partial}{}^1_y+\varepsilon$. Так как $\partial^1_y\geqslant\overset{\smile}{\partial}{}^1_y$, то $\underline{M}_1\partial^1_y\geqslant\underline{M}_1\overset{\smile}{\partial}{}^1_y$, и поэтому $r^1_\lambda(\partial)\leqslant r^1_\lambda(\overset{\smile}{\partial})+\varepsilon$, что и требовалось. Доказательство **2б** аналогично.

Комментарии. 1. Случай пессимизма $\varkappa=1$ привлекателен тем, что для правил **2а**, базирующихся на вторичных признаках гипотезы, сразу находится ошибка первого рода, а для правила **2б** — второго рода. В связи с этим для правил заданного уровня α рекомендуется форма **2а**, где фиксация уровня эквивалентна равенству $\overline{M}_0 g^0(y)=\alpha$ и осуществляется в виде линейного ограничения $c_{00}+\sum c^+_{i0}\overline{M}_0 g^0_i=\alpha$ на коэффициенты. А поиск оптимального правила сводится к минимизации ошибки второго рода $\overline{\beta}(\partial)$ (даваемой формулой продолжения средних для $\mathscr{M}^{y_1}$ и так или иначе определяемой коэффициентами) при ограничениях на коэффициенты: одного, фиксирующего уровень α, и другого, вытекающего из требования $\partial_y\leqslant 1$.

2. При $\varkappa>1$ и для правил **1а** имеем (используя факты из доказательства теоремы 7.1): $\overline{\alpha}(\partial)\leqslant\min\{\overline{M}_0 g^0,\ \overline{M}_0 g^1\}$ и можно удовлетвориться теми $g^0(y)$ и $g^1(y)$, для которых верно равенство. Здесь трудоемким может оказаться вычисление $\overline{M}_0 g^1$, если $g^1(y)$ — признаки «не свои», т. е. не класса $\mathscr{L}+\mathscr{G}^0$ гипотезы, и тогда для вычисления ошибки первого рода также потребуется применение формулы продолжения средних.

3. Если $\mathscr{G}^0=\mathscr{G}^1$, т. е. гипотеза и альтернатива базируются на одних и тех же первичных признаках, то правила **2а** (или **2б**) с теми же соотношениями для ошибок будут достаточными не только при $\varkappa=1$, но и при всех $\varkappa>1$, что следует из первой части теоремы (так как $g^0(y)=g^1(y)$).

4. Вместо $\mathscr{L}+\mathscr{G}^0$ и $\mathscr{L}+\mathscr{G}^1$ в условиях теоремы могут фигурировать любые более широкие классы признаков.

5. Если $\mathscr{L}+\mathscr{G}^0\cap\mathscr{L}+\mathscr{G}^1=\varnothing$, то достаточные классы правил, соответствующие пунктам **а** и **б** теоремы, не будут пересекаться между собой. Тогда оптимальное правило, поскольку оно может принадлежать как одному, так и другому достаточному классу, не будет единственным. Вообще нельзя говорить о минимальном достаточном классе вследствие того, что он очень часто не существует.

6. Достаточность, утверждаемая теоремой, казалось бы, специальна в том смысле, что относится к конкретному риску в виде линейной суммы взвешенных (коэффициентом пессимизма $\varkappa\geqslant 1$) ошибок. Но теорема верна для любого риска, монотонно связанного с этими ошибками, и в этом смысле универсальна для проверки гипотез. Например, для $r(\partial)=\max\{\overline{\alpha}(\partial),\ \overline{\beta}(\partial)\}$.

7. Достаточные при $\varkappa=1$ классы части 2 теоремы можно рекомендовать использовать и при оптимизме $\varkappa<1$.

Пример 7.1. Различие между случаями **2а** и **2б** теоремы выявим на простом примере, когда гипотеза и альтернатива заданы каждая одним первичным значением в виде верхних вероятностей $\overline{P}_0(A^0)=p_0$, $\overline{P}_1(A^1)=p_1$ соответственно событий A^0, A^1 из $\mathscr{Y}$. Тогда первичными будут $A^0(y)$ — для гипотезы и $A^1(y)$ — для альтернативы, а вторичными соответственно $\mathscr{L}+\mathscr{G}^0=$

$=\{c+c^+{}_0A^0(y)\}$, $\mathscr{L}+\mathscr{G}^1=\{c+c^+{}_1A^1(y)\}$, и достаточные классы в теореме 7.1 при $\varkappa=1$ примут (с заменой c на $1-c$) вид

2а. $\partial^0_y = c - c_0^+ A^0(y)$, $c_0^+ \leqslant c \leqslant 1$, $\overline{\alpha}(\partial) = 1 - c + c_0^+ p_0$,

2б. $\check{\partial}^1_y = \check{c} - c_1^+ A^1(y,)$ $c_1^+ \leqslant \check{c} \leqslant 1$, $\overline{\beta}(\check{\partial}) = 1 - \check{c} + c_1^+ p_1$.

Они изображены на рис. 7.1 в виде кривых $\partial^0{}_y$ для случая 2а и $\check{\partial}^1{}_y$ для 2б.

Для правил вида **2а** альтернативе отдается предпочтение лишь при $y \in A^0$, т. е. в области известной вероятности $P_0(A^0)$, позволяющей вычислить при $\varkappa=1$ величину ошибки первого рода, и следовательно, фиксировать уровень правила приравниванием $\alpha = 1 - c + c^+{}_0 p_0$. Уменьшение уровня достигается либо уменьшением $c^+{}_0$, либо увеличением c, иначе говоря, за счет «*расплывания*» вширь правила.

Примем нужное условие: $A^0 \cup A^1 = \mathscr{Y}$, $A^0 A^1 \neq \varnothing$. Оно отражено на рис. 7.1: множества A^0 и A^1, перекрываясь между собой, охватывают вместе все $\mathscr{Y}$. Это условие позволяет найти (мажорируя $\partial^0{}_y$ вторичным признаком альтернативы: $c - c^+{}_0 + c^+{}_0 A^1(y) \geqslant c - c^+{}_0 A^0(y)$ и подставляя его среднее) вероятность ошибки второго рода: $\overline{\beta}(\partial) = c - c^+{}_0 + c^+{}_0 p_1$ (без принятого условия мажорантой будут постоянное c и $\overline{\beta}(\partial) = c$).

Найдем оптимальное правило. Для этого находим коэффициенты c и c^+, минимизирующие $\overline{\beta}(\partial)$ (линейную комбинацию) при ограничении в виде равенства α. Имеем

1. $c = 1 - \alpha$, $c_0^+ = 0$, $\partial^0_y \equiv 1 - \alpha$, $\overline{\beta} = 1 - \alpha$ при $p_0 + p_1 \geqslant 1$;

2. $c = c_0^+ = \dfrac{1-\alpha}{1-p_0}$, $\partial^0_y = \dfrac{1-\alpha}{1-p_0}[1 - A^0(y)]$, $\overline{\beta} = \dfrac{1-\alpha}{1-p_0} p_1$

при $p_0 + p_1 < 1$, $p_0 \leqslant \alpha$

3. $c = 1$, $c_0^+ = \dfrac{\alpha}{p_0}$, $\partial^0_y = \left[1 - \dfrac{\alpha}{p_0} A^0(y)\right]$, $\overline{\beta} = 1 - \dfrac{\alpha}{p_0}(1 - p_1)$

при $p_0 + p_1 < 1$, $p_0 > \alpha$.

Аналогичен расчет правила типа 2б, да и $\check{\partial}^0{}_y$ получается таким же с той лишь разницей, что $A^0(y)$ заменяется на $1 - A^1(y)$ (A^0 заменяется дополнением к A^1). Проще говоря, область предпочтения нулевой гипотезе расширяется с дополнения к A^0 (см. рис. 7.1) до A^1. Ошибки у обоих типов правил совершенно одинаковы. Отсюда следует и более глубокий вывод, что любое правило

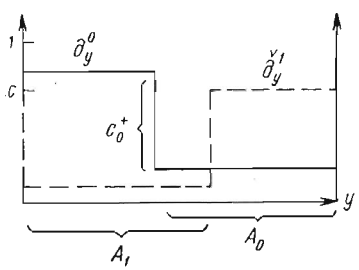


Рис. 7.1. Два типа достаточных правил

$\partial^0{}_y$, располагающееся между $\partial^0{}_y$ и $\check{\partial}^0{}_y$: $\partial^0{}_y \leqslant \leqslant \partial^0{}_y \leqslant \check{\partial}^0{}_y$ (т. е. в области пересечения $y \in A^0 A^1$ не скачком, а произвольно спадающее с одного уровня предпочтения на другой), будет иметь те же ошибки и тоже будет оптимальным.

Инвариантность и симметрия. А что, если при некоторых преобразованиях пространства $\mathscr{Y}$ на себя модели гипотезы и альтернативы остаются неизменными, т. е. строго в математическом смысле обнаруживают свойства инвариантности или симметрии к преобразованиям? Тогда ожидается передача похожих свойств к оптимальным правилам, что эквивалентно дополнительному сужению достаточного класса. Данная мысль полностью укладывается в рамки общих принципов инвариантности и симметрии § 5.4 при состоянии x, принимающем два значения: 0 или 1.

Теорема 7.2. *Пусть первичные признаки ИМ гипотезы* $\langle \bar{M}^y{}_0 \mathscr{G}^0 \rangle$ *либо альтернативы* $\langle \bar{M}^y{}_1 \mathscr{G}^1 \rangle$ *инвариантны к группе* $\mathscr{S}$ *преобразований пространства* $\mathscr{Y}$: $g(sy) = g(y)$, $s \in \mathscr{S}$, $g \in \mathscr{G}^j$, $j = 0$ *или* 1. *Тогда при* $\varkappa = 1$ *класс инвариантных к* $\mathscr{S}$ *правил (функций максимального инварианта) является достаточным.*

В самом деле, все функции класса $\mathscr{L}+\mathscr{G}^0$ будут инвариантны к $\mathscr{S}$. Отсюда инвариантными становятся решающие правила части **2а** теоремы 7.1, образующие при $\varkappa=1$ достаточный класс.

Теорема 7.2 будет справедлива при $\varkappa > 1$, *если обе модели, соответствующие как гипотезе, так и альтернативе, будут инвариантны к* $\mathscr{S}$.

Статистическая задача проверки гипотез называется *симметричной* к группе $\mathscr{S}$ преобразований пространства $\mathscr{Y}$, если $\mathscr{M}^y{}_0$ и $\mathscr{M}^y{}_1$ симметричны к $\mathscr{S}$ в смысле $\bar{M}_j f(sx) = \bar{M}_j f(x)$, $\forall f \in \mathscr{F}$, $s \in \mathscr{S}$, $j = 0, 1$.

Очевидно, из инвариантности обеих моделей к группе $\mathscr{S}$ будет следовать их симметрия (но не наоборот). Данное здесь понятие симметрии связывается с общим определением § 5.4. В самом деле, для этого нужно вместо пары $\mathscr{M}^y{}_0$, $\mathscr{M}^y{}_1$ перейти к СИМ $\mathscr{M}^{xy} = \mathscr{J}^x \mathscr{M}^y{}_x$, в которой x принимает одно из двух значений: 0 или 1 и никаких данных о x нет. Симметрия этой СИМ эквивалентна симметрии задачи проверки гипотез, поэтому применяя теорему 5.7, получаем следующее.

Теорема 7.3. *Пусть задача проверки гипотез симметрична к дискретной группе* $\mathscr{S}$ *преобразований пространства* $\mathscr{Y}$. *Тогда класс инвариантных к* $\mathscr{S}$ *правил является достаточным для этой задачи.*

Таким образом, при наличии симметрии задачи при поиске оптимальных правил можно ограничиться функциями максимального к $\mathscr{S}$ инварианта.

Пример 7.2. Пусть $y_1, \ldots, y_n$ — однородная выборка, заданная первичными значениями $\bar{M}_j g_k(y_i) = \bar{m}_{g,j}$, $g_k \in \mathscr{G}$, $i = 1, \ldots, n$, $j = 0, 1$. Задача симметрична к группе перестановок y_i между собой, и так как эта группа дискретна, то в классе инвариантных к перестановкам нужно искать оптимальное решающее правило. Согласно части 2а теоремы 7.1 структура оптимального при $\varkappa = 1$ правила будет иметь вид:

$$\partial^0_y = \Big[c_0 - \sum_k c_k^+ \hat{g}_k(y)\Big]^+, \quad \hat{g}_k(y) = \sum_{i=1}^n g_k(y_i)/n, \quad g_k \in \mathscr{G},$$

причем при таких коэффициентах $c^+{}_k$, что $\underline{\alpha}(\partial) = 1 - c_0 + \sum_k c^+{}_k \bar{m}_{g_k}, 0$.

Обнаружение сигнала по вероятностям превышений. Проиллюстрируем на одном частном случае, как формируется достаточный класс правил и как находится оптимальное.

Пусть выборка дискретная $y_1, \ldots, y_n$. Пусть нулевая гипотеза (интерпретируем как отсутствие сигнала) определяется первичными вероятностями $\bar{P}_0(|y_i|>h)=\bar{q}_0(h)$. При каждом h это есть та граница, которую не сможет никогда превзойти вероятность превышения уровня h при наихудшем для гипотезы стечении обстоятельств. Считаем $\bar{q}_0(h)$ неубывающими функциями h, они-то и определят ИМ $\mathscr{M}^{\mathbf{y}}_0$. При альтернативе (наличии сигнала) вероятности превышений возрастают и становятся не меньше $\underline{q}_1(h)=\underline{P}_1(|y_i|>h)$, они будут первичными для $\mathscr{M}^{\mathbf{y}}_1$. Причем мы не оговариваем здесь, каким является сигнал, а лишь опираемся на тот факт, что его наличие изменяет вероятности превышений уровней h, всех, либо по обстоятельствам некоторого набора уровней или только одного, что соответствует разным наборам первичных данных.

Такая постановка встречается, если характерными, отличительными чертами, сопутствующими сигналу, или просто которые мы знаем, является рост частоты превышений уровня или уровней.

Поскольку первичными и для $\mathscr{M}^{\mathbf{y}}_0$, и для $\mathscr{M}^{\mathbf{y}}_1$ являются индикаторные функции превышений $\{|y_i|>h\}$, согласно теореме 7.1 достаточным при $\varkappa=1$ будет класс правил вида:

$$\partial^0_{\mathbf{y}}=[c_0-\sum_i\sum_{h_j}c_i^+(h_j)\{|y_i|>h_j\}]^+.$$

А на основании симметрии задачи к перестановкам отсчетов коэффициенты $c_i(h_j)$ не должны зависеть от i и в результате правило упрощается:

$$\partial^0_{\mathbf{y}}=[c_0-\sum_{h_j}c^+(h_j)\,n(h_j)]^+,$$

где $n(h_j)=\sum_i\{|y_i|>h_j\}$ — число y_i, превысивших значение h_j. Причем ошибки первого и второго рода будут найдены, если взять $\bar{M}$ от выражений внутри квадратных скобок (ограничив результат снизу нулем, а сверху — единицей), и равны соответственно:

$$\bar{\alpha}(\partial)=1-c_0+\sum_{h_j}c^+(h_j)\,\bar{nq}_0(h_j),$$

$$\bar{\beta}(\partial)=c_0-\sum_{h_j}c^+(h_j)\,n\underline{q}_1(h_j).$$

Осталось при заданном $\bar{\alpha}(\partial)=\alpha$ выбором коэффициентов c_0 и $c^+(h_j)$ минимизировать $\bar{\beta}(\partial)$. Для этого нужно положить $c_0=1$ и все $c^+(h_j)=0$, кроме одного коэффициента, соответствующего некоторой конкретной «высоте» h^*. Тогда $\alpha=c^+(h^*)n\bar{q}_0(h^*)$, откуда

$c^+(h^*)=\alpha/n\bar{q}_0(h^*)$ и $\bar{\beta}(\partial)=1-c^+(h^*)n\underline{q}_1(h^*)=1-\alpha\underline{q}_1(h^*)/\bar{q}_0(h^*)$, а оптимальным правилом будет

$$\partial^0_{\mathbf{y}}=\left[1-\frac{\alpha}{\bar{q}_0(h^*)}\,\frac{n(h^*)}{n}\right]^+.$$

Высота h^* выбирается исходя из минимизации $\bar{\beta}(\partial)$, или же максимизации отношения: $\max\limits_h \underline{q}_1(h)/\bar{q}_0(h)$. Причем если это отношение меньше 1, то $c_0=1-\alpha$ и все $c^+(h_j)\equiv 0$, так что оптимальным становится тривиальное правило: $\partial_{\mathbf{y}}\equiv 1-\alpha$.

Рассмотрим функционирование этого правила. Если $n(h^*)=0$, т. е. ни одного превышения h^* не было, то $\partial^0_{\mathbf{y}}=1$ и принимается нулевая гипотеза. При увеличении числа $n(h^*)$ превышений h^* возникают сомнения в справедливости нулевой гипотезы, проявляемые в виде пропорционального уменьшения $\partial^0_{\mathbf{y}}$. Наконец, при $n(h^*)/n\geqslant\bar{q}_0(h^*)/\alpha$ принимается альтернатива, так как $\partial^0_{\mathbf{y}}=0$. Чем меньше α, тем больше $\bar{q}_0(h^*)/\alpha$, т. е. тем больше должна быть относительная частота превышений $n(h^*)/n$, при которой уверенно принимается альтернатива.

Если перейти от дискретной выборки y_i к реализации y_t, $0\leqslant t\leqslant T$, то при первичных вероятностях $\bar{P}_0(|y_t|>h)=\bar{q}_0(h)$, $\underline{P}_1(|y_t|>h)=\underline{q}_1(h)$ оптимальным будет правило, получаемое по аналогии с предыдущими и записываемое

$$\partial^0_{\mathbf{y}}=\left[1-\frac{\alpha}{\bar{q}_0(h^*)}\,\frac{\tau(h^*)}{T}\right]^+,$$

где $\tau(h^*)$ — суммарная длительность времени превышения реализацией y_t высоты h^*.

7.2. КОРРЕЛЯЦИОННАЯ ТЕОРИЯ ПРОВЕРКИ ГИПОТЕЗ

Получение оптимального правила при заданной средней мощности наблюдений. Здесь считается, что единственное, чем заданы и в чем отличается гипотеза от альтернативы, так это сдвиг и изменение средней мощности наблюдений.

Пусть $\mathbf{y}=(y_1, \ldots, y_n)$ и гипотеза задана следующим единственным первичным средним и первичным признаком:

$$\mathscr{M}^{\mathbf{y}}_0:\bar{M}_0\,\widehat{y^2}=\bar{\sigma}^2_0,\ \widehat{y^2}=\sum y_i^2/n,$$

т. е. задана верхняя граница усредненной мощности наблюдений. При $\varkappa=1$ согласно части **2а** теоремы 7.1 независимо от альтернативы оптимальным уровня α будет одно из правил класса

$$\partial^0_{\mathbf{y}}=[c_0-c\,\widehat{y^2}]^+,\ 1-c_0+c\,\bar{\sigma}^2_0=\alpha,\ c\geqslant 0,\ 0\leqslant c_0\leqslant 1.$$

Уровень α фиксируется указанной линейной связью коэффициентов c_0 и c, оптимальные значения которых могут искаться, если лишь задаться конкретной альтернативой. Ее и введем.

Пусть альтернатива $\mathcal{M}^{\mathbf{y}}_1$ также задана всего одним первичным средним

$$\mathcal{M}^{\mathbf{y}}_1 : \overline{M}_1 \ \Sigma\ (y_i - x_1)^2/n = \overline{\sigma}_1^2 .$$

Здесь x_1 — сдвиг выборки, соответствующий альтернативе, в дальнейшем для определенности считаемый неотрицательным $x_1 \geqslant 0$.

Вид правила задают коэффициенты c и c_0. Фиксируя их, находим ошибку второго рода $\overline{\beta}(c, c_0)$. Это не так просто, как первого, а требует привлечения формулы продолжения, по которой $\partial^0_{\mathbf{y}}$ мажорируется вторичными признаками альтернативы:

$$\overline{\beta}(c, c_0) = \overline{M}_1 [c_0 - c\,\widehat{y^2}]^+ = \inf \{(d_0 + d\,\overline{\sigma}_1^2) : d_0 + d \Sigma (y_i - x_1)^2/n \geqslant \geqslant c_0 - c\,\widehat{y^2}\}.$$

Будем ее искать. Здесь все коэффициенты должны быть неотрицательными. Неравенство под знаком инфимума переписывается так:

$$\left(d_0 - c_0 + dx_1^2\right) + (d + c)\,\widehat{y^2} - 2dx_1\,\hat{y} + (d + c)\,\hat{\sigma}^2_{\mathbf{y}} \geqslant 0,$$

где $\hat{y} = \Sigma y_i/n$, $\hat{\sigma}^2_{\mathbf{y}} = \Sigma (y_i - \hat{y})^2/n$. Это неравенство должно выполняться при всех $\mathbf{y}$. Последнее слагаемое левой части неотрицательно в силу неотрицательности d и c. Так как значения $\hat{\sigma}^2_{\mathbf{y}}$ никак не связаны с $\hat{y}$, то неотрицательной должна быть оставшаяся сумма, полученная, если положить $\hat{\sigma}^2_{\mathbf{y}} = 0$. Эта сумма образует квадратный трехчлен относительно $\hat{y}$ и условием неотрицательности будет неотрицательность дискриминанта: $(d+c)(d_0 - c_0 + dx^2_1) - (dx_1)^2 = (d+c)(d_0 - c_0) + cdx^2_1 \geqslant 0$, что с учетом $d \geqslant 0$ записывается $1/d \leqslant [-1/c + x^2_1/(c_0 - d_0)]^+$ и заменяет неравенство под знаком inf.

Теперь для вычисления ошибки $\overline{\beta}(c, c_0)$ требуется при условии выполнения последнего неравенства минимизировать $d_0 + d\overline{\sigma}^2_1$. Сначала минимизируем по коэффициенту d. Чем меньше d, тем меньше $d_0 + d\overline{\sigma}^2_1$, поэтому, заменяя последнее неравенство для $1/d$ равенством и подставляя найденное отсюда d в $d_0 + d\overline{\sigma}^2_1$, приходим к выражению для ошибки второго рода, в котором произведем сразу минимизацию по d_0:

$$\overline{\beta}(c, c_0) = \min_{d_0} \left(d_0 + \frac{\overline{\sigma}_1^2}{\left[\dfrac{x_1^2}{c_0 - d_0} - \dfrac{1}{c}\right]^+} \right) =$$

$$= \begin{cases} c_0 & \text{при } x_1 \leqslant \overline{\sigma}_1, \\ c_0 - c(x_1 - \overline{\sigma}_1)^2 & \text{при } x_1 - \dfrac{c_0}{cx_1} \leqslant \overline{\sigma}_1 \leqslant x_1, \\ \dfrac{\overline{\sigma}_1^2}{\left(\dfrac{x_1^2}{c_0} - \dfrac{1}{c}\right)} & \text{при } \overline{\sigma}_1 \leqslant x_1 - \dfrac{c_0}{cx_1}. \end{cases}$$

В верхнем случае минимум достигается при $d_0 = c_0$, в среднем — при $d_0 = c_0 - x_1(x_1 - \overline{\sigma}_1)c$, и наконец, в нижнем — при $d_0 = 0$.

Приступим теперь к поиску оптимального правила $\partial^0_{\mathbf{y}}$. Для этого нужно найти параметры c, c_0, минимизирующие $\beta(c, c_0)$ при условии $1 - c_0 + c\overline{\sigma}^2_0 = \alpha$. Очевидно, при $x_1 \leqslant \overline{\sigma}_1$ минимум достигается при $c = 0$, $1 - c_0 = \alpha$ и равен $\overline{\beta}(\partial) = 1 - \alpha$. Оптимальным здесь является тривиальное правило $\partial^0_{\mathbf{y}} \equiv 1 - \alpha$, согласно которому независимо от наблюдений с предпочтением (вероятностью) $1 - \alpha$ принимается нулевая гипотеза. К такому же тривиальному правилу приходим при $x_1 \leqslant \overline{\sigma}_0 + \overline{\sigma}_1$. Запишем полученный результат:

$$\partial^0_{\mathbf{y}} \equiv 1 - \alpha, \quad \overline{\beta}(\partial) = 1 - \alpha \quad \text{при} \quad x_1 \leqslant \overline{\sigma}_0 + \overline{\sigma}_1.$$

Пусть $x_1 > \overline{\sigma}_0 + \overline{\sigma}_1$. Тогда $\overline{\beta}(c, c_0) = b_1(c, c_0) = c_0 - c(x_1 - \overline{\sigma}_1)^2$ при $c_0 \geqslant cx_1(x_1 - \overline{\sigma}_1)$ и $\overline{\beta}(c, c_0) = b_2(c, c_0) = \overline{\sigma}^2_1/(x^2_1/c_0 - 1/c)$ при противоположном неравенстве. Таким образом, ошибка равна $\overline{\beta}(c, c_0) = b_1(c, c_0)$, если на плоскости значений коэффициентов точка (c, c_0) лежит выше прямой: *а)* $c_0 = cx_1(x_1 - \overline{\sigma}_1)$, и равна $b_2(c, c_0)$, — если ниже. На самой прямой верно неравенство $b_1(c_1, c_0) = b_2(c, c_0)$. Ограничение на уровень правила в свою очередь соответствует прямой: *б)* $c_0 = c\overline{\sigma}^2_1 + 1 - \alpha$. На этой прямой и нужно искать оптимальные c и c_0, минимизирующие $\overline{\beta}(c, c_0)$.

Здесь нужно рассмотреть два случая. Первый, когда прямая *б* лежит выше прямой *а* в области ограничений параметров $0 \leqslant c_0 \leqslant 1$, $c \geqslant 0$. Это эквивалентно неравенству $x_1 < \left(\overline{\sigma}_1 + \sqrt{\overline{\sigma}^2_1 + 4\overline{\sigma}^2_0/\alpha}\right)/2 = x_*$, где x_* обозначает правую часть ($x_* > \overline{\sigma}_0 + \overline{\sigma}_1$ при $\overline{\sigma}_1/\overline{\sigma}_0 + 1 < 1/\alpha$, и тогда интервал $\overline{\sigma}_0 + \overline{\sigma}_1 \leqslant x_1 \leqslant x_*$ будет непустым). В этом случае $\overline{\beta}(c, c_0) = b_1(c, c_0)$ и легко найти минимизирующие значения: $c_0 = 1$, $c = \alpha/\overline{\sigma}^2_0$, откуда оптимальным будет правило:

$$\partial^0_{\mathbf{y}} = [1 - \alpha\,\hat{y}^2/\overline{\sigma}^2_0]^+ \quad \text{при} \quad \overline{\sigma}_0 + \overline{\sigma}_1 \leqslant x_1 \leqslant x_*. \qquad (*)$$

Для него $\overline{\beta}(\partial) = 1 - \alpha(x_1 - \overline{\sigma}_1)^2/\overline{\sigma}^2_0$.

Пусть теперь прямые вида *а)* и *б)* пересекаются, т. е. $x_1 > x_*$. Тогда минимум, как нетрудно убедиться, достигается на прямой *б* правее ее точки пересечения с *а*, т. е. при $\overline{\beta}(\partial) = b_2(c, c_0)$. Подставляя в выражение для $b_2(c, c_0)$ уравнение прямой *б*, дифференцируя по c и приравнивая 0, находим точку минимума $c^* = (1 - \alpha)/(\overline{\sigma}_0(x_1 - \overline{\sigma}_0))$, $c^*_0 = (1 - \alpha)x_1/(x_1 - \overline{\sigma}_0)$, что и определит оптимальное при $c^* < \alpha/\overline{\sigma}^2_0$ (эквивалентно $0 \leqslant c^*_0 \leqslant 1$) или, что то же самое, при $x_1 > \sigma_0/\alpha$:

$$\partial^0_{\mathbf{y}} = \frac{1 - \alpha}{1 - (\overline{\sigma}_0/x_1)} \left[1 - \frac{\widehat{y^2}}{\overline{\sigma}_0 x_1}\right]^+$$

при $x_1 \geqslant \overline{\sigma}_0/\alpha$;

$$\overline{\beta}(\partial) = (1 - \alpha)\overline{\sigma}^2_1/(x_1 - \overline{\sigma}_0)^2.$$

Если же $x_* \leqslant x \leqslant \overline{\sigma}_0/\alpha$, то оптимальными будут $c^* = \overline{\sigma}/\alpha^2_0$, $c^*_0 = 1$ и, следовательно, правило (*), но с ошибкой, вычисляемой подстановкой оптимальных значений в выражение для $b_2(c, c_0)$, что дает

$$\partial_y^0 = [1 - \alpha\widehat{y^2}/\overline{\sigma}_0^2]^+ \quad \text{при} \quad x_* \leqslant x_1 \leqslant \overline{\sigma}_0/\alpha;$$

$$\beta(\partial) = \frac{\overline{\sigma}_1^2}{x_1^2 - \overline{\sigma}_0^2/\alpha}.$$

Подождем пока формулировать доказанный результат, а сделаем это чуть ниже (формулы (7.1), (7.2)) в более общем виде, так как он оказывается пригодным для многих случаев. Для унификации введем такие обозначения: $Y^2 = \widehat{y^2}$, $\overline{M}_0 Y^2 = \overline{\sigma}^2_0$, $q = \overline{\sigma}_1/\overline{\sigma}_0$ — *коэффициент изменения мощностей*, $\rho = x_1/\overline{\sigma}_0$ — *нормированный сдвиг* (аналог отношения сигнал-шум).

Оптимальные решающие правила и их свойства не зависят от объема выборки, и вообще от типа реализаций, вектор ли это или функция времени. Так, для процессов с непрерывным временем x_t, $0 \leqslant t \leqslant T$, определенных следующими гипотезой ($j=0$) и альтернативой ($j=1$): $M_j y_t = w_{jt}$, $\overline{M}_j \int_0^T (y_t - w_{jt})^2 dt/T = \overline{\sigma}^2_j$, оптимальные правила сохраняют тот же вид и ту же ошибку при обозначениях: $Y^2 = \int_0^T (y_t - w_{0t})^2 dt/T$, $x^2_1 = \int_0^T (w_{1t} - w_{0t})^2 dt/T$. Сказанное здесь и доказанное выше ведут к следующему.

Общая форма правила. *При известных только* $\overline{M}_0 Y^2 = b_0$, $\overline{M}_1 (Y - x_1)^2 = b_1$ *и при* $\rho^2 = x^2_1/b_0$, $q^2 = b_1/b_0$ *оптимальным уровня* α *в зависимости от величины нормированного сдвига* ρ *будет одно из следующих правил*:

$$\partial_y^0 = \begin{cases} 1 - \alpha & \text{при} \quad \rho < 1 + q, \\ [1 - \alpha Y^2/\overline{M}_0 Y^2]^+ & \text{при} \quad 1 + q \leqslant \rho \leqslant 1/\alpha, \\ \dfrac{(1-\alpha)}{\rho - 1} [\rho - Y^2/\overline{M}_0 Y^2]^+ & \text{при} \quad \rho > 1/\alpha. \end{cases} \tag{7.1}$$

Ошибка второго рода определяется выражениями:

$$\overline{\beta} = \begin{cases} 1 - \alpha & \text{при} \quad \rho < 1 + q, \\ 1 - \alpha(\rho - q)^2 & \text{при} \quad 1 + q \leqslant \rho \leqslant \rho_* = (q + \sqrt{q^2 + 4/\alpha})/2, \\ q^2/(\rho^2 - 1/\alpha) & \text{при} \quad \rho_* < \rho \leqslant 1/\alpha, \\ (1-\alpha)\, q^2/(\rho - 1)^2 & \text{при} \quad \rho > 1/\alpha. \end{cases} \tag{7.2}$$

Различные виды решающих правил (7.1) представлены сплошными линиями на рис. 7.2, а соответствующая им кривая ошибок второго рода как функция ρ — на рис. 7.3. Обращает на себя внимание параболический вид правил, унаследованный ими от первичного признака гипотезы, здесь единственного. Другая характерная особенность — расплывчатость оптимальных правил, в корнях своих обязанная исключительной бедности исходных данных о гипотезах. Это же есть причина низких качественных показателей (большой вероятности ошибки $\overline{\beta}$), объяснимых так-

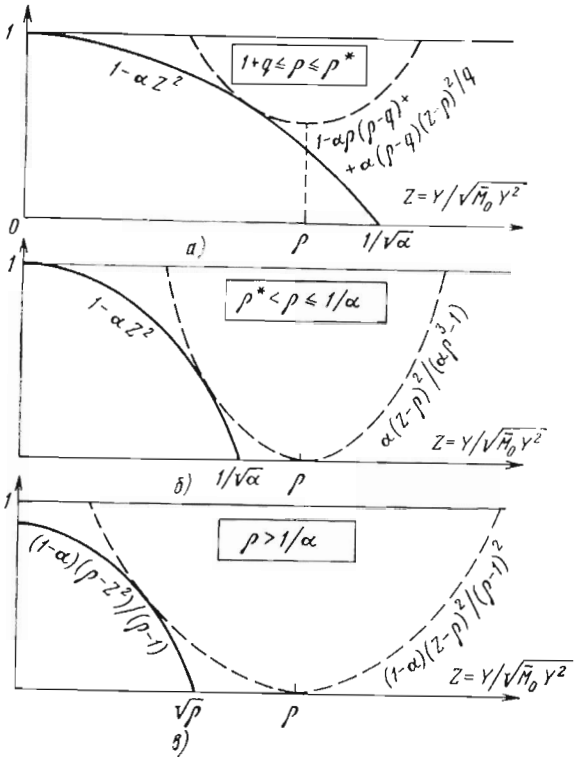


Рис. 7.2: *а*) слабый сигнал, *б*) средний, *в*) сильный

же тем, что пессимизм ($\varkappa = 1$) заставляет ориентировать расчеты на наименее благоприятный случай (в рамках исходных первичных данных), а таковой здесь зовет к постоянству реализаций $y_1 = y_2 = \ldots = y_n = Y$, т. е. сводит все к одному единственному наблюдению Y.

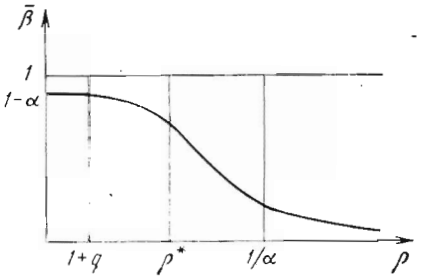


Рис. 7.3. График ошибки второго рода

Итак, правило (7.1) является оптимальным для задачи, когда наблюдение всего одно (или к нему все сводится) Y, а гипотеза задана одним первичным значением $\overline{M}_0 Y^2 = \overline{\sigma}^2_0$, как и альтернатива: $\overline{M}_1 (Y - x_1)^2 = \overline{\sigma}^2_1$. Это — суть мощности, измеренные при гипотезе без сдвига наблюдений, а при альтернативе — со сдвигом x_1.

Введем новые гипотезу и альтернативу:

$$\mathcal{M}_0 : M_0 Y = 0, \ \overline{M}_0 Y^2 = \overline{\sigma}^2_0;$$

$$\mathcal{M}_1 : M_1 Y = x_1, \ \overline{M}_1 (Y - x_1)^2 = \overline{\sigma}^2_1.$$

Здесь, в отличие от предыдущего, среднее считается известным, является точным и при альтернативе сдвигается на x_1, а дисперсия (она при известном среднем заменяет мощность) изменяется от $\bar{\sigma}^2_0$ до $\bar{\sigma}^2_1$. Несмотря на то, что новая гипотеза и альтернатива более узкие по сравнению с предыдущими, можно показать, что для них оптимальными уровня α будут точно те же решающие правила (7.1) с их ошибками (7.2).

Доказательство этого факта подменяется иллюстрацией рис. 7.2. Здесь представлены в зависимости от величины нормированного сдвига $\rho = x_1/\bar{\sigma}_0$ картины для оптимальных правил сплошной линией, а штриховой — для тех вторичных признаков альтернативы, с которых средние $\bar{M}_1$ переносятся на ∂^0_y, давая ошибку второго рода. Из рисунка выясняется, что как сами оптимальные правила, основанные на признаках вида $c_0 - cY^2$, так и вторичные признаки альтернативы, здесь имеющие вид $d_0 + d(Y - x_1)^2$, базируются каждая лишь на одном признаке, а именно том, который соответствует дисперсиям, и не используют совсем признака среднего Y. Это и объясняет, почему знание средних не меняет решения статистической задачи.

Сказанное дополняется следующим.

1. Знание нижних дисперсий $\underline{\sigma}^2_j$, $j = 0, 1$, или даже точных значений $\sigma_j = \bar{\sigma}_j = \sigma_j$ не меняет решения статистической задачи.

2. Штриховые кривые на рис. 7.2 дают другой возможный вариант решающих правил, соответствующих части **2б** теоремы 7.1 и ориентированных на первичные признаки альтернативы. Например, для случая последней строчки формулы (7.1) (рис. 7.2,*в*) это будет следующее контрастное правило:

$$\partial^0_y = 1 - \left[1 - \frac{1-\alpha}{(\rho-1)^2}\,(Y - x_1)^2/\bar{M}_0\, Y^2\right]^+ \text{ при } \rho > 1/\alpha.$$

Вообще любое правило, располагающееся между сплошной и штриховой линиями рис. 7.2, будет оптимальным. Можно выбрать его, например, в классе линейно-ломаных функций или какомнибудь другом сообразно дополнительным требованиям (например, разумно считать в только что выписанном правиле $\partial^0_y = 0$ при $Y > x_1$). Теперь становится понятной неконтрастность правила на рис. 7.2*в* как результат ограничения формы ∂^0_y классом парабол; в другом классе правил оно вполне уже может стать контрастным.

3. Детерминированные правила принимают только два значения: 0 или 1 и изображаются на рис. 7.2 прямоугольниками единичной высоты. Как видно из рисунков, прямоугольники нельзя расположить между сплошной и штриховой линиями, следовательно, детерминированные правила не могут быть оптимальными и переход к ним сопровождается ростом ошибок. Чтобы выдержать уровень α, нужно положить $\partial^0_y = 1$ при $Y < \bar{\sigma}_0/\sqrt{\alpha}$, а иначе приравнять правило 0, тогда основание прямоугольника на рис.

7.2,*а, б* должно совпасть с основанием положительной части параболы. Расчет для этого правила ошибки второго рода производится буквально как расчет вероятности события в примере 1.5 и зависит уже от того, известно ли среднее $M_1 Y = x_1$ или нет. Если известно, то $\bar{\beta} = [1 + (\rho - 1/\sqrt{\alpha})^2]^{-1}$, где $\rho > 1/\sqrt{\alpha}$, а если нет, то $\bar{\beta} = (\rho - 1/\sqrt{\alpha})^{-2}$, где $\rho \geqslant 1 + 1/\sqrt{\alpha}$, при других ρ ошибка равна 1.

Проверка гипотез по заданным корреляциям. Пусть $\mathbf{y} = (y_1, \ldots, y_n)^\tau$ и точно известными являются матрицы корреляций:

$$\mathcal{M}_0 : M_0\, y_i\, y_j = B_{ij},$$

$$\mathcal{M}_1 : M_1\,(y_i - w_i)\,(y_j - w_j) = K_{ij}.$$

В векторных обозначениях $\mathbf{w} = (w_1, \ldots, w_n)^\tau$ — сдвиг, а $\mathbf{B}$ и $\mathbf{K}$ — положительно определенные симметричные матрицы. Достаточный класс правил здесь записывается в виде $[c_0 - \sum\sum c_{ij} y_i y_j]^+$; выбором коэффициентов c_0 и c_{ij} нужно искать оптимальное из них.

Достаточным (см. § 7.6) здесь будет подкласс правил, основанных на линейных преобразованиях наблюдений: $\mathbf{g}^\tau \mathbf{y} = Y$, сводящих вектор $\mathbf{y}$ к одному значению Y при искомом (на заключительном этапе синтеза) векторе $\mathbf{g}$. Теперь для Y как нового наблюдения гипотеза и альтернатива будут определяться значениями $M_0 Y^2 = \mathbf{g}^\tau \mathbf{B} \mathbf{g}$, $M_1 (Y - x_1)^2 = \mathbf{g}^\tau \mathbf{K} \mathbf{g}$, где $x_1 = \mathbf{w}^\tau \mathbf{g}$. Таким образом, задача сводится к уже рассмотренной в этом параграфе: при заданном $\mathbf{g}$ оптимальным будет правило (7.1), (7.2) с подстановкой туда $Y = \mathbf{g}^\tau \mathbf{y}$, $M_0 Y^2 = \mathbf{g}^\tau \mathbf{B} \mathbf{g}$, $q^2 = \mathbf{g}^\tau \mathbf{K} \mathbf{g} / \mathbf{g}^\tau \mathbf{B} \mathbf{g}$, $\rho^2 = (\mathbf{w}^\tau \mathbf{g})^2 / \mathbf{g}^\tau \mathbf{B} \mathbf{g}$.

Перечисленные параметры зависят от $\mathbf{g}$, отсюда вид правила и его ошибка второго рода $\bar{\beta}_{\mathbf{g}}$ будут определяться выбранным $\mathbf{g}$. Осталось найти такое $\mathbf{g}^*$, которое минимизировало бы эту ошибку:

$$\mathbf{g}^* : \min_{\mathbf{g}} \bar{\beta}_{\mathbf{g}},$$

что и даст нам наилучшее правило.

Сложность выражения (7.2) не позволяет найти общий вид для $\mathbf{g}^*$. Пойдем по пути дальнейших упрощений, выбирая $\mathbf{g}$ максимизацией нормированного сдвига ρ

$$\mathbf{g}^{**} : \max (\mathbf{w}^\tau \mathbf{g})^2 / \mathbf{g}^\tau \mathbf{B} \mathbf{g}.$$

Основанием упрощений служит то, что ошибка $\bar{\beta}$, как это можно видеть из (7.2), монотонно убывает при увеличении ρ при каждом заданном q. Решение $\mathbf{g}^{**}$ последней задачи имеет вполне конкретный вид: $\mathbf{g}^{**} = \mathbf{B}^{-1}\mathbf{w}$.

При одинаковых корреляциях $\mathbf{K} = \mathbf{B}$ параметр $q = 1$ и найденный нами вектор будет строго минимизировать вероятность ошибки: $\mathbf{g}^* = \mathbf{g}^{**} = \mathbf{B}^{-1}\mathbf{w}$. Подстановкой этого вектора находятся

$$Y = \mathbf{w}^\tau \mathbf{B}^{-1} \mathbf{y}, \quad \rho^2 = M_0 Y^2 = \mathbf{w}^\tau \mathbf{B}^{-1} \mathbf{w},$$

и вместе с формулой (7.1) они приведут к оптимальному при одинаковых корреляциях правилу уровня α:

$$\partial_y^0 = \begin{cases} 1-\alpha & \text{при} \quad \rho < 2, \\ [1-\alpha\,(\mathbf{w}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{y})^2/\rho^2]^+ & \text{при} \quad 2 \leqslant \rho \leqslant 1/\alpha, \\ \dfrac{1-\alpha}{\rho-1}\left[\rho - \dfrac{(\mathbf{w}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{y})^2}{\rho^2}\right]^+ & \text{при} \quad \rho > 1/\alpha. \end{cases} \tag{7.3}$$

Дополнения и замечания. 1. Наличие данных $M_0\mathbf{y}=0$, $M_1\mathbf{y}=\mathbf{w}$ о средних значениях наблюдений не меняет решения задачи.

2. Случай одинаковых $\mathbf{K}=\mathbf{B}$ эквивалентен обнаружению сигнала $\mathbf{w}$, когда при его отсутствии действует только шум $\mathbf{y}=\xi$ с корреляционной матрицей $\mathbf{B}$, а при наличии — смесь детерминированного сигнала с шумом $\mathbf{y}=\mathbf{w}+\xi$.

3. Переход к процессам $y_t=w_t+\xi_t$, $0\leqslant t\leqslant T$, при известной $B(t,\tau)$ — корреляционной функции шума ξ_t сводит матричные выражения к операторным:

$$Y=\int w_t\,B^{-1}(t,\tau)\,y_t\,dt,\ \rho^2=\int\int w_t B^{-1}(t,\tau)\,w_\tau\,dt d\tau,$$

где $B^{-1}(t,\tau)$ — ядро обратного к корреляционному оператора.

При однородном шуме, т. е. $B(t,\tau)=B(t-\tau)$, обозначая $\dot B(\omega)=2\int_0^\infty B(\tau)\cos(\omega\tau)\,d\tau$ — энергетический спектр, а $\dot y_\omega$ и $\dot w_\omega$ — спектральные двойники (преобразования Фурье) y_t и w_t, и считая T достаточно большим, приходим к следующим значениям составляющих оптимального правила (7.1):

$$Y=\left|\int_{-\infty}^{\infty}\frac{\dot y_\omega\,\dot w_\omega}{\dot B(\omega)}\,d\omega\right|,\ \rho^2=\int_{-\infty}^{\infty}\frac{|\dot w_\omega|^2}{\dot B(\omega)}\,d\omega.$$

Знаменатель $\dot B(\omega)$ усиливает те участки спектра, где шум мал, и подавляет зашумленные.

Проследим, как повлияет на обнаружение присутствие гармонических помех на частотах ω_j. Их действие проявляется в возникновении дельта-выбросов в спектре: $\dot B(\omega)=\dot B_0(\omega)+\sum b_j\delta(\omega-\omega_j)$, где $\delta(\omega)$ — дельта-функция Дирака, равная 0 при $\omega\neq 0$ и ∞ при $\omega=0$. Подстановка $\dot B(\omega)$ в знаменатель формулы для Y убедит нас в том, что частоты гармоник будут режектироваться при преобразовании y_t в Y.

4. При некоррелированных наблюдениях: $B_{ii}=\sigma^2_{0i}$, $K_{ii}=\sigma^2_{1i}$, $B_{ij}=K_{ij}=0$ при $i\neq j$, имеем $g^*_i=w_i/\sigma^2_{0i}$ и правило проверки гипотез будет основываться на статистике $Y=\sum y_i w_i/\sigma^2_{0i}$ и даваться выражением (7.1), в котором $\rho^2=M_0Y^2=\sum w^2_i/\sigma^2_{0i}$, $q^2=(\sum w^2_i/\sigma^2_{1i})/\sum(w^2_i/\sigma^2_{0i})$. В частном случае постоянных значений $w_i=x_1$, $\sigma_{0i}=\sigma_0$, $\sigma_{1i}=\sigma_1$ имеем $g^*_i=\mathrm{const}$ и приходим (полагая $g^*_i=1/n$) к оптимальному правилу для однородной некоррелированной выборки, имеющему вид (7.1), (7.2) при $Y=\hat y$, $\bar M_0Y^2=\sigma^2_0/n$, $\rho=x_1\sqrt{n}/\sigma_0$, $q=\sigma_1/\sigma_0$.

27

5. Показательно, что при нулевом сдвиге $x_1=0$ оптимальным будет лишь тривиальное правило $\partial^0_y=1-\alpha$. Причина кроется в крайне слабых исходных допущениях, которые обязательно нужно дополнить знаниями четвертых моментов наблюдений, чтобы получить приемлемое правило проверки гипотез, различающее изменение одних только дисперсий (мощности шума).

Пусть стоит задача:

$$\mathcal{M}_k: M_k\,y_i^2=\sigma_k^2,\ M_k\,y_i^2\,y_j^2=\sigma_k^4 \text{ при } i\neq j;\ M_k\,y_i^4=m_{k4};\ k=0,1;\ i,j=1,\ldots,n.$$

Переходя к преобразованным наблюдениям $z_i=y_i^2-\sigma_0^2$, получаем:

$$M_0\,z_i\,z_j=M_1(z_i-x)(z_j-x)=0,\ i\neq j;\ M_0\,z_i^2=M_0(y_i^2-\sigma_0^2)^2=m_{04}-\sigma_0^4,\ M_1(z_1-x)^2=m_{14}-\sigma_1^4,\quad \text{где } x=\sigma_1^2-\sigma_0^2.$$

Задача в новых наблюдениях тождественна рассмотренной в конце предыдущего пункта, а оптимальным будет правило (7.1), (7.2) в обозначениях:

$$Y=\bar y^2-\sigma_0^2,\ M_0Y^2=(m_{04}-\sigma_0^4)/n,\ \rho^2=(\sigma_1^2-\sigma_0^2)n/(m_{04}-\sigma_0^4),$$
$$q^2=(m_{14}-\sigma_1^4)/(m_{04}-\sigma_0^4).$$

Неточные корреляции. Пусть корреляции не являются точными и заданы не все, а частично. Скажем, не все элементы B_{ij} и K_{ij} корреляционных матриц заданы, а часть, да еще и не точно, а их некоторые границы. Такая задача может решаться напрямую нашими методами. Собственно, одно такое решение было уже найдено в начале этого параграфа, когда известными были верхние границы сумм диагональных элементов корреляционных матриц: $\bar M_0y^2=\sum\bar B_{ii}/n=\bar\sigma^2_0$, $\bar M_1(y-x_1)^2=\sum\bar K_{ii}/n=\bar\sigma^2_1$.

Рассмотрим еще один случай. Пусть известны только «потолки» для мощности отдельных элементов наблюдений: $\bar M_0y_i^2=\bar B_{ii}$ и $\bar M_1(y_i-w_i)^2=\bar K_{ii}$, $i=1,\ldots,n$, а какая корреляция между ними — совсем не известно как при гипотезе, так и альтернативе. Сказанное не исключает, что наблюдения в наименее благоприятном случае ($\varkappa=1$) могут повторять друг друга. Теперь будет понятен и смысл оптимального правила, найденного решением этой задачи нашими методами. Оно имеет вид (7.1), в котором Y формируется как отбор только одного элемента $Y=y_l$ при таком $l=i$, при котором минимально значение $\bar\beta_i$, найденное по формуле (7.2) с подстановкой туда данных i-го элемента: $q_i=\bar K_{ii}/\bar B_{ii}$, $\rho^2_i=w_i^2/\bar B_{ii}$. Итак, вначале отбирается один наиболее «информативный» элемент y_l, который в сравнении с каждым другим обеспечит меньшую ошибку $\bar\beta_i$ при заданном α. Затем по нему строится правило

Подытожим оба частных случая. В первом осуществлялось квадратичное преобразование $\widehat{y^2}$ наблюдений, во втором его можно считать линейным, полагая $g_i=0$ для всех $i\neq l$, кроме одного $g_l=1$. А в общем, по теореме 7.1 о достаточности (и $\varkappa=1$) это будут преобразования вида $\sum\sum c_{ij}y_iy_j$ при суммировании по тем (i, j), для которых имеются данные о B_{ij}, остальные $c_{ij}=0$. Если так окажется, что оптимальные $c^*_{ij}=g_ig_j$ и $\sum\sum c^*_{ij}y_iy_j=(\sum g_iy_i)^2$, то преобразование сведется к линейному. В общем, это будет не так.

7.3. ИСПОЛЬЗОВАНИЕ ДОВЕРИТЕЛЬНЫХ ОЦЕНОК ДЛЯ ПРОВЕРКИ ГИПОТЕЗ

Описание способа. Нередко задача проверки гипотез сводится к принятию решений относительно значений параметра. Так, на отсутствие сигнала можно смотреть как на нулевое значение его амплитуды, оценку надежности аппаратуры можно перевести в проверку соответствия норме параметра надежности и т. д. Наша цель — связать правила проверки гипотез о параметре с доверительной оценкой этого параметра, чтобы затем использовать материалы предыдущей главы.

Пусть x есть числовой параметр и $\mathcal{M}^y{}_x$ — переходная ИМ, задающая модель на $\mathcal{Y}$ при фиксированных x. Пусть гипотеза и альтернатива соответствуют значениям x_0 и x_1 параметра x. Тогда $\mathcal{M}^y{}_0=\mathcal{M}^y{}_{x_0}$, $\mathcal{M}^y{}_1=\mathcal{M}^y{}_{x_1}$ — две гипотезы. В этом разделе считается $\varkappa=1$.

Возьмем СИМ $\mathcal{M}^{xy}=\mathcal{J}^x\mathcal{M}^y{}_x$ и пусть $\partial^\alpha{}_y(x)$ есть доверительная оценка уровня α параметра x: $\alpha=\overline{M}^{xy}(1-\partial^\alpha{}_y(x))=\sup\limits_x \overline{M}^y{}_x(1-\partial^\alpha{}_y(x))$. По доверительной оценке построим следующее правило проверки гипотез:

$$\partial^0{}_y=\partial^\alpha{}_y(x_0).$$

Принимается нулевая гипотеза с той же достоверностью, с какой значение x_0 включается в оценку $\partial^\alpha{}_y(x)$. Ошибка первого рода этого правила не больше уровня α оценки, так как

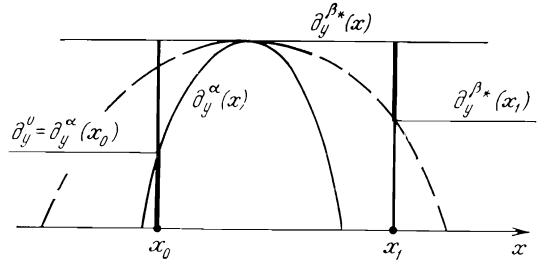
$$\overline{\alpha}(\partial)=\overline{M}_0(1-\partial^0_y)=\overline{M}^y_{x_0}(1-\partial^\alpha_y(x_0))\leqslant\sup_x\overline{M}^y_x(1-\partial^\alpha_y(x))=\alpha.$$

Следующее утверждение дает способ приближенного расчета ошибки второго рода правила (7.3).

Теорема 7.4. *Пусть $\partial^\beta{}_y(x)$ есть доверительная (при $\varkappa=1$) оценка уровня β параметра x и пусть β_* есть наименьший уровень β, при котором выполняется неравенство $\partial^\alpha{}_y(x_0)+\partial^\beta{}_y(x_1)\leqslant 1$. Тогда ошибка $\overline{\beta}(\partial)$ правила $\partial^0{}_y=\partial^\alpha{}_y(x_0)$ проверки гипотезы $x=x_0$ при альтернативе $x=x_1$ будет не больше значения β_*.*

Действительно, $\overline{\beta}(\partial)=\overline{M}_1\partial^0=\overline{M}^y_{x_1}\partial^0_y\leqslant\overline{M}^y_{x_1}[1-\partial^{\beta^*}_y(x_1)]\leqslant\sup\limits_x\overline{M}^y_x[1-\partial^{\beta^*}_y(x)]=\beta_*$.

Рис. 7.4. Нахождение прикидочной ошибки



Значение β_*, в общем, может отличаться в большую сторону от величины ошибки $\overline{\beta}(\partial)$, а потому называется *прикидочным* значением ошибки второго рода. Оно будет тем более точно приближать ошибку, чем меньше границы $\underline{M}^y{}_x\partial_y(x)$ и $\overline{M}^y{}_x\partial_y(x)$ зависят от x.

Дадим наглядное пояснение к теореме (рис. 7.4). Положение оценки $\partial^\alpha{}_y(x)$ зависит от наблюдения y, а сама оценка как функция x расширяется при уменьшении α. На рисунке сплошной линией и штриховой изображены два ее вида при разных уровнях, первый из которых равен α, а уровень β_* (на рисунке $\beta_*<\alpha$) подыскивается таким, чтобы сумма высот сплошной кривой в точке $x=x_0$ и штриховой в точке $x=x_1$ равнялась 1. Величина β_* и даст прикидочную ошибку.

Так как β_* будет тем меньше, чем менее расплывчатой является оценка $\partial^\beta{}_y(x)$, то правило проверки гипотез будет тем лучше, чем меньше ширина оценки в направлении изменения x, а поэтому в качестве исходной лучше брать оптимальную оценку, или хотя бы квазиоптимальную.

Следующие разделы дают примеры построения правил и расчета ошибок по теореме 7.4.

Асимптотическое правило при симметричных ограниченных флуктуациях. Пусть $y_i=x+\xi_i$, $i=1, \ldots, n$, где параметр x равен 0 при гипотезе и равен x_1 при альтернативе, а флуктуации ξ_i независимы, имеют нулевые средние $M\xi_i=0$ и ограниченную среднюю мощность $\overline{M}\xi^2_i=\overline{\sigma}^2$. Данная задача относится к рассмотренной в § 7.2, если положить там $\overline{\sigma}_1=\overline{\sigma}_0=\overline{\sigma}$. Введем дополнительные предположения: 1) ξ_i ограничены сверху и снизу уровнем H, т. е. $\overline{P}(|\xi_i|>H)=0$, 2) ξ_i симметричны, т. е. $M\xi_i^{2h-1}=0$, $i=1, 2, \ldots$ Тогда из доверительной оценки (6.22) степенного типа будет следовать правило

$$\partial^0_y=[1-(\hat{y}/\Delta)^{2k}]^+.$$

Расчет размаха 2Δ по уровню α был произведен там же в § 6.4; асимптотически при $n\to\infty$ размах, умноженный на $\sqrt{n}$, оценивается неравенством (6.23), правую часть которого обозначим $\Delta_\alpha(k)=\overline{\sigma}[(2k)!/(\alpha k!)]^{1/2k}/\sqrt{2}$.

В соответствии с теоремой 7.4 найдем прикидочное значение β_*. Это есть минимальное число β такое, что

$$\partial_y^0 + \partial_y^\beta(x_1) = \left[1 - \left(\frac{\hat{y}\sqrt{n}}{\Delta_\alpha(k)}\right)^{2k}\right]^+ + \left[1 - \left(\frac{(\hat{y}-x_1)\sqrt{n}}{\Delta_\beta(k)}\right)^{2k}\right]^+ \leqslant 1.$$

Полученное неравенство должно выполняться при всех $\hat{y}$. При $x_1\sqrt{n} > \Delta_\alpha(k)$ левая часть его как функция переменной $\hat{y}$ имеет три максимума: при $\hat{y}=0$, при $\hat{y}=x_1$ и в средней части между точками 0 и x_1. Нас интересует последний. Он достигается при $\hat{y} = x_1[(\Delta_\beta(k)/\Delta_\alpha(k))^{2k/(2k-1)}+1]^{-1}$ и равен $2-(x_1^2 n)^k[\Delta_\alpha(k)^{2k/(2k-1)} + \Delta_\beta(k)^{2k/(2k-1)}]^{1-2k}$. Этот максимум должен быть не больше 1. Отсюда находится размах искомого $\partial^\beta{}_y(x_1)$: $\Delta_{\beta_*}(k)^{2k/(2k-1)} = [(x_1\sqrt{n})^{2k/(2k-1)} - \Delta_\alpha(k)^{2k/(2k-1)}]^+$, а из него уже — прикидочная ошибка, которая применением формулы Стирлинга запишется в явном виде при условии $x_1\sqrt{n} \geqslant \Delta_\alpha(k)$:

$$\beta_* = \sqrt{2}\,(\sigma\sqrt{2k/e})^{2k} / [(x_1\sqrt{n})^{2k/(2k-1)} - \Delta_\alpha(k)^{2k/(2k-1)}]^{2k-1}.$$

Теперь нужно было бы искать оптимальное значение k, минимизирующее при заданном уровне α, и значении $x_1\sqrt{n}$ — прикидочную ошибку второго рода β_*. Мы не будем решать эту задачу, которая строго требует численных методов, а приближенно можно ограничиться найденным в § 6.5 значением k^*.

Поставим другую цель: сравнить прикидочное значение ошибки β_* с точным $\bar{\beta}(\partial)$, полагая для простоты $k=1$. Тогда $\Delta_\alpha = \bar{\sigma}/\sqrt{\alpha}$,

$$\partial_y^0 = [1 - \alpha n \hat{y}^2/\bar{\sigma}^2]^+, \quad \Delta^2_{\beta_*} = (x_1\sqrt{n})^2 - \Delta^2_\alpha, \quad \beta_* = (\rho^2 - 1/\alpha)^{-1},$$

где $\rho = x_1\sqrt{n}/\bar{\sigma}$.

Можно видеть, что результат совпадает с (7.2) при $q=1$ и при $\rho_* \leqslant \rho \leqslant 1/\alpha$. Значит, предлагаемая теоремой 7.4 методика привела к оптимальному правилу и его вероятности ошибки, что является показателем в пользу этой методики.

Проверка гипотез по мощности флуктуаций. Снова воспользуемся изложенной методикой, позволяющей переводить оценки в правила. Пусть $y_i = x\xi_i$, $i=1, \ldots, n$, где параметр масштаба x принимает при гипотезе значение x_0, а при альтернативе — x_1; считаем для конкретности $x_1 > x_0$. Пусть флуктуации ξ_i независимы, имеют нулевые средние $M\xi_i = 0$, единичные дисперсии $M\xi_i^2 = 1$ и ограниченные четвертые моменты $M\xi_i^4 = \bar{m}_4$. Воспользовавшись оценкой (6.27), получим

$$\partial_y^0 = [1 - \alpha n\,(1 - \hat{y}^2/x_0)^2/(\bar{m}_4 - 1)]^+.$$

Это есть правило уровня α. Ищем прикидочное значение ошибки второго рода β_* как минимальное β такое, что

$$\partial_y^0 + \partial_y^\beta(x_1) = \left[1 - \frac{\alpha n}{\bar{m}_4 - 1}\,(1 - \hat{y}^2/x_0)^2\right]^+ + \left[1 - \frac{\beta n}{\bar{m}_4 - 1}\,(1 - \hat{y}^2/x_1)^2\right]^+ \leqslant 1.$$

Сумма в левой части как функция $\hat{y}^2$ при интересующих нас β имеет три максимума: при $\hat{y}^2 = x_0$ и $\hat{y}^2 = x_1$, где эта сумма равна 1, и в промежуточной точке $\hat{y}^2 = (\alpha/x_0 + \beta/x_1)/(\alpha/x_0^2 + \beta/x_1^2)$, где эта сумма не должна превышать 1, что записывается

$$2 - n\alpha\beta\,(x_1 - x_0)^2\,(\beta/x_1^2 + \alpha/x_0^2)\,(\bar{m}_4 - 1)^{-1}\,(\alpha x_1/x_0 + \beta x_0/x_1)^{-2} \leqslant 1.$$

Из этого уравнения находится β_*:

$$\beta_* = \sqrt{[\alpha(\rho-1)^2/2 + q]^2 + q[\alpha\rho(\rho-1) + q]} - \alpha(\rho-1)^2/2 - q,$$

$$\rho = x_1/x_0, \quad q = \frac{\bar{m}_4 - 1}{n}.$$

Несколько более простое выражение для прикидочного значения будет получено, если искать минимальное β такое, что $\partial_y^0\,\partial_y^\beta(x_1) \equiv 0$. Тогда $\beta_{**} = q[1 - (1 + \sqrt{q/\alpha})/\rho]^{-2}$. Очевидно, что $\beta_{**} > \beta_*$. При увеличении $\rho = x_1/x_0$ имеем: $\beta_{**} \to q = (\bar{m}_4 - 1)/n$.

7.4. СПЕЦИАЛЬНЫЕ МЕТОДЫ СИНТЕЗА ПРАВИЛ

Задана формальная плотность альтернативы по отношению к гипотезе. Пусть существует плотность альтернативы по отношению к гипотезе $p(y) = \mathscr{M}_1/\mathscr{M}_0$. По определению формальной плотности § 1.4 это соответствует тождеству $\bar{M}_1 f(y) = \bar{M}_0 p(y) f(y)$, $\forall f \in \mathscr{F}_1$, справедливому для всех f из области существования $\mathscr{M}^y{}_1$. Если гипотеза $\langle \bar{M}_0 \mathscr{G} \rangle$ задана первичными значениями $\bar{M}_0 g(y)$, $g \in \mathscr{G}$, то $\mathscr{M}_1$ будет определено значениями $\bar{M}_1 g(y)/p(y) = \bar{M}_0 g(y)$, $g \in \mathscr{G}$, и $M_1(1/p(y)) = 1$.

Выпишем весовую сумму ошибок

$$r^\varkappa_\lambda(\partial) = \lambda\alpha^\varkappa(\partial) + \beta^\varkappa(\partial) = \lambda + \varkappa\left(\bar{M}_0\,\partial_y^0\,p(y) - \lambda\,\underline{M}_0\,\partial_y^0\right) + (1-\varkappa)\left(\underline{M}_0\,\partial_y^0\,p(y) - \lambda\,\bar{M}_0\,\partial_y^0\right).$$

На основании неравенств $\bar{M}_0\partial^0{}_y p(y) - \underline{M}_0\lambda\partial^0{}_y \geqslant \bar{M}_0\partial^0{}_y(p(y)-\lambda)$, $\underline{M}_0\partial^0{}_y p(y) - \bar{M}_0\lambda\partial^0{}_y \leqslant \underline{M}_0\partial^0{}_y(p(y)-\lambda)$ и в силу непрерывности $r^\varkappa{}_\lambda(\partial)$ по $\varkappa$ всегда можно подыскать такое $\varkappa^*$, $0 \leqslant \varkappa^* \leqslant 1$, что будет верно равенство

$$r^{\varkappa^*}_\lambda(\partial) = \lambda + \varkappa^*\,\bar{M}_0\,\partial_y^0(p(y) - \lambda) + (1-\varkappa^*)\,\underline{M}_0\,\partial_y^0(p(y) - \lambda) = \lambda + M^{\varkappa^*}\,\partial_y^0(p(y) - \lambda),$$

где $M^{\varkappa}=\varkappa\bar{M}+(1-\varkappa)M$ — взвешенное среднее. Тогда минимизирующим $r^{\varkappa *}{}_{\lambda}(\partial)$ будет правило $p(y)\underset{\gamma}{\leqslant}\lambda$ (в соответствии с которым при строгом неравенстве имеем $\partial^0{}_y=1$, при обратном неравенстве — $\partial^0{}_y=0$, а при $p(y)=\lambda$ считается $\partial^0{}_y=\gamma$). От значения γ величина $r^{\varkappa *}{}_{\lambda}(\partial)$ не зависит и определяется в результате выражением

$$r^{\varkappa *}_{\lambda}(\partial)=\lambda+M^{\varkappa *}(p(y)-\lambda)^{+}.$$

Варьируя величинами λ и γ, можно добиться того, чтобы уровень правила был равен α, что доказывает следующую теорему.

Теорема 7.5. *Если $p(y)$ есть формальная плотность альтернативы по отношению к гипотезе, то существует такой коэффициент $\varkappa^*$ в диапазоне оптимизма $0\leqslant\varkappa^*\leqslant 1$, что оптимальное правило уровня α будет состоять в сравнении порогом: $p(y)\underset{\gamma}{\leqslant}\lambda$, при таком выборе γ и λ, чтобы уровень равнялся α.*

Точные плотности вероятностей. Перейдем теперь к одному очень важному частному по отношению к предыдущему случаю, когда $\mathcal{M}_0=\mathcal{P}_0$ и $\mathcal{M}_1=\mathcal{P}_1$ — точные (на алгебре $\mathcal{A}_0$) распределения вероятностей, заданные своими плотностями вероятностей по какой-то мере. Следующая теорема в принципе хорошо известна, но она раскрашивается в цвета «оптимизма-пессимизма».

Теорема 7.6 (лемма Неймана-Пирсона). *Пусть $p_0(y)$, $p_1(y)$ есть плотности вероятностей точных на $\mathcal{A}_0$ конечно- или счетно-аддитивных распределений. Тогда при некотором $\varkappa^*$, $\varkappa^*\leqslant 1$, и при всех $\varkappa\geqslant 1$ оптимальным для проверки гипотезы $p_0(y)$ при альтернативе $p_1(y)$ будет правило Неймана — Пирсона, состоящее в сравнении отношения плотностей с порогом:*

$$p_1(y)/p_0(y)\underset{\gamma}{\leqslant}\lambda, \tag{7.4}$$

где $0<\lambda<\infty$ и $0\leqslant\gamma\leqslant 1$ выбираются исходя из нужного уровня $\alpha=1-M_0\partial^0{}_y$.

Доказательство. Существование $\varkappa^*\leqslant 1$, при котором правило (7.4) является оптимальным, следует из теоремы 7.5. Пусть теперь $\varkappa\geqslant 1$. Тогда достаточный класс образуют $\mathcal{A}_0$-измеримые правила. Оптимальность правила (7.16) в классе измеримых правил доказывается в [3, стр. 95].

Робастные методы. Их суть состоит в интерпретации гипотезы и альтернативы в виде семейств распределений вероятностей и подмене семейств на выбранные внутри них наименее благоприятные «экземпляры» распределений вероятностей.

Пара $\mathcal{P}^*{}_0\subset\mathcal{M}_0$, $\mathcal{P}^*{}_1\subset\mathcal{M}_1$ распределений вероятностей называется *наименее благоприятной* (нб) для статистической задачи проверки гипотезы $\mathcal{M}_0$, при альтернативе $\mathcal{M}_1$, если [13]

$$M_{\mathcal{P}^*_0}\{\mathcal{P}^*_1/\mathcal{P}^*_0\leqslant\lambda\}=\underline{M}_0\{\mathcal{P}^*_1/\mathcal{P}^*_0\underset{\gamma}{\leqslant}\lambda\}=1-\alpha,$$

$$M_{\mathcal{P}^*_1}\{\mathcal{P}^*_1/\mathcal{P}^*_0\underset{\gamma}{\leqslant}\lambda\}=\overline{M}_1\{\mathcal{P}^*_1/\mathcal{P}^*_0\underset{\gamma}{\leqslant}\lambda\}.$$

Здесь $\mathcal{P}^*{}_1/\mathcal{P}^*{}_0$ обозначает плотность $\mathcal{P}^*{}_1$ по отношению к $\mathcal{P}^*{}_0$ (оно же — частное плотностей $\mathcal{P}^*{}_1$ и $\mathcal{P}^*{}_0$ по отношению к одной и той же мере), а фигурными скобками оттенена индикаторная функция соответствующего события с рандомизацией на границе (т. е. при равенстве). Для существования нб-распределений, в общем, требуются A_0-измеримости первичных признаков гипотезы и альтернативы.

Утверждение 7.7. *Если наименее благоприятные распределения существуют, то $\mathcal{P}^*{}_1/\mathcal{P}^*{}_0\leqslant\lambda$ является оптимальным правилом при $\varkappa=1$.*

Доказательство. Обозначим ∂^* правило $\mathcal{P}^*{}_1/\mathcal{P}^*{}_0\leqslant\lambda$ уровня α (для обеспечения уровня подбираются λ и γ). Тогда для любого правила $\partial^0{}_y$ уровня α имеем

$$\overline{\beta}(\partial)=\overline{M}_1\,\partial^0_y\geqslant M_{\mathcal{P}^*_1}\,\partial^0_y\geqslant M_{\mathcal{P}^*_1}\{\mathcal{P}^*_1/\mathcal{P}^*_0\underset{\gamma}{\leqslant}\lambda\}=$$

$$=\overline{M}_1\{\mathcal{P}^*_1/\mathcal{P}^*_0\underset{\gamma}{\leqslant}\lambda\}=\overline{\beta}(\partial^*).$$

Таким образом, ошибка второго рода у ∂ не может быть меньше, чем у ∂^*, что и доказывает утверждение.

Существование нб-распределений — это целая проблема, устанавливаемая в каждой конкретной задаче. Ниже рассматривается один такой случай.

Проверка гипотез по заданным интервальным вероятностям. Пусть $\mathcal{M}^y{}_0=\langle\underset{\sim}{p}_0(y),\ \tilde{p}_0(y)\rangle$, $\mathcal{M}^y{}_1=\langle\underset{\sim}{p}_1(y),\ \tilde{p}_1(y)\rangle$ заданы своими интервальными плотностями (по мере-длине на $\mathcal{Y}=\mathcal{R}^n$, хотя это и не обязательно), что эквивалентно заданию интервальных вероятностей, порождающих аддитивные ИРВ.

Введем следующие плотности:

$$p^*_0(y)=\begin{cases}\tilde{p}_0(y) & \text{при}\quad \tilde{p}_0(y)<c_1\,\tilde{p}_1(y),\\ c_1\,\tilde{p}_1(y) & \text{при}\quad \underset{\sim}{p}_0(y)\leqslant c_1\,\tilde{p}_1(y)\leqslant\tilde{p}_0(y),\\ \underset{\sim}{p}_0(y) & \text{при}\quad c_1\,\tilde{p}_1(y)<\underset{\sim}{p}_0(y);\end{cases}$$

$$p^*_1(y)=\begin{cases}\tilde{p}_1(y) & \text{при}\quad \tilde{p}_1(y)<b_1\,\tilde{p}_0(y),\\ b_1\,\tilde{p}_0(y) & \text{при}\quad \underset{\sim}{p}_1(y)\leqslant b_1\,\tilde{p}_0(y)\leqslant\tilde{p}_1(y),\\ \underset{\sim}{p}_1(y) & \text{при}\quad b_1\,\tilde{p}_0(y)<\underset{\sim}{p}_1(y);\end{cases}$$

$$q^*_0(y)=\begin{cases}\tilde{p}_0(y) & \text{при}\quad \tilde{p}_0(y)<c_2\,\underset{\sim}{p}_1(y),\\ c_2\,\tilde{p}_1(y) & \text{при}\quad \underset{\sim}{p}_0(y)\leqslant c_2\,\underset{\sim}{p}_1(y)\leqslant\tilde{p}_0(y),\\ \underset{\sim}{p}_0(y) & \text{при}\quad c_2\,\underset{\sim}{p}_1(y)<\underset{\sim}{p}_0(y);\end{cases}$$

$$q_1^*(y)=\begin{cases}\tilde{p}_1(y) & \text{при} \quad \tilde{p}_1(y)<b_2\,\underset{\sim}{p}_0(y),\\ b_2\,\underset{\sim}{p}_0(y) & \text{при} \quad \underset{\sim}{p}_1(y)\leqslant b_2\,\underset{\sim}{p}_0(y)\leqslant\tilde{p}_1(y),\\ \underset{\sim}{p}_1(y) & \text{при} \quad b_2\,\underset{\sim}{p}_0(y)<\underset{\sim}{p}_1(y);\end{cases}$$

где коэффициенты c_1, c_2, b_1 и b_2 определяются однозначно из условия нормировки плотностей: $\int p^*_k(y)dy=\int q^*_k(y)dy=1$, $k=0,\ 1$.

Вид этих плотностей приведен схематически (для $y\in\mathcal{R}$) на рис. 7.5. Отличие $q^*_0(y)$ от $p^*_0(y)$ будет состоять в поведении на участке перехода с нижней границы плотности на верхнюю, где $p^*_0(y)$ движется пропорционально $\tilde{p}_1(y)$ и выше ее, а $q^*_0(y)$ — пропорционально $\underset{\sim}{p}_1(y)$ и ниже ее, причем коэффициент пропорциональности в первом случае больше единицы, во втором — меньше, но и там и там однозначно таков, что выполняется нормировка плотностей. Аналогичным является различие между $q^*_1(y)$ и $p^*_1(y)$. Уже с беглого взгляда видно, что эти плотности имеют характер «наименее благоприятных», так как в максимальной мере «наплывают» на сторону противоположной гипотезы. Этот факт облачим в строгую формулировку.

Теорема 7.8. *Пусть гипотеза и альтернатива заданы своими интервальными плотностями и пусть выполняются условия*

$$\int\limits_{\tilde{p}_0(y)>0}\tilde{p}_1(y)\,dy+\int\limits_{\tilde{p}_0(y)=0}\underset{\sim}{p}_1(y)\,dy>1,\quad \int\limits_{\tilde{p}_1(y)>0}\tilde{p}_0(y)\,dy+\int\limits_{\tilde{p}_1(y)=0}\underset{\sim}{p}_0(y)\,dy>1.$$

Тогда, если $c_1b_1\geqslant 1$, то нб-плотностями будут $p^*_0(y)$ и $p^*_1(y)$,

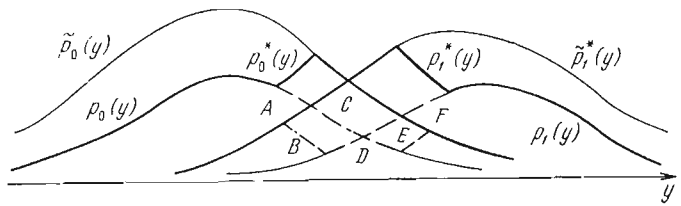


Рис. 7.5. Наименее благоприятные плотности при $c_1b_1\geqslant 1$ (сплошная линия) и $c_1b_1\leqslant$ (штриховая)

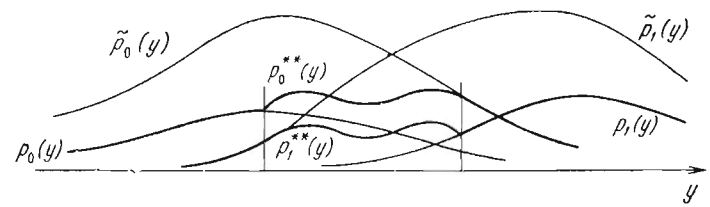


Рис. 7.6. Наименее благоприятные плотности при $c_1b_1<1<c_2b_2$

если $c_2b_2\leqslant 1$, то $q^*_0(y)$ и $q^*_1(y)$, наконец, если $c_1b_1<1$ и $c_2b_2>1$, то нб-плотности $p^{**}_0(y)$ и $p^{**}_1(y)$ таковы, что их отношение

$$\frac{p_1^{**}(y)}{p_0^{**}(y)}=\begin{cases}\tilde{p}_1(y)/\underset{\sim}{p}_0(y) & \text{при} \quad \tilde{p}_1(y)/\underset{\sim}{p}_0(y)<b_3,\\ b_3 & \text{при} \quad \underset{\sim}{p}_1(y)/\tilde{p}_0(y)\leqslant b_3\leqslant\tilde{p}_1(y)/\underset{\sim}{p}_0(y),\\ \underset{\sim}{p}_1(y)/\tilde{p}_0(y) & \text{при} \quad b_3<\underset{\sim}{p}_1(y)/\tilde{p}_0(y),\end{cases}$$

где постоянная b_3 определяется из уравнения

$$\int[\underset{\sim}{p}_1(y)-b_3\,\tilde{p}_0(y)]^+dy-\int[b_3\,\underset{\sim}{p}_0(y)-\tilde{p}_1(y)]^+\,dy+b_3=1.$$

Доказательство теоремы имеется в [13]. Таким образом, в зависимости от значений нормируемых констант c_1, b_1, c_2, b_2 разделяются три случая. Первый $c_1b_1>1$ соответствует на рис. 7.5 меньшей единице площади областей как $A+(B+C+D+E)$, так и $F+(B+C+D+E)$, и дает возможность линии перехода нб-плотностей с одной границы на другую пройти, как на рис. 7.5. Второй случай $c_2b_2<1$ эквивалентен превышению единицы площадей как $A+(B+D+E)$, так и $F+(B+D+E)$, и позволяет линии перехода пройти, как это обозначено на рис. 7.5 штриховой линией. Наконец, если оба этих случая исключаются, то наименее благоприятные плотности проходят так, как это показано на рис. 7.6: они просто в области перехода произвольны, пропорциональны друг другу, но лежат внутри границ и подчиняются условию нормировки плотностей.

Робастный алгоритм при независимых наблюдениях. Распространим понятие нб-плотности на независимую однородную выборку $\mathbf{y}=(y_1,\ \ldots,\ y_n)$, каждый элемент которой описывается интервальными плотностями $\underset{\sim}{p}_0(y_i)$, $\tilde{p}_0(y_i)$; $\underset{\sim}{p}_1(y_i)$, $\tilde{p}_1(y_i)$.

Утверждение 7.9. *Пусть $p^*_0(y_i)$ и $p^*_1(y_i)$ есть нб-плотности для одного элемента y_i. Тогда нб-плотностями для вектора* $\mathbf{y}$ *с независимыми компонентами y_i будут произведения* [13]

$$p_0^*(\mathbf{y})=\prod_i p_0^*(y_i),\qquad p_1^*(\mathbf{y})=\prod_i p_1^*(y_i).$$

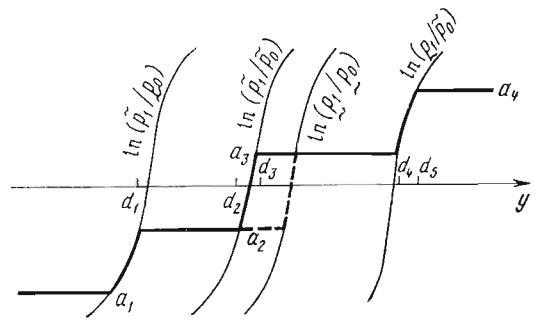


Рис. 7.7. Характерные нелинейные преобразования

Таким образом, оптимальное при $\varkappa=1$ правило проверки гипотез имеет вид $\prod\limits_i p^*_1(y_i)/p^*_0(y_i) \underset{\nu}{\leqslant} \lambda$, а после логарифмирования: $\sum\limits_i \ln[p^*_1(y_i)/p^*_0(y_i)] \underset{\nu}{\leqslant} \lambda$. Оно предполагает обработку каждого элемента выборки в соответствии с нелинейной функцией $f(y_i) = \ln[p^*_1(y_i)/p^*_0(y_i)]$, суммирование результатов обработки и сравнение их с порогом λ: $\sum\limits_i f(y_i) \underset{\nu}{\leqslant} \lambda$.

Вид нелинейной функции $f(y_i)$ зависит от исходных границ плотностей. В самом общем случае $f(y_i)$ записывается:

$$f(y_i) = \begin{cases} a_1 & \text{при} \quad y_i \leqslant d_1, \\ \ln[\widehat{p}_1(y_i)/\underset{\sim}{p}_0(y_i)] & \text{при} \quad d_1 < y_i \leqslant d_2, \\ a_2 & \text{при} \quad d_2 < y_i \leqslant d_3, \\ \ln[\widehat{p}_1(y_i)/\widehat{p}_0(y_i)] \\ \text{или } \ln[\underset{\sim}{p}_1(y_i)/\underset{\sim}{p}_0(y_i)] & \text{при} \quad d_3 < y_i \leqslant d_4, \\ a_3 & \text{при} \quad d_4 < y_i \leqslant d_5, \\ \ln[\underset{\sim}{p}_1(y_i)/\widetilde{p}_0(y_i)] & \text{при} \quad d_5 < y_i \leqslant d_6, \\ a_4 & \text{при} \quad d_6 < y_i. \end{cases}$$

Функция $f(y_i)$ имеет, в общем, как видно из рис. 7.7, четыре постоянных уровня: $a_1<a_2<a_3<a_4$. Крайние уровни со значениями a_1 и a_4 обязаны своим происхождением «затянутости хвостов» $\tilde{p}_0(y_i)$ и $\tilde{p}_1(y_i)$ при $|y_i|\to\infty$ и возникающей отсюда возможности «сверхвыбросов» наблюдений. Эти «сверхвыбросы» и ограничиваются уровнями a_1 и a_4.

В частном случае при $\tilde{p}_0(y_i) \equiv \tilde{p}_1(y_i) \equiv \infty$ (верхние границы не задаются) средние уровни a_2 и a_3 пропадают и функция $f(y_i)$ принимает вид ограничителя сверху и снизу:

$$f(y_i) = \begin{cases} a_1 & \text{при} \quad y_i \leqslant d_1, \\ \ln[\underset{\sim}{p}_1(y_i)/\underset{\sim}{p}_0(y_i)] & \text{при} \quad d_1 < y_i \leqslant d_6, \\ a_4 & \text{при} \quad d_6 < y_i. \end{cases}$$

Такой же вид нелинейности будет, если заданы только верхние границы плотностей и не заданы нижние $(\underset{\sim}{p}_0(y_i) \equiv \underset{\sim}{p}_1(y_i) \equiv 0)$; с той лишь разницей, что в этом случае на среднем участке $f(y_i)$ равна $\ln[\tilde{p}_1(y_i)/\tilde{p}_0(y_i)]$.

Возможен и другой вид нелинейности:

$$f(y_i) = \begin{cases} \ln[\widehat{p}_1(y_i)/\underset{\sim}{p}_0(y_i)] & \text{при} \quad y_i \leqslant d_2, \\ a_2 & \text{при} \quad d_2 < y_i \leqslant d_5, \\ \ln[\underset{\sim}{p}_1(y_i)/\widehat{p}_0(y_i)] & \text{при} \quad d_5 < y_i. \end{cases}$$

Здесь, наоборот, выделяются выбросы наблюдений и «смазывается» средняя неопределенная часть, т. е. работа производится по

«вершкам» наблюдений в отличие от предыдущего случая, где полезнее оказались «корешки».

Замечания. 1. Утверждение 7.9 справедливо и в том случае, если считать выборку стационарной [13] — значит, дополнительный тезис о стационарности наблюдений, хотя и сужает по сравнению с однородностью математическую модель, но в данной задаче не меняет ее решения (интересно, насколько это общий факт?).

2. Все изложенное нами в этом параграфе справедливо, если плотности определены не по длине, а по произвольной мере μ, только тогда интегралы Римана по dy (для счетно-аддитивных моделей — интегралы Лебега) заменяются на интегралы Римана — Стилтьеса (Лебега — Стилтьеса) по $d\mu$.

3. Заданы границы вероятностей событий. Под рассмотренный случай интервальных плотностей подпадает дискретный эксперимент, исходами которого являются J взаимноисключающих событий. Формально пронумеруем их числами от от 1 и до J: $\mathcal{Y}=\{1, \ldots, J\}$. Считаются заданными границы вероятностей $\underset{\sim}{P}_0(j)$, $\tilde{P}_0(j)$, $\underset{\sim}{P}_1(j)$, $\tilde{P}_1(j)$ этих событий $j=1, \ldots, J$, определяющие собой гипотезы. Имея результаты n независимых таких экспериментов в виде зарегистрированных чисел $n_1, n_2, \ldots, n_J$ выпадения каждого события, $\sum\limits_{j=1}^{J} n_j = n$, нужно проверить гипотезу. Совершенно формально здесь на вероятности надо смотреть как на плотность по считающей мере ($\mu(j) \equiv 1$) и, следуя замечанию 2, использовать результаты последнего раздела. Тогда нелинейность $f(y_i)$ заменится на указатель событий, какие произошли в i-м эксперименте, суммирование $f(y_i)$ по i от 1 до n превратится в счетчик чисел n_j и оптимальное правило примет вид

$$\sum_{j=1}^{J} n_j \ln\left(P^*_1(j)/P^*_0(j)\right) \underset{\nu}{\leqslant} \lambda.$$

Расчет значений отношений наименее благоприятных вероятностей в аргументе логарифма осуществляется по теореме 7.8 с подстановкой там границ вероятностей на место плотностей (вместо y подставляется j) и заменой интегралов по dy на суммы по j.

7.5. ПРОВЕРКА ГИПОТЕЗ О ЗАДАННОМ ЗНАЧЕНИИ ПАРАМЕТРА

Формулировка задачи. Если в § 7.3 при гипотезе $x=x_0$ альтернатива звучала как некоторое точное значение $x=x_1$ параметра, то здесь она конкретно не задается. Альтернативу будут составлять все значения x, не равные x_0, т. е. $x\neq x_0$. Решения выносятся такие: соответствует x значению x_0, т. е. $x=x_0$ или нет, т. е. $x\neq x_0$, а место альтернативы занимает направление отклонений, указываемое параметром x, а точнее, его физическим смыслом и связью с наблюдениями. Так, если нет сигнала, то значение его амплитуды — ноль ($x=0$), есть — не равно нулю.

Пусть каждому значению параметра x поставлена в соответствие переходная модель $\mathcal{M}^y{}_x$ на $\mathcal{Y}$. Тогда гипотеза $x=x_0$ о значении параметра и альтернатива $x\neq x_0$ в терминах ИМ принимает вид

$$\mathcal{M}^y_0=\mathcal{M}^y{}_{x_0},\ \mathcal{M}^y_1=\bigvee_{x\neq x_0}\mathcal{M}^y{}_x. \tag{7.5}$$

Замечание. Альтернатива (7.5) соответствует тому, что априори какие-либо данные о x, кроме $x\neq x_0$, отсутствуют. Тогда $\mathcal{M}^y{}_1$ будет частной ИМ к совместной $\mathcal{I}^x{}_1\mathcal{M}^y{}_x$, где $\mathcal{I}^x{}_1$ — голая на $x\in\mathcal{X}-x_0$ ($\mathcal{X}$ без точки x_0); если же априорные данные о x имеются в виде ИМ $\mathcal{M}^x{}_1$, то — будет частной к произведению $\mathcal{M}^x{}_1\mathcal{M}^y{}_x$.

Часто x имеет физическое толкование как параметр состояния среды. Например, при функциональном представлении $y=V_x\xi$ вектор (или процесс) ξ означает флуктуации (шум среды), а V_x — оператор, согласно которому влияние ξ на наблюдения зависит от x. Пусть $y_i=x+\xi_i$, тогда параметр состояния среды есть сдвиг на величину x всех наблюдений y_i, гипотеза $x=0$ соответствует отсутствию такого сдвига.

В случае непрерывности $\mathcal{M}^y{}_x$ по x (в смысле непрерывности $\overline{M}^y{}_x f(y)$ как функции x при любых $f\in\mathcal{F}$) или непрерывности оператора V_x в точке x_0, казалось бы, возникает безвыходная ситуация, так как альтернатива охватывает гипотезу: $\mathcal{M}^y{}_0\subset\mathcal{M}^y{}_1$, откуда для любого правила $\partial^0{}_y$ имеет место $\overline{\beta}(\partial)=\overline{M}_1\partial^0{}_y\geqslant\overline{M}_0\partial^0{}_y=1-\underline{M}_0\partial^1{}_y=1-\underline{\alpha}(\partial)$ и аналогично $\underline{\beta}(\partial)\leqslant 1-\overline{\alpha}(\partial)$. В результате при $\varkappa\geqslant 1$: $\beta^\varkappa(\partial)\geqslant 1-\varkappa\underline{\alpha}(\partial)+(\varkappa-1)\overline{\alpha}(\partial)\geqslant 1-\varkappa\overline{\alpha}(\partial)+(\varkappa-1)\overline{\alpha}(\partial)=1-\overline{\alpha}(\partial)$ и в любом случае оптимальным становится тривиальное правило $\partial^0{}_y\equiv 1-\alpha$. Задача вырождается.

В следующих разделах предлагаются различные способы выхода из создавшегося затруднения.

О правилах при оптимизме. Оптимизм $\varkappa<1$ в условиях охватывания альтернативой гипотезы позволяет строить оптимальные нетривиальные правила, что иллюстрируется примером.

Пример 7.3. Пусть в векторных обозначениях $y=1x+\xi$ и пусть флуктуации ξ заданы интервальной плотностью $\underset{\sim}{p}(z)$, $\tilde{p}(z)$ по отношению к мере-длине на $\mathcal{R}^n$. Будем считать, что границы $\underset{\sim}{p}(z)$ и $\tilde{p}(z)$ унимодальны и достигают максимума при $z=0$. Будем рассматривать детерминированные (измеримые относительно алгебры отрезков) правила, причем те, для которых

$$\underline{\alpha}(\partial)=\int\partial^1_z\,\underset{\sim}{p}(z)\,dz,\quad\overline{\alpha}(\partial)=\int\partial^1_z\,\tilde{p}(z)\,dz,$$

что будет иметь место при $\int\tilde{p}(z)dz<\infty$, если требуемый уровень α довольно мал. Тогда

$$\underline{\beta}(\partial)=\inf_x\underline{M}\,\partial^0_{1x+\xi}=\inf_x\max\left\{\int\partial^0_{1x+z}\,\underset{\sim}{p}(z)\,dz,\quad 1-\int\partial^1_{1x+z}\,\tilde{p}(z)\,dz\right\},$$

где использован тот факт, что $\underline{M}\partial^0_{1x+\xi}=\underline{P}\left(\partial^0_{1x+\xi}=1\right)$ есть для детерминированного правила $\partial^0{}_y$ вероятность соответствующего события и применена формула для вероятности событий аддитивных ИРВ. В силу унимодальности границ плотностей инфимум по x будет достигаться при неограниченном увеличении x. Положим $|x|\leqslant H$. Тогда инфимум достигается при $|x|=H$, что дает нам: $\underline{\beta}(\partial)=\max\left\{\int\partial^0_y\,\underset{\sim}{p}(y\pm 1H)\,dy,\ 1-\int\partial^1_y\,\tilde{p}(y\pm 1H)\,dy\right\}$, где справа берется тот знак $+H$ или $-H$, при котором окажется меньше $\underline{\beta}(\partial)$. Далее $\overline{\beta}(\partial)=\sup_x\overline{M}\,\partial^0_{1x+\xi}\geqslant\overline{M}\,\partial^0_\xi=1-\underline{\alpha}(\partial)$. При малом уровне α вполне можно считать $\overline{\beta}(\partial)=1$. В результате оптимальным будет правило, минимизирующее $(1-\varkappa)\underline{\beta}(\partial)$ при заданном $\alpha^\varkappa(\partial)=\int\partial^1{}_z[(1-\varkappa)\underset{\sim}{p}(z)+\varkappa\tilde{p}(z)]dz$. В соответствии с теоремой 7.6 (и выкладками, аналогичными сопровождающим теорему 7.8) оптимальным правилом будет одно из следующих четырех правил:

$$\frac{\underset{\sim}{p}(y\pm 1H)}{(1-\varkappa)\underset{\sim}{p}(y)+\varkappa\tilde{p}(y)}\leqslant\lambda,\qquad\frac{\tilde{p}(y\pm 1H)}{(1-\varkappa)\underset{\sim}{p}(y)+\varkappa\tilde{p}(y)}\leqslant\lambda.$$

Выбирается то из них, для которого минимальна ошибка $\underline{\beta}(\partial)$, рассчитываемая по составляющим каждое правило границам плотностей; по ним же находится уровень α, который фиксируется выбором λ. Отметим в заключение примера, что ошибка $\beta^\varkappa(\partial)$ для любого из приведенных правил при $\varkappa\to 1$ будет стремиться к $\overline{\beta}(\partial)$, причем $\overline{\beta}(\partial)\geqslant 1-\underline{\alpha}(\partial)\approx 1$, т. е. найденное правило работоспособно лишь при небольших $\varkappa$.

Равномерно-оптимальные правила. Правило $\partial^0{}_y$ называется *равномерно-оптимальным* для проверки гипотезы $x=x_0$ при альтернативе $x\neq x_0$, если оно является оптимальным правилом проверки гипотезы $x=x_0$ при альтернативе $x=x_1$, каким бы ни было $x_1\neq x_0$.

Платформу для существования равномерно-оптимальных правил при $\varkappa=1$ создает теорема 7.1, согласно которой структура оптимальных правил целиком определяется линейной комбинацией первичных признаков, составляющих гипотезу, а уровень α и вид альтернативы влияет лишь на коэффициенты линейной комбинации. При столь скудных исходных данных, что варьируемым становится всего один коэффициент, он сам полностью определится уровнем α. Тогда от альтернативы вид правила зависеть не будет и имеем равномерно-оптимальное правило.

Пример 7.4. Пусть $y_i=x+\xi_i$, $i=1,\ \ldots,\ n$; $M\xi_i\xi_j=0$, при $i\neq j$ и $M^\xi_x\,\xi^2{}_i=\overline{\sigma}^2{}_x$, т. е. флуктуации ξ_i некоррелированы, однородны и их мощность $\overline{\sigma}^2{}_x$ зависит от параметра сдвига x. Проверяем гипотезу $x=0$ при альтернативе $x\neq 0$. Если бы альтернативным значением было $x=x_1$, то оптимальное правило согласно дополнению 4 на с. 390 имело бы вид

$$\partial^0_y=[1-\alpha\hat{n}y^2/\overline{\sigma}^2_0]^+\ \text{ при }\ \varepsilon=(\overline{\sigma}_0+\overline{\sigma}_{x_1})/\sqrt{n}\leqslant|x_1|\leqslant\overline{\sigma}_0/\alpha\sqrt{n}.$$

В довольно широком (при малых α и больших n) диапазоне изменения пара-

метра x_1 оптимальным оказалось одно и то же правило, вид которого определился целиком данными о гипотезе. Оно и станет равномерно оптимальным для x в указанном диапазоне. Хотя нижняя граница этого диапазона зависит от $\overline{\sigma}_{x_1}$, но случай $|x_1|>\varepsilon$ не представляет практического интереса, так как при этом $\overline{\beta}(\partial)\geqslant 1-\alpha$, и поэтому может быть исключен из рассмотрения.

Введение защитного диапазона. Этот метод состоит в сужении альтернативы до неравенства: $0<\varepsilon\leqslant|x-x_0|\leqslant H<\infty$, где ε и H — числа, ограничивающие диапазон изменения x. Задача проверки гипотез, в общем, будет зависеть от выбора ε и H. В целом ряде задач значения $x_0\pm\varepsilon$ и $x_0\pm H$ будут соответственно наименее и наиболее благоприятными в том смысле, что для оптимальных правил $\partial^0{}_y$ уровня α справедливы равенства

$$\overline{\beta}(\partial)=\overline{M}^{y}_{x\pm\varepsilon}\,\partial^0_y,\quad \underline{\beta}(\partial)=\underline{M}_{x\pm H}\,\partial^y_0 .$$

В силу этих равенств при поиске оптимального правила и расчете ошибок первого и второго рода альтернатива $\mathscr{M}^y{}_1=\bigvee_{\varepsilon<|x-x_0|<H}\mathscr{M}^y{}_x$ вполне может быть заменена на $\mathscr{M}_{x_0-\varepsilon}\vee\mathscr{M}_{x_0+\varepsilon}\vee\mathscr{M}_{x_0-H}\vee\mathscr{M}_{x_0+H}$, что соответствует замене числового параметра x на отдельные точки: $x_0\pm\varepsilon$, $x_0\pm H$. При $\varkappa\geqslant 1$ «наиболее благоприятные» точки $x_0\pm H$ теряют смысл и останутся лишь «наименее благоприятные» $x_0\pm\varepsilon$ как наиболее приближенные к гипотезе. Если вдобавок к сказанному вид оптимального правила не будет зависеть от величины ε, то будем иметь равномерно оптимальное правило.

Пример 7.5. Пусть в примере 7.3 $\varkappa=1$. Тогда при сделанных там предположениях и альтернативе $x>\varepsilon$ задача нахождения оптимального правила сводится к минимизации

$$\overline{\beta}(\partial)=\overline{M}\,\partial^0{}_{1\,\varepsilon+\xi}=\min\left\{\int\partial^0_y\,\tilde{p}\,(\mathrm{y}-\mathbf{1}\,\varepsilon)\,d\mathrm{y},\quad 1-\int\partial^1_y\,\underset{\sim}{p}\,(\mathrm{y}-\mathbf{1}\,\varepsilon)\,d\,\mathrm{y}\right\}$$

при заданном $\overline{\alpha}(\partial)=\int\partial^1{}_{\mathrm{y}}\,\tilde{p}\,(\mathrm{y})\,d\mathrm{y}=\alpha$. Оптимальным будет то из двух правил $\tilde{p}(\mathrm{y}-\mathbf{1}\varepsilon)/\tilde{p}(\mathrm{y})\leqslant\lambda$, $\underset{\sim}{p}(\mathrm{y}-\mathbf{1}\varepsilon)/\tilde{p}(\mathrm{y})\leqslant\lambda$ (где λ выбирается по заданному уровню α), для которого минимальна ошибка $\overline{\beta}(\partial)$, рассчитанная по границам плотностей, составляющим правило. Правило будет равномерно оптимальным, если его вид не зависит от значения ε. Сказанное будет иметь место, например, в случае «нормальных» границ вида: $\underset{\sim}{p}(\mathrm{y})=\underline{\nu}p_0(\mathrm{y})$, $\tilde{p}(\mathrm{y})=\overline{\nu}p_0(\mathrm{y})$, $0\leqslant\underline{\nu}\leqslant 1\leqslant\overline{\nu}<\infty$, $p_0(\mathrm{y})=\exp\left\{-\frac{1}{2}\,\mathrm{y}^{\mathrm{T}}\,\mathbf{B}^{-1}\,\mathrm{y}\right\}\Big/(2\pi^{n/2}(\det\mathbf{B})^{1/2})$, где $\mathbf{B}$ есть симметричная положительно определенная матрица. Тогда правило $\mathrm{y}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{1}\leqslant\lambda$, где λ определяется из уравнения $\overline{\nu}\int\limits_{\mathrm{y}^{\mathrm{T}}\,\mathbf{B}^{-1}\,\mathbf{1}>\lambda}p_0(\mathrm{y})d\mathrm{y}=\alpha$, будет равномерно оптимальным. Его ошибка второго рода $\overline{\beta}(\partial)=\overline{\nu}\int\limits_{\mathrm{y}^{\mathrm{T}}\,\mathbf{B}^{-1}\,\mathbf{1}\leqslant\lambda}p_0(\mathrm{y}-\varepsilon\mathbf{1})d\mathrm{y}$, конечно же, зависит от величины ε и стремится к $1-\alpha$ при $\varepsilon\to 0$.

Минимизация интегральной ошибки. Пусть $\varkappa=1$. Введение защитного диапазона $|x-x_0|\geqslant\varepsilon$, отделяющего альтернативу от гипотезы, эквивалентно априорному допущению о невозможности

при альтернативе появления «очень близких» к x_0 значений x. Обобщением является задание соответствующего альтернативе вероятностного закона на $\mathscr{X}$ с помощью $\mathscr{M}^x{}_1$. Тогда $\mathscr{M}^y{}_1$ будет частной к произведению $\mathscr{M}^x{}_1\mathscr{M}^y{}_x$, а ошибки будут:

$$\overline{\alpha}(\partial)=\overline{M}^{y}_{x_0}\,\partial^1_y,\qquad \overline{\beta}(\partial)=\overline{M}^{x}_{1}\,\overline{M}^{y}_{x}\,\partial^0_y=\overline{M}^{x}_{y}\,\overline{\beta}_x(\partial),$$

где $\overline{\beta}_x(\partial)=\overline{M}^y{}_x\partial^0{}_y$. При точной априорной плотности $q(x)$ (по отношению к мере-длине), задающей $\mathscr{M}^x{}_1$, ошибка находится интегрированием:

$$\overline{\beta}(\partial)=\int\overline{\beta}_x(\partial)\,q\,(x)\,dx. \tag{7.6}$$

Задача поиска оптимального правила сводится к минимизации этой интегральной ошибки выбором $\partial^0{}_y$ при заданном уровне α. Отметим, что в (7.6) $q(x)$ может быть некоторой весовой функцией, характеризующей уровень предпочтения, отдаваемого различным значениям x при альтернативе. Например, допускается $q(x)\equiv 1$ (тогда $\overline{\beta}(\partial)$, по сути, уже не величина ошибки, а некоторый функционал от нее).

Использование доверительных оценок. Пусть $\partial^*{}_y(x)$ есть оптимальная доверительная оценка параметра x при коэффициенте пессимизма $\varkappa=1$. Если считается, что $\mathscr{M}^{xy}=\mathscr{J}^x\mathscr{M}^y{}_x$, то уровень оценки определяется формулой: $\alpha=\sup\limits_x\overline{M}^y{}_x[1-\partial^*{}_y(x)]$.

Для проверки гипотезы $x=x_0$ при альтернативе $x\neq x_0$ введем правило $\partial^0{}_y=\partial^*{}_y(x_0)$. Оно будет уровня не больше α, так как

$$\overline{\alpha}(\partial)=\overline{M}_0\left(1-\partial^0_y\right)=\overline{M}^{y}_{x_0}\left[1-\partial^*_y(x_0)\right]\leqslant\sup_x\overline{M}^{y}_{x}\left[1-\partial^*_y(x)\right]=\alpha.$$

Трудно говорить об оптимальности введенного правила $\partial^0{}_y$. Тем не менее на примере будет показано, что оптимальность оценки $\partial^*{}_y(x)$ в смысле минимума расплывчатости $\Omega(\partial)=\overline{M}^y\int\partial^*{}_y(x)dx$ в некотором смысле приводит к сокращению интегральной ошибки (7.6) второго рода.

Пример 7.6. Пусть в векторных обозначениях $\mathrm{y}=\mathbf{1}x+\xi$ и пусть $\varkappa=1$. Если об x и его связи с ξ ничего не известно, то оптимальная оценка будет функцией $\mathrm{y}-\mathbf{1}x$, поэтому правило $\partial^0{}_{\mathrm{y}}$ проверки гипотезы $x=x_0$ будет $\partial^0{}_{\mathrm{y}}=\partial^*(\mathrm{y}-\mathbf{1}x_0)$. Его ошибка первого рода равняется уровню α оценки $\partial^*(\xi)$; действительно,

$$\overline{\alpha}(\partial)=\overline{M}_0\left(1-\partial^0_{\mathrm{y}}\right)=\overline{M}_0\left[1-\partial^*(\mathrm{y}-\mathbf{1}\,x_0)\right]=\overline{M}\left[1-\partial^*(\xi)\right]=\alpha.$$

А ошибка второго рода при $x=x_1$ будет равна

$$\overline{\beta}_{x_1}(\partial)=\overline{M}^{\mathrm{y}}_{x_1}\partial^0_{\mathrm{y}}=\overline{M}^{\mathrm{y}}_{x_1}\partial(\mathrm{y}-\mathbf{1}\,x_0)=\overline{M}\,\partial(\xi+\mathbf{1}\,x_1-\mathbf{1}\,x_0).$$

С другой стороны, интегральная ширина оценки $\overline{M}\Omega(\partial)=\overline{M}\int\partial(\mathrm{y}-\mathbf{1}x)dx=\overline{M}\int\partial(\xi+\mathbf{1}x-\mathbf{1}x_0)dx\leqslant\int\overline{M}\partial(\xi+\mathbf{1}x-\mathbf{1}x_0)dx=\int\overline{\beta}_x(\partial)dx$. Если здесь вместо неравенства имеет место равенство, то правило $\partial^0{}_{\mathrm{y}}$ будет минимизировать интеграл от ошибки второго рода по параметру x, следовательно, будет оптимальным в смысле интегральной ошибки, что мы и хотели показать.

Таким образом, предлагаемая здесь методика позволяет прямо приложить все доверительные оценки параметров, полученные в гл. 6, к проверке гипотез о значении этих параметров. Выписать получающиеся отсюда правила предоставляется интересующемуся читателю.

7.6. РАЗЛИЧЕНИЕ НЕСКОЛЬКИХ ГИПОТЕЗ

Общие положения. Мы выше рассматривали задачу проверки нулевой (консервативной) гипотезы при противостоящей ей альтернативе. Пусть теперь гипотез, в общем, не две, а K, и они в какой-то степени равноправны, каждая определяется своей ИМ $\mathcal{M}_k$, $k=1, \ldots, K$ на $\mathcal{Y}$. По наблюдению y нужно установить, какая из гипотез имеет место, т. е. какой ИМ подчиняется y. Этот случай охватывает многие практические задачи классификации и распознавания образов, сюда входит прием ансамбля сигналов, различение букв рукописного текста, узнавание вида болезни по диагнозам и т. д.

Решающее правило составляется как совокупность $\partial_y = (\partial^1_y, \ldots, \partial^K_y)$ решений, каждое из которых ∂^k_y, $k=1, \ldots, K$, определяет тот уровень предпочтения, который при наблюдении y отдается k-й гипотезе. Суммарное предпочтение не должно превышать 1, что выражается в требовании

$$\sum_{k=1}^{K} \partial^k_y \leqslant 1, \ \forall y. \tag{7.7}$$

Зададимся вопросом, а можно ли сумме предпочтений разрешить быть меньше 1? Будем считать, что, в общем, да, оставляя незаполненной долю $\partial^?_y = 1 - \sum_{k=1}^{K} \partial^k_y$ предпочтений, называемую *нейтральным решением* (примеры дает рис. 7.2,*а, б, в*, где $\partial^?_y$ равняется длине отрезка по вертикали, заключенного между сплошной и штриховой линиями). Нейтральное решение не относится ни к одной из гипотез, т. е. никакого решения не выносится. Его введение вызвано исключительно удобством и никак не умаляет общности, а скорее наоборот: по крайней мере, как мы вскоре увидим, $\partial^?_y$ всегда можно «раздать» по ∂^k_y так, чтобы сумма предпочтений (7.7) при всех y строго равнялась 1, а качество сохранялось то же.

Предпочтения ∂^k_y, когда это не 1 (а остальные — 0) или не 0, представляют собой рекомендации, оставляющие все-таки последнее слово за человеком (способным при принятии окончательного решения воспользоваться дополнительными соображениями). При необходимости «механического» выбора в пользу конкретной гипотезы предпочтения заменяются на процедуру «слепой» рандомизации, согласно которой при каждом y разыгрывается игра с исходами $k=1, \ldots, K$, и вероятностями ∂^k_y исходов.

Обозначим $\mathcal{M}^k$ — модель на пространстве значений 1, ..., K, вбирающую в себя априори все статистические данные о гипотезах. Произведение $\mathcal{M}^{ky} = \mathcal{M}^k \mathcal{M}_k$ дает совместную модель гипотез и наблюдений, т. е. СИМ. Величина $\underline{M\partial} = \underline{M^k} \underline{M_k \partial^k_y}$ есть нижняя *вероятность правильного решения* (опознавания), усредненная по гипотезам. Это — совокупная характеристика надежности правила. Составными ее частями являются $\underline{M_k \partial^k_y}$ — нижние вероятности правильного принятия k-х гипотез (каждая вычисляется по своей $\mathcal{M}_k$ продолжением первичных средних), и обратные им величины $\bar{\alpha}_k(\partial) = 1 - \underline{M_k \partial^k_y}$, охватывающие все ошибки, связанные с k-й гипотезой.

Выделим два крайних случая. Первый, когда $\mathcal{M}^k = \mathcal{Y}^k$, т. е. априорных данных нет, тогда оператор $\underline{M^k}$ осуществляет минимизацию и получается $\underline{M\partial} = \min_k \underline{M_k \partial^k_y}$. Другой случай, когда априорные вероятности заданы точно и равны λ_k, $\sum_k \lambda_k = 1$, тогда $\underline{M\partial} = \sum_k \lambda_k \underline{M_k \partial^k_y}$. В последней сумме, чтобы она приобрела новую полезную окраску, полезно снять с λ_k «одежду» вероятностей (как правило, неизвестных) и наложить груз ответственности за последствия неверного отказа от гипотез, это общий прием.

Проблема оптимального синтеза состоит в поиске правила, минимизирующего *совокупную ошибку* $\bar{\alpha}(\partial) = 1 - \underline{M\partial}$ при ограничении (7.7). Методом множителей Лагранжа оптимальное правило получается минимизацией составного риска $\bar{\alpha}(\partial) + \lambda \max_y \sum_k \partial^k_y$ при выборе λ согласно (7.7). В этом плане формулируется и достаточность. В этом же плане усматривается явная аналогия с доверительным оцениванием дискретного параметра $x=k$ (с точностью до способа выбора λ). Осталось переформулировать теорему 6.1 (следствие 1) о достаточности к гипотезам.

Теорема 7.10. *Достаточный при $\varkappa=1$ класс правил в задаче различения нескольких гипотез $\mathcal{A}_k$, $k=1, \ldots, K$, определенных каждая своими первичными средними $\tilde{M}_k g_{kj}(y)$, $j=1, \ldots, J_k$, образуют усеченные снизу нулем, а сверху — совокупным требованием (7.7), вторичные признаки соответствующих гипотез*:

$$\partial^k_y = [c_k - \sum_j a^+_{kj} g_{kj}(y)]^+, \ k = 1, \ldots, K,$$

причем те, для которых ошибки определяются как суммы первичных средних:

$$\bar{\alpha}_k(\partial) = 1 - c_k + \sum_j a^+_{kj} \tilde{M}_k g_{kj}.$$

Обратим внимание, что решение ∂^k_y в пользу k-й гипотезы строится только по первичным признакам, определяющим k-ю ИМ $\mathcal{M}_k$ (в теореме 7.1 это соответствовало настройке на одну или

другую гипотезу), а по их первичным средним $\bar{M}_k g_{kj}$ находятся ошибки $\bar{\alpha}_k$, зависящие только от вида $\partial^k{}_y$.

После применения теоремы поиск оптимального правила сводится к нахождению коэффициентов $a^+{}_{kj}$, минимизирующих совокупную ошибку $\bar{\alpha}(\partial)=\bar{M}^k\bar{\alpha}_k(\partial)$, а в конечном счете, благодаря последней части теоремы, минимизирующих линейную форму от $a^+{}_{kj}$ при ограничении (7.7). Чем выше и шире как функция y каждое $\partial^k{}_y$, тем меньше будет ошибка $\bar{\alpha}_k$, поэтому совершенно ясно, что оптимальное правило будет стремиться увеличить сумму предпочтений $\sum\partial^k{}_y$ ближе к 1, делая тем самым как можно меньшим нейтральное решение $\partial^?{}_y$, и конечно же, равенство суммы 1 будет обязательно достигаться в каких-то точках y^*, называемых «горячими», в которых нейтральное решение отсутствует: $\partial^?{}_{y^*}=0$. Эти точки определят ошибки и вид правила.

Будь оптимальное правило построено, $\partial^k{}_y$ станут как можно широкими и дальше расширять их уже некуда, только лишь за счет «раздачи» нейтрального решения, что не меняет ни горячих точек, ни, следовательно, ошибок, поэтому все эти правила будут эквивалентными независимо от способа «раздачи».

Синтез детерминированных правил. Правило ∂_y, компоненты $\partial^k{}_y$ которого принимают только значения 0 или 1 (если одна — 1, то остальные — 0), называется детерминированным и эквивалентно детерминированной оценке $\hat{k}_y$ состояния k: $\partial^k{}_y=\delta_k(\hat{k}_y)$ (см. § 5.5). Риск оценки при дельта-потерях $\bar{M}\delta_k(\hat{k}_y)$ превращается в совокупную ошибку $\bar{\alpha}(\partial)$ правила различения гипотез. Детерминированные правила не являются оптимальными. Кроме случая, когда все $\mathcal{M}_k$, $k=1, \ldots, K$, представляют собой интервальные распределения вероятностей (что следует из индикаторной структуры первичных признаков ИРВ и теоремы 7.1).

Детерминированное правило эквивалентно разбиению пространства $\mathcal{Y}$ на непересекающиеся части $A_k=\{y:\partial^k{}_y=1\}$, $\sum\limits_k A_k=\mathcal{Y}$; при попадании наблюдений y в A_k принимается k-я гипотеза. Оптимальное (по крайней мере, в классе детерминированных) правило соответствует таким A_k, которые максимизируют $\underline{M\partial}=\underline{M}^k\underline{P}_k(A_k)$. Это может быть максимум по k стоящих справа нижних вероятностей правильных решений, или взвешенная их сумма $\sum\lambda_k\underline{P}_k(A_k)$. В последнем случае особенность оптимальных A_k, помогающая их нахождению, следующая: $\lambda_k\underline{P}_k(A_k)\geqslant\lambda_l\underline{P}_l(A_k)$, $\forall l$, k, т. е. взвешенная вероятность множества A_k при «своей», k-й гипотезе должна быть не меньше, чем при других.

Для ИРВ множества A_k должны составляться из первичных событий гипотез и строиться на объединенной (по $k=1, \ldots, K$) алгебре этих событий. Нужно отметить, что выбор оптимальных A_k, в общем, не является однозначным (намек на существование нейтральной области, уже как-то розданной по A_k). «Горячие» точки будут располагаться на границах областей и будут совершенно неподвижны: нейтральная область их не захватывает, что фиксирует ошибки.

Различение гипотез по заданным корреляциям. Рассмотрим и как иллюстрацию синтеза, и как важный для практики случай

задачу, где каждая из гипотез задана своими корреляциями, пусть точными:

$$\mathcal{M}_k: M_k(\mathbf{y}-\mathbf{w}_k)(\mathbf{y}-\mathbf{w}_k)^{\mathrm{T}}=\mathbf{B}_k,\ k=1,\ldots,K.$$

Здесь $\mathbf{y}$ и $\mathbf{w}_k$ — векторы-столбцы элементов y_i, $w_k(i)$, $i=1, \ldots, n$, а $\mathbf{B}_k$ — матрицы корреляций размерности $n\times n$. Таким образом, каждой гипотезе соответствует свой вектор сдвига $\mathbf{w}_k$ и заданная с учетом этого сдвига матрица корреляций. Критерием считаем минимум $\bar{\alpha}(\partial)=\sum\lambda_k\bar{\alpha}_k(\partial)$.

Определим, пользуясь теоремой 7.10, по признакам моделей форму оптимальных решений. Число первичных признаков для каждой отдельной гипотезы равно, очевидно, n^2 и согласно теореме 7.10 каждому из признаков должен быть приписан коэффициент, что в векторной форме выглядит следующим образом:

$$\partial^k_{\mathbf{y}}=[c_k-(\mathbf{y}-\mathbf{w}_k)^{\mathrm{T}}\mathbf{A}_k(\mathbf{y}-\mathbf{w}_k)]^+,\ k=1,\ldots,K, \tag{7.8}$$

$$\mathbf{M}_k\partial^k_y=1-\bar{\alpha}_k(\partial)=c_k-\sum_{k=1}^{K}\operatorname{tr}\mathbf{A}_k\mathbf{B}_k,$$

где $\bar{\alpha}_k(\partial)$ выписаны согласно последней части теоремы, а $\mathbf{A}_k$ и будут симметричными (по симметрии $\mathbf{B}_k$) матрицами неизвестных коэффициентов $A_k(i, j)$, выбор которых и составляет проблему синтеза.

Правило называется *контрастным*, если $\inf\limits_y\partial^k{}_y=0$, $\sup\limits_y\partial^k{}_y=1$, $\forall k$. Для (7.8) контрастность эквивалентна $c_k=1$: тогда каждое $\partial^k{}_{\mathbf{y}}$ достигает 1 при $\mathbf{y}=\mathbf{w}_k$ и совокупная ошибка запишется: $\bar{\alpha}(\partial)=\sum\lambda_k\operatorname{tr}\mathbf{A}_k\mathbf{B}_k$. Нетрудно видеть, что веса гипотез легко могут быть учтены умножением на них $\mathbf{B}_k$, поэтому допустимо (включив их в $\mathbf{B}_k$) считать далее $\lambda_k\equiv\mathrm{const}$ и минимизировать суммарную ошибку $\sum\operatorname{tr}\mathbf{A}_k\mathbf{B}_k$.

Контрастные правила оказываются достаточными при определенном удалении гипотез друг от друга (чтобы отойти от тривиального случая $\partial^k{}_{\mathbf{y}}\equiv c_k$). Как функции $\mathbf{y}$ они представляют собой совокупность параболоидов $\partial^k{}_{\mathbf{y}}$ с вершинами в точках $\mathbf{y}=\mathbf{w}_k$ высотой $\partial^k{}_{\mathbf{w}_k}=1$. Для наглядности интерпретации удобно мыслить $n=2$. Основаниями параболоидов, где они пересекают координатную плоскость, будут эллипсы (см. рис. 7.8) с серединами в $\mathbf{w}_k$. Чем шире параболоид k-й гипотезы, тем шире будет эллипс и тем меньше будет ошибка $\bar{\alpha}_k(\partial)$. Но ширина ограничивается требованием (7.7), соответствующим тому, что сумма параболоидов не должна нигде превышать 1. Контролировать это требование достаточно лишь в располагающихся где-то между $\mathbf{w}_k$ «горячих» точках, обозначенных на рис. 7.8 звездочками. Для оптимального правила, поскольку все $\partial^k{}_{\mathbf{y}}$ делаются как можно шире, «горячих» точек должно быть максимальное число, причем они будут располагаться где-то в серединах групп в разных сочетаниях векторов $\mathbf{w}_k$ (причем теми, для которых пересечение оснований $\partial^k{}_{\mathbf{y}}$ ненулевое).

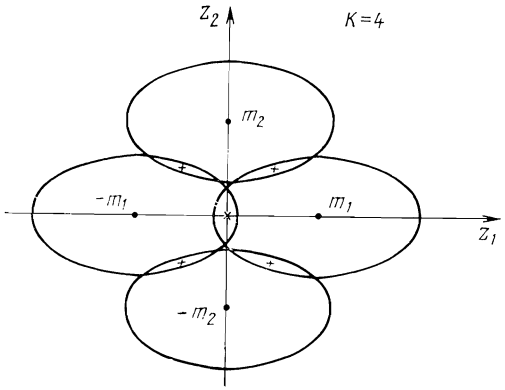


Рис. 7.8. Картина расположения «горячих» точек

Кроме «горячих» точек и центров $\mathbf{w}_k$ везде сумма параболоидов предпочтений окажется строго меньше 1. Наиболее это выражено в зоне на рис. 7.8, лежащей вне эллипсов, где решение полностью нейтральное: $\partial^2_{\mathbf{y}}=1$. «Раздача» нейтрального решения, хотя, в общем, и произвольная, но подчиняется разумным началам: например, все содержащееся в $\partial^2_{\mathbf{y}}$ предпочтение удобно отдать близлежащей гипотезе. Другой возможный путь: ввести во все матрицы $\mathbf{A}_k$ один постоянный множитель a и для каждого $\mathbf{y}$ подыскивать такое $a_{\mathbf{y}}$, при котором сумма предпочтений станет равной 1. Возможны иные варианты.

Оптимальность решений определяется формой лежащих в основаниях эллипсов, минимизирующих суммарную ошибку. Оси эллипсов и их ширина ставятся в зависимость от расположения $\mathbf{w}_k$ и вида $\mathbf{B}_k$ и управляются искомыми матрицами $\mathbf{A}_k$. Перейдем к наиболее простому случаю $K=2$.

Оптимальное правило различения двух гипотез. Пусть $K=2$ и будем искать оптимальное правило вида (7.8), т. е. искать матрицы $\mathbf{A}_1$ и $\mathbf{A}_2$, минимизирующие суммарную ошибку $\mathrm{tr}\,\mathbf{A}_1\mathbf{B}_1+\mathrm{tr}\,\mathbf{A}_2\mathbf{B}_2$. Требование $\partial^1_{\mathbf{y}}+\partial^2_{\mathbf{y}}\leqslant 1$ должно контролироваться в области ненулевых значений $\partial^1_{\mathbf{y}}$ и $\partial^2_{\mathbf{y}}$, где запишется как ограничение сверху единицей суммы двух параболоидов $2-(\mathbf{y}-\mathbf{w}_1)^{\mathrm{T}}\times\mathbf{A}_1(\mathbf{y}-\mathbf{w}_1)-(\mathbf{y}-\mathbf{w}_2)^{\mathrm{T}}\mathbf{A}_2(\mathbf{y}-\mathbf{w}_2)$ (квадратичных форм $\mathbf{y}$). Максимум по $\mathbf{y}$ достигается в «горячей» точке $\mathbf{y}_*$, получающейся решением уравнения $(\mathbf{A}_1+\mathbf{A}_2)\mathbf{y}_*=\mathbf{A}_1\mathbf{w}_1+\mathbf{A}_2\mathbf{w}_2$; в результате подстановки $\mathbf{y}^*$ требование запишется:

$$2-\mathbf{w}_1^{\mathrm{T}}\mathbf{A}_1\mathbf{w}_1-\mathbf{w}_2^{\mathrm{T}}\mathbf{A}_2\mathbf{w}_2+(\mathbf{A}_1\mathbf{w}_1+\mathbf{A}_2\mathbf{w}_2)^{\mathrm{T}}(\mathbf{A}_1+\mathbf{A}_2)^{-}(\mathbf{A}_1\mathbf{w}_1+$$
$$+\mathbf{A}_2\mathbf{w}_2)\leqslant 1,$$

или же после некоторых преобразований перепишется:

$$(\mathbf{w}_2-\mathbf{w}_1)^{\mathrm{T}}\mathbf{A}_1^{\mathrm{T}}(\mathbf{A}_1+\mathbf{A}_2)^{-}\mathbf{A}_2(\mathbf{w}_2-\mathbf{w}_1)\geqslant 1. \qquad (7.9)$$

Здесь минус в показателе означает, что матрица псевдообратная [1],

[1] Гантмахер Ф. Р. Теория матриц. — М.: Наука, 1966. — С. 34.

а именно, такая, что $\mathbf{A}\mathbf{A}^{-}\mathbf{A}=\mathbf{A}$ (обратная матрица в подпространстве собственных векторов, если она особая).

Сказанное важно вот почему. Минимизация $\mathrm{tr}\,\mathbf{A}_1\mathbf{B}_1+\mathrm{tr}\,\mathbf{A}_2\mathbf{B}_2$ при условии (7.9), замененном на равенство, производится методом множителей Лагранжа и ведет к матричным уравнениям

$$\gamma(\mathbf{A}_1+\mathbf{A}_2)\mathbf{B}_1(\mathbf{A}_1+\mathbf{A}_2)=\mathbf{A}_2\mathbf{w}^{\mathrm{T}}\mathbf{w}\mathbf{A}_2,\ \mathbf{w}=\mathbf{w}_2-\mathbf{w}_1,$$
$$\gamma(\mathbf{A}_1+\mathbf{A}_2)\mathbf{B}_2(\mathbf{A}_1+\mathbf{A}_2)=\mathbf{A}_1\mathbf{w}^{\mathrm{T}}\mathbf{w}\mathbf{A}_1,$$

в которых γ находится из (7.9). Отсюда, так как правая часть имеет ранг 1, сразу же следует вывод, что матрицы $\mathbf{A}_1$ и $\mathbf{A}_2$ также должны иметь ранг 1 и записываться в виде: $\mathbf{A}_k=a_k\mathbf{g}\mathbf{g}^{\mathrm{T}}$, $k=1,2$, где $\mathbf{g}$ — вектор-столбец. Тогда обратной матрицы не существует, а лишь псевдообратная: $(\mathbf{A}_1+\mathbf{A}_2)^{-}=\mathbf{g}\mathbf{g}^{\mathrm{T}}/(a_1+a_2)$. Обозначая $b_k=\mathbf{g}^{\mathrm{T}}\mathbf{B}_k\mathbf{g}$, $m_k=\mathbf{g}^{\mathrm{T}}\mathbf{w}_k$, сводим матричные уравнения вместе с (7.9) к системе скалярных уравнений:

$$m^2 a_1 a_2=a_1+a_2,\qquad m=m_2-m_1,$$
$$\gamma(a_1+a_2)^2 b_1=a_2^2 m^2,$$
$$\gamma(a_1+a_2)^2 b_2=a_1^2 m^2.$$

Решая их, находим оптимальные $a_1=(1+\sqrt{b_2/b_1})/m^2$, $a_2=(1+\sqrt{b_1/b_2})/m^2$.

В итоге подстановки найденных a_1 и a_2 правило запишется

$$\partial^1_{\mathbf{y}}=[1-(1+\sqrt{b_2/b_1})(\mathbf{g}^{\mathrm{T}}(\mathbf{y}-\mathbf{w}_1))^2/(\mathbf{g}^{\mathrm{T}}\mathbf{w})^2]+,$$

а $\partial^2_{\mathbf{y}}$ есть либо $1-\partial^1_{\mathbf{y}}$ (когда нейтральное решение отдается в пользу второй гипотезы), либо дается аналогичным записанному выражением с переменой индексов 1 и 2 местами. Суммарная ошибка правила будет равна

$$\bar{\alpha}_1+\bar{\alpha}_2=(\sqrt{\mathbf{g}^{\mathrm{T}}\mathbf{B}_1\mathbf{g}}+\sqrt{\mathbf{g}^{\mathrm{T}}\mathbf{B}_2\mathbf{g}})^2/(\mathbf{g}^{\mathrm{T}}\mathbf{w})^2.$$

Остался последний шаг — найти вектор линейной обработки $\mathbf{g}$ (на который проектируется $\mathbf{y}$), минимизирующий выписанную суммарную ошибку. Это сделать несложно при одинаковых корреляциях: $\mathbf{B}_1=\mathbf{B}_2=\mathbf{B}$, тогда $\mathbf{g}$ минимизирует отношение $\mathbf{g}^{\mathrm{T}}\mathbf{B}\mathbf{g}/(\mathbf{g}^{\mathrm{T}}\mathbf{w})^2$ и равен $\mathbf{g}=\mathbf{B}^{-1}\mathbf{w}$, $\mathbf{w}=\mathbf{w}_2-\mathbf{w}_1$.

Более двух гипотез. Пусть $K>2$ и нужно найти правило вида (7.8), минимизирующее суммарную ошибку. Требование (7.7) подстановкой туда (7.8) (конкретнее, суммированием содержимого квадратных скобок (7.8) и взятием максимума по $\mathbf{y}$, который достигается при $\sum\mathbf{A}_k\mathbf{y}=\sum\mathbf{A}_k\mathbf{w}_k$), обращается в систему неравенств:

$$\sum\mathbf{w}_k^{\mathrm{T}}\mathbf{A}_k\mathbf{w}_k-(\sum\mathbf{A}_k\mathbf{w}_k)^{\mathrm{T}}(\sum\mathbf{A}_k)^{-}(\sum\mathbf{A}_k\mathbf{w}_k)\geqslant(\sum c_k)-1, \qquad (7.10)$$

где суммы перебираются для любых сочетаний индексов в числе от двух (аналогично (7.9)) и вплоть до K. Причем для каких-то сочетаний будут равенства, соответствующие каждое своей «го-

рячей» точке. Само неравенство в (7.10) имеет в виду, что не между всеми сочетаниями $\mathbf{w}_k$ способны располагаться «горячие» точки (так, если все $\mathbf{w}_k$ стоят на одной прямой в ряд, то равенства будут лишь для смежных пар).

Облегчение синтезу дает следующее общее утверждение, являющееся аналогом теоремы 6.5.

Утверждение 7.11. *Пусть гипотезы симметричны между собой в том смысле, что существует преобразование наблюдений* $\mathbf{V}$, *оставляющее задачу на месте, меняя гипотезы между собой:* $\mathcal{M}_k^{\mathbf{Vy}} = \mathcal{M}_{vk}^{\mathbf{y}}$, $k=1, \ldots, K$, *где* vk *осуществляет перестановку индексов* k. *Тогда минимизирующее суммарную ошибку правило эквивариантно:* $\partial_{\mathbf{Vy}}^{k} = \partial_{\mathbf{y}}^{vk}$.

Если $\mathbf{V}$ — удовлетворяющее утверждению преобразование, то последовательное применение $\mathbf{V}^l = \mathbf{V} \ldots \mathbf{V}$ будет также удовлетворять ему, поэтому $\mathbf{V}^l$ образуют группу преобразований. Смысл утверждения 7.11 в том, что, построив какую-то одну из составляющих решающего правила, скажем, $\partial^1_{\mathbf{y}}$, мы переносим ее на другие составляющие $\partial^k_{\mathbf{y}}$, где $k = v^l 1$, подставляя $\mathbf{V}^l\mathbf{y}$ в $\partial^1_{\mathbf{y}}$ на место $\mathbf{y}$. Если $v^l 1$ при изменении l пробегает все $k=1, \ldots, K$, то оказывается достаточным построить всего одну составляющую правила, перенося ее преобразованием $\mathbf{V}^l\mathbf{y}$ на все остальные.

Для правила (7.8), если имеется $\mathbf{V}$, циклически переставляющее гипотезы $1\to2$, $2\to3$, ..., $K\to1$, сказанное соответствует: $\mathbf{A}_k = (\mathbf{V}^k)^{\mathrm{T}}\mathbf{A}_1\mathbf{V}^k$.

Перейдем к рассмотрению примеров синтеза.

Три ортогональных сигнала. Пусть $K=3$ и в векторной записи $\mathbf{y} = \mathbf{w}_k + \boldsymbol{\xi}$, $k=1, 2, 3$; «шум» $\boldsymbol{\xi}$ есть однородный некоррелированный $M\xi_i\xi_j = \sigma^2\delta_{ij}$ процесс, а «сигналы» $\mathbf{w}_k$ ортогональны между собой: $\mathbf{w}^{\mathrm{T}}_k\mathbf{w}_l = m^2\delta_{kl}$, m — их «амплитуда». Задача симметрична к перестановке гипотез между собой, причем плоскость, натянутая на векторы $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$ в $\mathcal{R}^n$, вращаясь, остается на месте (см. рис. 7.9). По этой причине и в силу инвариантности свойств шума к таким вращениям оптимальные $\mathbf{A}_k$ должны проектировать $\mathbf{y}$ на плоскость (z_1, z_2), где и достаточно решать задачу синтеза, направив новые оси согласно рис. 7.9:

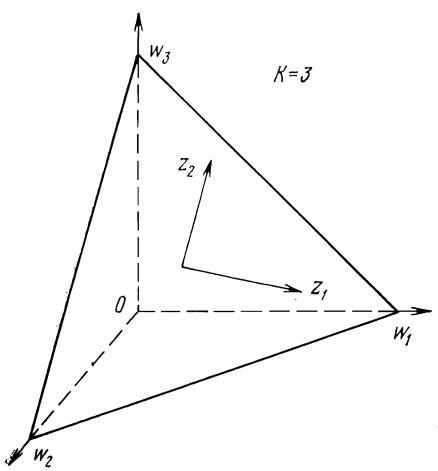


Рис. 7.9. Преобразования координат

$$z_1 = \mathbf{y}^{\mathrm{T}}(2\mathbf{w}_1 - \mathbf{w}_2 - \mathbf{w}_3)/(2m),$$

$$z_2 = \sqrt{3}\,\mathbf{y}^{\mathrm{T}}(\mathbf{w}_3 - \mathbf{w}_2)/(2m).$$

В новых координатах $\mathbf{z} = (z_1, z_2)^{\mathrm{T}}$ исходная задача переписывается:

$$\mathbf{z} = \mathbf{m}_k + \boldsymbol{\zeta},$$

$$\mathbf{m}_1 = (m, 0)^{\mathrm{T}},\ \mathbf{m}_2 = (-m/2, \sqrt{3}\,m/2)^{\mathrm{T}},\ \mathbf{m}_3 = (-m/2, -\sqrt{3}\,m/2)^{\mathrm{T}},$$

$$\boldsymbol{\zeta} = (\zeta_1, \zeta_2)^{\mathrm{T}},\ M\zeta_i\zeta_j = b\delta_{ij},\ b = 3\sigma^2/2.$$

Перестановка гипотез $1\to2$, $2\to3$, $3\to1$ соответствует вращению $\mathbf{z}$ на 120° по часовой стрелке. Обозначим $\mathbf{V}$ — унитарную матрицу вращения. Ее элементами будут: $V_{11} = V_{22} = -1/2$, $V_{21} = -V_{12} = \sqrt{3/2}$. Ищем матрицы $\breve{\mathbf{A}}_k$ в правиле вида $\partial^k\mathbf{z} = [1 - (\mathbf{z} - \mathbf{m}_k)\breve{\mathbf{A}}_k(\mathbf{z} - \mathbf{m}_k)]^+$, минимизирующие суммарную ошибку, равную $\sum_k b[\breve{A}_k(1,1) + \breve{A}_k(2,2)]$. Согласно утверждению 7.11 имеем: $\breve{\mathbf{A}}_2 = \mathbf{V}^{\mathrm{T}}\breve{\mathbf{A}}_1\mathbf{V}$, $\breve{\mathbf{A}}_3 = \mathbf{V}\breve{\mathbf{A}}_1\mathbf{V}^{\mathrm{T}}$ (где использовано $\mathbf{V}^{\mathrm{T}} = \mathbf{V}^2$), так что задача сводится к нахождению матрицы $\breve{\mathbf{A}}_1 = \breve{\mathbf{A}}$, определяющей $\partial^1\mathbf{z}$. В силу симметрии положения $\mathbf{m}_2$ и $\mathbf{m}_3$ по отношению к осям z_1, z_2 матрица $\breve{\mathbf{A}}$ должна быть диагональной: $\breve{A}(1,2) = \breve{A}(2,1) = 0$, откуда $\partial^1\mathbf{z} = [1 - \breve{A}(1,1)(z_1 - m)^2 - \breve{A}(2,2)z_2^2]^+$.

«Горячих» точек для $\partial^1\mathbf{z}$ будет три: одна — в начале новой координатной системы $z_1 = z_2 = 0$, относительно этой точки совершается вращение $\mathbf{V}$ и потому по симметрии $\partial^1{}_0 = \partial^2{}_0 = \partial^3{}_0 = 1 - \breve{A}(1,1)m^2$, откуда из требования $\partial^1{}_0 + \partial^2{}_0 + \partial^3{}_0 = 3 - 3\breve{A}(1,1)m^2 = 1$ следует $\breve{A}(1,1) = 2/(3m^2)$. Другие две «горячие» точки располагаются на серединах между парами $\mathbf{m}_1$, $\mathbf{m}_2$ и $\mathbf{m}_1$, $\mathbf{m}_3$ и дают одно уравнение $\mathbf{m}^{\mathrm{T}}_1(\mathbf{V} - \mathbf{I})^{\mathrm{T}}[(\mathbf{V}^{\mathrm{T}}\breve{\mathbf{A}}\mathbf{V})^{-1} + \breve{\mathbf{A}}^{-1}]^{-1}(\mathbf{V} - \mathbf{I})\mathbf{m}_1 = 1$, из которого находится $\breve{A}(2,2) = 2/(3m^2)$.

Таким образом, на плоскости (z_1, z_2) основаниями $\partial^1\mathbf{z}$, $\partial^2\mathbf{z} = \partial^1{}_{\mathbf{V}^{\mathrm{T}}\mathbf{z}}$, $\partial^3\mathbf{z} = \partial^1{}_{\mathbf{V}^{\mathrm{T}}\mathbf{z}}$ будут круги радиуса $\sqrt{3/2}\,m$ с центрами в $\mathbf{m}_1$, $\mathbf{m}_2$ и $\mathbf{m}_3$.

Четыре попарно ортогональных противоположных сигнала. Пусть $K=4$, $\mathbf{y} = \mathbf{w}_k + \boldsymbol{\xi}$, $M\xi_k\xi_l = \sigma^2\delta_{kl}$, $k, l = 1, \ldots, 4$, а «сигналы» попарно принимают противоположные значения $\mathbf{w}_3 = -\mathbf{w}_1$, $\mathbf{w}_4 = -\mathbf{w}_2$ и ортогональны в двух направлениях: $\mathbf{w}^{\mathrm{T}}_1\mathbf{w}_2 = 0$, $\|\mathbf{w}_1\| = m_1$, $\|\mathbf{w}_2\| = m_2$, имеют в них разные, в общем, «амплитуды» m_1 и m_2. Действует только один сигнал $\mathbf{w}_k$, нужно узнать какой, наблюдая $\mathbf{y}$. Очевидно, достаточной является проекция задачи на плоскость $z_1 = \mathbf{y}^{\mathrm{T}}\mathbf{w}_1/m$, $z_2 = \mathbf{y}^{\mathrm{T}}\mathbf{w}_2/m$ и на этой плоскости задача выглядит так: $\mathbf{z} = \mathbf{m}_k + \boldsymbol{\zeta}$, $\mathbf{m}_1 = -\mathbf{m}_3 = (m_1, 0)^{\mathrm{T}}$, $\mathbf{m}_2 = -\mathbf{m}_4 = (0, m_2)^{\mathrm{T}}$, $M\boldsymbol{\zeta}\boldsymbol{\zeta}^{\mathrm{T}} = \sigma^2\mathbf{I}$. Ищем, как и в предыдущей задаче, матрицы $\breve{\mathbf{A}}_k$, $k=1, \ldots, 4$, которые в силу симметрии задачи к перемене знака $\pm z_1$, $\pm z_2$ будут попарно одинаковыми и диаго-

нальными; обозначим их диагональные элементы соответственно $\breve{A}_1(j,\ j)=\breve{A}_3(j,\ j)=a_1(j)$, $\breve{A}_2(j,\ j)=\breve{A}_4(j,\ j)=a_2(j)$, $j=1,\ 2$. Будем их искать исходя из минимума суммы ошибок $2\sigma^2[a_1(1)+a_1(2)+a_2(1)+a_2(2)]$ при требовании (7.10), которое воплотится в три неравенства:

$$1)\ m_1^2\frac{a_1(1)\,a_2(1)}{a_1(1)+a_2(1)}+m_2^2\frac{a_1(2)\,a_2(2)}{a_1(2)+a_2(2)}\geqslant 1,$$

$$2)\ 2m_1^2\,a_1(1)\geqslant 1,\ 3)\ 2m_2^2\,a_2(2)\geqslant 1.$$

Пусть для определенности $m^2{}_2\geqslant m^2{}_1$. Тогда картину расположения «горячих» точек дает рис. 7.8, а в 1) и 2) будут равенства. По симметрии равенств и из 2) находится: $a_1(1)=a_2(1)=1/(2m^2{}_1)$, и далее подстановкой в 1): $a_1(2)=a_2(2)=3/(2m^2{}_2)$. Получаем следующее оптимальное правило:

$$\partial^1_{\mathbf{z}}=[1-(z_1-m_1)^2/(2m_1^2)-3z_2^2/(2m_2^2)]^+,$$

$$\partial^2_{\mathbf{z}}=[1-z_1^2/(2m_1^2)-3(z_2-m_2)^2/(2m_2^2)]^+,$$

а $\partial^3{}_{\mathbf{z}}$ отличается от $\partial^1{}_{\mathbf{z}}$ как и $\partial^4{}_{\mathbf{z}}$ от $\partial^2{}_{\mathbf{z}}$ переменой знака у m_j. Ошибки будут одинаковы: $\bar{\alpha}_k=\bar{\alpha}=\sigma^2/(2m^2{}_1)+3\sigma^2/(2m^2{}_2)$.

Замечания: 1. Разные значения дисперсий $M\zeta^2{}_j=M(\boldsymbol{\xi}^{\mathrm{T}}\mathbf{w}_j)^2=\sigma^2{}_j$, $j=1,\ 2$, оставляют тот же вид оптимального правила, если $m^2{}_2/\sigma^2{}_2\geqslant m^2{}_1/\sigma^2{}_1$ (вместо $m^2{}_2\geqslant m^2{}_1$), при этом $\bar{\alpha}=\sigma^2{}_1/(2m^2{}_1)+3\sigma^2{}_2/(2m^2{}_2)$.

2. Задача: $\mathbf{y}=\pm\mathbf{w}_j+\boldsymbol{\xi}$, $M\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}=\mathbf{B}$, $\mathbf{w}_j\mathbf{B}^{-1}\mathbf{w}_l=\delta_{jl}m^2{}_j/\sigma^2{}_j$, $j,\ l=1,\ 2$, где шум коррелирован, а $\mathbf{w}_j$ совпадают с направлениями собственных векторов $\mathbf{B}$ и $\sigma^2{}_j$ — собственные числа, эквивалентна предыдущей.

3. При равных амплитудах сигналов $m_1=m_2=m$ и $\sigma^2{}_1=\sigma^2{}_2=\sigma^2$ оптимальными будут значения $a_l(j)=1/m^2$. Основаниями оптимальных $\partial^k{}_{\mathbf{z}}$ станут круги (на подобие олимпийской эмблемы) и $\bar{\alpha}_k=2\sigma^2/m^2$.

4. Совершенно прост переход к непрерывному времени: векторные произведения заменяются на интегралы.

Система ортогонально-противоположных сигналов равной амплитуды. Обобщим предыдущее правило на случай любого четного числа сигналов, разбивающихся на пары сигналов, внутри каждой пары противоположных, а между парами — ортогональных, все одинаковой амплитуды:

$$\mathbf{y}=\mathbf{w}_k+\boldsymbol{\xi},\ k=1,\ldots,2K;\ \mathbf{w}_{K+j}=-\mathbf{w}_j,$$

$$\mathbf{w}_j^{\mathrm{T}}\mathbf{w}_l=m^2\,\delta_{jl},\ j=1,\ldots,K;\ M\,\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}=\sigma^2\,\mathbf{I}.$$

По аналогии с замечанием 3 оптимальным правилом будет

$$\partial^k_{\mathbf{z}}=[1-(z_k-m)^2/m^2-\sum_{j\neq k}z_j^2/m^2]^+;\ z_j=\mathbf{y}^{\mathrm{T}}\mathbf{w}_j;$$

$$\partial^{K+j}_{\mathbf{z}}=\partial^j_{-\mathbf{z}};\ \bar{\alpha}_k=K\sigma^2/m^2;\ j,\ k=1,\ldots,K.$$

Написанное осмыслено лишь в том случае, если $\bar{\alpha}_k\leqslant(2K-1)/2K$,

иначе оптимальным делается тривиальное правило $\partial^k\equiv 1/2K$ с $\bar{\alpha}_k=(2K-1)/2K$.

Неточно известные корреляции. Пусть корреляционные матрицы $\mathbf{B}_k$, определяющие $\mathcal{M}_k$, неточно известны, т. е. заданы приближенно в виде оценочных границ $\underline{B}_k(i.\ j)$, $\bar{B}_k(i,\ j)$ (это могут быть и доверительные границы, полученные по обучающему эксперименту). Для каких-то $i,\ j$ этих данных о границах, впрочем, может и не быть совсем, что лишь способно упростить задачу, так как по теореме 7.10 о достаточности для этих $i,\ j$ следует положить $A_k(i,\ j)=0$.

Пример 7.7. Пусть $K=2$ и заданы лишь диагональные элементы матриц корреляций $B_1(i,\ i)$, $B_2(i,\ i)$, $i=1,\ \ldots,\ n$, а относительно взаимных корреляций сведений нет. Тогда $A_k(i,\ j)=0$, $i\neq j$, а матрицы $\mathbf{A}_k$ станут диагональными. Их ранг должен быть равен 1 (аналогично вышеизложенному при $K=2$), что может быть, лишь когда все диагональные элементы матриц $\mathbf{A}_k$ нулевые, кроме одного. Сказанное эквивалентно тому, что из наблюдений $y_1,\ \ldots,\ y_n$ выбирается всего один элемент y_j, тот самый, для которого минимальной будет суммарная ошибка построенного по нему правила, равная (как это было показано при рассмотрении двух гипотез) величине: $\bar{\alpha}_1+\bar{\alpha}_2=(\sqrt{\bar{B}_1(j,\ j)}+\sqrt{\bar{B}_2(j,\ j)})/(w_2(j)-w_1(j))^2$. Все остальные наблюдения можно «забыть». Причина в том, что неизвестность взаимных корреляций вынуждает не исключать (в пессимистическом режиме $\varkappa=1$) случай, когда все элементы y_i повторяют друг друга, тогда остальные y_i ничего нового по сравнению с одним значением не несут.

Другой путь синтеза правил при неточных корреляциях базируется на том, что недоопределение корреляций, как и задание их границами, формирует собственные семейства $\mathfrak{B}_k$ корреляционных функций (матриц). Ошибки, очевидно, будут равны $\bar{\alpha}_k=\sup_{\mathbf{B}_k\in\mathfrak{B}_k}\operatorname{tr}\mathbf{A}_k\mathbf{B}_k$ и поиск оптимального правила сводится к минимизации суммы ошибок, т. е. к минимаксной задаче при ограничениях.

7.7. ЗАКЛЮЧЕНИЕ

Рассматриваются две гипотезы, нулевая и альтернативная, отличающиеся разными статистическими описаниями наблюдений в виде интервальных моделей средних (ИМ). За гипотезами стоят конкретные практические задачи типа есть сигнал или нет его при обнаружении, либо проверка неисправности устройства, соответствия его техническим требованиям и т. п. Назначение решающего правила в том, чтобы по результатам наблюдений сделать выбор в пользу одной из гипотез. И не обязателен конкретный выбор, а вполне допускается расплывчатый, при котором решения подаются неоднозначными в форме предпочтений (реализуемых рандомизацией).

Характеризуются правила проверки гипотез вероятностями ошибочного принятия одной, когда верна другая. Ошибки двух: первого рода, состоящая в неправильном отклонении нулевой гипотезы, и второго. Вероятности ошибок

находятся продолжением первичных средних ИМ на решающие правила (рассматриваемые как признаки). Итогом будут интервальные значения каждой вероятности: нижняя ее граница, дающая самые оптимистичные прогнозы на ошибку, и верхняя — пессимистичные. По ним выводятся промежуточные значения, зависящие от степени пессимизма.

Оптимальным для проверки нулевой гипотезы называется правило, минимизирующее величину ошибки второго рода при заданной первого (уровне правила). Хотя по определению приоритет отдается в пользу нулевой гипотезы, синтез эквивалентен решению смежной задачи минимизации взвешенной суммы ошибок с последующим подбором весов (заменяющих априорные вероятности гипотез). В этом плане определяется достаточность.

Центральный результат формулируется теоремой 7.1 и состоит в том, что достаточные при пессимизме классы правил, а следовательно, структура оптимального правила определяются исключительно линейными комбинациями первичных признаков гипотез. Остается отыскать коэффициенты, которых будет тем меньше, чем проще гипотеза в смысле их признакового состава. При этом вид правила может нацеливаться только на одну из гипотез, либо нулевую, либо альтернативную, причем первый случай предпочтительнее из-за простоты фиксации уровня. Симметрия и однородность гипотез дополнительно упрощает задачу.

Особенности оптимальных правил состоят, первая, в их расплывчатости (рандомизированности), характер которой всецело определяется формой первичных признаков гипотезы (либо альтернативы). И вторая особеность, являющаяся побочным фактором произвольности настройки на гипотезу, состоит в неоднозначности вида. Обе особенности обостряются при бедном исходном материале, составляющем гипотезы, и исчезают при переходе к «богатым» моделям в виде распределений вероятностей, где рандомизация может остаться лишь на границе по известной лемме Неймана — Пирсона.

Положения теории раскрываются нахождением в § 7.2 оптимальных правил при сдвигах наблюдений и известных корреляционных свойствах, сопутствующих гипотезам (как ни странно, такая нужная задача в классическом аппарате не имеет решения). Правила получаются расплывчатыми, характер их, как это следует из квадратичной формы первичных признаков, является параболическим.

Трудности определения ошибок правил толкают на поиски других способов синтеза, использующих пройденные пути. Один из них (§ 7.3), проторенный доверительными оценками предыдущей главы, требует записи гипотез через два разных значения одного и того же параметра. Правило состоит в отведении нулевой гипотезе той степени предпочтения, какое доверительная оценка дает соответствующему этой гипотезе значению параметра. Расплывчатость оценки породит расплывчатость правила, а величина ошибки оценки переходит в уровень правила и открывает дорогу прикидочному расчету ошибки второго рода. Так образуется сразу большое число правил, может быть не совсем оптимальных, но все же хороших по их «гинетической» близости к оценкам, если те взяты наилучшими.

Другой путь (§ 7.4) уже традиционный и состоит в интерпретации моделей как семейств точных распределений вероятностей, поиска внутри семейств наименее благоприятных и сравнения их отношения с порогом по предписанию известной леммы Неймана — Пирсона. Путь привлекает методы теории игр и охватывается робастным подходом. Автор приводит свои результаты по интервальным плотностям для демонстрации сравнительных возможностей робастного подхода внутри общих интервальных построений.

Качественно другой характер гипотезы приобретают, когда нужно принять решение в виде согласия с выдвинутым (гипотетическим) положением и не согласия. Например, исправен прибор или нет; конкретизируется гипотеза, а альтернативой будет все остальное, правда, не ясно что. Образно говоря, если гипотеза и конкретная альтернатива составляют направленный диполь, то тут он оказывается неориентированным, но закрепленным со стороны гипотезы. Постановка и методы подхода к этой задаче содержатся в § 7.5, где предлагается в качестве возможных путей использовать найденные ранее правила проверки гипотез и доверительные оценки.

Задача различения гипотез § 7.6 возникает при необходимости разделить или различить между собой несколько состояний объекта. Гипотезы формулируются в терминах ИМ, их первичные признаки определяют собой структуру оптимального правила (теорема 7.10), минимизирующего суммарную ошибку. Его неоднозначность учитывается введением нейтрального решения. Конкретные правила получены для гипотез, заданных корреляционными свойствами наблюдений. Проблема различения многих состояний приближает нас к их оцениванию и даже может решаться методами теории оценивания, вопрос весь упирается в критерии анализа качества и способы формирования риска.

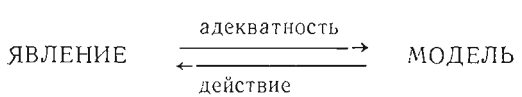
Глава 8.

НАДЕЖНОСТНЫЙ СИНТЕЗ

8.1. ОБЩИЕ ВОПРОСЫ СИНТЕЗА МОДЕЛЕЙ

Методология синтеза моделей. Наше житие можно сравнить с движением в купе скорого поезда, из окна которого проносятся мимо быстро сменяющие друг друга пейзажи. И нет ни малейшего времени задуматься, что каждый его фрагмент «дышит» собственной очень сложной и содержательной жизнью, детальное изучение закономерностей которой — удел многих и многих поколений (если не всех). Мы же, глядя в окно, ограничиваемся самым поверхностным представлением обо всем этом, внешней моделью, связывающей увиденное с нашими внутренними воззрениями, опытом. Это естественно, потому что здравый опыт в своих притязаниях и проявлениях суть экономное представление окружающего: выделение главного, игнорирование всего второстепенного (хотя детям, красивым девушкам и некоторым ученым иногда свойственна прямо противоположная тенденция).

Образный пример не является случайным, а объясняет характер непреходящей гносеологической связи:



Человек строит модели как отражения реальностей (причем на самой разной основе) и использует их для познания природы, а также воздействия на нее (поэтому-то отношение стрелками указано в ту и другую сторону). Как и природа, модель живет своей самостоятельной жизнью. Обе жизни родственны в том смысле, что они должны быть слаженными, как говорят, модель должна быть адекватной явлению (верхняя стрелка), что и позволит по состояниям модели прогнозировать, а затем и управлять состояниями явления (нижняя стрелка).

Прекрасной иллюстрацией сказанному является случайный процесс броуновского движения — классическая модель перемещения частицы при хаотических столкновениях ее с молекулами при тепловом движении. Наглядная физическая картина здесь породила математический образ, вероятностную модель, в которой как таковых частичек уже нет, про них забыли, а только отжатый результат — процесс перемещения в виде математического построения. Такая абстрактизация позволила найти вероятности уклонений, установить идеальные законы броуновского движения.

Вообще, нужно быть осторожным, так как на языке моделей удобным и отразимым оказывается далеко не все, что существует в природе. Идеализация приводит к моделям, надежным только на первый взгляд, отражающим желаемую для модели картину эксперимента. В дальнейшем, удобно в силу специфики настоящего изложения говорить только о математико-вероятностных моделях — символьного языка описания случайных явлений.

Глубина и безграничная сложность явлений реального мира вынуждает, казалось бы, такие же качества у моделей. Так и кажется подчас, что чем сложнее модель, богаче ее собственная жизнь, тем более сильной в своей отражательной потенции она является. Можно и согласиться, если бы при этом не разрушалась прямая связь модели с явлением. Усложнение модели требует настоятельной проверки каждого ее нового фрагмента на соответствие действительности. Словами «пусть» (столь привычными в современных математико-вероятностных изложениях: пусть процесс марковский, пусть плотность существует и дважды дифференцируема, пусть известна вероятность и пр. и пр.) достигается обособление жизни модели, превращается в самоцель ее изучение математиками, считающими себя совершенно чистыми.

Можно возразить, что история знает примеры, такие как теория групп или неевклидовая геометрия, когда то или иное сугубо математическое построение встречалось в конечном счете с практическими приложениями. Но рассчитывать каждый раз на случай — все равно, что делать ставку при покидании космического объекта на удар в него метеорита. Не надежнее ли нацелить в объект что-то управляемое во времени в трехкоординатном пространстве?

Мы убеждаемся здесь еще раз в нашей главной мысли, породившей и пронизывающей все содержание книги, что не нужны

оторванные суперсложные модели, нюансы которых кто-нибудь из добросовестных исследователей может воспринять всерьез как изведанные законы природы. Значительно экономнее и надежнее иметь арсенал простых моделей, связывающихся с явлением в небольшом числе сторон, своего рода каналов, и отражать явление не сразу все, а по частям, освещая каждый раз только ту сторону, которая представляет непосредственный интерес. В том-то сила и искусство познания: расчленить всю сферу деятельности на науки, каждой из которых отдается своя часть физического явления, затем науку — на предметные области и т. д.

Конкретизируем теперь связь: МАТЕМАТИЧЕСКАЯ ТЕОРИЯ → МОДЕЛЬ. Математическая теория развивает групповые законы моделей и организуется внутри себя системой аксиом — этой формальной конструкцией, сцепляющей каждую модель. На символьном материале аксиоматические связи должны повторять реальные, что и есть адекватность аксиом и гарант правомочности теории. А жизненность теории будет зависеть от того, какими сторонами ее представители (модели) связываются с явлениями, насколько эти связи легко наводятся, доступны, физически наглядны (интерпретируемы), привычны, наконец надежны. Это и поможет инженеру-исследователю выбрать ту конкретную модель, которая нужна для решения поставленной задачи, чтоб далее привлечь всю мощь теории, от упрощения моделей только выгадывающей.

Постановка задачи. Наша цель — рассмотреть проблему синтеза интервальных статистических моделей средних. Связующими для них с реальными явлениями будут некоторые *задающие параметры* $\theta = (\theta_1, \theta_2, \ldots)$, которые по их числу, содержанию и физическому смыслу могут быть самыми разнообразными (могут иметь только математический, формальный смысл). Теперь если модель искать в рамках исходной структуры $\mathcal{M}_\theta$ и найти параметр θ в виде детерминированной оценки $\hat{\theta}$, то получим модель $\mathcal{M}_{\hat{\theta}}$, а если в виде индикаторной $\hat{\Theta}$ (для одномерного θ — интервальной), то объединение $\bigvee_{\theta \in \hat{\theta}} \mathcal{M}_\theta$.

Оценивание θ должно производиться из ясного осознания конечных целей, под которыми подразумевается та последующая, наследственная задача, на решение которой призывается модель. Это может быть задача либо анализа, либо построения решающих правил. Например, модель шума привлекается для нахождения алгоритма оптимального обнаружения сигналов или оценивания параметров в аддитивном представлении. Возможен другой вариант, когда модели гипотезы и альтернативы не связаны между собой через шум и строятся отдельно одна от другой.

Нашу мысль, что решающие устройства должны через себя влиять на сам синтез модели в форме требований к оценкам задающих параметров, проиллюстрируем на примере.

Пример 8.1. Пусть θ — задающий (одномерный или многомерный) параметр; для каждого его значения вводится модель $\mathcal{M}_\theta$ и по ней синтезируется решающее правило ∂_θ (возможно, оптимальное). Вид этого правила определяется видом $\mathcal{M}_\theta$, следовательно, будет зависеть от θ. Но θ не известен. Считаем, что по обучающим реализациям $\mathbf{Z}$ он оценивается $\hat{\theta}=\hat{\theta}(\mathbf{Z})$ и подставляется в правило, что приводит к $\partial_{\hat{\theta}}$. Риск этого правила будет зависеть от истинного, но неизвестного θ и обозначается $\Pi_\theta(\partial_{\hat{\theta}})$. Теперь стоит вопрос, как оценивать $\hat{\theta}$? В среднем с учетом случайного разброса $\hat{\theta}$ риск всей указанной адаптивной процедуры (в общем, квазиоптимальной) равен $r(\partial, \hat{\theta})=M\times\Pi_\theta(\partial_{\hat{\theta}})$, где усреднение производится по совместной модели θ и $\mathbf{Z}$. Оценку $\hat{\theta}$ нужно выбирать так, чтобы минимизировать $r(\partial, \hat{\theta})$, откуда сразу видно, что $\Pi_\theta(\partial_{\hat{\theta}})$ и должна служить функцией потерь в задаче оценивания (синтеза) θ [1].

Выбор начальной структуры $\mathcal{M}_\theta$ — слабое место синтеза, способное в корне свести на нет наше стремление получить надежную модель. Так будет, если $\mathcal{M}_\theta$ описывается параметризованной точной плотностью вероятностей $p_\theta(x)$, при изменении θ очерчивающей лишь неосязаемую линию в пространстве всех распределений вероятностей. Надежные же модели должны быть объемными.

Наибольший интерес для нас представляет тот случай, когда $\mathcal{M}_\theta$ не надо выбирать, а оно само естественно определяется наведенными с явлением связями. А этот как раз тот самый случай, когда задающими параметрами являются статистические средние $\theta=M\mathcal{Q}$ набора $\mathcal{Q}$ признаков, составляющие в совокупности модель $\mathcal{M}=\bigvee\langle M\mathcal{Q}\rangle$ (где объединение обязано расплывчатости оценок средних). Нагрузка синтеза полностью перекладывается на выбор набора $\mathcal{Q}$ и оценивание соответствующих ему параметров $\theta_i=Mq_i$, $q_i\in\mathcal{Q}$. Не надо, что самое замечательное, ничего лишнего, никаких допущений, порождающих сомнения.

Здесь возникают две проблемы: выбор задающих признаков и оценивание их средних. Первая составляет свою сферу деятельности, и не всегда научную, а учитывая подавляющее разнообразие самих признаков и возможностей их выбора, превращающуюся подчас в искусство (начинается там, где заканчивается наука). Это инженерное искусство выбора связующих признаков, используя априорные сведения о явлении, какие есть знания его механизма, наблюдения за явлением, опираясь на практику, опыт и здравый смысл, сообразуясь с трудоемкостью самой процедуры синтеза модели, не «спуская прицел» с последующего применения модели. Путеводными здесь являются качества модели, проникающие настрой всей книги: простота, доступность, надежность.

[1] Кузнецов В. П. Байесов подход и оптимизация процесса обучения//Автоматика и телемеханика. — 1971. — № 4. — С. 66—71.

Вторая проблема состоит в оценивании задающих параметров. Она согласована с выбором задающих признаков и поэтому не исключает привлечения самых разнородных физических закономерностей, фактов. Но может и формально решаться на базе предварительного эксперимента, обучающих реализаций, что нас интересует в наибольшей мере. Здесь требование адекватности связи модели с явлением накладывает на обучающие реализации обязанность быть «полномочными представителями» интересующего нас явления, т. е. быть носителями одинаковых значений параметров (средних), тогда их можно оценивать. Это есть условие стационарности задающих параметров, которое сейчас обсудим.

Стационаризация статистических параметров. С позиций математических моделей средние как и вероятности есть фиксированные числа, тогда как интервальные средние и вероятности есть интервалы, границы которых удовлетворяют аксиомам ИМ. В эти понятия заложен смысл среднего арифметического или же частоты.

Сейчас не суть важно, каким является пространство $\mathcal{Z}$ элементарных исходов, на котором строится математическая модель; это может быть произведение $\mathcal{X}\times\mathcal{Y}$, либо пространство Ξ значений флуктуаций ξ. Пусть $z_1, z_2, \ldots, z_N$, есть реализация независимой последовательности *обучающих испытаний*, для которой параметр $Mq(z_i)$ является стационарным, т. е. одним и тем же для всех i. Такие испытания называются Mq-стационарными.

Для стационарного параметра точное среднее Mq достигается как предел среднего арифметического:

$$Mq=\lim_{N\to\infty}\frac{1}{N}\sum_{1}^{N}q(z_i)$$

в серии независимых стационарных испытаний z_i, $i=1, 2, \ldots$ Точная вероятность по тому же смыслу есть предел относительной частоты события

$$P(A)=\lim_{N\to\infty}\frac{1}{N}\sum_{1}^{N}\{z_i\in A\},$$

где фигурные скобки обозначают индикаторы событий, равные 1 при $z_i\in A$, а иначе — 0.

Интервальные средние и вероятности возникают тогда, когда точных их значений нет, либо по причине ограниченности числа испытаний (дефицит опыта), либо из-за возможной нестационарности испытаний, т. е. неустойчивости средних и частот. Последнее является неприятной «подножкой», но к счастью, не всему статистическому подходу, а тем формальным методам синтеза модели, к которым мы стремимся и которые волей-неволей требуют стационарности параметров. Укажем типичный для статистических приложений способ стационаризации, основанный на случайном выборе.

Пусть имеется совокупность в N независимых испытаний z_i, $i=1, ..., N$, причем средние $M_i q(z_i)$ в разных испытаниях, в общем, различны. «Перемешаем» теперь испытания слепым образом, а иначе говоря, произведем их последовательный равновозможный выбор. Тогда средние стабилизируются, станут одними и теми же, равными $Mq=\sum M_i q(z_i)/N$, так как равновозможно с вероятностями $1/N$ выбранным может оказаться любой индекс i. Как видно, произошла стационаризация испытаний, но увы, с возможной потерей независимости, хотя и из этого упущения есть свой выход, который проиллюстрируем на примере.

Пример 8.2. Пусть p_i есть вероятность события $z_i \in A$ в i-м испытании (A одно и то же) и пусть заведомо известно, что первые k раз A обязательно произойдет: $p_1=p_2= ... =p_k=1$, а остальные — нет: $p_{k+1}=p_{k+2}= ... =p_N=0$. Тогда случайный выбор (слепое перемешивание) испытаний ведет к вероятности $p=k/N$ события A в каждом из них. Но после перемешивания события A, увы, не обретут желаемой независимости, так как если стало известно, что первые k раз повторилось событие $z_i \in A$, то всеми последующими непременно будут противоположные события $z_i \in A^c$.

Если же после перемешивания из совокупности объема N произвести редкий выбор $z_1, ..., z_n$ небольшого их числа (объема) $n \ll N$, то эту выборку с достаточно большой точностью можно считать стационарной независимой. В самом деле, $\bar{M}_i q(z_i)$ от произвольного признака q будет одним и тем же $\bar{M}q$ и справедливо равенство $\bar{M}\Pi q_i(z_i) = \max\limits_{M^* \in \underline{M}, \bar{M}} \Pi M^* q_i(z_i)$.

Вывод такой: *случайный редкий выбор из совокупности есть средство стационаризации последовательности и причина независимости ее элементов.*

Стационаризация облегчает построение математической модели, так как сводит ее синтез к оцениванию одинаковых по течению испытаний задающих параметров, т. е. к своей новой статистической задаче, которую рассмотренными в предыдущих главах методами можно формализовать и решить. Для этого нужно знать, какого содержания оценки, задающие модель, хочется получить, каким основным показателям они должны удовлетворять.

Понятие доверительной модели. Проанализируем содержимое рис. 8.1. Построение математической модели — это отдельная задача оценивания параметров, а точнее сказать, *надзадача*, пищей которой служат испытания, и как всякая статистическая задача, она нуждается в исходной конструкции. В конструкцию надзадачи в первую очередь входит изначальная математическая модель самой последовательности испытаний, так сказать, *надмодель*, отличающаяся от искомой своей шириной и крайне непритязательными запросами: для нас это будут только предположения о независимости и стационарности. Интересно далее будет наблюдать за сужением ее в искомую модель, что достаточно сделать в направлении $\mathcal{Q}$ оцениваемых (задающих) параметров.

Детерминированные (точечные) задающие модель оценки $\hat{M}q$, $q \in \mathcal{Q}$, в надзадаче ведут к $\mathcal{Q}$-простым моделям $\langle M\mathcal{Q} \rangle$. Эти модели

с точными значениями задающих средних $Mq=\bar{M}q$, $q \in \mathcal{Q}$, будут ненадежными по причине ненадежности детерминированных оценок (кстати, включение детерминированных оценок в модель присуще, как это следует из теоремы 5.1, режиму оптимизма $\varkappa=1/2$). Нас сейчас всего более привлекает режим пессимизма и надежные модели, а для этого задающие оценки также должны быть надежными, доверительными, что ведет к следующему понятию.

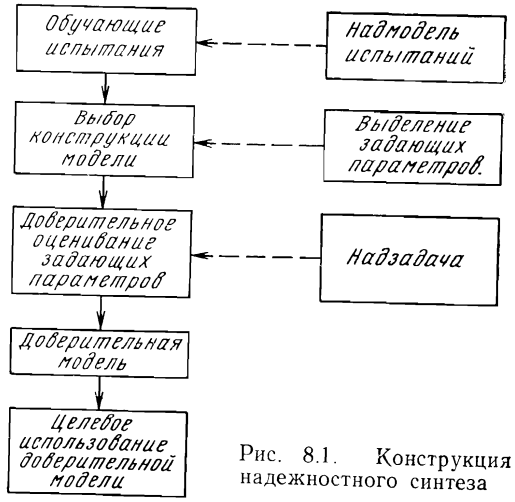


Рис. 8.1. Конструкция надежностного синтеза

Доверительной надежности ρ *называется модель, определенная пессимистичными* ($\varkappa=1$) *совместными доверительными оценками* ($\mathcal{Q}$) *уровня* $\alpha=1-\rho$ *набора* $M\mathcal{Q}=\{Mq, q \in \mathcal{Q}\}$ *задающих средних* (найденными по обучающей последовательности $\mathbf{z}=(z_1, ..., z_N)$ испытаний). Здесь нужно обратить внимание, что надежность доверительных оценок равна 1 минус верхняя граница $\bar{\alpha}(\mu)=\alpha$ вероятности ошибки, что соответствует пессимистичному оцениванию. Это принципиальный момент, что оптимизм, если и допустим при нахождении решающих правил, но ни в коем случае не при синтезе надежных моделей. Надежные модели пессимистичны.

Итак, доверительная модель равносильна расплывчатой оценке задающих модель параметров: средних значений признаков. Итогом станет интервальная модель средних $\mathcal{M}$, но только в том случае, если оценки $\mu_{\mathbf{z}}(\mathcal{Q})$ индикаторные или совместные интервальные, т. е. $\mu^{\mathbf{z}}(\mathcal{Q})$ принимает значения либо 0, либо 1. Причем значение $\mu_{\mathbf{z}}(\mathcal{Q})=1$ и выделяет как раз область Θ тех $\theta=M\mathcal{Q}$, которые включаются составляющими синтезируемой модели $\mathcal{M}$, записываемой как их объединение: $\mathcal{M}=\bigvee\limits_{MQ \in \Theta} \langle M\mathcal{Q} \rangle$. Представляется ИМ $\mathcal{M}$ как облако составляющих ее простых моделей $\langle M\mathcal{Q} \rangle$ одной концентрации в смысле одинакового доверия ко всем им внутри Θ (и нулевого доверия вне Θ).

Заметим, представив $\mathcal{M}$ через первичные средние: $\mathcal{M}=\langle \bar{M}\mathcal{G} \rangle$, что совпадение первичного набора $\mathcal{G}$ доверительной модели с задающими признаками $\mathcal{Q}$ будет лишь в случае, когда область Θ прямоугольная, т. е. направления ее граней совпадают с направлениями q. Например, $\mu_{\mathbf{z}}(\mathcal{Q})$ есть совместная интервальная оценка $\underset{\sim}{M}q$, $\bar{M}q$, $q \in \mathcal{Q}$, тогда $\Theta=\{Mq : \underset{\sim}{M}q \leqslant Mq \leqslant \bar{M}q, q \in \mathcal{Q}\}$. Для других индикаторных оценок $\mu_{\mathbf{z}}(\mathcal{Q})$ наборы $\mathcal{G}$ и $\mathcal{Q}$, в общем, различаются между собой: $\mathcal{G} \neq \mathcal{Q}$.

8.2. ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОЙ МОДЕЛИ НА ЗАДАННОМ НАБОРЕ СОБЫТИЙ

Исходные положения. Здесь рассматривается тот случай, когда задающими параметрами являются вероятности набора событий, поначалу непересекающихся, что применимо к тем задачам, в которых из вероятностей и слагаются характерные черты интересующего нас исследуемого явления. Такой выбор может руководствоваться простотой и удобством оценивания вероятностей. При этом четко нужно осознавать, что все первичные признаки итоговой доверительной модели будут обязательно постоянными на задающих событиях или их пересечениях и это же свойство перейдет дальше к решающим правилам (достаточного класса) как конечной цели построения модели.

Сформулируем формально задачу синтеза модели (надзадачу). Пусть $\mathbf{z}=(z_1, \ldots, z_N)$ есть обучающие испытания из пространства $\mathcal{Z}$: $z_j \in \mathcal{Z}$ (скалярного или векторного) и пусть $A_1, \ldots, A_k$ — непересекающиеся события, такие, что $\sum A_j \subset \mathcal{Z}$, и $A_{k+1} = \mathcal{Z} - \sum A_j \neq \varnothing$ — остаточное событие, введенное для общности. Вероятности $P(z_i \in A_j) = p_j$ считаются стационарными, не зависящими от номера i наблюдений, а события $z_i \in A_j$ и $z_{i'} \in A_{j'}$ при $i \neq i'$ нековариированными. Эти условия будут выполнены, если испытания являются независимыми стационарными.

Здесь задающим модель будет вектор $\mathbf{p}=(p_1, \ldots, p_k)$ вероятностей. Его и нужно оценить при заданной надежности ρ. Если обозначить доверительную оценку $\mu_{\mathbf{z}}(\mathbf{p})$, то ее надежность равна $\rho = \overline{M}\mu_{\mathbf{z}}(\mathbf{p})$.

При указанных условиях достаточными для оценивания являются векторы частот $\mathbf{r}=(r_1, \ldots, r_k)$, где r_j — число элементов z_i, попавших в A_j: $r_j = \sum_{i=1}^{N} \{z_i \in A_j\}$. Оценка $\mu_{\mathbf{z}}(\mathbf{p}) = \mu_{\mathbf{r}}(\mathbf{p})$ должна быть функцией вектора $\mathbf{r}$.

При каждом заданном векторе $\mathbf{p}$ вероятности частот $r_1, \ldots, r_k$ даются известной мультиномиальной формулой

$$P_{\mathbf{p}}(\mathbf{r}) = p_1^{r_1} \ldots p_{k+1}^{r_{k+1}} N!/(r_1! \ldots r_{k+1}!),$$

где считается $0^0=1$, $p_{k+1}=1-\sum_1^k p_j$, $r_{k+1}=N-\sum_1^k r_j$. Формула для $P_{\mathbf{p}}(\mathbf{r})$ определяет переходную надмодель $\mathcal{M}_{\mathbf{p}}^{\mathbf{r}}$, редуцированную от $\mathbf{z}$ к $\mathbf{r}$. Общей надмоделью испытаний будет произведение $\mathcal{M}^{\mathbf{pr}} = \mathcal{M}^{\mathbf{p}} \mathcal{M}_{\mathbf{p}}^{\mathbf{r}}$, куда в $\mathcal{M}^{\mathbf{p}}$ можно вложить априорные сведения о $\mathbf{p}$, если они есть. Их обычно нет, тогда $\mathcal{M}^{\mathbf{p}} = \mathcal{J}^{\mathbf{p}}$ и надежность доверительной оценки $\mu_{\mathbf{z}}(\mathbf{p})$ вектора $\mathbf{p}$ равна

$$\rho = \min_{\mathbf{p}} \sum_{\forall \mathbf{r}} \mu_{\mathbf{r}}(\mathbf{p}) P_{\mathbf{p}}(\mathbf{r}).$$

Шкала расплывчатости оценки вектора $\mathbf{p}$ во многом обязана желаемым свойствам получаемого по доверительной модели решающего правила. Если же не определять пока решаемую на базе модели статистическую задачу, то разумно использовать интегральную шкалу как наиболее простую:

$$\Omega(\mu) = \int_{\mathbf{I}} \ldots \int \mu_{\mathbf{r}}(\mathbf{p})\, d\mathbf{p}, \quad \mathbf{I} = \left\{ \mathbf{p}: \sum_1^{k+1} p_j = 1, p_j \geqslant 0 \right\}.$$

Частная $\mathcal{M}^{\mathbf{r}}$, дающая надмодель вектора частот $\mathbf{r}$, определяется границами:

$$\overline{M} f(\mathbf{r}) = \max_{\mathbf{p}} \sum_{\forall \mathbf{r}} f(\mathbf{r}) P_{\mathbf{p}}(\mathbf{r}),$$

поэтому составной риск запишется: $\Pi_\lambda(\mu) = 1 - \underline{M}\mu_{\mathbf{r}}(\mathbf{p}) + \lambda \overline{M}\Omega \times$

$$\times(\mu) = 1 - \min_{\mathbf{p}} \sum_{\forall \mathbf{r}} \mu_{\mathbf{r}}(\mathbf{p}) P_{\mathbf{p}}(\mathbf{r}) + \lambda \max_{\mathbf{p}^*} \sum_{\forall \mathbf{r}} P_{\mathbf{p}^*}(\mathbf{r}) \int_{\mathbf{I}} \ldots \int \mu_{\mathbf{r}}(\mathbf{p})\, d\mathbf{p}$$

где λ — весовой коэффициент. Нужно минимизировать риск выбором $\mu_{\mathbf{r}}(\mathbf{p})$.

Перепишем риск следующим образом: $\Pi_\lambda(\mu) = 1 - \inf_{w(\mathbf{p})} \sum_{\forall \mathbf{r}} \int_{\mathbf{I}} \ldots \int \mu_{\mathbf{r}} \times$

$$\times(\mathbf{p}) P_{\mathbf{p}}(\mathbf{r}) w(\mathbf{p})\, d\mathbf{p} + \lambda \sup_{w^*(\mathbf{p})} \sum_{\forall \mathbf{r}} \int_{\mathbf{I}} \ldots \int P_{\mathbf{p}}(\mathbf{r}) w^*(\mathbf{p})\, d\mathbf{p} \int_{\mathbf{I}} \ldots \int \mu_{\mathbf{r}}(\mathbf{p})\, d\mathbf{p}$$

Тогда цена надзадачи примет вид

$$v = 1 - \sup_{\mu} \inf_{w, w^*} \sum_{\forall \mathbf{r}} \int_{\mathbf{I}} \ldots \int \mu_{\mathbf{r}}(\mathbf{p}) [P_{\mathbf{p}}(\mathbf{r}) w(\mathbf{p}) - \lambda P_{w^*}(\mathbf{r})]\, d\mathbf{p},$$

где инфимум ищется по всевозможным априорным плотностям $w(\mathbf{p})$ и $w^*(\mathbf{p})$, а $P_{w^*}(\mathbf{r}) = \int_{\mathbf{I}} \ldots \int P_{\mathbf{p}}(\mathbf{r}) w^*(\mathbf{p})\, d\mathbf{p}$ — вероятность вектора $\mathbf{r}$ при плотности w^*.

В соответствии с общими принципами минимакса [22] супремум и инфимум в записи цены поменяем местами. При каждых заданных $w(\mathbf{p})$ и $w^*(\mathbf{p})$ оптимальная оценка $\mu_{\mathbf{r}}(\mathbf{p})$ вектора $\mathbf{p}$ будет индикаторной, принимающей значение 1 при

$$P_{\mathbf{p}}(\mathbf{r}) w(\mathbf{p}) / P_{w^*}(\mathbf{r}) \geqslant \lambda \tag{8.1}$$

и 0 в противном случае. С учетом найденной оценки цена запишется:

$$v = 1 - \inf_{w, w^*} \sum_{\forall \mathbf{r}} \int_{\mathbf{I}} \ldots \int [P_{\mathbf{p}}(\mathbf{r}) w(\mathbf{p}) - \lambda P_{w^*}(\mathbf{r})]^+\, d\mathbf{p}.$$

Отсюда нужно искать наименее благоприятные $w(\mathbf{p})$ и $w^*(\mathbf{p})$, которые затем подставляются в (8.1), что и приведет к искомой оптимальной оценке $\mathbf{p}$. Порог λ здесь должен быть выбран исходя из заданной надежности

$$\rho = \sum_{\forall \mathbf{r}} \int_{\mathbf{I}} \ldots \int \mu_{\mathbf{r}}(\mathbf{p}) P_{\mathbf{p}}(\mathbf{r}) w(\mathbf{p})\, d\mathbf{p}, \tag{8.2}$$

куда подставляется наименее благоприятная плотность $w(\mathbf{p})$.

Трудности нахождения наименее благоприятных плотностей вынуждают прибегать к некоторым «разумным» вариантам их подбора.

Модель наибольшего правдоподобия. Пусть априорное распределение вектора $\mathbf{p}$ равномерно $w(\mathbf{p})=\text{const}$ и пусть $P_{w^*}(\mathbf{r})=\text{const}$, т. е. априори никаким векторам частот предпочтения не отдается, что в определенном смысле отражает неблагоприятную ситуацию. Тогда формула (8.1) запишется $P_{\mathbf{p}}(\mathbf{r})\geqslant\lambda$, т. е. оценкой является индикаторная функция семейства векторов $\mathbf{p}$ максимальной вероятности (наибольшего правдоподобия). После логарифмирования обеих частей последнего неравенства, перемены знака и объединения между собой постоянных слагаемых оценка примет вид

$$F(\hat{\mathbf{p}},\mathbf{p})=-\sum_{j=1}^{k+1}\hat{p}_j\ln p_j\leqslant\lambda_1,\ \lambda_1 N=-\ln\lambda+\ln(N!)-\sum_{1}^{k+1}\ln(r_j!), \tag{8.3}$$

где $\hat{p}_j=r_j/N$ — есть относительные частоты выпадения событий A_j.

Отметим, что функция $F(\hat{\mathbf{p}},\mathbf{p})$ неотрицательна и принимает минимальное значение при $\mathbf{p}=\hat{\mathbf{p}}$, равное: $F(\hat{\mathbf{p}},\hat{\mathbf{p}})=-\sum_{1}^{k+1}\hat{p}_j\ln\hat{p}_j$. Пороговое значение λ (а отсюда и λ_1) находится по (8.2) с подстановкой $w(\mathbf{p})=\text{const}$. Интеграл при этом получается трудоемким, поэтому есть смысл определять λ исходя из нижней границы надежности:

$$\rho=\min_{\mathbf{p}}\sum_{\mathbf{r}:P_{\mathbf{p}}(\mathbf{r})\geqslant\lambda}P_{\mathbf{p}}(\mathbf{r}). \tag{8.4}$$

Неравенство (8.3) с нахождением λ_1 с помощью (8.4) формирует семейство $\mathscr{M}$ векторов $\mathbf{p}$, составляющих доверительную модель надежности ρ. Эта $\mathscr{M}$ иначе задается своими первичными средними, к поиску которых и приступаем. Здесь (так как семейство $\mathscr{M}$ непрямоугольное) первичными признаками не будут сами задающие события A_j (вероятности $\mathbf{p}$ которых оцениваются), а будут измеримые функции на них.

Обозначим $\mathbf{p}^\circ=(p^\circ_1,\ \ldots,\ p^\circ_k)$, $p^\circ_{k+1}=1-\sum_{1}^{k}p^\circ_j$ — решения уравнения $F(\hat{\mathbf{p}},\mathbf{p})=\lambda_1$ относительно $\mathbf{p}$ при выборе λ_1 по заданной надежности ρ. Эти решения образуют набор, задающий поверхность семейства $\mathscr{M}$. Первичные средние определяются гиперплоскостями в подмножестве I пространства $\mathscr{R}^{k+1}$, касающимися семейства $\mathscr{M}$ в точках $\mathbf{p}^\circ$. Так как $\partial F(\mathbf{r},\mathbf{p})/\partial p_j=-r_j/p_j$, то уравнения этих гиперплоскостей имеют вид $\sum_{1}^{k+1}\hat{p}_j(p_j-p^\circ_j)/p^\circ_j=0$. Эти гиперплоскости и определяют первичные признаки вида

$$g_{\mathbf{p}^\circ}(\mathbf{z})=\sum_{j=1}^{k+1}\hat{p}_j A_j(\mathbf{z})/p^\circ_j,\quad \hat{p}_j=r_j/N,$$

и соответствующие им средние

$$\overline{Mg}_{\mathbf{p}^\circ}=\max_{\mathbf{p}\in\mathscr{M}}\sum_{1}^{k+1}\hat{p}_j p_j/p^\circ_j=\sum_{1}^{k+1}\hat{p}_j p^\circ_j/p^\circ_j=1.$$

В результате различным $\mathbf{p}^\circ$ как решениям уравнения $\hat{F}(\mathbf{p},\mathbf{p})=\lambda$ соответствуют различные первичные признаки $g_{\mathbf{p}^\circ}(\mathbf{z})$ с одним и тем же у всех равным 1 верхним средним.

Таким образом, *первичными средними, определяющими доверительную модель, будут*

$$\overline{M}\sum_{1}^{k+1}\hat{p}_j A_j(z)/p^\circ_j=1,\ \forall\,\mathbf{p}^\circ:\ -\sum_{1}^{k+1}\hat{p}_j\ln p^\circ_j=\lambda_1. \tag{8.5}$$

Мы видим, что $\mathscr{M}$ зависит от полученных в испытаниях частот r_j (или относительных частот $\hat{p}_j$). При $N\to\infty$ модель $\mathscr{M}$ стягивается к точному распределению вероятностей, соответствующему предельным частотам $p_j=\lim_{N\to\infty}\hat{p}_j$.

Получим в явном виде некоторые из первичных средних, соответствующих набору (8.5). Для этого фиксируем $p^\circ_j=\hat{p}_j$, $j=2,\ \ldots,\ k$, и будем искать p°_1 из равенства $F(\hat{\mathbf{p}},\mathbf{p})=\lambda_1$. Получим p°_1 как решение уравнения

$$-\hat{p}_1\ln p_1-\hat{p}_{k+1}\ln(\hat{p}_{k+1}+\hat{p}_1-p_1)=\lambda_1+\sum_{2}^{k}\hat{p}_j\ln\hat{p}_j.$$

Этих решений будет два: $\underset{\sim}{p}{}^\circ_1$, $\tilde{p}^\circ_1$, где $\underset{\sim}{p}{}^\circ_1\leqslant\hat{p}_1\leqslant\tilde{p}^\circ_1$. Таким образом, искомый вектор $\mathbf{p}^\circ_k$ вероятностей будет иметь вид $p^\circ_1,\hat{p}_2,\ \ldots,\hat{p}_k$, $(1-\sum_{2}^{k}\hat{p}_j-p^\circ_1)$, где p°_1 равно либо $\underset{\sim}{p}{}^\circ_1$, либо $\tilde{p}^\circ_1$. Соответствующее каждому такому вектору первичное среднее находится из уравнения (8.5) и имеет вид

$$\overline{M}(\hat{p}_1-p^\circ_1)\,[A_1(\mathbf{z})/p^\circ_1-A_{k+1}(\mathbf{z})/(\hat{p}_{k+1}+\hat{p}_1-p^\circ_1)]=0.$$

Заменяя индексы 1 и $k+1$ разными комбинациями индексов l, m, $l\neq m$, от 1 до $k+1$, получим поднабор первичных средних вместе с уравнениями для нахождения $\underset{\sim}{p}{}^\circ_l$ и $\tilde{p}^\circ_l$.

Если склоняться к упрощениям, то от всего набора (8.5) первичных средних можно отказаться, оставив их какую-то часть, что приводит к расширению доверительной модели с увеличением надежности. За основу расширения могут быть взяты любые признаки вида $g(\mathbf{z})=\sum_{1}^{k+1}g_i A_i(z)$. Интересно то, что каждый из них с точностью до множителя совпадает с одним из первичных признаков (8.5) (для этого нужно подобрать соответствующее $\mathbf{p}^\circ$), поэтому такое расширение эквивалентно уменьшению числа первичных средних.

Использование критерия хи-квадрат. Определяющее доверительную модель $\mathcal{M}$ семейство векторов $\mathbf{p}$ может быть выбрано с позиций упрощенного расчета порога λ. Используем для этого статистику хи-квадрат

$$\sum_{1}^{k+1} (p_j - \hat{p}_j)^2/p_j \leqslant \lambda_2, \tag{8.6}$$

где λ_2 находится по заданной надежности ρ. При больших N левая часть неравенства имеет приближенно распределение хи-квадрат с k степенями свободы [25]. Тогда порог λ_2 будет критической точкой этого распределения. Семейству $\mathcal{M}$ векторов $\mathbf{p}$, определенному неравенством (8.6), соответствуют первичные значения

$$\overline{M} \sum_{1}^{k+1} \hat{p}_j^2 A_j(\mathbf{z})/(p_j^0)^2 = \sum_{1}^{k+1} \hat{p}_j^2/p_j^0,$$

где $\mathbf{p}^\circ$ есть всевозможные решения уравнения (8.6), в котором неравенство заменено на равенство.

Информационный критерий построения доверительной модели. Будем считать в (8.1) $w(\mathbf{p}) = \text{const}$ и остановимся на подборе плотности $w^*(\mathbf{p})$, входящей в знаменатель. Будем искать максимум знаменателя по $w^*(\mathbf{p})$ при каждом заданном векторе частот $\mathbf{r}$. Этот максимум достигается при дельта-функции Дирака $w^*(\mathbf{p}) = \delta(\hat{\mathbf{p}} - \mathbf{p})$ и определяется по формуле

$$P(\mathbf{r}) = \max_{w^*} P_{w^*}(\mathbf{r}) = \max_{\mathbf{p}} P_{\mathbf{p}}(\mathbf{r}) = P_{\mathbf{r}/N}(\mathbf{r}).$$

Оправданием нашим действиям служит то, что имея наблюденным вектор $\mathbf{r}$, мы рассматриваем наименее благоприятную плотность w^* применительно к $\mathbf{r}$.

С учетом найденного $P(\mathbf{r})$ после подстановки его в (8.1), логарифмирования обеих частей неравенства и перемены знака приходим к семейству $\mathcal{M}$ векторов $\mathbf{p}$, определяемому неравенством

$$J(\hat{\mathbf{p}}, \mathbf{p}) = \sum_{1}^{k+1} \hat{p}_j \ln(\hat{p}_j/p_j) \leqslant \lambda_3. \tag{8.7}$$

Левая часть неравенства (8.7) есть различающая информация [24], содержащаяся в векторе $\hat{\mathbf{p}}$ в пользу точной вероятностной модели, определяемой этим вектором, при конкурирующей альтернативе $\mathbf{p}$. Ясно, что чем меньше эта различающая информация, тем ближе $\mathbf{p}$ к $\hat{\mathbf{p}}$, а при $\mathbf{p} = \hat{\mathbf{p}}$ эта информация минимальна и равна 0. Таким образом, согласно (8.7) доверительную модель образуют такие векторы $\mathbf{p}$, которые в смысле различающей информации отстоят от $\hat{\mathbf{p}}$ не более, чем на число λ_3. Порог λ_3 находится из формулы (8.4) с подстановкой неравенства (8.7) под знак символа суммы как ограничения на $\mathbf{r}$ (вспомним, что $\hat{\mathbf{p}} = \mathbf{r}/N$).

В сравнении с (8.3) имеем

$$J(\hat{\mathbf{p}}, \mathbf{p}) = \sum_{1}^{k+1} \hat{p}_j \ln \hat{p}_j + F(\hat{\mathbf{p}}, \mathbf{p}),$$

поэтому (8.7) совпадает с (8.3) при $\lambda_1 = \lambda_3 - \sum_{1}^{k+1} \hat{p}_j \ln \hat{p}_j$. Заменим в (8.7) неравенство на равенство и обозначим какое-то его вектор-решение $\mathbf{p}^\circ$. Этот вектор совпадает с таким же определяемым правым равенством в (8.5), если подставить туда вновь пересчитанное значение λ_1, поэтому первичные средние будут совпадать с (8.5) при соответствующих $\mathbf{p}^\circ$, что ведет к следующей записи первичных средних доверительной модели и уравнений для $\mathbf{p}^\circ$ и λ_3:

$$\overline{M} \sum_{1}^{k+1} \hat{p}_j A_j(\mathbf{z})/p_j^0 = 1, \quad J(\hat{\mathbf{p}}, \mathbf{p}^0) = \lambda_3,$$

$$\rho = \min_{\mathbf{p}} \sum_{\mathbf{r}: J(\hat{\mathbf{p}}, \mathbf{p}) \leqslant \lambda_3} P_{\mathbf{p}}(\mathbf{r}).$$

Доверительные совместные оценки. Пусть $\mathcal{A}_k = \{A_1, \ldots, A_k\}$ — произвольный (в общем, пересекающийся) набор событий на $\mathcal{Z}$. Требуется получить совместные доверительные оценки вероятностей этих событий $p_j = P(A_j)$ в виде произведения оценок $\mu_{\mathbf{z}}(p_j)$ отдельных событий: $\mu_{\mathbf{z}}(\mathbf{p}) = \prod_{1}^{k} \mu_{\mathbf{z}}(p_j)$. Требование к надежности совместной оценки накладывает ограничения на надежности отдельных оценок. Этот вопрос здесь и рассматривается, а именно получить совместную доверительную оценку, имея отдельные оценки.

Для отдельного события A_j доверительная оценка $\underset{\sim}{p}_j$, $\tilde{p}_j$ уровня α_j дается как решение уравнений [25]

$$\sum_{i=0}^{r_j - 1} C_N^i \underset{\sim}{p}_j^i (1 - \underset{\sim}{p}_j)^{N-i} = 1 - \alpha_{j0}, \quad \sum_{i=0}^{r_j} C_N^i \tilde{p}_j^i (1 - \tilde{p}_j)^{N-i} = \alpha_{j1},$$

где $\alpha_{j0} + \alpha_{j1} = \alpha_j$ и r_j — частота этого события. В асимптотическом варианте при $N \to \infty$ и $\alpha_{j0} = \alpha_{j1} = \alpha_j/2$ доверительные границы будут приближенно равны

$$\hat{p}_j \pm \Phi^{-1}(1/2 - \alpha_j/2) \sqrt{\hat{p}_j (1 - \hat{p}_j)/N},$$

где $\Phi(x)$ — функция Лапласа.

Теорема 8.1. *Пусть* $\varkappa = 1$ *и* $\mu_{\mathbf{r}}(p_j)$, $j = 1, \ldots, k$, *есть доверительные интервальные оценки параметров* $p_j = P(A_j)$ *уровней* α_j, *определяемые границами* $\underset{\sim}{P}(A_j) = \underset{\sim}{p}_j$, $\tilde{P}(A_j) = \tilde{p}_j$. *Тогда эти границы, взятые за первичные, задают доверительную интервальную модель* $\mathcal{M}_\rho$ *надежности, по крайней мере,* $\rho \geqslant 1 - \sum_{1}^{k} \alpha_j$.

Доказательство. Для совместной оценки $\mu_{\mathbf{r}}(\mathbf{p})=\prod_1^k \mu_{r_j}(p_j)$, используя элементарное неравенство $\prod_1^k \mu_{r_j}(p_j) \geqslant 1-\sum_1^k [1-\mu_{r_j}(p_j)]$, имеем: $\rho=\underline{M}\mu_{\mathbf{r}}(\mathbf{p})=$ $=\underline{M}\prod_1^k \mu_{r_j}(p_j) \geqslant 1-\sum_1^k \bar{M}[1-\mu_{r_j}(p_j)]=1-\sum_1^k \alpha_j$, что и требовалось.

Замечания. 1. Теорема 8.1 имеет смысл для любых совместных доверительных оценок задающих параметров. Совершенно не обязательно, что это оценки вероятностей и что они интервальные. Так, если $\mu_{\mathbf{z}}(m_j)$, $j=1, \ldots, k$, есть расплывчатые доверительные оценки параметров $m_j=Mg_j(\mathbf{z})$, то размытая модель надежности $\rho \geqslant 1-\sum_1^k \alpha_j$ будет определяться первичными размытыми средними $\mu_{\mathbf{z}}(m_j)$ признаков $g_j(\mathbf{z})$, $j=1, \ldots, k$, и k — размерность модели (если нет избыточности в ее первичных средних).

2. Если уровни α_j, $j=1, \ldots, k$, считать равными между собой, то в соответствии с теоремой 8.1 нужно брать $\alpha_j=(1-\rho)/k$. Это значение является явно заниженным, особенно при большом k, что приведет к излишне расширенной доверительной модели, поэтому требуется пересчет надежности по уже выбранным первичным средним (доверительным оценкам) отдельных параметров. Это делается так. Пусть $m_j=M\sum_i g_{ji}A_i(\mathbf{z})=\sum_i g_{ji}p_i$, где A_i — непересекающиеся множества, образующие разбиение $\mathscr{Z}$, а p_i — их вероятности. Тогда пересчитанная надежность

$$\rho=\min_{\mathbf{p}} \sum_{\forall \mathbf{r}} P_{\mathbf{p}}(\mathbf{r})\, \mu_{\mathbf{r}}(m).$$

3. При увеличении числа первичных параметров надежность модели при заданных оценках отдельных параметров, в общем, падает. Кроме того случая, когда задающие параметры «сильно связаны» между собой, т. е. один лишь незначительно отличается от другого. Такая связь имеет место, если задающими являются вкладывающиеся друг в друга события, к рассмотрению чего и приступим.

Доверительная функция распределения. Пусть задающим (он же первичный) является набор вкладывающихся друг в друга событий $A_\theta \subset \mathscr{Z}$, $\theta \in \mathscr{R}$, $A_\theta \subset A_{\theta'}$ при $\theta \leqslant \theta'$; обозначим через $\hat{p}_\theta=$ $=\sum_1^N \{z_i \in A_\theta\}/N$ — значения относительных частот попадания последовательности z_i испытаний в множества A_θ. Если отобразить $\mathscr{Z}$ в $\mathscr{R}$ так, что A_θ отображается в полуинтервал $(-\infty, \theta)$, то $\hat{p}_\theta$ как функция θ будет выборочной функцией распределения. Обозначим $p_\theta=P(z \in A_\theta)$. Требуется найти доверительные границы для p_θ, такие, что $\underline{P}(\prod_\theta \{\hat{p}_\theta - c_N(\rho) \leqslant p_\theta \leqslant \hat{p}_\theta + c_N(\rho)\})=\rho$. Разность $|\hat{p}_\theta - p_\theta|$ — известная статистика Колмогорова, поэтому $c_N(\rho)$ есть табулированные процентные точки распределения этой статистики [25]. Таким образом, искомыми первичными значениями, определяющими доверительную модель, будут

$$\underset{\sim}{P}(A_\theta)=\hat{p}_\theta - c_N(\rho),\ \tilde{P}(A_\theta)=\hat{p}_\theta + c_N(\rho).$$

Мы видим, что несмотря на бесконечное в данном случае число задающих параметров p_θ доверительные границы отдельных параметров не становятся тем не менее тривиальными, не расширяются до интервала (0; 1), как это следовало бы из теоремы 8.1.

8.3. СОГЛАСОВАННЫЙ СИНТЕЗ МОДЕЛЕЙ И ПРАВИЛ

Надежность моделей и истинные ошибки правил. Процесс изготовления промышленных изделий «обрастает», хотим мы этого или не хотим, массой подготовительных работ и вспомогательных служб. Нужно раздобыть сырье, сделать заголовки (модели), скомплектовать их, доставить к месту, а для этого нужно иметь помещение, подготовить станки, технику, наконец, найти рабочих и заинтересовать их зарплатой, организовать экономические службы (хотя и это еще, конечно же, не все). И только потом можно приступать к самому изготовлению. На надежность, ритмичность работы нужно смотреть в комплексе с охватом всего отлаженного механизма предприятий в целом.

Примерно то же самое имеет место при синтезе модели, где подготовительные работы состоят в выборе $\mathscr{M}^{xy}$, $\mathscr{D}$, $[\pi]$, $\varkappa$, составляющих статистическую задачу, а самое главное — в выборе модели $\mathscr{M}^{xy}$, своего рода инструмента к будущему изделию — решающему правилу (заготовкой к которому является $\mathscr{D}$). Брак в части инструмента делает бессмысленным само дальнейшее «изготовление» (какая бы ни была заготовка), поэтому в первую очередь модель $\mathscr{M}^{xy}$ должна быть надежной, доверительной. В то же время нельзя и это требование доводить до полного абсурда, забывая о целевом назначении модели — прямо вести к изделию — решающему правилу. Чрезмерные издержки на инструмент поднимут и время изготовления, и суммарную стоимость. Здесь нужен компромисс.

Будем мыслимо под «изделиями» подразумевать правила доверительного оценивания (хотя это может быть проверка гипотез). У модели и у правила свои атрибуты: у модели — это надежность, у правила — ошибка (уровень значимости). Последняя рассчитывается по виду модели. Очевидно, имея ненадежную модель, нельзя уже доверять ошибке (расчетному уровню значимости α), рассчитанной (по модели) для правила, ибо истинная вероятность ошибки ожидается выше расчетной. В то же время слишком широкая и отсюда чрезмерно надежная модель приведет к неоправданному увеличению расчетных ошибок правил. Возникающее противоречие рождает потребность вскрыть

строгую связь надежности модели с ошибками правил, к чему и перейдем.

Пусть $\alpha^{\varkappa}(\partial)$ — уровень, или *расчетная вероятность ошибки* решающего правила ∂, рассчитанного по доверительной модели $\mathcal{M}_\rho$ надежности ρ. Тогда с вероятностью $1-\rho$, с какой модель «бракована», эта ошибка не соответствует истине, и неконтролируема. Пессимизм $\varkappa=1$ при расчете заставляет усугубить ситуацию, считать ошибку правил из-за «брака» модели максимально возможной, равной 1; а крайний оптимизм $\varkappa=0$ — наоборот, есть, по сути, вера, что все будет как нельзя лучше, т. е. ошибка минимальна и равна 0. В результате можно подметить, что неконтролируемая ошибка просто равна коэффициенту пессимизма $\varkappa$. Так как «брак» модели по ее построению закладывается с вероятностью $1-\rho$, то *истинная* (полная) *вероятность ошибки* правил составляется из неконтролируемой $\varkappa$ с вероятностью $1-\rho$ и расчетной с вероятностью ρ:

$$\alpha^{\varkappa}_{\text{и}}(\partial)=\rho\alpha^{\varkappa}(\partial)+(1-\rho)\,\varkappa. \tag{8.8}$$

Подставив в полученную формулу $\alpha^{\varkappa}(\partial)=\varkappa\overline{\alpha}(\partial)+(1-\varkappa)\underline{\alpha}(\partial)$, где $\overline{\alpha}(\partial)=1-M\partial$, $\underline{\alpha}(\partial)=1-\overline{M}\partial$, перепишем истинную вероятность ошибки в другом виде:

$$\alpha^{\varkappa}_{\text{и}}(\partial)=\varkappa\,[1-\rho\,(1-\overline{\alpha}(\partial)+\underline{\alpha}(\partial))]+\rho\underline{\alpha}(\partial).$$

Отсюда наблюдается связь истинной вероятности ошибки с коэффициентом пессимизма $\varkappa$. Чем больше пессимизм $\varkappa$, тем больше истинная ошибка $\alpha^{\varkappa}{}_{\text{и}}(\partial)$. Но эта ошибка всегда не выше $\alpha^{0}{}_{\text{и}}(\partial)=\rho\underline{\alpha}(\partial)$, что соответствует $\varkappa=0$. При $\varkappa=1$ истинная ошибка $\alpha^{1}{}_{\text{и}}(\partial)=\overline{\alpha}(\partial)+(1-\rho)[1-\overline{\alpha}(\partial)]$ складывается из верхней границы ошибки $\overline{\alpha}(\partial)$ и (при $\overline{\alpha}(\partial)\ll 1$) ненадежности $1-\rho$ (вероятности брака) доверительной модели.

Если записать (8.8) еще в одном виде: $\alpha^{\varkappa}{}_{\text{и}}(\partial)=\alpha^{\varkappa}(\partial)+(1-\rho)[\varkappa-\alpha^{\varkappa}(\partial)]$, то будет видно, что при $\varkappa>\alpha^{\varkappa}(\partial)$ истинная вероятность ошибки всегда больше расчетной $\alpha^{\varkappa}(\partial)$, причем чем меньше надежность ρ и больше пессимизм, тем существенней эта разница, тем больше истинная вероятность ошибки. При $\varkappa\geqslant 1-\alpha^{\varkappa}(\partial)$ истинная ошибка превышает расчетную, по крайней мере, на величину ненадежности $1-\rho$.

Установим связь между истинной ошибкой правила и надежностью модели, для чего запишем при $\varkappa=1$: $\alpha^{1}{}_{\text{и}}(\partial)=1-\rho[1-\overline{\alpha}(\partial)]$. Увеличение ρ вроде бы должно уменьшать $\alpha^{1}{}_{\text{и}}(\partial)$. На самом деле, при увеличении ρ доверительная модель расширяется, в результате $\alpha(\overline{\partial})$ возрастает, устремляясь к 1 при $\rho\to 1$. Тогда и $\alpha^{1}{}_{\text{и}}(\partial)\to 1$. Таким образом, брать большую надежность модели бессмысленно. А так как (в силу установленного выше) и малая надежность не имеет смысла, то *для каждой статистической задачи должно существовать оптимальное значение надежности* ρ (пример будет рассмотрен в последнем разделе).

Пусть требуется обеспечить заданный истинный уровень α оценки: $\alpha^{\varkappa}{}_{\text{и}}(\partial)\leqslant\alpha$. Подстановкой (8.8) находится ограничение снизу на надежность $\rho\geqslant(\varkappa-\alpha)/[\varkappa-\alpha^{\varkappa}(\partial)]$ и величина расчетного уровня $\alpha_{\text{рас}}$, на который нужно настраивать правило: $\alpha^{\varkappa}(\partial)\leqslant[\alpha-\varkappa(1-\rho)]/\rho=\alpha_{\text{рас}}$. Если $\alpha<\varkappa(1-\rho)$, то последнее неравенство невыполнимо: нельзя найти $\alpha_{\text{рас}}$, нацелив на которое правило, получили бы истинный уровень α, что говорит о невозможности получения по ненадежным моделям низкого уровня α оценок. Всегда $\alpha\geqslant\varkappa(1-\rho)$ — *истинный уровень правила больше ненадежности модели, взвешенной коэффициентом пессимизма.*

Сказанное точь-в-точь переносится на расчет ошибки первого рода $\alpha^{\varkappa}(\partial)$ при проверке гипотез. Но если оценка откликается на падение расчетного уровня (по сравнению с истинным) расширением, что своего рода реакция (протест) на ненадежность модели, то у правила проверки гипотез понижение расчетной ошибки $\alpha^{\varkappa}(\partial)$ по принципу качелей вызовет рост расчетной ошибки второго рода $\beta^{\varkappa}(\partial)$, и тем самым истинной, определяемой согласно формуле (8.8) будет: $\beta^{\varkappa}{}_{\text{и}}(\partial)=\rho\beta^{\varkappa}(\partial)+\varkappa(1-\rho)$.

Теперь понятен общий случай, включающий в себя рассмотренные частные и охватывающий также задачу фильтрации. Приведенные соображения приводят к следующему выражению для *истинного риска*:

$$\Pi^{\varkappa}_{\text{и}}(\partial)=\rho\Pi^{\varkappa}(\partial)+(1-\rho)\,[\varkappa\sup_{\partial}\overline{\Pi}(\partial)+(1-\varkappa)\inf_{\partial}\underline{\Pi}(\partial)], \tag{8.9}$$

где первое слагаемое есть риск, рассчитываемый по доверительной модели и взвешенный ее надежностью ρ, а слагаемые в квадратных скобках — это, в зависимости от степени пессимизма $\varkappa$, отражающего «настроение» правил, тот наибольший ущерб и соответственно наименьший, которого можно ждать от «бракованной» модели, причем эта часть риска от ∂ зависеть не будет.

Нужно отметить, что структура оптимальных правил от перерасчетов риска, в общем, защищена, так как всецело определяется первым слагаемым (8.9), и в конечном счете, первичными признаками доверительной модели. Гибкими останутся отдельные параметры этих правил, вариации которых помогают управлять расплывчатостью и ошибкой.

Итак, установлено, что *ненадежность модели ведет к необходимости введения поправок в ошибки правил: истинные ошибки будут, в общем, больше расчетных.* При потребности обеспечить фиксированную истинную ошибку (уровень) нужно предусмотрительно брать заведомо меньшее расчетное значение со скидкой на ненадежность, что соответствует и что приведет к падению реальных качеств правил как оптимальных, так и неоптимальных.

Размытые доверительные модели и решения. Расплывчатые оценки задающих параметров ведут к разным моделям, причем индикаторные оценки ведут к ИМ, а неиндикаторные — к более общим размытым моделям § 2.3. И тут возникает вопрос, как для размытых (неинтервальных) моделей организовать синтез опти-

мальных правил. Этот важный момент оставался вне рамок рассмотрения, поскольку при синтезе мы ограничивались строго интервальными моделями. К его освещению и перейдем.

Пусть имеется задающий параметр $\theta = Mq(z)$ (пока всего один) и $\mu_z(\theta)$ есть его расплывчатая оценка надежности $\rho = 1 - \bar{\alpha}(\mu)$. Считаем оценку $\mu_z(\theta)$ унимодальной контрастной функцией параметра θ (т. е. достигающей 0 и 1). При каждом числе $0 \leqslant \gamma \leqslant 1$, называемой высотой горизонтального среза, неравенство $\mu_z(\theta) \geqslant \gamma$ выделяет интервальную оценку (слой) $[\underset{\sim}{\theta}^\gamma, \tilde{\theta}^\gamma]$ задающего параметра θ, так что при непрерывной по θ функции $\mu_z(\theta)$ имеем $\mu_z(\underset{\sim}{\theta}^\gamma) = \mu_z(\tilde{\theta}^\gamma) = \gamma$. Каждый срез $[\underset{\sim}{\theta}^\gamma, \tilde{\theta}^\gamma]$ в свою очередь определяет свою интервальную модель $\mathcal{M}^z_{(\gamma)} = \langle \underset{\sim}{\theta}^\gamma, \tilde{\theta}^\gamma \rangle$, располагающуюся на высоте γ, причем $\mathcal{M}^z_{(\gamma')} \subset \mathcal{M}^z_{(\gamma)}$ при $\gamma' \geqslant \gamma$, а все вместе они, положенные друг на друга в соответствии с высотами, дадут размытую модель.

Пространство $\mathcal{Z}$ так или иначе связано с произведением $\mathcal{X} \times \mathcal{Y}$, поэтому по $\mathcal{M}^z_{(\gamma)}$ определяется СИМ $\mathcal{M}^{xy}_{(\gamma)}$ на этом произведении, и для каждой из них находится свое оптимальное правило $\partial_y^{(\gamma)}(x)$, зависящее от высоты среза γ модели (своего рода привилегий к ней), причем чем больше γ, тем (при $\varkappa \geqslant 1$) более узкой будет $\mathcal{M}^{xy}_{(\gamma)}$ и менее расплывчатым — правила $\partial_y^{(\gamma)}(x)$, т. е. $\partial_y^{(\gamma')}(x) \leqslant \leqslant \partial_y^{(\gamma)}(x)$ при $\gamma' \geqslant \gamma$. Теперь оптимальным правилом $\partial^*{}_y(x)$ при размытой модели $\mathcal{M}^{xy}_{(\gamma)}$, построенной по расплывчатой оценке $\mu_z(\theta)$, будет

$$\partial^*_y(x) = \int_0^1 \partial_y^{(\gamma)}(x)\, d\gamma. \qquad (8.10)$$

Поясним его, считая ∂ оценкой параметра $x \in \mathcal{R}$, Если $\partial_y^{(\gamma)}(x)$ есть при каждом γ интервальные оценки x, то так как они вкладываются друг в друга, будет иметь место равенство $\{\partial^*{}_y(x) \geqslant \gamma\} = = \partial_y^{(\gamma)}(x)$, где слева стоит индикаторная функция множества. Таким образом, интервальные оценки для моделей $\mathcal{M}^{xy}_{(\gamma)}$, рассматриваемые как положенные друг на друга слои на высотах срезов $0 \leqslant \gamma \leqslant 1$, формируют вместе расплывчатое правило $\partial^*{}_y(x)$. Это и есть наглядная интерпретация (8.10).

Истинный уровень или ошибка правила (8.10) рассчитывается по формуле (8.8), где $1-\rho$ есть уровень (вероятность ошибки) оценки $\mu_z(\theta)$, а следовательно, ρ — надежность доверительной модели.

Мы рассмотрели один параметр $\theta = Mq$, задающий модель. Очевидно, все сказанное будет верно и для произвольного их числа, когда с целью построения доверительной модели находится совместная расплывчатая оценка $\mu_z(\theta)$ «вектора» $\theta = M\mathcal{Q} = = \{Mq,\ q \in \mathcal{Q}\}$ и эта оценка, в общем, не является интервальной, а имеет расплывчатый вид.

Адаптация, надежностное оценивание среднего при неизвестной дисперсии. Принцип адаптации состоит в сужении доверительной модели в процессе поступления наблюдений $y_1, \ldots, y_n$. Особенность в том, что «обучение» производится по тем же наблюдениям, по которым принимаются решения относительно состояний (оценивание, проверка гипотез); это первое. А во-вторых, уточнение модели осуществляется последовательно по n.

Для описания схемы адаптации рассмотрим один случай, когда требуется оценить параметр сдвига независимой выборки при неизвестной дисперсии флуктуаций. Адаптация состоит в оценивании дисперсий и подстановке в оценку параметра сдвига.

Пусть $y_i = x + \xi_i$, $i = 1, 2, \ldots$, и пусть требуется оценить x, когда ξ_i независимы, имеют нулевые средние $M\xi_i = 0$, ограниченные дисперсии $m_2 = M\xi^2{}_i < \infty$ и четвертые моменты $M\xi^4{}_i = \bar{m}_4 m^2{}_2$ (полезно представить $\xi_i = \sqrt{m_2}\zeta_i$, $M\zeta_i = 0$, $M\zeta^2{}_i = 1$, $M\zeta^4{}_i = \bar{m}_4$). При заданном m_2 и $\varkappa = 1$ оптимальной расчетного уровня $\alpha_{\text{рас}}$ будет оценка (6.18):

$$\partial_y(x) = [1 - \alpha_{\text{рас}}\, n\, (\hat{y} - x)^2 / m_2]^+.$$

На самом же деле, m_2 не известно и нужно подставлять в эту формулу заведомо завышенное значение $\bar{m}_2$, что ведет к чрезмерному увеличению расплывчатости оценки. Здесь m_2 и будет задающим модель параметром, причем считается он стационарным. Адаптация состоит в оценивании дисперсии m_2 по наблюдениям и ее использовании для оценки среднего.

Оценка дисперсии должна быть доверительной, так как по ней строится СИМ, и при этом не должна зависеть от x. Используем для нее статистику $\hat{\sigma}^2{}_y = \hat{y}^2 - \hat{y}^2$, инвариантную к сдвигам x. Вид оценки при уровне $1-\rho$ ее значимости будет определяться формулой (6.27) (при $x = m_2$, $l = 1$, $\sigma^2 = M\zeta^2{}_i = 1$, $\varkappa = 1$)

$$\mu_y(m_2) = [1 - (1-\rho)\, n\, (1 - \hat{\sigma}^2_y / m_2)^2 / c]^+,$$

где $c \approx (\bar{m}_4 - 1)$. Оценка $\mu_y(m_2)$ как функция m_2 принимает максимальное значение 1 при $m_2 = \hat{\sigma}^2{}_y$, она убывает по мере отклонения m_2 от этого значения, а при $m_2 \leqslant \hat{\sigma}^2_y \left[1 - \sqrt{\frac{c}{(1-\rho)\, n}}\right]^{-1}$ или $m_2 \geqslant \hat{\sigma}^2_y \left[1 + \sqrt{\frac{c}{(1-\rho)\, n}}\right]^{-1}$ оценка равна 0. Эта оценка и определяет расплывчатую доверительную модель (причем отнюдь не интервальную).

Для получения оценки x по расплывчатой модели используем методику предыдущего раздела. Согласно ей из неравенства $\mu_y(m_2) \geqslant \gamma$ находятся интервалы $\underset{\sim}{m}{}^\gamma{}_2$, $\tilde{m}^\gamma{}_2$, соответствующие различным высотам γ срезов. Нас интересует лишь верхняя граница $\tilde{m}^\gamma{}_2$, так как оценка $\partial_y(x)$ определяется только ею. Имеем

$\tilde{m}v_2 = \hat{\sigma}^2_y[1+\sqrt{c(1-\gamma)/(1-\rho)n}]^{-1}$ и по формуле (8.10) получаем

$$\partial^*_y(x) = \int_0^1 \left[1-\alpha_{\text{рас}}\, n\left(1-\sqrt{\frac{c(1-\gamma)}{(1-\rho)\,n}}\right)\frac{(x-\hat{y})^2}{\hat{\sigma}^2_y}\right]^+ d\gamma. \qquad (8.11)$$

Эта оценка эквивариантна к сдвигу: $\partial_{y+1a}(x+a) = \partial_y(x)$, и инвариантна масштабным изменениям: $\partial_{by}(x) = \partial_y(x)$, $b>0$.

Расчетный уровень $\alpha_{\text{рас}}$ и надежность ρ, как это показано в начале параграфа, связаны соотношением $\rho(1-\alpha_{\text{рас}}) = 1-\alpha$, где α — истинный (требуемый) уровень правила. Выразив $1-\rho = (\alpha-\alpha_{\text{рас}})/(1-\alpha_{\text{рас}})$ и подставив в (8.11), можно было бы найти оптимальное значение $\alpha_{\text{рас}}$, минимизирующее расплывчатость оценки (8.11).

Чтобы обойти технические громоздкости, найдем значение $\alpha_{\text{рас}}$, минимизирующее расплывчатость при заданном γ. Для этого нужно минимизировать по $\alpha_{\text{рас}}$ коэффициент

$$\alpha_{\text{рас}}\, n\,[1-\sqrt{c(1-\gamma)(1-\alpha_{\text{рас}})/n\,(\alpha-\alpha_{\text{рас}})}.$$

Дифференцируя его по $\alpha_{\text{рас}}$ и приравнивая 0, приходим к уравнению: $2\sqrt{(1-\alpha_{\text{рас}})(\alpha-\alpha_{\text{рас}})^3} - \sqrt{c(1-\gamma)/n}[2(1-\alpha_{\text{рас}})(\alpha-\alpha_{\text{рас}}) + \alpha_{\text{рас}}(1-\alpha)] = 0$. Нам нужно выявить качественную сторону, поэтому приближенно полагая $1-\alpha \approx 1$, откуда $1-\alpha_{\text{рас}} \geqslant 1-\alpha \approx 1$, и считая n достаточно большим, так что $c(1-\gamma)/n$ мало, получаем

$$\alpha_{\text{рас}} \approx \alpha - \sqrt[3]{c(1-\gamma)\,\alpha^2/4n},\ \rho = 1-\sqrt[3]{c(1-\gamma)\,\alpha^2/4n}.$$

Мы видим, что *при увеличении n расчетный уровень $\alpha_{\text{рас}}$ устремляется к истинному со скоростью $1/\sqrt[3]{n}$, а оптимально выбранное ρ с той же скоростью стремится к 1*, так что суммарная ошибка $\alpha_{\text{рас}} + (1-\rho) = \alpha$ есть истинный уровень.

Для получения конкретных расчетных значений $\alpha_{\text{рас}}$ и ρ вместо неизвестного γ нужно подставить среднее между 0 и 1 число, например $\gamma = 1/2$, и положить $c \approx \bar{m}_4 - 1$. Здесь $\bar{m}_4$, если оно не известно, также может оцениваться по наблюдениям или находиться из других соображений.

8.4. ЗАКЛЮЧЕНИЕ

Какую же выгоду все же сулит использование интервальных моделей? В качественном отношении они проигрывают точным (распределениям вероятностей), и это как будто бы ставит на них крест.

Представьте на миг, что наблюдается доселе невиданный объект и требуется описать его форму. Причем объект плохо различим (находится в тумане или далеко от нас). Ясно, что придется оговаривать условность сделанного описания вставками типа «вроде бы», «кажется», либо удовлетвориться грубым образом, уняв воображение по отношению к тому, что не различимо (наука — не фантастика).

То же самое и для моделей, которые видятся в большинстве реальных задач весьма смутно из-за конечности времени и выделенных средств на изучение явления, наконец, сменности, неустойчивости самих реальных явлений.

Прекрасно осознавая, что ошибочность модели тут же сделает никчемной ее дальнейшую эксплуатацию, постараемся осторожнее, т. е. в размытой форме описать грани модели, отобрав при этом лишь наиболее «видимую» их часть. И приходим к «чистокровным» интервальным моделям в их конструктивном задании набором первичных средних.

Построение модели в реальных условиях — сфера многосторонних (подчас, многострадальных) исследований: физических, экспериментальных статистических. Нас интересуют формальные способы, когда при полной начальной неясности в распоряжение предоставляются обучающие реализации. Тогда построение модели суть совместное оценивание отобранной части параметров, задающих модель (§ 8.1). Точечные оценки ведут к точным моделям. А доверительные оценки формируют доверительную модель заданной надежности. Вопрос сведется к выбору задающих модель параметров. Ими могут быть вероятности, совместное оценивание которых рассматривается в § 8.2.

Другой путь выбора модели, свойственный классическому подходу, состоит в проверке согласия, что модель имеет выдвинутый конкретный вид (например, нормальная). Этот путь, обязанный во многом крайней узости рабочего арсенала точных моделей, заставляет с самого начала «довольствоваться» заранее выбранным гипотетическим в меру простым вариантом. Принятие последнего, если согласие имеет место, тем не менее не приведет к сколь-либо надежной модели, а всего лишь установит вхожесть гипотетического варианта в нашу доверительную ИМ-модель в качестве составной части. В конечном же счете это будет «выхватывание» из доверительной модели ее заранее сформированного кусочка — подход, свойственный режиму оптимизма.

Мы же поступаем значительно осторожнее, используя доверительную модель всю целиком (необъятный арсенал ИМ позволяет произвести любой выбор), регулируя ширину с помощью надежности. Чем больше положить надежность, тем шире доверительная модель, от чего пострадает конкретность и качество выводов при эксплуатации модели. Наоборот, меньшая надежность, казалось бы, выводы сделает более конкретными и качественными, но доверие к ним уменьшит из-за утраты верности модели. Выход из этого заколдованного круга состоит в совместном рассмотрении тандема модель-выводы (§ 8.3), беря в расчет оба типа ошибок: за счет ненадежности модели и при эксплуатации из-за случайности наблюдений. Совместное рассмотрение заставляет вводить поправки в расчетные ошибки эксплуатации, увеличивая их сообразно ненадежности модели, что ведет к истинным ошибкам, а при более широком изложении — к истинному риску. Это как раз и есть то, что объективно нужно для целей анализа и синтеза решающих правил.

Трудности в том, что доверительные модели по наследству от породивших их доверительных оценок, в общем, расплывчатых, неиндикаторных, становятся размытыми по форме средних, т. е. с размазанными итервалами средних. Определение понятия оптимального правила при размытых статистических моделях позволило в § 8.4 рассмотреть совместную картину синтеза модели по ее единственному задающему параметру — дисперсии, с последующим нахождением оптимальной оценки параметра сдвига. Все вместе это выглядит как доверительное оценивание дисперсии (уровень доверия которой и станет надежностью модели) с последующим использованием ее при доверительном оценивании сдвига при своем уже доверии (расчетной ошибке). Причем все про-

изводится по одной и той же выборке наблюдений, по мере удлинения которой происходит уточнение оценки дисперсии, т. е. адаптационное сужение модели.

Подчеркнем еще раз, что главным итогом рассмотрения тандема модель-правило явился истинный риск (истинные ошибки) синтезированного правила. А надежность модели есть всего лишь вспомогательный атрибут синтеза, приобретающий конкретное значение путем минимизации истинного риска. Такая оптимальная надежность существует и найдена в рассмотренной нами задаче адаптации § 8.4, где установлена ее тенденция с ростом длины выборки стремиться к единице со скоростью кубического корня.

Теперь мы можем дать обоснованный ответ на поставленный в начале вопрос: выгода интервальных моделей состоит в получении объективно надежных обоснованных во всех отношениях решающих правил с оценкой их истинных качеств. Это и есть главное достижение надежностного синтеза.

СПИСОК ЛИТЕРАТУРЫ

1. **Лоэв М.** Теория вероятностей/Пер. с англ. под ред. Ю. В. Прохорова. — М.: ИЛ, 1962. — 719 с.
2. **Уиттл П.** Вероятность/Пер. с англ. под ред. В. В. Сазонова. — М.: Наука, 1982. — 287 с.
3. **Леман Э.** Проверка статистических гипотез/Пер. с англ. под ред. Ю. В. Прохорова. — М.: Наука, 1964. — 498 с.
4. **Закс Ш.** Теория статистических выводов/Пер. с англ. под ред. Ю. К. Беляева. — М.: Мир, 1975. — 776 с.
5. **Wald A.** Statistical Decision Functions — N. Y.: Wily, 1950.
6. **Тарасенко Ф. П.** Непараметрическая статистика. — Томск: Изд-во Томского университета, 1976. — 291 с.
7. **Гаек Я., Шидак З.** Теория ранговых критериев/Пер. с англ. под ред. Л. И. Большева. — М.: ФМЛ, 1971. — 375 с.
8. **Кузнецов В. П.** Некоторые обобщения ранговых критериев//Мат. статистика и ее прилож.: Труды СФТИ. — Томск. 1974. — Вып. 6. — С. 70—108.
9. **Кузнецов В. П.** Инвариантность решений по отношению к мешающим параметрам//Проблемы передачи информации. — 1971. — № 4. — С. 36—44.
10. **Кузнецов В. П.** Инвариантность решений по методу максимального правдоподобия по отношению к мешающим параметрам//Проблемы передачи информации. — 1972. — № 3. — С. 38—47.
11. **Хьюбер П.** Робастность в статистике/Пер. с англ. под ред. И. Г. Журбенко. — М.: Мир, 1984. — 303 с.
12. **Кассам С., Пур Г.** Робастные методы обработки сигналов//ТИИЭР. — 1985. — Т. 73. — № 3. — С. 54—110.
13. **Кузнецов В. П.** Минимаксные критерии при ограниченных семействах плотностей распределения//Теория вероятностей и ее применения. — 1982. — Вып. 2. — С. 286—295.
14. **Де Гроот М.** Оптимальные статистические решения/Пер. с англ. под ред. Ю. В. Линника. — М.: Мир, 1974. — 491 с.
15. **Заде Л.** Понятие лингвистической переменной и его применения. — М.: Мир, 1976. — 168 с.
16. **Шокин Ю. И.** Интервальный анализ. — Новосибирск: Наука, 1981. — 112 с.
17. **Карлин С., Стадден В.** Чебышевские системы и их применение в анализе и статистике/Пер. с англ. под ред. С. М. Ермакова. — М.: Наука, 1976. — 567 с.
18. **Обработка** нечеткой информации в системах принятия решений/А. Н. Борисов, А. В. Алексеев и др. — М.: Радио и связь, 1989. — 304 с.
19. **Кузнецов В. П.** Интервальная мера и интеграл//Численный анализ и задачи интерпретации экспериментов: Межвузов. сборник. — Красноярск, 1987. — С. 75—95.
20. **Колмогоров А. Н., Фомин С. В.** Элементы теории функций и функционального анализа. — М.: Наука, 1972. — 496 с.
21. **Кузнецов В. П.** Интервальные модели вероятностей //Мат. статист. и ее прилож.: Труды СФТИ. — Томск. — 1986. — Вып. 10. — С. 128—151.
22. **Экланд И., Темам Р.** Выпуклый анализ и вариационные проблемы: Пер. с англ. — М.: Мир, 1979. — 399 с.
23. **Ахиезер Н. И.** Лекции по теории аппроксимации. — М.: Наука, 1965. — 407 с.
24. **Кульбак С.** Теория информации и статистика/Пер. с англ. под ред. А. Н. Колмогорова. — М.: Наука, 1967. — 408 с.
25. **Ван дер Варден Б.** Математическая статистика. — М.: ИЛ, 1960. — 518 с.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Адаптация 335
Аксиомы 15
— интервальных вероятностей 49, 50
— обращения 16
— переноса 15
— полуаддитивности 16
— сохранения порядка 15
— средних 15
Алгебры событий 44
— — изоморфные 74
— — счетные (сигма-алгебры) 45, 46, 51, 55
Арифметика интервальная 82
— размытая 91

Белый шум 188, 192

Вероятности 12
— интервальные 13
— относительные 49
— ошибок 204, 205
— — весовые суммы 276
— — интегральные 305
— — истинные 332
— — первого и второго рода 276
— — прикидочные 293
— — расчетные 332
— — совокупные 307
— первичные 35, 38, 40, 45, 109, 124
— правильных решений 307
— превышений (выбросов) 22, 163, 282
— размытые 90
— согласованные 34
— точные 12, 43
— условные 64, 65
Вершины модели 32, 56, 66, 81

Гипотеза нулевая 275
— альтернативная 275
Группы преобразований 214, 236, 281

Дискретизация 164
Дисперсия с. в. 58, 123, 266, 335
Доверительные модели *см.* СИМ
— интервалы 198, 227, 239, 247
Допредельная проблема 141
Допредельные неравенства 140, 147
Достаточность глобальная 207
— класса правил 207
— — — проверки гипотез 277
— — — различения гипотез 307
— — — расплывчатого оценивания 233
— множества решений 209
— набора признаков 211
— преобразований 210
— специальная 207, 232, 277

Задача надежностного синтеза 319
— проверки гипотез 276
— — — смежная 276
— различения гипотез 307
— расплывчатого оценивания 230
— статистическая 204
Закон больших чисел 132
— — — неустойчивый 135
— — — устойчивый 133

Измеримость 43
Изображения наблюдений 87
— признаков 84
— событий 73
— точек 73
ИМ (интервальные модели) *см.* модели интервальные
Инвариант к группе 214
— — — максимальный 215
Интервал корреляции процесса 165
ИРВ (интервальные распределения вероятностей) *см.* распределения вероятностей интервальные

Квантование 165
Кольцо событий 44
Ковариационные границы 173, 190
— функции 173
— — однородные 180
Корреляционная матрица
— — неточная 254, 291, 315
— — точная 221, 253, 289, 309
— функция 24, 163
— — наименее благоприятная 254, 255
Корреляционные свойства 170
Коэффициент пессимизма 204

Линдеберга-Феллера условие 153

Мера 42
Мера-длина 45
Модели 14, 317
— интервальные (ИМ) 16
— — абстрактно-условные 69
— — включение 29, 68, 80, 94, 122
— — голые 24, 29, 32, 67, 80, 94, 104, 198, 199
— — индикаторные 24, 67, 69, 82, 85, 135
— — моментные 62
— — объединения 31, 56, 68, 80, 94, 110, 179, 182, 234, 319, 320
— — пересечения 31, 68, 81, 94, 179, 182
— — переходные 78, 170, 197
— — предельные 37, 45, 51
— — простые 52
— — процесса 162
— — пустые 30
— — разложимые 96
— — размытые 90
— — совместные 92
— — стандартные 58
— — условные 63
— — частные 93
— статистические интервальные *см.* СИМ
Модифицированная формула продолжения 25
Моменты начальные 27, 139, 163, 256, 268
— — абсолютные 27, 137
— — центральные 58, 124, 145
Мощность случайной величины 21, 266
— процесса 24, 163
— средняя 244, 283
Мультипликативность интервальная 82, 106

Надежность доверительной модели 323
— — — оптимальная 332
Надмодель 322
Независимое произведение моделей *см.* произведение моделей
Независимость последовательности с. в. 125
— явлений 106
Нековариированность алгебр событий 116
— классов признаков 115
— признаков 108
— случайных величин 115
— элементарных исходов 113
Некоррелированость 115, 126, 220, 249, 251, 268, 290
Неравенства 129
Нормальная с. в. *см.* случайная величина нормальная

Область существования средних 14
— — — предельная 37
Оболочка линейная 26
— полулинейная 18, 26
Образ признака 75
— события 73
Обучающие испытания 321
Отображения *см.* преобразования
Оценки
— детерминированные 217, 218
— расплывчатые (доверительные) 228, 292
— — амплитуды сигнала 253, 254
— — вероятностей 324
— — дисперсии нормального распределения 240
— — интервальные 198, 227, 247
— — контрастные 246
— — масштаба 264
— — мощности 267
— — оптимальные 230
— — регрессии 239, 247, 252
— — сдвига 244, 249, 251, 335
— — совместные 329
— — степенного типа 256
Ошибки *см.* вероятности ошибок

Параметры задающие 319
— мешающие 199
— нормировки 266
— подчиняющие 101
— стационарные 127
Первичный набор 18
Плотность вероятностей 60, 70, 123, 200, 239, 296
— — апостериорная 231
— — интервальная 207, 297, 304
— — наименее благоприятная 298
— — совместная 207, 231
— — переходная 85
— — частная 231
— формальная 61, 295, 296
Подобие ИМ 77 (*см.* преобразования подобия)
— — случайное 86
— ИРВ 47
Полуаддитивность 16, 34
Последовательности *см.* случайные последовательности
Потери
— дельта 202, 218
— квадратичные 202, 220
— составные 202
Правила
— взвешенного правдоподобия 219
— детерминированные 217, 308
— квазиоптимальные 219
— контрастные 201, 276, 309
— минимаксные 205
— оптимальные 205
— — асимптотически 261
— — при оптимизме 234, 302
— — при полуоптимизме 206
— оценивания *см.* оценки
— проверки гипотез 277
— — — асимптотические 293
— — — контрастные 276, 288
— — — о значении параметра 301
— — — оптимальные 276
— — — равномерно оптимальные 276

— — рандомизированные 275
— — — уровня α 303
— различения гипотез 306
— — — детерминированные 308
— — — оптимальные 277, 309
— расплывчатые *см.* оценки доверительные
— решающие 202
— эквивариантные 237
— уровня α 228
Предельные теоремы 145, 148, 150, 152, 153, 157
Представление моделей 52
— — объединениями 54, 199
— — пересечениями 31
— — функциональные 59, 199, 212, 234
— последовательности аддитивно-мультипликативные 154
— — рекуррентные 128
— процесса аддитивные 167
— — линейные 166
— — мультипликативные 167
— — подчиненно-аддитивные 168, 173
— — расширенно-аддитивные 169, 174
— — свободно-аддитивные 169
— — функциональные 166
Преобразования 72
— детерминированные 73, 75, 76, 122
— изоморфные 74, 86
— индикаторные 82, 85
— линейные 188
— моделей 75
— нелинейные безынерционные 24, 179
— признаков 58, 74
— подобия 77, 86, 164, 165, 210, 215
— простые 83, 85
— случайные 78
— совместные 236
— частные 236
— числовые 211
— Фурье 184
Признаки 10
— вторичные 18
— гармонические 22, 123, 139, 162
— гибридные 162
— дельта 10, 96
— измеримые 43, 55, 69, 117
— инвариантные к группе преобразований 215
— индикаторные 10, 22, 162
— квадратичные 138, 161, 263
— нековариированные 108
— линейные 138, 161
— определяющие модель 30
— первичные 18
— представимые 75
— случайных величин 21
— — процессов 24, 160
— совместные 92
— стационарные 127
— степенные 27, 139
— центрированные 25
— частные 92
— экспоненциальные 139, 157
Проверка гипотез *см.* правила проверки гипотез
Полулинейность класса признаков 15
Продолжение вероятностей 35
— средних 19
— — модифицированная формула 25
— — предельное 36
Произведение моделей 95, 213
— — независимое 109, 117, 199
— — подчиненное 101
— — свободное 102, 109
Пространство предметное 87
— элементарных исходов 9
Процессы 160
— белый шум 192
— второго порядка 171, 183
— непрерывные 160, 165
— ограниченные 160, 165
— — сугубо 169
— однородные 178, 190
— — в широком смысле 180
— — частично 179
— первого порядка 180, 183
— скв-непрерывные 177, 186
— спектральные 187
— стационарные 181
— — в широком смысле 184, 187, 191
— — частично 181
— узкополосные 192
Разложения процессов по базису 175
— — спектральные (в ряды Фурье) 184
Размах оценки 256
Размерности моделей 21, 28
Рандомизация 275
Распределения вероятностей
— — интервальные (ИРВ) 35
— — — конечно-аддитивные 38, 77, 219
— — — счетно-аддитивные 41
— — точные (классические) 43, 169
— — — гибридные 51
— — — конечно-аддитивные 38, 44, 50
— — — Коши 51
— — — наименее и наиболее благоприятные 296
— — — равномерные 45, 51
— — — совместные 111
— — — счетно-аддитивные 45, 51, 55
— — — условные 64
— — — хи-квадрат 240, 242, 328
— — — частные 111
Расширение моделей 30
Решающие правила *см.* правила
Решения 201
— индикаторные 201
— интервальные 201
— контрастные 201
— нейтральные 306
Риск 228
— истинный 333
— составной 230, 233
— средний 203
Свободное произведение моделей *см.* произведение моделей
Свойства
— вероятностей интервальных 33
— — точных 44
— независимого произведения 110
— независимости 107
— средних 16, 17
— операций объединения и пересечения 32
— условных интервальных моделей 67
Семейства моделей 55, 212
— векторов вероятностей 27, 66, 111
— плотностей 200, 207, 297, 304
— распределений 48, 55, 81, 112, 198, 200, 220
— собственные *см.* собственные семейства
Сечения моделей 52, 69, 136
— — вероятностями 54, 64
— — задающие 56
СИМ (статистические интервальные модели) 197
— доверительные 323
— инвариантные 215, 280
— неразложимые 197
— подобные 210
— разложимые 197, 217, 234
— размытые 333
— робастные 200, 297
— симметричные 215, 280
— эквивариантные 237
Случайные величины 21, 121
— — дескриптивные 130
— — дискретные 21, 121
— — непрерывные 121
— — нормальные 123, 145, 153, 240, 259
— — — с интервальными средним и дисперсией 124, 154
— — — стандартные 123
— — ограниченные 121
— — симметричные 122, 143, 158, 258, 293
— — центрированные 58, 133
— последовательности 125
— — дескриптивные 131
— — зависимые 128
— — независимые 125, 137, 141, 147, 262, 264, 322
— — нековариированные 126
— — некоррелированные 126, 248, 249, 251, 258, 268
— — однородные 126, 141, 249, 258
— — свободные 126
— — стационарные 127, 322, 324
Смежная задача проверки гипотез 276
Собственные семейства ковариаций 172, 189, 223
— — средних 168, 172, 189
События 9
— вторичные 38
— нечеткие 10, 87
— первичные 35
— элементарные 9
Спектры энергетические 186, 191, 225, 290
Спектральные коэффициенты 184
— двойники 186, 290
— процессы 187
Среднее арифметическое 133
Среднеквадратическое *см.* мощность
Средние (статистические) 13
— абсолютные 139
— верхние 13
— гармонические 22, 124, 140, 261
— — абсолютные 140
— интервальные 13
— несогласованные 20
— нижние 13
— первичные 18
— — непротиворечивые 18
— признаков 12
— процессов 163
— размытые 89
— случайных величин 21, 22, 58, 122
— согласованные 16, 19, 20
— точные 12, 13
— условные 65, 104
Срезы 89, 334
Статистические интервальные модели *см.* СИМ
Стационарность 127, 128, 181
Суждения нечеткие 88
Сходимость ИМ 130
— случайных величин 131
— — в среднеквадратическом 131
— — в среднем 137
— — по вероятности 137
— — почти всюду 137
Теоремы
— о представлении ИМ 54
— продолжения 19
— разложимости 97

— факторизации 213
— характеризации нормальной с. в. 124

Универсальный класс признаков 139
Уровень 228, 276

Фильтр однородный 190
Фильтрация линейная 220
Функция
— Лапласа 123, 240, 259, 329
— инвариантная 214
— распределения интервальная 46
— — доверительная 330
Фурье ряды и преобразования 184

Цена задачи 205

Шкалы расплывчатости оценок 229
— — — взвешенная 229, 231
— — — интегральная 229
— — — обобщенная 229, 232
Штраф за расплывчатость 230

Энергетические спектры *см.* спектры энергетические

Явления случайные 9
— — независимые 106, 115
— — — подчиненно 112
— — свободные 115
— — статистически неустойчивые 12
— — статистически устойчивые 12

ОГЛАВЛЕНИЕ

Введение 3

ЧАСТЬ ПЕРВАЯ. ИНТЕРВАЛЬНЫЕ МОДЕЛИ 9

Глава 1. Описание случайных явлений 9

1.1. Интервальные вероятности и средние 9
Пространство исходов (9). Признаки явления (10). Средние значения признаков (11). Интервальные средние и вероятности (13). Математическая модель явления (14). Аксиоматика (15). Определение интервальной модели средних, основные свойства (16)

1.2. Продолжение первичных средних 17
Вступление (17). Первичные признаки и средние (18). Теорема продолжения и согласования средних (19). Согласованные первичные средние (20). Признаки случайных величин (21). Признаки случайных процессов (24). Голая модель (24). Модифицированная формула продолжения (24). Дополнения (25)

1.3. Отношения между интервальными моделями 27
Геометрическая иллюстрация ИМ (27). Обсуждение (29). Иерархия моделей (29). Пересечение ИМ (31). Объединение ИМ (31). Свойства операций (32). Дополнения (33)

1.4. Интервальные распределения вероятностей 33
Свойства интервальных вероятностей (33). Продолжение первичных вероятностей (35). Предельное продолжение средних (36). Иллюстрация ИРВ (37). Конечно-аддитивные ИРВ (38). Счетно-аддитивные ИРВ (40). Обобщения (42). Точные распределения вероятностей (43). Интервальные функции распределения (46). Подобие ИРВ (47). Семейства распределений (48). Относительные вероятности и средние (49). Дополнения (49)

1.5. Представления моделей 52
Предисловие (52). Сечения модели (52). Свойства сечений (53). Теорема о представлении ИМ (54). Определение ИМ задающими сечениями (56). Представление через стандартную ИМ (58). Функциональные представления (59). Плотность (59). Дополнения (62)

1.6. Условные интервальные модели 63
Постановка проблемы (63). Определение условной интервальной модели (64). Расчет условных моделей через вершины (66). Некоторые свойства условных интервальных моделей (67). О восстановлении безусловной модели по условным (68). Абстрактно-условные модели (69)

1.7. Заключение 71

Глава 2. Совместный анализ 72

2.1. Детерминированные преобразования исходов 72
Отображения (72). Преобразования признаков (74). Расчет средних (75). Подобие моделей (77)

2.2. Случайные преобразования 78
Переходные модели (78). Преобразования моделей (78). Свойства преобразований модели (80). Индикаторные преобразования, интервальная арифметика (82). Простые преобразования (83). Дополнения (85)

2.3. Нечеткие события и размытые вероятности 87
Наблюдения и их изображения (87). Размытые вероятности и средние (88). Размытые действия (91)

2.4. Совместные интервальные модели 91
Совместные и частные интервальные модели (91). Представление совместных моделей случайными преобразованиями (95). Восстановление сомножителей разложимой модели (96). Разложимость совместной модели (97). Первичные средние разложимых ИМ (98). Подчиненные произведения (101). Свободные произведения (102). Дополнения (104).
2.5. Независимость 105
Определение независимости (105). Свойства независимости (107). Независимое произведение (108). Независимые произведения на дискретных пространствах исходов (111). Геометрическая иллюстрация независимости (113). Нековариированность случайных величин (114). Независимость, свобода, нековариированность (115). Дополнения (117)
2.6. Заключение 119

Глава 3. Случайные величины, последовательности, суммы 121
3.1. Случайные величины, последовательности 121
Определения (121). Детерминированные преобразования (122). Нормальная случайная величина (123). Случайные последовательности (125). Однородность и стационарность последовательности (126). Зависимые последовательности (128)
3.2. Сходимости 129
Неравенства для случайных величин (129). Сходимость моделей (130). Сходимость случайных величин и сходимость их моделей (131). Сходимость среднего арифметического, закон больших чисел (132). Закон больших чисел для неустойчивых последовательностей (135). Дополнения (136)
3.3. Допредельная и предельная проблемы 137
Аппроксимация модели суммы независимых с. в. (137). Гармоническая аппроксимация (139). Допредельная проблема, однородный случай (141). Введение в предельную проблему (144). Дополнение (146)
3.4. Предельные модели сумм общего вида 147
Центральные допредельные неравенства (147). Первая ослабленная предельная теорема (148). Вторая ослабленная предельная теорема (149). Третья ослабленная предельная теорема (152). Центральная теорема нормальной сходимости (152). Интервальная нормальная сходимость (154). Дополнения (156)
3.5. Заключение 158

Глава 4. Случайные процессы 159
4.1. Описания случайных процессов 159
Принципы описаний (159). Реализации и признаки (160). Модель процесса (162). Характерные черты процессов (165). Дробление процесса на составляющие (166). Функциональные представления (166). Различные аддитивные представления (167). Дополнения (169)
4.2. Корреляционные свойства 170
Процессы второго порядка (170). Представление процессов второго порядка семействами средних и ковариационных функций (172). Интервальные ковариации и корреляции (173). Разложение процесса по базису (175)
4.3. Однородные и стационарные процессы 178
Однородные процессы (178). Стационарные процессы (181). Спектральные двойники процессов (184). Спектральные процессы (186)
4.4. Линейные преобразования процесса 188
Гладкость преобразований и непрерывность процессов (188). Расчет выхода фильтра (188). Линейное преобразование и представление стационарного процесса (190). Узкополосные процессы (192)

ЧАСТЬ ВТОРАЯ. СТАТИСТИЧЕСКИЙ СИНТЕЗ 195

Глава 5. Теория принятия решений 195
5.1. Статистические модели 195
Что такое математическая статистика? (195). Статистические интервальные модели (197). Функциональные представления наблюдений (199). Модели с мешающими параметрами (199). Робастные модели (199)
5.2. Оптимальные правила 200
Расплывчатые решения и решающие правила (200). Потери (202). Риск (203). Статистическая задача (204). Оптимальность и пессимизм (205). Проблема достаточности (207). Достаточность и функция потерь (208)
5.3. Достаточная редукция наблюдений 209
Теорема о представимости (209). Первичные признаки и достаточность (211). Достаточные преобразования и факторизация (212)
5.4. Редукция наблюдений и инвариантность 214
Инвариантные модели (214). Симметрия, инвариантность и достаточность (215)
5.5. Детерминированные решения и фильтрация 217
Общие соображения (217). Оптимальные решения при дельта-потерях (218). Постановка задачи линейной фильтрации сигнала при квадратичных потерях (220). Фильтрация сигнала с известными корреляционными свойствами из шума ограниченной мощности (221). Фильтрация при некоррелированном шуме (222). Корреляции заданы с погрешностями (223)
5.6. Заключение 225

Глава 6. Расплывчатое оценивание 227
6.1. Общие вопросы 227
Ошибки правил (227). Расплывчатость, риск (228). Оптимальные расплывчатые правила при заданных совместных плотностях вероятностей (230). Достаточные классы расплывчатых правил (232). Оптимизм и достаточность (235). Симметрия статистических моделей и эквивариантность расплывчатых правил (236)
6.2. Доверительное оценивание при заданных распределениях вероятностей флуктуаций 238
Предисловие (238). Оценка регрессии при известной плотности вероятностей (239). Доверительное оценивание дисперсии (241)
6.3. Оценка параметров регрессии по энергетическим и корреляционным данным о флуктуациях 243
Обоснование (243). Оценка параметров сдвига при заданной мощности флуктуаций (244). Развитие энергетического типа оценивания (246). Оптимальная оценка параметра сдвига при однородных некоррелированных флуктуациях (249). Оценивание сдвига при неоднородных некоррелированных флуктуациях (251). Обобщения оценок (252). Оценка амплитуды сигнала при колебаниях его формы и неточных корреляциях шума (254)
6.4. Оценивание параметра сдвига по моментам и гармоническим средним 256
Оценивание по моментам (256). Асимптотическая подстройка оценки степенного типа (258). Использование допредельных и предельных результатов (259). Синтез квазиоптимальных оценок по гармоническим средним (261). Об оценивании параметра сдвига при неоднородных флуктуациях (264)
6.5. Доверительное оценивание параметра масштаба 264
Общие соображения (264). Оценивание параметра масштаба по заданной мощности флуктуаций (266). Оценивание параметра масштаба по некоррелированной выборке (268). Развитие проблемы (270)
6.6. Заключение 272

Глава 7. Проверка гипотез 274
7.1. Общие положения 274
Введение (274). Математическое оформление задачи (275). Основная теорема о достаточности (277). Комментарии (279). Инвариантность и симметрия (280). Обнаружение сигнала по вероятностям превышений (282)
7.2. Корреляционная теория проверки гипотез 283
Получение оптимального правила при заданной средней мощности наблюдений (283). Общая форма правила (286). Проверка гипотез по заданным корреляциям (289). Неточные корреляции (291)
7.3. Использование доверительных оценок для проверки гипотез . . . 292
Описание способа (292). Асимптотическое правило при симметричных ограниченных флуктуациях (293). Проверка гипотез по мощности флуктуаций (294)
7.4. Специальные методы синтеза правил 295
Задана формальная плотность альтернативы по отношению к гипотезе (295). Точные плотности вероятностей (296). Робастные методы (296). Проверка гипотез по заданным интервальным вероятностям (297). Робастный алгоритм при независимых наблюдениях (299)
7.5. Проверка гипотез о заданном значении параметра 301
Формулировка задачи (301). О правилах при оптимизме (302). Равномерно оптимальные правила (303). Введение защитного диапазона (304). Минимизация интегральной ошибки (304). Использование доверительных оценок (305)
7.6. Различение нескольких гипотез 306
Общие положения (306). Различение гипотез по заданным корреляциям (308). Оптимальное правило различения двух гипотез (310). Более двух гипотез (311). Неточно известные корреляции (315)
7.7. Заключение 315

Глава 8. Надежностный синтез 317
8.1. Общие вопросы синтеза моделей 317
Методология синтеза моделей (317). Постановка задачи (319). Стационаризация статистических параметров (321). Понятие доверительной модели (322)
8.2. Построение доверительной модели на заданном наборе событий . . 324
Исходные положения (324). Модель наибольшего правдоподобия (326). Использование критерия хи-квадрат (328). Информационный критерий построения доверительной модели (328). Доверительные совместные оценки (329). Доверительная функция распределения (330)
8.3. Согласованный синтез моделей и правил 331
Надежность моделей и истинные ошибки правил (331). Размытые доверительные модели и решения (333). Адаптация, надежностное оценивание среднего при неизвестной дисперсии (335)
8.4. Заключение 336

Список литературы 339

Предметный указатель 340

Научное издание

КУЗНЕЦОВ Владимир Петрович

ИНТЕРВАЛЬНЫЕ СТАТИСТИЧЕСКИЕ МОДЕЛИ

Заведующий редакцией В. Н. Вяльцев. Редактор В. К. Старикова.
Переплет художника А. С. Дзуцева. Художественный редактор А. В. Проценко.
Технический редактор О. А. Гришкина. Корректор Т. Л. Кускова

ИБ № 1914

Сдано в набор 15.01.90 Подписано в печать 17.04.91
Формат 60×90¹/₁₆ Бумага типограф. № 2 Гарнитура литературная
Печать высокая Усл. печ. л. 22,0 Усл. кр.-отт. 22,0 Уч.-изд. л. 24,11
Тираж 2200 экз. Изд. № 22562 Зак. № 13 Цена 4 руб.
Издательство «Радио и связь». 101000 Москва, Почтамт, а/я 693

Типография издательства «Радио и связь» 101000 Москва, ул. Кирова, д. 40