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 $\mathcal{P}^{y}:\mathcal{M}^{y}{}_{x}=\bigvee_{\mathcal{P}^{y}\in\mathcal{M}^{y}_{x}}\mathcal{P}^{y}$, $x=x_0$ 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 $\mathcal{P}^{y}{}_{1}=\mathrm{S}\mathcal{P}^{y}{}_{0}$, получаем

$$\overline{M}f(x,\ y)=\sup_{\mathcal{P}_0\in\mathcal{M}^{y}_{0}}\overline{M}^{x}\left[\delta_{x_0}(x)\,\overline{M}^{y}_{\mathcal{P}_0}f(x,\ y)+\delta_{x_1}(x)\,\overline{M}^{y}_{\mathrm{S}\mathcal{P}_0}f(x,\ y)\right].$$

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: $\overline{M}f(x,\ y)=\sup_{x}\overline{M}^{y}{}_{x}f(x,\ y)$, 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

$$y=\mathrm{V}_{x}\xi,\ x\in\mathcal{X},\ y\in\mathcal{Y},\ \xi\in\Xi,\qquad(5.1)$$

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 $\mathcal{M}^{xy}=\bigvee_{\theta}\mathcal{M}_{\theta}{}^{xy}$. 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 $y=\mathrm{V}_{\theta,x}\xi$, 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: $(1-\varepsilon)p_0(z)+\varepsilon p_{\mathrm{V}}(z)$, где $p_0(z)$ — заданная, $p_{\mathrm{V}}(z)$ — совершенно неизвестная плотность, и тогда $\underset{\sim}{p}(z)=(1-\varepsilon)p_0(z)$. Семейство записывается через функции распределения: $(1-\varepsilon)F_0(z)+\varepsilon F_{\mathrm{V}}(z)$, где $F_{\mathrm{V}}(z)$ — 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_{V}. В D_{V} 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: $D_{\text{дет}}\subset D_{\text{и}}\subset D_{01}\subset D_{\mathrm{V}}$.

Вернемся к множеству D_{V} 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_{V} 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 $d(x') = [\underset{\sim}{x}, \hat{x}]$ можно было бы, скажем, положив $\pi(x, d(x')) = \min\{\|x - \underset{\sim}{x}\|^2, \|x - \hat{x}\|^2\}$.

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, останется 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* $\pi(x, \partial_y)$ *в режиме полуоптимизма* $\varkappa=1/2$ *средний риск любого правила* ∂_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 $\overline{\pi}(x, \hat{x}) = \int_0^T (x_t - \hat{x}_t)^2 dt$ равны среднему квадрату ошибки. Пусть СИМ $\mathscr{M}^{xy}$ задана корреляционными свойствами $\bar{M}(c + \sum c_i X_{t_i} + \sum\sum d_{ij} X_{t_i} X_{t_j})$. 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 $\tilde{M}\mathring{g} = 0$), положив для этого $\mathring{g} = g - \tilde{M}g$, $g \in \mathscr{G}$. Обозначим $\mathring{\mathscr{G}}$ — центрированный набор. Тогда согласно модифицированной формуле продолжения (1.3) нижний и верхний средние риски запишутся:

$$\underline{\Pi}(\partial) = - \inf_{g \in \mathscr{L} + \mathring{\mathscr{G}}} \sup_{x, y} [-\underline{\pi}(x, \partial_y) - g(x, y)];$$

$$\overline{\Pi}(\partial) = \inf_{g \in \mathscr{L} + \mathring{\mathscr{G}}} \sup_{x, y} [\overline{\pi}(x, \partial_y) - g(x, y)].$$

Отсюда видно, что риски в определенном смысле равны величине наилучшего приближения функции потерь (для верхнего — сверху, для нижнего — снизу) центрированными (с нулевыми $\tilde{M}\mathring{g}$) вторичными признаками СИМ из $\mathscr{L} + \mathring{\mathscr{G}}$. Достаточными будут классы таких правил, для которых $-\underline{\pi}(x, \partial_y)$ и $\overline{\pi}(x, \partial_y)$ лучше других аппроксимируются сверху функциями класса $\mathscr{L} + \mathring{\mathscr{G}}$. 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 $\varkappa \geqslant 1$).

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* $\mathscr{M}^{xy}=\langle\tilde{M}\mathscr{G}\rangle$ *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 $\mathscr{L}+\mathscr{G}$ 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 ∂^*_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:

$$\underline{\Pi}(\partial)=\underline{M}^{xy}\inf_{y'\in Q^{-1}Qy}\underline{\pi}(x,\ \partial_{y'})\leqslant\underline{M}^{xy}\underline{\pi}(x,\ \partial^*_y)=\underline{\Pi}(\partial^*);$$

$$\overline{\Pi}(\partial)=\overline{M}^{xy}\sup_{y'\in Q^{-1}Qy}\overline{\pi}(x,\ \partial_{y'})\geqslant\overline{M}^{xy}\overline{\pi}(x,\ \partial^*_y)=\overline{\Pi}(\partial^*).$$

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 ∂^*_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 $\mathscr{M}^{xy}=\langle\tilde{M}\mathscr{G}\rangle$. 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* $\tilde{M}_x\psi_x(y)$, $\psi_x(y)\in\Psi_x$. *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 $\bar{M}^{xy}f(x, y)=\bar{M}^x\bar{M}^\xi f(x, V_x\xi)=\bar{M}^x\bar{M}^y{}_xf(x, y)$, 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 $\bar{M}\psi(\int\xi_tdt)$, $\psi\in\Psi$. 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 $\int y_tdt$, 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* $\mathcal{M}^{xy}=\bigvee_\theta\mathcal{M}_\theta{}^{xy}$ *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: $\overline{\Pi}(\partial)=\bar{M}^{xy}\overline{\pi}(x, \partial_y)=\bar{M}^{xz}\bar{M}^y{}_z\overline{\pi}(x, \partial_y)\geqslant\bar{M}^{xz}\overline{\pi}(x, \bar{M}^y{}_z\partial_y)=\bar{M}^{xz}\overline{\pi}(x, \partial^*{}_z)$, где $\partial^*{}_z=\bar{M}^y{}_z\partial_y$; отсюда $\overline{\Pi}(\partial)\geqslant\overline{\Pi}(\partial^*)$.

Условие б) теоремы выполнится, если множество D решений есть выпуклое подмножество числового пространства, а $\mathcal{D}=\mathcal{D}_\forall$ 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* $\Upsilon=(\mathcal{M}^{xz}, \mathcal{M}^y{}_z, \mathcal{D}, [\underline{\pi}, \overline{\pi}], \varkappa)$, *где* $z=Qy$, *и переходные средние* $M^y{}_z\underline{\pi}(x, \partial_y)$, $M^y{}_z\overline{\pi}(x, \partial_y)$ *являются точными для всех* $d_y\in\mathcal{D}$, *причем* $d_y\in\mathcal{D}\Rightarrow M^y{}_zd_y\in\mathcal{D}$. *Тогда преобразование* Qy *будет достаточным для задачи* Υ *при всех* $\varkappa\leqslant 1$, *когда нижние* $\underline{\pi}(x, d)$ *и верхние* $\overline{\pi}(x, d)$ *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: $\underline{\Pi}(\partial)=\underline{M}^{xz}M^y{}_z\underline{\pi}(x, \partial_y)\geqslant\underline{M}^{xz}\underline{\pi}(x, M^y{}_z\partial_y)=\underline{M}^{xz}\underline{\pi}(x, \partial^*{}_z)=\underline{\Pi}(\partial^*)$, поэтому при $\varkappa\leqslant 1$ будет справедливо неравенство: $\Pi^\varkappa(\partial)\geqslant\underline{\Pi}^\varkappa(\partial)$. Наоборот, если эти потери вогнуты, то $\underline{\Pi}(\partial)\leqslant\underline{\Pi}(\partial^*)$ и при $\varkappa\geqslant 1$ имеем $\Pi^\varkappa(\partial)\geqslant\Pi^\varkappa(\partial^*)$.

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. $\underline{\pi}=\overline{\pi}$) 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 ∂, 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 $\inf\limits_{s} f(sy)=f(y)$ (как и sup) всегда инвариантен к группе $\mathscr{S}$, а если инфимум достигается и единствен: $\min\limits_{s} f(sy)=f(s_y(y))$, 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

$$\bar{M}f(x, sy)=\bar{M}f(x, y),\ \forall f\in\mathscr{F}.$$

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 ($\mathscr{X} = \mathscr{R}^k$). 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 $\bar{M}\sum_i (y_i-x)^2=b$. 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 $\int c_t y_t dt$, $\int c_t y^2_t dt$.

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 $P^{1-\varkappa}(x=\hat{x}_y)=\varkappa\underset{\sim}{P}(x=\hat{x}_y)+(1-\varkappa)\bar{P}(x=\hat{x}_y)$ — 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 $\underline{P}(x=\hat{x}_y)=0$, и тем более $\underset{\sim}{P}(x, y)\leqslant\underline{P}(x=\hat{x}_y)=0$, 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

$$\overline{M}\,\mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x}=\inf\Big\{\sum c_i^+\,\bar{M}\,\mathbf{x}^{\mathrm{T}}\mathbf{H}_i\,\mathbf{x}:\ \sum c_i^+\,\mathbf{H}_i-\mathbf{C}\geqslant 0\Big\}, \tag{5.6}$$

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

$$\Pi^{\varkappa}(\mathbf{L})=M^{\varkappa}\,\mathbf{x}^{\mathrm{T}}(\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\,\mathbf{x}+M^{\varkappa}\,\xi^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{L}\,\xi, \tag{5.7}$$

где $M^{\varkappa}$ есть взвешенное среднее, равное $(1-\varkappa)\underline{M}+\varkappa\overline{M}$.

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

$$\mathbf{L}^*=(\mathbf{K}+\mathbf{B})^{-1}\,\mathbf{K}, \qquad \Pi(\mathbf{L}^*)=\mathrm{tr}\,\mathbf{K}\,(\mathbf{K}+\mathbf{B})^{-1}\,\mathbf{B}, \tag{5.8}$$

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 $\mathbf{L}^T\mathbf{L}$. В результате средний риск

$$\Pi^{\varkappa}(\mathbf{L}) = \operatorname{tr}(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^T\mathbf{K} + (1-\varkappa)\lambda^2_{min}\, n\underline{\sigma}^2 + \varkappa\lambda^2_{max}\, n\overline{\sigma}^2.$$

Представим $\mathbf{K}=\mathbf{F}^T\mathbf{\Gamma}\mathbf{F}$, где $\mathbf{F}$ — матрица собственных векторов, а $\mathbf{\Gamma}$ — диагональная матрица собственных чисел γ_i матрицы $\mathbf{K}$. Тогда оптимальное правило запишется в виде $\mathbf{Ly}=\mathbf{F\Lambda}$, где $\mathbf{\Lambda}$ — диагональная матрица корней квадратных λ_i из собственных элементов λ^2_i матрицы $\mathbf{L}^T\mathbf{L}$. В результате средний риск станет

$$\Pi^{\varkappa}(\mathbf{L}) = \sum (1-\lambda_i)^2\gamma_i + (1-\varkappa)\lambda^2_{min}\, n\underline{\sigma}^2 + \varkappa\lambda^2_{max}\, n\overline{\sigma}^2.$$

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 $M\|\mathbf{L}\boldsymbol{\xi}\|^2 = \sup\{\sum c_i\sigma_i^2 : \sum c_i\xi_i^2 + \sum\limits_{i\neq j}\sum \xi_i c_{ij}\xi_i \leqslant$

$$\leqslant \sum\sum\sum \xi_i L_{ij}L_{jk}\xi_k\} = \sup\{\sum c_i\underline{\sigma}_i^2 : c_i \leqslant \sum_i L_{ij}^2\} = \sum\sum L_{il}^2\underline{\sigma}_l^2,$$

и точно так же $\overline{M}\|\mathbf{L}\boldsymbol{\xi}\|^2 = \sum\sum L^2_{ij}\overline{\sigma}^2_i$. В результате риск

$$\Pi^{\varkappa}(\mathbf{L}) = \operatorname{tr}(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^T\mathbf{K} + \operatorname{tr}\mathbf{L}\mathbf{L}^T\mathbf{D}_{\varkappa},$$

где $\sigma^2_{\varkappa i} = (1-\varkappa)\underline{\sigma}^2_i + \varkappa\overline{\sigma}^2_i$ и $\mathbf{D}_{\varkappa}$ 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

$$\Pi^{\varkappa}(\mathbf{L}) = \operatorname{tr}(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^T\mathbf{K} + \operatorname{tr}\mathbf{H}\mathbf{L}\mathbf{L}^T\mathbf{H}^T\mathbf{D}_{\varkappa},$$

и в результате минимизации по $\mathbf{L}$ получим

$$\mathbf{L}^* = (\mathbf{K}+\mathbf{H}\mathbf{D}_{\varkappa}\mathbf{H}^T)^{-1}\mathbf{K}, \qquad \Pi^{\varkappa}(\mathbf{L}^*) = \operatorname{tr}\mathbf{K}(\mathbf{K}+\mathbf{H}\mathbf{D}_{\varkappa}\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{D}_{\varkappa}\mathbf{H}^T.$$

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

$$\Pi^{\varkappa}(\mathbf{L}) = (1-\varkappa)\inf_{\mathbf{K}\in\mathscr{K}}\operatorname{tr}(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^T\mathbf{K} + \varkappa\sup_{\mathbf{K}\in\mathscr{K}}\operatorname{tr}(\mathbf{I}-\mathbf{L})(\mathbf{I}-\mathbf{L})^T\mathbf{K} +$$

$$+(1-\varkappa)\inf_{\mathbf{B}\in\mathscr{B}}\operatorname{tr}\mathbf{L}\mathbf{L}^T\mathbf{B} + \varkappa\sup_{\mathbf{B}\in\mathscr{B}}\operatorname{tr}\mathbf{L}\mathbf{L}^T\mathbf{B}.$$

Рассмотрим случай, когда $\mathcal{K}$ и $\mathcal{B}$ заданы метрическими ограничениями вида

$$\mathcal{K}=\{\mathbf{K}:\|\mathbf{K}-\mathbf{K}_0\|\leqslant\Delta_{\mathbf{K}}\},\ \mathcal{B}=\{\mathbf{B}:\|\mathbf{B}-\mathbf{B}_0\|\leqslant\Delta_{\mathbf{B}}\},$$

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 $\mathcal{K}$ и $\mathcal{B}$ 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 $\mathbf{K}_{\varkappa}=(1-\varkappa)(\mathbf{K}_0-\Delta_{\mathbf{K}}\mathbf{I})^+ +\varkappa(K_0+\Delta_{\mathbf{K}}\mathbf{I})$, 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.

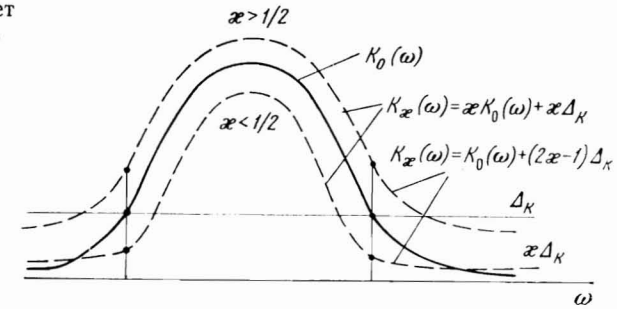
При $\mathbf{K}_0-\Delta_{\mathbf{K}}\mathbf{I}>0$ и $\mathbf{B}_0-\Delta_{\mathbf{B}}\mathbf{I}>0$ (что соответствует положительной определенности матриц) имеем

$$\mathbf{K}_{\varkappa}=\mathbf{K}_0+(2\varkappa-1)\Delta_{\mathbf{K}}\mathbf{I},\qquad \mathbf{B}_{\varkappa}=\mathbf{B}_0+(2\varkappa-1)\Delta_{\mathbf{B}}\mathbf{I}.$$

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.