Глава 8.

НАДЕЖНОСТНЫЙ СИНТЕЗ

8.1. ОБЩИЕ ВОПРОСЫ СИНТЕЗА МОДЕЛЕЙ

Методология синтеза моделей. Наше житие можно сравнить с движением в купе скорого поезда, из окна которого проносятся мимо быстро сменяющие друг друга пейзажи. И нет ни малейшего времени задуматься, что каждый его фрагмент «дышит» собственной очень сложной и содержательной жизнью, детальное изучение закономерностей которой — удел многих и многих поколений (если не всех). Мы же, глядя в окно, ограничиваемся самым поверхностным представлением обо всем этом, внешней моделью, связывающей увиденное с нашими внутренними воззрениями, опытом. Это естественно, потому что здравый опыт в своих притязаниях и проявлениях суть экономное представление окружающего: выделение главного, игнорирование всего второстепенного (хотя детям, красивым девушкам и некоторым ученым иногда свойственна прямо противоположная тенденция).

Образный пример не является случайным, а объясняет характер непреходящей гносеологической связи:

$$\text{ЯВЛЕНИЕ} \quad \underset{\text{действие}}{\overset{\text{адекватность}}{\rightleftarrows}} \quad \text{МОДЕЛЬ}$$

Человек строит модели как отражения реальностей (причем на самой разной основе) и использует их для познания природы, а также воздействия на нее (поэтому-то отношение стрелками указано в ту и другую сторону). Как и природа, модель живет своей самостоятельной жизнью. Обе жизни родственны в том смысле, что они должны быть слаженными, как говорят, модель должна быть адекватной явлению (верхняя стрелка), что и позволит по состояниям модели прогнозировать, а затем и управлять состояниями явления (нижняя стрелка).

Прекрасной иллюстрацией сказанному является случайный процесс броуновского движения — классическая модель перемещения частицы при хаотических столкновениях ее с молекулами при тепловом движении. Наглядная физическая картина здесь породила математический образ, вероятностную модель, в которой как таковых частичек уже нет, про них забыли, а только отжатый результат — процесс перемещения в виде математического построения. Такая абстрактизация позволила найти вероятности уклонений, установить идеальные законы броуновского движения.

Вообще, нужно быть осторожным, так как на языке моделей удобным и отразимым оказывается далеко не все, что существует в природе. Идеализация приводит к моделям, надежным только на первый взгляд, отражающим желаемую для модели картину эксперимента. В дальнейшем, удобно в силу специфики настоящего изложения говорить только о математико-вероятностных моделях — символьного языка описания случайных явлений.

Глубина и безграничная сложность явлений реального мира вынуждает, казалось бы, такие же качества у моделей. Так и кажется подчас, что чем сложнее модель, богаче ее собственная жизнь, тем более сильной в своей отражательной потенции она является. Можно и согласиться, если бы при этом не разрушалась прямая связь модели с явлением. Усложнение модели требует настоятельной проверки каждого ее нового фрагмента на соответствие действительности. Словами «пусть» (столь привычными в современных математико-вероятностных изложениях: пусть процесс марковский, пусть плотность существует и дважды дифференцируема, пусть известна вероятность и пр. и пр.) достигается обособление жизни модели, превращается в самоцель ее изучение математиками, считающими себя совершенно чистыми.

Можно возразить, что история знает примеры, такие как теория групп или неэвклидовая геометрия, когда то или иное сугубо математическое построение встречалось в конечном счете с практическими приложениями. Но рассчитывать каждый раз на случай — все равно, что делать ставку при ликвидации космического объекта на удар в него метеорита. Не надежнее ли нацелить в объект что-то управляемое во времени в трехкоординатном пространстве?

Мы убеждаемся здесь еще раз в нашей главной мысли, породившей и пронизывающей все содержание книги, что не нужны

оторванные суперсложные модели, нюансы которых кто-нибудь из добросовестных исследователей может воспринять всерьез как изведанные законы природы. Значительно экономнее и надежнее иметь арсенал простых моделей, связывающихся с явлением в небольшом числе сторон, своего рода каналов, и отражать явление не сразу все, а по частям, освещая каждый раз только ту сторону, которая представляет непосредственный интерес. В том-то сила и искусство познания: расчленить всю сферу деятельности на науки, каждой из которых отдается своя часть физического явления, затем науку — на предметные области и т. д.

Конкретизируем теперь связь: МАТЕМАТИЧЕСКАЯ ТЕОРИЯ → МОДЕЛЬ. Математическая теория развивает групповые законы моделей и организуется внутри себя системой аксиом — этой формальной конструкцией, сцепляющей каждую модель. На символьном материале аксиоматические связи должны повторять реальные, что и есть адекватность аксиом и гарант правомочности теории. А жизненность теории будет зависеть от того, какими сторонами ее представители (модели) связываются с явлениями, насколько эти связи легко наводятся, доступны, физически наглядны (интерпретируемы), привычны, наконец надежны. Это и поможет инженеру-исследователю выбрать ту конкретную модель, которая нужна для решения поставленной задачи, чтоб далее привлечь всю мощь теории, от упрощения моделей только выгадывающей.

Постановка задачи. Наша цель — рассмотреть проблему синтеза интервальных статистических моделей средних. Связующими для них с реальными явлениями будут некоторые *задающие параметры* $\theta = (\theta_1, \theta_2, \ldots)$, которые по их числу, содержанию и физическому смыслу могут быть самыми разнообразными (могут иметь только математический, формальный смысл). Теперь если модель искать в рамках исходной структуры $\mathscr{M}_\theta$ и найти параметр θ в виде детерминированной оценки $\hat{\theta}$, то получим модель $\mathscr{M}_{\hat{\theta}}$, а если в виде индикаторной $\hat{\Theta}$ (для одномерного θ — интервальной), то объединение $\bigvee_{\theta \in \hat{\Theta}} \mathscr{M}_\theta$.

Оценивание θ должно производиться из ясного осознания конечных целей, под которыми подразумевается та последующая, наследственная задача, на решение которой призывается модель. Это может быть задача либо анализа, либо построения решающих правил. Например, модель шума привлекается для нахождения алгоритма оптимального обнаружения сигналов или оценивания параметров в аддитивном представлении. Возможен другой вариант, когда модели гипотезы и альтернативы не связаны между собой через шум и строятся отдельно одна от другой.

Нашу мысль, что решающие устройства должны через себя влиять на сам синтез модели в форме требований к оценкам задающих параметров, проиллюстрируем на примере.

Пример 8.1. Пусть θ — задающий (одномерный или многомерный) параметр; для каждого его значения вводится модель $\mathcal{M}_\theta$ и по ней синтезируется решающее правило ∂_θ (возможно, оптимальное). Вид этого правила определяется видом $\mathcal{M}_\theta$, следовательно, будет зависеть от θ. Но θ не известен. Считаем, что по обучающим реализациям $\mathbf{Z}$ он оценивается $\hat{\theta}=\hat{\theta}(\mathbf{Z})$ и подставляется в правило, что приводит к $\partial_{\hat{\theta}}$. Риск этого правила будет зависеть от истинного, но неизвестного θ и обозначается $\Pi_\theta(\partial_{\hat{\theta}})$. Теперь стоит вопрос, как оценивать $\hat{\theta}$? В среднем с учетом случайного разброса $\hat{\theta}$ риск всей указанной адаптивной процедуры (в общем, квазиоптимальной) равен $r(\partial, \hat{\theta})=M^{\times}\Pi_\theta(\partial_{\hat{\theta}})$, где усреднение производится по совместной модели θ и $\mathbf{Z}$. Оценку $\hat{\theta}$ нужно выбирать так, чтобы минимизировать $r(\partial, \hat{\theta})$, откуда сразу видно, что $\Pi_\theta(\partial_{\hat{\theta}})$ и должна служить функцией потерь в задаче оценивания (синтеза) θ [1].

Выбор начальной структуры $\mathcal{M}_\theta$ — слабое место синтеза, способное в корне свести на нет наше стремление получить надежную модель. Так будет, если $\mathcal{M}_\theta$ описывается параметризованной точной плотностью вероятностей $p_\theta(x)$, при изменении θ очерчивающей лишь неосязаемую линию в пространстве всех распределений вероятностей. Надежные же модели должны быть объемными.

Наибольший интерес для нас представляет тот случай, когда $\mathcal{M}_\theta$ не надо выбирать, а оно само естественно определяется наведенными с явлением связями. А этот как раз тот самый случай, когда задающими параметрами являются статистические средние $\theta=MQ$ набора Q признаков, составляющие в совокупности модель $\mathcal{M}=\bigvee\langle MQ\rangle$ (где объединение обязано расплывчатости оценок средних). Нагрузка синтеза полностью перекладывается на выбор набора Q и оценивание соответствующих ему параметров $\theta_i=Mq_i$, $q_i\in Q$. Не надо, что самое замечательное, ничего лишнего, никаких допущений, порождающих сомнения.

Здесь возникают две проблемы: выбор задающих признаков и оценивание их средних. Первая составляет свою сферу деятельности, и не всегда научную, а учитывая подавляющее разнообразие самих признаков и возможностей их выбора, превращающуюся подчас в искусство (начинается там, где заканчивается наука). Это инженерное искусство выбора связующих признаков, используя априорные сведения о явлении, какие есть знания его механизма, наблюдения за явлением, опираясь на практику, опыт и здравый смысл, сообразуясь с трудоемкостью самой процедуры синтеза модели, не «спуская прицел» с последующего применения модели. Путеводными здесь являются качества модели, пронизывающие настрой всей книги: простота, доступность, надежность.

[1] Кузнецов В. П. Байесов подход и оптимизация процесса обучения//Автоматика и телемеханика. — 1971. — № 4. — С. 66—71.

Вторая проблема состоит в оценивании задающих параметров. Она согласована с выбором задающих признаков и поэтому не исключает привлечения самых разнородных физических закономерностей, фактов. Но может и формально решаться на базе предварительного эксперимента, обучающих реализаций, что нас интересует в наибольшей мере. Здесь требование адекватности связи модели с явлением накладывает на обучающие реализации обязанность быть «полномочными представителями» интересующего нас явления, т. е. быть носителями одинаковых значений параметров (средних), тогда их можно оценивать. Это есть условие стационарности задающих параметров, которое сейчас обсудим.

Стационаризация статистических параметров. С позиций математических моделей средние как и вероятности есть фиксированные числа, тогда как интервальные средние и вероятности есть интервалы, границы которых удовлетворяют аксиомам ИМ. В эти понятия заложен смысл среднего арифметического или же частоты.

Сейчас не суть важно, каким является пространство $\mathcal{Z}$ элементарных исходов, на котором строится математическая модель; это может быть произведение $\mathcal{X}\times\mathcal{Y}$, либо пространство Ξ значений флуктуаций ξ. Пусть $z_1, z_2, \ldots, z_N$, есть реализация независимой последовательности *обучающих испытаний*, для которой параметр $Mq(z_i)$ является стационарным, т. е. одним и тем же для всех i. Такие испытания называются Mq-стационарными.

Для стационарного параметра точное среднее Mq достигается как предел среднего арифметического:

$$Mq=\lim_{N\to\infty}\frac{1}{N}\sum_{1}^{N}q(z_i)$$

в серии независимых стационарных испытаний z_i, $i=1, 2, \ldots$ Точная вероятность по тому же смыслу есть предел относительной частоты события

$$P(A)=\lim_{N\to\infty}\frac{1}{N}\sum_{1}^{N}\{z_i\in A\},$$

где фигурные скобки обозначают индикаторы событий, равные 1 при $z_i\in A$, а иначе — 0.

Интервальные средние и вероятности возникают тогда, когда точных их значений нет, либо по причине ограниченности числа испытаний (дефицит опыта), либо из-за возможной нестационарности испытаний, т. е. неустойчивости средних и частот. Последнее является неприятной «подножкой», но к счастью, не всему статистическому подходу, а тем формальным методам синтеза моделей, к которым мы стремимся и которые волей-неволей требуют стационарности параметров. Укажем типичный для статистических приложений способ стационаризации, основанный на случайном выборе.

Пусть имеется совокупность в N независимых испытаний z_i, $i=1, \ldots, N$, причем средние $M_i q(z_i)$ в разных испытаниях, в общем, различны. «Перемешаем» теперь испытания слепым образом, а иначе говоря, произведем их последовательный равновозможный выбор. Тогда средние стабилизируются, станут одними и теми же, равными $Mq=\sum M_i q(z_i)/N$, так как равновозможно с вероятностями $1/N$ выбранным может оказаться любой индекс i. Как видно, произошла стационаризация испытаний, но увы, с возможной потерей независимости, хотя и из этого упущения есть свой выход, который проиллюстрируем на примере.

Пример 8.2. Пусть p_i есть вероятность события $z_i \in A$ в i-м испытании (A одно и то же) и пусть заведомо известно, что первые k раз A обязательно произойдет: $p_1=p_2=\ldots=p_k=1$, а остальные — нет: $p_{k+1}=p_{k+2}=\ldots=p_N=0$. Тогда случайный выбор (слепое перемешивание) испытаний ведет к вероятности $p=k/N$ события A в каждом из них. Но после перемешивания события A, увы, не обретут желаемой независимости, так как если стало известно, что первые k раз повторилось событие $z_i \in A$, то всеми последующими непременно будут противоположные события $z_i \in A^c$.

Если же после перемешивания из совокупности объема N произвести редкий выбор $z_1, \ldots, z_n$ небольшого их числа (объема) $n \ll N$, то эту выборку с достаточно большой точностью можно считать стационарной независимой. В самом деле, $\overline{M}_i q(z_i)$ от произвольного признака q будет одним и тем же $\overline{M}q$ и справедливо равенство $\overline{M}\Pi q_i(z_i) = \max\limits_{M^*=\underline{M},\, \overline{M}} \Pi M^* q_i(z_i)$.

Вывод такой: *случайный редкий выбор из совокупности есть средство стационаризации последовательности и причина независимости ее элементов.*

Стационаризация облегчает построение математической модели, так как сводит ее синтез к оцениванию одинаковых по течению испытаний задающих параметров, т. е. к своей новой статистической задаче, которую рассмотренными в предыдущих главах методами можно формализовать и решить. Для этого нужно знать, какого содержания оценки, задающие модель, хочется получить, каким основным показателям они должны удовлетворять.

Понятие доверительной модели. Проанализируем содержимое рис. 8.1. Построение математической модели — это отдельная задача оценивания параметров, а точнее сказать, *надзадача*, пищей которой служат испытания, и как всякая статистическая задача, она нуждается в исходной конструкции. В конструкцию надзадачи в первую очередь входит изначальная математическая модель самой последовательности испытаний, так сказать, *надмодель*, отличающаяся от искомой своей шириной и крайне непритязательными запросами: для нас это будут только предположения о независимости и стационарности. Интересно далее будет наблюдать за сужением ее в искомую модель, что достаточно сделать в направлении $\mathcal{Q}$ оцениваемых (задающих) параметров.

Детерминированные (точечные) задающие модель оценки $\hat{M}q$, $q \in \mathcal{Q}$, в надзадаче ведут к $\mathcal{Q}$-простым моделям $\langle M\mathcal{Q}\rangle$. Эти модели

29

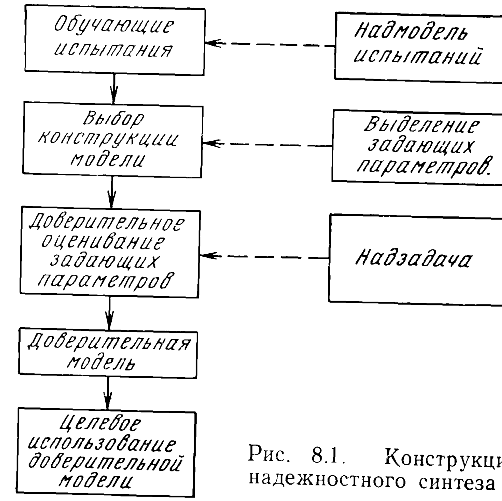

Рис. 8.1. Конструкция надежностного синтеза

с точными значениями задающих средних $Mq=\hat{M}q$, $q \in \mathcal{Q}$, будут ненадежными по причине ненадежности детерминированных оценок (кстати, включение детерминированных оценок в модель присуще, как это следует из теоремы 5.1, режиму оптимизма $\varkappa=1/2$). Нас сейчас всего более привлекает режим пессимизма и надежные модели, а для этого задающие оценки также должны быть надежными, доверительными, что ведет к следующему понятию.

Доверительной надежности ρ *называется модель, определенная пессимистичными* ($\varkappa=1$) *совместными доверительными оценками* ($\mathcal{Q}$) *уровня* $\alpha=1-\rho$ *набора* $M\mathcal{Q}=\{Mq,\ q \in \mathcal{Q}\}$ *задающих средних* (найденными по обучающей последовательности $\mathbf{z}=(z_1, \ldots, z_N)$ испытаний). Здесь нужно обратить внимание, что надежность доверительных оценок равна 1 минус верхняя граница $\overline{\alpha}(\mu)=\alpha$ вероятности ошибки, что соответствует пессимистичному оцениванию. Это принципиальный момент, что оптимизм, если и допустим при нахождении решающих правил, но ни в коем случае не при синтезе надежных моделей. Надежные модели пессимистичны.

Итак, доверительная модель равносильна расплывчатой оценке задающих модель параметров: средних значений признаков. Итогом станет интервальная модель средних $\mathcal{M}$, но только в том случае, если оценки $\mu_{\mathbf{z}}(\mathcal{Q})$ индикаторные или совместные интервальные, т. е. $\mu^{\mathbf{z}}(\mathcal{Q})$ принимает значения либо 0, либо 1. Причем значение $\mu_{\mathbf{z}}(\mathcal{Q})=1$ и выделяет как раз область Θ тех $\theta=M\mathcal{Q}$, которые включаются составляющими синтезируемой модели $\mathcal{M}$, записываемой как их объединение: $\mathcal{M}=\bigvee\limits_{M\mathcal{Q}\in\Theta} \langle M\mathcal{Q}\rangle$. Представляется ИМ $\mathcal{M}$ как облако составляющих ее простых моделей $\langle M\mathcal{Q}\rangle$ одной концентрации в смысле одинакового доверия ко всем им внутри Θ (и нулевого доверия вне Θ).

Заметим, представив $\mathcal{M}$ через первичные средние: $\mathcal{M}=\langle \tilde{M}\mathcal{G}\rangle$, что совпадение первичного набора $\mathcal{G}$ доверительной модели с задающими признаками $\mathcal{Q}$ будет лишь в случае, когда область Θ прямоугольная, т. е. направления ее граней совпадают с направлениями q. Например, $\mu_{\mathbf{z}}(\mathcal{Q})$ есть совместная интервальная оценка $\underset{\sim}{M}q$, $\tilde{M}q$, $q \in \mathcal{Q}$, тогда $\Theta=\{Mq : \underset{\sim}{M}q \leqslant Mq \leqslant \tilde{M}q,\ q \in \mathcal{Q}\}$. Для других индикаторных оценок $\mu_{\mathbf{z}}(\mathcal{Q})$ наборы $\mathcal{G}$ и $\mathcal{Q}$, в общем, различаются между собой: $\mathcal{G} \neq \mathcal{Q}$.

11*

30

8.2. ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОЙ МОДЕЛИ НА ЗАДАННОМ НАБОРЕ СОБЫТИЙ

Исходные положения. Здесь рассматривается тот случай, когда задающими параметрами являются вероятности набора событий, поначалу непересекающихся, что применимо к тем задачам, в которых из вероятностей и слагаются характерные черты интересующего нас исследуемого явления. Такой выбор может руководствоваться простотой и удобством оценивания вероятностей. При этом четко нужно осознавать, что все первичные признаки итоговой доверительной модели будут обязательно постоянными на задающих событиях или их пересечениях и это же свойство перейдет дальше к решающим правилам (достаточного класса) как конечной цели построения модели.

Сформулируем формально задачу синтеза модели (надзадачу). Пусть $\mathbf{z}=(z_1, \dots, z_N)$ есть обучающие испытания из пространства $\mathscr{Z}$: $z_j\in\mathscr{Z}$ (скалярного или векторного) и пусть $A_1, \dots, A_k$ — непересекающиеся события, такие, что $\sum A_j\subset\mathscr{Z}$, и $A_{k+1}=\mathscr{Z}-\sum A_j\neq\varnothing$ — остаточное событие, введенное для общности. Вероятности $P(z_i\in A_j)=p_j$ считаются стационарными, не зависящими от номера i наблюдений, а события $z_i\in A_j$ и $z_{i'}\in A_{j'}$ при $i\neq i'$ нековариированными. Эти условия будут выполнены, если испытания являются независимыми стационарными.

Здесь задающим модель будет вектор $\mathbf{p}=(p_1, \dots, p_k)$ вероятностей. Его и нужно оценить при заданной надежности ρ. Если обозначить доверительную оценку $\mu_{\mathbf z}(\mathbf p)$, то ее надежность равна $\rho=\underline{M}\mu_{\mathbf z}(\mathbf p)$.

При указанных условиях достаточными для оценивания являются векторы частот $\mathbf r=(r_1, \dots, r_k)$, где r_j — число элементов z_i, попавших в $A_j: r_j=\sum_{i=1}^{N}\{z_i\in A_j\}$. Оценка $\mu_z(\mathbf p)=\mu_{\mathbf r}(\mathbf p)$ должна быть функцией вектора $\mathbf r$.

При каждом заданном векторе $\mathbf p$ вероятности частот $r_1, \dots, r_k$ даются известной мультиноминальной формулой

$$P_{\mathbf p}(\mathbf r)=p_1^{r_1}\dots p_{k+1}^{r_{k+1}}\,N!/(r_1!\dots r_{k+1}!),$$

где считается $0^0=1$, $p_{k+1}=1-\sum^{k}p_j$, $r_{k+1}=N-\sum_{1}^{k}r_j$. Формула для $P_{\mathbf p}(\mathbf r)$ определяет переходную надмодель $\mathscr{M}^{\mathbf r}_{\mathbf p}$, редуцированную от $\mathbf z$ к $\mathbf r$. Общей надмоделью испытаний будет произведение $\mathscr{M}^{\mathbf{pr}}=\mathscr{M}^{\mathbf p}\mathscr{M}^{\mathbf r}_{\mathbf p}$, куда в $\mathscr{M}^{\mathbf p}$ можно вложить априорные сведения о $\mathbf p$, если они есть. Их обычно нет, тогда $\mathscr{M}^{\mathbf p}=\mathscr{J}^{\mathbf p}$ и надежность доверительной оценки $\mu_{\mathbf z}$ $(\mathbf p)$ вектора $\mathbf p$ равна

$$\rho=\min_{\mathbf p}\sum_{\forall\mathbf r}\mu_{\mathbf r}(\mathbf p)\,P_{\mathbf p}(\mathbf r).$$

Шкала расплывчатости оценки вектора $\mathbf p$ во многом обязана желаемым свойствам получаемого по доверительной модели решающего правила. Если же не определять пока решаемую на базе модели статистическую задачу, то разумно использовать интегральную шкалу как наиболее простую:

$$\Omega(\mu)=\int_{\mathrm I}\dots\int\mu_{\mathbf r}(\mathbf p)\,d\mathbf p,\quad \mathbf I=\left\{\mathbf p:\ \sum_{1}^{k+1}p_j=1,\ p_j\geqslant 0\right\}.$$

Частная $\mathscr{M}^{\mathbf r}$, дающая надмодель вектора частот $\mathbf r$, определяется границами:

$$\overline{M}f(\mathbf r)=\max_{\mathbf p}\sum_{\forall\mathbf r}f(\mathbf r)\,P_{\mathbf p}(\mathbf r),$$

поэтому составной риск запишется: $\Pi_\lambda(\mu)=1-\underline{M}\mu_{\mathbf r}(\mathbf p)+\lambda\overline{M}\Omega\times$

$$\times(\mu)=1-\min_{\mathbf p}\sum_{\forall\mathbf r}\mu_{\mathbf r}(\mathbf p)\,P_{\mathbf p}(\mathbf r)+\lambda\max_{\mathbf p^*}\sum_{\forall\mathbf r}P_{\mathbf p^*}(\mathbf r)\int_{\mathrm I}\dots\int\mu_{\mathbf r}(\mathbf p)\,d\mathbf p$$

где λ — весовой коэффициент. Нужно минимизировать риск выбором $\mu_{\mathbf r}(\mathbf p)$.

Перепишем риск следующим образом: $\Pi_\lambda(\mu)=1-\inf_{w(\mathbf p)}\sum_{\forall\mathbf r}\int_{\mathrm I}\dots\int\mu_{\mathbf r}\times$

$$\times(\mathbf p)\,P_{\mathbf p}(\mathbf r)\,w(\mathbf p)\,d\mathbf p+\lambda\sup_{w^*(\mathbf p)}\sum_{\forall\mathbf r}\int_{\mathrm I}\dots\int P_{\mathbf p}(\mathbf r)\,w^*(\mathbf p)\,d\mathbf p\int_{\mathrm I}\dots\int\mu_{\mathbf r}(\mathbf p)\,d\mathbf p.$$

Тогда цена надзадачи примет вид

$$v=1-\sup_{\mu}\inf_{w,w^*}\sum_{\forall\mathbf r}\int_{\mathrm I}\dots\int\mu_{\mathbf r}(\mathbf p)\,[P_{\mathbf p}(\mathbf r)\,w(\mathbf p)-\lambda P_{w^*}(\mathbf r)]\,d\mathbf p,$$

где инфимум ищется по всевозможным априорным плотностям $w(\mathbf p)$ и $w^*(\mathbf p)$, а $P_{w^*}(\mathbf r)=\int_{\mathrm I}\dots\int P_{\mathbf p}(\mathbf r)\,w^*(\mathbf p)\,d\mathbf p$ — вероятность вектора $\mathbf r$ при плотности w^*.

В соответствии с общими принципами минимакса [22] супремум и инфимум в записи цены поменяем местами. При каждых заданных $w(\mathbf p)$ и $w^*(\mathbf p)$ оптимальная оценка $\mu_{\mathbf r}(\mathbf p)$ вектора $\mathbf p$ будет индикаторной, принимающей значение 1 при

$$P_{\mathbf p}(\mathbf r)\,w(\mathbf p)/P_{w^*}(\mathbf r)\geqslant\lambda \qquad (8.1)$$

и 0 в противном случае. С учетом найденной оценки цена запишется:

$$v=1-\inf_{w,w^*}\sum_{\forall\mathbf r}\int_{\mathrm I}\dots\int[P_{\mathbf p}(\mathbf r)\,w(\mathbf p)-\lambda P_{w^*}(\mathbf r)]^{+}\,d\mathbf p.$$

Отсюда нужно искать наименее благоприятные $w(\mathbf p)$ и $w^*(\mathbf p)$, которые затем подставляются в (8.1), что и приведет к искомой оптимальной оценке $\mathbf p$. Порог λ здесь должен быть выбран исходя из заданной надежности

$$\rho=\sum_{\forall\mathbf r}\int_{\mathrm I}\dots\int\mu_{\mathbf r}(\mathbf p)\,P_{\mathbf p}(\mathbf r)\,w(\mathbf p)\,d\mathbf p, \qquad (8.2)$$

куда подставляется наименее благоприятная плотность $w(\mathbf p)$.

Трудности нахождения наименее благоприятных плотностей вынуждают прибегать к некоторым «разумным» вариантам их подбора.

Модель наибольшего правдоподобия. Пусть априорное распределение вектора $\mathbf{p}$ равномерно $w(\mathbf{p})=\mathrm{const}$ и пусть $P_{w^*}(\mathbf{r})=\mathrm{const}$, т. е. априори никаким векторам частот предпочтения не отдается, что в определенном смысле отражает неблагоприятную ситуацию. Тогда формула (8.1) запишется $P_{\mathbf{p}}(\mathbf{r})\geqslant\lambda$, т. е. оценкой является индикаторная функция семейства векторов $\mathbf{p}$ максимальной вероятности (наибольшего правдоподобия). После логарифмирования обеих частей последнего неравенства, перемены знака и объединения между собой постоянных слагаемых оценка примет вид

$$F(\hat{\mathbf{p}},\mathbf{p})=-\sum_{j=1}^{k+1}\hat{p}_j\ln p_j\leqslant\lambda_1,\ \lambda_1 N=-\ln\lambda+\ln(N!)-\sum_{1}^{k+1}\ln(r_j!),\tag{8.3}$$

где $\hat{p}_j=r_j/N$ — есть относительные частоты выпадения событий A_j.

Отметим, что функция $F(\hat{\mathbf{p}},\mathbf{p})$ неотрицательна и принимает минимальное значение при $\mathbf{p}=\hat{\mathbf{p}}$, равное: $F(\hat{\mathbf{p}},\hat{\mathbf{p}})=-\sum_{1}^{k+1}\hat{p}_j\ln\hat{p}_j$. Пороговое значение λ (а отсюда и λ_1) находится по (8.2) с подстановкой $w(\mathbf{p})=\mathrm{const}$. Интеграл при этом получается трудоемким, поэтому есть смысл определять λ исходя из нижней границы надежности:

$$\rho=\min_{\mathbf{p}}\sum_{\mathbf{r}:\,P_{\mathbf{p}}(\mathbf{r})\geqslant\lambda}P_{\mathbf{p}}(\mathbf{r}).\tag{8.4}$$

Неравенство (8.3) с нахождением λ_1 с помощью (8.4) формирует семейство $\mathscr{M}$ векторов $\mathbf{p}$, составляющих доверительную модель надежности ρ. Эта $\mathscr{M}$ иначе задается своими первичными средними, к поиску которых и приступаем. Здесь (так как семейство $\mathscr{M}$ непрямоугольное) первичными признаками не будут сами задающие события A_j (вероятности $\mathbf{p}$ которых оцениваются), а будут измеримые функции на них.

Обозначим $\mathbf{p}^\circ=(p^\circ_1,\ \dots,\ p^\circ_k)$, $p^\circ_{k+1}=1-\sum_1^k p^\circ_j$ — решения уравнения $F(\hat{\mathbf{p}},\mathbf{p})=\lambda_1$ относительно $\mathbf{p}$ при выборе λ_1 по заданной надежности ρ. Эти решения образуют набор, задающий поверхность семейства $\mathscr{M}$. Первичные средние определяются гиперплоскостями в подмножестве $\mathbf{I}$ пространства $\mathscr{R}^{k+1}$, касающимися семейства $\mathscr{M}$ в точках $\mathbf{p}^\circ$. Так как $\partial F(\mathbf{r},\mathbf{p})/\partial p_j=-r_j/p_j$, то уравнения этих гиперплоскостей имеют вид $\sum_1^{k+1}\hat{p}_j(p_j-p^\circ_j)/p^\circ_j=0$. Эти гиперплоскости и определяют первичные признаки вида

$$g_{\mathbf{p}^\circ}(\mathbf{z})=\sum_{j=1}^{k+1}\hat{p}_j A_j(\mathbf{z})/p_j^\circ,\ \ \hat{p}_j=r_j/N,$$

и соответствующие им средние

$$\overline{M}g_{\mathbf{p}^\circ}=\max_{\mathbf{p}\in\mathscr{M}}\sum_1^{k+1}\hat{p}_j p_j/p_j^\circ=\sum_1^{k+1}\hat{p}_j p_j^\circ/p_j^\circ=1.$$

В результате различным $\mathbf{p}^\circ$ как решениям уравнения $\hat{F}(\mathbf{p},\mathbf{p})=\lambda$ соответствуют различные первичные признаки $g_{\mathbf{p}^\circ}(\mathbf{z})$ с одним и тем же у всех равным 1 верхним средним.

Таким образом, *первичными средними, определяющими доверительную модель, будут*

$$\overline{M}\sum_1^{k+1}\hat{p}_j A_j(z)/p_j^\circ=1,\ \forall\,\mathbf{p}^\circ:\ -\sum_1^{k+1}\hat{p}_j\ln p_j^\circ=\lambda_1.\tag{8.5}$$

Мы видим, что $\mathscr{M}$ зависит от полученных в испытаниях частот r_j (или относительных частот $\hat{p}_j$). При $N\to\infty$ модель $\mathscr{M}$ стягивается к точному распределению вероятностей, соответствующему предельным частотам $p_j=\lim\limits_{N\to\infty}\hat{p}_j$.

Получим в явном виде некоторые из первичных средних, соответствующих набору (8.5). Для этого фиксируем $p^\circ_j=\hat{p}_j$, $j=2,\ \dots,\ k$, и будем искать p°_1 из равенства $F(\hat{\mathbf{p}},\mathbf{p})=\lambda_1$. Получим p°_1 как решение уравнения

$$-\hat{p}_1\ln p_1-\hat{p}_{k+1}\ln(\hat{p}_{k+1}+\hat{p}_1-p_1)=\lambda_1+\sum_2^k\hat{p}_j\ln\hat{p}_j.$$

Этих решений будет два: $\underset{\smile}{p}{}^\circ_1$, $\tilde{p}^\circ_1$, где $\underset{\smile}{p}{}^\circ_1\leqslant\hat{p}_1\leqslant\tilde{p}^\circ_1$. Таким образом, искомый вектор $\mathbf{p}^\circ_k$ вероятностей будет иметь вид $p^\circ_1,\hat{p}_2,\dots,\hat{p}_k$, $(1-\sum_2^k\hat{p}_j-p^\circ_1)$, где p°_1 равно либо $\underset{\smile}{p}{}^\circ_1$, либо $\tilde{p}^\circ_1$. Соответствующее каждому такому вектору первичное среднее находится из уравнения (8.5) и имеет вид

$$\overline{M}\,(\hat{p}_1-p_1^\circ)\,[A_1(\mathbf{z})/p_1^\circ-A_{k+1}(\mathbf{z})/(\hat{p}_{k+1}+\hat{p}_1-p_1^\circ)]=0.$$

Заменяя индексы 1 и $k+1$ разными комбинациями индексов l, m, $l\neq m$, от 1 до $k+1$, получим поднабор первичных средних вместе с уравнениями для нахождения $\underset{\smile}{p}{}^\circ_l$ и $\tilde{p}^\circ_l$.

Если склоняться к упрощениям, то от всего набора (8.5) первичных средних можно отказаться, оставив их какую-то часть, что приводит к расширению доверительной модели с увеличением надежности. За основу расширения могут быть взяты любые признаки вида $g(\mathbf{z})=\sum_1^{k+1}g_iA_i(z)$. Интересно то, что каждый из них с точностью до множителя совпадает с одним из первичных признаков (8.5) (для этого нужно подобрать соответствующее $\mathbf{p}^\circ$), поэтому такое расширение эквивалентно уменьшению числа первичных средних.

Использование критерия хи-квадрат. Определяющее доверительную модель $\mathscr{M}$ семейство векторов $\mathbf{p}$ может быть выбрано с позиций упрощенного расчета порога λ. Используем для этого статистику хи-квадрат

$$\sum_{1}^{k+1} (p_j-\hat{p}_j)^2/p_j \leqslant \lambda_2, \tag{8.6}$$

где λ_2 находится по заданной надежности ρ. При больших N левая часть неравенства имеет приближенно распределение хи-квадрат с k степенями свободы [25]. Тогда порог λ_2 будет критической точкой этого распределения. Семейству $\mathscr{M}$ векторов $\mathbf{p}$, определенному неравенством (8.6), соответствуют первичные значения

$$\overline{M} \sum_{1}^{k+1} \hat{p}_j^2 A_j(\mathbf{z})/(p_j^0)^2 = \sum_{1}^{k+1} \hat{p}_j^2/p_j^0 ,$$

где $\mathbf{p}^\circ$ есть всевозможные решения уравнения (8.6), в котором неравенство заменено на равенство.

Информационный критерий построения доверительной модели. Будем считать в (8.1) $w(\mathbf{p})=\mathrm{const}$ и остановимся на подборе плотности $w^*(\mathbf{p})$, входящей в знаменатель. Будем искать максимум знаменателя по $w^*(\mathbf{p})$ при каждом заданном векторе частот $\mathbf{r}$. Этот максимум достигается при дельта-функции Дирака $w^*(\mathbf{p})=\delta(\hat{\mathbf{p}}-\mathbf{p})$ и определяется по формуле

$$P(\mathbf{r}) = \max_{w^*} P_{w^*}(\mathbf{r}) = \max_{\mathbf{p}} P_{\mathbf{p}}(\mathbf{r}) = P_{\mathbf{r}/N}(\mathbf{r}).$$

Оправданием нашим действиям служит то, что имея наблюденным вектор $\mathbf{r}$, мы рассматриваем наименее благоприятную плотность w^* применительно к $\mathbf{r}$.

С учетом найденного $P(\mathbf{r})$ после подстановки его в (8.1), логарифмирования обеих частей неравенства и перемены знака приходим к семейству $\mathscr{M}$ векторов $\mathbf{p}$, определяемому неравенством

$$J(\hat{\mathbf{p}}, \mathbf{p}) = \sum_{1}^{k+1} \hat{p}_j \ln(\hat{p}_j/p_j) \leqslant \lambda_3. \tag{8.7}$$

Левая часть неравенства (8.7) есть различающая информация [24], содержащаяся в векторе $\hat{\mathbf{p}}$ в пользу точной вероятностной модели, определяемой этим вектором, при конкурирующей альтернативе $\mathbf{p}$. Ясно, что чем меньше эта различающая информация, тем ближе $\mathbf{p}$ к $\hat{\mathbf{p}}$, а при $\mathbf{p}=\hat{\mathbf{p}}$ эта информация минимальна и равна 0. Таким образом, согласно (8.7) доверительную модель образуют такие векторы $\mathbf{p}$, которые в смысле различающей информации отстоят от $\hat{\mathbf{p}}$ не более, чем на число λ_3. Порог λ_3 находится из формулы (8.4) с подстановкой неравенства (8.7) под знак символа суммы как ограничения на $\mathbf{r}$ (вспомним, что $\hat{\mathbf{p}}=\mathbf{r}/N$).

В сравнении с (8.3) имеем

$$J(\hat{\mathbf{p}}, \mathbf{p}) = \sum_{1}^{k+1} \hat{p}_j \ln \hat{p}_j + F(\hat{\mathbf{p}}, \mathbf{p}),$$

поэтому (8.7) совпадает с (8.3) при $\lambda_1=\lambda_3-\sum_{1}^{k+1} \hat{p}_j \ln \hat{p}_j$. Заменим в (8.7) неравенство на равенство и обозначим какое-то его вектор-решение $\mathbf{p}^\circ$. Этот вектор совпадает с таким же определяемым правым равенством в (8.5), если подставить туда вновь пересчитанное значение λ_1, поэтому первичные средние будут совпадать с (8.5) при соответствующих $\mathbf{p}^\circ$, что ведет к следующей записи первичных средних доверительной модели и уравнений для $\mathbf{p}^\circ$ и λ_3:

$$\overline{M} \sum_{1}^{k+1} \hat{p}_j A_j(\mathbf{z})/p_j^0 = 1, \quad J(\hat{\mathbf{p}}, \mathbf{p}^0) = \lambda_3,$$

$$\rho = \min_{\mathbf{p}} \sum_{\mathbf{r}:\, J(\hat{\mathbf{p}}, \mathbf{p}) \leqslant \lambda_3} P_{\mathbf{p}}(\mathbf{r}).$$

Доверительные совместные оценки. Пусть $\mathscr{A}_k=\{A_1, \ldots, A_k\}$ — произвольный (в общем, пересекающийся) набор событий на $\mathscr{Z}$. Требуется получить совместные доверительные оценки вероятностей этих событий $p_j=P(A_j)$ в виде произведения оценок $\mu_{\mathbf{z}}(p_j)$ отдельных событий: $\mu_{\mathbf{z}}(\mathbf{p})=\prod_{1}^{k} \mu_{\mathbf{z}}(p_j)$. Требование к надежности совместной оценки накладывает ограничения на надежности отдельных оценок. Этот вопрос здесь и рассматривается, а именно получить совместную доверительную оценку, имея отдельные оценки.

Для отдельного события A_j доверительная оценка $\underset{\smile}{p}_j$, $\tilde{p}_j$ уровня α_j дается как решение уравнений [25]

$$\sum_{i=0}^{r_j-1} C_N^i \underset{\smile}{p}_j^i (1-\underset{\smile}{p}_j)^{N-i} = 1-\alpha_{j0}, \quad \sum_{i=0}^{r_j} C_N^i \tilde{p}_j^i (1-\tilde{p}_j)^{N-i} = \alpha_{j1},$$

где $\alpha_{j0}+\alpha_{j1}=\alpha_j$ и r_j — частота этого события. В асимптотическом варианте при $N\to\infty$ и $\alpha_{j0}=\alpha_{j1}=\alpha_j/2$ доверительные границы будут приближенно равны

$$\hat{p}_j \pm \Phi^{-1}(1/2-\alpha_j/2) \sqrt{\hat{p}_j(1-\hat{p}_j)/N},$$

где $\Phi(x)$ — функция Лапласа.

Теорема 8.1. *Пусть $\varkappa=1$ и $\mu_{\mathbf{r}}(p_j)$, $j=1, \ldots, k$, есть доверительные интервальные оценки параметров $p_j=P(A_j)$ уровней α_j, определяемые границами $\underset{\smile}{P}(A_j)=\underset{\smile}{p}_j$, $\tilde{P}(A_j)=\tilde{p}_j$. Тогда эти границы, взятые за первичные, задают доверительную интервальную модель $\mathscr{M}_\rho$ надежности, по крайней мере, $\rho\geqslant 1-\sum_{1}^{k}\alpha_j$.*

Доказательство. Для совместной оценки $\mu_{\mathbf{r}}(\mathbf{p})=\prod_{1}^{k}\mu_{r_j}(p_j)$, используя элементарное неравенство $\prod_{1}^{k}\mu_{r_j}(p_j)\geqslant 1-\sum_{1}^{k}[1-\mu_{r_j}(p_j)]$, имеем: $\rho=\underline{M}\mu_{\mathbf{r}}(\mathbf{p})=$ $=\underline{M}\prod_{1}^{k}\mu_{r_j}(p_j)\geqslant 1-\sum_{1}^{k}\overline{M}[1-\mu_{r_j}(p_j)]=1-\sum_{1}^{k}\alpha_j$, что и требовалось.

Замечания. 1. Теорема 8.1 имеет смысл для любых совместных доверительных оценок задающих параметров. Совершенно не обязательно, что это оценки вероятностей и что они интервальные. Так, если $\mu_{\mathbf{z}}(m_j)$, $j=1, \ldots, k$, есть расплывчатые доверительные оценки параметров $m_j=Mg_j(\mathbf{z})$, то размытая модель надежности $\rho\geqslant 1-\sum_{1}^{k}\alpha_j$ будет определяться первичными размытыми средними $\mu_{\mathbf{z}}(m_j)$ признаков $g_j(\mathbf{z})$, $j=1, \ldots, k$, и k — размерность модели (если нет избыточности в ее первичных средних).

2. Если уровни α_j, $j=1, \ldots, k$, считать равными между собой, то в соответствии с теоремой 8.1 нужно брать $\alpha_j=(1-\rho)/k$. Это значение является явно заниженным, особенно при большом k, что приведет к излишне расширенной доверительной модели, поэтому требуется пересчет надежности по уже выбранным первичным средним (доверительным оценкам) отдельных параметров. Это делается так. Пусть $m_j=M\sum_{i}g_{ji}A_i(\mathbf{z})=\sum_{i}g_{ji}p_i$, где A_i — непересекающиеся множества, образующие разбиение $\mathscr{Z}$, а p_i — их вероятности. Тогда пересчитанная надежность

$$\rho=\min_{\mathbf{p}}\sum_{\forall\mathbf{r}}P_{\mathbf{p}}(\mathbf{r})\,\mu_{\mathbf{r}}(m).$$

3. При увеличении числа первичных параметров надежность модели при заданных оценках отдельных параметров, в общем, падает. Кроме того случая, когда задающие параметры «сильно связаны» между собой, т. е. один лишь незначительно отличается от другого. Такая связь имеет место, если задающими являются вкладывающиеся друг в друга события, к рассмотрению чего и приступим.

Доверительная функция распределения. Пусть задающим (он же первичный) является набор вкладывающихся друг в друга событий $A_\theta\subset\mathscr{Z}$, $\theta\in\mathscr{R}$, $A_\theta\subset A_{\theta'}$ при $\theta\leqslant\theta'$; обозначим через $\hat{p}_\theta=$ $=\sum_{1}^{N}\{z_i\in A_\theta\}/N$ — значения относительных частот попадания последовательности z_i испытаний в множества A_θ. Если отобразить $\mathscr{Z}$ в $\mathscr{R}$ так, что A_θ отображается в полуинтервал $(-\infty, \theta)$, то $\hat{p}_\theta$ как функция θ будет выборочной функцией распределения. Обозначим $p_\theta=P(z\in A_\theta)$. Требуется найти доверительные границы для p_θ, такие, что $\underline{P}(\prod_{\theta}\{\hat{p}_\theta-c_N(\rho)\leqslant p_\theta\leqslant\hat{p}_\theta+c_N(\rho)\})=\rho$. Разность $|\hat{p}_\theta-p_\theta|$ — известная статистика Колмогорова, поэтому $c_N(\rho)$ есть табулированные процентные точки распределения этой статистики [25]. Таким образом, искомыми первичными значениями, определяющими доверительную модель, будут

$$\underset{\sim}{P}(A_\theta)=\hat{p}_\theta-c_N(\rho),\quad \breve{P}(A_\theta)=\hat{p}_\theta+c_N(\rho).$$

Мы видим, что несмотря на бесконечное в данном случае число задающих параметров p_θ доверительные границы отдельных параметров не становятся тем не менее тривиальными, не расширяются до интервала (0; 1), как это следовало бы из теоремы 8.1.

8.3. СОГЛАСОВАННЫЙ СИНТЕЗ МОДЕЛЕЙ И ПРАВИЛ

Надежность моделей и истинные ошибки правил. Процесс изготовления промышленных изделий «обрастает», хотим мы этого или не хотим, массой подготовительных работ и вспомогательных служб. Нужно раздобыть сырье, сделать заготовки (модели), скомплектовать их, доставить к месту, а для этого нужно иметь помещение, подготовить станки, технику, наконец, найти рабочих и заинтересовать их зарплатой, организовать экономические службы (хотя и это еще, конечно же, не все). И только потом можно приступать к самому изготовлению. На надежность, ритмичность работы нужно смотреть в комплексе с охватом всего отлаженного механизма предприятий в целом.

Примерно то же самое имеет место при синтезе модели, где подготовительные работы состоят в выборе $\mathscr{M}^{xy}$, $\mathscr{D}$, $[\pi]$, $\varkappa$, составляющих статистическую задачу, а самое главное — в выборе модели $\mathscr{M}^{xy}$, своего рода инструмента к будущему изделию — решающему правилу (заготовкой к которому является $\mathscr{D}$). Брак в части инструмента делает бессмысленным само дальнейшее «изготовление» (какая бы ни была заготовка), поэтому в первую очередь модель $\mathscr{M}^{xy}$ должна быть надежной, доверительной. В то же время нельзя и это требование доводить до полного абсурда, забывая о целевом назначении модели — прямо вести к изделию — решающему правилу. Чрезмерные издержки на инструмент поднимут и время изготовления, и суммарную стоимость. Здесь нужен компромисс.

Будем мыслимо под «изделиями» подразумевать правила доверительного оценивания (хотя это может быть проверка гипотез). У модели и у правила свои атрибуты: у модели — это надежность, у правила — ошибка (уровень значимости). Последняя рассчитывается по виду модели. Очевидно, имея ненадежную модель, нельзя уже доверять ошибке (расчетному уровню значимости α), рассчитанной (по модели) для правила, ибо истинная вероятность ошибки ожидается выше расчетной. В то же время слишком широкая и отсюда чрезмерно надежная модель приведет к неоправданному увеличению расчетных ошибок правил. Возникающее противоречие рождает потребность вскрыть строгую связь надежности модели с ошибками правил, к чему и перейдем.

Пусть $\alpha^\varkappa(\partial)$ — уровень, или *расчетная вероятность ошибки* решающего правила ∂, рассчитанного по доверительной модели $\mathscr{M}_\rho$ надежности ρ. Тогда с вероятностью $1-\rho$, с какой модель «бракована», эта ошибка не соответствует истине, и неконтролируема. Пессимизм $\varkappa=1$ при расчете заставляет усугубить ситуацию, считать ошибку правил из-за «брака» модели максимально возможной, равной 1; а крайний оптимизм $\varkappa=0$ — наоборот, есть, по сути, вера, что все будет как нельзя лучше, т. е. ошибка минимальна и равна 0. В результате можно подметить, что неконтролируемая ошибка просто равна коэффициенту пессимизма $\varkappa$. Так как «брак» модели по ее построению закладывается с вероятностью $1-\rho$, то *истинная* (полная) *вероятность ошибки* правил составляется из неконтролируемой $\varkappa$ с вероятностью $1-\rho$ и расчетной с вероятностью ρ:

$$\alpha^\varkappa_и(\partial)=\rho\alpha^\varkappa(\partial)+(1-\rho)\varkappa. \qquad (8.8)$$

Подставив в полученную формулу $\alpha^\varkappa(\partial)=\varkappa\overline{\alpha}(\partial)+(1-\varkappa)\underline{\alpha}(\partial)$, где $\overline{\alpha}(\partial)=1-M\partial$, $\underline{\alpha}(\partial)=1-\overline{M}\partial$, перепишем истинную вероятность ошибки в другом виде:

$$\alpha^\varkappa_и(\partial)=\varkappa[1-\rho(1-\overline{\alpha}(\partial)+\underline{\alpha}(\partial))]+\rho\underline{\alpha}(\partial).$$

Отсюда наблюдается связь истинной вероятности ошибки с коэффициентом пессимизма $\varkappa$. Чем больше пессимизм $\varkappa$, тем больше истинная ошибка $\alpha^\varkappa_и(\partial)$. Но эта ошибка всегда не выше $\alpha^0_и(\partial)=\rho\underline{\alpha}(\partial)$, что соответствует $\varkappa=0$. При $\varkappa=1$ истинная ошибка $\alpha^1_и(\partial)=\overline{\alpha}(\partial)+(1-\rho)[1-\overline{\alpha}(\partial)]$ складывается из верхней границы ошибки $\overline{\alpha}(\partial)$ и (при $\overline{\alpha}(\partial)\ll 1$) ненадежности $1-\rho$ (вероятности брака) доверительной модели.

Если записать (8.8) еще в одном виде: $\alpha^\varkappa_и(\partial)=\alpha^\varkappa(\partial)+(1-\rho)[\varkappa-\alpha^\varkappa(\partial)]$, то будет видно, что при $\varkappa>\alpha^\varkappa(\partial)$ истинная вероятность ошибки всегда больше расчетной $\alpha^\varkappa(\partial)$, причем чем меньше надежность ρ и больше пессимизм, тем существенней эта разница, тем больше истинная вероятность ошибки. При $\varkappa\geqslant 1-\alpha^\varkappa(\partial)$ истинная ошибка превышает расчетную, по крайней мере, на величину ненадежности $1-\rho$.

Установим связь между истинной ошибкой правила и надежностью модели, для чего запишем при $\varkappa=1$: $\alpha^1_и(\partial)=1-\rho[1-\overline{\alpha}(\partial)]$. Увеличение ρ вроде бы должно уменьшать $\alpha^1_и(\partial)$. На самом деле, при увеличении ρ доверительная модель расширяется, в результате $\alpha(\overline{\partial})$ возрастает, устремляясь к 1 при $\rho\to 1$. Тогда и $\alpha^1_и(\partial)\to 1$. Таким образом, брать большую надежность модели бессмысленно. А так как (в силу установленного выше) и малая надежность не имеет смысла, то *для каждой статистической задачи должно существовать оптимальное значение надежности* ρ (пример будет рассмотрен в последнем разделе).

Пусть требуется обеспечить заданный истинный уровень α оценки: $\alpha^\varkappa_и(\partial)\leqslant\alpha$. Подстановкой (8.8) находится ограничение снизу на надежность $\rho\geqslant(\varkappa-\alpha)/[\varkappa-\alpha^\varkappa(\partial)]$ и величина расчетного уровня $\alpha_{рас}$, на который нужно настраивать правило: $\alpha^\varkappa(\partial)\leqslant[\alpha-\varkappa(1-\rho)]/\rho=\alpha_{рас}$. Если $\alpha<\varkappa(1-\rho)$, то последнее неравенство невыполнимо: нельзя найти $\alpha_{рас}$, нацелив на которое правило, получили бы истинный уровень α, что говорит о невозможности получения по ненадежным моделям низкого уровня α оценок. Всегда $\alpha\geqslant\varkappa(1-\rho)$ — *истинный уровень правила больше ненадежности модели, взвешенной коэффициентом пессимизма.*

Сказанное точь-в-точь переносится на расчет ошибки первого рода $\alpha^\varkappa(\partial)$ при проверке гипотез. Но если оценка откликается на падение расчетного уровня (по сравнению с истинным) расширением, что своего рода реакция (протест) на ненадежность модели, то у правила проверки гипотез понижение расчетной ошибки $\alpha^\varkappa(\partial)$ по принципу качелей вызовет рост расчетной ошибки второго рода $\beta^\varkappa(\partial)$, и тем самым истинной, определяемой согласно формуле (8.8) будет: $\beta^\varkappa_и(\partial)=\rho\beta^\varkappa(\partial)+\varkappa(1-\rho)$.

Теперь понятен общий случай, включающий в себя рассмотренные частные и охватывающий также задачу фильтрации. Приведенные соображения приводят к следующему выражению для *истинного риска*:

$$\Pi^\varkappa_и(\partial)=\rho\Pi^\varkappa(\partial)+(1-\rho)[\varkappa\sup_\partial\overline{\Pi}(\partial)+(1-\varkappa)\inf_\partial\underline{\Pi}(\partial)], \qquad (8.9)$$

где первое слагаемое есть риск, рассчитываемый по доверительной модели и взвешенный ее надежностью ρ, а слагаемые в квадратных скобках — это, в зависимости от степени пессимизма $\varkappa$, отражающего «настроение» правил, тот наибольший ущерб и соответственно наименьший, которого можно ждать от «бракованной» модели, причем эта часть риска от ∂ зависеть не будет.

Нужно отметить, что структура оптимальных правил от перерасчетов риска, в общем, защищена, так как всецело определяется первым слагаемым (8.9), и в конечном счете, первичными признаками доверительной модели. Гибкими останутся отдельные параметры этих правил, вариации которых помогают управлять расплывчатостью и ошибкой.

Итак, установлено, что *ненадежность модели ведет к необходимости введения поправок в ошибки правил: истинные ошибки будут, в общем, больше расчетных.* При потребности обеспечить фиксированную истинную ошибку (уровень) нужно предусмотрительно брать заведомо меньшее расчетное значение со скидкой на ненадежность, что соответствует и что приведет к падению реальных качеств правил как оптимальных, так и неоптимальных.

Размытые доверительные модели и решения. Расплывчатые оценки задающих параметров ведут к разным моделям, причем индикаторные оценки ведут к ИМ, а неиндикаторные — к более общим размытым моделям § 2.3. И тут возникает вопрос, как для размытых (неинтервальных) моделей организовать синтез оптимальных правил. Этот важный момент оставался вне рамок рассмотрения, поскольку при синтезе мы ограничивались строго интервальными моделями. К его освещению и перейдем.

Пусть имеется задающий параметр $\theta = Mq(\mathbf{z})$ (пока всего один) и $\mu_{\mathbf{z}}(\theta)$ есть его расплывчатая оценка надежности $\rho = 1-\bar{\alpha}(\mu)$. Считаем оценку $\mu_{\mathbf{z}}(\theta)$ унимодальной контрастной функцией параметра θ (т. е. достигающей 0 и 1). При каждом числе $0\leqslant\gamma\leqslant 1$, называемой высотой горизонтального среза, неравенство $\mu_{\mathbf{z}}(\theta)\geqslant\gamma$ выделяет интервальную оценку (слой) $[\underset{\smile}{\theta}{}^{\gamma}, \overset{\frown}{\theta}{}^{\gamma}]$ задающего параметра θ, так что при непрерывной по θ функции $\mu_{\mathbf{z}}(\theta)$ имеем $\mu_{\mathbf{z}}(\underset{\smile}{\theta}{}^{\gamma})=\mu_{\mathbf{z}}(\overset{\frown}{\theta}{}^{\gamma})=\gamma$. Каждый срез $[\underset{\smile}{\theta}{}^{\gamma}, \overset{\frown}{\theta}{}^{\gamma}]$ в свою очередь определяет свою интервальную модель $\mathscr{M}^{\mathbf{z}}_{(\gamma)}=\langle\underset{\smile}{\theta}{}^{\gamma}, \overset{\frown}{\theta}{}^{\gamma}\rangle$, располагающуюся на высоте γ, причем $\mathscr{M}^{\mathbf{z}}_{(\gamma')}\subset\mathscr{M}^{\mathbf{z}}_{(\gamma)}$ при $\gamma'\geqslant\gamma$, а все вместе они, положенные друг на друга в соответствии с высотами, дадут размытую модель.

Пространство $\mathscr{Z}$ так или иначе связано с произведением $\mathscr{X}\times\mathscr{Y}$, поэтому по $\mathscr{M}^{\mathbf{z}}_{(\gamma)}$ определяется СИМ $\mathscr{M}^{xy}_{(\gamma)}$ на этом произведении, и для каждой из них находится свое оптимальное правило $\partial_y{}^{(\gamma)}(x)$, зависящее от высоты среза γ модели (своего рода привилегий к ней), причем чем больше γ, тем (при $\varkappa\geqslant 1$) более узкой будет $\mathscr{M}^{xy}_{(\gamma)}$ и менее расплывчатым — правила $\partial_y{}^{(\gamma)}(x)$, т. е. $\partial_y{}^{(\gamma')}(x)\leqslant\partial_y{}^{(\gamma)}(x)$ при $\gamma'\geqslant\gamma$. Теперь оптимальным правилом $\partial^*{}_y(x)$ при размытой модели $\mathscr{M}^{xy}_{(\gamma)}$, построенной по расплывчатой оценке $\mu_{\mathbf{z}}(\theta)$, будет

$$\partial^*_y(x)=\int\limits_0^1 \partial^{(\gamma)}_y(x)\,d\gamma. \qquad (8.10)$$

Поясним его, считая ∂ оценкой параметра $x\in\mathscr{R}$. Если $\partial_y{}^{(\gamma)}(x)$ есть при каждом γ интервальные оценки x, то так как они вкладываются друг в друга, будет иметь место равенство $\{\partial^*{}_y(x)\geqslant\gamma\}=\partial_y{}^{(\gamma)}(x)$, где слева стоит индикаторная функция множества. Таким образом, интервальные оценки для моделей $\mathscr{M}^{xy}_{(\gamma)}$, рассматриваемые как положенные друг на друга слои на высотах срезов $0\leqslant\gamma\leqslant 1$, формируют вместе расплывчатое правило $\partial^*{}_y(x)$. Это и есть наглядная интерпретация (8.10).

Истинный уровень или ошибка правила (8.10) рассчитывается по формуле (8.8), где $1-\rho$ есть уровень (вероятность ошибки) оценки $\mu_{\mathbf{z}}(\theta)$, а следовательно, ρ — надежность доверительной модели.

Мы рассмотрели один параметр $\theta=Mq$, задающий модель. Очевидно, все сказанное будет верно и для произвольного их числа, когда с целью построения доверительной модели находится совместная расплывчатая оценка $\mu_{\mathbf{z}}(\theta)$ «вектора» $\theta=MQ=\{Mq,\ q\in Q\}$ и эта оценка, в общем, не является интервальной, а имеет расплывчатый вид.

Адаптация, надежностное оценивание среднего при неизвестной дисперсии. Принцип адаптации состоит в сужении доверительной модели в процессе поступления наблюдений $y_1, \ldots, y_n$. Особенность в том, что «обучение» производится по тем же наблюдениям, по которым принимаются решения относительно состояний (оценивание, проверка гипотез); это первое. А во-вторых, уточнение модели осуществляется последовательно по n.

Для описания схемы адаптации рассмотрим один случай, когда требуется оценить параметр сдвига независимой выборки при неизвестной дисперсии флуктуаций. Адаптация состоит в оценивании дисперсий и подстановке в оценку параметра сдвига.

Пусть $y_i=x+\xi_i$, $i=1, 2, \ldots$, и пусть требуется оценить x, когда ξ_i независимы, имеют нулевые средние $M\xi_i=0$, ограниченные дисперсии $m_2=M\xi^2{}_i<\infty$ и четвертые моменты $\bar{M}\xi^4{}_i=\bar{m}_4 m^2{}_2$ (полезно представить $\xi_i=\sqrt{m_2}\zeta_i$, $M\zeta_i=0$, $M\zeta^2{}_i=1$, $M\zeta^4{}_i=\bar{m}_4$). При заданном m_2 и $\varkappa=1$ оптимальной расчетного уровня $\alpha_{\text{рас}}$ будет оценка (6.18):

$$\partial_y(x)=[1-\alpha_{\text{рас}}\, n\,(\hat{y}-x)^2/m_2]^+.$$

На самом же деле, m_2 не известно и нужно подставлять в эту формулу заведомо завышенное значение $\bar{m}_2$, что ведет к чрезмерному увеличению расплывчатости оценки. Здесь m_2 и будет задающим модель параметром, причем считается он стационарным. Адаптация состоит в оценивании дисперсии m_2 по наблюдениям и ее использовании для оценки среднего.

Оценка дисперсии должна быть доверительной, так как по ней строится СИМ, и при этом не должна зависеть от x. Используем для нее статистику $\hat{\sigma}^2{}_y=\widehat{y^2}-\hat{y}^2$, инвариантную к сдвигам x. Вид оценки при уровне $1-\rho$ ее значимости будет определяться формулой (6.27) (при $x=m_2$, $l=1$, $\sigma^2=M\zeta^2{}_i=1$, $\varkappa=1$)

$$\mu_{\mathbf{y}}(m_2)=[1-(1-\rho)\,n\,(1-\hat{\sigma}^2_{\mathbf{y}}/m_2)^2/c]^+,$$

где $c\approx(\bar{m}_4-1)$. Оценка $\mu_{\mathbf{y}}(m_2)$ как функция m_2 принимает максимальное значение 1 при $m_2=\hat{\sigma}^2{}_y$, она убывает по мере отклонения m_2 от этого значения, а при $m_2\leqslant\hat{\sigma}^2_y\left[1-\sqrt{\dfrac{c}{(1-\rho)\,n}}\right]^{-1}$ или $m_2\geqslant\hat{\sigma}^2_y\left[1+\sqrt{\dfrac{c}{(1-\rho)\,n}}\right]^{-1}$ оценка равна 0. Эта оценка и определяет расплывчатую доверительную модель (причем отнюдь не интервальную).

Для получения оценки x по расплывчатой модели используем методику предыдущего раздела. Согласно ей из неравенства $\mu_{\mathbf{y}}(m_2)\geqslant\gamma$ находятся интервалы $\underset{\smile}{m}{}^{\gamma}{}_2$, $\overset{\frown}{m}{}^{\gamma}{}_2$, соответствующие различным высотам γ срезов. Нас интересует лишь верхняя граница $\overset{\frown}{m}{}^{\gamma}{}_2$, так как оценка $\partial_y(x)$ определяется только ею. Имеем

$\tilde{m}^{\gamma}{}_2 = \hat{\sigma}^2_y[1+\sqrt{c(1-\gamma)/(1-\rho)n}]^{-1}$ и по формуле (8.10) получаем

$$\partial^*_{\mathbf{y}}(x) = \int_0^1 \left[1 - \alpha_{\text{рас}}\, n\left(1 - \sqrt{\frac{c(1-\gamma)}{(1-\rho)\,n}}\right)\frac{(x-\hat{y})^2}{\hat{\sigma}_y^2}\right]^+ d\gamma. \qquad (8.11)$$

Эта оценка эквивариантна к сдвигу: $\partial_{\mathbf{y}+1a}(x+a) = \partial_{\mathbf{y}}(x)$, и инвариантна масштабным изменениям: $\partial_{b\mathbf{y}}(x) = \partial_{\mathbf{y}}(x)$, $b>0$.

Расчетный уровень $\alpha_{\text{рас}}$ и надежность ρ, как это показано в начале параграфа, связаны соотношением $\rho(1-\alpha_{\text{рас}}) = 1-\alpha$, где α — истинный (требуемый) уровень правила. Выразив $1-\rho = (\alpha-\alpha_{\text{рас}})/(1-\alpha_{\text{рас}})$ и подставив в (8.11), можно было бы найти оптимальное значение $\alpha_{\text{рас}}$, минимизирующее расплывчатость оценки (8.11).

Чтобы обойти технические громоздкости, найдем значение $\alpha_{\text{рас}}$, минимизирующее расплывчатость при заданном γ. Для этого нужно минимизировать по $\alpha_{\text{рас}}$ коэффициент

$$\alpha_{\text{рас}}\, n\,[1 - \sqrt{c(1-\gamma)(1-\alpha_{\text{рас}})/n(\alpha-\alpha_{\text{рас}})}.$$

Дифференцируя его по $\alpha_{\text{рас}}$ и приравнивая 0, приходим к уравнению: $2\sqrt{(1-\alpha_{\text{рас}})(\alpha-\alpha_{\text{рас}})^3} - \sqrt{c(1-\gamma)/n}[2(1-\alpha_{\text{рас}})(\alpha-\alpha_{\text{рас}}) + \alpha_{\text{рас}}(1-\alpha)] = 0$. Нам нужно выявить качественную сторону, поэтому приближенно полагая $1-\alpha\approx 1$, откуда $1-\alpha_{\text{рас}} \geqslant 1-\alpha \approx 1$, и считая n достаточно большим, так что $c(1-\gamma)/n$ мало, получаем

$$\alpha_{\text{рас}} \approx \alpha - \sqrt[3]{c(1-\gamma)\alpha^2/4n},\ \rho = 1 - \sqrt[3]{c(1-\gamma)\alpha^2/4n}.$$

Мы видим, что *при увеличении n расчетный уровень $\alpha_{\text{рас}}$ устремляется к истинному со скоростью $1/\sqrt[3]{n}$, а оптимально выбранное ρ с той же скоростью стремится к* 1, так что суммарная ошибка $\alpha_{\text{рас}} + (1-\rho) = \alpha$ есть истинный уровень.

Для получения конкретных расчетных значений $\alpha_{\text{рас}}$ и ρ вместо неизвестного γ нужно подставить среднее между 0 и 1 число, например $\gamma = 1/2$, и положить $c \approx \overline{m}_4 - 1$. Здесь $\overline{m}_4$, если оно не известно, также может оцениваться по наблюдениям или находиться из других соображений.

8.4. ЗАКЛЮЧЕНИЕ

Какую же выгоду все же сулит использование интервальных моделей? В качественном отношении они проигрывают точным (распределениям вероятностей), и это как будто бы ставит на них крест.

Представьте на миг, что наблюдается доселе невиданный объект и требуется описать его форму. Причем объект плохо различим (находится в тумане или далеко от нас). Ясно, что придется оговаривать условность сделанного описания вставками типа «вроде бы», «кажется», либо удовлетвориться грубым образом, уняв воображение по отношению к тому, что не различимо (наука — не фантастика).

То же самое и для моделей, которые видятся в большинстве реальных задач весьма смутно из-за конечности времени и выделенных средств на изучение явления, наконец, сменности, неустойчивости самих реальных явлений.

Прекрасно осознавая, что ошибочность модели тут же сделает никчемной ее дальнейшую эксплуатацию, постараемся осторожнее, т. е. в размытой форме описать грани модели, отобрав при этом лишь наиболее «видимую» их часть. И приходим к «чистокровным» интервальным моделям в их конструктивном задании набором первичных средних.

Построение модели в реальных условиях — сфера многосторонних (подчас, многострадальных) исследований: физических, экспериментальных статистических. Нас интересуют формальные способы, когда при полной начальной неясности в распоряжение предоставляются обучающие реализации. Тогда построение модели суть совместное оценивание отобранной части параметров, задающих модель (§ 8.1). Точечные оценки ведут к точным моделям. А доверительные оценки формируют доверительную модель заданной надежности. Вопрос сведется к выбору задающих модель параметров. Ими могут быть вероятности, совместное оценивание которых рассматривается в § 8.2.

Другой путь выбора модели, свойственный классическому подходу, состоит в проверке согласия, что модель имеет выдвинутый конкретный вид (например, нормальная). Этот путь, обязанный во многом крайней узости рабочего арсенала точных моделей, заставляет с самого начала «довольствоваться» заранее выбранным гипотетическим в меру простым вариантом. Принятие последнего, если согласие имеет место, тем не менее не приведет к сколь-либо надежной модели, а всего лишь установит вхожесть гипотетического варианта в нашу доверительную ИМ-модель в качестве составной части. В конечном же счете это будет «выхватывание» из доверительной модели ее заранее сформированного кусочка — подход, свойственный режиму оптимизма.

Мы же поступаем значительно осторожнее, используя доверительную модель всю целиком (необъятный арсенал ИМ позволяет произвести любой выбор), регулируя ширину с помощью надежности. Чем больше положить надежность, тем шире доверительная модель, от чего пострадает конкретность и качество выводов при эксплуатации модели. Наоборот, меньшая надежность, казалось бы, выводы сделает более конкретными и качественными, но доверие к ним уменьшит из-за утраты верности модели. Выход из этого заколдованного круга состоит в совместном рассмотрении тандема модель-выводы (§ 8.3), беря в расчет оба типа ошибок: за счет ненадежности модели и при эксплуатации из-за случайности наблюдений. Совместное рассмотрение заставляет вводить поправки в расчетные ошибки эксплуатации, увеличивая их сообразно ненадежности модели, что ведет к истинным ошибкам, а при более широком изложении — к истинному риску. Это как раз и есть то, что объективно нужно для целей анализа и синтеза решающих правил.

Трудности в том, что доверительные модели по наследству от породивших их доверительных оценок, в общем, расплывчатых, неиндикаторных, становятся размытыми по форме средних, т. е. с размазанными итервалами средних. Определение понятия оптимального правила при размытых статистических моделях позволило в § 8.4 рассмотреть совместную картину синтеза модели по ее единственному задающему параметру — дисперсии, с последующим нахождением оптимальной оценки параметра сдвига. Все вместе это выглядит как доверительное оценивание дисперсии (уровень доверия которой и станет надежностью модели) с последующим использованием ее при доверительном оценивании сдвига при своем уже доверии (расчетной ошибке). Причем все производится по одной и той же выборке наблюдений, по мере удлинения которой происходит уточнение оценки дисперсии, т. е. адаптационное сужение модели.

Подчеркнем еще раз, что главным итогом рассмотрения тандема модель-правило явился истинный риск (истинные ошибки) синтезированного правила. А надежность модели есть всего лишь вспомогательный атрибут синтеза, приобретающий конкретное значение путем минимизации истинного риска. Такая оптимальная надежность существует и найдена в рассмотренной нами задаче адаптации § 8.4, где установлена ее тенденция с ростом длины выборки стремиться к единице со скоростью кубического корня.

Теперь мы можем дать обоснованный ответ на поставленный в начале вопрос: выгода интервальных моделей состоит в получении объективно надежных обоснованных во всех отношениях решающих правил с оценкой их истинных качеств. Это и есть главное достижение надежностного синтеза.