Глава 6.

РАСПЛЫВЧАТОЕ ОЦЕНИВАНИЕ

6.1. ОБЩИЕ ВОПРОСЫ

Ошибки правил. Детерминированные (точечные) оценки $\hat{x}_y$ по своему внутреннему содержанию таковы, что обладают мизерной вероятностью $\bar{P}(\hat{x}_y=x)$ «угадывания» правильного состояния, а чаще всего и вовсе нулевой (соответственно вероятностью ошибки $P(\hat{x}_y \neq x)$ — единичной). И эта природная их черта не есть следствие того, что оценки плохи (для определения целей они могут быть очень даже хороши), а просто вызвана тем, что хоть как-то попасть «дрожащей» точкой $\hat{x}_y$ в точку x на числовой прямой или числовом пространстве невозможно, учитывая количество точек и ничтожность точечных размеров. Это можно сделать только тогда, когда истинное x «накалывается» не точкой, а накрывается чем-то ощутимым, например интервалом, что и подводит нас к расплывчатым оценкам.

Потребность в расплывчатых оценках возникает прежде всего там, где помимо значения состояния нужна точность, с которой оно оценивается. Например, ставится задача — сопровождать указание расстояния до цели величиной погрешности, диапазоном разброса, с которым оно измеряется. Так порождается *доверительный интервал.* Чем шире он, тем меньше вероятность, что при указании произойдет ошибка. Эта ошибка может входить как составляющая риска, но может быть частью исходных технических требований на оценку, аппаратуру, тогда ошибку нужно фиксировать, поддерживать на определенном уровне. В этом случае говорим об оценке фиксированного уровня α ошибки, или просто уровня α.

В наших рассуждениях сейчас не столь важно, является ли $\mathscr{X}$ числовым или каким другим, и даже не число элементов x в $\mathscr{X}$, поэтому говорим о правилах как общих случаях оценок.

Пусть $\partial_y(x)$ — расплывчатое правило — функция как x, так и y. Рассмотрим сначала его надежность — величину, обратную ошибке. Для заданного правила $\partial_y(x)$ нижняя и верхняя надежности соответственно равны $\underline{M}\partial_y$, $\overline{M}\partial_y$, где средние берутся по СИМ $\mathscr{M}^{xy}$. В самом деле, величина $\partial_y(x)$ есть уверенность, с которой x указывается как возможное значение искомого состояния по результату наблюдения y. При разных y и x эта величина колеблется от 0 до 1 сообразно виду правила, причем 0 соответствует тому, что x отвергается как приемлемый вариант состояний, а 1 — наоборот, соответствует бесприкословному включению; величина $\partial_y(x)=1/2$ означает, что при заданном y некоторое x принимается как возможное состояние с вероятностью 1/2, т. е. с половинчатой уверенностью. Надежность же есть уверенность в среднем, т. е. определенная в среднестатистическом смысле, что и заложено в $\overline{M}\partial_y$.

Ненадежность меряется как нижняя и верхняя вероятности ошибки: $\underline{\alpha}(\partial)=1-\overline{M}\partial_y$, $\overline{\alpha}(\partial)=1-\underline{M}\partial_y$, а взвешенная вероятность ошибки как объединенный показатель

$$\alpha^{\varkappa}(\partial)=(1-\varkappa)\,\underline{\alpha}\,(\partial)+\varkappa\,\overline{\alpha}\,(\partial)=1-(1-\varkappa)\,\overline{M}\,\partial_y-\varkappa\,\underline{M}\,\partial_y, \qquad (6.1)$$

где $\varkappa$ — коэффициент пессимизма.

Нулевой вероятностью ошибки обладают тривиальные правила $\partial_y(x)\equiv 1$, тождественно равные 1, соответствующие при любом x фразе: «Какое-то x из $\mathscr{X}$ имеет место» (но какое?). Его надежность равна 1, так как ошибиться здесь невозможно (по определению пространства элементарных исходов). Другая крайность — единичная вероятность ошибки, свойственная точечным (детерминированным) оценкам, причем чаще единичной будет оказываться именно верхняя ошибка $\overline{\alpha}(\partial)=1$, а нижняя при этом может вполне быть нулевой (тогде взвешенная ошибка будет равна $\varkappa$ — вере в неблагоприятный исход). Предметный интерес для теории представляют не крайности, а «золотая середина», т. е. унимодальные как функции x при каждом y контрастные (достигающие 0 и 1) правила ограниченной по x ширины.

Не ограждая класс $\mathscr{D}$ расплывчатых правил какими-либо барьерами, т. е. $\mathscr{D}=\mathscr{D}_{\forall}$, выделим из него подкласс следующим условием.

Правила ∂_y, для которых $\alpha^{\varkappa}(\partial)=\alpha$, называются *правилами уровня* α, а при числовом $\mathscr{X}$ — *доверительными оценками* x. В дальнейшем мы будем заниматься синтезом именно таких оценок.

Расплывчатость, риск. Помимо ошибок у рассматриваемых правил имеется другая «теневая сторона» — их расплывчатость.

Желательно, чтобы она была как можно меньше, т. е. при каждом y конкретнее указывалось бы искомое состояние x, но это вступает в противодействие с ошибкой правил, которая при этом увеличивается.

Расплывчатость характеризуется шириной правил как функций переменной x при заданном y; может измеряться разными способами. Прежде чем приступить к их рассмотрению, подумаем, а зачем здесь какое-то разнообразие, нужно ли оно? Оказывается, да, так как это своего рода регулировка, инструмент направленного воздействия на оптимальные правила, придания им тех или иных желаемых качеств, черт, оттенков.

Приведем разные шкалы расплычатости, пригодные как для скалярного, так и векторного параметров $x\in\mathscr{R}^k$ (лишь для векторного $x=(x_1, \ldots, x_k)$ и интегралы по $dx=dx_1\cdot\ldots\cdot dx_k$ становятся кратными).

1. *Интегральная шкала*: $\Omega(\partial_y)=\int\partial_y(x)\,dx$ — наиболее проста по смыслу, так как придает всем иксам одни и те же веса. Для одномерного x это будет приведенная (к прямоугольнику единичной высоты той же площади) ширина, а многомерного — приведенный объем.

2. *Взвешенная шкала*: $\Omega_q(\partial_y)=\int q(x)\,\partial_y(x)\,dx$. Весовая функция $q(x)$ нужна для выделения более важных (скажем, в смысловом или стратегическом отношении) состояний, вынуждая их более точное оценивание в ущерб остальным.

3. *Обобщенная шкала*: $\Omega^{\gamma}(\partial_y)=\int q(x)\,\partial_y(x)^{\gamma}dx$. Возведение решений в степень $\gamma<1$ увеличивает ∂^{γ}_y по сравнению с ∂_y, если последнее <1, тем самым увеличивая плату за неуверенность решений и стимулируя их повышенную категоричность типа 0 и 1 (нет, да). При $\gamma>1$, наоборот, будут поощряться неуверенные решения.

4. *Эффективная ширина:* $\int\partial_y(x)\,dx/\sup\limits_x\partial_y(x)$. Это есть при каждом y отношение площади (объема) под $\partial_y(x)$ как функции переменной x к высоте $\overline{\partial}_y$. Таким способом создаются благоприятные условия для контрастных категоричных правил, у которых $\overline{\partial}_y=1$, и для них это будет просто интегральная шкала.

5. *Периметр* свойствен векторному x и определяется как максимальная по компонентам x_i интегральная ширина правил:

$$\max_x\max_i\,[q_i\int\partial_y\,(x)\,dx_i],$$

где q_i — веса, выделяющие относительную «важность» разных компонент вектора x.

Введенные шкалы пригодны и для дискретных x, если интегралы обменять на суммы по x.

Величина расплывчатости по введенным шкалам зависит от наблюдений y. Так как расплывчатость правил $\partial_y(x)$ может быть, в общем, разной при разных наблюдениях y, поэтому нужно говорить о средней расплывчатости (ширине), усредняя $\Omega(\partial_y)$ по y согласно СИМ $\mathscr{M}^{xy}$ (по частной к ней $\mathscr{M}^{y}$). Средняя ширина есть

22

штраф за расплывчатость и меряется, в общем, сверху и снизу: $\overline{M}\Omega(\partial_y)$, $\underline{M}\Omega(\partial_y)$, где $\Omega(\ )$ — любая из указанных выше шкал.

Если брать взвешенную сумму $(1-\varkappa)\underline{M}\Omega(\partial_y)+\varkappa\overline{M}\Omega(\partial_y)$, то при пессимизме $\varkappa\geqslant 1$ нижнее значение будет с отрицательным знаком, и это уже, извините, будет не штраф, а поощрение расплывчатости, что сделает невозможным основные результаты о достаточности (справедливые, как увидим, при $\varkappa\geqslant 1$), поэтому за штраф будем брать верхнее значение $\overline{M}\Omega(\partial_y)$ (по крайней мере, годное при пессимизме).

Назовем *составным риском расплывчатых правил* взвешенную сумму вероятности ошибки и верхнего штрафа за расплывчатость (для скалярного x это средняя ширина):

$$\Pi^\varkappa_\lambda(\partial)=\alpha^\varkappa(\partial)+\lambda\overline{M}\Omega(\partial_y). \qquad (6.2)$$

Весовой коэффициент λ призван регулировать отношение между слагаемыми: при увеличении λ бо́льшая важность придается расплывчатости, нежели ошибке, в результате оптимальные правила, т. е. те, которые минимизируют $\Pi^\varkappa{}_\lambda(\partial)$, будут более точными и менее надежными. И наоборот.

Обратим внимание, что составной риск (6.2) не есть средний риск (среднее от потерь) в смысле предыдущей главы, а представляет собой некоторую очень простую форму характеризации негативных сторон правила. Тем не менее он сохраняет структуру риска как взвешенного коэффициентом пессимизма нижнего и верхнего его значений:

$$\Pi^\varkappa_\lambda(\partial)=(1-\varkappa)\underline{\Pi}(\partial)+\varkappa\overline{\Pi}(\partial)$$

при $\underline{\Pi}(\partial)=\underline{M}\partial_y$, $\overline{\Pi}(\partial)=\overline{M}\partial_y+\dfrac{\lambda}{\varkappa}\overline{M}\Omega(\partial_y)$.

Почему мы вдруг рассматриваем составной риск, а не средний? Да потому что составной риск делает простым синтез расплывчатых оценок и удовлетворяет (с помощью регулировочных шкал) всем практическим требованиям.

Регулируя λ, можно добиться любого фиксированного уровня α оптимального правила: если ошибка оказалась больше меньше α, то λ следует чуть уменьшить, стимулируя тем самым расплывчатость, а если меньше больше — то увеличить, и далее снова посмотреть, что при этом получится с ошибкой оптимального правила.

Минимизация по $\partial\in\mathcal{D}_V$ составного риска с параллельной «подгонкой» значения λ есть способ синтеза оптимальных правил фиксированного уровня α, обладающих минимальной средней шириной в классе правил уровня α, т. е. решающих задачу: $\min\limits_{\partial:\alpha^\varkappa(\partial)=\alpha}\overline{M}\Omega(\partial_y)$.

Оптимальные расплывчатые правила при заданных совместных плотностях вероятностей. Совместная плотность являет собой крайнее исключение, когда данные о практически всех вероятностях событий, связанных как с x, так и с y, имеются в виде точных значений. Выделение такого случая можно было бы

назвать отступлением от общей нашей тенденции иметь дело с грубыми моделями, описываемыми конечным числом данным. Тем не менее такой отход позволяет, во-первых, уяснить влияние шкалы расплывчатости на вид оптимального правила, и во-вторых, продемонстрировать простоту нового аппарата синтеза доверительных правил на классической «почве» распределений вероятностей.

Пусть $\mathscr{X}=\mathscr{R}^k$, т. е. $\mathbf{x}=(x_1,\ldots,x_k)$ и $\mathscr{Y}=\mathscr{R}^n$, т. е. $\mathbf{y}=(y_1,\ldots,y_n)$, и пусть СИМ задается точной совместной плотностью $p(\mathbf{x},\mathbf{y})$ по мере-длине на $\mathscr{R}^{k+n}$ (обобщение на другие пространства и меры для нас не принципиально). Тогда частные плотности будут равны: $p(\mathbf{y})=\int p(\mathbf{x},\mathbf{y})\,d\mathbf{x}$, $p(\mathbf{x})=\int p(\mathbf{x},\mathbf{y})\,d\mathbf{y}$. Считаем автоматически выполненными все необходимые условия измеримости по отношению к алгебре отрезков на $\mathscr{R}^{k+n}$, обязанной сопровождать плотность (см. § 1.5).

Для (измеримых) оценок $\partial_y(x)$ ошибка будет точной величиной, равной $\alpha(\partial)=1-\int\int\partial_y(\mathbf{x})p(\mathbf{x},\mathbf{y})\,d\mathbf{x}d\mathbf{y}$. Перейдем к составлению риска и нахождению оптимальных оценок для разных шкал расплывчатости.

Взвешенная шкала. Составной риск для нее записывается:

$$\Pi_\lambda(\partial)=1+\iint\partial_\mathbf{y}(\mathbf{x})\,[-p(\mathbf{x},\mathbf{y})+\lambda p(\mathbf{y})\,q(\mathbf{x})]\,d\mathbf{x}\,d\mathbf{y}.$$

Так как $\partial_\mathbf{y}(\mathbf{x})$ заключено в пределах от 0 до 1, то правилом, минимизирующим этот риск, будет:

$$\partial^*_\mathbf{y}(\mathbf{x})=\begin{cases}1 \text{ при } p(\mathbf{x},\mathbf{y})>\lambda p(\mathbf{y})\,q(\mathbf{x}),\\ 0 \text{ при } p(\mathbf{x},\mathbf{y})<\lambda p(\mathbf{y})\,q(\mathbf{x}).\end{cases} \qquad (6.3)$$

Оптимальным уровня α будет найденное правило при таком выборе λ, чтобы

$$1-\alpha=M^{\mathbf{xy}}\partial^*_\mathbf{y}(\mathbf{x})=\iint\limits_{p(\mathbf{x},\mathbf{y})>\lambda p(\mathbf{y})q(\mathbf{x})}p(\mathbf{x},\mathbf{y})\,d\mathbf{x}\,d\mathbf{y},$$

где неравенство снизу определяет совместную на $\mathscr{X}\times\mathscr{Y}$ область интегрирования. Обозначим $p_\mathbf{y}(\mathbf{x})=p(\mathbf{x},\mathbf{y})/p(\mathbf{y})$ и назовем *апостериорной плотностью*. Оптимальное решающее правило будет индикаторным и состоит в сравнении при каждом $\mathbf{y}$ отношения $p_\mathbf{y}(\mathbf{x})/q(\mathbf{x})$ с порогом λ и присвоении $\partial^*_\mathbf{y}$ значения 1, если это отношение выше порога, и 0, если ниже.

Функция $q(\mathbf{x})$ определяет вес штрафов в зависимости от $\mathbf{x}$. Для интегральной шкалы $q(\mathbf{x})\equiv 1$ и тогда $\partial^*{}_\mathbf{y}$ определяется сравнением апостериорной плотности $p_\mathbf{y}(\mathbf{x})$ с порогом λ. При $q(\mathbf{x})=p(\mathbf{x})$ с порогом будет сравниваться отношение $p(\mathbf{x},\mathbf{y})/[p(\mathbf{x})p(\mathbf{y})]$.

Возможна замена усредненного на максимальный по $\mathbf{y}$ штраф $\sup\limits_\mathbf{y}\int\partial_\mathbf{y}(\mathbf{x})\ q(\mathbf{x})dx$, имеющий смысл, если при любом $\mathbf{y}$ требуется контролировать ширину правила $\partial_{\mathbf{y}}(\mathbf{x})$ по переменной $\mathbf{x}$. Тогда составной риск приобретает вид:

$$\Pi_\lambda(\partial) = 1 - M^{\mathbf{xy}}\,\partial_{\mathbf{y}}(\mathbf{x}) + \lambda \sup_{\mathbf{y}} \int \partial_{\mathbf{y}}(\mathbf{x})\, q(\mathbf{x})\, d\mathbf{x}.$$

Записав $\inf\limits_{\partial} \Pi_\lambda(\partial) = \sup\limits_{p_0(y)} \inf\limits_{\partial} [1 - M^{\mathbf{xy}}\,\partial_{\mathbf{y}}(\mathbf{x}) + \lambda\, p_0(\mathbf{y}) \int \partial_{\mathbf{y}}(\mathbf{x})\, q(\mathbf{x})\, d\mathbf{x}]$ (где, используя соответствующие утверждения теории игр [22], inf и sup поменяли местами), мы придем к правилу (6.3), в котором $p(\mathbf{y})$ заменено на $p_0(\mathbf{y})$, определяемое минимизацией выражения $\iint [p(\mathbf{x}, \mathbf{y}) - \lambda p_0(\mathbf{y}) q(\mathbf{x})]\, d\mathbf{x} d\mathbf{y}$. В некотором смысле $p_0(\mathbf{y})$ выбирается таким образом, чтобы произведение $p_0(\mathbf{x}) q(\mathbf{x})$ было по возможности более «похожим» на $p(\mathbf{x}, \mathbf{y})$.

Обобщенная шкала. Составной риск

$$\Pi_\lambda(\partial) = 1 - \iint [-\partial_{\mathbf{y}}(\mathbf{x})\, p(\mathbf{x}, \mathbf{y}) + \lambda\, \partial_{\mathbf{y}}(\mathbf{x})^\gamma\, p(\mathbf{y})\, q(\mathbf{x})]\, d\mathbf{x}\, d\mathbf{y}.$$

Оптимальное решающее правило может быть найдено минимизацией при каждом $\mathbf{x}$ и $\mathbf{y}$ выражения в квадратных скобках. После несложных вычислений приходим к следующему его виду:

$$\partial^*_{\mathbf{y}}(\mathbf{x}) = \min\left\{1, \left[\frac{p(\mathbf{x}, \mathbf{y})}{\lambda\gamma\, p(\mathbf{y})\, q(\mathbf{x})}\right]^{1/(\gamma-1)}\right\}, \quad \gamma \geqslant 1,$$

где λ находится из уравнения

$$\alpha(\partial) = \iint p(\mathbf{x}, \mathbf{y}) \left[1 - \left(\frac{p(\mathbf{x}, \mathbf{y})}{\lambda\gamma\, p(\mathbf{y})\, q(\mathbf{x})}\right)^{1/(\gamma-1)}\right]^+ d\mathbf{x}\, d\mathbf{y} = \alpha.$$

Параметр γ определяет наше отношение к сомнительным решениям $\partial_{\mathbf{y}}(\mathbf{x}) < 1$. При $\gamma = 1$ возвращаемся к (6.3). При $\gamma = 2$ получаем $\partial_{\mathbf{y}}(\mathbf{x}) = \min\{1, p_{\mathbf{y}}(\mathbf{x})/[2\lambda q(\mathbf{x})]\}$. Интересно отметить, что при $q(\mathbf{x}) \equiv 1$ данное решающее правило совпадает по форме с апостериорной плотностью, возможно, усеченной сверху величиной 2λ.

Замечания. 1. Выводы настоящего раздела согласно теореме 5.1 останутся в силе, если синтез производится в режиме полуоптимизма $\varkappa = 1/2$, а СИМ задана интервальной плотностью $\underset{\smile}{p}(\mathbf{x}, \mathbf{y})$, $\tilde{p}(\mathbf{x}, \mathbf{y})$, так что $\iint [\underset{\smile}{p}(\mathbf{x}, \mathbf{y}) + \tilde{p}(\mathbf{x}, \mathbf{y})]\, d\mathbf{x} d\mathbf{y} = 2$. Тогда в предыдущие выражения нужно подставлять $p(\mathbf{x}, \mathbf{y}) = [\underset{\smile}{p}(\mathbf{x}, \mathbf{y}) + \tilde{p}(\mathbf{x}, \mathbf{y})]/2$.

2. При дискретных x и y плотности заменяются на вероятности, а интегралы — на суммы.

Достаточные классы расплывчатых правил. Нам неважно опять, каковыми являются $\mathscr{X}$ и $\mathscr{Y}$. Это могут быть числовые или дискретные пространства, поэтому вместо оценок в общем говорим о расплывчатых решениях. Неважно также, какой является при этом шкала расплывчатости; она может быть совсем другой, чем рассмотренные выше. Важно, что для шкал должно выполняться очевидное условие

$$d_1(x) > d_2(x) \Rightarrow \Omega(d_1) \geqslant \Omega(d_2), \tag{6.4}$$

т. е. более расплывчатым и более неопределенным решениям должно соответствовать большее (по крайней мере, не меньшее) число по шкале Ω. Как и выше, полагаем $\mathscr{D} = \mathscr{D}_\forall$ — класс всех оценок.

Выпишем выражение (6.2) для составного риска, раскрыв подробнее ошибку:

$$\Pi^\varkappa_\lambda(\partial) = 1 - \varkappa\, \underline{M}\, \partial_y + (\varkappa - 1)\, \overline{M}\, \partial_y + \lambda\, \overline{M}\, \Omega(\partial_y). \tag{6.5}$$

Следующая теорема о достаточности является ключевой для настоящей главы. Хотя достаточность здесь специальная, так как привязывается к конкретному риску (6.5), но зато напрямую обращается к структуре первичных признаков СИМ, а заодно и позволяет крайне просто найти верхнюю (наименее благоприятную) ошибку, наиболее важную для нас.

Теорема 6.1. *При СИМ $\langle \overline{M}\mathscr{G} \rangle$, пессимизме $\varkappa \geqslant 1$ и условии (6.4), достаточном в смысле составного риска (6.5) классом расплывчатых правил будет усеченный снизу осью абсцисс подкласс вторичных, не превышающих 1 признаков*

$$\mathscr{D}^* = \{\partial_y(x) = [c_0 - \textstyle\sum c_i^+ g_i(x, y)]^+ : c_0 \in \mathscr{R},\ c_i^+ \in \mathscr{R}^+,$$

$$g_i \in \mathscr{G},\ \partial_y(x) \leqslant 1\}. \tag{6.6}$$

Причем в него могут быть включены только те $\partial^*_y(x)$, для которых нижняя надежность равна

$$\underline{M}\, \partial^*_{\mathbf{y}} = c_0 - \textstyle\sum c_i^+\, \overline{M}\, g_i. \tag{6.7}$$

Доказательство. Для любого ∂_y согласно следствию к теореме 1.1, переписанному для нижнего среднего, имеем $\underline{M}\partial = \sup\limits_{-\partial \leqslant -g \in \mathscr{L}^+ \mathscr{G}} \underline{M}g$. Отсюда заданному $\varepsilon > 0$ можно всегда подыскать такую $-g_\varepsilon \in \mathscr{L}^+\mathscr{G}$, что $-g_\varepsilon \geqslant -\partial$ и $\underline{M}\partial \leqslant \underline{M}g_\varepsilon + \varepsilon$. С учетом неравенств $g_\varepsilon(x, y) \leqslant g_\varepsilon(x, y)^+ \leqslant \partial_y(x)$, где плюс означает взятие неотрицательной части, имеем $\underline{M}\partial - \varepsilon \leqslant \underline{M}g_\varepsilon \leqslant \underline{M}g^+_\varepsilon$, $\overline{M}\partial \geqslant \overline{M}g^+_\varepsilon$. Функция $g_\varepsilon(x, y)^+$ есть решающее правило, так как $0 \leqslant \overline{g}^+_\varepsilon \leqslant \overline{\partial} \leqslant 1$. Теперь из (6.5) на основании условия (6.4) и найденных отношений получаем:

$$\Pi^\varkappa_\lambda(\partial) = 1 - \varkappa\, \underline{M}\partial + (\varkappa - 1)\, \overline{M}\partial + \lambda \overline{M}\Omega(\partial) \geqslant 1 - \varkappa\, \underline{M}g^+_\varepsilon +$$

$$+ (\varkappa - 1)\, \overline{M}\, g^+_\varepsilon + \lambda \overline{M}\Omega(g^+_\varepsilon) - \varkappa\varepsilon = \Pi^\varkappa_\lambda(g^+_\varepsilon) - \varkappa\varepsilon,$$

что и доказывает достаточность класса $\mathscr{D}^*$ правил.

Осталось доказать последнюю часть теоремы. Если в (6.7) вместо равенства стоит неравенство, то по следствию к теореме 1.1 существует другой вторичный признак $-g \in \mathscr{L}^+\mathscr{G}$, для которого $g(x, y)^+ \leqslant \partial_y(x)$ и $\underline{M}\partial = \underline{M}g + \varepsilon$, а так как в силу первого неравенства $\overline{M}g^+ \leqslant \overline{M}\partial$, то риск (6.5) у правила g^+ будет (при $\varepsilon \to 0$) не больше, чем у ∂, и последнее может быть исключено из достаточного класса.

Заметим, что обозначение $\overline{M}g_i$ в правой части формулы (6.7) вместо $\tilde{M}g_i$ означает, что оставляются лишь полулинейные комбинации согласованных первичных признаков.

Следующие ниже утверждения получаются из теоремы 6.1 расшифровкой класса вторичных признаков при разных способах задания СИМ. Эти утверждения тут же сопровождаются поясняющими примерами.

Следствие 1. *Пусть СИМ разложима* $\mathcal{M}^{xy}=\mathcal{M}^x\mathcal{M}^y{}_x$, *причем* $\mathcal{M}^x=\langle \tilde M\mathcal{H}\rangle$, $\mathcal{M}^y{}_x=\langle \tilde M_x\Psi\rangle$ *определены своими первичными средними* $\tilde M h_i(x)$, $h_i\in\mathcal{H}$; $\tilde M_x\psi_j(y)$, $\psi_j\in\Psi$. *Тогда при* $\varkappa\geqslant 1$ *достаточным будет класс правил*

$$\mathcal{D}^*=\{\partial_y(x)=[c_0-\sum a_i^+ h_i(x)-\sum c_i^+(x)(\psi_i(x,y)-$$

$$-\tilde M_x\psi_i)]^+ : h_i\in\mathcal{H},\ \psi_i\in\Psi\},$$

причем таких, что $\partial_y(x)\leqslant 1$, $\underline{M}^y{}_x\partial_y(x)=c_0-\sum a^+{}_i\bar M h_i$.

Пример 6.1. Пусть $\mathcal{M}^y{}_x$ определяется всего одним первичным признаком $\psi(y)$ и значением $\tilde M_x\psi$, зависящим от x, а $\mathcal{M}^x=\mathcal{J}^x$ — голая модель. Тогда $\mathcal{D}^*$ составляют $\partial_y(x)=[c_0-c^+(x)(\psi(y)-\tilde M_x\psi)]^+$ при выборе коэффициентов c_0 и $c^+(x)\geqslant 0$, удовлетворяющих неравенству $c_0+c^+(x)(\psi(y)-\tilde M_x\psi)\leqslant 1$, причем $\underline{M}\partial=c_0$. Мы видим, что так как $c^+(x)$ есть произвольная неотрицательная функция, то правила $\partial_y(x)$ как функции переменной x при каждом заданном y могут иметь произвольный вид, но как функций переменной y их вид целиком определяется первичной функцией $\psi(y)$.

Следствие 2. *Пусть СИМ задана функциональным представлением* $y=V_x\xi$ *и моделью* $\mathcal{M}^{x\xi}=\langle\tilde M^x\mathcal{H}\rangle\langle\tilde M^\xi\Psi\rangle$ *(т. е. флуктуации* ξ *свободны от* x*). Пусть оператор* V_x *обратим:* $\xi=V_x^{-1}y$. *Тогда при* $\varkappa\geqslant1$ *достаточными в смысле составного риска (6.5) будут решающие правила*

$$\mathcal{D}^*=\{\partial_y(x)=[c_0-\sum a_i^+h_i(x)-\sum c_i^+(x)(\psi_i(V_x^{-1}y)-$$

$$-\tilde M\psi_i)]^+ : h_i\in\mathcal{H},\ \psi_i\in\Psi\},$$

причем такие, что $\partial_y(x)\leqslant1$, $\underline{M}\partial=c_0-\sum a^+{}_i\bar M h_i$.

Пример 6.2. Пусть наблюдения есть смесь сигнала x и шума: $y_i=x+\xi_i$, $i=1,\dots,n$, $x\in\mathcal{R}$, $\xi_i\in\mathcal{R}$, или в векторном виде $\mathbf{y}=\mathbf{1}x+\xi$, где $\mathbf{1}=(1,\dots,1)^T$ — единичный вектор. Считаем, что шум ξ свободен от x и имеет модель $\mathcal{M}^\xi$, заданную первичными значениями $\tilde M\psi_j(\xi)$, $\psi_j\in\Psi$, а модель сигнала x — значениями $\tilde M h_i(x)$, $h_i\in\mathcal{H}$. Тогда класс $\mathcal{D}^*$ образуют правила вида $\partial_y(x)=[c_0-\sum a^+{}_i h_i(x)-\sum c^+{}_j(x)(\psi_j(\mathbf{y}-\mathbf{1}x)-\tilde M\psi_j)]^+$, при $\partial_y(x)\leqslant1$, $\underline{M}\partial=c_0-\sum a^+{}_i\bar M h_i$.

Откажемся в следствии 2 от предположения свободы ξ от x.

Следствие 3. *Пусть* $y=V_x\xi$, *причем оператор* V_x *обратим, и пусть первичными средними заданы частные* $\mathcal{M}^x=\langle\tilde M\mathcal{H}\rangle$, *и* $\mathcal{M}^\xi=\langle\tilde M\Psi\rangle$. *Тогда при* $\varkappa\geqslant1$ *и риске (6.4) достаточным будет следующий класс правил:*

$$\mathcal{D}^*=\{\partial_y(x)=[c_0-\sum a_i^+h_i(x)-\sum c_i^+\psi_i(V_x^{-1}y)]^+ : h_i\in$$

$$\in\mathcal{H},\ \psi_i\in\Psi\},$$

причем таких, что $\partial_y(x)\leqslant1$ *и* $\underline{M}\partial=c_0-\sum a^+{}_i\bar M h_i-\sum c^+{}_i\bar M\psi_i$.

Пример 6.3. Пусть $y_i=\sum_{j=1}^k w_{ij}x_j+\xi_i$, $i=1,\dots,n$, или в векторных обозначениях $\mathbf{y}=\mathbf{W}\mathbf{x}+\xi$, где $\mathbf{W}$ — матрица w_{ij}. Пусть заданы только первичные средние $\tilde M\psi_i$ флуктуаций ξ, а о связи ξ с x совершенно ничего не известно. Тогда достаточный (при $\varkappa\geqslant1$) класс образуют правила вида $\partial_{\mathbf{y}}(\mathbf{x})=[c_0-\sum c^+{}_i\psi_i(\mathbf{y}-\mathbf{W}\mathbf{x})]^+$ при таком выборе коэффициентов, что $\partial_{\mathbf{y}}(\mathbf{x})\leqslant1$ и $\underline{M}\partial=c_0-\sum c^+{}_i\bar M\psi_i$. Нетрудно видеть, если перейти к $k=1$, $\mathbf{W}=\mathbf{1}$, что этот класс проще того, что получился в предыдущем примере.

Следствие 4. *Пусть* $y=V_x\xi$, *причем* V_x *обратим, и пусть об* x *ничего не известно* $\mathcal{M}^x=\mathcal{J}^x$, *как и о его связи с* ξ *(либо* x *свободен от* ξ*), а* $\mathcal{M}^\xi=\langle\tilde M\Psi\rangle$ *основывается на первичных признаках* $\psi_i(\xi)$, $\psi_i\in\Psi$. *Тогда достаточный класс образуют правила вида* $[c_0-\sum c^+{}_i\psi_i(\xi)]^+\leqslant1$, *куда подставляется* $\xi=V_x^{-1}y$, *причем* $\underline{M}\partial=c_0-\sum c^+{}_i\bar M\psi_i$.

Например, если $\mathbf{y}=\mathbf{W}\mathbf{x}+\xi$, то оптимальное правило есть функция $\mathbf{y}-\mathbf{W}\mathbf{x}$: $\partial_y(\mathbf{x})=\partial(\mathbf{y}-\mathbf{W}\mathbf{x})$, в этом виде его и нужно искать.

Замечание. Если СИМ задается в виде семейства $\bigvee\mathcal{M}_\theta^{xy}$, или в виде пересечений $\bigwedge\mathcal{M}_\theta^{xy}$, а класс $\mathcal{D}^*$ достаточен при каждой $\mathcal{M}_\theta^{xy}$, то он будет достаточным для объединений и для пересечений. Это верно хотя бы потому, что классы вторичных признаков, если они одни и те же для $\mathcal{M}_\theta^{xy}$, при объединениях и пересечениях сохраняются.

Оптимизм и достаточность. Согласно предыдущему разделу режим пессимизма (т. е. $\varkappa\geqslant1$) позволяет произвести предварительную обработку, состоящую в редукции к достаточному классу $\mathcal{D}^*$ расплывчатых правил, что облегчает решение статистической задачи в плане нахождения оптимального правила.

Возникают два вопроса. Один — возможна ли при оптимизме $\varkappa<1$ такая предварительная обработка? Общего ответа на этот вопрос дать нельзя. Причина в том, что любая редукция наблюдений ведет к потере данных и расширению интервальной модели. В свою очередь, это уменьшает $\underline{M}\partial$ и увеличивает $\bar M\partial$, т. е. эти величины как слагаемые риска (6.5) будут меняться в противоположных направлениях и при $\varkappa<1$ трудно уследить, в какую сторону поведет их взвешенную сумму, тем более делать общие утверждения на этот счет, составляющие смысл достаточности.

Отсюда возникает другой вопрос, насколько целесообразно пользоваться определенными при $\varkappa\geqslant1$ достаточными классами $\mathcal{D}^*$, если $\varkappa<1$? Отчасти ответ освещается следующим легко проверяемым утверждением.

Утверждение 6.2. *Пусть для всех правил достаточного (при* $\varkappa\geqslant1$*) класса* $\mathcal{D}^*$ *верно:* $\bar M\partial^*{}_y=\sup_{x,y}\partial^*{}_y(x)$. *Тогда класс* $\mathcal{D}^*$ *будет достаточным и при* $\varkappa<1$.

Итак, видим, что рассмотренные в предыдущем пункте достаточные классы правил $\mathcal{D}^*$ будут достаточными и при $\varkappa<1$, если в $\alpha^{\varkappa}(\partial)$ «нейтрализовано» слагаемое с $\overline{M}\partial_y$. Иначе возникает чисто математическая возможность уменьшить ошибку $\alpha^{\varkappa}(\partial)$ за счет искусственного увеличения $\overline{M}\partial_y$ при сохранении неизменным $\underline{M}\partial_y$. Приведем пример реализации такой возможности, чтобы осмыслить всю абсурдность, к которой можно прийти, если пойти возникшим путем.

Пример 6.4. Допустим СИМ задана точными вероятностями $p_{ij}=P(x\in A_i,\ y\in B_j)$ на произведении разбиений $A_i\in\mathscr{A}^x{}_\Sigma$ и $B_j\in\mathscr{B}^y{}_\Sigma$ соответственно пространств $\mathscr{X}$ и $\mathscr{Y}$. Оптимальное правило по теореме 6.1 записывается в виде $\partial_y(x)=1-\sum\sum c_{ij}A_i(x)B_j(y)$, где $0\leqslant c_{ij}\leqslant 1$. Пусть теперь $\varkappa<1$, а $\mathscr{X}$ и $\mathscr{Y}$ есть линейные пространства (скажем, $\mathscr{X}\in\mathscr{R}$, $\mathscr{Y}\in\mathscr{R}^n$), причем множества A_i и B_j содержат каждое более, чем по одной точке этого пространства. Тогда, добавив к каждому слагаемому $c_{ij}A_i(x)B_j(y)$, у которого $c_{ij}<1$, произведение дельта-функций $(1-c_{ij})\delta_{x_{(i)}}(x)\delta_{y_{(j)}}(y)$, где $x_{(i)}$ и $y_{(j)}$ есть произвольные представители множеств A_i и соответственно B_j, приходим к правилу

$$\partial^*{}_y(x)=1-\sum\sum[c_{ij}A_i(x)B_j(y)+(1-c_{ij})\delta_{x_{(i)}}(x)\delta_{y_{(j)}}(y)]$$

с дельта-выбросами, у которого за счет нулевой площади выбросов штраф тот же: $\int\partial^*{}_y(x)dx=\int\partial_y(x)dx$, но $\overline{M}\partial^*{}_y(x)=1\geqslant\overline{M}\partial_y(x)=\underline{M}\partial_y(x)=1-\sum\sum c_{ij}p_{ij}$. Цель выбросов — обеспечить в наиблагоприятнейшем случае единичную надежность $\overline{M}\partial^*=1$ и тем самым уменьшить риск при оптимизме.

Таким образом, отход при $\varkappa<1$ от достаточного класса $\mathscr{D}^*$ (соответствующего $\varkappa\geqslant 1$) может привести к «нарушению гармонии» в правилах, вряд ли практически оправданному, поэтому использование класса $\mathscr{D}^*$ представляется разумным и при $\varkappa<1$, даже если условия утверждения 6.2 не выполняются.

Симметрия статистических моделей и эквивариантность расплывчатых правил. Часто в задаче обнаруживается симметрия следующего содержания: синхронные изменения наблюдений эквивалентны соответствующему сдвигу параметра состояний. Мысль подтверждается на примере, когда $y_i=x+\xi_i$, где подъем всех y_i на число a равносилен сдвигу x на то же самое число. Нужно ожидать переноса названного свойства на оптимальные оценки x, а в общем, на правила, о чем и пойдет речь. Но в широком плане, когда синхронные изменения y и x математически постулируются как связанные между собой преобразования пространств $\mathscr{X}$ и $\mathscr{Y}$, или просто как особый класс преобразований произведения пространств $\mathscr{X}\times\mathscr{Y}$. Причем важно не только, чтобы эти преобразования не перекрещивали между собой $\mathscr{X}$ и $\mathscr{Y}$, но и чтобы они образовывали группу, куда входили последовательные преобразования, равно как и обратные. Это нужно для последующего применения принципов инвариантности.

Перейдем к формальному изложению. Рассмотрим группу $\mathscr{S}$ преобразований пространства $\mathscr{X}\times\mathscr{Y}$, сохраняющую на месте по отдельности $\mathscr{X}$ и $\mathscr{Y}$, т. е. такую группу, что каждое $\mathrm{s}\in\mathscr{S}$ отображает $\mathscr{X}$ на $\mathscr{X}$ и $\mathscr{Y}$ на $\mathscr{Y}$: $\mathrm{s}(x,\ y)=(\mathrm{s}_1x,\ \mathrm{s}_2y)$. Любое преобразование s записывается как совместная пара $(\mathrm{s}_1,\ \mathrm{s}_2)$ преобразова-

ний, первое из них s_1 действует только на $\mathscr{X}$, а второе s_2 — только на $\mathscr{Y}$. Называются s_1 и s_2 частными к s преобразованиями. Дадим обобщение на совместные преобразования $\mathrm{s}=(\mathrm{s}_1,\ \mathrm{s}_2)$ результатов § 5.4 (где рассматривались преобразования лишь одного пространства наблюдений $\mathscr{Y}$), используя введенные там понятия и некоторые результаты, интерпретированные к нашему случаю.

Статистическая интервальная модель называется *симметричной к группе* $\mathscr{S}$, если для любого преобразования из этой группы

$$\overline{M}f(\mathrm{s}_1x,\ \mathrm{s}_2y)=\overline{M}f(x,\ y),\quad \forall f\in\mathscr{F}.$$

Симметрия достигается, если симметричен первичный набор $\mathscr{G}$ модели в том смысле, что каждой $g(x,\ y)\in\mathscr{G}$ и каждому преобразованию $\mathrm{s}\in\mathscr{S}$ может быть указан другой первичный признак $g_{\mathrm{s}}(x,\ y)=g(\mathrm{s}_1x,\ \mathrm{s}_2y)\in\mathscr{G}$ такой, что $\overline{M}g_{\mathrm{s}}(x,\ y)=\overline{M}g(x,\ y)$.

Утверждение 6.3. *Если* $\mathscr{M}^{xy}$ *симметрична к группе* $\mathscr{S}$, *то частные модели* $\mathscr{M}^x$ *и* $\mathscr{M}^y$ *будут симметричны к своим частным преобразованиям* s_1x *и* s_2y, $(\mathrm{s}_1,\ \mathrm{s}_2)\in\mathscr{S}$.

Утверждение следует из определения и соответствует равенствам $\overline{M}f(\mathrm{s}_1x)\equiv\overline{M}f(x)$, $\overline{M}\varphi(\mathrm{s}_2y)\equiv\overline{M}\varphi(y)$.

Функция $f(x,\ y)$ называется *эквивариантной* по отношению к группе преобразований $\mathscr{S}$ пространства $\mathscr{X}\times\mathscr{Y}$, если $f(\mathrm{s}_1x,\ \mathrm{s}_2y)\equiv\equiv f(x,\ y)$, $(\mathrm{s}_1,\ \mathrm{s}_2)\in\mathscr{S}$. Модель $\mathscr{M}^{xy}$ называется *эквивариантной к группе* $\mathscr{S}$, если все ее первичные признаки эквивариантны к $\mathscr{S}$. Смысл эквивариантности — компенсировать изменения y соответствующим изменением x. Инвариантность СИМ есть особый случай эквивариантности, когда $(\mathrm{s}_1x,\ \mathrm{s}_2y)=(x,\ \mathrm{s}_2y)$, т. е. каждое $\mathrm{s}\in\mathscr{S}$ преобразует только пространство $\mathscr{Y}$ (s_1 оставляет x четко на месте).

Если СИМ $\langle\overline{M}\mathscr{G}\rangle$ *эквивариантна к* $\mathscr{S}$, *то она будет симметрична к* $\mathscr{S}$. В самом деле все функции класса $\mathscr{L}^+\mathscr{G}$ будут эквивариантными и

$$\overline{M}f(\mathrm{s}_1x,\ \mathrm{s}_2y)=\inf\{\overline{M}g:f(\mathrm{s}_1x,\ \mathrm{s}_2y)\leqslant g(x,\ y)\in\mathscr{L}^+\mathscr{G}\}=$$

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

В соответствии с определением *решающее правило* будет (как f) *эквивариантным к* $\mathscr{S}$, если $\partial_{\mathrm{s}_2y}(\mathrm{s}_1x)\equiv\partial_y(x)$. Иначе может быть записано $\partial_{\mathrm{s}_2y}(x)\equiv\partial_y(\mathrm{s}_1{}^{-1}x)$, $(\mathrm{s}_1,\ \mathrm{s}_2)\in\mathscr{S}$. Например, для детерминированных правил имеем $\hat{x}_{\mathrm{s}_2y}=\mathrm{s}_1\hat{x}_y$, отсюда преобразование s_2y выборки y ведет к «смещению» решения до решения $\mathrm{s}_1\hat{x}$ в пространстве состояний.

Класс эквивариантных функций замкнут относительно линейных операций и нелинейных преобразований, т. е. из эквивариантности всех $g(x,\ y)\in\mathscr{G}$ следует эквивариантность их линейных комбинаций (а в общем — эквивариантность любых функций от них $\varphi(g(x,\ y))$). Отсюда на основании теоремы 6.1 имеем:

Утверждение 6.4. *При $\varkappa \geqslant 1$ достаточными для эквивариантных СИМ в смысле составного риска (6.5) будут эквивариантные правила.*

Значит, и оптимальные правила будут эквивариантны. Сейчас покажем, что в ряде случаев эквивариантными оптимальные правила должны быть и при симметричных СИМ. Назовем *шкалу инвариантной к* $\mathscr{S}$, если $\Omega(\partial_y(s_1 x)) = \Omega(\partial_y(x))$, $\forall (s_1, s_2) \in \mathscr{S}$. В случае $\Omega(\partial_y(x)) = \int \partial_y(x) dx$ шкала будет инвариантной, пожалуй, только к сдвигам $sx = s + x$.

Теорема 6.5. *Пусть СИМ симметрична к группе преобразований $\mathscr{S}$, а риск составной, определенный формулой (6.5), причем шкала Ω инвариантна к $\mathscr{S}$. Допустим, также выполнено любое из двух условий:*

А. *Оптимальное правило единственно или число этих правил конечно*;

Б. *Группа $\mathscr{S}$ дискретна. Тогда при любых $\varkappa$ оптимальное правило (а если их несколько, то хотя бы одно из них) будет эквивариантным к $\mathscr{S}$.*

Доказательство. Пусть оптимальное правило является единственным. Тогда, используя сначала инвариантность шкалы, а затем следующую из утверждения 6.3 симметрию частной модели, получим $\bar{M}\Omega(\partial_y(x)) = \bar{M}\Omega(\partial_y(s_1 x)) = \bar{M}\Omega(\partial_{s_2 y}(s_1 x))$. Отсюда

$$\Pi^{\varkappa}_{\lambda}(\partial_y(x)) = 1 - M^{(1-\varkappa)}\,\partial_y(x) + \lambda \cdot \bar{M}\,\Omega(\partial_y(x)) =$$

$$= 1 - M^{(1-\varkappa)}\,\partial_{s_2 y}(s_1 x) + \lambda \cdot \bar{M}\,\Omega(\partial_{s_2 y}(s_1 x)) = \Pi^{\varkappa}_{\lambda}(\partial_{s_2 y}(s_1 x))$$

и в силу единственности оптимального правила $\partial_y(x) = \partial_{s_2 y}(s_1 x)$.

Если теперь оптимальных правил не одно, а несколько, скажем, $\partial_y(x)_i$, $i = 1, \ldots, l$, то преобразованиями $s \in \mathscr{S}$ одно переходит в другое, так что заданным $s \in \mathscr{S}$ и i найдется такое j, что $\partial_{s_2 y}(s_1 x)_i = \partial_y(x)_j$. Производя равномерную рандомизацию между $\partial_y(x)_i$ (т. е. выбирая $\partial_y(x)_i$ с вероятностями $1/l$), приходим к эквивариантному решающему правилу. То же самое необходимо проделать, если группа $\mathscr{S}$ дискретна: $\mathscr{S} = (s^{(1)}, \ldots, s^{(l)})$. Теорема доказана.

Теорема проста для случая, когда $\mathscr{S}$ дискретна. Если же нет, то трудности может вызвать проверка условия **А**.

Максимальный инвариант относительно группы $\mathscr{S}$, поскольку по определению $\mathscr{S}$ не перемешивает между собой $\mathscr{X}$ и $\mathscr{Y}$, записывается как пара $I_1(x)$, $I_2(y)$, и тогда эквивариантное правило записывается через максимальный инвариант: $\partial_y(x) = \partial'_{I_2(y)}(I_1(x))$. При условиях утверждения 6.4 и теоремы 6.5 оптимальные правила нужно искать именно в таком виде.

6.2. ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ ПРИ ЗАДАННЫХ РАСПРЕДЕЛЕНИЯХ ВЕРОЯТНОСТЕЙ ФЛУКТУАЦИЙ

Предисловие. Отсюда мы начинаем «осаду» проблемы доверительного оценивания числового параметра $x \in \mathscr{R}$, априорных данных о котором совсем нет, с целью получения конкретных расплывчатых оценок $\partial_y(x)$, минимизирующих штраф $\bar{M}\Omega(\partial)$ за расплывчатость при заданной ошибке (уровне) $\alpha^{\varkappa}(\partial) = \alpha$, или что то же самое, минимизирующих составной риск $\Pi^{\varkappa}{}_{\lambda}(\partial) = \alpha^{\varkappa}(\partial) + \lambda\bar{M}\Omega(\partial)$ при соответствующем подборе λ, обеспечивающем уровень α. Основная задача — научиться пользоваться аппаратом. Здесь в качестве первого шага не лишне выяснить, как работает предлагаемый аппарат на привычной и хорошо «утоптанной» многими исследователями почве заданных распределений вероятностей и их семейств, и к чему он приводит. Этому и посвящен данный параграф. Неожиданным оказалось то, что и на «утрамбованном грунте» наш подход дает пробиться свежим «росткам» в виде как общих выражений для доверительных оценок векторного параметра x, так и получения новых доверительных интервалов для скалярного параметра.

Оценка регрессии при известной плотности вероятностей. Пусть $y_i = \sum_{j=1}^{k} W_{ij} x_j + \xi_i$, $i = 1, \ldots, n$, или в векторных обозначениях $\mathbf{y} = \mathbf{W}\mathbf{x} + \boldsymbol{\xi}$. Считается, что относительно вектора $\mathbf{x}^{\mathrm{T}} = (x_1, \ldots, x_k)$ ничего не известно, а вектор $\boldsymbol{\xi}^{\mathrm{T}} = (\xi_1, \ldots, \xi_n)$, именуемый флуктуациями, определен n-мерной плотностью $p_{\xi}(\mathbf{z})$, $\mathbf{z}^{\mathrm{T}} = (z_1, \ldots, z_n)$, по отношению к мере-длине на $\mathscr{R}^n$. Требуется оценить вектор $\mathbf{x}$. Относительно связи $\mathbf{x}$ и $\boldsymbol{\xi}$ никаких предположений не делается, она не известна (или так нам удобно это считать).

Согласно следствию 4 к теореме 6.1 достаточный (при $\varkappa \geqslant 1$) класс образуют правила вида $\partial(\mathbf{y} - \mathbf{W}\mathbf{x})$, где $\partial(\mathbf{z})$ измеримы относительно алгебры отрезков. Считая за штраф верхнюю среднюю площадь под $\partial_{\mathbf{y}}(\mathbf{x})$, что соответствует интегральной шкале расплывчатости, и учитывая отсутствие сведений о $\mathbf{x}$, преобразуем штраф:

$$\bar{M}\Omega(\partial) = \bar{M}^{\mathbf{y}} \int \partial(\mathbf{y} - \mathbf{W}\mathbf{u})\, d\mathbf{u} = \bar{M}^{\xi} \sup_{\mathbf{x}} \int \partial(\mathbf{W}\mathbf{x} + \boldsymbol{\xi} - \mathbf{W}\mathbf{u})\, d\mathbf{u} =$$

$$= \bar{M}^{\xi} \int \partial(\boldsymbol{\xi} - \mathbf{W}\mathbf{u})\, d\mathbf{u} = \int p_{\xi}(\mathbf{z})\, d\mathbf{z} \int \partial(\mathbf{z} - \mathbf{W}\mathbf{u}) d\mathbf{u} =$$

$$= \iint p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u})\, \partial(\mathbf{z})\, d\mathbf{z}\, d\mathbf{u}.$$

В силу измеримости $\partial(\mathbf{z})$ ошибка будет точной, равной $\alpha(\partial) = 1 - M\partial(\boldsymbol{\xi}) = 1 - \int \partial(\mathbf{z}) p_{\xi}(\mathbf{z}) d\mathbf{z}$, и составной риск

$$\Pi^{\varkappa}_{\lambda}(\partial) = 1 - \int \partial(\mathbf{z})\, p_{\xi}(\mathbf{z})\, d\mathbf{z} + \lambda \iint p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u})\, \partial(\mathbf{z})\, d\mathbf{z}\, d\mathbf{u} =$$

$$= 1 + \int \partial(\mathbf{z})\, [\lambda \int p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u})\, d\mathbf{u} - p_{\xi}(\mathbf{z})]\, d\mathbf{z}.$$

Минимизирует риск интервальная оценка $\partial^*(\mathbf{z})$, равная 1 при $\lambda \int p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u}) d\mathbf{u} - p_{\xi}(\mathbf{z}) \leqslant 0$ и 0 в противном случае. Цена будет $v = 1 - \int [p_{\xi}(\mathbf{z}) - \lambda \int p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u}) d\mathbf{u}]^{+} d\mathbf{z}$. После замены $\lambda_{\alpha} = 1/\lambda$ оценка записывается $\int p_{\xi}(\mathbf{z} + \mathbf{W}\mathbf{u}) d\mathbf{u} / p_{\xi}(\mathbf{z}) \leqslant \lambda_{\alpha}$ (указываются значения $\mathbf{z}$, при которых $\partial^*(\mathbf{z}) = 1$). Здесь λ_{α} находится по заданной ошибке из уравнения $\alpha(\partial^*) = 1 - \int \partial^*(\mathbf{z}) p_{\xi}(\mathbf{z}) d\mathbf{z} = \alpha$.

Выражение для доверительного интервала в явном виде получается заменой $\mathbf{z}=\mathbf{y}-\mathbf{Wu}$ и приобретает окончательный общий вид:

$$\partial^*_{\mathbf{y}}(\mathbf{x}) = 1 \text{ при } \frac{\int p_\xi(\mathbf{y}-\mathbf{Wx}+\mathbf{Wu})\,d\,\mathbf{u}}{p_\xi(\mathbf{y}-\mathbf{Wx})} \leqslant \lambda_\alpha, \tag{6.8}$$

иначе значение правила равно 0.

Нужно разрешить это неравенство относительно $\mathbf{x}$ и подобрать λ_α, что и ведет к доверительной индикаторной оценке заданного уровня α, а при одномерном $\mathbf{x}$ — к доверительному интервалу. Отметим, что, во-первых, постоянные множители левой части (6.8), не зависящие от $\mathbf{y}-\mathbf{Wx}$, удобно перенести в правую часть, и во-вторых, целесообразно искать упрощения с помощью монотонных нелинейных преобразований обеих частей, что сразу же и продемонстрируем.

Рассмотрим тот случай, когда ξ есть нормальный вектор с нулевым средним и матрицей корреляций $\mathbf{B}$, что соответствует плотности $p_\xi(\mathbf{z})=((2\pi)^n \det \mathbf{B})^{-1/2}\exp(-\mathbf{z}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{z}/2)$. Тогда по (6.8) после несложных вычислений находим

$$\frac{\int p_\xi(\mathbf{z}+\mathbf{Wu})\,d\,\mathbf{u}}{p_\xi(\mathbf{z})} = \frac{(\sqrt{2\pi})^k}{\sqrt{\det \mathbf{W}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{W}}} \times$$

$$\times \exp\left\{\frac{1}{2}\mathbf{z}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{W}(\mathbf{W}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{z}\right\}.$$

Отсюда отбрасыванием постоянного множителя, логарифмированием и подстановкой $\mathbf{z}=\mathbf{y}-\mathbf{Wx}$ получаем индикаторную оценку, равную 1 в области значений $\mathbf{x}$, ограниченной неравенством

$$(\mathbf{y}-\mathbf{Wx})^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{W}(\mathbf{W}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{y}-\mathbf{Wx}) \leqslant \chi^2_\alpha(k), \tag{6.9}$$

где $\chi^2{}_\alpha(k)$ — критическая точка распределения хи-квадрата с k степенями свободы, и равную 0 вне этой области.

Пример 6.5. Рассмотрим простейший случай одного неизменного по i параметра x, который нужно оценить по выборке $y_i=x+\xi_i$, $i=1, \ldots, n$, где ξ_i независимы и нормально распределены с $M\xi_i=0$, $M\xi^2{}_i=1$. Тогда $k=1$, и из (6.9) получим доверительный интервал для x, записанный в форме неравенства $|\hat{y}-x|\leqslant\lambda_\alpha$, где $\hat{y}=\sum\limits_1^n y_i/n$ и $\lambda_\alpha=\dfrac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\dfrac{1-\alpha}{2}\right)$, а $\Phi(u)=$ $=\dfrac{1}{\sqrt{2\pi}}\int\limits_0^u \exp(-u^2/2)\,du$ — функция Лапласа. В результате получается хорошо известный доверительный интервал для среднего при заданной дисперсии нормальной выборки

$$|\hat{y}-x| \leqslant \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1-\alpha}{2}\right).$$

Пусть теперь $y_i=xw_i+\xi_i$, $i=1, \ldots, n$, где ξ_i нормальны, зависимы и заданы корреляционной матрицей $\mathbf{B}=\{M\xi_i\xi_j,\ i,\ j=1, \ldots, n\}$,

а $M\xi_i=0$. Тогда оптимальной доверительной оценкой уже будет

$$\left|x-\frac{\mathbf{y}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{w}}\right| \leqslant (\mathbf{w}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{w})^{-1/2}\,\Phi^{-1}\left(\frac{1-\alpha}{2}\right). \tag{6.10}$$

Применительно к процессам $y_t=xw_t+\xi_t$, $M\xi_t=0$, $B(t,\tau)=$ $=M\xi_t\xi_\tau$ формула (6.10) переписывается:

$$\left|x-\frac{\int y_t h_t\,dt}{\int w_t h_t\,dt}\right| \leqslant \left(\int w_t h_t\,dt\right)^{-1/2}\Phi^{-1}\left(\frac{1-\alpha}{2}\right),$$

где h_t есть решение уравнения $\int B(t,\tau)h_\tau d\tau=w_t$.

При взвешенной шкале расплывчатости в знаменателе левой части неравенства в (6.8) появляется весовой множитель $q(\mathbf{y}-\mathbf{Wx})$, расширяющий интервал для тех x, которые имеют «повышенный приоритет» при сужении для остальных x.

В случае обобщенной шкалы $\Omega(\partial)=\int q(x)\partial_{\mathbf{y}}(x)^\gamma dx$, $\gamma>1$, оптимальная оценка уже не будет доверительным интервалом, а повторяет с некоторыми искажениями по форме плотность

$$\partial^*(\mathbf{z}) = \min\left\{1;\ \left[\frac{p_\xi(\mathbf{z})}{\lambda_\alpha q(\mathbf{z})\int p_\xi(\mathbf{z}-\mathbf{Wu})\,d\,\mathbf{u}}\right]^{1/(\gamma-1)}\right\},$$

где λ_α ищется по заданному уровню α. Например, при одном параметре x, при $q(\mathbf{z})\equiv 1$, нормальной плотности и независимых однородных флуктуациях с $M\xi_i=0$, $M\xi^2{}_i=\sigma^2$, имеем (рис. 6.1)

$$\partial^*_y(x) = \min\left\{1;\ \lambda\exp\left[-\frac{n}{2\sigma^2(\gamma-1)}(x-\hat{y})^2\right]\right\}.$$

Доверительное оценивание дисперсии. Пусть $y_i=x\xi_i$, $i=1, \ldots, n$, или в векторах $\mathbf{y}=x\xi$, где $x>0$ — оцениваемый параметр, свободный от ξ и описываемый голой моделью $\mathcal{I}^x$ (что соответствует полному отсутствию данных о x). Пусть известна плотность $p_\xi(\mathbf{z})$. Тогда достаточным при $\varkappa\geqslant 1$ будет класс всех измеримых правил вида $\partial_{\mathbf{y}}(x)=\partial(\mathbf{y}/x)$. Ошибка равна $\alpha(\partial)=1-M\partial(\mathbf{y}/x)=$ $=1-\int p_\xi(\mathbf{z})\partial(\mathbf{z})d\mathbf{z}$, а штраф

$$\overline{M}^{\mathbf{y}}\int\limits_0^\infty \partial(\mathbf{y}/u')\,du' = \overline{M}^{\xi}\sup_{x>0}\int\limits_0^\infty \partial(x\,\xi/u')\,du' =$$

$$= \int p_\xi(\mathbf{z})\sup_{x>0}\int\limits_0^\infty \partial(x\mathbf{z}/u')\,du'\,d\mathbf{z} = \overline{\sigma}\int p_\xi(\mathbf{z})\int\limits_0^\infty \partial(\mathbf{z}/u)\,du\,d\mathbf{z},$$

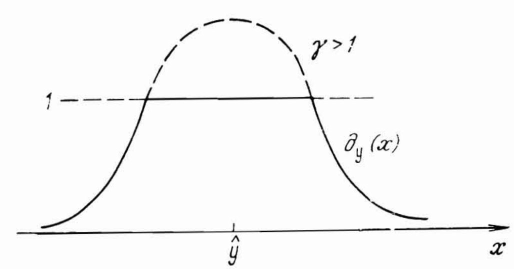


Рис. 6.1. Доверительная оценка среднего нормальной выборки при обобщенном штрафе

где произведена замена $u=u'/x$ и сделано допущение $\sup\limits_{x>0} x=\bar{\sigma}$. Необходимость такого допущения состоит в том, что ширина доверительной оценки растет с увеличением x, стремясь к ∞ при $x\to\infty$, вот и приходится вводить ограничение $x<\bar{\sigma}$ (причем $\bar{\sigma}$ можно считать любым сколь угодно большим числом). Тогда составной риск запишется

$$\Pi_\lambda(\partial)=1-\int \partial(\mathbf{z})\, p_\xi(\mathbf{z})\, d\mathbf{z}+\lambda\bar{\sigma}\int p_\xi(\mathbf{z})\int_0^\infty \partial(\mathbf{z}/u)\, du\, d\mathbf{z}.$$

Обозначая $\lambda\bar{\sigma}=1/\lambda_\alpha$ и производя во втором интеграле замену $\mathbf{z}'=\mathbf{z}/u$, учитывая при этом, что $d\mathbf{z}=u^n d\mathbf{z}'$, приводим риск к виду

$$\Pi_\lambda(\partial)=1-\int \partial(\mathbf{z})\left[p_\xi(\mathbf{z})-\lambda_\alpha^{-1}\int_0^\infty u^n p_\xi(u\mathbf{z})\, du\right] d\mathbf{z}.$$

Отсюда сразу же следует, что оптимальное доверительное решающее правило должно быть интервальным и иметь вид

$$\int_0^\infty u^n p_\xi(u\mathbf{y}/x)\, du/p_\xi(\mathbf{y}/x)\leqslant\lambda_\alpha, \tag{6.11}$$

где неравенство указывает значения x, при которых $\partial^*(\mathbf{y}/x)=1$ (иначе — 0); λ_α находится из уравнения $\int\partial^*(\mathbf{z})p_\xi(\mathbf{z})d\mathbf{z}=1-\alpha$, а при нахождении доверительного интервала совершена подстановка $\mathbf{z}=\mathbf{y}/x$.

В частном случае нормальной стандартной плотности флуктуаций оптимальный доверительный интервал в терминах переменной $\mathbf{z}=\mathbf{y}/x$ приобретает вид

$$\int_0^\infty u^n \exp\left(\|\mathbf{z}\|^2\,\frac{1-u^2}{2}\right) du=\frac{\exp(\|\mathbf{z}\|^2/2)}{\|\mathbf{z}\|^{n+1}}\int_0^\infty u^n\exp(-u^2/2)\, du\leqslant\lambda.$$

Отнесем интеграл к правой части, обозначив ее λ_n, и пусть $\underline{u}(\lambda_n)$ и $\bar{u}(\lambda_n)$ есть решения уравнения $\exp(u^2/2)=\lambda_n u^{n+1}$. Тогда искомый доверительный интервал запишется $\underline{u}^2(\lambda_n)\leqslant\|\mathbf{z}\|^2\leqslant\bar{u}^2(\lambda_n)$, если верно неравенство $\lambda_n\geqslant\exp[(n+1)^2/2]/(n+1)^{n+1}$. При равенстве доверительный интервал вырождается в точку, причем $\underline{u}^2(\lambda_n)=\bar{u}^2(\lambda_n)=n+1$, а при обратном неравенстве — бессмыслен.

Найдем λ_n в соответствии с заданной ошибкой α. Для этого обратим внимание на тот факт, что $\|\xi\|^2$ по определению есть случайная величина, распределенная по закону хи-квадрата с n степенями свободы. Подставляя плотность этого распределения, выводим

$$P\,[\underline{u}^2(\lambda_n)\leqslant\|\xi\|^2\leqslant\bar{u}^2(\lambda_n)]=\int_{\underline{u}^2(\lambda_n)}^{\bar{u}^2(\lambda_n)}\frac{v^{n/2}-1}{2^{n/2}\,\Gamma(n/2)}\,e^{-v/2}\, dv=1-\alpha,$$

где $\Gamma(\ )$ — гамма-функция. Отсюда находится значение $\lambda_{\alpha,n}$ как решение этого уравнения. Обозначая $\underline{u}_{\alpha,n}=\underline{u}(\lambda_{\alpha,n})$, $\bar{u}_{\alpha,n}=$

$=\bar{u}(\lambda_{\alpha,n})$ и подставляя $\|\mathbf{z}\|^2=\|\mathbf{y}\|^2/x^2$, получим окончательно следующий доверительный интервал уровня α:

$$\|\mathbf{y}\|^2/\bar{u}^2_{\alpha,n}\leqslant x^2\leqslant\|\mathbf{y}\|^2/\underline{u}^2_{\alpha,n}.$$

При $\alpha\to1$ имеем $\underline{u}^2{}_{\alpha,n}\uparrow n+1$, $\bar{u}^2{}_{\alpha,n}\downarrow n+1$ и доверительный интервал стягивается к точке $\|\mathbf{y}\|^2/(n+1)=\sum y^2{}_i/(n+1)$.

Таким образом, *получен доверительный интервал для дисперсии нормальной выборки, обладающий вообще минимальной расплывчатостью*, в том числе среди всех известных. Математическое ожидание наблюдений $\mathbf{y}$ считалось нулевым:

$$My_i=0,\ i=1,\ \dots,\ n.$$

Пусть теперь среднее $My_i=\theta$ совершенно неизвестно, т. е. $y_i=x\xi_i+\theta$. Считая θ свободным параметром, а ξ_i — независимыми стандартными нормальными с. в., путем вовлечения понятия инвариантности (к сдвигу) можно прийти к тому, что оптимальной оценкой дисперсии является доверительный интервал:

$$n\hat{\sigma}^2_{\mathbf{y}}/\bar{u}^2_{\alpha,n-1}\leqslant x^2\leqslant n\hat{\sigma}^2_{\mathbf{y}}/\underline{u}^2_{\alpha,n-1}, \tag{6.12}$$

где числа $\underline{u}_{\alpha,n-1}$ и $\bar{u}_{\alpha,n-1}$ были определены выше, а $\hat{\sigma^2}_{\mathbf{y}}=\sum\limits_1^n(y_i-\hat{y})^2/n$.

6.3. ОЦЕНКА ПАРАМЕТРОВ РЕГРЕССИИ ПО ЭНЕРГЕТИЧЕСКИМ И КОРРЕЛЯЦИОННЫМ ДАННЫМ О ФЛУКТУАЦИЯХ

Обоснование. Наиболее обиходными, доступными сведениями о процессе помимо среднего являются его энергетические характеристики: средняя мощность, текущая мощность, а также корреляционные свойства, спектр. Причина их распространения в том, что верхняя граница средней мощности есть просто предел энергетических возможностей источника излучения. А корреляционные свойства обычно обязаны фильтру, будь он объект естественной природы как инерционность среды или искусственной в виде начальных каскадов приемника, который преодолевает процесс прежде, чем попасть на устройство обработки и принятия решений. Кстати, энергетические данные есть часть корреляционных, как и некоррелированность выборки, обычно достигаемая, если отсчеты процесса разнесены достаточно широко друг от друга.

Здесь строятся оптимальные доверительные оценки для регрессионных параметров, в частности, параметра сдвига, когда известны те или иные данные указанного сорта о флуктуациях. Оценки получаются расплывчатые, но никак не индикаторные, т. е. не в виде доверительных интервалов. Любопытно по ходу изложения проследить, как связывается расплывчатая форма этих оценок с видом первичных признаков, незримо присутствующих в энергетических, корреляционных и других исходных (первичных) данных, а также как по мере увеличения количества и улучшения качества (точности) этих данных улучшаются оценки, становясь более точными, более надежными.

Оценка параметра сдвига при заданной мощности флуктуаций. Пусть $y_i = x + \xi_i$, $i = 1, \ldots, n$, или в векторных обозначениях $\mathbf{y} = \mathbf{1}x + \xi$. Пусть единственное, что известно, так это $\overline{M}^{\xi}\|\xi\|^2/n = \overline{\sigma^2}$ — верхняя средняя по отсчетам (и по ансамблю) мощность флуктуаций. Тогда СИМ $\mathscr{M}^{x\mathbf{y}}$ будет определена единственным первичным средним, получаемым подстановкой $\xi_i = y_i - x$:

$$\overline{M}^{x\mathbf{y}} \sum (y_i - x)^2/n = \overline{\sigma^2}. \tag{6.13}$$

Здесь сумма (далее все суммы по i от 1 до n) составляет первичный признак. Ищем при этих данных оптимальную оценку параметра x. Достаточным для поставленной задачи согласно следствию 4 к теореме 6.1 будет следующий класс оценок:

$$\partial_{\mathbf{y}}(x) = [c_0 - c \sum (y_i - x)^2/n]^+, \quad 0 \leqslant c_0 \leqslant 1, \quad c > 0, \tag{6.14}$$

причем для них $\underline{M}\partial_{\mathbf{y}}(x) = [c_0 - c\overline{\sigma^2}]^+$.

Чтобы искать оптимальную оценку, вычислим штраф. Обратим внимание, что оценка (6.14) как функция x есть парабола, что видно, если переписать входящий в (6.14) первичный признак следующим образом: $\sum(y_i - x)^2/n = \hat{\sigma}^2_{\mathbf{y}} + (x - \hat{y})^2$, где $\hat{y} = \sum y_i/n$ — выборочное среднее, $\hat{\sigma}^2_{\mathbf{y}} = \sum(y_i - \hat{y})^2/n$ — выборочная дисперсия. Парабола усечена снизу осью абсцисс. Интегрируя ее, получаем

$$\Omega(\partial) = \int \partial_{\mathbf{y}}(x)\,dx = \int\limits_{\hat{y} - \sqrt{[c_0/c - \hat{\sigma}^2_{\mathbf{y}}]^+}}^{\hat{y} + \sqrt{[c_0/c - \hat{\sigma}^2_{\mathbf{y}}]^+}} [c_0 - c\hat{\sigma}^2_{\mathbf{y}} - c(x - \hat{y})^2]\,dx =$$

$$= \frac{4([c_0 - c\hat{\sigma}^2_{\mathbf{y}}]^+)^{3/2}}{3\sqrt{c}},$$

где пределы интегрирования соответствуют положительной части $\partial_{\mathbf{y}}(x)$. Осталось найти $\overline{M}^{\mathbf{y}}\Omega(\partial) = \frac{4}{3\sqrt{c}}\overline{M}^{\mathbf{y}}([c_0 - c\hat{\sigma}^2_{\mathbf{y}}]^+)^{3/2}$. Частная модель $\mathscr{M}^{\mathbf{y}}$ согласно теореме 2.1 основывается на первичном значении:

$$\overline{M}^{\mathbf{y}} \inf_x \sum (y_i - x)^2/n = \overline{M}^{\mathbf{y}}\hat{\sigma}^2_{\mathbf{y}} = \overline{\sigma^2}.$$

Здесь $\hat{\sigma}^2_{\mathbf{y}}$ есть первичный признак, а так как он всего один, то $\underline{M}^{\mathbf{y}}\hat{\sigma}^2_{\mathbf{y}} = \inf_{\mathbf{y}} \hat{\sigma}^2_{\mathbf{y}} = 0$, и поэтому

$$\overline{M}^{\mathbf{y}}([c_0 - c\hat{\sigma}^2_{\mathbf{y}}]^+)^{3/2} = \sup_{\mathbf{y}}([c_0 - c\hat{\sigma}^2_{\mathbf{y}}]^+)^{3/2} =$$

$$= (c_0 - c\inf_{\mathbf{y}}\hat{\sigma}^2_{\mathbf{y}})^{3/2} = c_0^{3/2}.$$

23

Осталось найти $\overline{M}^{x\mathbf{y}}\partial_{\mathbf{y}}(x)$. Правило $\partial_{\mathbf{y}}(x)$ не мажорируется первичным признаком (6.13), записанным $\hat{\sigma}^2_{\mathbf{y}} + (x - \hat{y})^2$, а он всего один, поэтому

$$\overline{M}^{x\mathbf{y}}\partial_{\mathbf{y}}(x) = \sup_{x,\mathbf{y}}[c_0 - c\sum(y_i - x)^2/n]^+ = c_0.$$

В результате $\alpha^{\varkappa}(\partial) = 1 - \varkappa\underline{M}\partial_{\mathbf{y}}(x) + (\varkappa - 1)\overline{M}\partial_{\mathbf{y}}(x) = 1 - \varkappa[c_0 - c\overline{\sigma^2}]^+ + (\varkappa - 1)c_0$.

Теперь записываем составной риск:

$$\Pi_{\lambda}(\partial) = 1 - \varkappa[c_0 - c\overline{\sigma^2}]^+ + (\varkappa - 1)c_0 + \lambda c_0\frac{4}{3}\sqrt{c_0/c}. \tag{6.15}$$

Здесь нужно рассматривать два случая: $c_0 - c\overline{\sigma^2} \geqslant 0$ и $c_0 - c\overline{\sigma^2} < 0$. Последний из них не представляет интереса, так как соответствует ошибке $\alpha^{\varkappa}(\partial) = (\varkappa - 1)c_0$, определяемой исключительно пессимизмом (а не знаниями о свойствах флуктуаций). Полагая $c_0 - c\overline{\sigma^2} \geqslant 0$, найдем сначала оптимальное значение c при заданном c_0. Для этого, дифференцируя правую часть составного риска по c и приравнивая 0, получаем

$$c = c_0(2\lambda/(3\varkappa\overline{\sigma^2}))^{2/3} \quad \text{и}$$

$$\min_c \Pi_{\lambda}(\partial) = 1 + c_0[(12\lambda^2\varkappa\overline{\sigma^2})^{1/3} - 1].$$

Отсюда видно, что оптимальное значение c_0 должно быть равным либо 0, либо 1. Поскольку неравенство $c_0 - c\overline{\sigma^2} \geqslant 0$ исключает значение $c_0 = 0$, полагаем $c_0 = 1$. Коэффициент c находится по заданной ошибке α из уравнения $\alpha = 1 - \varkappa + c\varkappa\overline{\sigma^2} + \varkappa - 1 = c\varkappa\overline{\sigma^2}$ и равен $c^* = \alpha/(\varkappa\overline{\sigma^2})$. Причем условие $c_0 - c^*\overline{\sigma^2} > 0$ соответствует $\alpha/\varkappa < 1$.

Таким образом, *при $\varkappa \geqslant 1$ оптимальная оценка имеет вид*

$$\partial^*_{\mathbf{y}}(x) = [1 - c^*\sum(y_i - x)^2/n]^+ = [1 - c^*\hat{\sigma}^2_{\mathbf{y}} - c^*(x - \hat{y})^2]^+, \quad c^* = \frac{\alpha}{\varkappa\overline{\sigma^2}}. \tag{6.16}$$

Как функция переменной x, есть парабола, нанесенная на рис. 6.2 штриховой линией, максимальная при значении x, равном выборочному среднему $\hat{y}$, и усеченная снизу осью абсцисс. Максимальное значение параболы $1 - c^*\hat{\sigma}^2$, не равно 1, что говорит о неконтрастности оценки, а размах параболы у основания, равный $2\sqrt{1/c^* - \hat{\sigma}^2_{\mathbf{y}}}$, характеризует ее расплывчатость. Здесь неприятным является влияние $\hat{\sigma}^2_{\mathbf{y}}$, от которого, как оказывается, нетрудно избавиться, что сейчас и будет рассмотрено.

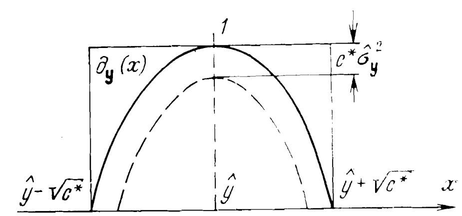


Рис. 6.2. Доверительная парабола при энергетических данных

…дем контрастную оценку (т. е. $\sup\limits_x \partial_y \equiv 1$, $\inf\limits_x \partial_y \equiv 0$) вида

$$\partial_y(x) = [1 - c^*(x - \hat{y})^2]^+, \quad c^* = \alpha/(\varkappa\overline{\sigma}^2). \tag{6.17}$$

Для нее $\underline{M}\,\partial_y(x) = \underline{M}\,[1 + c^*\hat{\sigma}_y^2 - c^*\Sigma(y_i - x)^2/n]^+ = 1 + c^*\underline{M}\hat{\sigma}_y^2 - c^*\overline{M}\,\Sigma\,(y_i - x)^2/n = 1 - c^*\overline{\sigma}^2 = 1 - \alpha/\varkappa$ и $\overline{M}\,\partial_y(x) = 1$, поэтому уровень $\alpha^\varkappa(\partial) = 1 - \varkappa\,\underline{M}\,\partial + (\varkappa - 1)\,\overline{M}\,\partial = \alpha$, а штраф $\overline{M}^y \int \partial_y(x)\,dx = M^y\, 4/(3\sqrt{c^*}) = 4/(3\sqrt{c^*})$.

Таким образом, и штраф, и уровень оценок (6.16) и (6.17) совпадают, *контрастная оценка* (6.17) *также является оптимальной*. Это парабола по переменной x, нанесенная сплошной линией на рис. 6.2.

Можно сказать, что любая оценка, заключенная в пределах между оценками (6.16) и (6.17)

$$[1 - c^*\hat{\sigma}_y^2 - c^*(x - \hat{y})^2]^+ \leqslant \partial_y(x) \leqslant [1 - c^*(x - \hat{y})^2]^+,$$

будет оптимальной. Причина такой неоднозначности оптимальной оценки объясняется тем, что как $\underline{M}\partial_y$, так и $\overline{M}\partial_y$, а следовательно, риск ориентирован на такой наименее благоприятный процесс, у которого $\hat{\sigma}^2{}_y = 0$, что означает постоянство реализаций флуктуаций $\xi_1 = \xi_2 = \ldots = \xi_n$, при этом оценки (6.16) и (6.17) совпадают. Это же и причина, по которой точность оценки, ее ширина не зависит от объема выборки n.

Оценки (6.16) и (6.17) называются оценками энергетического типа, так как используют только среднемощностные данные о флуктуациях. Сейчас мы выявим некоторые их дополнительные стороны и дадим обобщения.

Развитие энергетического типа оценивания. 1. Если помимо $\overline{\sigma}^2$ задана нижняя граница $\underline{\sigma}^2 = \underline{M}^\xi\|\xi\|^2/n$ мощности флуктуаций, то оптимальная оценка не изменится, так как новое среднее, взятое за первичное, не меняет ни частной модели $\mathscr{M}^y$, ни шкалы $\Omega(\partial)$, ни значений $\underline{M}\partial_y$, $\overline{M}\partial_y$.

2. Пусть помимо верхней мощности $\overline{\sigma}^2$ известно, что флуктуации имеют нулевые средние, а именно, либо $M\xi_i = 0$, $i = 1, \ldots, n$, либо $M\hat{\xi} = 0$. В обоих случаях в силу симметрии СИМ к перестановкам y_i оптимальный алгоритм будет принадлежать следующему достаточному классу:

$$\partial_y(x) = [c_0 + c_1\,\Sigma\,(y_i - x)/n - c\,\Sigma\,(y_i - x)^2/n]^+.$$

Слагаемое с коэффициентом c_1 даст лишь смещение в сторону по оси x оценки $\partial_y(x)$ как функции x, что, как нетрудно убедиться, приведет к увеличению риска, поэтому $c_1 = 0$. Частная модель $\mathscr{M}^y$ как при нулевых данных $M\xi_i = 0$, так и когда этих данных о $M\xi_i$ нет, одна и та же, поэтому штраф рассматриваемой оценки, а следовательно, и риск будут такими же, как у оценок (6.16), (6.17). Таким образом, *данные о нулевом среднем значении флуктуаций не меняют решения исходной задачи*: оптимальная оценка будет иметь вид (6.16) или (6.17). Сказанное переносится и на случай, когда заданы границы: а) $\underline{M}\xi_i = -m$, $\overline{M}\xi_i = m$, $i = 1, \ldots, n$, либо б) $\underline{M}\hat{\xi} = -m$, $\overline{M}\hat{\xi} = m$.

3. Оценки (6.16) и (6.17) «настроены» на максимальный штраф $\overline{M}\Omega(\partial)$ и соответствуют $\varkappa \geqslant 1$. Тем не менее они остаются оптимальными и при $\varkappa < 1$, а также при взвешенном (коэффициентом пессимизма) штрафе вида

$$M^\varkappa\,\Omega(\partial) = \varkappa\,\overline{M}\int\partial_y(x)\,dx + (1 - \varkappa)\,\underline{M}\int\partial_y(x)\,dx.$$

В самом деле, $\overline{M}\partial_y(x) = \sup\partial_y(x)$ и согласно утверждению 6.2 достаточным при $\varkappa < 1$ будет тот же класс (6.14) оценок. Полагая $c_0 = 1$ (оптимальное значение), нетрудно убедиться, что переход к взвешенному штрафу не может повлиять на вид оптимальной оценки, так как значение c однозначно определяется ошибкой α.

4. Если искать оптимальную оценку в классе интервальных, то она примет вид $\hat{y} \pm \overline{\sigma}\sqrt{\varkappa/\alpha}$ и получается заменой в (6.17) усеченной параболы на прямоугольник единичной высоты с тем же основанием, как это выглядит на рис. 6.2. Ошибка этой оценки будет также равна α, однако расплывчатость ее по сравнению с (6.17) возрастет за счет увеличения площади под ней.

5. Мы считали, что о связи x и ξ ничего не известно. То же самое согласно следствию 4 теоремы 6.1 будет, если считать, что в представлении $\mathbf{y} = \mathbf{1}x + \boldsymbol{\xi}$ о параметре x известно только, что он свободен от ξ, т. е. $\mathscr{M}^{x\xi} = \mathscr{M}^\xi\mathscr{I}^x$ и теперь уже x может произвольно подстраиваться под значения ξ. Можно показать, что допущение о свободе x от ξ не меняет оптимальной оценки.

6. Оценивание параметра регрессии. Пусть СИМ задана следующим образом: $y_i = w_i x + \xi_i$, $i = 1, \ldots, n$, $\overline{M}\|\xi\|^2/n = \overline{\sigma}^2$, где w_i — функция регрессии (скажем, форма детерминированного сигнала), x — параметр регрессии (амплитуда сигнала). Достаточный класс в этом случае образуют оценки вида

$$\partial_y(x) = [c_0 - c\,\|\mathbf{y} - \mathbf{w}\,x\|^2/n]^+, \quad \mathbf{w} = (w_1, \ldots, w_n)^{\mathrm{T}}.$$

Представим здесь квадрат нормы в виде $\|\mathbf{w}\|^2[\hat{\hat{\sigma}}^2{}_y - (x - \hat{y})^2]$ в обозначения $\hat{\hat{\mathbf{y}}} = \Sigma y_i w_i/\|\mathbf{w}\|^2$, $\hat{\hat{\sigma}}^2{}_y = \|\mathbf{y}\|^2/\|\mathbf{w}\|^2 - \hat{y}^2$ и запишем штраф (верхнюю среднюю ширину): $\lambda\sup\limits_{\mathbf{y}}(4c_0/3)\sqrt{c_0 n/(c\|\mathbf{w}\|^2) - \hat{\hat{\sigma}}^2{}_y}$. Супремум достигается при $\hat{\hat{\sigma}}^2{}_y = 0$, поэтому штраф равен $\lambda 4c_0\sqrt{c_0 n/c/}$

$(3\|\mathbf{w}\|)$. Из сравнения с (6.15) видим, что оптимальной должна быть оценка (6.17) (или (6.16)), если заменить в ней $\hat{y}$ на $\hat{\hat{y}}$ ($\hat{\sigma}_{\mathbf{y}}$ на $\hat{\hat{\sigma}}_{\mathbf{y}}$), а c^* — на $\alpha\|\mathbf{w}\|^2/(\varkappa\overline{\sigma}^2 n)$.

7. Оценивание векторного параметра регрессии. Пусть $y_i=\sum_{j=1}^{k}w_{ij}x_j+\xi_i$, $i=1, \ldots, n$, или в векторном виде $\mathbf{y}=\mathbf{W}\mathbf{x}+\xi$, и пусть $\widetilde{M}\|\xi\|^2/n=\overline{\sigma}^2$ определяет $\mathscr{M}^\xi$. Оптимальной уровня α (при $\varkappa\geqslant1$) будет оценка $[1-c^*\|\xi\|^2]^+$, $c^*=\alpha/(\varkappa n\overline{\sigma}^2)$, куда подставляется $\xi=\mathbf{y}-\mathbf{W}\mathbf{x}$.

Записанная оценка неконтрастна, так как $\max\limits_{\mathbf{x}}\|\mathbf{y}-\mathbf{W}\mathbf{x}\|=\|\mathbf{y}_\perp\|\geqslant0$, откуда $\max\limits_{\mathbf{x}}\partial_{\mathbf{y}}(\mathbf{x})\leqslant1$, где $\mathbf{y}_\perp$ — проекция $\mathbf{y}$ в подпространство, ортогональное вектор-столбцам матрицы $\mathbf{W}$. Разложим y на две ортогональные составляющие: $\mathbf{y}=\mathbf{y}_{\mathbf{W}}+\mathbf{y}_\perp$, где $\mathbf{y}_{\mathbf{W}}=\mathbf{W}(\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{y}$ — проекция $\mathbf{y}$ в подпространство вектор-столбцов $\mathbf{W}$. Замена в вышезаписанной оценке $\mathbf{y}$ на $\mathbf{y}_{\mathbf{W}}$ не изменяет $\overline{\alpha}$ (поскольку $\widetilde{M}\|\xi\|^2=\widetilde{M}\|\xi_{\mathbf{W}}\|^2$, где $\xi_{\mathbf{W}}$ — аналогичная проекция ξ), поэтому доверительная оценка многомерного параметра регрессии

$$\partial_{\mathbf{y}}(x)=[1-\alpha\,\|\mathbf{y}_{\mathbf{W}}-\mathbf{W}\,x\|^2/(\varkappa\, n\,\overline{\sigma}^2)]^+$$

будет оптимальной уровня α и одновременно контрастной.

В частном случае, когда параметр регрессии скалярный, результат, как нетрудно убедиться, совпадает с рассмотренным в предыдущем пункте.

8. Неоднородные флуктуации. Пусть СИМ задана следующим образом:

$$\mathbf{y}=\mathbf{W}\mathbf{x}+\xi,\qquad \widetilde{M}\,\|\mathbf{A}\,\xi\|^2=n\overline{\sigma}^2,$$

где $\mathbf{A}$ — матрица $n\times n$.

Совершая преобразование $\mathbf{z}=\mathbf{A}\mathbf{y}=\mathbf{A}\mathbf{W}\mathbf{x}+\mathbf{A}\xi=\mathbf{A}\mathbf{W}\mathbf{x}+\eta$, видим, что $\widetilde{M}\|\eta\|^2=n\overline{\sigma}^2$, и задача в новых наблюдениях $\mathbf{z}$ сводится к рассмотренной в предыдущем пункте с заменой $\mathbf{W}$ на $\mathbf{A}\mathbf{W}$.

В частном случае, когда заданы $\widetilde{M}\sum a^2{}_i\xi^2{}_i=n\overline{\sigma}^2$, матрица $\mathbf{A}$ становится диагональной с элементами a_i по главной диагонали. Такой результат легко переносится на процессы:

$$y_t=w_t\,x+\xi_t,\ t\in[0, T],\qquad \widetilde{M}\,\frac{1}{T}\int\limits_0^T a_t^2\,\xi_t^2\,dt=\overline{\sigma}^2,$$

где w_t и a_t — известные функции. Тогда для получения оптимальной оценки нужно в формулах (6.16), (6.17), $\hat{y}$, $\hat{\sigma}_{\mathbf{y}}$ и c^* заменить на

$$\hat{\hat{y}}_T=\frac{\int\limits_0^T y_t\,w_t\,a_t^2\,dt}{\|\mathbf{wa}\|^2},\quad \hat{\hat{\sigma}}{}^2_{\mathbf{y}}=\frac{\int\limits_0^T y_t^2\,a_t^2\,dt}{\|\mathbf{wa}\|^2}-\hat{\hat{y}}{}^2_T,\quad c^*=\frac{\alpha\,\|\mathbf{wa}\|^2}{\varkappa\,\overline{\sigma}^2\,T},$$

где $\|\mathbf{wa}\|^2=\int\limits_0^T w^2{}_t a^2{}_t dt$. Так, оценка типа (6.17) принимает вид

$$\partial_{\mathbf{y}}^*(x)=[1-\alpha\,\|\mathbf{wa}\|^2\,(x-\hat{\hat{y}}_T)^2/(\varkappa\overline{\sigma}^2\,T)]^+.$$

9. Пусть теперь исходным является не задание «мощности» флуктуаций, а среднее значение некоторого неотрицательного функционала от реализаций флуктуаций: $\widetilde{M}F\{\xi\}=\mu$. Тогда оценка будет иметь вид $\partial_{\mathbf{y}}(x)=[c_0-cF\{\mathbf{y}-\mathbf{W}\mathbf{x}\}]^+$, причем $\underline{M}\partial=[c_0-c\overline{\mu}]$. Рассмотрим два примера таких оценок.

Пример 6.6. Пусть x — параметр сдвига и заданным является среднее эффективное значение реализации флуктуаций $\widetilde{M}|\xi|=\mu$, $|\xi|=\Sigma|\xi_i|$. Тогда оптимальная оценка уровня α записывается: $\partial(\xi)=[c_0-c^*|\xi|]^+$, $c_0=1$, $c^*=\alpha/(\varkappa\mu)$. Сюда нужно подставить $\xi=\mathbf{y}-\mathbf{1}x$. Оценка как функция переменной x обретает вид линейно-ломаной с узлами в точках y_i. Форма записи оценки не изменится при заданной любой другой норме $|\xi|$.

Пример 6.7. Пусть единственным первичным значением является

$$\widetilde{M}\Sigma\{|\xi_i|>h\}/n=p_h,$$

т. е. исходным является задание верхнего среднего относительной частоты превышений абсолютными значениями ξ_i порога h. Тогда оптимальная оценка имеет вид

$$\partial_{\mathbf{y}}(x)=[c_0-c\,\Sigma\{|\xi_i|>h\}/n]^+,\quad c_0=c=\min\{1,\ (1-\alpha)/(1-\varkappa\,p_h)\}.$$

Оптимальная оценка параметра сдвига при однородных некоррелированных флуктуациях. Дополним сведения относительно флуктуаций, обогащая первичный набор новыми признаками, и посмотрим, как видоизменится оценка, какие новые свойства у нее появятся. Пусть $y_i=x+\xi_i$, $i=1, \ldots, n$, причем известно только, что ξ_i некоррелированы $M\xi_i\xi_j=0$, $i\neq j$, и имеют одинаковые «мощности» $\widetilde{M}\xi^2{}_i=\overline{\sigma}^2$. Это соответствует СИМ с первичными средними, получаемыми подстановкой $\xi_i=y_i-x$ и приобретающими вид:

$$M\,(y_i-x)\,(y_j-x)=0,\ i\neq j;\ \ \overline{M}\,(y_i-x)^2=\overline{\sigma}^2,$$

где i и j пробегают значения от 1 до n.

Поскольку СИМ симметрична к перестановкам y_i между собой, а группа перестановок дискретна, то достаточным будет класс инвариантных к перестановкам правил (см. утверждение 6.5). Вместе с теоремой 6.1 это позволяет получить следующий достаточный класс оценок:

$$\partial_{\mathbf{y}}(x)=[c_0-c_1\sum_{i\neq j}\sum(y_i-x)\,(y_j-x)/n^2-c_2\,\sum(y_i-x)^2/n]^+=$$

$$=[c_0-c_1(\hat{y}-x)^2+(c_1/n-c_2)\,(\hat{y}^2-2\,x\,\hat{y}+x^2)]^+,\ \hat{y}^2=\sum y_i^2/n,$$

и для них $\underline{M}\partial_{\mathbf{y}}(x)=[c_0-c_2\overline{\sigma}^2]^+$. Осталось найти коэффициенты c_0, c_1 и c_2, минимизирующие риск и удовлетворяющие ограничению $\partial_{\mathbf{y}}(x)\leqslant1$.

Будем искать $\Omega(\partial)$, для чего перепишем $\partial_{\mathbf{y}}(x)$ в виде:

$$\partial_{\mathbf{y}}(x) = [c_0 - c_1\hat{y}^2 + (c_1/n - c_2)\hat{y^2} +$$

$$+ 2x(c_1(n-1)/n + c_2)\hat{y} - x^2(c_1(n-1)/n + c_2)]^+ = [\hat{a}_0 + \hat{a}_1 x - a_2 x^2]^+,$$

где $\hat{a}_0 = c_0 - a_2\hat{y}^2 + (c_1/n - c_2)\hat{\sigma}^2_{\mathbf{y}}$, $\hat{a}_1 = 2a_2\hat{y}$, $a_2 = c_1(n-1)/n + c_2$.

Условие $\partial_{\mathbf{y}}(x) \leqslant 1$ соответствует двум неравенствам $a_2 \geqslant 0$ и $\sup_x \partial_{\mathbf{y}}(x) = \hat{a}_0 + \hat{a}^2_1/4a_2 \geqslant 1$. Так как $\partial_{\mathbf{y}}(x) \geqslant 0$ при $\hat{a}_0 + \hat{a}^2_1/2a_2 \geqslant 0$ и $|x - \hat{a}_1/2a_2| \leqslant \sqrt{[\hat{a}^2_1 + 4\hat{a}_0 a_2]^+/2a_2}$, то, проинтегрировав $\partial_{\mathbf{y}}(x)$ по x в пределах, соответствующих последнему неравенству, найдем

$$\Omega(\partial) = \int \partial_{\mathbf{y}}(x)\,dx = (2/3a_2)\sqrt{[\hat{a}_1^2 + 4\hat{a}_0 a_2]^+}\,[\hat{a}_0 + \hat{a}_1^2/4a_2]^+.$$

Учитывая теперь, что $\overline{M}\partial(x) = \overline{M}^{\mathbf{y}}[\hat{a}_0 + \hat{a}^2_1/4a_2]^+ = \sup_{\mathbf{y}}[\hat{a}_0 + \hat{a}^2_1/4a_2]^+$, сформируем выражение для составного риска:

$$\Pi^{\varkappa}_{\lambda}(\partial) = 1 - \varkappa[c_0 - c_2\overline{\sigma}^2]^+ + (\varkappa - 1)S + 4\lambda S^{3/2}/3\sqrt{a_2}.$$

Найдем в этом выражении $S = \sup_{\mathbf{y}}[\hat{a}_0 - \hat{a}^2_1/4a_2]^+ = c_0 + \inf_{\hat{y}} a_2\hat{y}^2 + \sup_{\hat{\sigma}^2_{\mathbf{y}}}(c_1/n - c_2)\hat{\sigma}^2_{\mathbf{y}} = c_0$. Отметим, что для справедливости последнего выражения должно выполняться $c_1/n \leqslant c_2$ (иначе супремум от $\hat{\sigma}^2_{\mathbf{y}}$ равен ∞). В результате

$$\Pi^{\varkappa}_{\lambda}(\partial) = 1 - \varkappa[c_0 - c_2\overline{\sigma}^2]^+ + (\varkappa - 1)c_0 + 4\lambda c_0^{3/2}/3\sqrt{a_2}.$$

Оптимальным c_1, минимизирующим риск при заданных c_0 и c_2, будет такое значение, которое максимизирует $a_2 = c_1(n-1)/n + c_2$ при условии $c_1/n \leqslant c_2$. Так как чем больше c_1, тем больше a_2 и тем меньше риск, то оптимальным будет наибольшее возможное значение: $c_1 = c_2 n$, откуда сразу же $a_2 = c_1$, и после подстановки этих значений оценка приобретает вид $\partial_{\mathbf{y}}(x) = [c_0 - c_1(\hat{y} - x)^2]^+$, а ее риск становится равным

$$\Pi^{\varkappa}_{\lambda}(\partial) = 1 - \varkappa[c_0 - c_1\overline{\sigma}^2/n]^+ + (\varkappa - 1)c_0 + \lambda c_0(4/3)\sqrt{c_0/c_1}.$$

Полученное выражение отличается от (6.15) лишь тем, что вместо $\overline{\sigma}^2$ здесь стоит $\overline{\sigma}^2/n$, поэтому оптимальная оценка будет иметь вид (6.17) с заменой $\overline{\sigma}^2$ на $\overline{\sigma}^2/k$, что мы и сформулируем как результат.

При однородной некоррелированной выборке и $\varkappa \geqslant 1$ оптимальная расплывчатая оценка уровня α имеет вид

$$\partial^*_{\mathbf{y}}(x) = [1 - c^*(x - \hat{y})^2]^+, \quad c^* = \alpha n/\varkappa\overline{\sigma}^2. \tag{6.18}$$

Иллюстрация ее дана на рис. 6.3. При увеличении n оценка становится все более узкой, менее расплывчатой, что очевидно должно быть, так как при получении $\hat{y}$ как суммы некоррелированные флуктуации складываются в беспорядке, стохастически, сглаживая в известном смысле друг друга, в результате разброс уменьшается, что позволяет для фиксации α сообразно этому

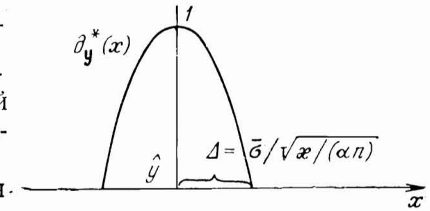


Рис. 6.3. Расплывчатая оценка при некоррелированных флуктуациях

сократить ширину. Причем такой же оценка будет, если дополнительно известно, что $M\xi_i = 0$, эти новые данные оказываются излишними.

Оценка (6.18) инвариантна к перестановкам наблюдений. Она же эквивариантна к преобразованию сдвига в том смысле, что одновременное прибавление ко всем y_i и к x любого числа a не меняет ее значения: $\partial_{\mathbf{y}}(x) \equiv \partial_{\mathbf{y}+\mathbf{1}a}(x+a)$. Последнее и позволяет исследовать (6.18) при $x = 0$, исследуя тем самым $\partial(\xi)$.

Оценивание сдвига при неоднородных некоррелированных флуктуациях. Пусть выборка не является однородной, но остается некоррелированной, т. е. $y_i = x + \xi_i$, $M\xi_i\xi_j = 0$, $i \neq j$, $\overline{M}\xi^2_i = \overline{\sigma}^2_i$, $i = 1, \ldots, n$. Определенная таким образом СИМ уже не будет симметричной к перестановкам. Достаточным будет следующий класс оценок:

$$\partial_{\mathbf{y}}(x) = [c_0 - \sum\sum c_{ij}(y_i - x)(y_j - x)]^+,$$

где $c_{ij} = c_{ji}$.

Перепишем: $\partial_{\mathbf{y}}(x) = [\hat{a}_0 + \hat{a}_1 x - a_2 x^2]^+$, где $\hat{a}_0 = c_0 - \sum\sum c_{ij}y_i y_j$, $\hat{a}_1 = \sum\sum c_{ij}(y_i + y_j) = 2\sum c_i y_i$, $c_i = \sum_j c_{ij}$, $a_2 = \sum\sum c_{ij}$. При введенных обозначениях штраф

$$\overline{M}^{\mathbf{y}}\int \partial_{\mathbf{y}}(x)\,dx = \frac{4\lambda}{3a_2^2}[\sup_{\mathbf{y}}(\hat{a}_0 a_2 + \hat{a}_1^2/4)^+]^{3/2}.$$

Ищем внутренний супремум в квадратных скобках:

$$\sup_{\mathbf{y}}(\hat{a}_0 a_2 + \hat{a}^2_1/4) = c_0 a_2 + \inf_{\mathbf{y}}[a_2\sum\sum c_{ij}y_i y_j - \sum\sum c_i c_j y_i y_j].$$

Чтобы инфимум не равнялся $-\infty$, матрица $\mathbf{Q}$ элементов $Q_{ij} = (\sum\sum c_{kl})c_{ij} - c_i c_j$, $i, j = 1, \ldots, n$, должна быть неотрицательно определенной, и тогда $\sup_{\mathbf{y}}(\hat{a}_0 a_2 + a^2_1/4) = c_0 a_2$. Далее $\underline{M}\partial_{\mathbf{y}}(x) = [c_0 - \sum_i c^+_{ii}\sigma^2_i]^+$, где c^+_{ii} есть положительные части c_{ii}. Наконец, $\overline{M}\partial_{\mathbf{y}}(x) = \sup_{x,\mathbf{y}}\partial_{\mathbf{y}}(x) = (1/a_2)\sup_{x,\mathbf{y}}(\hat{a}_0 a_2 + \hat{a}^2_1/4)^+ = c_0$, а $\partial_{\mathbf{y}}(x) \leqslant 1$ эквивалентно условию $c_0 \leqslant 1$. Используя найденные значения и полагая $c_0 - \sum c^+_{ii}\overline{\sigma}^2_i > 0$, запишем риск

$$\Pi^{\varkappa}_{\lambda}(\partial) = 1 - c_0 + \varkappa\sum c^+_{ii}\overline{\sigma}^2_i + \lambda\, 4\cdot c_0^{3/2}/3\sqrt{\sum\sum c_{ij}}.$$

Нужно минимизировать риск c_{ij} при условии неотрицательной определенности матрицы $\mathbf{Q}$. Разложим $\mathbf{Q}$ на сумму матриц ранга 1: $\mathbf{Q} = \sum \mathbf{b}_{(k)}\mathbf{b}^{\mathrm{T}}_{(k)}$, где $\mathbf{b}_{(k)} = (b_{k1}, \ldots, b_{kn})^{\mathrm{T}}$ — векторы. Тогда нужно будет минимизировать $\Pi_{\lambda}(\partial)$ независимым выбором векторов $\mathbf{b}_{(k)}$. Совершенно ясно, что все эти векторы све-

24

дутся к одному, так что $c_{ij}=b_i b_j$. Отсюда $\sum\sum c_{ij}=(\sum b_i)^2$ и $c_{ii}=b^2{}_i$. Оценка и риск запишутся так:

$$\partial_{\mathbf{y}}(x)=[c_0-(\sum b_i^2 (y_i-x))^2]^+,$$

$$\Pi_\lambda^\varkappa(\partial)=1-c_0+\varkappa\sum b_i^2\bar\sigma_i^2+\lambda\, 4\cdot c_0^{3/2}/3\sum b_i.$$

Осталось минимизировать риск по b_i и c_0. Дифференцируя по b_i и приравнивая 0, получаем $b_i=2\lambda c_0{}^{3/2}/[3\varkappa\bar\sigma^2{}_i(\sum b_i)^2]$. Суммируя по i, выводим уравнение, из которого находим сумму $\sum b_i=[2\lambda(\sum(1/\bar\sigma^2{}_i))/(3\varkappa)]^{1/3}c_0{}^{1/2}$. Подставляя эту сумму в выражение для b_i и далее в $\Pi^\varkappa{}_\lambda(\partial)$, имеем

$$b_i=c_0^{1/2}\left(\frac{2\lambda}{3\varkappa}\right)^{1/3}\Big/\left[\bar\sigma_i^2\left(\sum\left(1/\bar\sigma_i^2\right)\right)^{2/3}\right] \text{ и}$$

$$\Pi_\lambda(\partial)=1-c_0\left[1-(12\lambda^2\varkappa)^{1/3}/\left(\sum\left(1/\bar\sigma_i^2\right)\right)^{1/3}\right].$$

Очевидно, c_0 будет равно 1 или 0 в зависимости от того, больше или меньше 0 выражение в квадратных скобках. Нас интересует лишь первый случай, поэтому $c_0=1$ и тогда $b_i=\lambda/\bar\sigma^2{}_i$. Определяя λ по заданной ошибке $\alpha=\varkappa\sum b^2{}_i\bar\sigma^2{}_i=\varkappa\lambda\sum(1/\bar\sigma^2{}_i)$, находим $\lambda^2=\alpha/[\varkappa\sum(1/\bar\sigma^2{}_i)]$. Сформулируем окончательный результат.

Оптимальная (при $\varkappa\geqslant 1$) уровня α оценка параметра сдвига неоднородной некоррелированной выборки имеет вид

$$\partial_{\mathbf{y}}^*(x)=\left[1-\frac{\alpha}{\varkappa}\left(\sum\frac{1}{\bar\sigma_i^2}\right)\left(x-\frac{\sum y_i/\bar\sigma_i^2}{\sum 1/\bar\sigma_i^2}\right)^2\right]^+.$$

В обозначениях $\bar\sigma^2{}_{\text{ср}}=n/(\sum 1/\bar\sigma^2{}_i)$ и $\hat y=\bar\sigma^2{}_{\text{ср}}\sum(y_i/(n\bar\sigma^2{}_i))$ оптимальная оценка записывается

$$\partial_{\mathbf{y}}^*(x)=\left[1-\frac{\alpha n}{\varkappa\bar\sigma_{\text{ср}}^2}(x-\hat y)^2\right]^+. \tag{6.19}$$

Отметим, что при одинаковых дисперсиях $\bar\sigma_i=\bar\sigma$ имеем $\bar\sigma_{\text{ср}}=\bar\sigma$ и $\hat{\hat y}=\hat y$, поэтому из (6.19) частным случаем будет (6.17).

Обобщения оценок 1. Оценка параметра регрессии. Пусть $y_i=w_i x+\xi_i$, $M\xi_i\xi_j=0$, $i\neq j$, $\bar M\xi^2{}_i=\bar\sigma^2{}_i$, $i=1, \ldots, n$. Переходя к новым наблюдениям $z_i=y_i/w_i=x+(\xi_i/w_i)=x+\eta_i$, $\bar M\eta^2{}_i=\bar\sigma^2{}_i/w^2{}_i$, мы сводим задачу к (6.19), в которой y_i заменяются на z_i, а $\bar\sigma^2{}_i$ — на $\bar\sigma^2{}_i/w^2{}_i$. Получается следующая оптимальная оценка:

$$\partial_{\mathbf{y}}^*(x)=\left[1-\frac{\alpha}{\varkappa}\left(\sum\frac{w_i^2}{\bar\sigma_i^2}\right)\left(x-\frac{\sum y_i w_i/\bar\sigma_i^2}{\sum w_i/\bar\sigma_i^2}\right)^2\right]^+. \tag{6.20}$$

Отметим, во-первых, что отсчеты y_i индекса $i: w_i=0$ исключаются. Во-вторых, оценка эквивариантна к преобразованиям сдвига вида $\mathbf{s}(x, y)=(x+a, \mathbf{y}+\mathbf{w}a)$. В-третьих, оценка (6.20) базируется на $\hat y=(\sum y_i w_i/\bar\sigma^2{}_i)/(\sum w_i/\bar\sigma^2{}_i)$ — такой детерминированной оценке x, которая максимизирует в классе линейных оценок вида

$\sum c_i y_i$ отношение $(\sum c_i w_i)^2/(\sum c^2{}_i\bar\sigma^2{}_i)$ (интерпретируемое в задачах радиотехники как выходное после линейной обработки отношение энергии полезного сигнала к мощности шума).

2. Рассмотрим более общую задачу. Пусть известны корреляции флуктуаций: $y_i=w_i x+\xi_i$, $M\xi_i\xi_j=B_{ij}$, i, $j=1, \ldots, n$. Здесь элементы B_{ij} образуют неотрицательно определенную симметричную матрицу $\mathbf{B}$. Разлагая ее по собственным функциям, получим $\mathbf{B}=\mathbf{F}^\tau\mathbf{\Sigma}\mathbf{F}$, где $\mathbf{\Sigma}$ — диагональная матрица с элементами $\sigma^2{}_i$ по главной диагонали, а $\mathbf{F}$ — унитарная матрица F_{ij}: $\mathbf{F}^\tau=\mathbf{F}^{-1}$.

Переходя к векторам-столбцам $\mathbf{w}$ и $\boldsymbol{\xi}$ элементов w_i и ξ_i, запишем $\mathbf{y}=\mathbf{w}x+\boldsymbol{\xi}$, $\mathbf{B}=M\boldsymbol{\xi}\boldsymbol{\xi}^\tau$. Совершим преобразование $\mathbf{z}=\mathbf{F}\mathbf{y}=x\mathbf{F}\mathbf{w}+\mathbf{F}\boldsymbol{\xi}=x\check{\mathbf{w}}+\boldsymbol{\eta}$, где $\check w_i=\sum_j F_{ij}w_j$, $\eta_i=\sum_j F_{ij}\xi_j$. Нетрудно видеть, что $M\boldsymbol{\eta}\boldsymbol{\eta}^\tau=M\mathbf{F}\boldsymbol{\xi}\boldsymbol{\xi}^\tau\mathbf{F}^\tau=\mathbf{F}\mathbf{B}\mathbf{F}^\tau=\mathbf{F}\mathbf{F}^\tau\mathbf{\Sigma}\mathbf{F}\mathbf{F}^\tau=\Sigma$, так что η_i некоррелированы между собой и $M\eta^2{}_i=\sigma^2{}_i$. Таким образом, в наблюдениях z_i задача сводится к уже рассмотренной, только в формуле (6.20) нужно заменить y_i на z_i, w_i на $\check w_i$, а $\bar\sigma^2{}_i$ на $\sigma^2{}_i$. Так как

$$\sum\check w_i^2/\sigma_i^2=\check{\mathbf{w}}^\tau\,\Sigma^{-1}\,\check{\mathbf{w}}=\mathbf{w}^\tau\,\mathbf{F}^\tau\,\mathbf{F}\,\mathbf{B}^{-1}\,\mathbf{F}^\tau\,\mathbf{F}\mathbf{w}=\mathbf{w}^\tau\,\mathbf{B}^{-1}\,\mathbf{w},$$

$$\sum z_i\check w_i/\sigma_i^2=\mathbf{z}^\tau\,\Sigma^{-1}\,\check{\mathbf{w}}=\mathbf{y}^\tau\,\mathbf{F}^\tau\,\mathbf{F}\mathbf{B}^{-1}\,\mathbf{F}^\tau\,\mathbf{F}\mathbf{w}=\mathbf{y}^\tau\,\mathbf{B}^{-1}\,\underline{w},$$

то после соответствующей подстановки в (6.20) получаем оптимальную оценку

$$\partial_{\mathbf{y}}^*(x)=\left[1-\frac{\alpha}{\varkappa}\,\mathbf{w}^\tau\,\mathbf{B}^{-1}\,\mathbf{w}\,(x-\hat{\hat x}_{\mathbf{y}})^2\right]^+, \tag{6.21}$$

где $\hat{\hat x}_{\mathbf{y}}=\mathbf{y}^\tau\mathbf{B}^{-1}\mathbf{w}/\mathbf{w}^\tau\mathbf{B}^{-1}\mathbf{w}$ (полезно сравнить ее с (6.10)). Если учесть, что верно $M\hat{\hat x}^2{}_\xi=1/\mathbf{w}^\tau\mathbf{B}^{-1}\mathbf{w}$, то оценка (6.21) записывается иначе:

$$\partial_{\mathbf{y}}^*(x)=[1-(\alpha/\varkappa)\,(x-\hat{\hat x}_{\mathbf{y}})^2/M\,\hat{\hat x}_\xi^2]^+.$$

Отметим, что $\hat{\hat x}_{\mathbf{y}}$ при $M\xi_i=0$ есть детерминированная оценка x, минимизирующая в классе линейных оценок $\mathbf{c}^\tau\mathbf{y}$ отношение «сигнал-шум»: $(M\mathbf{c}^\tau\mathbf{y})^2/M(\mathbf{c}^\tau\mathbf{y})^2=(\mathbf{c}^\tau\mathbf{w})^2/(\mathbf{c}^\tau\mathbf{B}\mathbf{c})$.

3. Оценивание амплитуды детерминированного сигнала. Пусть требуется оценить x по данным

$$y_t=w_t\,x+\xi_t,\quad M\,\xi_t\,\xi_{t'}=B\,(t, t'),\quad t, t'\in[0, T],$$

где $B(t, t')$ — корреляционная функция шума, это есть ядро положительно определенного обратимого оператора. Оценка имеет точно приведенный в (6.21) вид с той лишь разницей, что квадратичные формы $\mathbf{w}^\tau\mathbf{B}^{-1}\mathbf{w}$ и $\mathbf{y}^\tau\mathbf{B}^{-1}\mathbf{w}$ заменяются на интегралы

$$\mathbf{w}^\tau\,\mathbf{B}^{-1}\,\mathbf{w}=\int_0^T w_t\,h_t\,dt,\qquad \mathbf{y}^\tau\,\mathbf{B}^{-1}\,\mathbf{w}=\int_0^T y_t\,h_t\,dt,$$

где h_t получается решением интегрального уравнения

$$\int_0^T B(t, t')\, h_t\, dt = w_{t'}.$$

Отметим, что точное знание корреляций есть некоторая идеализация реальных знаний, если ошибки невелики. В следующем разделе будет изучено влияние ошибок в знании корреляций на способы формирования оценок и оптимальный вид их.

Оценка амплитуды сигнала при колебаниях его формы и неточных корреляциях шума. Пусть $y_t = w_t x + \xi_t$, где $t = t_1, t_2, \ldots, t_n$ (или $t \in [0, T]$), w_t определяет форму сигнала, а x — его амплитуду. Пусть шум ξ_t имеет нулевое среднее $M\xi_t = 0$, а его корреляционная функция $B(t, \tau)$ неизвестна и лежит внутри заданного выпуклого ограниченного семейства $\mathfrak{B}$:

$$M\,\xi_t\,\xi_\tau = B(t, \tau) \in \mathfrak{B}, \text{ т.е. } \mathscr{M}^\xi = \bigvee_{\mathbf{B} \in \mathfrak{B}} \mathscr{M}^\xi_{\mathbf{B}}.$$

Будем искать оптимальную оценку x в достаточном для указанной задачи классе правил вида $\partial(y_t - w_t x)$. Запишем для этого риск

$$\Pi^\varkappa_\lambda(\partial) = 1 - \varkappa \underline{M}^\xi \partial(\xi) + (\varkappa - 1)\overline{M}^\xi \partial(\xi) + \lambda \overline{M}^\xi \sup_{x'} \int \partial(\xi + \mathbf{w}(x' - x))\, dx =$$

$$= 1 - \varkappa \inf_{\mathbf{B}} \underline{M}^\xi_{\mathbf{B}} \partial(\xi) + (\varkappa - 1) \sup_{\mathbf{B}} \overline{M}^\xi_{\mathbf{B}} \partial(\xi) + \lambda \sup_{\mathbf{B}} \overline{M}^\xi_{\mathbf{B}} \int \partial(\xi - \mathbf{w}x)\, dx =$$

$$= 1 - \varkappa \inf_{\mathbf{B}} \underline{M}^\xi_{\mathbf{B}} \partial(\xi) + (\varkappa - 1)\overline{\partial} + \lambda \sup_{\xi} \int \partial(\xi - \mathbf{w}x)\, dx,$$

где использовано равенство $\overline{M}^\xi_{\mathbf{B}} f(\xi) = \sup f(\xi) = \overline{f}$ и индекс t у ξ_t и w_t для краткости опущен. Задача синтеза оптимального правила обретает вид

$$\inf_\partial \sup_{\mathbf{B}} \left[1 - \varkappa \underline{M}^\xi_{\mathbf{B}} \partial(\xi) + (\varkappa - 1)\overline{\partial} + \lambda \sup_\xi \int \partial(\xi - \mathbf{w}\, x)\, dx\right].$$

Так как

$$\underline{M}^\xi_{\gamma \mathbf{B}_1 + (1-\gamma)\mathbf{B}_2} \partial(\xi) = \sup \{ \sum_t \sum_\tau [\gamma B_1(t, \tau) +$$

$$+ (1 - \gamma) B_2(t, \tau)]\, c_{t,\tau} : \sum_t c_t \xi_t + \sum_t \sum_\tau c_{t,\tau} \xi_t \xi_\tau \leqslant \partial(\xi)\} =$$

$$= \gamma \underline{M}^\xi_{\mathbf{B}_1} \partial(\xi) + (1 - \gamma) \underline{M}^\xi_{\mathbf{B}_2} \partial(\xi), \quad 0 \leqslant \gamma \leqslant 1,$$

то риск является вогнутой функцией $\mathbf{B}$. Учитывая выпуклость и ограниченность семейства $\mathfrak{B}$, мы можем инфимум и супремум поменять местами и сначала отыскать оценку при заданной корреляционной функции $B(t, \tau)$, обозначая риск $\Pi^\varkappa_{\lambda, \mathbf{B}}(\partial)$, а затем, максимизируя риск, определить наименее благоприятное $\mathbf{B}_{\text{н.б}}$ и подставить в оценку.

Оптимальная оценка, минимизирующая $\Pi^\varkappa_{\lambda, \mathbf{B}}(\partial)$, и риск для нее записываются в матричном виде (при непрерывном t легко переписывающимся в соответствующие интегральные аналоги):

$$\partial_{\mathbf{y}}(x) = \left[1 - \left(\frac{2\lambda}{3\varkappa}\, \mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{w}\right)^{2/3} \left(x - \frac{\mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{y}}{\mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{w}}\right)^2\right]^+,$$

$$\Pi^\varkappa_{\lambda, \mathbf{B}}(\partial) = (12\lambda^2 \varkappa)^{1/3} / (\mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{w})^{1/3},$$

где при некотором λ это — формула (6.21), и было сделано допущение, что $\Pi^\varkappa_{\lambda, \mathbf{B}}(\partial) \leqslant 1$. Теперь из выражения для риска видно, что *наименее благоприятной* $\mathbf{B}_{\text{н.б}}$ *будет корреляционная функция, минимизирующая квадратичную форму*:

$$\mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1}_{\text{н.б}} \mathbf{w} = \min_{\mathbf{B} \in \mathfrak{B}} \mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{w}.$$

Окончательно *оптимальная оценка уровня* α *примет вид* (6.21), *куда вместо* $\mathbf{B}$ *подставляется* $\mathbf{B}_{\text{н.б}}$. Рассмотрим примеры получения таких оценок для различных семейств $\mathfrak{B}$.

Пример 6.8. Пусть заданы границы корреляционной функции $\underline{B}(t, \tau)$, $\bar{B}(t, \tau)$, являющиеся неотрицательно определенными ядрами. Тогда наименее благоприятным будет значение $B_{\text{н.б}}(t, \tau) = \bar{B}(t, \tau)$, соответствующее максимальной мощности $\bar{B}(t, t) = \bar{b}^2{}_t$ шума и максимальным корреляциям.

В другом случае пусть $\mathfrak{B} = \{\mathbf{B} : \|\mathbf{B} - \mathbf{B}_0\| \leqslant \nu\}$, где норма $\|\mathbf{B}\|$ есть максимальное собственное число оператора с ядром $B(t, \tau)$. Здесь $B_0(t, \tau)$ есть некоторое предполагаемое (оценочное) значение корреляционной функции, а ν — величина ошибки. В этом случае

$$\min_{\|\mathbf{B} - \mathbf{B}_0\| \leqslant \nu} \mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{w} = \mathbf{w}^{\mathrm{T}} (\mathbf{B}_0 + \nu \mathbf{I})^{-1} \mathbf{w},$$

где $\mathbf{I}$ — единичный оператор, соответствующий ядру в виде дельта-функции Дирака. Ядро $\nu\mathbf{I}$ интерпретируется как корреляционная функция «белого шума» спектральной интенсивности ν. Поэтому сумма $\mathbf{B}_0 + \nu\mathbf{I}$ равносильна добавлению к флуктуациям «белого шума». Таким образом, *ошибка в знании корреляционной функции флуктуаций компенсируется прибавлением добавки в виде «белого шума»* (подобный вывод, если вспомнить, мы уже получали в задаче фильтрации конца § 5.5).

Теперь обсудим тот случай, когда неточно известна функция w_t, определяющая форму полезного сигнала. Оценка (6.21) уже не будет в этом случае оптимальной, так как она призывает к линейной обработке наблюдений y_t, тогда как при неточно известном сигнале не исключено, что следует перейти к нелинейной обработке. По крайней мере, оценки вида $[c_0 - \sum\sum c_{ij}(y_i - x) \times \times (y_j - x)]^+$ уже не образуют достаточного класса. Тем не менее, если изменения сигнала малы по сравнению с некоторым гипотетическим значением $w^0{}_t$, то все же можно ограничиться линейными оценками, подбирая в $\hat{x}_{\mathbf{y}} = \mathbf{c}^{\mathrm{T}}\mathbf{y}$ значение $\mathbf{c}$ таким образом, чтобы максимизировать отношение сигнал-шум на выходе фильтра линейной обработки при наименее благоприятном отклике w_t, т. е. решая задачу

$$\max_{\mathbf{c}} \min_{\mathbf{w} \in \mathscr{W}} (\mathbf{c}^{\mathrm{T}} \mathbf{w})^2 / \mathbf{c}^{\mathrm{T}} \mathbf{B} \mathbf{c} \text{ при } \mathbf{c}^{\mathrm{T}} \mathbf{w} = 1.$$

Пример 6.9. Пусть w_t может колебаться около значения $w^\circ{}_t$, так что $\|\mathbf{w} - \mathbf{w}^\circ\|^2 = \sum_t (w_t - w^\circ{}_t)^2 \leqslant \Delta^2$. Тогда

$$\min_{\mathbf{w} \in \mathscr{W}} (\mathbf{c}^{\mathrm{T}} \mathbf{w})^2 = \min_{\|\mathbf{w} - \mathbf{w}^\circ\| \leqslant \Delta} (\sum c_t w_t)^2 = \left(\sum c_t \overset{\circ}{w}_t - \|\mathbf{c}\| \Delta\right)^2 = (\mathbf{c}^{\mathrm{T}} \mathbf{w}^\circ - \|\mathbf{c}\|\, \Delta)^2,$$

где минимум достигается при $w_t = w^\circ_t - c_t\Delta/\|\mathbf{c}\|$. Максимизация по $\mathbf{c}$ отношения сигнал-шум $(\mathbf{c}^\mathrm{T}\mathbf{w}^\circ - \|\mathbf{c}\|\Delta)^2/\mathbf{c}^\mathrm{T}\mathbf{B}\mathbf{c}$ ведет к значению

$$\mathbf{c}^* = (\mathbf{B} - \nu\,\mathbf{I})^{-1}\,\mathbf{w}/\mathbf{w}^\mathrm{T}\,(\mathbf{B} - \nu\,\mathbf{I})^{-1}\,\mathbf{w},$$

где ν находится как решение уравнения $\nu^2\mathbf{w}^\mathrm{T}(\mathbf{B}+\nu\mathbf{I})^{-2}\mathbf{w} = \Delta^2$.

При непрерывном времени t суммы заменяются на интегралы, а матричные выражения — на интегральные аналоги.

Таким образом, *наличие колебаний сигнала в случае линейной обработки компенсируется введением дополнительного аддитивного «белого шума» спектральной интенсивности* ν[1]. Полезно провести сравнение с предыдущим примером, где такая же добавка являлась следствием неточного знания корреляционных свойств шума.

6.4. ОЦЕНИВАНИЕ ПАРАМЕТРА СДВИГА ПО МОМЕНТАМ И ГАРМОНИЧЕСКИМ СРЕДНИМ

Оценивание по моментам. Какова будет оценка параметра x по наблюдениям $y_i = x + \xi_i$, $i = 1, \ldots, n$, когда известны моменты более высоких, чем первого и второго порядков? Тогда вид оценки, в общем, будет отличаться от усеченной снизу параболы, каковой она являлась выше. Начнем с примера.

Пример 6.10. Допустим, что исходными для ξ являются: $M\xi^3_i\xi^3_j = 0$, $i \neq j$, $\overline{M}\xi^6_i = \overline{m}_6$. Тогда при $\varkappa \geqslant 1$ достаточным будет следующий класс оценок: $\partial(\xi) = [c_0 - c_1\sum\sum_{i\neq j}\xi^3_i\xi^3_j - c_2\sum\xi^6_i]^+$ (куда для получения $\partial_\mathbf{y}(x)$ нужно подставить $\xi_i = y_i - x$). В частности, к этому классу относится оценка вида

$$\partial(\xi) = [1 - (\widehat{\xi^3}/\Delta_n)^2]^+,\ \widehat{\xi^3} = \sum\xi_i^3/n,$$

которая заменой $\xi^3_i = \eta_i$ выводится из (6.18), причем $M\eta_i = 0$, $\overline{M}\eta^2_i = \overline{m}_6$, поэтому для фиксации уровня должно быть $\Delta_n = \sqrt{\varkappa\overline{m}_6/(\alpha n)}$. Если в той же самой задаче дополнительно известны промежуточные моменты (вплоть до шестого), то, как можно будет видеть из сравнения со следующим примером, более хорошими могут оказаться оценки вида $[1 - (\hat{\xi}/\Delta)^6]^+$, так что форма оценок во многом определяется видом исходных (первичных) признаков.

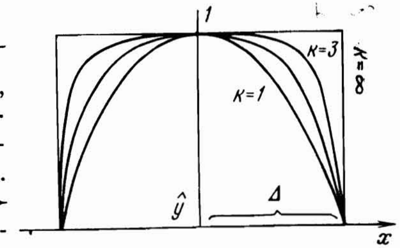
Перейдем от примера к общему случаю. Пусть ограниченными являются моменты вплоть до порядка $2k$. Зададимся оценками вида

$$\partial(\xi) = [1 - (\hat{\xi}/\Delta_n)^{2k}]^+,\ k = 1, 2, \ldots, \tag{6.22}$$

и назовем их *оценками степенного типа*. Как функция переменной x оценка степенного вида $\partial(\mathbf{y} - \mathbf{1}x)$ при $k > 1$ уже не перевернутая

[1] Впервые показано в статье: Кузнецов В. П. О синтезе линейных обнаружителей при неточно заданном сигнале и неполностью известных свойствах нормального шума//Радиотехника и электроника. — 1974. — № 12. — С. 2529—2538.

Рис. 6.4. Оценки степенного типа



усеченная снизу парабола, а кривая, более тупая, чем парабола у вершины, и более круто спадающая у основания, как это показано на рис. 6.4, трансформирующаяся при $k \to \infty$ к индикаторной форме, т. е. доверительному интервалу.

Величина $2\Delta_n$ — *размах* оценки у основания. Очевидно $\underline{M}\partial(\xi) \geqslant 1 - \overline{M}\hat{\xi}^{2k}/\Delta^{2k}_n$ (так как $\partial(\mathbf{z}) \geqslant 1 - (\hat{z}/\Delta_n)^{2k}$), $\overline{M}\partial(\xi) \leqslant 1$ (так как $\partial(\mathbf{z}) \leqslant 1$), откуда при $\varkappa \geqslant 1$ имеем

$$\alpha^\varkappa(\partial) = 1 - \varkappa\,\underline{M}\,\partial + (\varkappa - 1)\,\overline{M}\,\partial \leqslant \varkappa\,\overline{M}\,\hat{\xi}^{2k}/\Delta_n^{2k}.$$

Теперь, чтобы уровень $\alpha^\varkappa(\partial)$ оценки был не больше α, нужно брать $\Delta_n \leqslant (\varkappa\overline{M}\hat{\xi}^{2k}/\alpha)^{1/2k}$. Для определения размаха оценки Δ_n нужно знать $\overline{M}\hat{\xi}^{2k}$. В свою очередь, для нахождения $\overline{M}\hat{\xi}^{2k}$, как это следует из неравенства $\overline{M}\hat{\xi}^{2k} \leqslant \sum\ldots\sum|\overline{M}|\,\xi_{i_1}\ldots\xi_{i_k}/n^{2k}$, достаточно знать смешанные моменты вплоть до порядка $2k$.

Чтобы получить содержательные результаты, нужно сделать предположения относительно существования моментов. Характер этих предположений и их влияния на размах оценки установим сначала на примере.

Пример 6.11. Пусть $k = 2$. Тогда для независимой симметричной последовательности с конечным четвертым моментом:

$$|\overline{M}|\ \xi_{i_1}\,\xi_{i_2}\,\xi_{i_3}\,\xi_{i_4} = \begin{cases} 0 & \text{при}\quad i_1 \neq i_2,\quad i_1 \neq i_3,\quad i_1 \neq i_4, \\ \overline{\sigma}^4 & \text{при}\quad i_1 = i_2 \neq i_3 = i_4, \\ \overline{m}_4 & \text{при}\quad i_i = i_2 = i_3 = i_4, \end{cases}$$

где $\overline{\sigma}^2 = \overline{M}\xi^2_i$, $\overline{m}_4 = \overline{M}\xi^4_i$. Отсюда имеем

$$\overline{M}\,\hat{\xi}^4 \leqslant \frac{\overline{m}_4}{n} + \frac{3\,\overline{\sigma}^4\,(n-1)}{n^3},\quad \Delta_n \leqslant \frac{1}{\sqrt{n}}\left[\frac{\varkappa}{\alpha}\left(\frac{\overline{m}_4}{n} + \frac{3\,\overline{\sigma}^4\,(n-1)}{n}\right)\right]^{1/4}.$$

Аналогично, если $k = 3$, то

$$\overline{M}\,\hat{\xi}^6 \leqslant \frac{1}{n^6}\left(n\,\overline{m}_6 + C_6^2\,n\,(n-1)\,\overline{m}_4\,\overline{\sigma}^2 + n\,(n-1)\,(n-2)\,\overline{\sigma}^6\,\frac{6!}{(2!)^3\,3!}\right),$$

$$\Delta_n \leqslant \frac{1}{\sqrt{n}}\left[\frac{\varkappa}{\alpha}\left(\frac{\overline{m}_6}{n^2} + C_6^2\frac{n-1}{n^2}\overline{m}_4\,\overline{\sigma}^2 + 15\,\frac{(n-1)\,(n-2)}{n^2}\,\overline{\sigma}^6\right)\right]^{1/6}.$$

Здесь видно, что чем больше n, тем в меньшей мере влияют величины старших моментов на ширину оценки (6.22). При $n \to \infty$ существенным оказывается лишь значение второго момента $\overline{\sigma}^2$.

Сделаем предположения. Пусть k — некоторое целое число и обратимся к оценке степенного типа (6.22). Считаем ξ_i однородной последовательностью симметричных с.в., у которой $\overline{m}_{2k}=\overline{M}\xi_i^{2k}<\infty$. Для нее $M\xi_i^{2j-1}=0$ при $j=1, \ldots, k$, и $\overline{m}_{2j}$ конечны при $j\leqslant k$. Нужны также предположения относительно некоррелированности между собой степеней отсчетов, а именно, считаем $|\overline{M}|\xi_1^{j1}\xi_2^{j2}\ldots\xi_n^{jn}=0$, если нечетно хотя бы одно j_i, где j_i принимают целые значения от 0 до $\sum j_i\leqslant 2k$ (требуемая некоррелированность будет, если ξ_i независимы и симметричны). Тогда для расчета $\overline{M}\hat{\xi}^{2k}$, а отсюда и для расчета ошибки $\alpha^{\varkappa}(\partial)=\varkappa\overline{M}\hat{\xi}^{2k}/\Delta^{2k}{}_n$ оценки (6.22) уместно воспользоваться допредельной формулой (3.11), а в асимптотике при $n\to\infty$ — утверждением 3.11, согласно которому $\lim\limits_{n\to\infty}\overline{M}\hat{\xi}^{2k}n^k\leqslant\overline{\sigma}^{2k}(2k)!/(k!2^k)$.

Подставляя правую часть в ошибку и приравнивая последнюю α, находим асимптотическое значение размаха Δ_n оценки степенного типа, нормированное множителем $\sqrt{n}$:

$$\lim_{n\to\infty}\Delta_n\sqrt{n}=\frac{\overline{\sigma}}{\sqrt{2}}\left[\frac{\varkappa(2k)!}{\alpha k!}\right]^{1/2k}. \tag{6.23}$$

Обратим внимание, что при $n\to\infty$ размах Δ_n уменьшается пропорционально $1/\sqrt{n}$, так как правая часть от n не зависит. Как функция числа k правая часть имеет явно выраженный минимум, дающий наименьший размах оценки при данном уровне α. Это рождает задачу асимптотически (при $n\to\infty$) оптимального выбора числа k, т. е. подстройке формы оценки по параметру k в рамках структуры степенного типа (6.22), к чему и перейдем.

Асимптотическая подстройка оценки степенного типа. Считаем ξ_i симметричными, ограниченными (поэтому существуют моменты любых порядков) и независимыми. В результате при любом целом k будут выполнены предпосылки формулы (6.23), дающей асимптотическое при $n\to\infty$ выражение для размаха оценки степенного типа.

Ширина оценки (6.22) получается интегрированием

$$\Omega_n(\partial)=\int\limits_{-\Delta_n}^{\Delta_n}[1-(x/\Delta_n)^{2k}]\,dx=2\Delta_n\frac{2k}{2k+1}.$$

Подстановка сюда значения Δ_n из (6.23), обеспечивающего по крайней мере асимптотически уровень α, приводит к выражению для *асимптотически нормированной ширины*

$$\Omega_*=\lim_{n\to\infty}\Omega_n(\partial)\sqrt{n}=\sqrt{2}\,\overline{\sigma}\left[\frac{\varkappa(2k)!}{\alpha k!}\right]^{1/2k}\frac{2k}{2k+1}.$$

Поставим задачу выбора такого k, при котором достигается минимум правой части. Ее решение — сфера численных методов. Но на качественном уровне вполне приемлемы и аналитические, воспользоваться которыми позволяет формула Стирлинга:

$k!\approx(k/e)^k\sqrt{2\pi k}$. Подстановка ее на место факториалов в Ω_* дает:

$$\Omega_*=2\overline{\sigma}(\sqrt{2}\varkappa/\alpha)^{1/2k}\sqrt{2k/e}\frac{2k}{2k+1},$$

а минимизация по k ведет к уравнению $l_\alpha=k(2k+3)/(2k+1)$, где $l_\alpha=\ln(\sqrt{2}\varkappa/\alpha)$, решение которого есть $k^*=l_\alpha/2-3/4+\sqrt{l^2_\alpha-3l_\alpha/2+9/2}$. Нужно взять ближайшее целое к нему значение $[k^*]$, подстановка которого в Ω_* даст асимптотически нормированную ширину оценки (6.22). Ее приведенное (к единичной дисперсии) значение $\Omega_*/\overline{\sigma}$ при $\varkappa=1$ сведено в таблицу для сравнения с приведенной шириной $2\Phi^{-1}(1/2-\alpha/2)$ доверительного интервала нормальной выборки примера 6.5. При составлении таблицы брались по очереди целые значения $k^*=1, 2, \ldots$, по ним находились l_α, затем α (оптимальные уровни, соответствующие заданному k^*) и вычислялись $\Omega_*/\overline{\sigma}$, а параллельно для сравнения находились $2\Phi^{-1}(1/2-\alpha/2)$.

Степень оценки	1	2	3	4	5	6
Оптимальный уровень	0,267	0,086	0,03	0,01	$3{,}84\cdot10^{-3}$	$1{,}4\cdot10^{-3}$
Расплывчатость $\Omega_*/\overline{\sigma}$	2,62	3,9	4,84	5,62	6,3	6,9
$2\Phi^{-1}(1/2-\alpha/2)$	2,22	3,44	4,36	5,08	5,76	6,4

Мысль, проводимая нами и подкрепленная таблицей, вот в чем. При увеличении n нормированная сумма $\sum\xi_i/\sqrt{n}$ ограниченной симметричной независимой последовательности с.в. в соответствии с центральной предельной теоремой 3.16 должна сходиться к нормальному закону, это факт. Но тогда следует ожидать, что ширина оценки (6.22) тоже должна приближаться к ширине доверительного интервала для нормальной последовательности ξ_i (сумма которых, как известно, тоже нормальна). Эти ожидания, как явствует из таблицы, полностью оправдываются. Оправдываются и доказывают принципиальную возможность получения качественных оценок при крайне скудных знаниях о флуктуациях ξ_i. Отсутствие же полного совпадения объясняется тем, что из предельных свойств суммы $\sum\xi_i$ «выхвачены» только моменты, да и сама оценка по постановке заключена внутри форм степенного типа. Еще важно, что наши оценки «впитывают» в себя все имеющиеся и следующие из них данные, даже побочные.

Использование допредельных и предельных результатов. Приступим к более методичному привлечению допредельных и предельных результатов гл. 3 для синтеза оценок. В предыдущих двух разделах рассматривались только степенные признаки сумм и, как следствие, моменты. В то же время не использованы данные о гармонических признаках, возникающие при суммировании независимых с. в. согласно исследованиям гл. 3.

Прежде чем высказать соображения на этот счет, сделаем небольшую остановку, чтобы резюмировать характерные свойства полученных выше оценок параметра сдвига x. Во-первых, оценки $\partial_{\mathbf{y}}(x)$ являются функциями $\xi_i = y_i - x$, что равносильно их эквивариантности к преобразованиям сдвига: $\partial_{\mathbf{y}+\mathbf{1}a}(x+a) = \partial_{\mathbf{y}}(x) = \partial(\mathbf{y} - \mathbf{1}x) = \partial(\boldsymbol{\xi})$. Первопричина такого свойства в том, что любые значения x находятся в одинаковом по привилегиям априорном положении, если данных об x нет (данные, скажем, о диапазоне значений x умышленно можно игнорировать с целью достичь упрощений) и если шкала расплывчатости не взвешена. Во-вторых, для однородной выборки $y_i = x + \xi_i$ наблюдений оценка является функцией суммы $\sum \xi_i$ как последствие симметрии задачи к перестановкам наблюдений, в результате все наблюдения оказываются одинаково равноценными участниками в поиске скрытого x.

И, наконец, самое важное в настоящий момент, что оценки являются функциями не просто суммы, а *нормированной суммы* ζ_n. Это либо среднее арифметическое $\hat{\xi}$ в оценке (6.17), когда данных о корреляциях ξ_i нет, либо нормированная сумма $\hat{\xi}\sqrt{n} = \sum \xi_i / \sqrt{n}$, если ξ_i некоррелированная или независимая однородная последовательность. Последнее легко вытекает из (6.22), (6.23), если учесть, что оценка зависит от отношения $\hat{\xi}/\Delta_n$, а размах Δ_n пропорционален $1/\sqrt{n}$.

В общем, оценки записываются: $\partial_{\mathbf{y}}(x) = v(\zeta_n)$, где $v(z)$ есть некоторая унимодальная симметричная функция z, равная 1 при $z=0$ и спадающая к 0 при отклонении z от 0. Интегральная ширина функции $v(z)$, а для независимых флуктуаций с поправкой на множитель $1/\sqrt{n}$ дает расплывчатость оценки. В упомянутом случае $\zeta_n = \hat{\xi}\sqrt{n} = (\hat{y} - x)\sqrt{n}$ и

$$\int \partial_{\mathbf{y}}(x)\, dx = \int v((\hat{y} - x)\sqrt{n})\, dx = \int v(z)\, dz / \sqrt{n}.$$

В аргументе функции $v(\zeta_n)$ стоит нормированная сумма ζ_n, в поведении которой выделяется то, что ζ_n при увеличении n стабилизируется как случайная величина, т. е. не вырождается ни к нулю, ни к бесконечности. Удобно представлять себе с. в. ζ_n, как будто совершающую ограниченные колебания в диапазоне, меняющемся при увеличении n несущественно, в то время, как среднестатистические данные об ζ_n претерпевают изменения, да еще какие! Базу исследованиям дает гл. 3, где показано, что нормированная сумма сообразно исходным данным стремится к разным по ширине моделям, первичными для которых являются

степенные и гармонические признаки (вплоть до нормальной и интервальной нормальной моделей). Здесь интересно напомнить, что сведения о степенных и гармонических признаках сумм ζ_n возникают и уточняются даже тогда, когда похожих данных о слагаемых вроде бы и не было.

Сказанное наводит на мысль для независимых флуктуаций ξ_i при расчете ошибки оценки $\partial(\boldsymbol{\xi}) = v(\zeta_n)$, $\zeta_n = \sum \xi_i / \sqrt{n}$, и синтезе оптимальной оценки как можно шире пользоваться данными о ζ_n, предоставляемыми допредельными и предельными теоремами в виде средних степенных $\overline{M}\zeta_n^k$ и гармонических $\overline{M}\sin u\zeta_n$, $\overline{M}\cos u\zeta_n$ признаков, взятых за первичные для ζ_n. При этом вроде бы «забываются» исходные данные о флуктуациях после того, как они были использованы для получения необходимых сведений относительно ζ_n.

Теперь расчет ошибки будет осуществляться так: для вычисления $\overline{M}\partial_{\mathbf{y}}(x) = \overline{M}v(\zeta_n)$ функция $v(z)$ мажорируется полиномами, смесью гармоник или теми, и другими, а при расчете $\underline{M}\partial_{\mathbf{y}}(x)$ (соответствующей верхней ошибке), наоборот, мажорирует их в соответствии с общим принципом продолжения средних. А если вспомнить принцип достаточности, формулируемый теоремой 6.1, то $v(z)$ и вовсе сама должна записываться в виде степенного ряда, гармонического ряда Фурье или получаться смешением того и другого. Так, в общем-то, и делалось в предыдущем разделе, но только для степенных признаков. Нам же предстоит это проделать еще и для гармонических.

Замечания. 1. Нужно отметить, что оценивание по допредельным неравенствам и предельным результатам нельзя назвать совсем оптимальными хотя бы потому, что мы «забываем» данные о ξ_i и пользуемся только тем, что узнаем по ним о ζ_n. И далее, хотя по данным о ζ_n оценка строится наилучшим образом, но такие оценки все же точнее назвать квазиоптимальными, а при $n \to \infty$ — *асимптотически оптимальными*.

2. Нормироваться сумма $\sum \xi_i$ в оценке $v(\zeta_n)$ при независимых флуктуациях не обязательно должна коэффициентом $1/\sqrt{n}$, это может быть и $1/n$, если данных об отдельных слагаемых почти нет.

Синтез квазиоптимальных оценок по гармоническим средним. Пусть ξ_i, $i = 1, 2, \ldots, n$, — однородная последовательность независимых с. в. и пусть $\partial(\boldsymbol{\xi}) = v(\zeta_n)$ есть оценка параметра сдвига, где $\zeta_n = \sum \xi_i / \sqrt{n}$ — нормированная сумма. Функцию $v(z)$, считая ее симметричной по отношению к началу координат, запишем через преобразование Фурье $V(u)$:

$$v(z) = \int_{-\infty}^{\infty} V(u) \cos(uz)\, du, \qquad V(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} v(z) \cos(uz)\, dz.$$

Это прямая и обратная формулы преобразования Фурье.

Раскладывая $V(u)=V^+(u)-V^-(u)$ на положительную и отрицательную части, имеем

$$\underline{M}\, v(\zeta_n) \geqslant \int\limits_U V^+(u)\, \underline{M} \cos(u\zeta_n)\, du - \int\limits_U V^-(u)\, \overline{M} \cos(u\zeta_n)\, du.$$

Мы видим, что ошибка $\overline{\alpha}(\partial)=1-\underline{M}v(\zeta_n)$ оценки выражается через гармонические средние $\underline{M}\cos(\overline{u\zeta_n})$ и $\overline{M}\cos(u\zeta_n)$ нормированных сумм. Данные гармонических средних доставляются допредельными и предельными теоремами главы 3, в общем, не для всех u, а для некоторой области U, учтенной при интегрировании.

Для простоты конкретизируем задачу. Пусть ξ_i независимы, имеют нулевые средние $M\xi_i=0$ и конечные дисперсии $\overline{M}\xi^2_i=\overline{\sigma}^2$. При этих данных последовательность ξ_i, в общем, статистически неустойчива, сходимости ζ_n к нормальной модели не будет. Тем не менее значения $\underline{M}\cos(u\zeta_n)$ могли бы быть получены из допредельной теоремы дополнения к § 3.3, которые и позволили бы в принципе рассчитать ошибку $\overline{\alpha}(\partial)$ при конечном n (и следовательно, взвешенную ошибку, так как $\underline{\alpha}(\partial)=0$). Мы не будем выписывать все громоздкие следующие этому пути формулы хотя бы потому, что они принципиально мало чем отличаются от более простых вытекающих из них при $n\to\infty$ асимптотических формул, тех, что даются следствием 1 теоремы 3.13, на которых и остановимся.

Последние предоопределяют такие асимптотические данные о ζ_n:

$$M\,\zeta_n=0,\ \overline{M}\,\zeta_n^2=\overline{\sigma}^2,\ \underset{\sim}{M}\cos(u\,\zeta_n)=\exp(-u^2\overline{\sigma}^2/2)\cos(c^*u^2\overline{\sigma}^2), \quad (6.24)$$

где $c^*=0{,}3184$, а $|u|^2\leqslant\pi/(2c^*\overline{\sigma}^2)$. Это неравенство для u и определяет область U, так что если $v(z)$ такова, что $V(u)\geqslant 0$ для всех u, то имеем асимптотическое значение

$$\lim_{n\to\infty} \underline{M}\, v(\zeta_n) \geqslant 2 \int\limits_0^{\pi/(2c^*\overline{\sigma}^2)} V(u)\exp(-u^2\overline{\sigma}^2/2)\cos(c^*u^2\overline{\sigma}^2)\, du,$$

где при выставлении пределов использована симметрия подынтегрального выражения.

Замечание. Гармонические средние в (6.22) может быть не столь значительно, но все же уточняют модель суммы по сравнению с $\overline{M}\zeta^2_n=\overline{\sigma}^2$. По одному только последнему значению $\overline{\sigma}^2$ продолжением имели бы: $\underline{M}\cos(u\zeta_n)\geqslant 1-u^2\overline{\sigma}^2/2$, а то, что содержится о ζ_n в выражении (6.24), несколько лучшее, хотя при $u\to 0$ они и стремятся друг к другу.

Пример 6.12. Пусть $v(z)$ есть по форме «гауссов колокол»: $v_\delta(z)=\exp(-\pi z^2/\delta^2)$, где δ — его интегральная ширина: $\int v_\delta(z)dz=\delta$. Тогда $V(u)=(\delta/2\pi)\exp(-u^2\delta^2/4\pi)\geqslant 0$ и ошибка асимптотически будет не больше:

$$\overline{\alpha}(\partial)\leqslant 1-\frac{\delta}{\pi}\int\limits_0^{\pi/(2c^*\overline{\sigma}^2)} \exp\left\{-\frac{u^2}{2}\left(\overline{\sigma}^2+\frac{\delta^2}{2\pi}\right)\right\}\cos(c^*u^2\overline{\sigma}^2)\, du.$$

Приравнивая правую часть значению $\alpha/\varkappa$, рассчитывается ширина δ, обеспечивающая нужный уровень α. При этом напомним, что ширина самой оценки $v(\zeta_n)$ будет равна $\delta/\sqrt{n}$, убывая с ростом n пропорционально $\sqrt{n}$ при фиксированном асимптотически α.

Вопрос, конечно, почему в примере в качестве исходной формы $v(\zeta_n)$ взят «гауссов колокол»? Именно потому, что преобразование Фурье от него есть всюду неотрицательная функция. Этому факту имеется разумное обоснование. В самом деле, при принятых данных имеем $\overline{M}\cos u\zeta_n=1$, поэтому наличие отрицательной части $V(u)$ внесет довольно неприятный вклад в ошибку оценки в виде слагаемого $\int V^-(u)du$. В результате те $v(z)$, преобразование Фурье которых неотрицательно, должны иметь определенное преимущество в плане достаточности.

Последнее суждение требует некоторой корректировки, так как не принимает в счет еще одного обстоятельства: согласно (6.24) помимо гармонических средних имеются данные о квадратичном признаке $\overline{M}\zeta^2_n=\overline{\sigma}^2$. Если только на нем основывать оценку, то это привело бы нас обратно к оценке (6.18) параметра сдвига по некоррелированной выборке (и значительно более простым путем, правда, при дополнительном условии $M\xi_i=0$). А целесообразно, по-видимому, использовать все данные о ζ_n, гармонические и степенные, приводящие к следующей смешанной форме оценок:

$$v(z)=[c_0+\sum c_i^+\cos(u_i z)-c^+z^2]^+,\ c_0+\sum c_i^+=1,$$

при последующем выборе u_i, c_0, c^+_i и c^+. Оценки этой формы симметричны и при $z=0$ согласно условию справа достигают максимального, равного 1, значения. Для них

$$\underline{M}\, v(\zeta_n)\geqslant c_0+\sum c_i^+\underline{M}\cos(u_i\zeta_n)-c^+\overline{\sigma}^2=1-\alpha/\varkappa,$$

а ширина, нормированная множителем $1/\sqrt{n}$, найдется интегрированием $v(z)$.

Пример 6.13. Пусть форма оценки задана выражением

$$v(z)=[\cos(\pi z/2\Delta)-z^2/(16\Delta^2)]^+,$$

где делитель при z^2 подобран так, чтобы выражение в квадратных скобках не имело побочных выбро.ов выше 0, кроме одного основного при $z=0$. Тогда подстановка на место аргумента $\zeta_n=\hat{\xi}\sqrt{n}$ и использование (6.24) приводит к уравнению

$$\underline{M}\,\partial(\zeta_n)=\exp(-\pi^2 b/2)\cos(c^*\pi^2 b)-b/4=1-\alpha/\varkappa,$$

где $b=\overline{\sigma}^2/(4\Delta^2)$. Решением уравнения находится b, а отсюда и Δ.

Применение других допредельных и предельных результатов принципиально мало отличается от только что проделанного. По этому поводу сейчас выскажем некоторые соображения применительно к неоднородной выборке.

Об оценивании параметра сдвига при неоднородных флуктуациях. Пусть $y_i = x + \xi_i$, $i = 1, \ldots, n$, где ξ_i — независимые с. в. с нулевыми средними $M\xi_i = 0$ и разными дисперсиями: $\overline{M}\xi^2{}_i = \overline{\sigma}^2{}_i$. Тогда нормированную сумму ζ_n нужно заменить на взвешенную $\zeta_n = \sum_{i=1}^{n} c_{in}\xi_i$, $c_{in} \geqslant 0$, записав оценку через $v(\zeta_n)$. Интегральной шириной оценки по x будет величина $\int v(\sum c_{in} y_i - x \sum c_{in})\,dx = (\sum c_{in})^{-1} \int v(z)\,dz$, обратно пропорциональная сумме $\sum c_{in}$. Ошибка монотонно зависит в конечном счете от дисперсии: $\overline{M}\xi^2{}_n = \sum c^2{}_{in}\overline{\sigma}^2{}_i$, поэтому выбор коэффициентов c_{in} будет, по крайней мере, близким к оптимальному, если минимизирует $\sum c^2{}_{in}\overline{\sigma}^2{}_i$ при заданном $\sum c_{in}$. Это ведет к значениям $c_{in} = \lambda/\overline{\sigma}^2{}_i$, откуда $\zeta_n = \lambda \sum \xi_i/\overline{\sigma}^2{}_i$, т. е. весовые коэффициенты оказываются обратными дисперсиями. Становится ясным, что чем больше $\overline{\sigma}^2{}_i$, тем с меньшим весом должен участвовать i-й отсчет в сумме ζ_n, ограждая ее тем самым от влияния этого излишне интенсивного, излишне шумящего отсчета.

С теми же весами получалась оценка (6.19), но для некоррелированных ξ_i. У нас ξ_i независимы, что позволяет достичь дополнительного эффекта оценки применением к сумме $\sum \xi_i/\overline{\sigma}^2{}_i$ допредельных и предельных теорем § 3.4. Дальнейший путь строго следует «колее» предыдущего раздела, и мы подкрепим его примером.

Пример 6.14. Пусть ξ_i независимы, ограничены, заданы $\overline{\sigma}^2{}_i$ и симметричны. Рассмотрим оценку вида $\partial(\xi) = [1 - (\zeta_n/\delta)^{2k}]^+$, $\zeta_n = \lambda \sum \xi_i/\overline{\sigma}^2{}_i$. Пусть $\overline{\sigma}^2{}_i \geqslant \varepsilon > 0$. Применение дополнения 3 к § 3.4 (самый конец главы 3) дает

$$\lim_{n\to\infty} \overline{M}\zeta_n^{2k} \leqslant \lim_{n\to\infty} \frac{(2k)!}{k!\,2^k} [\lambda^2 \Sigma (1/\overline{\sigma}_i^2)]^k,$$

откуда $\delta = \lambda[\varkappa(2k)!/(k!\alpha)]^{1/2k} \sqrt{\Sigma(1/2\overline{\sigma}_i^2)}$. При $\lambda = [\Sigma(1/\overline{\sigma}^2{}_i)]^{-1}$ случайная величина ζ_n будет стабилизирована по n, а асимптотическая нормированная ширина оценки

$$\Omega_* = \lim_{n\to\infty} \Omega_n(\partial)\sqrt{n} = \sqrt{2} \left[\frac{\varkappa(2k)!}{\alpha k!}\right]^{1/2k} \overline{\sigma}_{\text{ср}},$$

где $\overline{\sigma}^2_{\text{ср}} = \left[\frac{1}{n} \Sigma\left(1/\overline{\sigma}_i^2\right)\right]^{-1}$. Ее минимум по k ищется точно так, как это проделывалось в начале настоящего параграфа при асимптотической подстройке оценки степенного типа.

6.5. ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРА МАСШТАБА

Общие соображения. Параметр масштаба определяет энергетические свойства наблюдений: мощность, эффективное значение. Задача его оценивания возникает, во-первых, когда уровень интенсивности процесса несет полезную информацию об интересующем нас положении объекта. Например, величина гидроакустических шумов может характеризовать расстояние до косяка рыб или до корабля. Сам шум в этом случае в некотором смысле обретает ранг полезности. А во-вторых, потребность измерения уровня шумов возникает при определении и текущем контроле показателей работающей аппаратуры, например, для фиксации вероятности ложных тревог в радиолокационных станциях. И в этом случае оценивание уровня шума нельзя считать «второсортным», второстепенным делом, так как оценки составным элементом входят в единый синтезируемый комплекс функционирования аппаратуры.

С позиций современной теории вероятностей оценивание масштабных параметров занимает особое положение. Обязано оно непрерывному времени и допущению о точном знании корреляционных свойств, так как тогда, разлагая процесс конечной временной протяженности от 0 до T в ряд по собственным функциям корреляционного ядра, получаем бесконечное число некоррелированных коэффициентов разложения, которые нормировкой (делением на корни из их дисперсий) все приводятся к одинаковым дисперсиям. В результате происходит превращение наблюдений в бесконечную последовательность преобразованных, новых значений, некоррелированных между собой и одинаково реагирующих на изменение масштаба, и таким образом, несущих неограниченный запас сведений об этом параметре; запас, позволяющий сколь угодно точно его выделить (причем по отрезку $(0, T)$ наблюдений сколь угодно малой протяженности!). Этот парадокс — эффект точного знания — получил название сингулярности [1].

Явление сингулярности оказывается серьезной преградой в реальном оценивании параметра масштаба, так как все оценки рефлекторно тянутся в ее сторону как сулящую «райские кущи» неограниченного улучшения их качественных показателей, перекрывая поиски других путей и напрочь заставляя забыть о том, что сокращение длины интервала $(0, T)$ наблюдений предъявляет все более жесткие требования к точности корреляционной функции, чтобы хоть как-то гарантировать некоррелированность даже небольшого числа компонентов разложения.

Наша цель — строить оценки параметров масштаба по реальным, конечным данным, что само собой исключит любую возможность таких чисто математических «фокусов», как сингулярность, поставит теорию оценивания на почву реальности. Причем, как всегда, за начало возьмем самые слабые исходные данные (о явлении) энергетического толка, а затем будем их постепенно усиливать в нужном направлении, не забывая следить за происходящими при этом видоизменениями оптимальной оценки. В конце концов, полезно знать, какие реальные сведения должны иметься,

[1] Гихман И. И., Скороход А. В. Теория случайных процессов. — М.: Наука, 1971. — Т. 1. — 664 с.

чтобы с нужной точностью оценить параметр масштаба, в частности, его разновидность — дисперсию.

Оценивание параметра масштаба по заданной мощности флуктуаций. Пусть $y_i = x^{l/2}\xi_i$, $i=1, ..., n$, $x>0$, $l \neq 0$. Требуется оценить параметр x, если известно, что ограничена средняя мощность флуктуаций: $\overline{M}\widehat{\xi^2} = \overline{\sigma}^2$, $\widehat{\xi^2} = \sum \xi^2_i/n$, и более ничего не дано.

Отметим, что при $\overline{\sigma}=1$ и $l=1$ параметр x будет *мощностью* (при $M\xi_i=0$ — *дисперсией*) наблюдений, а при $l=2$ — *средним эффективным* их значением. При $l=-2$ это будет *параметр нормировки.* В общем случае $l\neq 0$ будем называть x *параметром масштаба.*

Согласно следствию 4 теоремы 6.1 достаточный класс образуют оценки, зависящие от $\widehat{\xi^2}$ с подстановкой сюда $\xi_i = y_i/x^{l/2}$, что ведет к их виду $\partial_y(x) = [c_0 - c^+ \widehat{y^2}/x^l]^+$, и для них $\underline{M}\partial_y(x) = [c_0 - c^+\overline{\sigma}^2]^+$, $\overline{M}\partial_y(x) = c_0$.

Найдем интегральную ширину и штраф, рассматривая сначала случай $l<0$,

$$\Omega(\partial) = \int\limits_0^\infty [c_0 - x^{-l} c^+ \widehat{y^2}]^+ dx = \frac{c_0 l}{l-1}\left(\frac{c_0}{c^+\widehat{y^2}}\right)^{-1/l},$$

$$\overline{M}\Omega(\partial) = \frac{c_0 l}{l-1} \sup_{\widehat{y^2}} \left(\frac{c_0}{c^+\widehat{y^2}}\right)^{-1/l}.$$

Правая часть равна ∞ при $\widehat{y^2}=0$, поэтому есть смысл предположить $\widehat{y^2} \geqslant \varepsilon > 0$, считая ε сколь угодно малым числом. Тогда составной риск запишется

$$\Pi^\varkappa_\lambda(\partial) = 1 - \varkappa[c_0 - c^+\overline{\sigma}^2]^+ + (\varkappa - 1)c_0 + \lambda c_0 \frac{l}{l-1}\left(\frac{c_0}{c^+\varepsilon}\right)^{-1/l}.$$

Из этого выражения после некоторых выкладок следует, что оптимальным будет значение $c_0=1$. Значение же c^+ будет определяться уровнем α, что при $\alpha<\varkappa$ даст $c^+=\alpha/(\varkappa\overline{\sigma}^2)$. После подстановки найденных значений *оптимальная оценка параметра* x примет вид

$$\partial^*_y(x) = [1 - x^{-l}\alpha\,\widehat{y^2}/(\varkappa\overline{\sigma}^2)]^+, \quad x \geqslant 0. \tag{6.25}$$

Так, при $l=-2$ (тогда $y_i=\xi_i/x$) оптимальная оценка параметра нормировки x получится равной: $\partial^*_y(x) = [1 - x^2\alpha\widehat{y^2}/(\varkappa\overline{\sigma}^2)]^+$, $x>0$. Как видно из рис. 6.5,*а*, это есть полупарабола с максимумом в точке $x=0$, усеченная снизу осью абсцисс и слева — осью ординат.

Пусть теперь $l>0$. Оценка будет иметь вид рис. 6.5,*б* и имеет бесконечную ширину. Тогда, чтобы штраф не равнялся бесконечности, нужно рассматривать взвешенную шкалу интегрирования по x с весовым множителем $q(x)$, убывающим при $x\to\infty$, по крайней мере, быстрее $1/x$. Считая $q(x) = x^{-1-\gamma}$, $\gamma>0$, (т. е. чем больше

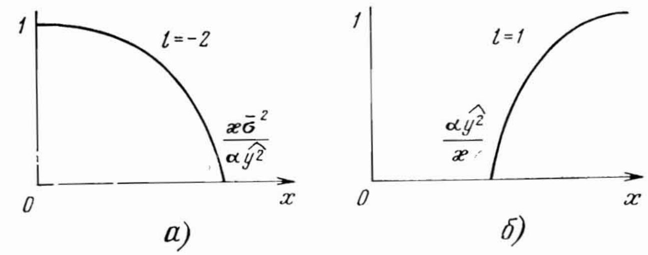


Рис. 6.5, *а*, *б*. Оцена параметра масштаба по энергетическим данным

x, тем меньше берется вес), найдем штраф как верхнее среднее взвешенной ширины:

$$\overline{M}^y\Omega(\partial) = \overline{M}^y \int\limits_0^\infty \left[c_0 - \frac{c^+\widehat{y^2}}{x^l}\right]^+ \frac{dx}{x^{1+\gamma}} = \overline{M}^y \int\limits_{\left(\frac{c^+}{c_0}\widehat{y^2}\right)^{1/l}}^\infty \left(\frac{c_0}{x^{1+\gamma}} -\right.$$

$$\left. - \frac{c^+\widehat{y^2}}{x^{1+\gamma+l}}\right)dx = \frac{c_0 l}{\gamma + l}\max_{\widehat{y^2}}\left(\frac{c_0}{c^+\widehat{y^2}}\right)^{\gamma/l} = \frac{c_0 l}{\gamma+l}\left(\frac{c_0}{c^+\varepsilon}\right)^{\gamma/l}.$$

Составляя риск и производя его минимизацию (точно так же, как это делалось при $l<0$), получим оценку (6.25), которая оказывается оптимальной при любых l.

Так, при $l=1$ и $\overline{\sigma}^2=1$ (x есть мощность наблюдений) доверительная оценка x примет вид (см. рис. 6.5,*б*): $\partial_y(x) = [1 - \alpha\widehat{y^2}/(x\varkappa)]^+$. Как функция переменной x оценка равна 0 при $x \leqslant \alpha\widehat{y^2}/\varkappa$ и стремится к 1 при $x\to\infty$.

Дополнения. 1. Тот же самый вид (6.25) будет иметь оптимальная оценка для непрерывных реализаций $y_t = x^{l/2}\xi_t$, $0 \leqslant t \leqslant T$, с той лишь разницей, что $\widehat{y^2} = \frac{1}{T}\int\limits_0^T y_t^2\,dt$.

2. Сведения о точной мощности (дисперсии) $\widehat{M\xi^2} = \sigma^2$, а также о нулевом среднем флуктуаций $M\xi_i=0$ (или же $M\xi=0$) не меняют вида оценки, т. е. эти сведения оказываются ненужными. То же самое, если x является свободным от ξ.

3. Столь большая расплывчатость получающихся оценок есть следствие крайне скудных данных о флуктуациях. Дело в том, что поставленная задача, когда известна только средняя мощность флуктуаций, в пессимистическом настрое эквивалентна оцениванию параметра масштаба по одному-единственному наблюдению, так как не исключается случай постоянных реализаций $\xi_1 = \xi_2 = ... = \xi_n$.

4. Если заданы точное значение $M\widehat{\xi^2}=\sigma^2$ и верхняя граница $\overline{M}(\widehat{\xi^2})^2=\overline{m}_4$, то доверительная оценка уровня α примет вид

$$\partial_{\mathbf{y}}(x)=[1-c^*(1-x^{-l}\,\widehat{y^2})^2]^+,\ \ c^*=\alpha/[\varkappa\,(1+\overline{m}_4-2\overline{\sigma^2})].$$

Эта оценка, как видно из рис. 6.6, равна 1 при $x=(\widehat{y^2})^{1/l}$ и убывает по мере отклонения x от этого значения. Некоторое улучшение оценки здесь достигается за счет данных о моментах четвертого порядка. Это в некотором роде общее свойство оценок моментов: для получения более удовлетворительных оценок момента k-го порядка нужно иметь сведения о моментах порядка $2k$. Сейчас мы лишний раз в этом убедимся.

Оценивание параметра масштаба по некоррелированной выборке. Пусть $y_i=x^{l/2}\xi_i$, $i=1,\ \ldots,\ n$, $x>0$, $l\neq 0$, и нужно оценить параметр масштаба x, если флуктуации некоррелированы и имеют одинаковые мощности:

$$M\,\xi_i^2=\sigma^2,\ \ M\,\xi_i\,\xi_j=0,\ \ i\neq j,\ \ i,j=j=1,\ldots,n.$$

Оптимальную оценку будем искать в следующем достаточном классе: $\partial_{\mathbf{y}}(x)=[c_0-x^{-l}(c_1\hat{y}^2+c_2\hat{\sigma}^2_{\mathbf{y}})]^+$, $\hat{\sigma}^2_{\mathbf{y}}=\widehat{y^2}-\hat{y}^2$. При нахождении этого класса учтены первичные данные и симметрия поставленной задачи к перестановкам наблюдений. Для этого класса $M\partial_{\mathbf{y}}(x)=[c_0-c_1\sigma^2/n-c_2\sigma^2(n-1)/n]^+$, $\overline{M}\partial_{\mathbf{y}}(x)=c_0$ при $c_1,\ c_2\geqslant 0$.

Точно так же, как это делалось в предыдущем разделе, только с заменой $c^+\hat{y}^2$ на $c_1\hat{y}^2+c_2\hat{\sigma}^2_{\mathbf{y}}$, вычисляется штраф. При $l<0$ он будет равен

$$\overline{M}\,\Omega\,(\partial)=c_0^{(l-1)/l}\,l/[(l-1)\inf\,(c_1\,\hat{y}^2+c_2\,\hat{\sigma}^2_{\mathbf{y}})]=c_0^{(l-1)/l}\,l/(l-1)\,c_2\,\varepsilon,$$

где инфимум достигается при $\hat{y}=0$ и $\hat{\sigma}^2_{\mathbf{y}}=\varepsilon^2$, причем допущено $\widehat{y^2}\geqslant\varepsilon>0$ (иначе ширина будет бесконечна), считая ε сколь можно малым. Оптимальным будет значение $c_0=1$, и тогда при $\alpha<\varkappa$ риск записывается двумя уравнениями: $\Pi^{\varkappa}{}_{\lambda}(\partial)=\alpha+\lambda/c_2$, $\alpha=\varkappa(c_1\sigma^2/n+c_2\sigma^2(n-1)/n)$. Чем больше c_2, тем меньше риск. Поэтому учитывая последнее равенство, оптимальными будут $c_1=0$,

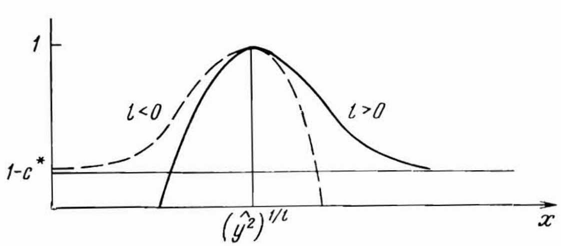


Рис. 6.6. Оценка параметра масштаба при заданном четвертом моменте

$c_2=\alpha n/(\varkappa(n-1)\sigma^2)$, в результате чего оптимальная оценка приобретает вид

$$\partial^*_{\mathbf{y}}(x)=\left[1-x^{-l}\,\frac{\alpha\,n}{\varkappa\,(n-1)\,\sigma^2}\,\hat{\sigma}^2_{\mathbf{y}}\right]^+.$$

Точно такой же вид будет иметь оценка при $l>0$ (если брать взвешенную множителем $x^{-1-\gamma}$ шкалу интегрирования в штрафе).

Таким образом, *данные о некоррелированности наблюдений при отсутствии другой информации лишь незначительно меняют оценку и ее свойства по сравнению с рис. 6.5,а и б*, а именно, оценка сужается в $n/(n-1)$ раз, что при увеличении n становится практически не заметным. Причина этого неприятного эффекта кроется в том, что для получения хороших оценок требуются данные о четвертых моментах флуктуаций.

Допустим теперь, что помимо некоррелированности $M\xi_i\xi_j=0$, $i\neq j$, и $M\xi^2{}_i=\sigma^2$ дано $M\xi^2{}_i\xi^2{}_j=\sigma^4$, $i\neq j$, $\overline{M}\xi^4{}_i=\overline{m}_4$. Тогда достаточный класс образуют оценки вида $\partial_{\mathbf{y}}(x)=[c_0+c_1x^{-l}\widehat{y^2}+c_2x^{-l}\sum\sum_{i\neq j}y_iy_j-c_3x^{-2l}\widehat{y^4}-c_5x^{-2l}\sum\sum_{i\neq j}y^2{}_iy^2{}_j]^+$, где $\widehat{y^4}=\sum y^4{}_i/n$. Постараемся более или менее «угадать» вид оптимальной оценки из этого класса, записав ее в виде

$$\partial_{\mathbf{y}}(x)=[1-c\,(1-x^{-l}\,\widehat{y^2}/a)^2]^+. \tag{6.26}$$

Очевидно следующее: $\overline{M}\partial_{\mathbf{y}}(x)=\overline{\partial}=1$, причем максимум оценки достигается при $x^l=\widehat{y^2}/a$. Если расписать оценку через ξ, раскрыв круглые скобки, то получим

$$\partial(\xi)=\left[1-c+\frac{2c}{a}\,\widehat{\xi^2}-\frac{c}{a^2\,n^2}\left(\sum\xi_i^4+\sum_{i\neq j}\sum\xi_i\,\xi_j\right)\right]^+$$

и отсюда будет следовать, что

$$\underline{M}\,\partial_{\mathbf{y}}(x)=[1-c+2\,c\,\sigma^2/a-c\overline{m}_4/na^2-c\,(n-1)\,\sigma^4/n\,a^2]^+.$$

Считая выражение в квадратных скобках неотрицательным, ищем сначала оптимальное значение параметра a, тогда как c потом будет найдено исходя из заданного уровня α. Если вычислить штраф, полагая $\widehat{y^2}\geqslant\varepsilon>0$, то он будет равен $\overline{M}\Omega(\partial)=\varepsilon^{1/l}a^{-1/l}r(c)$, где $r(c)=(1-c)(c_+{}^{1/l}-c_-{}^{-1/l})+2c\,(c_+{}^{(l-1)/l}-c_-{}^{(l-1)/l})/(1-l)-c\,(c_+{}^{(2l-1)/l}-c_-{}^{(2l-1)/l})/(1-2l)$, а $c_+=1+1/\sqrt{c}$, $c_-=1-1/\sqrt{c}$. Подставляя найденные значения в составной риск, дифференцируя его по a и приравнивая 0, получим значение a как решение уравнения: $\lambda\,l\,\varepsilon^{1/l}\,r\,(c)\,a^{(2l-1)/l}+\sigma^2\,a-(\overline{m}_4-\sigma^4)/n-\sigma^4=0$. Чем меньше уровень α, тем меньше будет значение λ. Поэтому при малых α первым слагаемым левой части равенства можно пренеберчь, положив $\lambda=0$. В результате получим значение параметра a в явном виде: $a=\sigma^2+(\overline{m}_4/\sigma^2-\sigma^2)/n$. Дальнейшие упрощения могут быть получены, если считать n большим, и тогда $a\approx\sigma^2$. Отметим, что увеличение a по сравнению с σ^2 ведет к увеличению средней ширины при $l<0$ и уменьшению при $l>0$.

Полагая для простоты $a=\sigma^2$, получаем исходя из заданной ошибки α значение $c^*=\alpha n\sigma^4/\varkappa(\overline{m}_4-\sigma^4)$. В результате искомая оценка приобретает вид

$$\partial_{\mathbf{y}}(x)=\left[1-\frac{\alpha n\sigma^4}{\varkappa(\overline{m}_4-\sigma^4)}\,(1-x^{-l}\,\widehat{y^2}/\sigma^2)^2\right]^+ . \qquad (6.27)$$

Оценка имеет максимум при $x^l=\widehat{y^2}/\sigma^2$ и принимает ненулевые значения в интервале с границами $\left[\frac{\sigma^2}{\widehat{y^2}}\left(1\pm\sqrt{\frac{\varkappa}{\alpha n}\left(\frac{\overline{m}_4}{\sigma^4}-1\right)}\right)\right]^{-1/l}$.

Размах этого интервала, характеризующий расплывчатость оценки, будет тем меньше, чем больше α, а самое главное, эта ширина будет стремиться к 0 при увеличении n со скоростью $1/\sqrt{n}$. Это то самое, чего мы добивались.

Развитие проблемы. 1. Для расчета оценки (6.27) использовалось точное значение мощности σ^2. В случае неточного значения: $\underline{\sigma}^2=\underline{M}\xi^2_i$, $\overline{\sigma}^2=\overline{M}\xi^2_i$ ошибка в (6.26) равна $\alpha^\varkappa(\partial)=\varkappa(1-\underline{M}\partial_{\mathbf{y}})=\varkappa c[1-2\underline{\sigma}^2/a+\overline{m}_4/na^2+\overline{\sigma}^4(n-1)/na^2]$. Оптимальным значением a, минимизирующим ошибку, будет $a=\overline{\sigma}^4/\underline{\sigma}^2+(\overline{m}_4-\overline{\sigma}^4)/n\sigma^2$.

2. Данные о $\overline{m}_4-\sigma^4$, необходимые для оценки (6.27), могут быть получены косвенным путем. Например, если ξ_i ограничены числом H, т. е. $\underline{P}(|\widehat{\xi_i}|<H)=1$, то неравенство $\overline{M}(\xi^2-\sigma^2)^2\leqslant H^4/4$ позволяет $\overline{m}_4-\sigma^4$ заменить на $H^4/4$.

3. Рассмотрим вопрос об асимптотически (при $n\to\infty$) наилучших оценках, если известно, что $\overline{M}\xi^{2k}_j<\infty$, $\forall k>0$. Считая ξ_i независимыми и полагая для простоты $M\xi^2_i=\sigma^2=1$, введем следующий класс степенных оценок

$$\partial(\xi)=[1-(1-\widehat{\xi^2})^{2k}/\Delta^{2k}]^+,$$

куда для получения $\partial_{\mathbf{y}}(x)$ вместо ξ_i нужно подставить $x^{-l/2}y_i$. Если обозначить $\eta_i=\xi^2_i-1$, то $M\eta_i=M\xi^2_i-1=0$ и тогда $\partial(\eta)=[1-(\hat{\eta}/\Delta)^{2k}]^+$. Мы видим, что в новых обозначениях оценка совпадает с (6.22). Тогда при $n\to\infty$ по аналогии с оценкой (6.22) имеем

$$1-\underline{M}\,\partial=[\overline{M}\,(\xi_i^2-1)^2/\Delta^2 n]^k\,(2k)!/k!\,2^k+0\,(1/n).$$

Отсюда, пренебрегая последним слагаемым правой части и используя то, что $\overline{M}\partial=1$, получаем ширину Δ оценки по заданному уровню α:

$$\Delta=\left(\frac{\varkappa\,(2k)!}{\alpha\,k!\,2^k}\right)^{1/2k}\sqrt{\frac{\overline{M}\,(\xi_i^2-1)^2}{n}}\,.$$

Если воспользоваться теперь в целях упрощений формулой Стирлинга, то получим $[(2k)!/k!2^k]^{1/2k}\approx\sqrt{2k/e}\,2^{1/4k}$, откуда

$$\Delta\approx\frac{1}{\sqrt{n}}\left(\sqrt{2}\,\frac{\varkappa}{\alpha}\right)^{1/2k}\sqrt{2k/e}\,H,\quad H=\sqrt{\overline{M}\,(\xi_i^2-1)^2}.$$

При оптимальном значении $k^*=\ln(\sqrt{2}\varkappa/\alpha)$, минимизирующем ширину Δ оценки, полагая k^* ближайшим целым числом, имеем окончательно $\Delta_*=(H/\sqrt{n})\sqrt{2\ln(\sqrt{2}\varkappa/\alpha)}$ и в результате оценка $\partial_{\mathbf{y}}(x)$ приобретает вид:

$$\partial_{\mathbf{y}}(x)=[1-(1-x^{-l}\,\widehat{y^2})^{2k^*}/\Delta_*^{2k^*}]^+.$$

4. Данные о нулевом среднем $M\xi_i=0$ не меняют найденных оценок. Сказанное не означает, что средние $M\xi_i$ могут быть произвольными, так как некоррелированность $M\xi_i\xi_j=0$, $i\neq j$, сама по себе ограничивает возможные вариации среднего,

5. Пусть $y_i=x^{l/2}(\xi_i+\theta)$, $i=1,\ldots,n$, где $\theta\in\mathscr{R}$ — скалярный мешающий параметр. Тогда оценку следует искать в виде: $\partial_{\mathbf{y}}(x)=[1-c(1-x^{-l}\hat{\sigma}^2_{\mathbf{y}}/\sigma^2)^2]^+$. Чтобы ошибка этой оценки стремилась к 0 при увеличении n, требуются следующие условия: $M\xi_i\xi_j=0$, $i\neq j$; $M\xi^2_i=\sigma^2$; $M\xi^2_i\xi^2_j=\sigma^4$, $i\neq j$; $\overline{M}\xi^4_i=\overline{m}_4$; $M\xi_i\xi_j\xi_k\xi_l=0$, $i\neq j$, $i\neq k$, $i\neq l$. При этих условиях $1-\underline{M}\partial_{\mathbf{y}}(x)=c[\overline{m}_4(1/n-2/n^2+1/n^3)/\sigma^4-1/n+3/n^2-1/n^3]=\alpha/\varkappa$, откуда и находится коэффициент c. Такой же останется оценка для наблюдений $y_i=x^{l/2}\xi_i+m$, $m\in\mathscr{R}$.

6. Рассмотрим случай, когда мощности флуктуаций ξ_i по известному закону меняются: $M\xi^2_i=\sigma^2_i$, $\overline{M}\xi^4_i=\sigma^4_i\overline{m}_4$. Тогда переходя к новым наблюдениям $z=y_i/\sigma_i=x^{l/2}\xi_i/\sigma_i=x^{l/2}\eta_i$, мы сводим задачу к рассмотренной с $M\eta^2_i=1$. Оценка будет отличаться от найденных в этом пункте лишь тем, что вместо $\widehat{y^2}$ подставляется $\sum y^2_i/(\sigma^2_i n)$. Аналогично, если задана корреляционная матрица $\mathbf{B}$ вектора ξ, то вектор $\boldsymbol{\eta}=\mathbf{B}^{-1/2}\xi$ будет обладать только что рассмотренными свойствами, поэтому $\widehat{y^2}$ заменяется на $\mathbf{y}^{\mathsf{T}}\mathbf{B}^{-1}\mathbf{y}/n$.

Мы не сможем в этой задаче перейти к процессам $y_t=x^{l/2}\xi_t$, пока не изучим влияние ошибок в знании σ^2_i на свойства оценок. Дело в том, что для процессов с точным значением корреляционной функции $B(t,\tau)=M\xi_t\xi_\tau$, разлагая ядро по собственным функциям $B(t,\tau)=\sum_i\sigma^2_i\varphi_i(t)\varphi_i(\tau)$ в новых наблюдениях $z_i=\int y_t\varphi_i(t)\,dt/\sigma_i=x^{l/2}\eta_i$ получаем бесконечно длинную некоррелированную последовательность случайных величин η_i, $M\eta^2_i=1$, что соответствует $n=\infty$ и приведет к абсолютно точной оценке x. Этот парадокс сингулярности, о котором говорилось во введении к параграфу, есть следствие допущения о точном знании $B(t,\tau)$, предположения, увы, практически не выполнимого. Выход содержится в следующем пункте.

7. Пусть мощности флуктуаций не являются точно известными, т. е. $y_i=x^{l/2}\xi_i$, $M\xi_i\xi_j=0$, $\underline{M}\xi^2_i=\underline{\sigma}^2_i$, $\overline{M}\xi^2_i=\overline{\sigma}^2_i$, $\overline{M}\xi^2_i\xi^2_j=\overline{\sigma}^2_i\overline{\sigma}^2_j$,

$\overline{M}\xi^4_i = m_i$, $i \neq j = 1, \ldots, n$. Ищем оценку в виде $\partial_{\mathbf{y}}(x) = [1 - c(1 - \sum y^2_i/d_i)^2]^+$, $d_i \geqslant 0$. Для нее $\overline{M}\partial_{\mathbf{y}} = 1$ и

$$\overline{\alpha}(\partial) = 1 - \underline{M}\,\partial_y = c\left[1 - 2\sum\frac{\underline{\sigma}_i^2}{d_i} + \sum\frac{\overline{m}_i - \overline{\sigma}_i^4}{d_i^2} + \left(\sum\frac{\overline{\sigma}_i^2}{d_i}\right)^2\right]^+ .$$

Определим d_i исходя из минимизации ошибки $\overline{\alpha}(\partial)$. Для этого продифференцируем правую часть равенства по d_i и приравняем 0. Получим $d_i = (\overline{m}_i - \overline{\sigma}^4_i)/[\underline{\sigma}^2_i - \gamma\overline{\sigma}^2_i]^+$, где γ находится из уравнения $\gamma = \sum\overline{\sigma}^2_i[\underline{\sigma}^2_i - \gamma\overline{\sigma}^2_i]^+/(\overline{m}_i - \overline{\sigma}^4_i)$. Отметим, что $\gamma < 1$. Мы видим, что $d_i = \infty$ при $\underline{\sigma}^2_i/\overline{\sigma}^2_i < \gamma$. Таким образом, если ошибка в знании σ^2_i достаточно велика, то наблюдение y_i делится на ∞ и исключается. *Накапливаются лишь те y^2_i/d_i, для которых $\underline{\sigma}^2_i/\overline{\sigma}^2_i > \gamma$.*

Сказанное относится и к процессам, так как в разложении неточной корреляционной функции в ряд по собственным функциям ядра коэффициенты разложения, обозначенные так же σ^2_i, станут неизвестными. Причем чем больше индекс i, тем меньше их абсолютные значения и тем меньше будет отношение $\underline{\sigma}^2_i/\overline{\sigma}^2_i$. Поэтому, как и выше, придется ограничиться лишь конечным усеченным рядом, включающим только «главные компоненты» разложения.

6.6. ЗАКЛЮЧЕНИЕ

Логика такова, что на неясный вопрос не жди вразумительного ответа. Естественным ответом на размытость моделей, вызванную априорной бедностью данных, будет расплывчатость оценок.

Оценки искомых параметров должны подчиняться двум диаметрально противоположным требованиям. С одной стороны, быть надежными, и следовательно, в нужной мере расплывчатыми, а с другой — потребительский интерес вынуждает их к конкретности, т. е. по возможности к наибольшей точности. Разрешение этого противоречия ведет к оптимальным доверительным оценкам фиксированной надежности.

Введение составного риска как суммы штрафа за расплывчатость (в виде площади под кривой решений) плюс величина (вероятность) взвешенной ошибки достигает своей цели — существенного упрощения методов синтеза доверительных оценок. Даже при классических распределениях вероятностей минимизация составного риска ведет к новым по содержанию общим формулам для доверительных интервалов (§ 6.2) (для нормальных распределений превращающимся в известные), ставить и решать новые задачи, например, найти доверительную оценку дисперсии, обладающую минимальной расплывчатостью; получить оптимальные доверительные оценки параметров регрессии.

Самое важное, что составной риск, органично сочетаясь с конструкцией интервальных моделей в виде первичных средних, выходит на простые структуры оптимальных оценок, которыми будут усеченные снизу линейные комбинации первичных признаков, о чем гласит теорема 6.1. В зависимости от первичных данных получаются самые разнообразные по форме оценки, дающие

по шкале 0—1 картину предпочтений значениям параметра. В частности, при вероятностных моделях это будут доверительные интервалы (предпочтение полное 1 внутри и 0 вне интервала) ввиду того, что первичные признаки здесь индикаторные.

Меньше данных — легче поиск оптимальной оценки. Разве не это правильная пока непривычная концепция? Нет данных — никакого поиска не надо, оценки тривиальны (всем значениям будет приписано одинаковое предпочтение).

Однородность исходных данных, их симметрия откликается в структурах оценок уменьшением числа варьируемых коэффициентов, подлежащих оптимизации. Коэффициенты связываются линейным уравнением, фиксирующим надежность, а в остальном находятся из минимума интегральной расплывчатости оценки. Их поиск вкладывается в типовую параметрическую задачу оптимизации с ограничениями. Мы ее решаем аналитически, не исключая полезности численных способов в виде стандартных программ.

Оптимальные доверительные оценки параметра сдвига находятся в § 6.3, где считается известной только усредненная мощность флуктуаций. Более точными они становятся для некоррелированных флуктуаций, а затем обобщаются на случай заданных корреляций и на оценивание параметра (одномерного) регрессии. Примечательно, что параболический вид все эти оценки наследуют от квадратичных признаков, составляющих свойства второго порядка наблюдений. И характерно для нашей теории, что оценки по сложности одинаковы как для последовательности наблюдений, так и для процессов, если состав первичных данных одинаков. По мере роста числа данных оценки несколько усложняются, зато и уточняются, становятся по форме более узкими при той же надежности.

Предугадывается потребительский вопрос, лучше ли полученные оптимальные оценки, чем привычные доверительные интервалы для нормального распределения? На наивный вопрос ответ неутешителен: конечно же, хуже! И могут быть намного, так как разные оценки рассчитаны на разные условия. Одни условия — это режим полного априорного благополучия в виде нормального распределения вероятностей, по нему считается доверительный интервал. Полученные нами оценки оптимальны в условиях полного неблагополучия, свойственного бедным знаниями моментным задачам, поэтому более расплывчаты.

И все же, неблагополучие их относительное. При независимых отсчетах по мере увеличения числа наблюдений без каких бы то ни было дополнительных предположений наши оценки приближаются по свойствам к нормальным (§ 6.4). Можно лишь подозревать, что этому мы обязаны внутренним законам нормальной сходимости, формально при синтезе нигде не задействованным. Данный результат демонстрирует великолепное «пищеварение» разработанного аппарата, позволяющего при крайне бедном априорном пайке усваивать любые питательные вещества. Внедрение допредельных и предельных результатов в оценку на систематическую основу поставлено в конце § 6.4.

Конец главы посвящен задаче оценивания параметра масштаба (в частности, мощности) по данным о свойствах второго порядка наблюдений. Неожиданным оказывается здесь то, что отсутствие ограничений на моменты четвертого порядка не позволяет получить сколь-либо хорошие оценки даже при неограниченном росте длины наблюдений. Но исправить их при наличии нужных ограничений оказывается очень просто. И совершенно правильно ожидалось, что проблема сингулярности (состоящая в абсолютной точности оценки дисперсии, а в общем-то и сдвига) сама собой становится невозможной при отказе от идеальных вероятностных моделей и переходе к реальным интервальным. Сингулярность — побочный плод теоретизированного изобилия и недосягаемая мечта для привычной практическим задачам априорной бедности.