Класс сниженных оценок – обоснован или эвристичен?

2. Устойчивое оценивание при байесовском точечном засорении

Александр Михайлович Шурыгин

2.1. Модели засорения и критерии качества

Пусть реальное распределение наблюдений является ε-засоренным с функцией

$$G(y,\theta) = (1-\varepsilon)F(y,\theta) + \varepsilon H(y,\theta)$$

и используется M-оценка параметра как решение уравнения

$$\sum_{i=1}^{N} \psi\left(y_i, \hat{\theta}\right) = 0.$$

В классическом статистическом анализе в качестве показателя неточности оценивания используется асимптотическая дисперсия $V(\psi, F)$.

Робастная оценка $\hat{\theta}$ должна быть в некотором смысле устойчивой к замене модельного распределения на реальное.

Качество оценивания будем характеризовать величиной математического ожидания квадрата отклонения оценки $\hat{\theta} = \hat{\theta}_N$ от истинного значения θ. Обозначим эту величину через τ_N:

$$\tau_N = \mathbf{E}(\hat{\theta}_N - \theta)^2.$$

Обозначим смещение оценки как

$$b(\hat{\theta}) = \mathbf{E}\hat{\theta}_N - \theta.$$

Для дисперсии оценки справедливо представление

$$\mathbf{D}\hat{\theta}_N = \mathbf{E}(\hat{\theta}_N^2) - (\mathbf{E}\hat{\theta}_N)^2.$$

Отсюда имеем

$$\tau_N = \mathbf{E}(\hat{\theta}_N^2) - 2\theta\mathbf{E}\hat{\theta}_N + \theta^2 = \mathbf{D}\hat{\theta}_N + (\mathbf{E}\hat{\theta}_N)^2 - 2\theta\mathbf{E}\hat{\theta}_N + \theta^2 =$$

$$= \mathbf{D}\hat{\theta}_N + (\mathbf{E}\hat{\theta}_N - \theta)^2 = \mathbf{D}\hat{\theta}_N + b^2(\hat{\theta}_N). \qquad (2.1)$$

Оценка, асимптотически несмещенная при модельном засорении, в случае засорения в общем случае становится асимптотически смещенной. Приближенное значение асимптотического смещения при малом ε и некоторых условиях регулярности имеет вид (1.39):

$$b(\psi, H) \approx \varepsilon \int_Y IF(y, \theta, F)\, dH(y, \theta).$$

Чтобы получить асимптотически несмещенную оценку в засоренной модели, уровень засорения ε необходимо при увеличении N устремить к нулю:

$$\varepsilon = \varepsilon_N \to 0, \ N \to \infty.$$

Такую конструкцию называют **моделью сужающихся окрестностей** (Хьюбер).

Привлекательность этой модели состоит еще и в том, что с ростом N при малом, но конечном значении ε смесь может быть идентифицирована.

В регулярных моделях при больших N дисперсия оценки имеет порядок малости $1/N$. Выделим три случая, зависящих от порядка малости смещения, который предположим равным

$$\frac{1}{N^{u/2}}, \quad u > 0.$$

1. Если квадрат смещения имеет более высокий порядок малости, чем дисперсия, т. е.

$$u > 1,$$

то в выражении (2.1) с ростом N второе слагаемое становится пренебрежимо мало по сравнению с первым. При этом задача сводится к классической задаче минимизации асимптотической дисперсии, которая приводит к методу максимального правдоподобия, часто являющемуся неустойчивым. Требование устойчивости оценки не учитывается, поскольку в асимптотике отклонением от модели можно пренебречь.

2. При

$$0 < u < 1$$

имеем обратную ситуацию: в формуле (2.1) асимптотически остается только квадрат смещения. Критерий не учитывает неточность оценок, поскольку изменчивостью оценок в асимптотике можно пренебречь.

3. При $u = 1$ оба слагаемых в (2.1) имеют порядок малости $1/N$. Критерий учитывает как дисперсию, так и смещение оценки.

Рассмотрим этот случай более подробно.

Так как квадрат смещения имеет порядок малости $1/N$, то можно установить связь между ε и N. Из (1.39) с точностью до произвольной положительной константы k имеем

$$\varepsilon_N = \frac{k}{\sqrt{N}}.$$

При оперировании бесконечно малыми величинами вводят их асимптотические аналоги – конечные величины, полученные путем коррекции порядка малости. В нашем случае асимптотический аналог величины τ_N с учетом (2.1) и (1.39) имеет вид

$$\tau(\psi, F, H, k) = \lim_{N\to\infty} N\,\tau_N = V(\psi, F) + k^2 \left[\int_Y IF(y, \psi, F)\, dH(y, \theta) \right]^2 . \qquad (2.2)$$

Назовем эту величину **квадратичной ошибкой оценки**.

Таким образом, при рассмотренном порядке малости ε вклад засорения сосредоточен во втором слагаемом, связанном со смещением оценки. Поэтому смысл второго слагаемого в (2.2) – это чувствительность оценки к малому засорению плотности (характеризует неустойчивость оценивания), а первое слагаемое характеризует неточность оценивания.

В формуле (2.2) важную роль играют константа k и функция H. Обсудим их.

Величина k на практике неизвестна и вряд ли может быть определена по выборке. Однако показатель (2.2) можно рассматривать как **компромиссный**, учитывающий некоторый баланс между показателем неточности $V(\psi, F)$ и показателем неустойчивости, выражающимся через функцию влияния $IF(y, \psi, F)$. В зависимости от величины k в показателе (2.2) будет преобладать влияние одного из двух слагаемых.

Обратимся к засоряющему распределению H. Простейшим видом засорения является точечное, когда

$$H(y,\theta) = \Delta(y - y^*),\ h(y,\theta) = \delta(y - y^*),$$

где Δ – функция единичного скачка, $h(y,\theta)$ – плотность засоряющего распределения, δ – дельта-функция Дирака. Таким образом, засорение представляет собой точечный импульс, возникающий в некоторой точке $y = y^*$ и не зависящий от параметра θ.

Согласно (1.39), имеем

$$b(\psi, H) = b(\psi, y^*) \approx \varepsilon \int_Y IF(y,\theta,F)\delta(y - y^*)dy = \varepsilon IF(y^*,\psi,F) \tag{2.3}$$

и квадратичная ошибка (2.2) принимает вид

$$\tau(\psi, F, y^*, k) = V(\psi, F) + k^2 IF^2(y^*,\psi,F). \tag{2.4}$$

Обратим внимание на вопрос полезности для практики модели точечного засорения.

Пусть функция влияния является ограниченной (т. е. оценка B-робастна), тогда справедливо

$$IF_* \le IF(y,\psi,F) \le IF^*,$$

где $IF_* = \inf\limits_y IF(y,\psi,F)$, $IF^* = \sup\limits_y IF(y,\psi,F)$.

При произвольном засорении с плотностью $h(y,\theta)$ имеем

$$IF_* \le \int\limits_Y IF(y,\psi,F)h(y,\theta)dy \le IF^*.$$

Пусть функция влияния непрерывна. Из ее непрерывности и ограниченности по теореме о среднем значении следует, что для любой плотности $h(y,\theta)$ найдется значение y^* такое, что будут **равны** значения квадратичной ошибки (2.2) при ε-засорении – точечном и произвольном.

В результате зависимость асимптотического смещения от функции H мы перевели в зависимость от значения y^*. От последней зависимости можно избавиться, рассмотрев наихудший случай, т.е. максимизировав квадратичную ошибку (2.4) по y^*:

$$\max_{y^*} \tau(\psi,F,y^*,k) = V(\psi,F) + k^2 \max_{y^*} IF^2(y^*,\psi,F) = V(\psi,F) + k^2 \left[\gamma^*(\psi,F)\right]^2.$$

Последний показатель отражает компромисс между асимптотической дисперсией и чувствительностью к грубой ошибке, и в этом смысле он связан с принципом получения оптимальных B-робастных оценок. Однако рассмотренные выше оптимальные B-робастные оценки обычно неустойчивы при асимметричном засорении.

Для получения оценок, устойчивых при асимметричном засорении, будем использовать другой показатель качества, который построен на основе модели **байесовского точечного засорения**, введенной А.М. Шурыгиным.

Мониторинг с помощью беспроводных сенсорных сетей.

Предположим, что величина y^* для каждой конкретной выборки постоянна, однако имеется серия выборок, в которой y^* является случайной величиной с плотностью $s(y^*,\theta)$.

Совместная плотность y и y^* имеет вид

$$g(y,y^*,\theta)=g(y,\theta\,|\,y^*)s(y^*,\theta),$$

где $g(y,\theta\,|\,y^*)=(1-\varepsilon)f(y,\theta)+\varepsilon\delta(y-y^*)$.

Определим маргинальное распределение наблюдений в серии выборок:

$$g(y,\theta)=\int\limits_Y g(y,y^*,\theta)dy^*=\int\limits_Y g(y,\theta\,|\,y^*)s(y^*,\theta)dy^*=$$

$$=(1-\varepsilon)f(y,\theta)\int\limits_Y s(y^*,\theta)dy^*+\varepsilon\int\limits_Y \delta(y-y^*)s(y^*,\theta)dy^*=(1-\varepsilon)f(y,\theta)+\varepsilon s(y,\theta).$$

Таким образом, в серии выборок распределение наблюдений является **обычным** ε-засоренным распределением.

2.2. Оптимальные оценки при байесовском точечном засорении

Рассмотрим функционал квадратичной ошибки (2.4), т. е. компромиссный функционал (2.2), в условиях байесовского точечного засорения.

Математическое ожидание функционала (по плотности s) имеет вид

$$\mathbf{E}_s\tau(\psi, F, y^*, k) = \mathbf{E}_s\left[V(\psi, F) + k^2 IF^2(y^*, \psi, F)\right] = V(\psi, F) + k^2\mathbf{E}_s IF^2(y^*, \psi, F), \quad (2.11)$$

где $\mathbf{E}_s$ – символ математического ожидания по плотности $s(y, \theta)$.

Таким образом, функционал (2.11) – это компромисс между классическим показателем неточности и показателем неустойчивости в модели байесовского точечного засорения, для которого введем обозначение $U_s(\psi)$:

$$U_s(\psi) = U_s(\psi, F) = \mathbf{E}_s IF^2(y, \psi, F) = \int_Y IF^2(y, \psi, F) s(y, \theta) dy = \frac{\mathbf{E}_s \psi^2(y, \theta)}{d^2}. \quad (2.5)$$

Если функция $s(y,\theta)$ ненормированная, но интегрируемая, то справедливо

$$\mathbf{E}_s\tau(\psi,F,y^*,k) = s_0 V(\psi,F) + k^2 U_s(\psi,F),$$

где $s_0 = \int\limits_Y s(y,\theta)dy$, откуда можно перейти к эквивалентному функционалу

$$\frac{1}{s_0}\mathbf{E}_s\tau(\psi,F,y^*,k) = V(\psi,F) + \frac{k^2}{s_0}U_s(\psi,F) = V(\psi,F) + k^2 U_{\tilde{s}}(\psi,F),$$

где $\tilde{s}(y,\theta) = s(y,\theta)/s_0$ – (нормированная) плотность распределения.

Распишем функционал (2.11):

$$\mathbf{E}_s\tau(\psi,F,y^*,k) = \frac{\mathbf{E}\psi^2}{d^2} + k^2\frac{\mathbf{E}_s\psi^2}{d^2} =$$

$$= \frac{1}{d^2}\int\limits_Y \psi^2(y,\theta)\left[f(y,\theta) + k^2 s(y,\theta)\right]dy = U_{\tilde{s}}(\psi,F), \qquad (2.12)$$

здесь функция $\tilde{s}(y,\theta) = f(y,\theta) + k^2 s(y,\theta)$ не удовлетворяет условию нормировки плотности распределения, однако она с точностью до нормировки может быть проинтерпретирована как плотность смеси распределений:

$$\tilde{s}(y,\theta) \sim (1-\alpha)f(y,\theta) + \alpha s(y,\theta),$$

где $0 < \alpha < 1$, $k^2 = \alpha/(1-\alpha)$ и $\alpha = k^2/(1+k^2)$.

Минимизируя функционал $U_s(\psi, F)$ на множестве оценочных функций, удовлетворяющих условию

$$\mathbf{E}_s \psi^2(y, \theta) < \infty, \tag{2.6}$$

будем получать наилучшие оценочные функции при байесовском точечном засорении.

Будем считать выполненными следующие условия:

1) $\mathbf{E} N \cdot (\hat{\theta}_N - \theta)^2 < \infty$ при модельном распределении;

2) условие асимптотической несмещенности оценки (1.6):

$$\int_Y \psi(y, \theta) f(y, \theta) dy = 0;$$

3) $d = -\int_Y \psi'_\theta(y, \theta) f(y, \theta)\, dy = \int_Y \psi(y, \theta) f'_\theta(y, \theta)\, dy$ (следствие (1.6)).

Теорема 2.1. Функционал $U_s(\psi, F)$ достигает минимума на оценочной функции

$$\psi_s(y,\theta) = \arg\min_{\psi} U_s(\psi, F) = c\left[\frac{\partial}{\partial\theta}\ln f(y,\theta) + \beta\right]\frac{f(y,\theta)}{s(y,\theta)}, \qquad (2.7)$$

где $c = c(\theta) \neq 0$ задает множество эквивалентных оценочных функций, а $\beta = \beta(\theta)$ обеспечивает справедливость условия асимптотической несмещенности оценки (1.6).

Из условия асимптотической несмещенности оценки (1.6)

$$\int_Y \psi(y,\theta) f(y,\theta) dy = 0$$

легко определить вид величины β:

$$\beta = -\frac{\displaystyle\int_Y \frac{\partial}{\partial\theta}\ln f(y,\theta)\frac{f^2(y,\theta)}{s(y,\theta)}dy}{\displaystyle\int_Y \frac{f^2(y,\theta)}{s(y,\theta)}dy}.$$

Заметим, что для справедливости теоремы функция $s(y,\theta)$ не обязательно должна быть нормированной. Т.е. она может быть применена для компромиссного функционала (2.12) $V(\psi,F)+k^2\mathbf{E}_s IF^2(y^*,\psi,F)$. В результате получим наилучшую оценочную функцию вида

$$\psi_s(y,\theta)=c\left[\frac{\partial}{\partial\theta}\ln f(y,\theta)+\beta\right]\frac{f(y,\theta)}{f(y,\theta)+k^2 s(y,\theta)}. \qquad (2.13)$$

Таким образом, переход от оптимизации функционала (2.5) к функционалу (2.12) не требует новой техники получения решений, меняется лишь их интерпретация.

Можно отказаться от условия **интегрируемости** функции $s(y,\theta)$, но тогда необходимо отказаться и от функционала

$$\mathbf{E}_s\tau(\psi,F,y^*,k)=V(\psi,F)+k^2U_s(\psi,F).$$

Однако можно ввести функционал $U_s(\psi,F)=\mathbf{E}_s IF^2(y,\psi,F)$ как квадрат L_2-нормы (с весом $s(y,\theta)$) функции влияния на множестве функций, удовлетворяющих условию (2.6) $\mathbf{E}_s\psi^2(y,\theta)<\infty$, а затем – **компромиссный** функционал

$$V(\psi,F)+k^2U_s(\psi,F)=U_{\tilde{s}}(\psi,F), \qquad (2.14)$$

уже не имеющий смысла $\mathbf{E}_s\tau(\psi,F,y^*,k)$. При этом теорема 2.1 вновь остается справедливой, давая оптимальную оценочную функцию (2.13).

Конкретизируя функцию $s(y,\theta)$, можно получать различные решения. Плотность $s(y,\theta)$, однако, неизвестна на практике.

Один из путей решения этой проблемы – сформулировать минимаксную задачу, подобную рассмотренной в разделе 1.4:

$$\psi^*(y,\theta) = \arg\inf_{\psi}\sup_{s\in S} U_s(\psi,F), \tag{2.9}$$

где S – множество возможных плотностей, а функционал $U_s(\psi,F)$ имеет два оптимизируемых аргумента, при этом должно учитываться условие асимптотической несмещенности оценки (1.6).

Ее решением будет оценочная функция **медианной** оценки.

Медианная оценка имеет ограниченную функцию влияния, т. е. является *B*-робастной, но неустойчива при асимметричном засорении.

Другой, оказавшийся более интересным, путь – определить наихудшую плотность $s(y,\theta)$ для наилучшей оценочной функции (2.7):

$$s_*(y,\theta) = \arg\max_{s\in S} U_s(\psi_s, F) = \arg\max_{s\in S} \min_{\psi} U_s(\psi, F). \tag{2.10}$$

Стойкой называется оценка θ_s с оценочной функцией, которая есть решение максиминной задачи (2.10) при множестве S, являющемся параметрическим семейством плотностей, содержащим модельную плотность.

Пример 2.1. Рассмотрим модель нормального распределения $N(\mu,\sigma^2)$ с неизвестным математическим ожиданием и известной дисперсией. Стойкие оценки будем получать исходя из S в виде множества плотностей нормальных распределений $N(\mu,\omega^2\sigma^2)$, $\omega \in R^1$. Наихудшее значение ω^2 получается в результате решения максиминной задачи равно $\omega^2 = 2$. Имеем оценочное уравнение:

$$\sum_{i=1}^{N}(y_i - \mu_s)\exp\left[-\frac{(y_i-\mu_s)^2}{4\sigma^2}\right] = 0.$$

Параметр β равен нулю, поскольку плотность – четная функция, а оценочная функция нечетная. Это справедливо и для всех последующих примеров.

Из-за экспоненциально убывающего веса оценка оказывается сниженной.

С другой стороны, исходя из вида наилучшей оценочной функции в теореме 2.1 можно целенаправленным выбором функции $s(y,\theta)$ обеспечить предпочтительные свойства оценки $\hat{\theta}$.

Пример 2.2. Выберем

$$s(y,\theta) = f(y,\theta),$$

при этом распределение наблюдений в серии выборок является незасоренным, но каждая отдельная выборка засорена, тогда

$$U_s(\psi,F) = V(\psi,F)$$

и согласно теореме 2.1 оптимальной становится оценка по методу максимального правдоподобия. Действительно, при $\beta = 0$ имеем

$$\psi_0(y,\theta) = c\frac{\partial}{\partial\theta}\ln f(y,\theta).$$

При возможности изменения порядка дифференцирования и интегрирования получаем

$$\int_Y \psi_0(y,\theta) f(y,\theta)dy = c\int_Y \frac{\partial}{\partial\theta}\ln f(y,\theta) f(y,\theta)dy = c\int_Y \frac{\partial}{\partial\theta} f(y,\theta)dy =$$

$$= c\frac{\partial}{\partial\theta}\int_Y f(y,\theta)dy = 0.$$

В результате убедились, что оценка является асимптотически несмещенной. Оценка по методу максимального правдоподобия, как правило, неустойчива к засорению.

Пример 2.3. Пусть наблюдения имеют нормальное распределение с оцениваемым математическим ожиданием μ и известной дисперсией σ^2. Если $s(y,\mu)$ есть плотность нормального распределения с тем же математическим ожиданием и дисперсией $\omega^2\sigma^2$, то наилучшая оценочная функция

$$\psi(y,\mu)=(y-\mu)\exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\left(1-\frac{1}{\omega^2}\right)\right]$$

приведет к неустойчивой оценке при $\omega^2<1$ (когда дисперсия s меньше дисперсии f) и к устойчивой, в том числе при асимметричном засорении, при $\omega^2>1$ (когда дисперсия s больше дисперсии f).

Например, стойкая оценка соответствует выбору $\omega^2=2$ и является устойчивой. Выбирая $\omega^2=1$, получаем условия предыдущего примера и ОМП как результат.

2.4. Максимально неопределенное байесовское точечное засорение

Рассматриваемый подход отличается от подходов Хьюбера и Хампеля тем, что для него требуется бо́льшая априорная информация – информация о виде плотности s. Можно избавиться от этого ограничения, рассматривая минимаксную задачу. Однако ее решение дает медианную оценку – неустойчивое решение, уже известное в рамках подходов Хьюбера и Хампеля. Стойкая оценка соответствует оптимальному решению лишь в ограниченном параметрическом классе плотностей (само его наличие уже является использованием априорной информации). Использование параметрической информации выводит стойкие оценки за пределы теории робастности.

Еще один путь – использование **максимально неопределенного** байесовского точечного засорения, когда засоряющее наблюдение равномерно распределено по области значений случайной величины Y.

Выберем

$$s(y,\theta)=1$$

(функция интегрируема, когда Y – конечный интервал, в иных случаях возможно обоснование решений в некотором обобщенном смысле).

Показатель $U_s(\psi, F)$ в этих условиях обозначим $W(\psi, F)$:

$$W(\psi) = W(\psi, F) = \int\limits_Y IF^2(y, \psi, F)\, dy = \frac{\int\limits_Y \psi^2(y, \theta)\, dy}{d^2}.$$

Функционал $W(\psi, F)$ используется в качестве меры неустойчивости оценки.

Тогда оценка и соответствующая оценочная функция могут быть названы **устойчивыми**, если

$$W(\psi, F) < \infty,$$

и неустойчивыми – в противном случае.

Для случая $Y = R^1$ требование устойчивости автоматически приводит к свойству сниженности.

Удобным свойством показателя $W(\psi, F)$ является, например, то, что для параметра сдвига сниженные оценки для нормального распределения и ОМП стьюдентовского распределения (оценки, устойчивые при асимметричном засорении) являются устойчивыми, а ОМП нормального распределения, выборочная медиана, оценка Хьюбера, неустойчивые при асимметричном засорении, – нет.

С другой стороны, показатель $W(\psi, F)$ можно рассматривать вне модели байесовского точечного засорения, просто как квадрат L_2-нормы функции влияния. Таким образом, полученные в этих условиях результаты являются еще одним подходом в теории робастности.

Согласно теореме 2.1 на множестве устойчивых оценок показатель неустойчивости $W(\psi, F)$ достигает минимума на оценочной функции

$$\psi_*(y,\theta) = c\left[\frac{\partial}{\partial\theta}\ln f(y,\theta) + \beta\right] f(y,\theta).$$

Оценка с оценочной функцией ψ_* называется *оценкой максимальной устойчивости* (ОМУ).

Классическая относительная характеристика точности оценивания – это эффективность

$$\operatorname{eff}\ \psi = \frac{1}{V(\psi,F)}\bigg/\frac{1}{V_0} = \frac{V_0}{V(\psi,F)},$$

где V_0 определяет асимптотическую дисперсию оценки максимального правдоподобия. Таким образом, справедливо

$$0 \le \operatorname{eff}\ \psi \le 1.$$

Аналогично эффективности введем относительную характеристику устойчивости

$$\operatorname{stb}\ \psi = \frac{1}{W(\psi, F)} \bigg/ \frac{1}{W_*} = \frac{W_*}{W(\psi, F)},$$

где $W_* = W(\psi_*, F)$, причем $\operatorname{stb}\ \psi_* = 1$, для неустойчивой оценки $\operatorname{stb}\ \psi = 0$.

Если ОМП может иметь низкую (часто нулевую) устойчивость, то ОМУ может иметь низкую эффективность (для нашего примера $\operatorname{eff}\psi_* \approx 0{,}65$).

В аналогичной ситуации в подходе Хампеля возникла идея оптимальных *B*-робастных оценок, которые минимизируют асимптотическую дисперсию при заданной верхней грани значений чувствительности к грубой ошибке γ^*. Мы тоже обратимся к компромиссу между эффективностью и устойчивостью.

Рассмотрим следующие две связанные задачи:

1) $\max\limits_{\psi}\ \operatorname{eff}\psi$ при ограничении снизу на значение $\operatorname{stb}\psi$;

2) $\max\limits_{\psi}\ \operatorname{stb}\psi$ при ограничении снизу на значение $\operatorname{eff}\psi$.

Теорема 2.4. Во множестве устойчивых оценок наибольшая эффективность при ограничении снизу на значение устойчивости, или наибольшая устойчивость при ограничении снизу на значение эффективности, достигается при оценке с оценочной функцией вида

$$\psi_{co}(y,\theta) = c\left[\frac{\partial}{\partial\theta}\ln f(y,\theta) + \beta\right]\frac{f(y,\theta)}{f(y,\theta)+k^2}, \qquad (2.17)$$

где $k^2 = k^2(\theta)$ определяется выбранным уровнем устойчивости или эффективности.

Оценочная функция, полученная в теореме 2.4, и соответствующая оценка называются **условно оптимальными**.

Назовем *компромиссной* оценку θ_c с оценочной функцией, минимизирующей функционал

$$\frac{1}{\text{eff } \psi}+\frac{1}{\text{stb } \psi}=\frac{V\left(\psi,F\right)}{V_0}+\frac{W\left(\psi,F\right)}{W_*}.$$

Эквивалентный ему функционал имеет вид

$$V\left(\psi,F\right)+\frac{V_0}{W_*}W\left(\psi,F\right),$$

это компромиссный функционал (2.14)

$$V(\psi,F)+k^2U_s(\psi,F)$$

при $s(y,\theta)=1$ и $k^2=\dfrac{V_0}{W_*}$.

Вид оценочной функции (2.13) в этих условиях будет

$$\psi_c(y,\theta)=c\left[\frac{\partial}{\partial\theta}\ln f(y,\theta)+\beta\right]\frac{f(y,\theta)}{f(y,\theta)+k^2}.$$

Легко заметить, что компромиссная оценка является частным случаем условно оптимальной оценки.

Пример 2.5. Рассмотрим модель нормального распределения $N(\mu, \sigma^2)$ с оцениваемым математическим ожиданием μ и известной дисперсией σ^2. Компромиссной оценке μ_c соответствует значение $k^2 = 1 \big/ 4\sqrt{\pi}\sigma$. Имеем оценочное уравнение

$$\sum_{i=1}^{N} (y_i - \mu_c) \Bigg/ \left\{ 2\sqrt{2} + \exp\left[\frac{(y_i - \mu_c)^2}{2\sigma^2} \right] \right\} = 0$$

и характеристики $\mathrm{eff} \approx 0{,}82$, $\mathrm{stb} \approx 0{,}92$.

Хотя в компромиссной оценке эффективность и устойчивость учитываются с равным весом, это не обеспечивает равенства их значений.

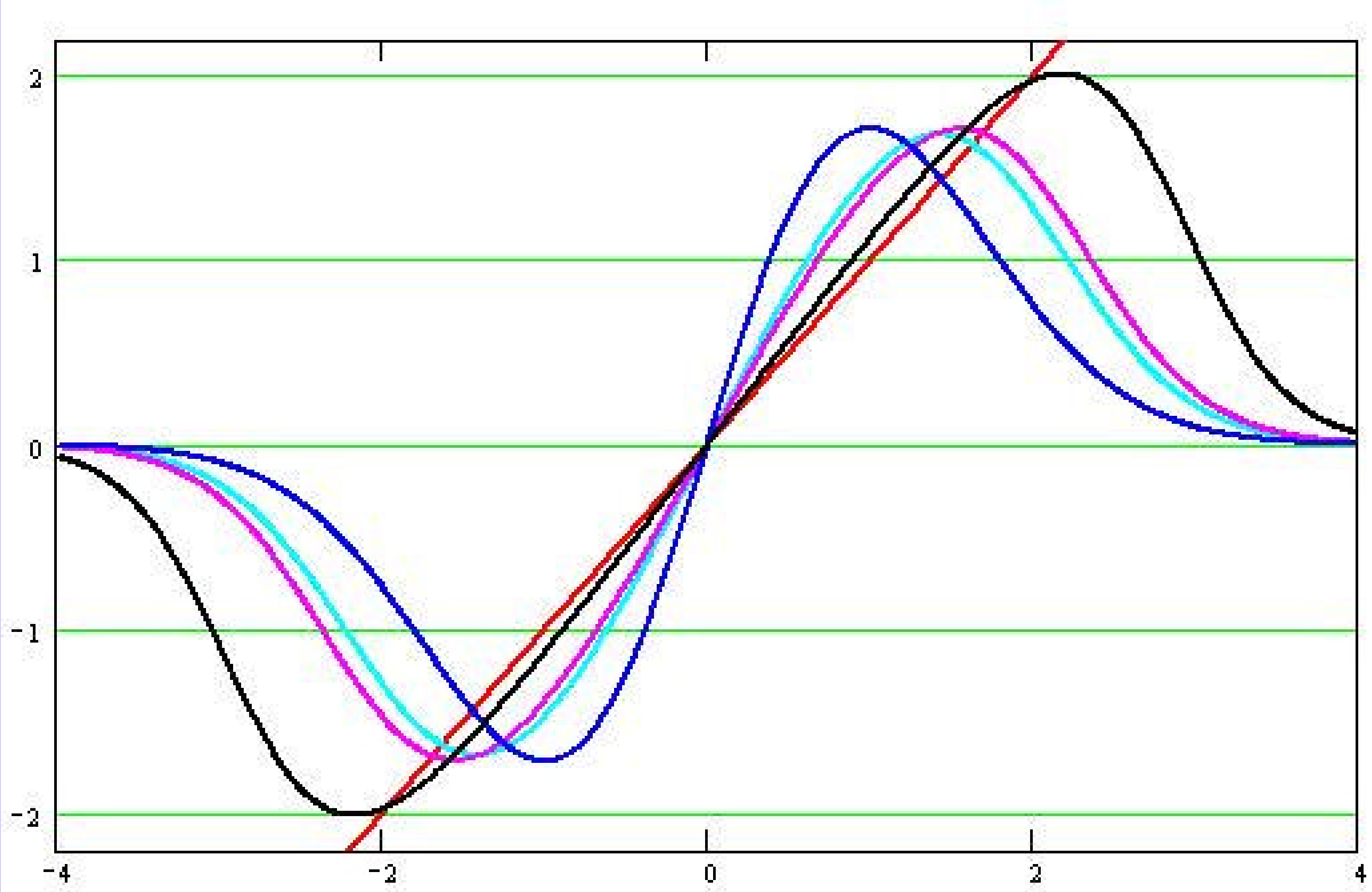
Поэтому введем условно оптимальную оценку, удовлетворяющую условию

$$\mathrm{eff}\ \ \psi = \mathrm{stb}\ \ \psi, \qquad (2.18)$$

назовем ее *равнооптимальной* и обозначим θ_{eq}.

Пример 2.6. В модели нормального распределения $N(\mu, \sigma^2)$ с оцениваемым математическим ожиданием μ и известной дисперсией σ^2 равнооптимальным оценкам μ_{eq} соответствует значение $k^2 \approx \dfrac{1}{6{,}8\sqrt{\pi}\sigma}$ с $\mathrm{eff} = \mathrm{stb} \approx 0{,}85$.

Компромиссная оценка проигрывает равнооптимальной 3% эффективности, но выигрывает 7% устойчивости.



Стандартное нормальное распределение, оценивается параметр μ. Линии: красная – ОМП ($\mathrm{eff}=1,\ \mathrm{stb}=0$), синяя – ОМУ ($\mathrm{eff}\approx 0,65,\ \mathrm{stb}=1$), бирюзовая – компромиссная с $k^2\approx 0{,}14$ ($\mathrm{eff}\approx 0,82,\ \mathrm{stb}\approx 0,92$), лиловая – равнооптимальная с $k^2\approx 0{,}08$ ($\mathrm{eff}=\mathrm{stb}\approx 0,85$), черная – $k^2\approx 0{,}01$ ($\mathrm{eff}\approx 0,96,\ \mathrm{stb}\approx 0,55$).

2.5. Обобщенные радикальные оценки

Условно оптимальные оценки являются семейством оценок, промежуточных между ОМП и ОМУ, включая их как крайние свои элементы. Рассмотрим семейство обобщенных радикальных оценок, имеющих такие же свойства, но более удобные с точки зрения нахождения оценок. При его введении будем использовать некоторые элементы **теории информации**, активно применяемые в статистике.

Будем рассматривать модель байесовского точечного засорения и определять **наихудшую** плотность $s(y,\theta)$, исходя из **принципа максимальной энтропии**, согласно которому (в данном случае) должна максимизироваться **дифференциальная энтропия** вида

$$H(s) = \int_Y s(y,\theta) \ln \frac{1}{s(y,\theta)}\, dy = -\int_Y s(y,\theta) \ln s(y,\theta)\, dy \qquad (2.5.1)$$

при ограничении на нормировку плотности $s(y,\theta)$

$$\int_Y s(y,\theta)\, dy = 1 \qquad (2.5.2)$$

и, обычно, при некоторых дополнительных ограничениях.

Дифференциальная энтропия может интерпретироваться как мера **неопределенности** случайной величины, или, в более узком смысле, мера ее **равномерности**. Таким образом, использование принципа максимальной энтропии для выбора наихудшего байесовского точечного засорения позволяет формализовать недостаток априорной информации о засорении, т.е. наихудшее засорение находится как максимально неопределенное.

Этот принцип уже использован нами при формулировке ОМУ как оценки при максимально неопределенном байесовском точечном засорении. Действительно, равномерное распределение с $s(y,\theta)\sim 1$ максимизирует дифференциальную энтропию при отсутствии дополнительных ограничений.

Для получения семейства оценок мы будем использовать дополнительное ограничение на **дивергенцию Кульбака–Лейблера** между распределением засоряющей точки (с плотностью $s(y,\theta)$) и модельным распределением (с плотностью $f(y,\theta)$). Ограничение имеет вид

$$D(s,f)\le\Delta, \tag{2.5.3}$$

где $D(s,f)=\int\limits_Y s(y,\theta)\,\ln\dfrac{s(y,\theta)}{f(y,\theta)}\,dy$ – дивергенция Кульбака–Лейблера, Δ – неотрицательная величина.

Дивергенция Кульбака–Лейблера является частным случаем **вероятностного расстояния** – характеристики «непохожести» случайных явлений, описываемых двумя распределениями вероятностей.

Дивергенция Кульбака–Лейблера не является расстоянием, однако всегда неотрицательна и равна нулю тогда и только тогда, когда $s = f$. Хотя, как правило, $D(s, f) \neq D(f, s)$.

Таким образом, ограничение (2.5.3) $D(s, f) \leq \Delta$ задает некоторую окрестность модельного распределения, в которой плотности s до некоторой степени «похожи» на плотность f.

Оптимальная плотность имеет вид

$$s_*(y, \theta) = k\, f(y, \theta)^{1-\delta}, \tag{2.5.7}$$

где $k = e^{\delta b - 1}$ и $\delta = 1/(a+1)$, причем $0 \leq \delta \leq 1$.

Заметим, что для получения s **не используется параметрическая информация**, как это и требуется в теории робастности.

Критерий качества $U_s(\psi)$ с точностью до несущественного сомножителя имеет вид

$$U_\delta(\psi) = \int_Y IF^2(y,\psi,F)\, f^{1-\delta}(y,\theta)dy = \frac{\int_Y \psi^2(y,\theta) f^{1-\delta}(y,\theta)\, dy}{d^2} = \frac{\mathbf{E}\left[\psi^2 / f^\delta\right]}{d^2} \quad (2.5.8)$$

При $\delta = 0$ получаем асимптотическую дисперсию

$$U_0(\psi, F) = V(\psi, F), \quad (2.5.9)$$

при $\delta = 1$ получаем меру неустойчивости

$$U_1(\psi, F) = W(\psi, F), \quad (2.5.10)$$

в диапазоне $0 < \delta < 1$ получаем некоторые промежуточные решения.

Согласно теореме 2.1 минимум функционалу $U_\delta(\psi)$ доставляет оценочная функция

$$\psi_\delta(y,\theta) = c\left[\frac{\partial}{\partial\theta}\ln f(y,\theta) + \beta\right] f^\delta(y,\theta).$$

Оценки с $\delta = 1/2$ А.М. Шурыгин назвал **радикальными**. Поэтому, подчеркивая связь с радикальными, мы и называем данные оценки **обобщенными радикальными**.

Радикальные оценки имеют особую важность среди обобщенных радикальных оценок. Для них имеют место следующий результат.

Теорема 2.8. Пусть θ – параметр сдвига. Тогда, если функция $f(y,\theta)$ непрерывна при $y \in Y$ и имеет равные значения в граничных точках интервала Y, то для радикальной оценочной функции справедливо равенство

$$\mathrm{eff}\ \psi_{1/2} = \mathrm{stb}\ \psi_{1/2}. \tag{2.19}$$

Для других параметров и условий радикальная оценка не обязательно дает равенство (2.19).

Таким образом, с точки зрения наличия свойства (2.18) $\mathrm{eff}\ \psi = \mathrm{stb}\ \psi$ радикальные оценки близки к равнооптимальным оценкам. И те, и другие представляют собой «золотую середину» между ОМП и ОМУ. При этом равнооптимальные оценки, как условно оптимальные, будут иметь бóльшие значения eff и stb.

Пример 2.8. Для модели нормального распределения $N(\mu,\sigma^2)$ оценки, названные впоследствии обобщенными радикальными, ввел Л.Д. Мешалкин. В случае оцениваемого математического ожидания μ и известной дисперсии σ^2 оценочная функция для оценки μ_δ имеет вид

$$\psi_\delta(y,\mu_\delta)=(y_i-\mu_\delta)\exp\left[-\delta\frac{(y_i-\mu_\delta)^2}{2\sigma^2}\right].$$

Введенная выше стойкая оценка совпадает с радикальной оценкой и является, таким образом, частным случаем оценок Мешалкина.

Пример 2.9. В модели нормального распределения $N(\mu,\sigma^2)$ радикальная оценка $\mu_{1/2}$ имеет характеристики $\mathrm{eff}=\mathrm{stb}\approx 0{,}84$ и, таким образом, хуже равнооптимальных на 1%.

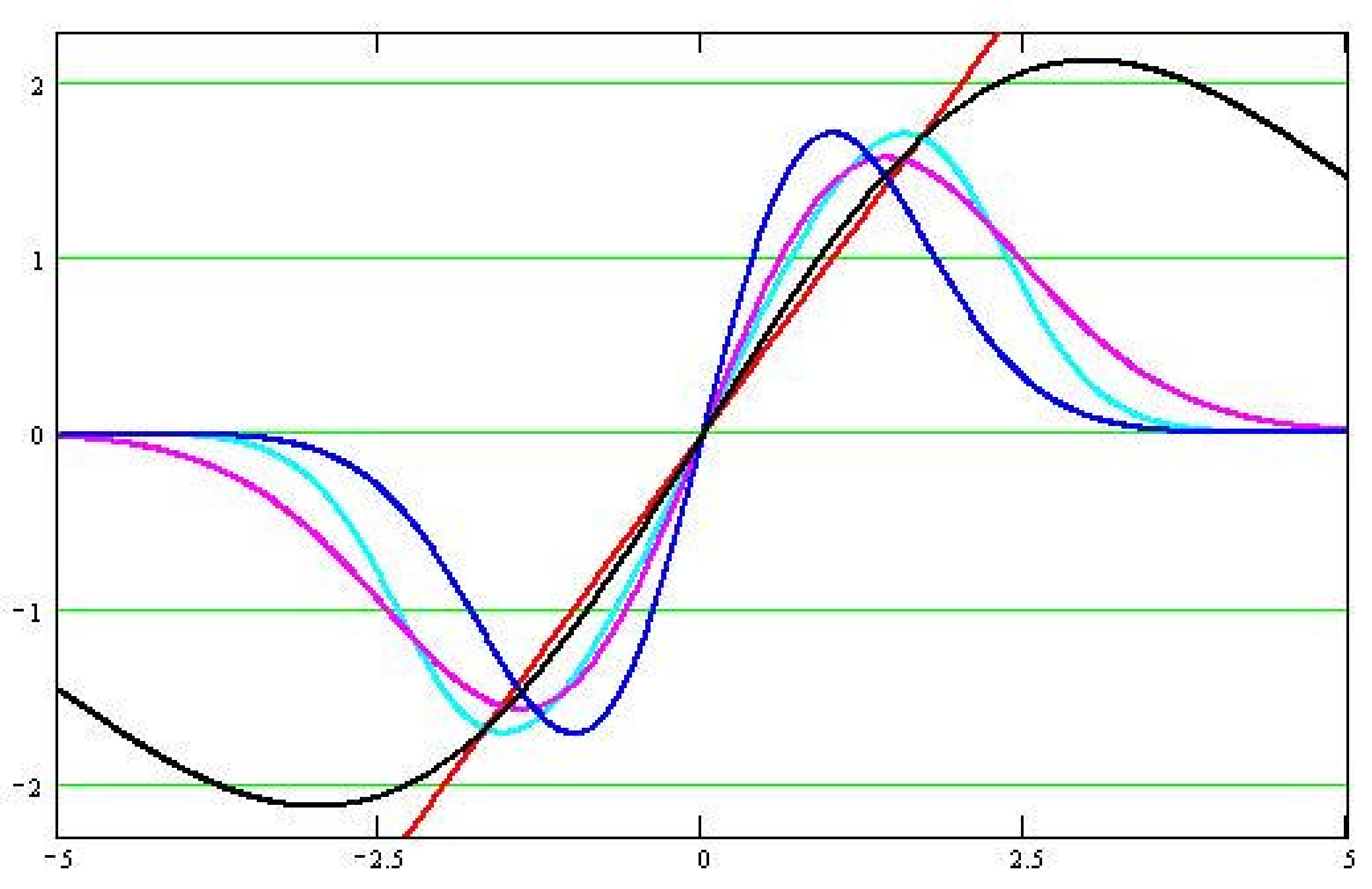
Пример 2.10. В таблицу 2.1 сведены ранее полученные результаты об оценках математического ожидания нормального распределения $N(\mu, \sigma^2)$.

Таблица 2.1

Оценки параметра сдвига нормального распределения и их характеристики

Оценка		eff	stb
ОМП	μ_0	1	0
ОМУ	$\mu_* = \mu_1$	0,65	1
медианная	μ_{med}	0,64	0
стойкая, радикальная	$\mu_s = \mu_{1/2}$	0,84	0,84
равнооптимальная	μ_{eq}	0,85	0,85
компромиссная	μ_c	0,82	0,92

В подходе Шурыгина наибольшую устойчивость имеет ОМУ, в подходе Хампеля – медианная. Сравним их: обе оценки имеют примерно равную эффективность, но с точки зрения показателя устойчивости оценки имеют противоположные значения.



Стандартное нормальное распределение, оценивается параметр μ. Линии: красная – ОМП ($\delta = 0$), синяя – ОМУ ($\delta = 1$), бирюзовая – равнооптимальная, лиловая – радикальная–стойкая ($\delta = 0.5$), черная – $\delta = 1/9 \approx 0.11$ (значение, рекомендованное Л.Д. Мешалкиным, $\mathrm{eff} \approx 0{,}99$, $\mathrm{stb} \approx 0{,}22$).