Министерство образования и науки Российской Федерации
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Д.В. ЛИСИЦИН

УСТОЙЧИВЫЕ МЕТОДЫ ОЦЕНИВАНИЯ ПАРАМЕТРОВ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ

Утверждено
Редакционно-издательским советом университета
в качестве учебного пособия

НОВОСИБИРСК
2013

УДК 519.246.2(075.8)
Л 632

Рецензенты:

д-р техн. наук, профессор *В.И. Денисов*
д-р техн. наук, профессор *В.С. Тимофеев*

Работа подготовлена на кафедре прикладной математики для магистрантов ФПМиИ, обучающихся по направлению 010400.68 – «Прикладная математика и информатика»

Лисицин Д.В.
Л 632 Устойчивые методы оценивания параметров статистических моделей : учеб. пособие / Д.В. Лисицин. – Новосибирск: Изд-во НГТУ, 2013. – 76 с.

ISBN 978-5-7782-2196-3

Рассматриваются устойчивые методы оценивания параметров моделей по статистическим данным. Основное внимание уделяется адаптивным и робастным методам оценивания неизвестных параметров распределений. Рассмотрены как традиционные подходы П. Хьюбера и Ф. Хампеля к робастному оцениванию, так и подход А.М. Шурыгина, связанный с байесовским точечным засорением.

Учебное пособие предназначено для магистрантов, обучающихся по направлению «Прикладная математика и информатика». Оно будет полезно аспирантам и научным работникам, разрабатывающим или использующим статистические методы анализа данных.

УДК 519.246.2(075.8)

ISBN 978-5-7782-2196-3

ВВЕДЕНИЕ

В основе классических статистических процедур лежат некоторые предположения о свойствах наблюдений, например, их независимости, однородности (одинаковой распределенности). Известный метод максимального правдоподобия требует знания закона распределения наблюдений. На практике некоторые предположения могут не выполняться, они могут выступать в роли идеализации, удобного приближения реальной ситуации. В результате соответствующие методы теряют оптимальность. Для решения этой проблемы разработаны различные подходы, приводящие к устойчивым процедурам. Описанию ряда подходов к устойчивому оцениванию параметров распределений и посвящено данное учебное пособие.

Пособие состоит из двух разделов, в первом из которых отражены традиционные подходы к устойчивому оцениванию (адаптивное оценивание, в частности L_ν -оценивание, подходы к робастному оцениванию П. Хьюбера и Ф. Хампеля). Во втором разделе излагается новый подход, введенный А.М. Шурыгиным и связанный с байесовским точечным засорением. Заметим, что изложение последнего подхода ведется оригинально с привлечением идей подхода Ф. Хампеля.

Ряд результатов, отраженных в учебном пособии, получен автором единолично или в соавторстве с канд. техн. наук К.В. Гавриловым; работа над значительной частью этих результатов происходила при поддержке гранта Президента РФ (№ МД-2690.2008.9), Министерства образования и науки РФ (аналитическая ведомственная целевая программа «Развитие научного потенциала высшей школы (2006-2008 гг.)», код проекта РНП-2.1.2.43), гранта администрации Новосибирской области (договор № ФГМ-4-05).

Учебное пособие предназначено для магистрантов ФПМиИ, обучающихся по направлению 010400.68 – «Прикладная математика и информатика», в рамках курса «Современные проблемы прикладной математики и информатики». Оно будет полезно также аспирантам и научным работникам, разрабатывающим или использующим статистические методы анализа данных.

1. ТРАДИЦИОННЫЕ ПОДХОДЫ К УСТОЙЧИВОМУ ОЦЕНИВАНИЮ

1.1. УСТОЙЧИВОСТЬ СТАТИСТИЧЕСКИХ ПРОЦЕДУР

1.1.1. ВВЕДЕНИЕ В ПРОБЛЕМУ УСТОЙЧИВОСТИ

Рассмотрим задачу оценивания параметра сдвига распределения вероятностей. Модель наблюдений имеет вид

$$y_i = \theta + e_i,\ i = 1,...,N, \tag{1.1}$$

где y_i – i-е наблюдение исследуемой случайной величины; θ – параметр сдвига; e_i – i-е значение случайной величины с нулевым параметром сдвига; N – количество наблюдений.

В условиях нормального распределения *среднее арифметическое* является для параметра сдвига асимптотически эффективной оценкой (оценкой по методу максимального правдоподобия – ММП-оценкой).

Отметим характерную особенность нормального распределения: основная масса распределения сосредоточена на конечном интервале (-3σ, $+3\sigma$), где σ – корень квадратный из дисперсии. Вне этого интервала находится лишь 0,27 % распределения. Другими словами, нормальное распределение имеет «легкие хвосты» [4]. Таким образом, принимая гипотезу нормальности, мы автоматически предполагаем, что основная масса наблюдений сосредоточена на некотором интервале. Вероятность большого отклонения при этом весьма мала.

В реальной ситуации эта гипотеза – чересчур жесткая: предполагаемая модель редко бывает абсолютно точно специфицированной; в наборе данных могут присутствовать *выбросы* – грубые ошибки. Выбросы могут быть результатом нарушения условия эксперимента, неправильного измерения, появления посторонних данных и т. п. Поэтому необходимо предположить, что отклонения с большой вероятностью могут принимать и большие значения.

В таких случаях более адекватны распределения с тяжелыми хвостами, а среднее арифметическое не всегда оказывается подходящей оценкой. С другой стороны, в некоторых ситуациях более адекватным будет предположение о распределении наблюдений, имеющем хвосты более легкие, чем у нормального распределения, среднее арифметическое при этом также теряет некоторые из положительных свойств.

Для распределений с тяжелыми хвостами более эффективными, чем среднее арифметическое, будут оценки, которые не меняют резко своих значений при возникновении больших отклонений. Если распределение не засорено (вероятность больших отклонений мала), устойчивые оценки будут менее эффективны, зато если наблюдения содержат выбросы, то эти оценки будут малочувствительны к ним, а потому более удовлетворительными.

Пример 1.1. Пусть имеется следующий набор данных [19]:

$$0.96 \quad 1.01 \quad 0.97 \quad 1.02 \quad 1.04 \quad 1.00 \quad 10.52.$$

Последнее число, очевидно, является выбросом.

Среднее арифметическое элементов выборки равно 2.36 и, таким образом, сильно отличается от истинного значения, равного единице.

Более устойчивой оценкой будет *усеченное среднее*, которое вычисляется следующим образом: отбрасывается некоторое количество минимальных и максимальных наблюдений в выборке. На основе оставшихся наблюдений находится среднее арифметическое.

В нашем случае усеченное среднее (при отбрасывании одного минимального и одного максимального наблюдений) равно 1.008.

Рассмотрим другую устойчивую оценку – *медиану*. Напомним, что медиана последовательности есть величина, по левую и по правую стороны от которой лежит одинаковое количество элементов. Медиана нашей последовательности равна 1.01.

Для обоснования применения методов используется принцип непрерывности или устойчивости, когда малая ошибка в математической модели не должна приводить к существенной ошибке в окончательных выводах.

Оказывается, что многие распространенные статистические процедуры не обладают свойством устойчивости.

Известно, что за редким исключением не обладают устойчивостью процедуры, основанные на методе максимального правдоподобия. Будучи оптимальными для точной модели, при отклонении от нее свойства резко ухудшаются.

Пример 1.2. Сравним две оценки разброса: $s_N = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})^2}$ – стандартное отклонение и $d_N = \frac{1}{N}\sum_{i=1}^{N}|\, y_i - \overline{y}\,|$ – средний модуль. Оценка s_N для нормального распределения является ММП-оценкой, она асимптотически эффективна, но уже при двух «плохих» наблюдениях (с дисперсией, большей в 9 раз) на тысячу преимущество (в асимптотическом смысле) имеет d_N [20].

Заметим, что в естественных науках типичные выборки «хороших» данных содержат от 1 до 10 % подобных «плохих» наблюдений [17].

1.1.2. ОСНОВНЫЕ ПОДХОДЫ К УСТОЙЧИВОМУ ОЦЕНИВАНИЮ

Точные параметрические модели – это абстракция, их выбор для описания конкретных данных часто до определенной степени произволен, поэтому истинное распределение в большей или меньшей степени отличается от модельного. Среди других важных типов отклонений назовем зависимость данных. Однако именно отклонение от модельного распределения считается наиболее важным и наиболее изучено. Его мы и будем в дальнейшем рассматривать.

Фактически приходим к следующей ситуации. Имеется выборка независимых одинаково распределенных случайных величин y_i, $i = 1,...,N$, из неизвестного распределения G с функцией $G(y)$. Выборке соответствует эмпирическая функция распределения

$$G_N(y) = \frac{1}{N}\sum_{i=1}^{N}\Delta(y_i - y),$$

где $\Delta(y)$ – функция единичного скачка.

Поскольку G нам неизвестно, мы пользуемся некоторым параметрическим семейством $\{F_\theta, \theta \in \Theta\}$, подгоняя параметр θ под данные. Вообще говоря, параметрическая «идеальная» модель также может зависеть от N: F_θ есть $F_{N,\theta}$ и при $N \to \infty$ мы, получая все больше информации, приближаем идеальную модель к истинной $F_{N,\theta} \to G$.

Наша цель состоит в том, чтобы найти такую оценку неизвестного параметра θ, которая была бы в некотором смысле «хорошей» для распределения G_N.

Поскольку точные параметрические модели не всегда оказываются устойчивыми, при построении устойчивых процедур необходимо использовать модели других типов.

Имеется три основных способа ухода от точных параметрических моделей.

1. *Непараметрические процедуры* требуют наложить на распределение лишь некоторые весьма слабые ограничения, в условиях которых возможен анализ получаемых решений (выяснение таких свойств, как нормальность, состоятельность и т. д.). Примеры ограничений – непрерывность распределения, независимость, одинаковая распределенность наблюдений. Фактически в непараметрических процедурах используется ограниченное непараметризованное множество распределений. Или, говоря иначе, используется бесконечномерный параметр. Однако непараметрические процедуры не обязательно приводят к устойчивым решениям.

2. *Адаптивные процедуры* сами подстраиваются к имеющемуся эмпирическому распределению. Цель подстройки – получение оценок с высокой эффективностью (асимптотически эффективных оценок в условиях неизвестного распределения). Тот или иной способ оценивания может выбираться при этом в зависимости от значений некоторых характеристик эмпирического распределения. Адаптивные процедуры также не всегда приводят к устойчивым решениям.

Часто адаптивные процедуры бывают по характеру непараметрическими. С другой стороны, при оценивании может использоваться конечное число параметров, за счет которых происходит адаптация, это приводит к частично-адаптивным процедурам. В дальнейшем будем рассматривать именно такие процедуры, называя их «адаптивными».

3. В *теории робастности* идея обеспечения устойчивости становится основной. Рассматривается устойчивость в (непараметрической) окрестности параметрической модели.

В теории робастности имеется несколько различных подходов, и каждый конкретный результат о робастности конкретной процедуры необходимо сопровождать указанием, в каком смысле следует робастность понимать.

В частности, надо зафиксировать следующие моменты [23].

1. Указать идеальную модель F_θ: $F(y,\theta)$. Например, $N(\mu,1)$ – нормальное распределение с оцениваемым математическим ожиданием μ и единичной дисперсией, т. е. $\theta=\mu$.

2. Описать учитываемые отклонения от модели в виде некоторой окрестности $\tilde{F}$ идеального распределения F_θ, так что ей принадлежит неизвестное истинное распределение: $G\in\tilde{F}$. Например, часто используется «окрестность» в виде ε-засоренного распределения

$$\tilde{F}=\tilde{F}_\varepsilon=\left\{G_\theta : G(y,\theta)=(1-\varepsilon)F(y,\theta)+\varepsilon H(y)\right\},$$

где $H(y)$ – произвольное (неизвестное) распределение; ε – интенсивность засорения (для этой «окрестности» используется также название «модель большой ошибки»).

3. Указать критерий качества процедуры и сформулировать требования к его поведению в заданной окрестности. Например, асимптотическая дисперсия оценки должна быть некоторым образом ограничена.

1.2. *М*-ОЦЕНКИ

В дальнейшем будем рассматривать простой и удобный вид оценок параметров (*М*-оценки) для модели одномерной непрерывной случайной величины, имеющей модельное распределение $F(y,\theta)$ с плотностью $f(y,\theta)$ и неизвестным скалярным параметром θ, причем область значений случайной величины – интервал Y – не зависит от параметра θ.

М-оценка $\hat{\theta}$ параметра θ находится как решение задачи минимизации:

$$\sum_{i=1}^{N}\rho\left(y_i,\hat{\theta}\right)=\min, \tag{1.2}$$

где ρ – функция потерь.

Альтернативное определение *М*-оценки требует решения неявного уравнения

$$\sum_{i=1}^{N}\psi\left(y_i,\hat{\theta}\right)=0, \tag{1.3}$$

где ψ – оценочная функция.

Обозначив $\psi(y,\theta)=\rho'_\theta(y,\theta)$, переходим от (1.2) к (1.3). Однако при невыпуклой функции потерь может иметься несколько локальных минимумов, соответствующих нескольким решениям системы (1.3). По этой причине два указанных определения *M*-оценки не эквивалентны.

Отметим, что одна и та же *M*-оценка может быть задана эквивалентными оценочными функциями, отличающимися сомножителями, не зависящими от данных, т. е. оценочная функция $c(\theta)\psi(y,\theta)$, $c(\theta)\neq 0$, задает ту же оценку, что $\psi(y,\theta)$.

При некоторых условиях регулярности [20, 23] *M*-оценка будет:

- асимптотически несмещенной, т. е. $\mathbf{E}\hat{\theta}\to\theta$ при $N\to\infty$, где $\mathbf{E}$ – символ математического ожидания;
- состоятельной;
- асимптотически нормальной.

Найдем условия асимптотической несмещенности оценки [18]. Пусть функция ψ непрерывно дифференцируема по θ. Разлагая ψ в ряд Тейлора в окрестности θ, получим из (1.3)

$$\sum_{i=1}^{N}\psi(y_i,\theta)+(\hat{\theta}-\theta)\sum_{i=1}^{N}\psi'_\theta(y_i,\theta)\approx 0,$$

откуда

$$\hat{\theta}-\theta\approx-\frac{\sum_{i=1}^{N}\psi(y_i,\theta)}{\sum_{i=1}^{N}\psi'_\theta(y_i,\theta)}=-\frac{\frac{1}{N}\sum_{i=1}^{N}\psi(y_i,\theta)}{\frac{1}{N}\sum_{i=1}^{N}\psi'_\theta(y_i,\theta)}.$$

Если случайные величины $\psi(y_i,\theta)$, $\psi'_\theta(y_i,\theta)$ имеют математические ожидания, то при $N\to\infty$ числитель и знаменатель правой части последнего выражения сходятся почти наверное к своим математическим ожиданиям:

$$\hat{\theta}-\theta\to-\frac{\mathbf{E}\psi(y_i,\theta)}{\mathbf{E}\psi'_\theta(y_i,\theta)}=-\frac{\int_Y\psi(y,\theta)dF(y,\theta)}{\int_Y\psi'_\theta(y,\theta)dF(y,\theta)}. \tag{1.4}$$

Исходя из представления (1.4) оценка будет асимптотически несмещенной, если не равен нулю знаменатель и равен нулю числитель:

$$\int_Y \psi(y,\theta)dF(y,\theta)=0 \tag{1.5}$$

или

$$\int_Y \psi(y,\theta)f(y,\theta)dy=0\,. \tag{1.6}$$

Дифференцируя (1.6) по θ, получим

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

откуда

$$\int_Y \psi'_\theta(y,\theta)f(y,\theta)dy=-\int_Y \psi(y,\theta)f'_\theta(y,\theta)dy\,. \tag{1.7}$$

Величина $\sqrt{N}(\hat{\theta}-\theta)$ асимптотически нормальна с нулевым математическим ожиданием и дисперсией

$$V(\psi,F)=\frac{\mathbf{E}\psi^2(y,\theta)}{d^2}\,, \tag{1.8}$$

где $d=d(\theta)=\frac{\partial}{\partial t}\mathbf{E}\,\psi(y,t)\big|_{t=\theta}=\mathbf{E}\psi'_\theta(y,\theta)$ (для выполнения последнего равенства требуются дополнительные условия).

Величина $V(\psi,F)$ называется асимптотической дисперсией. Функционал $V(\psi)=V(\psi,F)$ достигает минимума на оценочной функции

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

где $c=c(\theta)\neq 0$, соответствующей ММП-оценке. При этом функция потерь имеет вид $\rho_0(y,\theta)=-\ln f(y,\theta)$. Таким образом, выбирая оценочную функцию ψ, отличную от оценочной функции ММП-оценки, мы увеличиваем асимптотическую дисперсию в идеальной модели, но можем при должном выборе ψ приобрести свойство устойчивости.

Воспользовавшись формулой (1.7), мы можем записать асимптотическую дисперсию в виде

$$V(\psi, F) = \frac{\int\limits_Y \psi^2(y,\theta)dF(y,\theta)}{\left[\int\limits_Y \psi(y,\theta) f'_\theta(y,\theta)dy\right]^2} = \frac{\int\limits_Y \psi^2(y,\theta)dF(y,\theta)}{\left[\int\limits_Y \psi(y,\theta)\psi_0(y,\theta)dF(y,\theta)\right]^2} =$$

$$= \frac{\mathbf{E}\psi^2(y,\theta)}{\left[\mathbf{E}\psi(y,\theta)\psi_0(y,\theta)\right]^2}$$

и $d = \mathbf{E}\psi(y,\theta)\psi_0(y,\theta)$.

1.3. АДАПТИВНОЕ ОЦЕНИВАНИЕ

1.3.1. ОБЩИЕ ПРИНЦИПЫ ПОЛУЧЕНИЯ АДАПТИВНЫХ ОЦЕНОК

При адаптивном оценивании конкретный вид статистической процедуры выбирается на основе оценки некоторых характеристик неизвестной функции распределения наблюдений, т. е. по имеющейся выборке. Таким образом, процедура оценивания адаптируется к свойствам имеющейся выборки.

В качестве характеристик, которыми можно пользоваться для адаптации, назовем тяжесть хвостов и асимметрию распределения.

В качестве меры асимметрии может использоваться коэффициент асимметрии

$$\beta_3 = \frac{\gamma_3}{\sigma^3},$$

где γ_3 – центральный момент третьего порядка, σ – квадратный корень из дисперсии.

Для симметричного распределения коэффициент асимметрии равен нулю. Если коэффициент асимметрии положителен, то распределение имеет положительную (правую) асимметрию, а если отрицателен – то отрицательную (левую). Для унимодального распределения, имеющего плотность, в случае положительной (отрицательной) асимметрии более тяжелая часть плотности находится правее (левее) моды.

Для измерения тяжести хвостов используются разные меры (само понятие тяжести хвостов может быть определено по-разному [23]). В ряде случаев может быть использован показатель островершинности – эксцесс

$$\beta_4 = \frac{\gamma_4}{\sigma^4},$$

где γ_4 – центральный момент четвертого порядка. Для нормального распределения он равен 3, если он больше 3, то распределение имеет более острую вершину, чем нормальное, а если меньше 3 – то более плоскую вершину.

Для построения адаптивной оценки необходимо вычислить выборочные значения коэффициента асимметрии и эксцесса и на их основе выбрать некоторую функцию потерь.

Например, пусть распределение считается симметричным и необходимо оценить параметр сдвига, тогда можно построить следующую, впрочем, весьма грубую процедуру: если эксцесс больше 3, то воспользоваться медианой (как более устойчивой оценкой), в противном случае – средним арифметическим.

Заметим, однако, что выборочные значения моментов высоких порядков, в том числе третьего и четвертого, могут оказаться неустойчивыми оценками соответствующих теоретических характеристик. По этой причине в адаптивных процедурах часто используются иные оценки показателей асимметрии и тяжести хвостов.

Другой, более современный, подход к адаптивному оцениванию состоит в построении ММП-оценок при использовании некоторого параметризованного по форме семейства распределений, независимо от того, принадлежит ли реальное распределение этому семейству (подобные оценки называют квазиправдоподобными). Благодаря наличию параметра формы имеется возможность для методов оценивания адаптироваться к свойствам ошибок измерений. Среди таких семейств можно назвать двустороннее экспоненциальное распределение и распределение Стьюдента.

Пример 1.3. Двустороннее экспоненциальное (ДСЭ) распределение имеет плотность

$$f(y, \mu, \lambda, \nu) = \frac{1}{2\lambda\Gamma(1+1/\nu)} \exp\left[-\left(\frac{|y-\mu|}{\lambda}\right)^{\nu}\right],$$

где μ – параметр сдвига; $\lambda > 0$ – параметр масштаба; $\nu > 0$ – параметр формы; Γ – гамма-функция. Параметр масштаба λ в общем случае определяется соотношением $F(y,\lambda)=F(y/\lambda,1)$ и в случае нормального распределения совпадает с корнем квадратным из дисперсии.

Заметим, что несмотря на широкое применение, общепринятого названия у распределения нет, его называют также обобщенным гауссовским, экспоненциально-степенным, обобщенным распределением ошибок и др.

Впервые это распределение описал в одной из своих ранних работ [29] М.Ф. Субботин (советский астроном, директор Института теоретической астрономии АН СССР, член-корреспондент АН СССР), в связи с чем данное распределение называют также распределением Субботина.

Распределение является симметричным (коэффициент асимметрии нулевой); его частные случаи: нормальное при $\nu=2$, Лапласа при $\nu=1$, предельное при $\nu\to\infty$ – равномерное распределение.

Распределение имеет следующие характеристики:

- математическое ожидание μ;
- дисперсию $\sigma^2=\lambda^2\dfrac{\Gamma(3/\nu)}{\Gamma(1/\nu)}$;
- центральный абсолютный момент q-го порядка (не обязательно целого) $m_q=\int\limits_{-\infty}^{\infty}|y-\mu|^q\ f(y)dy=\lambda^q\,\Gamma\left(\dfrac{q+1}{\nu}\right)\Big/\Gamma\left(\dfrac{1}{\nu}\right)$;
- эксцесс $\beta_4=\dfrac{m_4}{m_2^2}=\dfrac{\Gamma(5/\nu)\Gamma(1/\nu)}{\left[\Gamma(3/\nu)\right]^2}$.

Чем больше значение параметра формы, тем более плоской будет вершина и тем более легкими будут хвосты.

Часто данное распределение можно считать адекватным для описания ошибок измерений, встречающихся на практике. Так, в одном исследовании погрешностей измерительных приборов оказалось, что ДСЭ-распределение адекватно в 50 % случаев [15].

Пример 1.4. Распределение Стьюдента имеет плотность

$$f(y,\mu,\lambda,\nu)=\frac{\nu^{\nu/2}\Gamma\left((\nu+1)/2\right)}{\pi^{1/2}\lambda\Gamma\left(\nu/2\right)}\left[\nu+\frac{(y-\mu)^2}{\lambda^2}\right]^{-\frac{\nu+1}{2}},$$

где μ – параметр сдвига; $\lambda>0$ – параметр масштаба; $\nu>0$ – параметр формы.

Распределение имеет следующие характеристики:

- математическое ожидание существует при $\nu > 1$ и равно μ;
- дисперсия существует при $\nu > 2$ и равна $\lambda^2 \dfrac{\nu}{\nu - 2}$;
- эксцесс существует при $\nu > 4$ и равен $\dfrac{3(\nu - 2)}{\nu - 4}$.

Чем больше значение параметра формы, тем более легкими становятся хвосты распределения. Распределение является симметричным. Его частный случай при $\nu = 1$ – это распределение Коши, предельное при $\nu \to \infty$ – нормальное распределение. Таким образом, распределение имеет хвосты, более тяжелые, чем у нормального.

Распределение Стьюдента часто используют на практике, поскольку ММП-оценки его параметров сочетают в себе свойства адаптивности и робастности [6].

1.3.2. АДАПТИВНЫЕ L_ν-ОЦЕНКИ

1.3.2.1. L_ν-ОЦЕНКИ И ИХ СВОЙСТВА

Пусть модель имеет вид (1.1) с величинами e_i, подчиненными двустороннему экспоненциальному распределению с параметром масштаба λ и параметром формы ν.

Запишем логарифм функции правдоподобия, взятый со знаком «минус» (с точностью до несущественной константы):

$$Q(\tilde{\theta}, \tilde{\lambda}, \tilde{\nu}) = \frac{1}{\tilde{\lambda}^{\tilde{\nu}}} \sum_{i=1}^{N} \left| r_i(\tilde{\theta}) \right|^{\tilde{\nu}} + N \ln\left[\tilde{\lambda} \Gamma\left(1 + 1/\tilde{\nu}\right) \right], \tag{1.9}$$

где $r_i(\tilde{\theta}) = y_i - \tilde{\theta}$ – остаток для i-го измерения при параметре сдвига, равном $\tilde{\theta}$.

В условиях известного значения параметра формы ν ММП-оценка параметра θ находится в результате минимизации функции вида [4]

$$Q(\tilde{\theta}) = \sum_{i=1}^{N} \left| r_i(\tilde{\theta}) \right|^{\nu}. \tag{1.10}$$

Минимум функции (1.10) задает *М*-оценку с функцией потерь $\rho(y-\theta)=\left|y-\theta\right|^{\nu}$. Полученные в результате оценки носят название L_{ν}-оценок [4], поскольку фактически минимизируют по $\tilde{\theta}$ вектор остатков в L_{ν}-метрике.

Заметим, что при $\nu=2$ L_{ν}-оценка совпадает со средним арифметическим. При $\nu=1$ мы получаем другую широко распространенную оценку – медиану.

L_{ν}-оценки параметра θ являются несмещенными при $\nu \geq 1$ и математическом ожидании наблюдений, равном θ [4].

L_{ν}-оценки параметра θ обладают еще одним важным свойством – масштабной инвариантностью, т. е. они не зависят от параметра масштаба [см. формулу (1.10)]. Это позволяет проводить оценивание параметра θ независимо от оценивания параметра масштаба λ.

После получения оценок параметра θ можно оценить параметр масштаба λ. Логарифм функции правдоподобия, взятый со знаком «минус», с точностью до постоянных, несущественных при оценивании параметра λ, имеет вид

$$Q(\tilde{\lambda})=\frac{1}{\tilde{\lambda}^{\nu}}\sum_{i=1}^{N}|\,r_i(\hat{\theta})\,|^{\nu}+N\ln\tilde{\lambda}\,.$$

Приравнивая его производную нулю в точке $\tilde{\lambda}=\hat{\lambda}$ и разрешая полученное уравнение относительно неизвестного, найдем

$$\hat{\lambda}=\left[\frac{\nu}{N}\sum_{i=1}^{N}|\,r_i(\hat{\theta})\,|^{\nu}\right]^{\frac{1}{\nu}}. \qquad (1.11)$$

Таким образом, налицо еще одно преимущество L_{ν}-оценок: явная формула для оценок параметра масштаба.

Заметим, ММП-оценка параметра λ совпадает с оценкой по методу моментов, в котором приравниваются теоретическое и выборочное значения m_{ν}.

Чтобы построить адаптивную оценку параметра θ, необходимо оценить также и параметр формы. Соответствующее уравнение правдоподобия имеет вид

$$\frac{1}{\hat{\lambda}^{\hat{\nu}}}\sum_{i=1}^{N}\left[\left|r_i(\hat{\theta})\right|^{\hat{\nu}}\ln\left(\frac{\left|r_i(\hat{\theta})\right|}{\hat{\lambda}}\right)\right]-\frac{N}{\hat{\nu}^2}DG\left(1+\frac{1}{\hat{\nu}}\right)=0\,, \qquad (1.12)$$

где $DG(z)=\dfrac{d\ln\Gamma(z)}{dz}$ – диграмма-функция. Уравнение является нелинейным и требует применения соответствующих методов поиска корня.

1.3.2.2. ВЫЧИСЛЕНИЕ L_ν-ОЦЕНОК

При построении вычислительных алгоритмов оценивания будем использовать покомпонентный метод, который выгодно применять, когда фиксация некоторых элементов множества оцениваемых параметров существенно упрощает максимизацию функции правдоподобия относительно других элементов. Согласно данному методу множество параметров разбивается на ряд подмножеств. Решение находится итеративно, при этом итерация включает несколько этапов, каждый из которых состоит в нахождении очередного приближения оценок для одного подмножества параметров из соответствующей подсистемы уравнений правдоподобия при фиксированных значениях параметров остальных подмножеств.

В нашей задаче будем считать каждый параметр отдельным подмножеством. Таким образом, одна итерация будет состоять из трех этапов: на первом находится очередное приближение для параметра θ при фиксированных остальных параметрах, на втором – для параметра λ, на третьем – для параметра ν.

Рассмотрим вычисление оценки параметра сдвига.

Уравнение правдоподобия имеет вид

$$\hat{\nu}\sum_{i=1}^{N}\left|r_i(\hat{\theta})\right|^{\hat{\nu}-1}\operatorname{sign}\left[r_i(\hat{\theta})\right]=0$$

или

$$\hat{\nu}\sum_{i=1}^{N}\left|r_i(\hat{\theta})\right|^{\hat{\nu}-2} r_i(\hat{\theta}) = 0 .$$

Обозначая

$$w(y-\hat{\theta}) = \left|y-\hat{\theta}\right|^{\hat{\nu}-2} \qquad (1.13)$$

и отбрасывая несущественный сомножитель $\hat{\nu}$, получим

$$\sum_{i=1}^{N} w\left[r_i(\hat{\theta})\right]\left[y_i-\hat{\theta}\right] = 0 ,$$

откуда

$$\hat{\theta} = \sum_{i=1}^{N} w\left[r_i(\hat{\theta})\right] y_i \Big/ \sum_{i=1}^{N} w\left[r_i(\hat{\theta})\right] .$$

Таким образом, оценка параметра сдвига имеет вид взвешенного среднего, однако вес сам зависит от значения оценки.

Для вычисления оценки воспользуемся методом последовательных приближений, на s-й итерации которого определяется очередное приближение $\hat{\theta}^{(s)}$ в виде

$$\hat{\theta}^{(s)} = \sum_{i=1}^{N} w\left[r_i(\hat{\theta}^{(s-1)})\right] y_i \Big/ \sum_{i=1}^{N} w\left[r_i(\hat{\theta}^{(s-1)})\right] . \qquad (1.14)$$

Однако итеративный процесс сходится, если весовая функция $w(y-\hat{\theta})$ будет невозрастающей при $(y-\hat{\theta}) \geq 0$ [4, 14].

Рассмотрим случай, когда очередное приближение параметра формы $\hat{\nu}^{(s-1)} > 2$. Здесь сходимость не гарантируется. С целью обеспечения сходимости очередное полученное по формуле (1.14) приближение оценки, которое обозначим $\hat{\theta}_0^{(s)}$, модифицируют следующим образом [4]:

$$\hat{\theta}^{(s)} = (1-\alpha)\hat{\theta}^{(s-1)} + \alpha^{(s)}\hat{\theta}_0^{(s)} ,$$

где $\hat{\theta}^{(s)}$ – окончательное приближение оценки на s-й итерации, $\alpha^{(s)} = 1\big/(\hat{\nu}^{(s-1)}-1)$.

Рассмотрим случай $\hat{\nu}^{(s-1)} < 2$. Условие сходимости выполняется, однако весовая функция (1.13) в окрестности нуля является неограниченно возрастающей. В этом случае весовую функцию в окрестности нуля необходимо ограничить [14]:

$$w\left(y-\hat{\theta}\right)=\begin{cases}\left|y-\hat{\theta}\right|^{\hat{\nu}-2}, & \left|y-\hat{\theta}\right|>\varepsilon,\\ \left|\varepsilon\right|^{\hat{\nu}-2}, & \left|y-\hat{\theta}\right|\le\varepsilon,\end{cases}$$

где ε – малая положительная константа.

Минимизируемая функция (1.10) при $\nu > 1$ имеет единственный локальный минимум, который совпадает с глобальным. Для $0 < \nu < 1$ это неверно и минимизируемая функция может иметь несколько локальных минимумов. Для $\nu = 1$ функция может иметь неединственный глобальный минимум.

Функция (1.9) является многоэкстремальной и при $\nu > 1$. В связи с этим большое значение имеет выбор начального приближения. Хорошим начальным приближением для $\hat{\theta}$ здесь будет медиана.

Этап оценивания параметра λ заключается в вычислении очередного приближения по явной формуле (1.11). Этап оценивания параметра ν состоит в решении уравнения (1.12) и является итеративным.

1.3.2.3. АЛЬТЕРНАТИВНЫЕ СПОСОБЫ ОЦЕНИВАНИЯ ПАРАМЕТРА ФОРМЫ

Параметр сдвига и параметр масштаба ДСЭ-распределения с фиксированным параметром формы обычно оценивают методом максимального правдоподобия.

Однако вопрос об оценивании параметра формы остается открытым [27]. Отмечается, что использовать ММП при оценивании параметра формы целесообразно при выборках объемом $N > 50...100$. Необходимость использования альтернативных методов продиктована как стремлением уменьшить трудоемкость вычислений, так и низким качеством ММП-оценки при малом числе наблюдений. Заметим также, что выпуклость функции правдоподобия по параметру формы не гарантирована.

Рассмотрим такие методы, в которых в системе уравнений правдоподобия уравнение, соответствующее параметру формы, заменяется некоторым другим оценочным уравнением. В большинстве случаев в оценочном уравнении приравниваются теоретические и выборочные значения некоторой характеристики распределения.

Вычисление оценок в таких случаях также производится покомпонентным методом, причем вместо уравнения правдоподобия для параметра ν на третьем этапе решается альтернативное уравнение.

Основной метод – это использование оценочного уравнения, приравнивающего теоретическое и выборочное значения эксцесса [15]:

$$\beta_4 = \frac{m_4}{m_2^2}.$$

Оценочное уравнение имеет вид

$$\beta_4(\hat{\nu}) = \hat{\beta}_4, \tag{1.15}$$

где $\beta_4(\hat{\nu})$ – теоретическое значение эксцесса ДСЭ-распределения с параметром формы $\hat{\nu}$; $\hat{\beta}_4 = \dfrac{\hat{m}_4}{\hat{m}_2^2}$ – выборочное значение эксцесса, а выборочные значения моментов вычисляются по формуле

$$\hat{m}_q = \frac{1}{N}\sum_{i=1}^{N}\left|r_i(\hat{\theta})\right|^q.$$

При вычислении значения оценки параметра формы вместо поиска корня уравнения (1.15) в качестве его приближенного значения можно использовать аппроксимирующее выражение [15]

$$\hat{\nu} = 1,46\left[\ln\left(\hat{\beta}_4 - 2/9 - 10,7/\hat{\beta}_4^7\right) - 0,289\right]^{-1}.$$

Оно обеспечивает в диапазоне $0,25 \le \nu \le 4$ погрешность не более $\pm 0,3\,\%$, для $4 \le \nu \le 7$ – около 1 %, для $10 \le \nu \le 20$ – до $\pm 6\,\%$, а при $\beta_4 = 1,8$, соответствующем равномерному распределению, вместо $\nu = \infty$ дает значение $\nu = 29$. Такое ДСЭ-распределение уже мало отличается от равномерного (по значению эксцесса на 2,2 %). Поэтому данная формула может с успехом использоваться на практике.

Другая характеристика, используемая при оценивании, названа обобщенным эксцессом q-го порядка и имеет вид [27]

$$\beta_q^{(1)} = \frac{m_{2q}}{m_q^2}.$$

При $q = 2$ получаем традиционный эксцесс $\beta_2^{(1)} = \beta_4$.

Для двустороннего экспоненциального распределения q-й обобщенный эксцесс $\beta_q^{(1)}$ определяется формулой

$$\beta_q^{(1)}(\nu) = \frac{\Gamma\big((2q+1)/\nu\big)\Gamma\big(1/\nu\big)}{\Big[\Gamma\big((q+1)/\nu\big)\Big]^2}.$$

Оценивание можно производить, приравнивая теоретическое и выборочное значения этой величины:

$$\beta_q^{(1)}(\hat{\nu}) = \hat{\beta}_q^{(1)}. \tag{1.16}$$

В [27, 28] используется значение $q = \hat{\nu}$, а также описываются приемы решения получаемого уравнения.

Можно ввести еще один обобщенный эксцесс q-го порядка ($q \neq 2$) [16]:

$$\beta_q^{(2)} = \frac{m_q}{\sigma^q} = \frac{m_q}{m_2^{q/2}}.$$

При $q = 4$ получаем традиционный эксцесс.

Для двустороннего экспоненциального распределения $\beta_q^{(2)}$ определяется формулой

$$\beta_q^{(2)}(\nu) = \frac{\Gamma\big((q+1)/\nu\big)}{\Gamma\big(1/\nu\big)}\left[\frac{\Gamma\big(1/\nu\big)}{\Gamma\big(3/\nu\big)}\right]^{q/2}.$$

Оценивание можно производить, приравнивая теоретическое и выборочное значения этой величины:

$$\beta_q^{(2)}(\hat{\nu}) = \hat{\beta}_q^{(2)}. \tag{1.17}$$

Для случая тяжелых хвостов распределений используется значение $q = 1$ [16, 27].

В более широком диапазоне свойств распределений предпочтительным представляется использование значения $q = \hat{\nu}$ [7, 16]. Рассмотрим этот случай подробнее.

Уравнение имеет вид

$$\beta_{\hat{\nu}}^{(2)}(\hat{\nu}) = \hat{\beta}_{\hat{\nu}}^{(2)},$$

где $\beta_{\nu}^{(2)}(\nu) = \frac{1}{\nu}\left[\frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}\right]^{\nu/2}$.

Тогда имеем оценочное уравнение

$$\frac{1}{\hat{\nu}}\left[\frac{\Gamma(1/\hat{\nu})}{\Gamma(3/\hat{\nu})}\right]^{\hat{\nu}/2} = \frac{\frac{1}{N}\sum_{i=1}^{N}|r_i(\hat{\theta})|^{\hat{\nu}}}{\left[\frac{1}{N}\sum_{i=1}^{N} r_i^2(\hat{\theta})\right]^{\hat{\nu}/2}}. \tag{1.18}$$

Это же уравнение можно получить другим способом. Воспользуемся для оценивания параметра формы методом моментов. Приравнивая теоретическую и выборочную дисперсии, получаем

$$\hat{\lambda}^2 \frac{\Gamma(3/\hat{\nu})}{\Gamma(1/\hat{\nu})} = \frac{1}{N}\sum_{i=1}^{N} r_i^2(\hat{\theta}).$$

Подставляя в уравнение ММП-оценку параметра λ (1.11), получим

$$\left[\frac{\hat{\nu}}{N}\sum_{i=1}^{N}|r_i(\hat{\theta})|^{\hat{\nu}}\right]^{2/\hat{\nu}} \frac{\Gamma(3/\hat{\nu})}{\Gamma(1/\hat{\nu})} = \frac{1}{N}\sum_{i=1}^{N} r_i^2(\hat{\theta}).$$

Легко увидеть, что данное уравнение эквивалентно уравнению (1.18).

Заметим, что уравнение (1.18) наследует от метода моментов (при использовании ММП-оценки параметра λ) тривиальный корень, равный порядку момента, в нашем случае $\hat{\nu} = 2$.

Решение уравнения (1.18) трудоемко с вычислительной точки зрения, поскольку на итерациях требуется пересчитывать суммы N элементов. Упростить вычисления и одновременно избавиться от тривиального корня можно, применяя метод расщепления [5]. Согласно ему вместо исходного уравнения (1.18) с неизвестным $\hat{\nu}$ будем решать покомпонентно систему уравнений

$$\begin{cases} \beta_q^{(2)}(p) = \hat{\beta}_q^{(2)}, \\ q = p \end{cases}$$

с двумя неизвестными p и q, по смыслу совпадающими с параметром формы ($\hat{\nu} = q = p$), уточняя сначала первый, а затем второй параметр (первое уравнение системы совпадает с уравнением (1.17)) [7]. Эту покомпонентную процедуру удобно встроить в исходную покомпонентную процедуру, в которой каждая итерация становится четырехэтапной. Однако четвертый этап выделять нет смысла, если третий этап представить для s-й итерации покомпонентной процедуры в виде решения уравнения

$$\beta_{\hat{\nu}^{(s-1)}}^{(2)}(\hat{\nu}^{(s)}) = \hat{\beta}_{\hat{\nu}^{(s-1)}}^{(2)}.$$

Заметим, что подобный прием может применяться и для решения уравнения (1.16) при $q = \hat{\nu}$, в частности можно использовать процедуру

$$\beta_{\hat{\nu}^{(s-1)}}^{(1)}(\hat{\nu}^{(s)}) = \hat{\beta}_{\hat{\nu}^{(s-1)}}^{(1)}.$$

1.4. МИНИМАКСНЫЙ ПОДХОД

Перейдем к рассмотрению подходов к робастному оцениванию. И начнем с минимаксного подхода, который хронологически был предложен первым, его автор – П. Хьюбер.

Минимаксный подход основан на использовании наилучшей оценки в наихудшей точке окрестности. Критерием качества при этом служит либо асимптотическое смещение (точнее, его модуль), либо асимптотическая дисперсия.

В первом случае имеем задачу

$$\inf_{\psi} \sup_{G \in \tilde{F}} |b(\psi, G)|, \tag{1.19}$$

где $b(\psi, G)$ – асимптотическое смещение, во втором –

$$\inf_{\psi} \sup_{G \in \tilde{F}} V(\psi, G), \tag{1.20}$$

причем в обоих случаях функция ψ удовлетворяет условию (1.6).

1.4.1. МИНИМАКСНОЕ СМЕЩЕНИЕ

Рассмотрим первую задачу [18]. В качестве окрестности выберем ε-засоренное распределение, точнее, реальное распределение будет иметь функцию

$$G(y, \theta) = (1-\varepsilon) F(y, \theta) + \varepsilon H(y, \theta). \tag{1.21}$$

Подставляя выражение для G в формулу (1.4) и применяя формулы (1.5) и (1.7), получим

$$b(\psi, G) = -\frac{(1-\varepsilon) \int\limits_Y \psi(y, \theta) dF(y, \theta) + \varepsilon \int\limits_Y \psi(y, \theta) dH(y, \theta)}{(1-\varepsilon) \int\limits_Y \psi'_\theta(y, \theta) dF(y, \theta) + \varepsilon \int\limits_Y \psi'_\theta(y, \theta) dH(y, \theta)} =$$

$$= \frac{\varepsilon \int\limits_Y \psi(y, \theta) dH(y, \theta)}{(1-\varepsilon) \int\limits_Y \psi(y, \theta) f'_\theta(y, \theta) dy - \varepsilon \int\limits_Y \psi'_\theta(y, \theta) dH(y, \theta)}. \tag{1.22}$$

Таким образом, смещение оценки имеет порядок ε. Будем предполагать, что ε конечно, но мало, и будем искать наилучшую оценочную функцию с точностью до членов порядка ε.

Тогда, пренебрегая членами порядка ε в знаменателе, получим

$$b(\psi, G) = b(\psi, H) \approx \varepsilon \frac{\int\limits_Y \psi(y, \theta) dH(y, \theta)}{\int\limits_Y \psi(y, \theta) f'_\theta(y, \theta) dy}. \tag{1.23}$$

Возвращаясь к минимаксной задаче (1.19), имеем

$$\inf_{\psi} \sup_{H} | b(\psi, H) |.$$

Легко видеть, что наихудшее засорение H, дающее наибольшее смещение, сосредоточено в точках, где достигается наибольшее по модулю значение ψ :

$$\sup_H \left| \int_Y \psi(y,\theta)\, dH(y,\theta) \right| = \sup_y |\psi(y,\theta)|.$$

Тогда оптимальная оценочная функция $\psi(y,\theta)$ есть решение задачи

$$\inf_\psi \varepsilon \frac{\sup_y |\psi(y,\theta)|}{\left| \int_Y \psi(y,\theta) f'_\theta(y,\theta)\, dy \right|} \tag{1.24}$$

при ограничении (1.6).

Поскольку одну и ту же *М*-оценку задает бесконечное множество эквивалентных оценочных функций $\psi(y,\theta)$, решение задачи не единственно. Поэтому введем следующую нормировку:

$$\sup_y |\psi(y,\theta)| = 1. \tag{1.25}$$

Таким образом, переходим к задаче

$$\sup_\psi \left| \int_Y \psi(y,\theta) f'_\theta(y,\theta)\, dy \right|$$

при ограничениях (1.6) и (1.25). Модуль можно опустить, так как если $\psi(y,\theta)$ – решение, то и $-\psi(y,\theta)$ – тоже решение. В результате имеем задачу

$$\sup_\psi \int_Y \psi(y,\theta) f'_\theta(y,\theta)\, dy \tag{1.26}$$

при ограничениях (1.6) и (1.25).

Введем функцию Лагранжа для учета ограничения (1.6):

$$\int_Y \psi(y,\theta) f'_\theta(y,\theta)\, dy - \beta(\theta) \int_Y \psi(y,\theta) f(y,\theta)\, dy =$$

$$= \int_Y \psi(y,\theta) \left[f'_\theta(y,\theta) - \beta(\theta) f(y,\theta) \right] dy.$$

Супремум этой функции при ограничении (1.25) достигается на функции $\psi(y,\theta)$, которая равна 1 или –1 в зависимости от того, положительно или отрицательно выражение в квадратных скобках:

$$\psi(y,\theta)=\operatorname{sign}\left[f'_\theta(y,\theta)-\beta(\theta)f(y,\theta)\right]=\operatorname{sign}\left[f'_\theta(y,\theta)/f(y,\theta)-\beta(\theta)\right].$$

Оценку с такой оценочной функцией будем называть *медианной*. Функция $\beta(\theta)$ определяется из условия (1.6):

$$\int_Y \operatorname{sign}\left[f'_\theta(y,\theta)/f(y,\theta)-\beta(\theta)\right]f(y,\theta)dy=0 .$$

Последнее уравнение определяет β как медиану случайной величины $\psi_0(\xi,\theta)=f'_\theta(\xi,\theta)/f(\xi,\theta)$, где ξ – изучаемая случайная величина.

Найдем максимальное асимптотическое смещение. Согласно (1.24), (1.25), (1.6) имеем

$$\sup_H |b(\psi,H)|\approx\frac{\varepsilon}{\left|\int_Y \psi(y,\theta)f'_\theta(y,\theta)dy\right|}=\frac{\varepsilon}{\left|\int_Y \psi(y,\theta)\left[f'_\theta(y,\theta)-\beta(\theta)f(y,\theta)\right]dy\right|}=$$

$$=\frac{\varepsilon}{\left|\int_Y \operatorname{sign}(f'_\theta(y,\theta)-\beta(\theta)f(y,\theta))\left[f'_\theta(y,\theta)-\beta(\theta)f(y,\theta)\right]dy\right|}=$$

$$=\frac{\varepsilon}{\int_Y |f'_\theta(y,\theta)-\beta(\theta)f(y,\theta)|\,dy}. \qquad (1.27)$$

Замечание. Оптимальная функция ψ как функция θ является разрывной, следовательно, отброшенный в знаменателе (1.22) член порядка ε может привести к существенному ухудшению оценки. Однако в рассматриваемом случае этого не происходит. В частности, доказано, что при некоторых ограничениях на плотность $f(y,\theta)$ с точностью до членов порядка ε^2 смещение, большее, чем (1.27), в общем случае невозможно [18].

Пример 1.5. Задана нормальная модель $F_\theta : N(\mu, \sigma^2)$ с известной дисперсией σ^2 и оцениваемым математическим ожиданием μ. Плотность распределения имеет вид

$$f(y, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(y-\mu)^2}{2\sigma^2} \right\}. \tag{1.28}$$

В нашем случае $f(y, \mu) = f(y, \mu, \sigma)$ и

$$f'_\mu(y, \mu) = f(y, \mu) \frac{y-\mu}{\sigma^2},$$

$$\psi(y, \mu) = \operatorname{sign}\left[\frac{f'_\mu(y, \mu)}{f(y, \mu)} - \beta(\mu) \right] = \operatorname{sign}\left[\frac{y-\mu}{\sigma^2} - \beta(\mu) \right].$$

Поскольку $f(y, \mu) = f(y-\mu)$ – четная функция, а $\psi(y, \mu) = \psi(y-\mu)$ при $\beta(\mu) = 0$ – нечетная, при $\beta(\mu) = 0$ условие (1.6) выполняется. Следовательно, $\beta(\mu) = 0$. Окончательно, отбрасывая несущественный знаменатель, имеем

$$\psi(y, \mu) = \operatorname{sign}[y - \mu],$$

$\hat{\mu} = \operatorname*{med}_i y_i$ – выборочная медиана.

Пример 1.6. Задана нормальная модель $F_\theta : N(\mu, \sigma^2)$ с плотностью (1.28), известным математическим ожиданием μ и оцениваемым параметром σ. Имеем $f(y, \sigma) = f(y, \mu, \sigma)$ и

$$f'_\sigma(y, \sigma) = -\frac{f(y, \sigma)}{\sigma} + f(y, \sigma) \frac{(y-\mu)^2}{\sigma^3} = f(y, \sigma) \left[\frac{(y-\mu)^2}{\sigma^3} - \frac{1}{\sigma} \right],$$

$$\psi(y, \sigma) = \operatorname{sign}\left[\frac{(y-\mu)^2}{\sigma^3} - \frac{1}{\sigma} - \beta(\sigma) \right].$$

Слагаемое $\beta(\sigma)$ является медианой случайной величины $\frac{(\xi-\mu)^2}{\sigma^3}-\frac{1}{\sigma}$, т. е.

$$\int\limits_{\frac{(y-\mu)^2}{\sigma^3}-\frac{1}{\sigma}\le\beta} \frac{1}{\sqrt{2\pi}\cdot\sigma}\exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}dy=\frac{1}{2}.$$

Сделаем замену переменных в интеграле $x=\frac{y-\mu}{\sigma}$, получим

$$\int\limits_{\frac{x^2-1}{\sigma}\le\beta}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\int\limits_{-\sqrt{1+\beta\sigma}}^{\sqrt{1+\beta\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=2\int\limits_{0}^{\sqrt{1+\beta\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\frac{1}{2},$$

следовательно,

$$\int\limits_{0}^{\sqrt{1+\beta\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\frac{1}{4},\qquad \int\limits_{-\infty}^{\sqrt{1+\beta\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\frac{3}{4},$$

$\sqrt{1+\beta\sigma}$ – 75-процентная квантиль стандартного нормального распределения, т. е. $\sqrt{1+\beta\sigma}=a\approx 0{,}67449$, $\beta(\sigma)=\frac{a^2-1}{\sigma}$. Тогда имеем

$$\psi(y,\sigma)=\operatorname{sign}\left[\frac{(y-\mu)^2}{\sigma^3}-\frac{1}{\sigma}-\frac{a^2-1}{\sigma}\right]=\operatorname{sign}\left[\frac{(y-\mu)^2}{a^2}-\sigma^2\right]. \quad (1.29)$$

Отсюда $\sum\limits_{i=1}^{N}\operatorname{sign}\left[\frac{(y_i-\mu)^2}{a^2}-\hat{\sigma}^2\right]=0$, $\hat{\sigma}^2$ – выборочная медиана случайной величины $\frac{(\xi-\mu)^2}{a^2}$, следовательно, имеет место

$$\hat{\sigma}=\operatorname*{med}_i\frac{|y_i-\mu|}{a}. \quad (1.30)$$

Такая оценка называется MAD-оценкой (медианой абсолютного отклонения).

Пример 1.7. Распределение Коши с оцениваемым параметром сдвига μ и известным параметром масштаба λ. Функция плотности имеет вид

$$f(y,\mu,\lambda)=\frac{1}{\pi\lambda}\frac{1}{1+(y-\mu)^2/\lambda^2}=\frac{1}{\pi}\frac{\lambda}{\lambda^2+(y-\mu)^2}.$$

В этом случае $f(y,\mu)=f(y,\mu,\lambda)$. Рассмотрим случай $\lambda=1$.

Распределение Коши – это частный случай распределения Стьюдента при $\nu=1$ (см. пример 1.4). Распределение симметрично относительно μ, но случайная величина не имеет математического ожидания. Оценка μ в виде выборочного среднего несостоятельна, ММП приводит к состоятельной и устойчивой оценке с оценочной функцией

$$\psi(y,\mu)=\frac{y-\mu}{1+(y-\mu)^2}.$$

Найдем оценочную функцию для оценки с минимаксным смещением:

$$\psi(y,\mu)=\mathrm{sign}\left[\frac{f'_\mu(y,\mu)}{f(y,\mu)}-\beta(\mu)\right]=\mathrm{sign}\left[\frac{2(y-\mu)}{1+(y-\mu)^2}\right]=\mathrm{sign}\left[y-\mu\right].$$

Таким образом, $\hat{\mu}$ – выборочная медиана.

Найдем наибольшее смещение оценки $\hat{\mu}$. Найдем знаменатель в формуле (1.27):

$$\int_{-\infty}^{\infty}|f'_\theta(y,\mu)-\beta(\mu)f(y,\mu)|\,dy=\int_{-\infty}^{\infty}|f'_\theta(y,\mu)|\,dy=\int_{-\infty}^{\infty}\left|\frac{1}{\pi}\frac{2(y-\mu)}{\left[1+(y-\mu)^2\right]^2}\right|dy=$$

$$=\frac{2}{\pi}\int_{-\infty}^{\infty}\frac{|x|}{\left[1+x^2\right]^2}dx=\frac{2}{\pi}\int_{0}^{\infty}\frac{2x}{\left[1+x^2\right]^2}dx=\frac{2}{\pi}\int_{0}^{\infty}\frac{d(1+x^2)}{\left[1+x^2\right]^2}=\frac{2}{\pi}\left[\frac{-1}{1+x^2}\right]\Bigg|_0^{\infty}=\frac{2}{\pi}.$$

По ходу преобразований в интеграле сделана такая же замена переменных, как в примере 1.6.

В результате $\sup\limits_H|b(\psi,H)|\approx\frac{\pi\varepsilon}{2}$.

Пример 1.8. Распределение Коши с известным параметром сдвига и оцениваемым параметром масштаба. В этом случае $f(y,\lambda)=f(y,\mu,\lambda)$. Рассмотрим случай $\mu=0$. Найдем оценочную функцию для оценки с минимаксным смещением.

Имеем

$$f'_\lambda\left(y,\lambda\right)=\frac{1}{\pi}\frac{y^2-\lambda^2}{\left[\lambda^2+y^2\right]^2},\quad \frac{f'_\lambda\left(y,\lambda\right)}{f\left(y,\lambda\right)}=\frac{1}{\lambda}\frac{y^2-\lambda^2}{\lambda^2+y^2},$$

$$\psi(y,\lambda)=\operatorname{sign}\left[\frac{f'_\lambda\left(y,\lambda\right)}{f\left(y,\lambda\right)}-\beta(\lambda)\right]=\operatorname{sign}\left[\frac{1}{\lambda}\frac{y^2-\lambda^2}{\lambda^2+y^2}-\beta(\lambda)\right],$$

где $\beta(\lambda)$ есть медиана случайной величины $\frac{1}{\lambda}\frac{\xi^2-\lambda^2}{\lambda^2+\xi^2}$.

Определим $\beta(\lambda)$. Имеем

$$\int\limits_{\frac{1}{\lambda}\frac{y^2-\lambda^2}{\lambda^2+y^2}\le\beta}\frac{1}{\pi\lambda}\frac{1}{1+y^2/\lambda^2}\,dy=\frac{1}{2}.$$

Обозначим $z=y/\lambda$, тогда получим

$$\int\limits_{\frac{1}{\lambda}\frac{z^2-1}{1+z^2}\le\beta}\frac{1}{\pi}\frac{1}{1+z^2}\,dz=\frac{1}{2}.$$

При $1-\lambda\beta\ge0$ имеем $z^2\le\frac{\lambda\beta+1}{1-\lambda\beta}$ и

$$\int\limits_{-\sqrt{\frac{\lambda\beta+1}{1-\lambda\beta}}}^{\sqrt{\frac{\lambda\beta+1}{1-\lambda\beta}}}\frac{1}{\pi}\frac{1}{1+z^2}\,dz=\frac{1}{2}. \tag{1.31}$$

При $1-\lambda\beta<0$ имеем $z^2 \geq \dfrac{\lambda\beta+1}{1-\lambda\beta}$ и

$$1-\int\limits_{-\sqrt{\frac{\lambda\beta+1}{1-\lambda\beta}}}^{\sqrt{\frac{\lambda\beta+1}{1-\lambda\beta}}} \frac{1}{\pi}\frac{1}{1+z^2}dz=\frac{1}{2},$$

следовательно, в этом случае также справедливо (1.31).

Далее, имеем

$$\int\limits_{-\sqrt{\frac{\lambda\beta+1}{1-\lambda\beta}}}^{\sqrt{\frac{\lambda\beta+1}{1-\lambda\beta}}} \frac{1}{\pi}\frac{1}{1+z^2}dz=2\int\limits_{0}^{\sqrt{\frac{\lambda\beta+1}{1-\lambda\beta}}} \frac{1}{\pi}\frac{1}{1+z^2}dz=\frac{1}{2},$$

откуда

$$\int\limits_{0}^{\sqrt{\frac{\lambda\beta+1}{1-\lambda\beta}}} \frac{1}{\pi}\frac{1}{1+z^2}dz=\frac{1}{4}, \qquad \int\limits_{-\infty}^{\sqrt{\frac{\lambda\beta+1}{1-\lambda\beta}}} \frac{1}{\pi}\frac{1}{1+z^2}dz=\frac{3}{4}.$$

Следовательно, $a=\sqrt{\dfrac{\lambda\beta+1}{1-\lambda\beta}}$ – 75-процентная квантиль стандартного распределения Коши, равная $a=1$. Имеем $\sqrt{\dfrac{\lambda\beta+1}{1-\lambda\beta}}=1$, откуда $\beta=0$, т. е. специальная коррекция оценочной функции для выполнения условия (1.6) не требуется.

Окончательно получаем

$$\psi(y,\lambda)=\operatorname{sign}\left[\frac{1}{\lambda}\frac{y^2-\lambda^2}{\lambda^2+y^2}\right]=\operatorname{sign}\left[y^2-\lambda^2\right].$$

1.4.2. МИНИМАКСНАЯ ДИСПЕРСИЯ

1.4.2.1. ОЦЕНИВАНИЕ ПАРАМЕТРА СДВИГА

Перейдем к обсуждению задачи (1.20) [18]. Рассмотрим решение задачи для оценивания параметра сдвига. В этом случае функция и плотность распределения имеют вид $F(y,\theta) = F(y-\theta)$ и

$$f(y,\theta) = f(y-\theta) \tag{1.32}$$

соответственно.

Решение задачи состоит в том, чтобы найти наименее благоприятное распределение, т. е. распределение с функцией

$$G^* = \arg\sup_{G\in\tilde{F}} V(\psi_0, G), \tag{1.33}$$

где ψ_0 – оценочная функция ММП-оценки, тогда ММП-оценка при распределении G^* и будет решением рассматриваемой задачи. Обоснованием данного пути для класса ε-засоренных распределений служит следующая теорема Хьюбера.

Теорема 1.1. Пусть $\tilde{F}$ – класс ε-засоренных распределений

$$\tilde{F} = \left\{G : G(y-\theta) = (1-\varepsilon)F(y-\theta) + \varepsilon H(y-\theta)\right\},$$

где ε – известная фиксированная интенсивность засорения, распределение F симметрично, имеет дважды непрерывно дифференцируемую плотность $f(y)$, для которой $-\ln f(y)$ – выпуклая функция, распределение H симметрично, имеет ограниченную плотность, $Y = R^1$.

В этих условиях справедливо следующее.

1. Функционал $V(\psi, G)$ имеет седловую точку, т. е. существуют такая функция распределения $G^*(y) = (1-\varepsilon)F(y) + \varepsilon H^*(y)$ и такая оценочная функция ψ^*, что

$$V(\psi^*, G) \le V(\psi^*, G^*) \le V(\psi, G^*),$$

а G таково, что $\int\limits_{-\infty}^{\infty} \psi(y)dG(y) = 0$.

2. Пусть $y_0 < y_1$ – концы интервала, где $\left|\frac{f'(y)}{f(y)}\right| \le k$, причем один или оба конца могут обратиться в бесконечность, тогда наименее благоприятная плотность распределения имеет вид

$$g^*(y) = \begin{cases} (1-\varepsilon) f(y_0) e^{k(y-y_0)}, \ y \le y_0, \\ (1-\varepsilon) f(y), \ y_0 \le y \le y_1, \\ (1-\varepsilon) f(y_1) e^{-k(y-y_1)}, \ y \ge y_1, \end{cases}$$

а величина $k = k(\varepsilon)$ определяется из условия нормировки плотности $g^*(y)$, которое имеет вид

$$\frac{1}{1-\varepsilon} = \int_{y_0}^{y_1} f(y)dy + \frac{f(y_0) - f(y_1)}{k}.$$

Таким образом, в центральной части наименее благоприятная плотность совпадает (с точностью до константы) с плотностью модельного распределения, а хвосты являются лапласовскими.

3. Функция $\psi^*(y) = -\frac{\left[g^*(y)\right]'}{g^*(y)}$ монотонна, ограниченна и имеет вид

$$\psi^*(y) = \begin{cases} -k, \ y \le y_0, \\ -\frac{f'(y)}{f(y)}, \quad y_0 \le y \le y_1, \\ k, \ y \ge y_1. \end{cases}$$

4. Асимптотическая дисперсия оценки имеет вид

$$V(\psi^*, G^*) = \frac{(1-\varepsilon)\int_{-\infty}^{\infty}\left[\psi^*(y)\right]^2 f(y)dy + \varepsilon k^2}{\left\{(1-\varepsilon)\int_{-\infty}^{\infty}\left[\psi^*(y)\right]' f(y)dy\right\}^2}.$$

Теорему Хьюбера можно применять к различным симметричным распределениям, получая для них наименее благоприятную плотность и соответствующую оценочную функцию.

Пример 1.9. Для нормального распределения с плотностью (1.28) имеем

$$g^*\left(\frac{y-\mu}{\sigma}\right)=\begin{cases}\dfrac{1-\varepsilon}{\sqrt{2\pi}\cdot\sigma}\exp\left\{\dfrac{k^2}{2}-\dfrac{k}{\sigma}|\,y-\mu\,|\right\}, \dfrac{|\,y-\mu\,|}{\sigma}\ge k,\\ \dfrac{1-\varepsilon}{\sqrt{2\pi}\cdot\sigma}\exp\left\{-\dfrac{(y-\mu)^2}{2\sigma^2}\right\}, \dfrac{|\,y-\mu\,|}{\sigma}\le k,\end{cases}$$

$$\psi^*\left(\frac{y-\mu}{\sigma}\right)=\begin{cases}-k, \dfrac{y-\mu}{\sigma}\le -k,\\ \dfrac{y-\mu}{\sigma}, \left|\dfrac{y-\mu}{\sigma}\right|\le k,\\ k, \dfrac{y-\mu}{\sigma}\ge k.\end{cases}$$

Получаемая в результате оценка носит название оценки Хьюбера.

Заметим, что помимо класса ε-засоренных распределений рассматривают также классы распределений, лежащих в ε-окрестности модельного распределения относительно некоторой метрики (например, Колмогорова, Леви).

1.4.2.2. ОЦЕНИВАНИЕ ПАРАМЕТРА МАСШТАБА

Перейдем к задаче оценивания параметра масштаба. Рассмотрим семейство распределений с параметром масштаба σ, для упрощения положим параметр сдвига нулевым. Плотность имеет вид

$$f(y,\sigma)=\frac{1}{\sigma}f\left(\frac{y}{\sigma}\right), \tag{1.34}$$

где $f(y)$ – плотность стандартного распределения с единичным масштабом.

Рассмотрим *M*-оценки параметра масштаба σ в классе засоренных нормальных распределений, определяемые путем решения уравнения

$$\sum_{i=1}^{N} \chi\left(y_i / \hat{\sigma}\right) = 0, \tag{1.35}$$

где χ – оценочная функция.

Оценка, соответствующая минимаксной дисперсии, находится с помощью перехода к величинам $z_i = \ln|y_i|$ и использования теории, разработанной для параметра сдвига.

Решение имеет вид

$$\chi^*(y) = \begin{cases} y_0^2 - 1, |y| \le y_0, \\ y^2 - 1, y_0 \le |y| \le y_1, \\ y_1^2 - 1, |y| \ge y_1, \end{cases}$$

где y_0 и y_1 – границы интервала, для которого $\left|\frac{f'_\sigma(y,\sigma)}{f(y,\sigma)}\right|_{\sigma=1} = \left|y^2 - 1\right| \le k$,

а k связана с ε соотношением

$$\frac{1}{2(1-\varepsilon)} = \int_{y_0}^{y_1} f(y)dy + \frac{y_0 f(y_0) - y_1 f(y_1)}{k}.$$

Для сравнения найдем ММП-оценку σ^2 исходя из плотности $g^*\left(\frac{y-\mu}{\sigma}\right)$, наименее благоприятной при оценивании параметра сдвига. Получаем

$$-\ln g^*\left(\frac{y-\mu}{\sigma}\right) \sim \begin{cases} \ln\sigma + \frac{k}{\sigma}|y-\mu|, \ \frac{|y-\mu|}{\sigma} \ge k, \\ \ln\sigma + \frac{(y-\mu)^2}{2\sigma^2}, \ \frac{|y-\mu|}{\sigma} \le k. \end{cases}$$

Тогда соответствующая оценочная функция имеет вид

$$\chi\left(\frac{y-\mu}{\sigma}\right)=\begin{cases}\dfrac{1}{\sigma}-\dfrac{k}{\sigma^2}|y-\mu|, & \dfrac{|y-\mu|}{\sigma}\ge k,\\ \dfrac{1}{\sigma}-\dfrac{(y-\mu)^2}{2\sigma^3}, & \dfrac{|y-\mu|}{\sigma}\le k.\end{cases}$$

Полученная оценочная функция, хотя и возрастает медленнее оценочной функции ММП-оценки при нормальном распределении для больших значений наблюдений, однако не будет ограниченной. Следовательно, большие значения наблюдений будут приводить к большому изменению оценки (хотя и меньшему, чем для ММП-оценки). Данная ситуация вполне объяснима: при получении наименее благоприятной плотности мы ориентировались на получение робастной оценки параметра сдвига, а не параметра масштаба. Оценочная функция $\chi^*(y)$ приводит к робастной оценке.

Для практического применения может использоваться упрощенная по сравнению с $\chi^*(y)$, асимптотически несмещенная при нормальном распределении оценочная функция

$$\chi\left(\frac{y-\mu}{\sigma}\right)=\begin{cases}k^2-\beta, & \dfrac{|y-\mu|}{\sigma}\ge k,\\ \dfrac{(y-\mu)^2}{2\sigma^2}-\beta, & \dfrac{|y-\mu|}{\sigma}\le k,\end{cases}$$

где параметр β определяется исходя из условия асимптотической несмещенности (1.6).

1.4.3. ОПТИМАЛЬНОСТЬ НА КЛАССЕ

К задаче поиска дисперсионно-минимаксных оценок можно прийти в более общей ситуации, не связанной напрямую с теорией робастности. В частности, нет необходимости формулировать идеальное распределение и его окрестность. Достаточно постулировать класс распределений, которому принадлежит распределение изучаемой случайной величины [3, 21, 22].

При формулировке класса используется имеющаяся априорная информация: максимальная информация позволяет зафиксировать единственное точное распределение, при уменьшении априорной информации класс распределений расширяется.

Приведем примеры классов распределений с симметричными унимодальными плотностями, конечной информацией Фишера о параметре сдвига и соответствующие им наименее благоприятные плотности (плотности распределений с наименьшей информацией Фишера о параметре сдвига). Все фигурирующие в определении класса константы предполагаются известными.

1. Класс *невырожденных распределений* определяется ограничением на значение плотности распределения при $y = 0$ вида

$$g(0) \geq 1/(2s) > 0 .$$

Это наиболее широкий класс, соответствующий случаю, когда у исследователя нет (почти) никакой информации о распределении. Наименее благоприятная плотность является лапласовской с параметром масштаба s. Оптимальная оценка параметра сдвига – медиана.

2. Класс *распределений с ограниченной дисперсией* определяется условием

$$\sigma^2 \leq \Delta .$$

Распределение Коши в него не входит. Наименее благоприятная плотность является нормальной с дисперсией Δ. Оптимальная оценка параметра сдвига – среднее арифметическое.

3. Класс *финитных распределений* определяется ограничением на значение плотности распределения вида

$$\int_{-\lambda}^{\lambda} g(y)dy = 1 .$$

Наименее благоприятная плотность для него имеет вид

$$g(y) = \begin{cases} \dfrac{1}{\lambda}\cos^2\left(\dfrac{\pi y}{2\lambda}\right), |y| \leq \lambda, \\ 0, |y| > \lambda. \end{cases}$$

Оптимальная оценочная функция параметра сдвига имеет вид

$$\psi(y)=\begin{cases}\operatorname{tg}\dfrac{\pi y}{2\lambda}, & |y|\le\lambda,\\ 0, & |y|>\lambda\end{cases}$$

и является неограниченной.

4. Класс *приближенных финитных распределений* определяется ограничением на значение плотности распределения вида

$$\int_{-\lambda}^{\lambda} g(y)dy\ge 1-\nu\,,$$

где параметр ν характеризует степень отступления от финитности. Наименее благоприятная плотность для него имеет вид

$$g(y)=\begin{cases}\dfrac{\beta_2}{\lambda}\cos^2\dfrac{\beta_1 y}{\lambda}, & |\,y\,|\le\lambda,\\ \dfrac{\beta_3}{\lambda}\exp\left(-\dfrac{\beta_4\,|\,y\,|}{\lambda}\right), & |\,y\,|\ge\lambda,\end{cases}$$

где константы β_1, ..., β_4 – неотрицательные числа, которые зависят от ν и определяются из уравнений

$$1+\beta_1\operatorname{tg}\beta_1-\nu^{-1}\cos^2\beta_1=0\,,\qquad \beta_4=2\beta_1\operatorname{tg}\beta_1\,,$$

$$\beta_3=\frac{\nu\beta_4}{2}\exp\beta_4\,,\qquad \beta_2=\frac{\beta_4}{2+\beta_4}\,.$$

Оптимальная оценочная функция параметра сдвига имеет вид

$$\psi(y)=\begin{cases}\operatorname{tg}\dfrac{\beta_1 y}{\lambda}\,, & |\,y\,|\le\lambda,\\ \operatorname{sign}(y)\operatorname{tg}\beta_1, & |\,y\,|\ge\lambda.\end{cases}$$

5. Различные классы засоренных распределений. Оптимальные оценочные функции для них задаются теоремой Хьюбера.

Заметим, что некоторые из полученных решений неустойчивы. Таким свойством обладают решения в классе распределений с ограниченной дисперсией и классе финитных распределений. Это демонстрирует различие между идеями робастности и оптимальности на классе. Второй подход не ориентирован на получение робастных решений, и все зависит от свойств класса распределений.

1.5. ЛОКАЛЬНЫЙ ПОДХОД

В минимаксном подходе рассматриваются конечные, хотя и малые, окрестности точной параметрической модели. Локальный (инфинитезимальный) подход, предложенный Ф. Хампелем, имеет дело с бесконечно малыми окрестностями.

1.5.1. ОСНОВНЫЕ ПОНЯТИЯ

1.5.1.1. КАЧЕСТВЕННАЯ РОБАСТНОСТЬ

Будем рассматривать оценки параметра θ как функционалы от функции распределения: $\hat{\theta} = \theta(F)$.

M-оценка как функционал от эмпирической функции распределения $G_N(y)$ задается уравнением (1.3), что эквивалентно

$$\int_Y \psi\left(y, \hat{\theta}_N\right) dG_N(y) = 0 , \tag{1.36}$$

при $\hat{\theta} = \hat{\theta}_N = \hat{\theta}(G_N)$, где обозначение $\hat{\theta}_N$ подчеркивает зависимость от объема выборки.

Многие оценки зависят от выборки только через эмпирическую функцию распределения. Тогда можно ввести функциональный аналог (1.36) как $\int_Y \psi\,[y, \theta(G)] dG(y) = 0$, а для модельного распределения

$$\int_Y \psi\,[y, \theta(F)] dF(y) = 0 . \tag{1.37}$$

С оценкой $\hat{\theta}_N$ связывается ее распределение с функцией $L_F(\hat{\theta}_N)$ при распределении наблюдений с функцией $F(y)$. Тогда на качественном уровне мы хотели бы, чтобы при малом изменении функции $F(y)$

функция $L_F(\hat{\theta}_N)$ также изменялась мало. Подходящей формализацией такого свойства будет непрерывность [17, 23]. Рассмотрим случай $Y = R^1$.

Последовательность оценок $\left\{\hat{\theta}_N\right\}$ называется *качественно робастной в* F, если $\forall \varepsilon > 0 \;\; \exists \delta > 0 :\; \forall G$ и $\forall N$

$$d(F, G) < \delta \Rightarrow d(L_F(\hat{\theta}_N), L_G(\hat{\theta}_N)) < \varepsilon,$$

где d – метрика Леви

$$d(F, G) = \inf\left\{\varepsilon : G(y-\varepsilon) - \varepsilon \le F(y) \le G(y+\varepsilon) + \varepsilon,\; \forall y \in R^1\right\}.$$

Последовательность оценок $\left\{\hat{\theta}_N\right\}$ называется *непрерывной в* F, если $\forall \varepsilon > 0 \;\; \exists \delta > 0$, $\exists N_0 :\; \forall n, m > N_0 \;\forall F_n,\, F_m$

$$d(F, F_n) < \delta \wedge d(F, F_m) < \delta \Rightarrow |\, \hat{\theta}(F_n) - \hat{\theta}(F_m) \,| < \varepsilon.$$

Теорема 1.2. Пусть последовательность $\left\{\hat{\theta}_N\right\}$ такая, что:

1) $\hat{\theta}_N$ – непрерывная функция в R^N для каждого N (т. е. непрерывная функция наблюдений);

2) последовательность $\left\{\hat{\theta}_N\right\}$ – непрерывная в F.

Тогда $\left\{\hat{\theta}_N\right\}$ качественно робастна в F.

M-оценка качественно робастна, если оценочная функция ограниченна и решение уравнения (1.37) единственно.

1.5.1.2. КОЛИЧЕСТВЕННАЯ РОБАСТНОСТЬ

Качественная робастность отвечает на вопрос: робастна ли оценка? Однако на практике требуется как-то измерить величину робастности. Среди мер количественной робастности в первую очередь используется *функция влияния*, которая определяется формулой

$$IF(y, \theta(F), F) = \lim_{t \to 0} \frac{\theta[(1-t)F + t\Delta_y] - \theta(F)}{t},$$

где Δ_y – функция вырожденного распределения, сосредоточенного в точке y [17]. Функция влияния определяет воздействие на оценку, оказываемое добавлением к очень большой выборке одного наблюдения в точке y. В результате она отражает асимптотическое смещение оценки, вызываемое засорением наблюдений.

На основе функции влияния вводятся следующие характеристики робастности: чувствительность к большой ошибке, чувствительность к локальному сдвигу и точка удаления [17, 23].

Чувствительность к большой ошибке определяется формулой

$$\gamma^* = \sup_y |IF(y, \theta(F), F)|.$$

Данная величина измеряет наибольшее (приближенное) влияние на значение оценки небольшого засорения фиксированного объема. Поэтому с ней можно соотносить верхнюю границу для (нормированного) асимптотического смещения оценки. Желательно, чтобы величина γ^* была конечна, и если это так, то оценка θ называется *B-робастной* в F.

Чувствительность к локальному сдвигу. Данная характеристика связана с малыми флуктуациями в наблюдениях. При округлении и группировке наблюдения случайной величины слегка искажаются и оценка претерпевает определенные изменения. Эффект (приближенный и нормированный), вызываемый заменой наблюдения y_1 близким ему наблюдением y_2, можно оценить величиной *чувствительности к локальному сдвигу*

$$\lambda^* = \sup_{y_1 \neq y_2} |IF(y_1, \theta(F), F) - IF(y_2, \theta(F), F)| / |y_1 - y_2|.$$

Желательна конечность данной характеристики. Отметим, однако, что даже бесконечное значение λ^* из-за нормировки может относиться к очень небольшому изменению в действительности.

Точка удаления для симметричного модельного распределения F (с нулевым сдвигом) определяется формулой

$$\rho^* = \inf\left\{r > 0 : IF(y, \theta(F), F) = 0 \text{ при } |y| > r\right\}.$$

Наблюдения, значения которых превышают порог ρ^*, при построении оценки полностью игнорируются, что реализует жесткое правило отбраковки резко выделяющихся наблюдений.

1.5.1.3. ПОРОГОВАЯ ТОЧКА

Функция влияния, будучи производной, предлагает нам локальное линеаризованное представление оценки в точке идеального распределения. Пороговая точка показывает, на каком расстоянии от модели еще можно пользоваться локальной линеаризацией.

Пороговая точка – максимально возможное отклонение, при котором смещение оценки еще остается ограниченным. Таким образом, пороговая точка является глобальной характеристикой.

Пороговая точка для множества засоренных распределений – максимальная доля засоряющих наблюдений, при которой смещение оценки еще остается ограниченным.

Другая формализация понятия пороговой точки имеет вид

$$\varepsilon^* = \sup\left\{ \varepsilon : \sup_{G: d(G,F) < \varepsilon} \left| \theta(G) - \theta(F) \right| < \infty \right\}.$$

Максимальное значение пороговой точки есть $\varepsilon^* = 1/2$.

Пример 1.10. Рассмотрим случай *M*-оценивания параметра сдвига симметричного распределения F. В этом случае естественно взять нечетную оценочную функцию ψ. Если оценочная функция ψ вдобавок ограниченна и строго монотонна, то $\varepsilon^* = 1/2$. Для неограниченной оценочной функции $\varepsilon^* = 0$.

Имеется эмпирическое правило для *M*-оценок сдвига, согласно которому линейная экстраполяция по функции влияния остается очень точной до $\varepsilon^*/4$, а до $\varepsilon^*/2$ – вполне приемлемой.

Вычисляя некоторые пороговые точки, следует иметь в виду, что «переступать порог» оценка может подчас неоднозначно. В ряде случаев возможность достижения оценкой неограниченных значений («взрыв») – это не единственная проблема. К примеру, при оценивании параметра масштаба часто необходимо в той же мере позаботиться о том, чтобы оценка была отделена от нуля (в нуле имеем «коллапс»). В результате пороговая точка должна обеспечивать выполнение условия

$$\inf_{G: d(G,F) < \varepsilon^*} \theta(G) > 0.$$

Понятия качественной робастности, функции влияния и пороговой точки Хьюбер сравнил с характеристиками устойчивости моста:

1) качественная робастность – устойчивость к малым возмущениям, влекущим за собой лишь небольшие эффекты;

2) функция влияния – измеритель величины эффектов, возникающих из-за бесконечно малых возмущений;

3) пороговая точка – величина возмущения, которое мост может выдержать, не разрушившись.

1.5.2. ФУНКЦИЯ ВЛИЯНИЯ *M*-ОЦЕНКИ

Найдем функцию влияния *M*-оценки. Подставим в уравнение (1.37) функцию распределения $F_t = (1-t)F + t\Delta_y$ вместо F и возьмем производную по t в точке $t = 0$. Имеем

$$\int\limits_Y \psi\,[z, \theta(F_t)]\, dF_t(z) = 0\,,$$

$$\left.\frac{\partial}{\partial t}\theta(F_t)\right|_{t=0} \times \int\limits_Y \left.\frac{\partial}{\partial \tau}\psi[z, \tau]\right|_{\tau=\theta(F)} dF(z) + \int\limits_Y \psi[z, \theta(F)]d(\Delta_y - F(z)) = 0\,,$$

если порядок интегрирования и дифференцирования можно менять местами. Учитывая, что $\left.\frac{\partial}{\partial t}\theta(F_t)\right|_{t=0}$ есть функция влияния, получаем

$$IF(y, \psi, F) = -\frac{\psi\,[y, \theta(F)]}{\int\limits_Y \left.\frac{\partial}{\partial \tau}\psi[z, \tau]\right|_{\tau=\theta(F)} dF(z)}$$

или, используя (1.7), при модельном распределении с фиксированным значением параметра θ:

$$IF(y, \psi, F) = \frac{\psi\,[y, \theta]}{\int\limits_Y f'_\theta(z, \theta)\psi(z, \theta)dz} =$$

$$= \frac{\psi\,[y, \theta]}{\mathbf{E}\psi(z, \theta)\psi_0(z, \theta)} = \frac{\psi\,[y, \theta]}{d}\,, \tag{1.38}$$

где ψ_0 – оценочная функция ММП-оценки, d определяется в формуле (1.8).

Таким образом, функция влияния *M*-оценки как функция первого аргумента пропорциональна оценочной функции. В результате *B*-робастная *M*-оценка должна иметь ограниченную оценочную функцию.

При некоторых условиях регулярности можно установить связь между функцией влияния и асимптотической дисперсией. Она имеет вид

$$V(\psi, F) = \mathbf{E} IF^2(y, \psi, F) = \int_Y IF^2(y, \psi, F)\,dF(y, \theta).$$

Пример 1.11. Рассмотрим две оценки параметра сдвига стандартного нормального распределения $N(0,1)$.

1. Арифметическое среднее является *M*-оценкой с $\psi(y) = y$, а для модельного распределения оно является ММП-оценкой. В этом случае $IF(y) = y$, поэтому $\gamma^* = \infty$ – оценка не *B*-робастна, $\lambda^* = 1$ – оценка нечувствительна к малому изменению наблюдений, $\rho^* = \infty$ – оценка не удаляет резко выделяющиеся наблюдения.

2. Выборочная медиана является *M*-оценкой с $\psi_{\text{med}}(y) = \text{sign}(y)$. Для нормального распределения $IF(y) = \sqrt{\pi/2}\,\text{sign}(y)$, поэтому $\gamma^* = \sqrt{\pi/2}$ – оценка *B*-робастна, $\lambda^* = \infty$ – оценка чувствительна к малому изменению наблюдений, $\rho^* = \infty$ – оценка не удаляет резко выделяющиеся наблюдения, и они оказывают известное (фиксированное) влияние.

1.5.3. ОПТИМАЛЬНЫЕ ОЦЕНКИ

Поскольку один из основных показателей качества оценивания в подходе Хампеля – это чувствительность к большой ошибке γ^*, представляет интерес оценка, наилучшая с этой точки зрения. Оценка θ^*, доставляющая минимальное значение чувствительности к большой ошибке γ^*, называется *наиболее B-робастной* [17].

Низкая чувствительность к большой ошибке обычно входит в противоречие с требованием эффективности (т. е. малой асимптотической дисперсии). Некоего компромисса позволяют достигнуть *оптимальные B-робастные оценки* – это наиболее эффективные оценки при наложении на величину γ^* ограничения сверху. Для *M*-оценок можно записать [17]:

$$\min_{\psi} V(\psi, F) \text{ при } \gamma^*(\psi, F) \le c .$$

Далее мы изучим наиболее *B*-робастные и оптимальные *B*-робастные *M*-оценки для параметров сдвига и масштаба [17].

1.5.3.1. ОПТИМАЛЬНЫЕ *M*-ОЦЕНКИ ПАРАМЕТРА СДВИГА

Будем рассматривать множество распределений, удовлетворяющих следующим условиям.

1. Распределение $F(y-\theta)$ имеет дважды дифференцируемую плотность $f(y-\theta)$, симметричную около нуля и удовлетворяющую неравенству $f > 0$ при всех $y \in Y = R^1$.

2. Отображение $\Lambda = -f' / f = (-\ln f)'$ удовлетворяет при всех $y \in Y$ неравенству $\Lambda' > 0$ и $\int_Y \Lambda' f dy = -\int_Y \Lambda f' dy < \infty$.

В частности, этим условиям удовлетворяют симметричные унимодальные распределения с конечной информацией Фишера о параметре сдвига.

Будем рассматривать класс оценочных функций, состоящий из всех вещественных функций $\psi(y-\theta)$, удовлетворяющих следующим условиям.

1. Функция ψ определена корректно и непрерывна на множестве $R^1 \setminus C(\psi)$, где множество $C(\psi)$ конечно. В каждой точке $C(\psi)$ существуют правый и левый пределы ψ, которые конечны и различны; кроме того, $\psi(-y) = -\psi(y)$, если $\{y, -y\} \subset R^1 \setminus C(\psi)$, и $\psi(y) \ge 0$ при $y \ge 0$, не принадлежащих $C(\psi)$.

2. Множество $D(\psi)$ точек, в которых ψ непрерывна, но в которых ψ' не определена или не непрерывна, конечно.

3. $\int_Y \psi^2 dF < \infty$.

4. $0 < \int\limits_Y \psi'\, dF = -\int\limits_Y \psi(y) f'(y) dy = \int\limits_Y \Lambda\psi dF < \infty$.

Тогда справедливы следующие результаты.

Наиболее B-робастной оценкой является медиана с оценочной функцией

$$\psi_{\text{med}}(y) = \text{sign}(y),$$

и для оценочной функции ψ справедливо

$$\gamma^*(\psi) \geq \frac{1}{2f(0)} = \gamma^*_{\text{med}} .$$

Асимптотическая дисперсия и чувствительность к большой ошибке, таким образом, имеют положительные нижние значения: $\min V$ – граница Рао–Крамера, достигаемая на ММП-оценке, $\min \gamma^* = \dfrac{1}{2f(0)}$ соответствует медиане. Но одновременное их уменьшение невозможно.

Пример 1.12. Для параметра сдвига нормального распределения среднее арифметическое эффективно, но не B-робастно, медиана имеет наибольшую B-робастность, но ее эффективность составляет 64 %.

В результате интерес представляют оптимальные B-робастные оценки.

Обозначим

$$\psi_b(y) = \begin{cases} -b, & \Lambda(y) < -b, \\ \Lambda(y), & -b \leq \Lambda(y) \leq b, \\ b, & \Lambda(y) > b, \end{cases}$$

где $0 < b < \overline{\Lambda} = \sup\limits_y |\Lambda(y)|$.

Фактически это оценочная функция, получаемая в теореме Хьюбера и доставляющая асимптотической дисперсии минимаксное значение.

Имеет место следующий результат.

Если функция Λ неограниченна, то оптимальные робастные оценки задаются множеством $\left\{\psi_{\text{med}}, \psi_b(0 < b < \infty)\right\}$. Если функция Λ ограниченна: $|\Lambda| < \overline{\Lambda}$, то указанное множество имеет вид

$$\left\{\psi_{\text{med}}, \psi_b(0 < b < \overline{\Lambda}), \Lambda\right\} .$$

1.5.3.2. ОПТИМАЛЬНЫЕ *М*-ОЦЕНКИ ПАРАМЕТРА МАСШТАБА

Рассмотрим задачу оценивания параметра масштаба σ для распределения с плотностью (1.34). *М*-оценка параметра σ находится как решение уравнения (1.35).

Множество распределений удовлетворяет тем же условиям, что в предыдущем пункте.

Будем рассматривать класс оценочных функций χ, удовлетворяющих следующим условиям.

1. Функция χ определена корректно и непрерывна на множестве $R^1 \setminus C(\chi)$, где $C(\chi)$ конечно. В каждой из точек $C(\chi)$ существуют правый и левый пределы χ, причем эти пределы различны. Кроме того, $\chi(-y) = \chi(y)$, если $\{y, -y\} \subset R^1 \setminus C(\chi)$ и существует такое $r > 0$, что $\chi(y) \le 0$ на $y \in (0, r)$ и $\chi(y) \ge 0$ на $y \in (r, \infty)$.

2. Множество $D(\psi)$ точек, в которых χ непрерывна, но в которых χ' либо не определена, либо разрывна, конечно.

3. $\int\limits_Y \chi dF = 0$, $\int\limits_Y \chi^2 dF < \infty$.

4. $0 < \int\limits_Y y\chi'(y)dF(y) = \int\limits_Y [y\Lambda(y) - 1]\chi(y)dF(y) < \infty$.

Пример 1.13. Рассмотрим ММП-оценку. Для нее имеем

$$\frac{\partial}{\partial\sigma}\left[-\ln\frac{1}{\sigma}f\left(\frac{y}{\sigma}\right)\right] = \frac{\partial}{\partial\sigma}\left[\ln\sigma - \ln f(y/\sigma)\right] = \left[\frac{1}{\sigma} + \frac{f'(y/\sigma)}{f(y/\sigma)}\frac{y}{\sigma^2}\right],$$

тогда оценочная функция ММП-оценки есть $\chi(y/\sigma) = \left[\dfrac{y}{\sigma}\Lambda(y/\sigma) - 1\right]$ или $\chi(y) = \left[y\Lambda(y) - 1\right]$.

Таким образом, для стандартного нормального распределения $N(0,1)$ $\chi(y) = y^2 - 1$, а константа r равна единице.

Наиболее *B*-робастной оценкой является нормированная медиана отклонений (MAD-оценка, нормированная для обеспечения асимптотической несмещенности). Оценочную функцию MAD-оценки обозначим χ_{MAD}.

Например, для нормального распределения оценочная функция MAD-оценки имеет вид (1.29).

Для оценочной функции χ справедливо неравенство

$$\gamma^*(\chi, F) \geq 1 \Big/ \left[4F^{-1}(3/4) f(F^{-1}(3/4)) \right],$$

причем равенство имеет место тогда и только тогда, когда χ эквивалентна χ_{MAD}.

Оптимальные *B*-робастные оценки задаются множеством

$$\left\{ \chi_{\text{MAD}}, \chi_b (0 < b < \infty) \right\},$$

где

$$\chi_b(y) = \begin{cases} -b, & \bar{\chi}(y) < -b, \\ \bar{\chi}(y), & -b \leq \bar{\chi}(y) \leq b, \\ b, & \bar{\chi}(y) > b, \end{cases}$$

$\bar{\chi}(y) = y\Lambda(y) - 1 - \beta_b$, β_b обеспечивает асимптотическую несмещенность оценки при модельном распределении.

1.5.4. СРАВНЕНИЕ ПОДХОДОВ ХЬЮБЕРА И ХАМПЕЛЯ

Рассмотрим для *M*-оценки модель ε-засоренного распределения. Приближенное значение асимптотического смещения согласно (1.23), (1.38) имеет вид

$$b(\psi, H) \approx \varepsilon \frac{\int\limits_Y \psi(y, \theta) dH(y, \theta)}{\int\limits_Y \psi(y, \theta) f'_\theta(y, \theta) dy} = \varepsilon \int\limits_Y IF(y, \theta, F) dH(y, \theta), \qquad (1.39)$$

откуда

$$\sup_H |b(\psi, H)| \approx \varepsilon \sup_H \left| \int\limits_Y IF(y, \theta, F) dH(y, \theta) \right| = \varepsilon \sup_y |IF(y, \theta, F)| = \varepsilon \gamma^*(\psi).$$

Минимизация левой части последней цепочки равенств дает оценку с минимаксным смещением. Минимизация правой части цепочки дает наиболее B-робастную оценку. Например, для параметра сдвига симметричного распределения в обоих случаях эта оценка – медиана (для наиболее B-робастной оценки – в условиях раздела 1.5.3.1). Таким образом, имеется связь этих двух задач в подходах Хьюбера и Хампеля.

Заметим, что связь между данными подходами проявляется также в совпадении дисперсионно-минимаксных оценок в подходе Хьюбера с оптимальными B-робастными оценками в подходе Хампеля (при некоторых условиях). Подробно причины этого совпадения рассмотрены в [17].

Многие результаты, полученные в рамках подходов Хьюбера и Хампеля, существенно опираются на симметричность засорения. Для изучения влияния асимметричного засорения на параметр сдвига в Принстонском университете было проведено обширное исследование методом Монте-Карло. Оно показало в частности неустойчивость медианы и оценки Хьюбера при оценивании параметра сдвига нормального распределения. Для решения указанной задачи был предложен ряд оценок с конечной точкой удаления $r < \infty$, когда оценочная функция является нулевой вне отрезка $[-r, r]$ (так называемые сниженные оценки). Для данного класса оценок была развита ветвь теории Хампеля.

Заметим, что сниженные оценки – не единственный класс оценок, устойчивых к асимметричному засорению. Для обеспечения устойчивости в указанном смысле достаточно использовать оценочные функции, стремящиеся к нулю при увеличении/уменьшении аргумента [2]. Эти функции не осуществляют полного удаления резко выделяющихся наблюдений, а лишь сильно уменьшают их влияние.

Пример 1.14. ММП-оценка параметра сдвига распределения Стьюдента с параметром формы ν имеет оценочную функцию

$$\psi(y) = y / (\nu + y^2)$$

и обладает устойчивостью к асимметричному засорению.

Во втором разделе пособия мы сосредоточимся на изучении нового подхода, который часто приводит к оценкам, устойчивым к асимметричному засорению.

1.6. КОНТРОЛЬНЫЕ ВОПРОСЫ И УПРАЖНЕНИЯ

1. Что такое выбросы и каковы их причины?

2. Сравните непараметрические, адаптивные и робастные процедуры.

3. Предложите конкретные примеры *М*-оценок.

4. Каковы условия асимптотической несмещенности *М*-оценок?

5. Как изменяется дисперсия *М*-оценки при росте объема выборки?

6. Приведите примеры адаптивных оценок.

7. Предложите процедуру получения адаптивных оценок параметров регрессии.

8. Сравните свойства традиционного и альтернативных способов оценивания параметра формы ДСЭ-распределения.

9. Проанализируйте свойства оптимизационной задачи (1.26).

10. С какой целью в формуле (1.30) производится деление на константу a? Проиллюстрируйте ответ на графиках оценочных функций исходной оценки и оценки, не содержащей деление на a.

11. Сравните устойчивость ММП- и медианной оценок параметров распределения Коши. Проиллюстрируйте ответ на графиках оценочных функций.

12. Объясните возможность перехода от задачи (1.20) к задаче (1.33).

13. Постройте графики оптимальных оценочных функций, даваемых теоремой Хьюбера.

14. Постройте графики оценочных функций, приведенных в разделе 1.4.2.2. Сравните свойства соответствующих оценок.

15. Постройте графики оптимальных оценочных функций для классов, приведенных в разделе 1.4.3, проанализируйте соответствующие оценки.

16. Опишите свойства оценки с ограниченной функцией влияния.

17. Постройте график зависимости оценки от интенсивности засорения в модели засоренного распределения, укажите на графике пороговую точку. Можно ли на этом графике отобразить функцию влияния или какое-то ее значение?

18. Выпишите оценочные функции наиболее *B*-робастных и оптимальных *B*-робастных оценок параметров нормального распределения, постройте графики их оценочных функций. Сравните их с оценками, полученными в рамках подхода Хьюбера.

2. УСТОЙЧИВОЕ ОЦЕНИВАНИЕ ПРИ БАЙЕСОВСКОМ ТОЧЕЧНОМ ЗАСОРЕНИИ

2.1. МОДЕЛИ ЗАСОРЕНИЯ И КРИТЕРИИ КАЧЕСТВА

Пусть реальное распределение наблюдений является ε-засоренным с функцией (1.21) и используется *M*-оценка параметра как решение уравнения (1.3).

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:

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

Обозначим смещение оценки как

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

Для дисперсии оценки справедливо представление

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

Отсюда имеем

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

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

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).

Чтобы получить асимптотически несмещенную оценку в засоренной модели, уровень засорения ε необходимо при увеличении N устремить к нулю [24]:

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

Привлекательность такой 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 [2, 10], 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 [10]

$$\tau(\psi, F, H, k) = \lim_{N\to\infty} N\ \tau_N =$$

$$= V(\psi, F) + k^2 \left[\int_Y IF(y, \psi, F) dH(y, \theta) \right]^2 . \tag{2.2}$$

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 плотность 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 засоряющей плотности. 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

$$b(\psi, H) = b(\psi, y^*) = \frac{\varepsilon}{d}\psi(y^*, \theta) = \varepsilon IF(y^*, \psi, F) \tag{2.3}$$

и показатель (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 [24].

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

$$\psi_* \le \psi(y, \theta) \le \psi^*,$$

где $\psi_* = \inf_y \psi(y, \theta)$, $\psi^* = \sup_y \psi(y, \theta)$. В том числе при точечном засорении справедливо

$$\psi_* \le \psi(y^*, \theta) \le \psi^*,$$

тогда согласно (1.38) при $d > 0$ имеем

$$\frac{\psi_*}{d} \le IF(y^*, \psi, F) \le \frac{\psi^*}{d}$$

(при $d < 0$ знаки неравенств меняются на $\ge$). При произвольном засорении имеем

$$\psi_* \le \int_Y \psi(y, \theta) h(y, \theta) dy \le \psi^*$$

и

$$\frac{\psi_*}{d} \le \int_Y IF(y, \psi, F) h(y, \theta) dy \le \frac{\psi^*}{d}.$$

Пусть функция ψ непрерывна. Из непрерывности и ограниченности ψ по теореме о среднем значении следует, что для любой плотности $h(y, \theta)$ найдется значение y^* такое, что будут равны значения показателя (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, воспользовавшись некоторой нормой функции $IF^2(y^*, \psi, F)$. Например, используя L_∞-норму, т. е. $\max_{y^*} IF^2(y^*, \psi, F)$, получим показатель (2.2) в виде [10, 11]

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

Последний показатель отражает компромисс между асимптотической дисперсией и чувствительностью к большой ошибке, и в этом смысле он связан с принципом получения оптимальных B-робастных оценок. Однако рассмотренные выше оптимальные B-робастные оценки 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 [24–26] (см. также [6]).

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

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

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

Определим маргинальное распределение наблюдений в серии выборок:

$$g(y, \theta) = (1-\varepsilon) f(y, \theta) \int_Y s(y^*, \theta) dy^* + \varepsilon \int_Y \delta(y - y^*) s(y^*, \theta) dy^* =$$

$$= (1-\varepsilon) f(y, \theta) + \varepsilon s(y, \theta).$$

Таким образом, в серии выборок распределение наблюдений является обычной смесью распределений.

2.2. ОПТИМАЛЬНЫЕ ОЦЕНКИ ПРИ БАЙЕСОВСКОМ ТОЧЕЧНОМ ЗАСОРЕНИИ

Вначале сосредоточимся на показателе неустойчивости – функции влияния.

Функция влияния согласно (2.3) пропорциональна асимптотическому смещению при точечном засорении. В свою очередь квадрат асимптотического смещения, как показатель неустойчивости, возникает при порядке малости смещения с $0 < u \le 1$. Но он может использоваться и при постоянном, но малом значении ε, если поведением асимптотической дисперсии сознательно пренебрегают.

На основе функции влияния Хампелем были введены различные меры неустойчивости. В частности, чувствительность к большой ошибке γ^* можно рассматривать как L_∞-норму функции влияния.

Выберем критерием качества (точнее, некачественности) оценки другой показатель – квадрат L_2-нормы (с весом $s(y,\theta)$) функции влияния, или, иначе, математическое ожидание (по плотности $s(y,\theta)$) квадрата функции влияния:

$$U_s(\psi) = U_s(\psi, F) = \mathbf{E}_s IF^2(y, \psi, F) = \frac{\mathbf{E}_s \psi^2(y,\theta)}{d^2} =$$

$$= \int_Y IF^2(y, \psi, F) s(y,\theta) dy \,, \tag{2.5}$$

где $\mathbf{E}_s$ – математическое ожидание по плотности $s(y,\theta)$ [9].

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.

Для получаемых в дальнейшем задач функциональной оптимизации рассмотрим технику их решения [24–26].

Будем считать выполненными следующие условия:

1) $\mathbf{E} N \cdot \left(\hat{\theta}_N - \theta\right)^2 < \infty$ при модельном распределении, имеет место (1.6);

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

Лемма 2.1. Пусть:

1) функционал K имеет вид

$$K\left(\theta, f, \psi, \psi'_\theta\right) = Q\left(\theta, f, \psi\right) + \lambda_1 \psi'_\theta\left(y,\theta\right) f\left(y,\theta\right),$$

где $Q(\theta, f, \psi)$ – дважды непрерывно дифференцируемый по ψ функционал, $\lambda_1 = \lambda_1(\theta) \neq 0$;

2) при всех значениях θ функции ψ, f, f'_θ и $\dfrac{\partial Q}{\partial \psi}$ принадлежат L_2 (т. е. интегрируемы квадраты функций).

Тогда необходимым условием экстремума по ψ функционала $\int\limits_Y K(\theta, f, \psi, \psi'_\theta)dy$ при вариациях $\delta\psi$ оценочной функции, удовлетворяющих условию (1.6), будет выполнение равенства

$$\frac{\partial Q}{\partial \psi} - \lambda_1 \frac{\partial f}{\partial \theta} + \lambda_0 f = 0,$$

где коэффициент $\lambda_0 = \lambda_0(\theta)$ не зависит от y.

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

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

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

Доказательство. Чтобы получить единственное решение, введем условие нормировки

$$d = 1. \qquad (2.8)$$

Тогда $U_s(\psi) = \mathbf{E}_s\psi^2$. Рассмотрим функционал

$$\int\limits_Y \left[\psi^2(y, \theta)s(y, \theta) + \lambda_1\psi'_\theta(y, \theta)f(y, \theta) + \lambda_2\psi(y, \theta)f(y, \theta)\right]dy,$$

где с множителем Лагранжа λ_1 введено условие (2.8), а с λ_2 – условие (1.6). Согласно лемме 2.1 необходимое условие экстремума есть

$$2\psi(y, \theta)s(y, \theta) - \lambda_1 f'_\theta(y, \theta) + \tilde{\lambda}_2 f(y, \theta) = 0,$$

где $\tilde{\lambda}_2 = \lambda_2 + \lambda_0$, откуда

$$\psi(y, \theta) = \frac{\lambda_1}{2}\left[\frac{f'_\theta(y, \theta)}{f(y, \theta)} - \frac{\tilde{\lambda}_2}{\lambda_1}\right]\frac{f(y, \theta)}{s(y, \theta)},$$

что с точностью до обозначений совпадает с доказываемым результатом.

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 [9], 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 учитывается ограничение (1.6).

Легко увидеть, что наихудшая плотность $s(y,\theta)$, дающая наибольшее значение функционала (2.5), отлична от нуля в точках, где достигается наибольшее значение квадрата оценочной функции:

$$\sup_{s\in S}U_s(\psi,F)=\sup_{s\in S}\int_Y IF^2(y,\psi,F)s(y,\theta)dy=$$

$$=\frac{1}{d^2(\psi)}\sup_{s\in S}\int_Y \psi^2(y,\theta)s(y,\theta)dy=\frac{1}{d^2(\psi)}\sup_{y}\psi^2(y,\theta)\,.$$

Для обеспечения единственности решения введем нормировку оценочной функции вида (1.25). Как следствие, от исходной задачи (2.9) можно перейти к задаче (1.26) при ограничениях (1.6) и (1.25). Согласно результатам п. 1.4.1 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 конструктивным, путь – определить наихудшую плотность $s(y,\theta)$ для наилучшей оценочной функции (2.7) [24]:

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

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 $N(\mu,\sigma^2)$. 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 и равно $\omega^2 = 2$ для μ и $\omega^2 = 3$ для σ^2. Имеем оценочные уравнения:

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

$$\sum_{i=1}^{N}\left[\frac{(y_i - \mu)^2}{\sigma_s^2} - \frac{3}{5}\right]\exp\left[-\frac{(y_i - \mu)^2}{3\sigma_s^2}\right] = 0 .$$

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

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

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 $\omega^2 \le 1$ и к устойчивой (в том числе при асимметричном засорении) при $\omega^2 > 1$. Например, стойкая оценка соответствует выбору $\omega^2 = 2$ и является устойчивой.

Теорема 2.2 [12]. Для наилучшей оценки, определяемой в теореме 2.1, имеют место равенства

$$d=\frac{1}{c}\int_Y \psi^2(y,\theta)s(y,\theta)dy,$$

$$U_s(\psi)=\frac{c}{d}=\frac{c^2}{\int_Y \psi^2(y,\theta)s(y,\theta)dy}.$$

Пример 2.4. Применим теорему 2.2 к случаю $s(y,\theta)=g(y,\theta)$. Мы уже увидели в примере 2.2, что $U_s(\psi,F)=V(\psi,F)$ и оптимальная оценочная функция соответствует методу максимального правдоподобия.

Согласно теореме 2.2

$$V(\psi,F)=\frac{c^2}{\int_Y \psi_0^2(y,\theta)f(y,\theta)dy},$$

т. е.

$$V(\psi,F)=\frac{c^2}{\int_Y\left[c\frac{\partial}{\partial\theta}\ln f(y,\theta)\right]^2 f(y,\theta)dy}=\left\{\int_Y\left[\frac{\partial}{\partial\theta}\ln f(y,\theta)\right]^2 f(y,\theta)dy\right\}^{-1}.$$

В фигурных скобках стоит информация Фишера. Таким образом, теорема 2.2 в данном случае устанавливает известную связь между асимптотической дисперсией и информацией Фишера.

Заметим, что для справедливости теорем 2.1 и 2.2 функция $s(y, \theta)$ не обязательно должна быть нормированной и, более того, интегрируемой.

При использовании оценочных функций вида (2.7) в случае оценивания параметра сдвига существует одно достаточное условие того, что величина $\beta(\theta)$ в оценочной функции равна нулю [13].

Теорема 2.3. Пусть оценочная функция параметра сдвига θ имеет вид (2.7), а для плотностей f, s справедливо представление (1.32), причем s является сложной функцией от f: $s(y) = s(f(y))$. Тогда, если функция $f(y, \theta)$ непрерывна при $y \in Y$ и имеет равные значения в граничных точках интервала Y, то $\beta(\theta) = 0$.

2.3. ОПТИМАЛЬНЫЕ ОЦЕНКИ ПРИ КОМПРОМИССНОМ ФУНКЦИОНАЛЕ

Вернемся к анализу функционала (2.4), т. е. компромиссного функционала (2.2) в условиях байесовского точечного засорения [8].

Математическое ожидание функционала (по плотности s) имеет вид

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

$$= V(\psi, F) + k^2 U_s(\psi, F). \tag{2.11}$$

Таким образом, функционал (2.11) – это компромисс между классическим показателем неточности и показателем неустойчивости (2.5) в модели байесовского точечного засорения.

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):

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

$$= \frac{1}{d^2}\int_Y \psi^2(y,\theta)\Big[f(y,\theta) + k^2 s(y,\theta)\Big]dy = U_{s_1}(\psi, F), \qquad (2.12)$$

здесь функция $s_1(y,\theta) = f(y,\theta) + k^2 s(y,\theta)$ не удовлетворяет условию нормировки плотности распределения. Для функционала (2.12) справедливы теоремы 2.1 и 2.2.

Заметим, что функция $s_1(y,\theta)$ в (2.12) с точностью до (несущественной) нормировки может быть проинтерпретирована как плотность смеси распределений:

$$s_1(y,\theta) \sim (1-\alpha)f(y,\theta) + \alpha s(y,\theta),$$

где $0 < \alpha < 1$.

Применяя теорему 2.1 для функционала (2.12), получим наилучшую оценочную функцию вида

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

Таким образом, переход от функционала (2.5) к компромиссному функционалу (2.11) 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 $\mathbf{E}_s\tau(\psi, F, y^*, k)$. Однако можно вновь ввести функционал (2.5) на множестве функций, удовлетворяющих условию (2.6), а затем – компромиссный функционал

$$\tilde{U}_s(\psi) = V(\psi, F) + k^2 U_s(\psi, F) = U_{s_1}(\psi, F), \qquad (2.14)$$

уже не имеющий смысла $\mathbf{E}_s\tau(\psi, F, y^*, k)$. При этом теоремы 2.1 и 2.2 вновь остаются справедливыми.

Выбор подходящего значения величины k^2 – отдельная задача. Кроме того, работать с k^2, величиной неограниченной ($k^2 \in [0, +\infty)$), неудобно на практике. По этим причинам целесообразно рассмотреть компромиссный показатель, основанный на относительных характеристиках.

Классическая относительная характеристика точности оценивания – это эффективность

$$\mathrm{eff}\ \psi = \frac{V_0}{V(\psi, F)},$$

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

$$0 \le \mathrm{eff}\,\psi \le 1.$$

Аналогично эффективности введем относительную характеристику устойчивости при байесовском точечном засорении

$$\mathrm{stb}_s\psi = \frac{U_s^*}{U_s(\psi, F)},$$

где $U_s^* = \min\limits_{\psi} U_s(\psi, F)$. Таким образом, справедливо

$$0 \le \mathrm{stb}_s\psi \le 1.$$

Составим компромиссный функционал вида

$$\frac{1-\alpha}{\mathrm{eff}\,\psi} + \frac{\alpha}{\mathrm{stb}_s\,\psi}, \tag{2.15}$$

где $0 \le \alpha \le 1$, или иначе

$$\frac{1-\alpha}{V_0} V(\psi, F) + \frac{\alpha}{U_s^*} U_s(\psi, F).$$

Этот функционал эквивалентен функционалу (2.14). Действительно, умножая (2.14) на $\dfrac{1-\alpha}{V_0}$, получим

$$\frac{1-\alpha}{V_0} V(\psi, F) + \frac{1-\alpha}{V_0} k^2 U_s(\psi, F),$$

откуда

$$k^2 = \frac{\alpha}{1-\alpha} \frac{V_0}{U_s^*}. \tag{2.16}$$

Следствие из теоремы 2.1. Минимум функционала (2.15) достигается на оценочной функции (2.13) со значением k^2, определяемым формулой (2.16).

2.4. МАКСИМАЛЬНО НЕОПРЕДЕЛЕННОЕ БАЙЕСОВСКОЕ ТОЧЕЧНОЕ ЗАСОРЕНИЕ

Рассматриваемый подход отличается от подходов Хьюбера и Хампеля тем, что для него требуется бóльшая 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. Максиминное решение – стойкая оценка – соответствует оптимальному решению лишь в ограниченном классе плотностей (само его наличие уже является использованием априорной информации).

Еще один путь – использование *максимально неопределенного байесовского точечного засорения*, когда засоряющее наблюдение равномерно распределено по области значений случайной величины Y [6, 10, 11].

Выберем

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

(функция не интегрируема при $Y=R^1$), а показатель $U_s(\psi,F)$ в этих условиях обозначим $W(\psi,F)$:

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

Функционал $W(\psi,F)$ используется в качестве меры неустойчивости оценки [24].

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

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

и неустойчивыми – в противном случае.

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

С другой стороны, показатель $W(\psi,F)$ можно рассматривать как квадрат L_2-нормы функции влияния (без веса) [10].

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) [24].

Относительный показатель устойчивости вводится как частный случай $\mathrm{stb}_s\ \psi$ [24]:

$$\mathrm{stb}\ \psi = W_* / W(\psi, F)\,,$$

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

Если ММП-оценка может иметь низкую (часто нулевую) устойчивость, то ОМУ может иметь низкую эффективность. Снова необходимо обратиться к компромиссу между эффективностью и устойчивостью.

Рассмотрим следующие две задачи [24]:

1) $\max\limits_{\psi}\ \mathrm{eff}\,\psi$ при фиксированном значении $\mathrm{stb}\,\psi$;

2) $\max\limits_{\psi}\ \mathrm{stb}\,\psi$ при фиксированном значении $\mathrm{eff}\,\psi$.

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

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

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

Доказательство. Пусть выполняются условия асимптотической несмещенности (1.6) и нормировки (2.8) и зафиксирована эффективность, тогда ψ минимизирует функционал

$$\int\left[\psi^2(y,\theta) + \lambda_1\psi'_\theta(y,\theta)f(y,\theta) + \lambda_2\psi(y,\theta)f(y,\theta) + \lambda_3\psi^2(y,\theta)f(y,\theta)\right]dy\,,$$

где с множителем Лагранжа λ_1 введено условие (2.8), с λ_2 – условие (1.6), множитель λ_3 отвечает за фиксированную эффективность.

Используя лемму 2.1, находим

$$2\psi(y,\theta)\left[1+\lambda_3 f(y,\theta)\right]-\lambda_1 f'_\theta(y,\theta)+\tilde{\lambda}_2 f(y,\theta)=0,$$

где $\tilde{\lambda}_2=\lambda_2+\lambda_0$, откуда

$$\psi(y,\theta)=\frac{\lambda_1}{2\lambda_3}\left[\frac{f'_\theta(y,\theta)}{f(y,\theta)}-\frac{\tilde{\lambda}_2}{\lambda_1}\right]\frac{f(y,\theta)}{f(y,\theta)+1/(\lambda_3)},$$

что с точностью до обозначений совпадает с доказываемым результатом.

Минимизация эффективности при фиксированной устойчивости проводится аналогично и дает тот же результат.

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

Сравнение условно оптимальной оценочной функции (2.17) с оценочной функцией (2.13) показывает, что это одни и те же оценочные функции, различающиеся только способом выбора функции k^2.

Заметим также, что имеется определенная аналогия между данными оценками и оптимальными B-робастными оценками в подходе Хампеля, которые минимизируют асимптотическую дисперсию при заданной верхней грани для значений чувствительности к большой ошибке γ^* [9].

Для величины $\beta=\beta(\theta)$ обычно не существует аналитического представления. Однако для параметра сдвига применима теорема 2.3, поскольку в этом случае $s(y)$ является сложной функцией $f(y)$:

$$s(f(y))=f(y)+k^2,$$

если коэффициент k^2 не зависит от θ. Таким образом, в условиях теоремы 2.3 для условно оптимальной оценочной функции параметра сдвига имеет место

$$\beta=0.$$

Поиск значения k^2, соответствующего накладываемым условиям, – отдельная задача, а, поскольку k^2 в общем случае зависит от оцениваемого параметра ($k^2=k^2(\theta)$), поиск значений $\hat{\theta}$ и $k^2(\hat{\theta})$ требуется производить одновременно, т. е. к решаемому оценочному уравнению необходимо добавить еще одно уравнение. Поэтому удобно рассматривать также и компромиссный функционал вида (2.15).

В частности, назовем *компромиссной* оценку θ_c с оценочной функцией, минимизирующей компромиссный функционал (2.15) при $s(y, \theta) = 1$ и $\alpha = 0.5$ [24], т. е. функционал

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

Согласно следствию из теоремы 2.1 компромиссная оценка имеет оценочную функцию (2.13) с

$$k^2 = \frac{V_0}{W_*},$$

и, таким образом, она является условно оптимальной.

Пример 2.5. Рассмотрим модель нормального распределения $N(\mu, \sigma^2)$. Компромиссным оценкам соответствует значение $k^2 = \dfrac{1}{4\sqrt{\pi}\sigma}$ для μ_c и $k^2 = \dfrac{1}{8\sqrt{\pi}\sigma}$ для σ_c [24]. Имеем для μ_c оценочное уравнение

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

и характеристики $\text{eff} \approx 0,82$, $\text{stb} \approx 0,92$. Для σ_c имеем оценочное уравнение

$$\sum_{i=1}^{N}\left[\frac{(y_i - \mu\ )^2}{\sigma_c^2} - 0,745\right] \Big/ \left\{4\sqrt{2} + \exp\left[\frac{(y_i - \mu\ )^2}{2\sigma_c^2}\right]\right\} = 0$$

и характеристики $\text{eff} \approx 0,7$, $\text{stb} \approx 0,8$.

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

Поэтому введем условно оптимальную оценку, удовлетворяющую условию

$$\text{eff}\,\psi = \text{stb}\,\psi, \tag{2.18}$$

назовем ее *равнооптимальной* и обозначим θ_{eq} [6, 10, 11].

Теорема 2.5 [12]. Пусть модельная плотность $f(y,\theta)$ такова, что эффективность и устойчивость условно оптимальной оценки – непрерывные функции параметра k^2.

Тогда равнооптимальная оценочная функция будет решением оптимизационной задачи

$$\max_{\psi}\ \min\ \{\mathrm{eff}\,\psi,\ \mathrm{stb}\,\psi\}.$$

Для используемых на практике распределений условие теоремы часто выполнено.

Пример 2.6. В модели нормального распределения $N(\mu,\sigma^2)$ равнооптимальным оценкам соответствует значение $k^2 \approx \dfrac{1}{6{,}8\sqrt{\pi}\cdot\sigma}$ с $\mathrm{eff} \approx \mathrm{stb} \approx 0{,}85$ для μ_{eq} и $k^2 \approx \dfrac{1}{13\sqrt{\pi}\cdot\sigma}$ с $\mathrm{eff} \approx \mathrm{stb} \approx 0{,}73$ для σ_{eq}.

2.5. ОБОБЩЕННЫЕ РАДИКАЛЬНЫЕ ОЦЕНКИ

Часто нахождение оценочных функций условно оптимальных оценок оказывается технически непростым. Поэтому рассмотрим еще один, более удобный, вид оценок, аппроксимирующих условно оптимальные.

В качестве синтетического показателя, отражающего в неявном виде компромисс между эффективностью и устойчивостью, может быть использован показатель качества при байесовском точечном засорении с плотностью $s(y,\theta)$, пропорциональной $f^{1-\delta}(y,\theta)$, где $0<\delta<1$ – константа [24]:

$$U_\delta(\psi,F) = \int\limits_Y IF^2(y,\psi,F)\,f^{1-\delta}(y,\theta)\,dy =$$

$$= \frac{\int\limits_Y \psi^2(y,\theta)\,f^{1-\delta}(y,\theta)\,dy}{d^2} = \frac{\mathbf{E}\left[\psi^2/f^{\delta}\right]}{d^2}.$$

Пример 2.7. Возведем в степень $1-\delta$ плотность двустороннего экспоненциального распределения:

$$f^{1-\delta}(y,\mu,\lambda,\nu)=\left\{\frac{1}{2\lambda\Gamma(1+1/\nu)}\exp\left[-\left(\frac{|y-\mu|}{\lambda}\right)^{\nu}\right]\right\}^{1-\delta}=$$

$$=\frac{1}{\left[2\lambda\Gamma(1+1/\nu)\right]^{1-\delta}}\exp\left[-(1-\delta)\left(\frac{|y-\mu|}{\lambda}\right)^{\nu}\right]=$$

$$=\frac{1}{\left[2\lambda\Gamma(1+1/\nu)\right]^{1-\delta}}\exp\left[-\left(\frac{|y-\mu|}{\lambda/\sqrt[\nu]{1-\delta}}\right)^{\nu}\right].$$

Таким образом, возведение данной плотности в степень при $0<\delta<1$ приводит к увеличению масштаба (помимо естественного нарушения условия нормировки плотности), при этом распределение становится «менее определенным».

В общем случае возведение плотности $f(y,\theta)$ в степень, если переменная y является только аргументом показательной функции, приводит к изменению масштаба. Если плотность зависит от переменной y иным способом, то изменяется форма распределения.

При $\delta=0$ получаем асимптотическую дисперсию

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

при $\delta=1$ получаем меру неустойчивости

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

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

Заметим, что для функционала $U_\delta(\psi,F)$ могут использоваться и значения параметра δ, лежащие вне диапазона $0\le\delta\le1$. Ограничения на параметр δ будут определяться для каждой конкретной модели условиями регулярности. Например, для случая нормального распределения справедливо $\delta>-\frac{1}{2}$ [1, 24].

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 $\delta = 1/2$ называются *радикальными* [24]. В общем случае, подчеркивая связь с радикальными, будем называть данные оценки *обобщенными радикальными* [6, 11].

Большее удобство нахождения оценочных функций для обобщенных радикальных оценок, по сравнению с условно оптимальными, обусловлено более простым определением величины β.

Для параметра сдвига применима теорема 2.3, поскольку в этом случае $s(y)$ является сложной функцией $f(y)$:

$$s(f(y)) = f(y)^{1-\delta}.$$

Таким образом, в условиях теоремы 2.3 для обобщенной радикальной оценочной функции параметра сдвига имеет место

$$\beta = 0.$$

Для параметра масштаба справедливо следующее утверждение [11].

Теорема 2.6. Пусть θ – параметр масштаба. Тогда, если в граничных точках интервала Y значения функции

$$y \cdot f^{1+\delta}(y,\theta)$$

равны между собой, то для обобщенной радикальной оценочной функции имеет место

$$\beta = \frac{1}{\theta}\frac{\delta}{1+\delta}.$$

Для произвольного параметра модели можно использовать следующий результат.

Теорема 2.7. Для обобщенной радикальной оценочной функции имеет место

$$\beta = \frac{\int\limits_Y \frac{\partial}{\partial\theta}\ln f(y,\theta) f^{1+\delta}(y,\theta)\, dy}{\int\limits_Y f^{1+\delta}(y,\theta)\, dy}.$$

Пример 2.8. Для модели нормального распределения $N(\mu, \sigma^2)$ обобщенные радикальные оценки ввел Л.Д. Мешалкин (см. [1, 2, 24, 26]). Их оценочные функции имеют вид:

- для оценки μ_δ параметра сдвига

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

- для оценки σ_δ параметра масштаба

$$\chi_\delta(y, \sigma_\delta) = \left[\frac{(y_i - \mu)^2}{\sigma_\delta^2} - \frac{1}{1+\delta}\right]\exp\left[-\delta\frac{(y_i - \mu)^2}{2\sigma_\delta^2}\right].$$

Многие введенные выше оценки для нормального распределения – это частные случаи оценок Мешалкина.

Особую важность среди обобщенных радикальных оценок имеют радикальные оценки. Для них имеют место следующие результаты [11].

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

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

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

Однако имеется, по крайней мере, два частных случая, даваемых следующей теоремой.

Теорема 2.9. Пусть θ – параметр масштаба. Тогда, если плотность принадлежит семейству распределений

$$f(y, \theta) = \frac{1}{\theta \cdot \Gamma\left(1 + \frac{1}{\nu}\right)}\exp\left[-\frac{y^\nu}{\theta^\nu}\right],\ \nu > 0,\ y \ge 0,$$

или семейству двусторонних экспоненциальных распределений

$$f(y, \theta) = \frac{1}{2\theta \cdot \Gamma\left(1 + \frac{1}{\nu}\right)}\exp\left[-\frac{|y|^\nu}{\theta^\nu}\right],\ \nu > 0,\ -\infty < y < \infty,$$

то для радикальной оценочной функции имеет место (2.19).

Таким образом, с точки зрения наличия свойства (2.18) радикальные оценки близки к равнооптимальным оценкам, и те, и другие представляют собой «золотую середину» между ММП-оценкой и ОМУ. При этом равнооптимальные оценки, как условно оптимальные, будут иметь бóльшие значения eff и stb .

Пример 2.9. В модели нормального распределения $N(\mu, \sigma^2)$ радикальные оценки имеют характеристики $\text{eff} = \text{stb} \approx 0,84$ для оценки μ и $\text{eff} = \text{stb} \approx 0,72$ для оценки σ. Таким образом, радикальные оценки хуже равнооптимальных на 1 %.

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

Т а б л и ц а 2.1

Оценки параметров сдвига и масштаба нормального распределения и их характеристики

Оценка		eff	stb
ММП	μ_0 σ_0	1 1	0 0
ОМУ	$\mu_* = \mu_1$ $\sigma_* = \sigma_1$	0,65 0,43	1 1
Медиана	μ_{med} σ_{med}	0,64 0,37	0 0
Радикальная	$\mu_{1/2}$ $\sigma_{1/2}$	0,84 0,72	0,84 0,72
Равнооптимальная	μ_{eq} σ_{eq}	0,85 0,73	0,85 0,73
Компромиссная	μ_c σ_c	0,82 0,7	0,92 0,8
Стойкая	$\mu_s = \mu_{1/2}$ $\sigma_s = \sigma_{2/3}$	0,84 0,6	0,84 0,89

Компромиссная оценка проигрывает радикальной 2 % эффективности, но выигрывает 8 % устойчивости. Равнооптимальные оценки имеют характеристики на 1 % больше по сравнению с радикальными.

В подходе Шурыгина наибольшую устойчивость имеет ОМУ, в подходе Хампеля – медиана. Сравним их.

Для параметра сдвига обе оценки имеют примерно равную эффективность, для параметра масштаба ОМУ – на 6 % эффективнее. С точки зрения показателя устойчивости оценки имеют противоположные значения.

2.6. КОНТРОЛЬНЫЕ ВОПРОСЫ И УПРАЖНЕНИЯ

1. Опишите возможные действия исследователя, строящего модель в условиях засорения с фиксированной интенсивностью, при увеличении объема исследуемой выборки.

2. В чем полезность для практики модели точечного засорения?

3. Сравните подходы Шурыгина и Хампеля. В чем их сходство и в чем различие?

4. Как зависит устойчивость оценки от свойств плотности s? Ответ проиллюстрируйте на конкретном примере.

5. Сравните оптимальные B-робастные оценки в подходе Хампеля с условно оптимальными оценками в подходе Шурыгина.

6. Постройте графики оценочных функций ОМУ и наиболее B-робастных оценок параметров нормального распределения. Сравните свойства этих оценок.

7. Почему равнооптимальные оценки параметров нормального распределения имеют лучшие значения eff и stb по сравнению с радикальными оценками?

8. Выпишите оценочные функции радикальных оценок параметров распределения Лапласа, постройте их графики, сравните свойства радикальных и равнооптимальных оценок.

9. Выпишите оценочные функции ОМУ параметров распределения Коши. Сравните их с ММП- и медианными оценками, постройте графики оценочных функций.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Айвазян С.А.* Прикладная статистика. Основы моделирования и первичная обработка данных / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1983.

2. *Айвазян С.А.* Прикладная статистика. Исследование зависимостей / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1985.

3. *Вапник В.Н.* Восстановление зависимостей по эмпирическим данным / В.Н. Вапник. – М.: Наука, 1979.

4. *Демиденко Е.З.* Линейная и нелинейная регрессии / Е.З. Демиденко. – М.: Финансы и статистика, 1981.

5. *Демиденко Е.З.* Оптимизация и регрессия / Е.З. Демиденко. – М.: Наука, 1989.

6. *Денисов В.И.* Методы построения многофакторных моделей по неоднородным, негауссовским, зависимым наблюдениям: монография / В.И. Денисов, Д.В. Лисицин. – Новосибирск: Изд-во НГТУ, 2008.

7. *Лисицин Д.В.* Адаптивные методы построения многооткликовых регрессионных моделей // Труды V Междунар. конф. «Актуальные проблемы электронного приборостроения» АПЭП-2000 / Д.В. Лисицин. – Новосибирск, 2000. – Т. 3. – С. 14–18.

8. *Лисицин Д.В.* Об оценивании параметров модели при байесовском точечном засорении // Доклады АН ВШ РФ / Д.В. Лисицин. – 2009. – № 1(12). – С. 41–55.

9. *Лисицин Д.В.* Оценивание при байесовском точечном засорении: связь с подходом Хампеля и минимаксная оценка / Д.В. Лисицин // Сб. науч. тр. НГТУ. – Новосибирск: Изд-во НГТУ, 2011. – Вып. 3(65). – С. 61–66.

10. *Лисицин Д.В.* О локально устойчивом оценивании параметров распределений / Д.В. Лисицин, К.В. Гаврилов // Сб. науч. тр. НГТУ. – Новосибирск: Изд-во НГТУ, 2004. – Вып. 2(36). – С. 37–46.

11. *Лисицин Д.В.* Устойчивое оценивание параметров модели при асимметричном засорении данных / Д.В. Лисицин, К.В. Гаврилов // Изв. Междунар. академии наук высшей школы. – 2006. – № 1(35). – С. 60–73.

12. *Лисицин Д.В.* Об устойчивом оценивании параметров моделей при асимметричном засорении данных / Д.В. Лисицин, К.В. Гаврилов // Научный вестник НГТУ. – Новосибирск: Изд-во НГТУ, 2008. – № 1(30). – С. 33–40.

13. *Лисицин Д.В.* О некоторых свойствах *М*-оценок // Сб. науч. тр. НГТУ / Д.В. Лисицин, К.В. Гаврилов. – Новосибирск: Изд-во НГТУ, 2011. – Вып. 2(64). – С. 61–68.

14. *Мудров В.И.* Методы обработки измерений: Квазиправдоподобные оценки / В.И. Мудров, В.Л. Кушко. – М.: Радио и связь, 1983.

15. *Новицкий П.В.* Оценка погрешностей результатов измерений / П.В. Новицкий, И.А. Зограф. – Л.: Энергоатомиздат, 1991.

16. *Редько М.Ю.* Квазиправдоподобные оценки для линейной регрессии / М.Ю. Редько. – Новосибирск: Новосиб. электротехн. ин-т, 1988. – Деп. в ВИНИТИ, № 4821–В88.

17. Робастность в статистике. Подход на основе функций влияния / Ф. Хампель, Э. Рончетти, П. Рассеу, В. Штаэль. – М.: Мир, 1989.

18. *Смоляк С.А.* Устойчивые методы оценивания. (Статистическая обработка неоднородных совокупностей) / С.А. Смоляк, Б.П. Титаренко. – М.: Статистика, 1980.

19. Устойчивые статистические методы оценки данных / под ред. Р.Л. Лонера, Г.Н. Уилкинсона. – М.: Машиностроение, 1984.

20. *Хьюбер П.* Робастность в статистике / П. Хьюбер. – М.: Мир, 1984.

21. *Цыпкин Я.З.* Основы информационной теории идентификации / Я.З. Цыпкин. – М.: Наука, 1984.

22. *Цыпкин Я.З.* Информационная теория идентификации / Я.З. Цыпкин. – М. Наука. Физматлит, 1995.

23. *Шуленин В.П.* Введение в робастную статистику / В.П. Шуленин. – Томск: Изд-во Том. ун-та, 1993.

24. *Шурыгин А. М.* Прикладная стохастика: робастность, оценивание, прогноз / А.М. Шурыгин. – М.: Финансы и статистика, 2000.

25. *Шурыгин А.М.* Асимптотическая теория устойчивого оценивания: Автореф. дис. ... д-ра техн. наук. – М., 2002.

26. *Шурыгин А. М.* Математические методы прогнозирования / А.М. Шурыгин. – М.: Горячая линия – Телеком, 2009.

27. *Mineo A.M.*, *Ruggieri M.* A software tool for the exponential power distribution: The normalp package // Journal of Statistical Software. – 2005. – Vol. 12, issue 4.

28. *Song K.-Sh.* A globally convergent and consistent method for estimating the shape parameter of a generalized gaussian distribution // IEEE Trans. Inform. Theory. – 2006. – Vol. 52. – P. 510–527.

29. *Subbotin M.Th.* On the law of frequency of error // Математический сборник. – 1923. – Т. 31. – С. 296–301.

ОГЛАВЛЕНИЕ

Лисицин Даниил Валерьевич

УСТОЙЧИВЫЕ МЕТОДЫ
ОЦЕНИВАНИЯ ПАРАМЕТРОВ
СТАТИСТИЧЕСКИХ МОДЕЛЕЙ

Учебное пособие

Редактор *И.Л. Кескевич*
Выпускающий редактор *И.П. Брованова*
Корректор *И.Е. Семенова*
Дизайн обложки *А.В. Ладыжская*
Компьютерная верстка *А.В. Сухарева*

Подписано в печать 15.03.2013. Формат 60×84 1/16. Бумага офсетная. Тираж 60 экз.
Уч.-изд. л. 4,41. Печ. л. 4,75. Изд. № 321/12. Заказ № Цена договорная

Отпечатано в типографии
Новосибирского государственного технического университета
630073, г. Новосибирск, пр. К. Маркса, 20