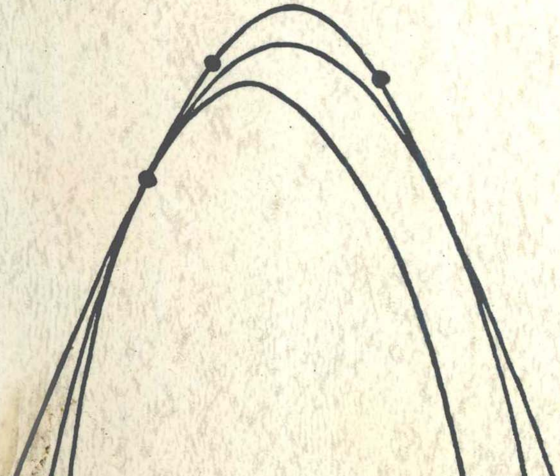
П. ХЬЮБЕР

Робастность в статистике



УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., 2, издательство «Мир».

Питер Хьюбер

РОБАСТНОСТЬ В СТАТИСТИКЕ

Научные редакторы А. А. БрянДинская, И. А. Маховая
Мл. научный редактор Н. С. Полякова
Художник С. А. Бычков
Художественный редактор В. И. Шаповалов
Технический редактор Е. В. Ящук
Корректор Н. А. Гиря

ИБ № 3447

Сдано в набор 12.04.83. Подписано к печати 23.11.83. Формат $60 \times 90^{1}/_{16}$. Бумага типографская № 2. Гарнитура литературная. Печать высокая. Объем 9,50 бум. л. Усл. печ. л. 19 Усл. кр.-отт. 19 Уч.-изд. л. 17, 51 Изд. № 1/2374. Тираж 6250 экз. Зак. 617. Цена 2 руб.

ИЗДАТЕЛЬСТВО «МИР»
129820, Москва, И-110, ГСП, 1-й Рижский пер., 2.

Ленинградская типография № 2 головное предприятие ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгении Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 198052, г. Ленинград, Л-52, Измайловский проспект, 29.

ИЗДАТЕЛЬСТВО • МИР •

Wiley Series in Probability and Mathematical Statistics

Robust Statistics

PETER J. HUBER

Professor of Statistics
Harvard University
Cambridge, Massachusetts

John Wiley and Sons
New York • Chichester • Brisbane • Toronto
1981

Робастность в статистике

Перевод с английского
И. А. Маховой и
В. И. Хохлова

под редакцией
И. Г. Журбенко

Москва «Мир» 1984

ББК 22.17
Х 98
УДК 519.24

Хьюбер Дж. П.

Х 98 Робастность в статистике: Пер. с англ.— М.: Мир, 1984. — 304 с., ил.

Первое систематическое изложение теории робастных оценок — важного и интенсивно развивающегося направления современной математической статистики. Монография написана американским специалистом — одним из создателей этой теории. В ней обобщены разрозненные методы проверки устойчивости конкретных статистических процедур. Часть результатов публикуется впервые. Приведены алгоритмы вычислений робастных оценок, а также таблицы, количественно характеризующие робастность нескольких оценок.

Для научных работников, инженеров и студентов, специализирующихся в области математической и прикладной статистики.

Х $\frac{1702060000-071}{041(01)-84}$ 33—84, ч. 1

ББК 22.17
517.8

Редакция литературы по математическим наукам

Предисловие редактора перевода

Любые статистические выводы всегда основываются на тех или иных начальных предположениях. К числу таких предположений очень часто относят ограничения на форму рассматриваемых распределений. Например, нередко используемым априорным допущением является предположение о нормальности рассматриваемых распределений. Естественно, что такие ограничения на практике могут выполняться лишь с некоторой степенью точности. Остаются ли при этом верными статистические выводы, полученные в начальных предположениях для «невозмущенных» данных? Оказывается, что существует целый класс так называемых «робастных» процедур, которые обладают этим свойством. Термин «робастность» был введен для выделения класса статистических процедур, слабочувствительных к небольшим изменениям начальных предположений. Для этих же целей некоторые авторы используют термин «устойчивость», излишняя популярность которого во многих областях математики делает его неудобным для читателя.

Предлагаемая книга является первой монографией, целиком посвященной изучению проблемы робастности статистических процедур. Тематика книги вплотную примыкает к классическому направлению теории вероятностей — сходимости вероятностных мер, основы которого были заложены и развиты советской вероятностной школой. Результаты книги в основном относятся к области теоретической статистики, но в настоящее время ни одно практическое применение статистических методов не обходится без проверки робастности этих методов. В этом плане данная монография играет исключительную роль в прикладной статистике, обобщая разрозненные методы проверки робастности многочисленных конкретных статистических процедур, предоставляя исследователю точный и строгий математический инструмент проверки качества используемых статистик. Книга служит превосходным примером того, какие широкие возможности открывает применение глубоких математических теорий в практических областях. Таким образом она, несомненно, дает толчок как новым исследованиям в области теоретической статистики, так и обоснованным применениям статистики в самых разнообразных областях.

В настоящее время существует обширная отечественная и зарубежная литература по данному вопросу, что связано как с теоретической, так и с практической важностью рассматриваемого круга вопросов. Однако большая часть литературы, особенно советских авторов, оказалась не затронута. Безусловно, этот недостаток будет восполнен с появлением отечественных монографий на эту тему.

Книга написана достаточно строгим математическим языком, рассуждения и доказательства прозрачны и лаконичны. Текст постоянно сопровождается пояснениями и примерами, необходимыми для ясного понимания прикладной сущности рассматриваемых вопросов. Книга вполне доступна студентам старших курсов математических специальностей, аспирантам, инженерам и научным сотрудникам, интересующимся развитием и применением статистических методов на практике. Она будет интересна и математикам-теоретикам, и статистикам-прикладникам.

При подготовке рукописи перевода были внесены необходимые исправления к тексту оригинала, любезно предоставленные нам автором книги, а также устранены замеченные переводчиками опечатки.

И. Г. Журбенко

Предисловие

Предлагаемая монография — первая попытка систематического изложения теории робастных оценок. Словом «робастность» как термином стали пользоваться лишь с 1953 г. (ввел его Дж. Е. П. Бокс) и только к середине шестидесятых годов названное этим словом свойство начали рассматривать как предмет специального исследования, однако саму эту проблематику по существу нельзя назвать принципиально новой. Среди ведущих ученых конца девятнадцатого и начала двадцатого столетий было несколько статистиков-практиков (назовем хотя бы некоторых из них: астроном С. Ньюкомб, астрофизик А. Эддингтон и геофизик Г. Джеффрис), которые в своих исследованиях проявляли совершенно отчетливое понимание идеи робастности. Они знали об опасностях, порождаемых длинными хвостами функций распределения ошибок; ими были предложены вероятностные модели для грубых ошибок и даже придуманы превосходные робастные варианты стандартных оценок, заново переоткрытые лишь совсем недавно. Вместе с тем статистики-теоретики долгое время старались избегать этой, как считалось, расплывчатой и «недостойной» темы. Хочется думать, что рассеять это предубеждение помогла моя работа 1964 г. Забавно (хотя это и вызывает беспокойство), что с недавних пор, по-видимому, ударились в другую крайность и ныне слово «робастность» становится чем-то вроде магического заклинания, призванного добавить респектабельности.

Эта книга задумана как довольно полное введение в теорию робастности, предназначенное в равной мере и специалистам по теоретической статистике, и статистикам-прикладникам. Хотя монография написана в теоретическом ключе, упор в ней сделан не на математическую завершенность, а на идейное содержание предмета. Уровень изложения намеренно неоднороден: в некоторых главах простые случаи разобраны со всей математической строгостью, в других (например, для оценок регрессии в многопараметрической постановке и оценок ковариационной матрицы) результаты, подробно разобранные для простых случаев, по аналогии переносятся на более сложные ситуации, в которых доказательства не всегда проходят, или проходят лишь при некоторых искусственных допущениях. В книгу включен также ряд алгоритмов для вычисления робастных оценок: там, где это возможно, приводятся доказательства сходимости процедур.

Глава 1 содержит общее введение и обзор: она обязательна для всех читателей. В гл. 2 под идеи качественной и количественной робастности подводится формальная математическая база. Эту главу читатель, готовый принять на веру некоторые результаты, может пропустить (или бегло просмотреть). В гл. 3 вводятся и рассматриваются три основных типа оценок (*M*-, *L*- и *R*-оценки), а в гл. 4 изложена теория асимптотической минимаксности для оценок сдвига: обе эти главы также обязательны. Дальнейшее изложение охватывает разные направления, так что последующие главы достаточно независимы и замкнуты. Их можно читать или изучать в более или менее произвольном порядке.

В книге отсутствуют упражнения — я пришел к выводу, что в затронутой области придумать достаточное количество задач, которые были бы и нетривиальными, и не слишком трудными, довольно сложно, поэтому она не отвечает некоторым формальным критериям учебника. Тем не менее материал разных частей рукописи довольно успешно использовался при чтении специальных курсов.

Книга не претендует на энциклопедический подход. Мне хотелось лишь осветить те аспекты и методы, которые лично я считаю наиболее важными. Некоторые пропуски и пробелы отчасти сделаны преднамеренно, а отчасти объясняются просто острым дефицитом времени: больше нельзя было откладывать выход книги (первый вариант рукописи появился еще в 1972 г.). Например, были исключены адаптивные оценки, которые я теперь предпочитаю относить не к робастным, а к непараметрическим оценкам (в части, касающейся непараметрического эффективного оценивания). Так называемый байесовский подход к робастности связывает рассматриваемый предмет с допустимым оцениванием в соответствующей параметрической супермодели, и до сих пор не найдено столь надежных правил выбора супермодели и приоритета, которые в конечном счете имели бы некое отношение к робастности. Изложение *L*- и *R*-оценок было сокращено по сравнению с более ранними планами, поскольку эти оценки не допускают удобных обобщений и в многопараметрических ситуациях вызывают трудности при вычислениях и обработке.

Значительная часть окончательного варианта рукописи была написана в Гарвардском университете осенью 1977 г.; я приношу благодарность студентам, в частности П. Розенбауму и И. Ёсидзоэ, активно посещавшим семинар и сделавшим много ценных замечаний.

П. Дж. Хьюбер

Кембридж, Массачусетс
Июль 1980 г.

ГЛАВА 1

Общие основы

1.1. ЗАЧЕМ НУЖНЫ РОБАСТНЫЕ ПРОЦЕДУРЫ?

Статистические выводы лишь отчасти основываются на наблюдениях. Столь же важную основу этих выводов составляют исходные предположения об исследуемой ситуации. Даже в самых простых случаях делаются явные или неявные допущения о случайности и независимости, о виде тех или иных распределений в изучаемой модели, например о виде исходных распределений для некоторых неизвестных параметров, и т. д.

От такого рода предположений не требуется абсолютной точности. Они представляют собой математически целесообразные приближения, отвечающие зачастую не вполне точным знаниям или представлениям. Как и во всякой другой отрасли прикладной математики, такие приближения или упрощения крайне необходимы; при оценке обоснованности их применения обращаются к принципу своего рода непрерывности или устойчивости: малая ошибка в математической модели не должна приводить к существенной ошибке окончательных выводов.

К сожалению, этому принципу отвечают далеко не все модели. На протяжении последних десятилетий росло понимание того факта, что некоторые наиболее распространенные статистические процедуры (в том числе те, которые оптимальны в предположении о нормальности распределения) весьма чувствительны к довольно малым отклонениям от предположений. Вот почему теперь в изобилии появились иные процедуры — «робастные» [1]).

Со словом «робастность» связывают — и подчас безосновательно — различные понятия. В настоящей книге это слово используется в относительно узком смысле, диктуемом нашими целями: *робастность означает нечувствительность к малым отклонениям от предположений.*

Обратимся сначала к *робастности по распределению*, т. е. к ситуациям, в которых истинная функция распределения незначительно отличается от предполагаемой в модели (как правило, гауссовской функции распределения). Это не только наи-

[1]) От англ. robust — крепкий, здоровый, дюжий. — *Прим. перев.*

более важный случай, но и наиболее полно изученный. Гораздо меньше известно о том, что происходит в тех ситуациях, когда несколько нарушаются прочие стандартные допущения статистики, и о том, какие меры защиты должны предусматриваться в подобных случаях.

Приведем пример (из работы Тьюки (1960)), показывающий, к каким последствиям приводит отсутствие робастности по распределению в некоторых классических процедурах.

Пример 1.1. Предположим, что имеется набор, составленный большим числом n «хороших» и «плохих» случайно перемешанных наблюдений x_i некоторой величины μ. Каждое «хорошее» наблюдение появляется с вероятностью $1-\varepsilon$, «плохое» — с вероятностью ε, где ε — малое число. В первом случае наблюдения x_i имеют нормальное распределение $\mathcal{N}(\mu, \sigma^2)$, во втором — нормальное распределение $\mathcal{N}(\mu, 9\sigma^2)$. Иначе говоря, все наблюдения имеют одно и то же среднее, а ошибка для некоторых из них в три раза больше, чем у остальных.

Можно дать следующее эквивалентное описание приведенной ситуации: величины x_i независимы и имеют одно и то же распределение

$$F(x) = (1-\varepsilon)\Phi\left(\frac{x-\mu}{\sigma}\right) + \varepsilon\Phi\left(\frac{x-\mu}{3\sigma}\right), \tag{1.1}$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy \tag{1.2}$$

— функция стандартного нормального распределения.

Рассмотрим две широко известные оценки разброса — среднее абсолютное отклонение

$$d_n = \frac{1}{n} \sum |x_i - \bar{x}| \tag{1.3}$$

и среднее квадратичное отклонение

$$s_n = \left[\frac{1}{n} \sum (x_i - \bar{x})^2\right]^{1/2}. \tag{1.4}$$

По поводу относительных преимуществ оценок d_n и s_n противоположных точек зрения придерживались Эддингтон (1914, с. 147) и Фишер (1920, сноска на с. 762). Эддингтон защищал использование первой из них: «Хотя это входит в противоречие с рекомендацией большинства учебников, как нетрудно показать, правильнее пользоваться именно ею». Фишер решал вопрос, по-видимому, на основе того факта, что для нормально распределенных наблюдений величина s_n примерно на 12 % более эффективна, чем d_n.

Разумеется, статистиками s_n и d_n измеряются разные характеристики распределения ошибки. Например, если ошибки имеют в точности нормальное распределение, то величина s_n сходится к σ, в то время как величина d_n стремится к $\sqrt{2/\pi}\,\sigma \approx 0.80\sigma$. Поэтому нам следует уточнить, как нужно проводить сравнение качества этих статистик. Изберем асимптотическую относительную эффективность (АОЭ) статистики d_n по статистике s_n, определяемую следующим образом:

$$\text{АОЭ}(\varepsilon) = \lim_{n\to\infty} \frac{\mathrm{D}(s_n)/(\mathrm{E}s_n)^2}{\mathrm{D}(d_n)/(\mathrm{E}d_n)^2} = \frac{[3(1+80\varepsilon)/(1+8\varepsilon)^2-1]/4}{\pi(1+8\varepsilon)/(2(1+2\varepsilon)^2)-1}.$$

Значения этого показателя при различных ε приведены на рис. 1.1.1.

ε	АОЭ(ε)
0	0.876
0.001	0.948
0.002	1.016
0.005	1.198
0.01	1.439
0.02	1.752
0.05	2.035
0.10	1.903
0.15	1.689
0.25	1.371
0.5	1.017
1.0	0.876

Рис. 1.1.1. Асимптотическая эффективность среднего абсолютного отклонения относительно среднего квадратичного отклонения. Из работы Хьюбера (1977б) с уведомлением издателя.

Полученная картина удручает: уже двух плохих наблюдений на тысячу достаточно, чтобы свести на нет 12 %-ное преимущество средней квадратичной ошибки, причем наибольшее значение, которое показатель АОЭ(ε) принимает вблизи значения $\varepsilon = 0.05$, превосходит 2.

Такое положение усугубляется, в частности, тем фактом, что, оказывается, в естественных науках типичные выборки «хороших данных» довольно точно моделируются законом распределения ошибки вида (1.1), где ε лежит в пределах от 0.01 до 0.1. (Это не значит, что такие выборки содержат от 1 до 10 % больших ошибок, хотя довольно часто это именно так; приведенный выше закон (1.1) может служить просто удобным описанием функции распределения с несколько более длинными хвостами, нежели у функции нормального распределения.) Поэтому приходится с сожалением констатировать, что таких естественно появляющихся отклонений от идеальной модели бывает достаточно, чтобы лишить смысла применение традиционной теории асимптотической оптимальности: на практике, безусловно, следует отдать предпочтение статистике d_n перед s_n, поскольку она лучше при всех ε от 0.002 до 0.5.

Чтобы избежать недоразумений, сразу же укажем, какие выводы *не* следует делать из приведенных здесь соображений. Во-первых, не следует делать вывод, что мы защищаем использование среднего абсолютного отклонения (в то же время оно дает лучшие оценки масштаба). Во-вторых, некоторые считают, что приведенный пример не соответствует реальности, поскольку «плохие» наблюдения будут выступать в качестве выделяющихся наблюдений, а всякий добросовестный статистик до вычисления средней квадратичной ошибки что-нибудь предпримет относительно этих наблюдений. Здесь это возражение не по существу; при предварительном усечении выделяющихся наблюдений средняя квадратичная ошибка может по достоинствам в значительной мере превзойти среднюю абсолютную ошибку, но мы ограничились выше рассмотрением только *немодифицированных* классических оценок.

Приведенный пример побуждает вплотную заняться вопросом о длине хвостов распределения: удлинение хвостов распределения дискредитирует дисперсию оценки s_n (и в значительно меньшей мере влияет на d_n). С другой стороны, укорачивание хвостов оказывает на распределения этих оценок незначительное влияние. (Оно может привести к потере абсолютной эффективности за счет понижения асимптотической границы Крамера — Рао, но последняя настолько неустойчива при малых изменениях распределения, что указанному влиянию можно не придавать особого значения.)

Такая чувствительность к удлинению хвостов распределения типична для классических процедур и не исчерпывается данным примером. Вследствие этого «робастность по распределению» и «защищенность от выделяющихся наблюдений» — понятия, различные в идейном плане, — на практике выступают как синонимы. Всякая разумная формальная или неформальная процедура для усечения выделяющихся наблюдений призвана защитить от возможных ошибок в выводах.

Теперь имеются основания задать следующие вопросы. Нужны ли вообще робастные процедуры; может быть, достаточно руководствоваться подходом, в котором предусмотрены два шага:

(1) «редактирование» данных усечением выделяющихся наблюдений по некоторому правилу,

(2) последующее применение для полученных данных классических критериев и процедур оценивания?

Будут ли эти два шага кратчайшим путем к результату?

К сожалению, не будут по следующим причинам.

(1) Указанные шаги довольно редко удается четко разграничить; например, в задачах многопараметрической регрессии выделяющиеся наблюдения распознать трудно, если не иметь надежных робастных оценок для параметров.

(2) Даже если исходный набор составляют перемешанные с некоторым числом больших ошибок наблюдения, имеющие нормальное распределение, то данные отредактированного набора не будут иметь нормальное распределение (что объясняется статистическими ошибками двоякого рода — неверными усечениями и ошибочными сохранениями); еще хуже ситуация, когда исходный набор, за исключением вкраплений больших ошибок, получен из данных, подлинное распределение которых не гауссовское. Поэтому теоретические выводы, основанные на предположении о нормальном распределении наблюдений, неприменимы для отредактированных данных и оценка действительной ценности рассматриваемой процедуры из двух шагов может оказаться более сложной, чем для непосредственной робастной процедуры.

(3) Практикой установлено, что лучшие процедуры усечения не достигают в полной мере качеств лучших робастных процедур. По-видимому, предпочтение следует отдать последним, так как они допускают плавный переход от полной неизменности до усечения всех наблюдений; см. Хэмпел (1974 а; 1976).

1.2. ЧТО ТРЕБУЕТСЯ ОТ РОБАСТНОЙ ПРОЦЕДУРЫ?

Мы будем придерживаться точки зрения, которую можно назвать «прикладной параметрической» точкой зрения, т. е. будем считать, что имеется параметрическая модель, пусть (и даже наверное) не абсолютно точно, но, надо полагать, хорошо приближающая истинную изучаемую ситуацию. Подобная точка зрения предполагает у любой статистической процедуры наличие таких желательных особенностей.

(1) Для выбранной модели процедура должна иметь достаточно хорошую (оптимальную или почти оптимальную) эффективность.

(2) Процедура обязана быть робастной, иначе говоря, малые отклонения от предположений о модели должны ухудшать качество процедуры лишь в малой степени, т. е. характеристики процедуры (например, асимптотика дисперсии или уровень значимости и мощность критерия) должны быть близки к номинальным величинам, вычисленным для принятой модели.

(3) Несколько бóльшие отклонения от допущений модели не должны приводить к катастрофическим последствиям.

Использование критериев, основанных на асимптотических результатах, необходимо сопроводить некоторыми дополнительными оговорками. В частности, необходимо, чтобы сходимость в окрестности принятой модели была равномерной или по крайней мере была бы там односторонне равномерно ограниченной, поскольку в противном случае нельзя гарантировать робастность при каждом конечном n независимо от того, насколько велико n. Эта тонкость в прошлом нередко упускалась из виду.

Следует еще раз подчеркнуть, что появление больших ошибок в наблюдениях, составляющих небольшую долю выборки, нужно трактовать как малое отклонение и что главное назначение робастных процедур в отличие от ряда крайне чувствительных классических процедур — исключить влияние больших ошибок.

В литературе можно встретить и немало других явных и неявных требований к робастным процедурам. Например, от них требуют высокой асимптотической *относительной эффективности* (по отношению к некоторым конкретным классическим процедурам) или высокой *абсолютной эффективности* как для совершенно произвольных (достаточно гладких) предполагаемых распределений, так и для конкретных параметрических семейств.

Однако, на наш взгляд, требования такого рода имеют второстепенное значение и никогда не должны занимать главенствующее положение по отношению к трем упомянутым выше.

Робастные и непараметрические процедуры. Свободные от распределения критерии. В традиционной классификации робастные процедуры относят к тому же классу, который содержит непараметрические и свободные от распределения процедуры. По нашему мнению, эти три вида процедур имеют очень мало общего.

Процедуру называют *непараметрической*, если она предназначена для применения к ограниченному непараметризованному множеству допустимых распределений. Например, выборочные среднее и дисперсия служат непараметрическими оценками истинных среднего и дисперсии соответственно. Будучи непараметрической оценкой, выборочное среднее весьма чувствительно к выделяющимся наблюдениям и, следовательно, не робастно. В тех относительно редких случаях, когда оценкой истинного среднего интересуются специально, ничего больше не остается, как воспользоваться выборочным средним «на авось».

Критерий называют *свободным от распределения*, если вероятность ошибочно отвергнуть нулевую гипотезу одна и та же для всех возможных непрерывных распределений основной гипотезы (оптимальная робастность справедливости гипотезы). Типичными примерами критериев такого рода служат двухвыборочные ранговые критерии проверки гипотезы идентичности распределений. Как оказалось, бо́льшая часть критериев, свободных от распределения, обладает достаточно стабильной мощностью, а следовательно, имеет хорошую робастность всех характеристик. Но этот факт — не более чем счастливое совпадение, так как свобода от распределения никак не отражается на поведении функции мощности.

Оценки, получаемые в свободных от распределения критериях, иногда тоже называют свободными от распределения. Однако здесь налицо неправильное употребление термина: случайное поведение точечных оценок тесно связано с мощностью (но не уровнем) критериев, в которых они возникают, и зависит от истинного распределения. Единственным исключением служат интервальные оценки, получаемые в ранговых критериях. Например, истинная медиана попадает в интервал между двумя определенными выборочными квантилями с фиксированной вероятностью (но зависимость распределения длины этого интервала от действительного распределения остается).

Робастные методы, как они понимаются в этой книге, в идейном плане намного ближе к классическим параметрическим методам, чем к непараметрическим или свободным от распределения методам. Особенность робастных методов, предназначенных для работы с параметрическими моделями, состоит в том, что эти модели уже не предполагаются совершенно точными, и это влечет за собой в свою очередь необходимость соответствующих формальных построений.

Приведенные соображения оправдывают наше намерение относить робастные оценки к *состоятельным оценкам* неизвестных параметров *в идеальной модели*. В силу робастности эти оценки не могут отклоняться слишком сильно, если модель верна лишь приближенно. В этом случае вне модели можно *определять* параметр, который должен оцениваться, по предельному значению оценки. Например, если используется выборочная медиана, то ею естественно оценивать истинную медиану и т. д.

Адаптивные процедуры. Возможность создания эффективных непараметрических критериев и оценок обнаружена Стейном (1956). Позднее ряд авторов, в частности Такеути (1971), Биран (1974), Сакс (1975) и Стоун (1975), рассмотрели конкретные оценки сдвига, асимптотически эффективные для всех достаточно гладких симметричных плотностей. Поскольку эти оценки, если можно так выразиться, сами адаптируются к основному распределению, они получили известность под