Math-Net.Ru

Общероссийский математический портал

М. В. Бурнашев, Асимптотические разложения для медианной оценки параметра, *Теория вероятн. и ее примен.*, 1996, том 41, выпуск 4, 738–754

DOI: <https://doi.org/10.4213/tvp3193>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 107.22.136.117

26 апреля 2023 г., 17:45:15

© 1996 г.

БУРНАШЕВ М. В.*

АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ
ДЛЯ МЕДИАННОЙ ОЦЕНКИ ПАРАМЕТРА¹⁾

Получены асимптотические разложения для функции распределения и моментов медианной оценки параметра по наблюдениям в аддитивном шуме с симметричной плотностью. Для лапласовской (т.е. двусторонней экспоненциальной) плотности эта оценка совпадает с оценкой максимального правдоподобия. Приведенные численные сравнения с точными данными показывают, что использование асимптотических разложений значительно улучшает точность статистических выводов уже при сравнительно небольших объемах выборок.

Ключевые слова и фразы: медианная оценка, асимптотические разложения.

1. Постановка задачи и основные результаты

Пусть наши наблюдения $x_1, \dots, x_n$ имеют вид

$$x_i = \theta + \xi_i; \quad i = 1, \dots, n; \quad \theta \in (-\infty, \infty),$$

где $\xi_1, \dots, \xi_n$ суть независимые одинаково распределенные случайные величины с известной симметричной плотностью $f(x)$ и соответствующей функцией распределения $F(x)$, а θ — неизвестный параметр сдвига, подлежащий оцениванию.

Упорядочим наши наблюдения в возрастающем порядке, т.е. $x_{[1]} \leq x_{[2]} \leq \dots \leq x_{[n]}$, и определим медианную оценку (МО) $\hat{\theta}_n$, положив $\hat{\theta}_n = x_{[(n+1)/2]}$ для нечетных $n = 2k+1$ и $\hat{\theta}_n = (x_{[n/2]} + x_{[(n+2)/2]})/2$ для четных $n = 2k$. Другими словами, медианная оценка $\hat{\theta}_n$ для нечетных n есть центральный член вариационного ряда $x_{[1]} \leq x_{[2]} \leq \dots \leq x_{[n]}$.

Будем использовать также обозначения

$$F_n(x) = P_{\theta}\{\hat{\theta}_n - \theta \leq x\}, \quad p_n(x) = \frac{d}{dx} F_n(x).$$

*Институт проблем передачи информации РАН, ул. Еромоловой, 19, 101447 Москва, Россия.

¹⁾Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант 95-01-00136а, и Международного научного фонда, грант J3M100.

Хорошо известно [2], [3, п. 28.5], [4, п. II.1.2], что при некоторых естественных предположениях справедливы асимптотические соотношения

$$F_n \left(\frac{x}{2f(0)\sqrt{n}} \right) = P_\theta \left\{ 2f(0)\sqrt{n}(\hat{\theta}_n - \theta) < x \right\} = \Phi(x) + o(1), \quad n \rightarrow \infty, \quad (1)$$

($\Phi(x)$ — стандартная гауссовская функция распределения) и

$$4f^2(0)nE_\theta \left(\hat{\theta}_n - \theta \right)^2 = 1 + o(1), \quad n \rightarrow \infty. \quad (2)$$

В данной работе получены асимптотические разложения (по степеням $n^{-1/2}$) для остаточных членов $o(1)$ в соотношениях (1), (2). Чтобы сохранить необходимую точность, но избежать излишне громоздких выражений, мы ограничиваемся получением членов порядков $n^{-1/2}$ и n^{-1} .

Одна из главных причин нашего интереса к медианным оценкам состоит в том, что в случае лапласовской плотности

$$l(x) = \frac{1}{2} e^{-|x|}, \quad x \in (-\infty, \infty),$$

эта оценка совпадает с оценкой максимального правдоподобия $\hat{\theta}_n$ (мы не смогли найти опубликованное доказательство этого простого и хорошо известного факта и поэтому для полноты доказываем его в приложении). Лапласовская же плотность уже сама по себе заслуживает специального внимания в силу определенных теоретических причин [1] (см. ниже) и в силу ее распространенности в определенных статистических приложениях и в радиолокации [10]–[12].

Следует сразу отметить, что рассматриваемые здесь типы асимптотических разложений хорошо известны [5]–[9] для оценок максимального правдоподобия и байесовских в том случае, если плотность $f(x)$ является достаточно гладкой функцией. Например, если плотность $f(x)$ имеет четыре непрерывные ограниченные производные и конечную фишеровскую информацию $I(f) = \int (f'^2/f) dx$, то остаточный член $o(1)$ в (2) (с оценкой максимального правдоподобия $\hat{\theta}_n$ вместо $\check{\theta}_n$ и $I(f)$ вместо $2f(0)$) имеет вид $A/n + O(n^{-2})$ с некоторым явным выражением для постоянной $A = A(f, f', f'')$. Такая достаточно быстрая скорость сходимости (как n^{-1}) в (2) для достаточно гладких плотностей $f(x)$ служит разумным основанием для простой аппроксимации $E_\theta(\hat{\theta}_n - \theta)^2 \approx 1/(nI(f))$ уже при умеренных объемах выборок n (например, для $n \sim 20$ –50) (хотя использование и следующего члена A/n дает все же лучшую аппроксимацию).

Так как лапласовская плотность $l(x)$ имеет разрывную первую производную $l'(x)$, скорость сходимости в (1), (2) для нее значительно медленнее. Еще Фишер [1] более 70 лет назад заметил, что в этом случае остаточный член $o(1)$ в (2) имеет порядок $n^{-1/2}$ и поэтому пренебрежение остаточными членами $o(1)$ в (1), (2) может вызвать ощутимую неточность при умеренных объемах выборок. Он писал в этой связи

[1, с. 717], что «этот пример подчеркивает важность исследования действительного поведения статистики в конечных выборках, вместо того чтобы полностью полагаться на ее вычисленное поведение в бесконечно больших выборках».

Действительно, численные примеры из п. 4 ниже показывают, что для лапласовской плотности остаточный член $o(1)$ из (2) вносит вклад менее 10% истинного значения среднеквадратичного риска только для объема выборки $n > 250$. С другой стороны, остаточный член $O(n^{-3/2})$ из аппроксимации (4) ниже вносит вклад менее 10% истинного значения этого риска для выборок, начиная с объема $n = 8$. Другие численные примеры такой медленной скорости сходимости для лапласовской плотности можно найти в [10]–[12], где рассматривалась близкая задача обнаружения постоянного сигнала на фоне аддитивного шума (эта задача эквивалентна проверке нулевой гипотезы против односторонней альтернативы).

Насколько нам известно, по определенным причинам (поясненным ниже) не было серьезных попыток получить асимптотические разложения для остаточных членов $o(1)$ в соотношениях (1), (2), хотя лапласовский случай стал стандартным примером медленной скорости сходимости. В этой связи мы очень рекомендуем прекрасно написанную обзорную работу [7].

Поясним некоторые математические трудности, отличающие изучаемую здесь задачу от рассмотренных ранее в [5]–[9] подобных же задач. При получении асимптотических разложений для остаточных членов $o(1)$ в соотношениях (1), (2) для достаточно гладких плотностей $f(x)$ обычно используется один из следующих двух подходов.

Если возникающий процесс отношения правдоподобия имеет достаточное число непрерывных производных, то можно разложить его по формуле Тейлора (в малой окрестности истинного значения параметра) и исследовать затем остаточный член этого разложения (как случайный процесс). В лапласовском случае соответствующий процесс отношения правдоподобия является абсолютно непрерывной функцией, но его производная имеет n точек разрыва, что делает такой подход довольно неудобным.

Другой подход является гораздо более трудоемким. Он состоит в применении асимптотических разложений в центральной предельной теореме непосредственно к процессу отношения правдоподобия в «фиксированной параметрической точке» (тогда для каждой такой точки возникает свой остаточный член) и дальнейшем получении равномерно «хорошей» (одновременно для всех «параметрических» точек) оценки остаточного члена (как случайного процесса). В лапласовском случае этот подход выглядит еще более неприятным, так как случайные величины, дающие здесь в сумме логарифм отношения правдоподобия, имеют распределение с ненулевыми абсолютно непрерывной и решетчатой компо-

нентами, причем абсолютно непрерывная компонента может быть произвольно малой. Другими словами, мы оказываемся в точности между двумя крайними исследованными случаями в теории асимптотических разложений в центральной предельной теореме [13]. (Отметим, что автору не известна ни одна работа, в которой второй подход был бы доведен до конца.)

Следует также упомянуть здесь работы [14], [15], где на неформальном эвристическом уровне для лапласовского случая были получены члены порядка $n^{-1/2}$ в соотношениях (1), (2). Их авторы пытались применить нечто подобное второму подходу, однако они получили разложения только в «фиксированной» точке и не уделили никакого внимания равномерным оценкам, что составляет здесь основную трудность (см. также критические комментарии в [7, с. 33–34]).

В данной работе используется прямой комбинаторный подход. Мы не используем никаких предельных теорем. Более того, все аналитические методы в статье были известны уже в 18-м веке (что упрощает проблемы ссылок). Оказывается, что естественным нормировочным коэффициентом в данной задаче является $\sqrt{2k}$, где $k = [n/2]$ (здесь $[x]$ обозначает целую часть числа x). Поэтому при нечетном n нормировочный коэффициент есть $\sqrt{n-1}$, а при четном n он есть $\sqrt{n}$. Это позволит нам, например, в лапласовском случае получить следующие асимптотические разложения для оценки максимального правдоподобия $\hat{\theta}_n$:

$$P_{\theta} \left\{ \sqrt{2k}(\hat{\theta}_n - \theta) < x \right\} = \Phi(x) - \frac{x|x|e^{-x^2/2}}{4\sqrt{\pi k}} + \frac{x(18 + 10x^2 - 3x^4)e^{-x^2/2}}{48k\sqrt{2\pi}} + x^2(1 + x^6)e^{-x^2/2}O(n^{-3/2}) + O(n^{-2}) \quad (3)$$

и

$$2kE_{\theta}(\hat{\theta}_n - \theta)^2 = 1 + \frac{2}{\sqrt{\pi k}} - \frac{1}{8k} + O(n^{-3/2}), \quad (4)$$

где все $O(\cdot)$ зависят только от n .

Немедленным следствием такой нормировки является тот факт, что среднеквадратичная ошибка $E_{\theta}(\hat{\theta}_n - \theta)^2$ должна убывать значительно быстрее после каждого четного наблюдения, чем после смежного нечетного наблюдения. Этот любопытный факт ясно виден из численных результатов в п. 4.

Вообще говоря, несмотря на очевидные достоинства, медианная оценка не является эффективной (за исключением лапласовской плотности). Тем не менее, асимптотические разложения для нее интересны и с чисто теоретической точки зрения. Дело в том, что, к сожалению, в случае аналогичных разложений для эффективных оценок (максимального правдоподобия и байесовских) [5]–[9] нет примеров численных сравнений с действительными данными (по крайней мере, мы никогда их не видели), хотя имеется определенная уверенность (см., например, [7]),

что использование таких разложений может значительно улучшить точность статистических выводов. В случае же медианной оценки такие сравнения можно провести (см. п. 4), и вполне благоприятные выводы, которые следуют из этих сравнений, служат в качестве дополнительных аргументов в поддержку использования асимптотических разложений и для эффективных оценок.

Мы считали также, что если А. Н. Колмогоров [2] проявлял определенный интерес к медианным оценкам, то для нас тем более является большим удовольствием развить его результаты в этом направлении, причем используя аналогичный комбинаторный подход.

Статья организована следующим образом. В п. 2 мы получаем асимптотические разложения для функции распределения $F_n(x)$ (теорема 1). Пункт 3 содержит результаты о «больших отклонениях» (теорема 2) и асимптотические разложения для среднеквадратичной ошибки (следствие). В п. 4 в качестве примеров мы рассматриваем распределения Лапласа и Коши и сравниваем точные среднеквадратичные ошибки с различными аппроксимациями для них. Пункт 5 содержит некоторые замечания.

Предположения. В статье предполагается, что:

(A1) $f(x)$, $x \in \mathbf{R}^1$, есть симметричная плотность с $f(0) > 0$;

(A2) $f(x)$ имеет три непрерывные ограниченные производные в некоторой окрестности $(0, x_0]$;

(A3) существуют такие положительные постоянные K и δ , что функция распределения $F(x)$ удовлетворяет следующему условию: $1 - F(x) \leq Kx^{-\delta}$, $x > 0$.

Мы также используем обозначения $f = f(0)$, $f' = f'(+0)$, $f'' = f''(+0)$.

2. Функция распределения

Следующая теорема составляет основной результат данного раздела.

Теорема 1. Если выполнены условия (A1), (A2), то функция распределения $F_n(x)$ медианной оценки $\hat{\theta}_n$ имеет следующее асимптотическое разложение ($k = [n/2]$):

$$F_n\left(\frac{x}{2f\sqrt{2k}}\right) = \Phi(x) + \frac{f'x|x|e^{-x^2/2}}{8f^2\sqrt{\pi k}} + \frac{xe^{-x^2/2}}{8k\sqrt{2\pi}} \left[3 + x^2 + \frac{x^2 f''}{6f^3} - \frac{x^4 (f')^2}{8f^4} \right] + x^2(1 + x^6)e^{-x^2/2}O(n^{-3/2}) + O(n^{-2}). \quad (5)$$

В частности,

$$F_{2k}(x) = F_{2k+1}(x) + O(n^{-3/2}), \quad (6)$$

причем все $O(\cdot)$ в (5), (6) не зависят от x .

Замечания. 1) Хотя разложение (5) справедливо для любого

$x \in (-\infty, \infty)$, оно представляет интерес только для $x = O(\sqrt{\log n})$. Большие значения x рассматриваются в следующем разделе.

2) Из соотношения (6) вытекает, что для медианной оценки каждое нечетное наблюдение добавляет «количество информации» порядка $n^{-3/2}$ (а не n^{-1} , как обычно!). См. также предсказуемое следствие для моментов в следующем разделе.

Для доказательства этой теоремы необходимо различать случаи нечетного и четного объема выборки n . Хотя оба эти случая похожи, второй из них требует несколько больших вычислений. Поэтому мы рассмотрим эти случаи отдельно и начнем с нечетного n . Часть доказательства ниже будет использована в следующем разделе для получения результатов о «больших отклонениях».

а) n — нечетное. Пусть $n = 2k + 1$ — нечетное число. Тогда, очевидно,

$$F_{2k+1}(a) = P_0\{\hat{\theta}_n \leq a\} = \sum_{i=0}^k \binom{2k+1}{i} [1 - F(a)]^i F^{2k+1-i}(a). \quad (7)$$

Мы, однако, используем другое — более удобное — представление для $F_n(a)$, которое получается с помощью плотности $p_n(a)$. Легко проверить, что

$$p_n(a) = (2k+1)f(a) \binom{2k}{k} [F(a)(1 - F(a))]^k. \quad (8)$$

Действительно, для $\delta \rightarrow +0$ имеем

$$\begin{aligned} P_\theta \{a < \hat{\theta}_n - \theta < a + \delta\} &= (2k+1) \binom{2k}{k} [F(a)(1 - F(a + \delta))]^k \\ &\quad \times \int_a^{a+\delta} f(x) dx + O(\delta^2), \end{aligned}$$

откуда следует формула (8). Кроме того, в силу непрерывности функции $p_n(a)$, функция $F_n(a)$ является абсолютно непрерывной с производной $p_n(a)$. (См. также [3, 28.5], [4, II.1.2].) Теперь из (8) получаем

$$\begin{aligned} F_{2k+1}(a) &= (2k+1) \binom{2k}{k} \int_{-\infty}^a [F(x)(1 - F(x))]^k dF(x) \\ &= (2k+1) \binom{2k}{k} \int_0^{F(a)} [t(1-t)]^k dt \\ &= (2k+1) 2^{-2k-1} \binom{2k}{k} \int_{1-2F(a)}^1 (1-x^2)^k dx. \end{aligned} \quad (9)$$

Эквивалентность соотношений (7) и (9) можно проверить также и независимо с помощью основного комбинаторного тождества

$$\binom{n}{m} = \binom{n-1}{m} + \binom{n-1}{m-1}.$$

Формулу (9) можно записать в терминах неполной бета-функции $B_x(\cdot, \cdot)$ [16]:

$$F_{k+1}(a) = \frac{B_{F(a)}(k+1, k+1)}{B(k+1, k+1)}, \quad (2k+1) \binom{2k}{k} B(k+1, k+1) = 1. \quad (10)$$

Теперь, учитывая, что

$$(2k+1) 2^{-2k-1} \binom{2k}{k} \int_0^1 (1-x^2)^k dx = \frac{1}{2},$$

получаем из (9) для $a \geq 0$

$$\begin{aligned} F_{2k+1}(a) &= \frac{1}{2} + (2k+1) 2^{-2k-1} \binom{2k}{k} \int_0^{2F(a)-1} (1-x^2)^k dx \\ &= \frac{1}{2} + A_{2k+1} \int_0^{z(a)} \left(1 - \frac{x^2}{2k}\right)^k dx, \end{aligned} \quad (11)$$

где (это следует из формулы Стирлинга)

$$\begin{aligned} A_{2k+1} &= \frac{(2k+1) 2^{-2k}}{\sqrt{8k}} \binom{2k}{k} = \frac{(2k+1)}{2k\sqrt{2\pi}} \exp \left\{ -\frac{1}{8k} + O(k^{-3}) \right\}, \\ z &\equiv z(a) = (2F(a) - 1)\sqrt{2k}. \end{aligned} \quad (12)$$

Ясно, что асимптотическое поведение функции $F_n(a)$ будет зависеть от величины $a = a(n)$. При относительно небольших значениях величины $a(n)$ мы получим «нормальную» аппроксимацию, а при достаточно больших значениях величины $a(n)$ получим аппроксимацию через «большие отклонения». Нашей основной целью является получение асимптотических разложений в «нормальной» зоне и просто точной асимптотики для «больших отклонений».

Начнем с «нормального» поведения. С помощью формулы Тейлора получаем

$$\begin{aligned} \left(1 - \frac{x^2}{2k}\right)^k &= \left(1 - \frac{x^4}{8k} + \frac{r_1(x)}{k^2}\right) e^{-x^2/2}, \\ -\frac{x^6}{24} \left(1 - \frac{x^2}{2k}\right)^{-3} &\leq r_1(x) \leq \frac{x^8}{128} \left(1 - \frac{x^2}{2k}\right)^{-4}, \quad x^2 < 2k. \end{aligned} \quad (13)$$

Учитывая, что

$$\int_0^x u^4 e^{-u^2/2} du = -(x^3 + 3x)e^{-x^2/2} + 3 \int_0^x e^{-u^2/2} du$$

и

$$\int_0^z x^{2m} e^{-x^2/2} dx \leq \min \{z^{2m}, (2m-1)!!\} \int_0^z e^{-x^2/2} dx,$$

получаем из (13)

$$\begin{aligned} \int_0^z \left(1 - \frac{x^2}{2k}\right)^k dx &= \left(1 - \frac{3}{8k} + \frac{r_2(z)}{k^2}\right) \int_0^z e^{-x^2/2} dx \\ &\quad + \frac{(z^3 + 3z)}{8k} e^{-z^2/2}, \\ |r_2(z)| &\leq \frac{\min(1, z^6)}{(1 - z^2/(2k))^4} = \frac{\min(1, z^6)}{256[F(a)(1 - F(a))]^4} \\ &\leq \frac{\min(1, z^6)}{16[1 - F(a)]^4}. \end{aligned} \quad (14)$$

В результате из (11), (12) и (14) мы получаем

$$\begin{aligned} F_{2k+1}(a) &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left(1 + \frac{r(z)}{k^2} + O(k^{-2})\right) \int_0^{z(a)} e^{-x^2/2} dx \\ &\quad + \frac{(z^3 + 3z)}{8k\sqrt{2\pi}} e^{-z^2/2} (1 + O(k^{-1})) \\ &= \Phi(z(a)) + \frac{(z^3 + 3z)}{8k\sqrt{2\pi}} e^{-z^2/2} + \frac{r(z)}{k^2} + O(k^{-2}), \\ |r(z)| &\leq \frac{1}{32[1 - F(a)]^4}, \end{aligned} \quad (15)$$

где $z = z(a)$ определено в (12) и $O(k^{-2})$ не зависит от a .

Формула (15) уже дает асимптотическое разложение для функции распределения $F_{2k+1}(a)$ через функцию $z(a)$. Чтобы представить ее в более привычной нормализованной форме, заменим аргумент a величиной $b/(2f(0)\sqrt{2k})$ и воспользуемся формулой Тейлора (имея в виду предположения (A1), (A2.)) (Из (11), (12) ясно, что величина $\sqrt{2k} = \sqrt{n-1}$ является здесь более естественным нормировочным коэффициентом, чем традиционное $\sqrt{n}$.) Тогда для $0 \leq b \leq x_0\sqrt{2k}$ получаем

$$\begin{aligned} z\left(\frac{b}{2f(0)\sqrt{2k}}\right) &= \left(2F\left(\frac{b}{2f(0)\sqrt{2k}}\right) - 1\right) \sqrt{2k} \\ &= b + \frac{b^2 f'}{4f^2\sqrt{2k}} + \frac{b^3 f''}{48kf^3} + b^4 O(n^{-3/2}), \end{aligned} \quad (16)$$

где $f' = f'(+0)$, $f'' = f''(+0)$. Кроме того, справедлива формула ($\varepsilon \rightarrow 0$)

$$\Phi(x + \varepsilon) = \Phi(x) + \frac{\varepsilon}{\sqrt{2\pi}} \left(1 - \frac{\varepsilon x}{2}\right) e^{-x^2/2} + (x^2 + 1) e^{-x^2/2 - \varepsilon x} O(\varepsilon^3). \quad (17)$$

В результате после некоторых вычислений получаем из (15)–(17) для любого $b \geq 0$ (в силу монотонности функции $F_n(x)$)

$$\begin{aligned} F_{2k+1}\left(\frac{b}{2f\sqrt{2k}}\right) &= \Phi(b) + \frac{f'b^2 e^{-b^2/2}}{8f^2\sqrt{\pi k}} + \frac{be^{-b^2/2}}{8k\sqrt{2\pi}} \left[3 + b^2 + \frac{b^2 f''}{6f^3} - \frac{b^4 f'^2}{8f^4}\right] \\ &\quad + (1 + b^6) e^{-b^2/2} O(b^2 n^{-3/2}) + O(n^{-2}). \end{aligned} \quad (18)$$

Учитывая, что $F(x) = 1 - F(-x)$, получаем из (18) разложение (5).

б) n — ч е т н о е. Пусть теперь $n = 2k$ — четное число. Обозначим $p_{2k}(x, y)$ совместную плотность распределения пары случайных величин $(x_{[k]}, x_{[k+1]})$. Тогда, аналогично представлению (8), имеем для $x \leq y$

$$p_{2k}(x, y) = n(n-1)f(x)f(y) \binom{2k-2}{k-1} [F(x)(1-F(y))]^{k-1}. \quad (19)$$

Для плотности $p_{2k}(a)$ медианной оценки $\check{\theta}_n = (x_{[k]} + x_{[k+1]})/2$ получаем из (19) после интегрирования

$$p_{2k}(a) = nk \binom{2k}{k} \int_{-\infty}^a f(x)f(2a-x) [F(x)(1-F(2a-x))]^{k-1} dx. \quad (20)$$

Для функции распределения получаем из (20) после некоторых вычислений:

$$F_{2k}(z) = k \binom{2k}{k} \left\{ 2^{-n} \int_{-1}^{2F(z)-1} (1-t^2)^{k-1} dt + \frac{1}{n} [F(z)F(-z)]^k - \int_{-\infty}^z f(t)F^{k-1}(t)F^k(t-2z) dt \right\}. \quad (21)$$

При $z \geq 0$ правую часть (21) можно продолжить следующим образом:

$$\begin{aligned} F_{2k}(z) &= \frac{1}{2} + k \binom{2k}{k} \left\{ 2^{-n} \int_0^{2F(z)-1} (1-t^2)^{k-1} dt + \frac{1}{n} [F(z)F(-z)]^k - \int_{-\infty}^z f(t)F^{k-1}(t)F^k(t-2z) dt \right\} \\ &= \frac{1}{2} + k \binom{2k}{k} \left\{ 2^{-n} \frac{(n+1)}{n} \int_0^{2F(z)-1} (1-t^2)^k dt + \frac{1}{k} F(-z)[F(z)F(-z)]^k - \int_{-\infty}^z f(t)F^{k-1}(t)F^k(t-2z) dt \right\}. \end{aligned}$$

Сравнивая это выражение с представлением (11) и выполняя некоторые преобразования, получаем

$$\begin{aligned} F_{2k}(z) &= F_{2k+1}(z) + \frac{1}{2} \binom{2k}{k} \int_{-\infty}^z g(t, z) d[F(t)F(t-2z)]^k, \quad (22) \\ g(t, z) &= \frac{f(t-2z)F(t) - f(t)F(t-2z)}{f(t-2z)F(t) + f(t)F(t-2z)} + 1 - 2F(z). \end{aligned}$$

Нам остается оценить интеграл в правой части (22). Для этого заметим сначала, что так как $|g(t, z)| \leq 1$, то

$$\left| \int_{-\infty}^{-t_0} g(t, z) d[F(t)F(t-2z)]^k \right| \leq [F(-t_0)]^{2k} = 2^{-n} O(n^{-1}), \quad (23)$$

$$t_0 = \frac{\log n}{2fn}.$$

С другой стороны, с помощью формулы Тейлора имеем

$$g(t, z) = \frac{f'}{2f} (2z - t - |t|) + O(t^2 + z^2) \quad (t, z \rightarrow 0). \quad (24)$$

Теперь, интегрируя и используя формулы (23), (24), получаем

$$\begin{aligned} 2^n \int_{-\infty}^z g(t, z) d[F(t) F(t - 2z)]^k &= \\ &= 2^{n-1} \frac{f'}{f} \int_{-t_0}^z (2z - t - |t|) d[F(t) F(t - 2z)]^k + O(n^{-1}) \\ &= 2^n \frac{f'}{f} \int_0^z [F(t) F(t - 2z)]^k dt + O(n^{-1}) = O(n^{-1}). \end{aligned} \quad (25)$$

Так как $\binom{2k}{k} 2^{-n} = O(n^{-1/2})$, то в результате из (22) и (25) получаем

$$F_{2k}(z) = F_{2k+1}(z) + O(n^{-3/2}), \quad (26)$$

откуда следуют соотношения (5), (6).

3. Большие уклонения и моменты

Следующая теорема составляет основной результат данного раздела.

Теорема 2. Для любой симметричной плотности $f(z)$, любых $x > 0$ и $n \geq 1$ справедлива следующая оценка свертку:

$$F_n(-x) = 1 - F_n(x) \leq \sqrt{\frac{2}{\pi n}} \frac{F(x)[4F(x)F(-x)]^{n/2}}{2F(x) - 1}. \quad (27)$$

Кроме того, для $n = 2k + 1$ и $x > 0$ справедливо равенство

$$1 - F_{2k+1}(x) = \frac{[4F(x)F(-x)]^{k+1}}{(2F(x) - 1)\sqrt{2\pi n}} \left[1 + \left(1 + \frac{1}{x^2} \right) O(n^{-1}) \right]. \quad (28)$$

В частности, из теорем 1, 2 вытекает

Следствие. Если выполнены предположения (A1)–(A3), то справедливо следующее асимптотическое разложение ($k = [n/2]$):

$$8f^2 k E_0(\tilde{\theta}_n^2) = 1 - \frac{f'}{f^2 \sqrt{\pi k}} - \frac{1}{2k} \left(3 + \frac{f''}{4f^3} - \frac{15f'^2}{16f^4} \right) + O(n^{-3/2}). \quad (29)$$

З а м е ч а н и я. 1) Легко видеть, что теорема 2 дает более точный результат, чем теорема 1, если $x^{-2} = O(\log n)$.

2) Из формулы (29) следует, что

$$E_0(\tilde{\theta}_{2k}^2) - E_0(\tilde{\theta}_{2k+1}^2) = O(k^{-5/2}),$$

в то время как

$$E_0(\theta_{2k-1}^2) - E_0(\theta_{2k}^2) = \frac{1}{8f^2k^2} + O(n^{-5/2}).$$

Другими словами, среднеквадратичная ошибка убывает в основном после четных наблюдений. К сожалению, мы не можем иметь только четные наблюдения. Этот эффект можно ясно видеть из численных результатов в следующем разделе.

Доказательство теоремы 2 начнем с нечетных $n = 2k + 1$. Тогда из (10), (11) для $x > 0$ имеем

$$1 - F_{2k+1}(x) = (2k + 1)2^{-2k-1} \binom{2k}{k} \int_{2F(x)-1}^1 (1 - u^2)^k du. \quad (30)$$

Для любого $k \geq 1$ справедливы следующие неравенства:

$$\frac{k+1}{\sqrt{\pi(k+2)}} \leq (2k+1)2^{-2k-1} \binom{2k}{k} \leq \sqrt{\frac{k+1}{\pi}}. \quad (31)$$

Например, чтобы доказать второе из неравенств (31), достаточно показать, что

$$B_k = \frac{2k+1}{\sqrt{k+1}} 2^{-2k-1} \binom{2k}{k} \leq \frac{1}{\sqrt{\pi}}. \quad (32)$$

Легко проверяется, что $B_{k+1}/B_k > 1$ для любого k . Поэтому с помощью формулы Стирлинга получаем

$$B_k < \lim_{k \rightarrow \infty} B_k = \frac{1}{\sqrt{\pi}},$$

откуда следуют неравенство (32) и второе из неравенств (31). Аналогично доказывается и первое из неравенств (31).

Оценим теперь интеграл в правой части (30). Сначала оценим его сверху:

$$\int_A^1 (1 - u^2)^k du \leq \frac{(1 - A^2)^{k+1}}{2A(k+1)}, \quad (33)$$

что вместе с (31) дает неравенство

$$1 - F_{2k+1}(x) \leq \frac{[4F(x)F(-x)]^{k+1}}{2(2F(x) - 1)\sqrt{\pi(k+1)}}, \quad x > 0, \quad (34)$$

откуда следует оценка сверху (27) для нечетных n .

Чтобы оценить интеграл в (30) снизу, заметим, что

$$\max_{A \leq u \leq 1} u(1 - u^2)^m = A(1 - A^2)^m, \quad m = \frac{1 - A^2}{2A^2}.$$

Тогда имеем

$$\begin{aligned} \int_A^1 (1-u^2)^k du &\geq \frac{1}{A(1-A^2)^m} \int_A^1 u(1-u^2)^{k+m} du \\ &= \frac{(1-A^2)^{k+1}}{2A(k+m+1)} \geq \frac{(1-A^2)^{k+1}}{2A(k+1)} \left[1 - \frac{1}{2A^2k} \right]. \end{aligned} \quad (35)$$

Теперь из (30), (31) и (35) получаем

$$\begin{aligned} 1 - F_{2k+1}(x) &\geq \frac{[4F(x)F(-x)]^{k+1}}{2(2F(x)-1)\sqrt{\pi(k+2)}} \\ &\times \left[1 - \frac{1}{2k(2F(x)-1)^2} \right], \quad x > 0. \end{aligned} \quad (36)$$

В результате, из (34) и (36) следует формула (28).

Рассмотрим теперь четные $n = 2k$. Тогда из (21), (22) имеем

$$\begin{aligned} 1 - F_{2k} &= 1 - F_{2k+1} + \binom{2k}{k} [4F(x)F(-x)]^k r(x, k), \\ -F(-x) &\leq r(x, k) \leq F(x). \end{aligned} \quad (37)$$

Используя в (37) оценку (31), получаем в результате из (34) и (37) оценку (27) для четных n .

Для доказательства следствия заметим, что в силу предположений (A1), (A2) функция $4F(x)F(-x)$ при малых x ведет себя как $1 - 4f^2(0)x^2$. Из этого факта и предположения (A3) следует, что для некоторой строго положительной постоянной K_1 и для любых x выполняется следующее условие:

$$4F(x)F(-x) \leq \frac{1}{1 + K_1 \min\{x^2, |x|^\delta\}}. \quad (38)$$

Оценим теперь вклад в среднеквадратичный риск больших значений величины $|\sqrt{n}\check{\theta}_n|$. Используя обычное обозначение

$$E[\xi, B] = \int_B x dF_\xi(x)$$

и интегрируя по частям, имеем: для любого $A > 0$

$$E_0 \left[|\sqrt{n}\check{\theta}_n|^2, |\sqrt{n}\check{\theta}_n| > A \right] \leq 2A^2 [1 - F(A)] + 4 \int_A^\infty x [1 - F(x)] dx. \quad (39)$$

Используя в последнем соотношении оценки (27) и (38), получаем, что существует такая постоянная $D = D(K, \delta)$, что

$$E_0 \left[|\sqrt{n}\check{\theta}_n|^2, |\sqrt{n}\check{\theta}_n| > D\sqrt{\log n} \right] = O(n^{-2}). \quad (40)$$

Следовательно, чтобы получить первые три члена асимптотического разложения для среднеквадратичного риска, мы можем принимать во

внимание только значения $|\sqrt{n}\hat{\theta}_n| \leq D\sqrt{\log n}$ и использовать для этой цели формулу (5). В результате, после стандартных вычислений получаем соотношение (29), завершающее доказательство следствия.

Аналогично можно рассмотреть и другие моменты.

4. Примеры и численные сравнения

В этом разделе мы рассмотрим частные случаи теорем 1, 2 для лапласовской плотности (наш основной пример) и плотности Коши (как гладкой плотности с «тяжелыми» хвостами).

а) Лапласовская плотность. Для лапласовской плотности $l(x) = e^{-|x|}/2$ с фишеровской информацией $I(l) = 1$ оценка максимального правдоподобия $\hat{\theta}_n$ совпадает с медианной оценкой $\check{\theta}_n$ (см. приложение) и обе являются эффективными. Поэтому из теоремы 1 и следствия получаем

Предложение. Для лапласовской плотности $l(x)$ справедливы следующие асимптотические разложения ($k = [n/2]$):

$$\begin{aligned} P_{\theta} \left\{ \sqrt{2k}(\hat{\theta}_n - \theta) < x \right\} = \Phi(x) - \frac{x|x|e^{-x^2/2}}{4\sqrt{\pi k}} + \frac{x(18 + 10x^2 - 3x^4)e^{-x^2/2}}{48k\sqrt{2\pi}} \\ + x^2(1 + x^6)e^{-x^2/2}O(n^{-3/2}) + O(n^{-2}) \end{aligned} \quad (41)$$

и

$$2kE_{\theta} \left(\hat{\theta}_n - \theta \right)^2 = 1 + \frac{2}{\sqrt{\pi k}} - \frac{1}{8k} + O(n^{-3/2}). \quad (42)$$

Чтобы провести некоторые численные сравнения, найдем истинную среднеквадратичную ошибку $e_n = E_0(\hat{\theta}_n^2)$ и различные аппроксимации для нее. Заметим сначала, что с помощью формул (8) и (20) можно получить точные формулы для e_n (для $n = 2k + 1$ и $n = 2k$ соответственно):

$$e_n = n2^{-2k} \binom{2k}{k} \int_0^1 (1-x^2)^k (\log(1-x))^2 dx \quad (43)$$

и

$$e_n = n2^{-n} \binom{2k}{k} \int_0^1 (1-x^2)^{k-1} \log^2(1-x) dx - \frac{1}{nk} - \frac{1}{nk} 2^{1-n} \binom{2k}{k}. \quad (44)$$

В силу следствия естественно называть $e_n^{(1)} = 1/(2k)$ аппроксимацией 1-го порядка для e_n . Тогда из (41) следуют аппроксимации 2-го и 3-го порядков ($k = [n/2]$):

$$e_n^{(2)} = \frac{1}{2k} + \frac{1}{k\sqrt{\pi k}}, \quad e_n^{(3)} = \frac{1}{2k} + \frac{1}{k\sqrt{\pi k}} - \frac{1}{16k^2}. \quad (45)$$

Ниже в таблице приведены точные значения величин $e_n(L)$ и трех аппроксимаций $e_n^{(1)}(L)$, $e_n^{(2)}(L)$ и $e_n^{(3)}(L)$ для них для объема выборки

$n \leq 31$. Легко видеть, что разности $e_{2k} - e_{2k-1}$ гораздо больше, чем смежные разности $e_{2k+1} - e_{2k}$, что соответствует результатам следствия.

б) Плотность Коши. Для плотности Коши

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in (-\infty, \infty),$$

с фишеровской информацией $I(f) = \frac{1}{2}$ имеем из формулы (29)

$$\frac{8k}{\pi^2} E_0 \check{\theta}_n^2 = 1 + \frac{\pi^2 - 6}{4k} + O(n^{-3/2}). \quad (46)$$

С другой стороны, аналогично (42), (43) можно получить точные формулы для $n = 2k + 1$ и $n = 2k$ соответственно:

$$E_0 \check{\theta}_n^2 = n 2^{-2k} \binom{2k}{k} \int_0^1 (1-x^2)^k \operatorname{tg}^2\left(\frac{\pi x}{2}\right) dx \quad (47)$$

и

$$E_0 \check{\theta}_n^2 = nk 2^{-(n+2)} \binom{2k}{k} \times \int_0^1 (1-x)^{k-1} \int_{-x}^x \left[\operatorname{tg}\left(\frac{\pi x}{2}\right) + \operatorname{tg}\left(\frac{\pi z}{2}\right) \right]^2 (1+z)^{k-1} dz dx. \quad (48)$$

Аналогично лапласовскому случаю введем две аппроксимации (1-го и 3-го порядков):

$$e_n^{(1)}(C) = \frac{\pi^2}{8k}, \quad e_n^{(3)}(C) = \frac{\pi^2}{8k} + \frac{\pi^2(\pi^2 - 6)}{32k^2}. \quad (49)$$

Правая часть таблицы содержит точные значения величин $e_n(K)$ и аппроксимаций $e_n^{(1)}(K)$, $e_n^{(3)}(K)$. Как и следовало ожидать, в отличие от лапласовского случая, для плотности Коши скорость сходимости обеих аппроксимаций к истинному значению значительно быстрее. Например, аппроксимация $e_n^{(1)}(C)$ достигает 10% точности, начиная с объема выборки $n = 23$. Тем не менее, следующая аппроксимация $e_n^{(3)}(C)$ достигает такой же точности быстрее (начиная с $n = 11$). Разности $e_{2k} - e_{2k-1}$ здесь также значительно больше, чем соответствующие разности $e_{2k+1} - e_{2k}$.

5. Замечания

1) Хотя медианная оценка, вообще говоря, не является эффективной, результаты (27)–(29) о «больших отклонениях» для нее выглядят довольно привлекательными. Например, неасимптотическая и «робастная» форма оценки сверху (27) побуждает использовать медианную оценку в качестве некоторой предварительной оценки параметра, а затем скорректировать ее (каким-либо образом) в некоторой малой окрестности для достижения эффективности.

2) Численные примеры из п. 4 для плотностей с изолированными особенностями (разрыв первой производной у лапласовской плотности и «тяжелый хвост» у плотности Коши) позволяют рискнуть сделать заключение, что (исключая некоторые «очень патологические случаи») использование первых двух значащих членов асимптотического разложения дает хорошую точность (скажем, 10–15 %), начиная с объема выборки $n = 10$.

3) Нетрудно объяснить геометрически качественную разницу в скоростях сходимости к асимптотическому пределу для гладких плотностей (с $f'(0) = 0$) и плотностей с $f'(+0) \neq 0$. По существу, в асимптотическом анализе мы заменяем область под кривой $f(x)$, $-c/\sqrt{n} < x < c/\sqrt{n}$, прямоугольником (или соответствующей частью гауссовской плотности). Если $f'(0) = 0$, то относительная точность такой замены имеет порядок n^{-1} . Если же $f'(+0) \neq 0$, то эта относительная точность имеет порядок $n^{-1/2}$ (см. также [12]).

Автор благодарит Д. М. Чибисова и Винсента Пура (H. Vincent Poor) за полезные дискуссии и замечания.

Приложение

Заметим, что для лапласовской плотности $l(x)$ и $\theta \geq 0$ мы имеем

$$\log \frac{l(x-\theta)}{l(x)} = \begin{cases} -\theta, & x \leq 0, \\ 2x - \theta, & 0 \leq x \leq \theta, \\ \theta, & x \geq \theta, \end{cases}$$

и аналогично для $\theta < 0$

$$\log \frac{l(x-\theta)}{l(x)} = \begin{cases} -\theta, & x \leq \theta, \\ \theta - 2x, & \theta \leq x \leq 0, \\ \theta, & x \geq 0. \end{cases}$$

Полагая без потери общности, что истинное значение параметра $\theta_0 = 0$, получаем для логарифма отношения правдоподобия (ниже $\#\{x_i > \theta\}$ означает число наблюдений, превышающих θ , и т.д.)

$$\begin{aligned} Z_n(\theta) &= \sum_{i=0}^n \log \frac{l(x_i - \theta)}{l(x_i)} = \theta [\#\{x_i > \theta\} - \#\{x_i \leq \theta\}] + 2 \sum_{i=0}^n x_i I_{[0, \theta]} \\ &= \theta [\#\{x_i \geq 0\} - \#\{x_i < 0\}] + 2 \sum_{i=0}^n (x_i - \theta) I_{[0, \theta]}, \quad \theta \geq 0, \end{aligned}$$

и

$$Z_n(\theta) = \theta [\#\{x_i \geq \theta\} - \#\{x_i < \theta\}] - 2 \sum_{i=0}^n x_i I_{[\theta, 0]}, \quad \theta < 0.$$

Тогда для любого $\theta \neq x_i$, $i = 1, \dots, n$, имеем

$$Z'_n(\theta) = \#\{x_i > \theta\} - \#\{x_i < \theta\} = 2\#\{x_i > \theta\} - n.$$

Поэтому $Z'_n(\theta)$ является невозрастающей кусочно-постоянной функцией с отрицательными скачками величины 2 в точках $x_i, i = 1, \dots, n$, и, следовательно, $Z_n(\theta)$ является непрерывной кусочно-линейной вогнутой функцией.

Ясно, что при нечетном n процесс $Z_n(\theta)$ достигает максимального значения в точке $\hat{\theta}_n = x_{[(n+1)/2]}$. При четном n процесс $Z_n(\theta)$ принимает постоянное максимальное значение на целом интервале $(x_{[n/2]}, x_{[(n+2)/2]})$. Поэтому оценка максимального правдоподобия для лапласовской плотности совпадает с медианной оценкой.

СПИСОК ЛИТЕРАТУРЫ

1. Fisher R. A. Theory of statistical estimation. — Proc. Cambridge Philos. Soc., 1925, v. 98, p. 39–82.
2. Колмогоров А. Н. Метод медианы в теории ошибок. — Матем. сб., 1931, т. 38, № 3/4, с. 47–50.
3. Крамер Г. Математические методы статистики. М.: Мир, 1975, 648 с.
4. Гаек Я., Шудак З. Теория ранговых критериев. М.: Наука, 1971, 376 с.
5. Linnik Ju. V., Mitrofanova N. M. Some asymptotic expansions for the distribution of the maximum likelihood estimate. — Sankhyā, Ser. A, 1965, v. 27, p. 73–82.
6. Чибисов Д. М. Асимптотическое разложение для одного класса оценок, включающего оценки максимального правдоподобия. — Теория вероятн. и ее примен., 1973, т. XVIII, № 4, с. 689–702.
7. Pfanzagl J. Asymptotic expansions in parametric statistical theory. — In: Developments in Statistics. Ed. by P. R. Krishnaiah. New York–London, Academic Press, 1980, v. 3, p. 1–97.
8. Гусев С. И. Асимптотические разложения, связанные с некоторыми статистическими оценками в гладком случае. I. — Теория вероятн. и ее примен., 1975, т. XX, № 3, с. 488–514.
9. Бурнашев М. В. Исследование свойств второго порядка статистических оценок в схеме независимых наблюдений. — Изв. АН СССР, сер. матем., 1981, т. 45, № 3, с. 509–539.
10. Michalsky D. L., Wise G. L., Poor H. V. A relative efficiency study of some popular detectors. — J. Franklin Inst., 1982, v. 313, p. 135–148.
11. Dadi M. I., Marks R. J. Detector relative efficiencies in the presence of Laplace noise. — IEEE Trans. Aerospace. Electron. Systems, 1987, v. 23, p. 568–582.
12. Burnashev M. V., Poor H. V. Asymptotic expansions and rate of convergence in signal detection. — IEEE Trans. Inform. Theory, 1995, v. 41, № 6.
13. Петров В. В. Суммы независимых случайных величин. М.: Наука, 1972, с. 416.
14. Takeuchi K. Tōkei-teki suitei no Zenkinriron (Asymptotic Theory of Statistical Estimation). Tokyo, Kyoiku-Shuppan, 1974 (in Japanese).
15. Takeuchi K., Akahira M. On the second-order asymptotic efficiencies of estimators. — Lect. Notes Math., 1976, v. 550, p. 604–638.
16. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. М.: Наука, 1971.

Поступила в редакцию
11.VII.1995

Таблица. Точные значения величин e_n и аппроксимаций
для них для плотностей Лапласа и Коши

n	$e_n(L)$	$e_n^1(L)$	$e_n^2(L)$	$e_n^3(L)$	$e_n(C)$	$e_n^1(C)$	$e_n^{(3)}(C)$
2	1	.5	1.064	1.002	∞	1.234	2.427
3	.639	.5	1.064	1.002	∞	1.234	2.427
4	.420	.25	.449	.434	∞	.617	.915
5	.351	.25	.449	.434	1.22	.617	.915
6	.261	.166	.275	.268	.860	.411	.544
7	.236	.166	.275	.268	.612	.411	.544
8	.187	.125	.195	.192	.478	.308	.383
9	.175	.125	.195	.192	.409	.308	.383
10	.145	.100	.150	.148	.336	.247	.294
11	.138	.100	.150	.148	.307	.247	.294
12	.118	.083	.122	.120	.261	.206	.239
13	.114	.083	.122	.120	.246	.206	.239
14	.0993	.0714	.1019	.1006	.214	.176	.201
15	.0963	.0714	.1019	.1006	.205	.176	.201
16	.0855	.0625	.0874	.865	.181	.154	.173
17	.0834	.0625	.0874	.865	.176	.154	.173
18	.0749	.0556	.0765	.0757	.158	.137	.152
19	.0734	.0556	.0765	.0757	.154	.137	.152
20	.0666	.0500	.0678	.0672	.140	.123	.135
21	.0654	.0500	.0678	.0672	.137	.123	.135
22	.0599	.0455	.0609	.0604	.125	.112	.122
23	.0590	.0455	.0609	.0604	.123	.112	.122
24	.0544	.0417	.0552	.0548	.113	.103	.111
25	.0537	.0417	.0552	.0548	.112	.103	.111
26	.0498	.0385	.0505	.0501	.104	.095	.102
27	.0492	.0385	.0505	.0501	.103	.095	.102
28	.0459	.0357	.0465	.0462	.0957	.088	.0942
29	.0454	.0357	.0465	.0462	.0947	.088	.0942
30	.0425	.0333	.0430	.0428	.0887	.0823	.0876
31	.0421	.0333	.0430	.0428	.0879	.0823	.0876