РОССИЙСКАЯ АКАДЕМИЯ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ
АВТОМЕТРИЯ

№ 6 2000

УДК 519.816

С. Н. Моисеев
(Воронеж)

ОЦЕНКИ ПАРАМЕТРОВ СИММЕТРИЧНЫХ УСТОЙЧИВЫХ РАСПРЕДЕЛЕНИЙ МЕТОДОМ МОМЕНТОВ

Методом моментов получены простые оценки параметров симметричных устойчивых распределений. Выведены теоретические асимптотически точные с ростом объема выборки формулы для смещений и рассеяний оценок, границы применимости которых определены моделированием. Проведен сравнительный анализ оценок.

Постоянно растущие требования к точности описания вероятностными моделями реальных данных все чаще вынуждают обращаться к негауссовским моделям. Характерным примером в этом плане могут служить устойчивые распределения вероятностей, которые в последнее время стали использовать при анализе телекоммуникационных трафиков [1], эконометрических данных [2] и геофизических наблюдений [3], при описании случайных радиофизических сигналов [4] и т. д. Интерес к этим моделям связан прежде всего с тем, что они составляют полный класс предельных распределений для линейно нормированных сумм независимых одинаково распределенных случайных величин. Однако их более широкое применение сдерживается тем, что плотности вероятностей устойчивых распределений в общем случае не выражаются через элементарные функции. Это приводит, в частности, к практически неприемлемым на сегодняшний день вычислительным затратам при оценивании параметров распределений методом максимального правдоподобия. Поэтому весьма актуальным является поиск более простых оценок.

В данной работе методом моментов получены несколько простых оценок параметров симметричных устойчивых распределений, проведен их статистический анализ и сравнение между собой.

Оценки на основе логарифмических моментов. Характеристическая функция симметричной устойчивой случайной величины $Y(a,b,\alpha)$ имеет вид [4]:

$$\theta(u)=\exp(jua-b|u|^{\alpha}),\qquad a\in(-\infty,\infty),\ b>0,\ \alpha\in(0,2]. \tag{1}$$

Гауссовские случайные величины получаются из $Y(a,b,\alpha)$ как частный случай при $\alpha=2$, случайные величины Коши – при $\alpha=1$. Это единственные случайные величины из симметричных устойчивых, плотности вероятностей которых выражаются через элементарные функции. При $\alpha<2$ случайная величина $Y(a,b,\alpha)$ не имеет моментов выше α. Параметр a распределения (1) –

медиана: $a = \operatorname{med}Y(a, b, \alpha)$. Параметр b характеризует ширину плотности вероятностей симметричного устойчивого распределения:

$$E = (Y_{(0,75)} - Y_{(0,25)})/2 = \frac{1}{2}\operatorname{med}|Y(0, b, \alpha)| = \lambda_\alpha b^{1/\alpha},$$

где E – срединное (вероятное) отклонение; $Y_{(p)}$ – квантиль порядка p случайной величины $Y(a, b, \alpha)$; λ_α – корень уравнения $\frac{4}{\pi}\int\limits_0^\infty \exp[-(x/\lambda_\alpha)^\alpha]\frac{\sin x}{x}dx = 1.$ Характеристический показатель α определяет тяжесть хвостов распределения: $\mathrm{P}[Y(0, b, \alpha) > x] = O(x^{-\alpha})$, $x \to \infty$, $\alpha < 2$.

В дальнейшем будем полагать $a = 0$, так как выборочные значения всегда можно центрировать на выборочную медиану, которая является достаточно хорошей оценкой параметра сдвига a, лишь незначительно уступающей по эффективности оценке максимального правдоподобия [5]. При известном α (например, из физических соображений [3]) оценить параметр b можно по методу квантилей: $b^* = Y_p^\alpha$, где Y_p – выборочная квантиль порядка $p = \frac{1}{2} + \frac{1}{\pi}\int\limits_0^\infty \exp(-x^\alpha)\frac{\sin x}{x}dx$. Однако основную трудность представляет совместное оценивание параметров α и b распределения (1), которое наиболее актуально в практических приложениях.

Рассмотрим характеристическую функцию случайной величины $\xi = \ln|Y(0, b, \alpha)|$:

$$\theta_\xi(u) = \frac{2}{\pi}\int\limits_0^\infty \theta_d(u; \alpha, bx^\alpha)\frac{\sin x}{x}dx, \tag{2}$$

где

$$\theta_d(u; \alpha, \beta) = \beta^{ju/\alpha}\Gamma(1 - ju/\alpha) \tag{3}$$

– характеристическая функция двойного показательного распределения; $\Gamma(\cdot)$ – гамма-функция. Подставляя (3) в (2), получаем характеристическую функцию величины ξ в виде

$$\theta_\xi(u) = \frac{b^{ju/\alpha}\,\Gamma(1 - ju/\alpha)}{\Gamma(1 - ju)\cos(ju\pi/2)}. \tag{4}$$

Отсюда по известной формуле

$$\chi_k = j^{-k}\left[\frac{d^k \ln\theta_\xi(u)}{du^k}\right]_{u=0}$$

находим кумулянты χ_k случайной величины ξ:

$$\begin{gathered} \chi_1 = (C + \ln b)/\alpha - C, \\ \chi_k = \zeta(k)(k-1)![\alpha^{-k} + (-1)^k - 2^{-k} - (-2)^{-k}], \qquad k = 2, 3, \ldots, \end{gathered} \tag{5}$$

где $C = 0{,}5772\ldots$ – постоянная Эйлера; $\zeta(x)$ – дзета-функция Римана [6]. Заменяя в (5) теоретические кумулянты χ_k на выборочные χ_k^*, получаем целый набор оценок для параметров α и b:

$$\alpha^* = \{\chi_k^* / [\zeta(k)(k-1)!] + 2^{-k} + (-2)^{-k} - (-1)^k\}^{-1/k},$$
$$b^* = \exp\{\alpha^*(\chi_1^* + C) - C\}. \tag{6}$$

Аналогичная идея получения оценок устойчивых распределений в более частном виде использована в [4], где найдены дисперсии некоторых функций от оценок (6).

Наиболее простые из оценок (6) имеют вид:

$$\alpha_1^* = \pi / \sqrt{6\mu_2^* - \pi^2/2}, \qquad b_1^* = \exp\{\alpha_1^*(m^* + C) - C\}, \tag{7}$$

где m^* и μ_2^* – соответственно выборочные среднее и дисперсия случайной величины ξ. Раскладывая статистики (7) в ряд Тейлора по степеням m^* и μ_2^* вблизи теоретических значений среднего m и дисперсии μ_2 и ограничиваясь первыми членами разложения, получаем выражения для смещений

$$\beta(\alpha_1^*) = \mathrm{M}\alpha_1^* - \alpha = \frac{\alpha}{2n}\left\{1 + \frac{\alpha^2}{2} + \frac{27\alpha^4}{\pi^4}(\mu_4 - \mu_2^2)\right\} + O(n^{-2}),$$

$$\beta(b_1^*) = \mathrm{M}b_1^* - b =$$

$$= \frac{3\alpha^3 b}{\pi^2 n}\left\{\frac{\mu_2}{\alpha}\left(s + \frac{\pi^2}{6}\right) + \frac{3\alpha s}{2\pi^2}(s+3)(\mu_4 - \mu_2^2) - \mu_3(1+s)\right\} + O(n^{-2})$$

и рассеяний

$$V(\alpha_1^*) = \mathrm{M}(\alpha_1^* - \alpha)^2 = \frac{9\alpha^6}{\pi^4 n}(\mu_4 - \mu_2^2) + O(n^{-2}),$$
$$V(b_1^*) = \mathrm{M}(b_1^* - b)^2 = \frac{\alpha^2 b^2}{n}\{\mu_2 + 9\alpha^2 s^2(\mu_4 - \mu_2^2)/\pi^4 - 6\alpha s\mu_3/\pi^2\} + O(n^{-2}) \tag{8}$$

оценок (7). В (8) $s = C + \ln b$; М – оператор математического ожидания; μ_k – центральные моменты случайной величины ξ; n – объем независимой выборки значений ξ, по которой производится оценивание параметров α и b.

Оценки на основе тригонометрических моментов. По определению характеристическая функция симметричной устойчивой случайной величины $Y(0, b, \alpha)$ выражается через среднее значение косинуса

$$\theta(u) = \mathrm{M}\cos\{uY(0, b, \alpha)\} = \theta_u. \tag{9}$$

Выбирая различные u и заменяя теоретические тригонометрические моменты θ_u их выборочными аналогами θ_u^* из (9), получаем следующий набор ценок для параметров α и b:

$$\alpha^* = \ln^{-1}|u/v| \ln \frac{\ln \theta_u^*}{\ln \theta_v^*}, \tag{10}$$

$$b^* = -\frac{\ln \theta_x^*}{|x|^{\alpha^*}}. \tag{11}$$

Наиболее простой вид оценки (10), (11) имеют при $u = e = \exp(1), v = x = 1$:

$$\alpha_2^* = \ln \frac{\ln \theta_e^*}{\ln \theta_1^*}, \qquad b_2^* = -\ln \theta_1^*. \tag{12}$$

Аналогично оценкам (7) находим выражения для смещений

$$\beta(\alpha^*) = \frac{1}{4n \ln|u/v|} \left\{ \frac{(1+\ln\theta_v)(1+\theta_{2v}-2\theta_v^2)}{\theta_v^2 \ln^2 \theta_v} - \frac{(1+\ln\theta_u)(1+\theta_{2u}-2\theta_u^2)}{\theta_u^2 \ln^2 \theta_u} \right\} + O(n^{-2}),$$

$$\beta(b_2^*) = \frac{1}{4n\theta_1^2}(1+\theta_2-2\theta_1^2) + O(n^{-2})$$

и рассеяний

$$V(\alpha^*) = \frac{1}{2n \ln^2|u/v|} \left\{ \frac{1+\theta_{2v}-2\theta_v^2}{\theta_v^2 \ln^2 \theta_v} + \frac{1+\theta_{2u}-2\theta_u^2}{\theta_u^2 \ln^2 \theta_u} - \right.$$

$$\left. -2\frac{\theta_{u+v}+\theta_{u-v}-2\theta_u\theta_v}{\theta_u\theta_v \ln\theta_u \ln\theta_v} \right\} + O(n^{-2}), \tag{13}$$

$$V(b_2^*) = \frac{1}{2n\theta_1^2}(1+\theta_2-2\theta_1^2) + O(n^{-2})$$

оценок α^* (10) и b_2^* (12).

Сравнение оценок. Из анализа выражений (8), (13) следует, что точность полученных оценок существенно зависит от истинных значений параметров α и b. Чтобы определить области значений параметров, где одна оценка точнее другой, введем понятие выигрыша оценки параметра α (7) по сравнению с оценкой (12) $\chi_\alpha = V(\alpha_2^*)/V(\alpha_1^*)$ и выигрыша оценки параметра b (7) по сравнению с оценкой (12) $\chi_b = V(b_2^*)/V(b_1^*)$.

На рис. 1 показано характерное поведение выигрышей χ_α в зависимости от значений параметра α для различных значений параметра b: кривая 1 – $b = 1$, кривая 2 – $b = 0,3$, кривая 3 – $b = 0,05$. При больших $b > 1$ оценка α_1^* (7) имеет меньшее рассеяние по сравнению с оценкой α_2^* (12) при любых значениях α. Выигрыш χ_α в этом случае возрастает при $\alpha \to 0$ и $\alpha \to 2$. Однако при малых $b < 0,5$ оценка α_1^* проигрывает α_2^* по величине рассеяния при $\alpha \to 2$.

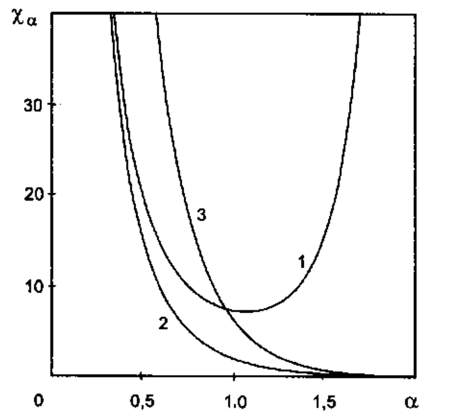

Рис. 1. Выигрыш оценки α (7) по сравнению с (12)

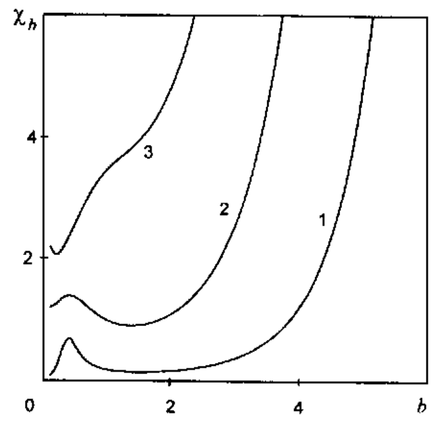

Рис. 2. Выигрыш оценки b (7) по сравнению с (12)

На рис. 2 изображены выигрыши χ_b в зависимости от значений параметра b для разных α: кривая 1 – $\alpha = 2$, кривая 2 – $\alpha = 1$, кривая 3 показывает предельное поведение χ_b при $\alpha \to 0$. Из рисунка видно, что оценка b_1^* (7) имеет меньшее рассеяние, чем b_2^* (12), для любых b при $\alpha \le 1$. Однако при $\alpha > 1$ оценка b_2^* оказывается предпочтительней для небольших значений b. Так, при $\alpha = 2$ оценка b_2^* имеет меньшее рассеяние, чем b_1^*, для значений $b < 4$.

Моделирование. С целью определения границ применимости полученных асимптотически точных при $n \to \infty$ выражений для смещений и рассеяний оценок проводилось моделирование алгоритмов оценивания (7) и (12) при разных наборах параметров α и b. Оказалось, что скорость сходимости модельных точек к теоретическим кривым (8), (13) существенно зависит от конкретных значений параметров α и b. Для иллюстрации этого на рис. 3 в логарифмическом по оси абсцисс масштабе изображено поведение модельных и теоретических рассеяний $V(\alpha_2^*)$ в зависимости от объема выборки n для двух разных значений α. Прямая 1 соответствует теоретическому значению $\lambda V(\alpha_2^*)$ при $\alpha = 2$, $\lambda = n$, $b = 0{,}1$, квадратики – значениям $\lambda V(\alpha_2^*)$, полу-

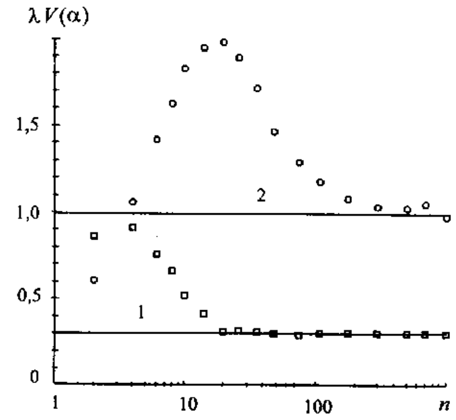

Рис. 3. Рассеяние оценки α

ченным моделированием. Прямой 2 показаны теоретические значения $\lambda V(\alpha_2^*)$ при $\alpha = 1$, $\lambda = n/10$, $b = 0{,}1$, кружками – соответствующие этому случаю модельные значения. Из рисунка хорошо видно, что если для кривой 1 удовлетворительное соответствие теоретических и модельных значений наблюдается уже при $n = 10$, то для кривой 2 аналогичное соответствие наблюдается только начиная со значений $n = 100$.

Полученные в работе выражения для оценок параметров симметричных устойчивых распределений имеют очень простой вид, однако для наиболее эффективного их использования желательно иметь априорную информацию об интервалах, содержащих истинные значения параметров.

СПИСОК ЛИТЕРАТУРЫ

1. **Resnick S. I.** Heavy tail modeling and teletraffic data. N. Y., 1995. (Prepr. /School of ORIE. Cornell University, Ithaca).
2. **Mandelbrot B. B.** The Pareto – Levy law and the distribution of income // Intern. Economic Rev. 1960. N 1. P. 79.
3. **Моисеев С. Н.** Механизм образования и вероятностное распределение максимальной элктронной концентрации слоя E_s // Геомагнетизм и аэрономия. 1997. **37**, № 3. С. 107.
4. **Золотарев В. М.** Одномерные устойчивые распределения. М.: Наука, 1983.
5. **Боровков А. А.** Математическая статистика. М.: Наука, 1984.
6. **Янке Е., Эмде Ф., Леш Ф.** Специальные функции. М.: Наука, 1977.

Воронежский государственный университет,
E-mail: mois@rf.main.vsu.ru

Поступила в редакцию
13 июля 1999 г.