ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ С ТЯЖЕЛЫМИ ХВОСТАМИ С ПОМОЩЬЮ ЭМПИРИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

Беврани Х., Аничкин К.

(Иран, Табриз; Россия, Москва)

Распределения с тяжелыми хвостами наблюдаются во многих естественных явлениях, гидрологии, геологии, электротехнике, информатике, физике и страховании.

Наиболее очевидной и естественной задачей статистического оценивания, связанной с тяжелыми хвостами, является задача оценивания параметра γ*. Данный параметр является показателем тяжести хвостов, показывает скорости убывания хвостов распределений.*

В данной статье мы рассмотрим оценивание параметра γ *по методу наименьших квадратов с помощью эмпирического распределения.*

1. Введение

Пусть $F(x)$ — функция распределения случайной величины X. Тогда $F(x)$, как говорят, является распределением с тяжелыми хвостами, если существуют положительные константы c и $0 < \gamma \le 2$ такие, что

$$\lim_{x\to\infty} x^{-\gamma}\bar{F}(x) = c\,, \tag{1}$$

где $\bar{F}(x) = 1 - F(x)$.

Распределения с тяжелыми хвостами наблюдаются во многих естественных явлениях, гидрологии, геологии, электротехнике, информатике, физике, страховании, описывают поведение финансовых рынков (см.[3–7]).

Часто при анализе статистических зависимостей мы пренебрегаем возможностью крупных событий, лежащих на "хвосте" распределения. Распределения с тяжелыми хвостами — распределения, хвост которых нельзя отрезать, т.е. нельзя пренебрегать крупными, но редкими событиями. Под крупными событиями понимаются явления, ущерб от которых может превосходить ущерб от всех остальных событий этого класса вместе взятых. Например, землетрясения, авиакатастрофы и аварии на атомные электростанции.

Далее мы рассмотрим распределения Парето и Стьюдента, как представителей класса распределений с тяжелыми хвостами.

1.1 Распределение Парето

Как известно, закон распределения Парето имеет вид:

$$p_X(x) = \frac{\alpha\beta^{\alpha}}{x^{\alpha+1}} \;; \quad x > \beta\,,\ \alpha > 0\,, \tag{2}$$

и функция распределения будет:

$$F(x) = 1 - \left(\frac{\beta}{x}\right)^{\alpha}. \tag{3}$$

В случае $\beta = 1$ это распределение считают распределением с одним параметром.

Математическое ожидание и дисперсия этого распределения будут:

$$E(X) = \frac{\alpha\beta}{\alpha-1}; \quad \alpha > 1, \quad \mathrm{Var(X)} = \frac{\alpha\beta^2}{\alpha-2} - \frac{\alpha\beta}{(\alpha-1)^2}; \quad \alpha > 2.$$

Как видно, математическое ожидание и дисперсия будут бесконечны, если $\alpha \le 1$ и $\alpha \le 2$. И особенно, в случае $\alpha \le 2$, когда дисперсия очень большая, распределение Парето является распределением с тяжелыми хвостами. Сравнение соотношения (1) и (3) показывает, что определение распределения с тяжелыми хвостами похоже на распределение Парето, поэтому часто первое определение называют распределением с тяжелыми хвостами вида Парето.

1.2 Распределение Стьюдента

Распределение Стьюдента, как известно, задается плотностью вида

$$p(x) = \frac{\Gamma((\gamma+1)/2)}{\sqrt{\pi\gamma}\,\Gamma(\gamma/2)}(1+\frac{x^2}{\gamma})^{-\frac{\gamma+1}{2}}, \qquad -\infty < x < \infty. \tag{4}$$

Если Y_γ — случайная величина с плотностью Стьюдента (4), то при $x \to \infty$,

$$x^\gamma P(|Y_\gamma| > x) \approx C_{(\gamma)}, \tag{5}$$

где $C_{(\gamma)} = \dfrac{2\Gamma((\gamma+1)/2)\gamma^{-\frac{\gamma}{2}-1}}{\sqrt{\pi}\,\Gamma(\gamma/2)}$.

Соотношение (5) соответствует определению распределений с тяжелыми хвостами.

Поэтому надо отметить, что распределение Стьюдента принадлежит к классу законов с тяжелыми хвостами, а параметр γ является показателем тяжести хвостов.

Наиболее очевидной и естественной задачей статистического оценивания, связанной с тяжелыми хвостами, является задача оценивания параметра γ. Для оценивания этого параметра могут также применяться известные процедуры оценивания степени тяжести хвостов, например, оценки Хилла [2], Де Хаана (см, например, [1]) и другие (см., например, [4]). В данной статье мы рассмотрим оценивание параметра γ по методу наименьших квадратов с помощью эмпирического распределения.

2. Оценивание параметра γ с помощью эмпирического распределения

Обозначим $X_i = |Y_i|$, $i = 1,...,n$. Пусть $X_{(1)},...,X_{(n)}$ — соответствующие порядковые статистики. Пусть $F_n(x)$ — эмпирическая функция распределения, соответствующая распределению. Тогда $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i < x)$, $-\infty < x < \infty$. В силу теоремы Гливенко при больших n справедливо соотношение

$$P(|Y_{\alpha,c}| > x) \approx 1 - F_n(x), \tag{6}$$

несложно видеть, что

$$F_n(X_i) = \frac{i-1}{n}, \quad i = 1, \ldots, n. \tag{7}$$

Сопоставляя (4), (6) и (7), мы замечаем, что для достаточно больших i, выполняется соотношение

$$cX_{(i)}^{\gamma} \approx \frac{n-i+1}{n}$$

или

$$\log(c) - \gamma \log(X_{(i)}) \approx \log(\frac{n-i+1}{n}) \tag{8}$$

или

$$\log(X_{(i)}) \approx \frac{\log(c)}{\gamma} - \frac{1}{\gamma} \log(\frac{n-i+1}{n}) \tag{9}$$

Естественно считать, что соотношения (8) и (9) выполняется для всех i, больших или равных некоторому k, то есть для $i \geq k$.

Воспользовавшись соотношениями (8) и (9) для $i = k, \ldots, n$, будем искать оценку $\hat{\gamma}_k$ параметра γ из условия

$$\hat{\gamma}_k = \arg\min_{\gamma} \sum_{i=k}^{n} \left[\log(c) - \gamma \log(X_{(i)}) - \log(\frac{n-i+1}{n}) \right]^2 \tag{10}$$

или

$$\hat{\gamma}_k = \arg\min_{\gamma} \sum_{i=k}^{n} \left[\log(X_{(i)}) - \frac{\log(c)}{\gamma} - \frac{1}{\gamma} \log(\frac{n-i+1}{n}) \right]^2. \tag{11}$$

Условия (10) и (11) представляют собой простейшую двухпараметрическую линейную модель наименьших квадратов. Решением задачи (10) является оценка

$$\hat{\gamma}_k = -\frac{(n-k+1)\sum_{i=k}^{n} \log X_{(i)} \log(\frac{n-i+1}{n}) - \sum_{i=k}^{n} \log X_{(i)} \sum_{i=k}^{n} \log(\frac{n-i+1}{n})}{(n-k+1)\sum_{i=k}^{n} (\log X_{(i)})^2 - (\sum_{i=k}^{n} \log X_{(i)})^2} \tag{12}$$

и для (11) оценка

$$\frac{\hat{1}}{\gamma_k} = \frac{(n-k+1)\sum_{i=k}^{n} \log X_{(i)} \log(\frac{n-i+1}{n}) - \sum_{i=k}^{n} \log X_{(i)} \sum_{i=k}^{n} \log(\frac{n-i+1}{n})}{(n-k+1)\sum_{i=k}^{n} (\log(\frac{n-i+1}{n}))^2 - (\sum_{i=k}^{n} \log(\frac{n-i+1}{n}))^2} \quad (13)$$

По второй оценке (13) можно найти несмещенную оценку для $\frac{1}{\gamma_k}$. В силу свойства порядковых статистик (см.,[9])

$$E(\log X_{(i)}) = \sum_{r=1}^{i} \frac{E(\log X_i)}{n-r+1} = E(\log X_1) \sum_{r=1}^{i} \frac{1}{n-r+1}, \quad (14)$$

но математическое ожидание $\log X_i$ будет:

$$E(\log X_i) = \frac{1+\log c}{\gamma}, \quad (15)$$

поэтому несмещенная оценка для $\frac{1}{\gamma_k}$ будет:

$$\frac{\hat{\hat{1}}}{\gamma_k} = \frac{\hat{1}}{\gamma_k} \frac{(n-k+1)\sum_{i=k}^{n}\left(\log(\frac{n-i+1}{n})\right)^2 - \left(\sum_{i=k}^{n}\log(\frac{n-i+1}{n})\right)^2}{(n-k+1)(1+\log\hat{c})\sum_{i=k}^{n}\sum_{r=1}^{i}\frac{1}{n-r+1}\left(\log(\frac{n-i+1}{n}) - \sum_{i=k}^{n}\log(\frac{n-i+1}{n})\right)} \quad (16)$$

Так как при умеренных значениях аргумента функция распределения Стьюдента отнюдь не является гиперболой, то включение в оценки (12), (13) и (16) порядковых статистик с малыми или умеренными номерами может сильно исказить окончательный результат. Однако, поскольку оба коэффициента линейной модели (10) и (11) зависят от неизвестного параметра γ, условие (10) и (11) является в некотором смысле избыточным, что позволяет использовать оценку второго коэффициента c, получаемую наряду с (12), (13) и (16) из условий (10) и (11). Для отыскания оптимального значения k, то есть числа порядковых статистик, используемых для нахождения оценки параметра γ из условий (8) и (9). А именно, оценка значения c, получаемая по методу наименьших квадратов из условий (10) и (11), имеет вид

$$\hat{c}_k = \exp\left\{\frac{1}{n-k+1}\left[\sum_{i=k}^{n}\log(\frac{n-i+1}{n}) + \hat{\gamma}_k \sum_{i=k}^{n}\log X_{(i)}\right]\right\} \qquad (17)$$

и

$$\hat{c}_k = \exp\left\{\frac{\hat{\gamma}_k}{n-k+1}\left[\sum_{i=k}^{n}\log X_{(i)} + \frac{\hat{1}}{\gamma_k}\sum_{i=k}^{n}\log(\frac{n-i+1}{n})\right]\right\}. \qquad (18)$$

3. Анализ полученных результатов

В приложении для наглядности приведены графики $\bar{\hat{\gamma}}$ и $\bar{\hat{\gamma}} \pm S_{\bar{\hat{\gamma}}}$, где $\bar{\hat{\gamma}}$ — среднее значение для оцениваемого параметра по выборке из *n* = 100 оценок, $S_{\bar{\hat{\gamma}}}$ — стандартное отклонение $\bar{\hat{\gamma}}$. Кривая посередине отражает поведение $\bar{\hat{\gamma}}$, верхняя кривая отражает поведение $\bar{\hat{\gamma}} + S_{\bar{\hat{\gamma}}}$, нижняя — $\bar{\hat{\gamma}} - S_{\bar{\hat{\gamma}}}$. По оси абсцисс отложено значение *k* —номер порядковой статистики, с которой начинается суммирование в формулах (12) и (16).

Было замечено, что для распределения Стьюдента при $\gamma \le 2$ получаются наиболее точные оценки. С ростом γ точность оценок резко снижается. В отличие от распределения Стьюдента оценки, полученные предложенным методом для распределения Парето, останутся точными и при $\gamma > 2$.

Формулы (12) и (13) дают почти одинаковые результаты. Однако (16) дают более точный результат. Нетрудно заметить, что отклонение быстро возрастает при увеличении k, особенно для $k \ge 90$. Данный результат справедлив для распределения Стьюдента и Парето.

4. Заключение

В данной статье мы оценили параметр распределения с тяжелыми хвостами с помощью эмпирического распределения. Результаты вычислений этих оценок сведены в графиках (См.

приложение). Мы получили наиболее точные оценки параметров распределений при $\gamma \leq 2$. Легко заметить, что можно использовать наш метод не только для оценивания параметра Стьюдента и Парето, но для параметров других распределений, принадлежащих классу распределений с тяжелыми хвостами.

Список литературы:

1. de Hann, L. On regular variation and its application to the weak convergence of sample extremes. Mathematisch centrum amesterdam, 1970.
2. Hill, B. A simple general approach to inference abaut the tail of a distribution // Ann . Statist. 1975. N 3. P. 1163–1174.
3. Mandelbort The fractal geometry of nature. W. H. Freeman, San Francisco, 1982.
4. Hosking J. and Wallis J. Parameter and quintile estimation for the generalized Pareto distribution // Technometrics. 1987. N 29. P. 339–349.
5. Brockwell P. and Davis R. Time series: Theory and methods, 2nd Ed. Springer, New york, 1991.
6. Janicki A. and Weron A. Can one see alpha stable variables and processes? // Statist. Sci. 1994. N 9. P. 109–126.
7. Nikias C. and Shao M. Signal Processing with alpha stable distributions and applications. Wiley, New York, 1995.
8. Adler R., Feldman R. A practical guide to heavy tails. Birkhauser, Boston, 1998.
9. Arnold C., Balakrishnan N., Nagaraja H. N. A first course in order statistics.
10. Bevrani H. Parameter estimation for a class of heavy-tailed distribution, Book of abstracts 12th Iranian researchers' conference in Europe, Manchester, United Kingdom, 2004. P. 376.

Приложение

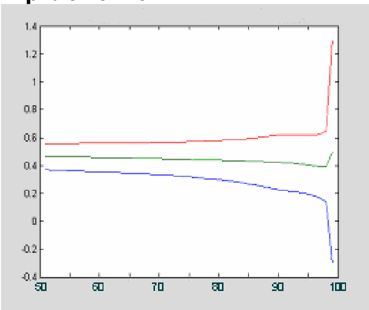

Рис. 1. Оценка параметра распределения Парето по формуле (12) при $\gamma = 0{,}5$

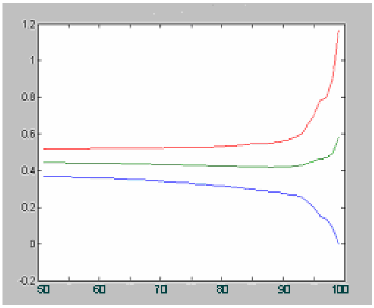

Рис. 2. Оценка параметра распределения Стьюдента по формуле (12) при $\gamma = 0{,}5$

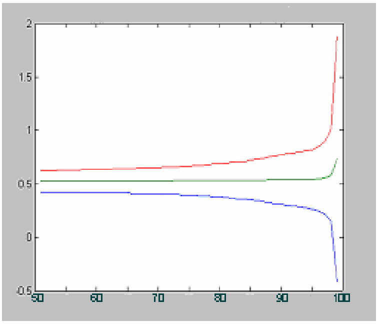

Рис. 3. Оценка параметра распределения Парето по формуле (16) при $\gamma = 0{,}5$

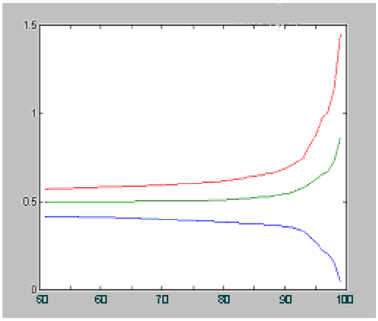

Рис. 4. Оценка параметра распределения Стьюдента по формуле (16) при $\gamma = 0{,}5$

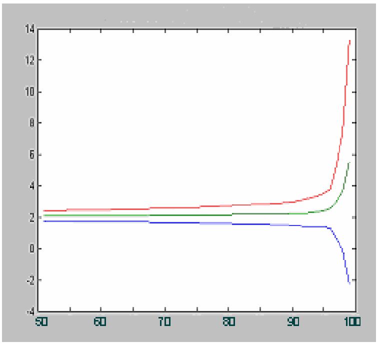

Рис. 5. Оценка параметра распределения Парето по формуле (16) при $\gamma = 2$

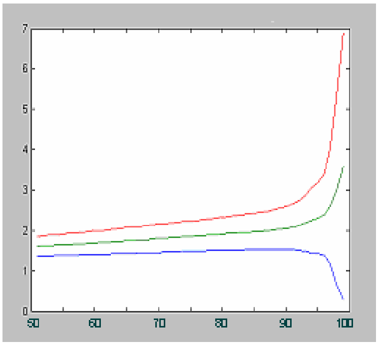

Рис. 6. Оценка параметра распределения Стьюдента по формуле (12) при $\gamma = 2$

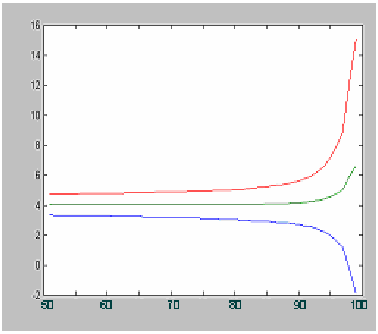

Рис. 7. Оценка параметра распределения Парето по формуле (16) при $\gamma = 4$

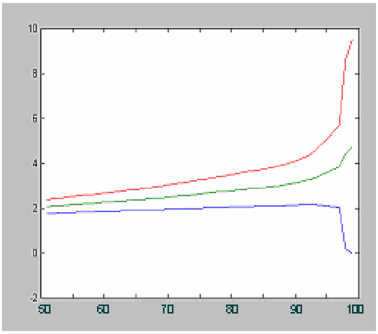

Рис. 8. Оценка параметра распределения Стьюдента по формуле (12) при $\gamma = 4$

PARAMETER ESTIMATION FOR THE HEAVY TAILED DISTRIBUTIONS WITH THE EMPIRICAL DISTRIBUTION

Bevrani H., Anichkin K.

(Iran, Tabriz; Russia, Moscow)

Heavy- tailed distributions have been observed in many natural phenomena including hydrology, geology, electro technology, informatics, physics and insurance. Heavy- tailed distributions are the distributions, the tail of which cannot be cut off. So, we cannot neglect the large-scale but rare events.

Therefore the purpose of the given article is the estimation of parameter for heavy tailed distributions using the empirical distribution.