УДК 519.22./.25

В.Ю. Дубницкий, А.И. Ходырев

Харьковский институт банковского дела Университета банковского дела НБУ, Харьков

ОЦЕНКА ПАРАМЕТРОВ РАНГОВЫХ РАСПРЕДЕЛЕНИЙ: РАСПРЕДЕЛЕНИЕ ПАРЕТО

Методом максимума правдоподобия получены оценки параметров распределения Парето и обобщенного распределения Парето. Показаны примеры применения этих распределений для решения задач актуарной математики.

Ключевые слова: *ранговые распределения, распределение Парето, обобщённое распределение Парето, метод максимума правдоподобия, оценка параметров распределений.*

Введение

Постановка проблемы. В последние два десятилетия в математической статистике происходит, в определённой мере, смена парадигмы. В связи с проблемами страховых выплат вследствие катастроф с большим количеством пострадавших возникла проблема прогнозирования как количества катастроф, так и количества пострадавших. Традиционная статистика, с её правилом «трёх сигм», здесь бессильна, ибо происходят именно маловероятные события. Это обстоятельство побудило искать новые статистические методы, ориентированные на возникшие задачи. К таким методам можно отнести теорию ранговых распределений.

Анализ последних исследований и публикаций. К классу ранговых распределений относят распределение Парето. Модели, приводящие к его возникновению, описаны в работах [1, 2, 3, 5]. В справочнике [4] приведены сведения о связи этого распределения с остальными и правила оценки его параметров. Распределение Парето допускает несколько способов записи.

Цель данного сообщения – получение оценок распределения Парето и обобщённого распределения Парето методом максимума правдоподобия. Вид функции распределения в настоящей работе принят таким, какой использован в задачах анализа последствий техногенных и естественных катастроф, а также финансовых кризисов.

Изложение основного материала

Распределение Парето в литературе записывают в различных вариантах.

В работе [1] дано следующее выражение для плотности распределения вероятности, подчиняющейся закону Парето:

$$f(x)=\frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{1+\alpha},\ x>x_0;\ \alpha>0\,. \qquad (1)$$

В работах [2, 3] функция распределения Парето приведена в виде:

$$F(x)=\begin{cases}1-\alpha x^{-\alpha};\ x\geq 1;\\ 0;\ x<1.\end{cases}; \qquad (2)$$

Соответственно плотность примет вид:

$$f(x)=\begin{cases}\alpha x^{-\alpha-1};\\ 0;\ x\leq 1.\end{cases} \qquad (3)$$

В виде, аналогичном работе [1], плотность этого распределения приведена в справочнике [4].

Принципиальное отличие этих подходов в том, что в одном случае для применения этого распределения требуется знание минимального значения x_0, в другом случае – максимального. Далее рассмотрим плотность распределения Парето в виде (1).

Определим 1, 2,… k-й начальные моменты этого распределения:

$$\int_{x_0}^{\infty} x\frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{1+\alpha}dx=\frac{x_0\alpha}{\alpha-1}, \qquad (4)$$

$$\int_{x_0}^{\infty} x^2\frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{1+\alpha}dx=\frac{x_0^2\alpha}{\alpha-2}, \qquad (5)$$

………………………………………………

$$\int_{x_0}^{\infty} x^k\frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{1+\alpha}dx=\frac{x_0^k\alpha}{\alpha-k}. \qquad (6)$$

Моменты k-го порядка для распределения Парето вида [1] с плотностью вида (1) будут иметь смысл при $\alpha>k$.

Используя метод максимума правдоподобия, определим параметр α для распределения Парето с плотностью вида (1) в предположении, что величина x_0 известна из физических соображений.

Функция правдоподобия для выборки из

$x_1, x_2, \ldots, x_i, \ldots, x_n$

наблюдений в этом случае примет вид:

$$L=\prod_{i=1}^{n}\frac{\alpha}{x_0}\left(\frac{x_0}{x_i}\right)^{1+\alpha}=\left(\frac{\alpha}{x_0}\right)^n\cdot\frac{x_0^{n(1+\alpha)}}{\prod_{i=1}^{n}x_i^{1+\alpha}}. \qquad (7)$$

Логарифм функции правдоподобия примет вид:

$$\ln L =$$
$$= n\left(\ln\alpha - \ln x_0\right) + n\left(1+\alpha\right)\ln x_0 - \left(1+\alpha\right)\sum_{i=1}^{n}\ln x_i . \qquad (8)$$

Из условия

$$\frac{d}{d\alpha}\ln L = \frac{n}{\alpha} + n\ln x_0 - \sum_{i=1}^{n}\ln x_i = 0 \qquad (9)$$

получим оценку $\hat{\alpha}$ параметра α в виде:

$$\hat{\alpha} = \left[\frac{\sum_{i=1}^{n}\ln x_i - n\ln x_0}{n}\right]^{-1} . \qquad (10)$$

Принимая во внимание, что логарифм среднего геометрического значения

$$\overline{x_\Gamma} = \sqrt[n]{\prod_{i=1}^{n} x_i} \qquad (11)$$

равен

$$\overline{\ln x_\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\ln x_i , \qquad (12)$$

окончательно получим, что:

$$\hat{\alpha} = \ln\left(\overline{x_\alpha}/x_0\right)^{-1} . \qquad (13)$$

Полученная таким образом оценка (13) совпадает с оценкой, приведенной в работе [4].

Рассмотрим функцию распределения Парето в виде:

$$F(x) = \begin{cases} 1 - x^{-\alpha} \;;\; x > 1; \\ \quad 0 \;;\; x < 1. \end{cases} \qquad (14)$$

Тогда ее плотность равна величине

$$f(x) = \alpha x^{\alpha - 1} . \qquad (15)$$

Начальный момент k-го порядка примет вид:

$$M\left[x^k\right] = \int_1^\infty x^k \cdot \alpha x^{\alpha-1} \alpha x = \lim_{x\to\infty}\frac{\alpha x^{-\alpha+k}}{-\alpha+k} - \frac{\alpha}{-\alpha+k} . \qquad (16)$$

Моменты порядка k для распределения Парето в виде (15) будут существовать только при $k < \alpha$.

Оценим параметр α для этого распределения методом максимума правдоподобия.

Функция правдоподобия примет вид:

$$L = \prod_{i=1}^{n}\alpha x_i^{-\alpha-1} , \qquad (17)$$

её логарифм равен:

$$\ln L = n\ln\alpha - \left(\alpha+1\right)\sum_{i=1}^{n}\ln x_i , \qquad (18)$$

откуда следует

$$\hat{\alpha} = \left[\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right]^{-1} = \left(\ln\overline{x}_\Gamma\right)^{-1} . \qquad (19)$$

Применение распределения вида (14) рассмотрим на примере важных для практики задач об определении общей суммы крупных страховых премий и прогнозирования размера самого крупного страхового вознаграждения.

Связь между этими задачами показана в работе [2] и приведено ее приближенное решение. Ниже приведен уточнённый вариант решения этой задачи.

Пусть случайная величина x_i – сумма i-й страховой премии. Тогда

$$S_n = x_1 + x_2 + ... + x_n , \forall x_i > x_0 . \qquad (20)$$

Пусть

$$m_n = \max\left\{x_1, x_2, ... x_n\right\} . \qquad (21)$$

С учетом условия (21) и, принимая приведенный в [2] результат, получим, что:

$$P\ \left\{m_n < x\right\} = F^n\left(x\right) = \left(1 - x^{-\alpha}\right)^n . \qquad (22)$$

Далее определим медиану последовательности (22). Для этого, исходя из определения медианы, примем, что:

$$F^n\left(x\right) = 0{,}5 . \qquad (23)$$

То есть:

$$\left(1 - x^{-\alpha}\right)^n = 0{,}5. \qquad (24)$$

Отсюда медиана

$$\text{med } m_n = \left(\frac{1}{1 - \alpha^{1/n}}\right)^{1/\alpha} . \qquad (25)$$

В работе [2] указано, что при больших значениях n величина

$$\text{med } m_n \approx \left(\frac{n}{\ln 2}\right)^{1/\alpha} . \qquad (26)$$

В свою очередь оценка среднего значения суммы вида (20) может быть получена по асимптотической формуле:

$$\left\langle S_n \right\rangle = \text{med}\ \ m_n \cdot R_n . \qquad (27)$$

Используя (25), получим, что

$$\left\langle S_n \right\rangle = \left(\frac{1}{1 - \alpha^{1/n}}\right)^{1/\alpha} R_n . \qquad (28)$$

В свою очередь, согласно [2] величина

$$R_n = 1 + \frac{n\beta\left(n; 1/n\right) - \alpha}{\alpha - 1} , \qquad (29)$$

где $\beta\left(n;\ 1/n\right)$ – бета-функция, то есть равенство вида:

$$\beta\left(n;\ \frac{1}{n}\right) = \frac{\Gamma(n)\cdot\Gamma\left(1/n\right)}{\Gamma\left(\frac{n^2+1}{n}\right)} . \qquad (30)$$

Для вычисления бета-функции можно использовать либо специальные программы, имеющиеся в некоторых математических пакетах, либо используя известное соотношение:

$$B(x,y)=\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \qquad (31)$$

то есть

$$\beta\left(n;\ \frac{1}{n}\right)=\frac{\Gamma(n)\cdot\Gamma\left(1/n\right)}{\Gamma\left(\frac{n^2+1}{n}\right)}. \qquad (32)$$

Из условия (26) следует, что среднее значение отношения

$$\delta=\frac{\langle S_n\rangle}{R_n}=\frac{\langle S_n\rangle}{\text{med}\ \ m_n}. \qquad (33)$$

Таблица значений $\rho(\alpha,n)$ приведена в [2], поверхность и линии уровня функции $\rho(\alpha,n)$ показаны на рис. 1 и 2.

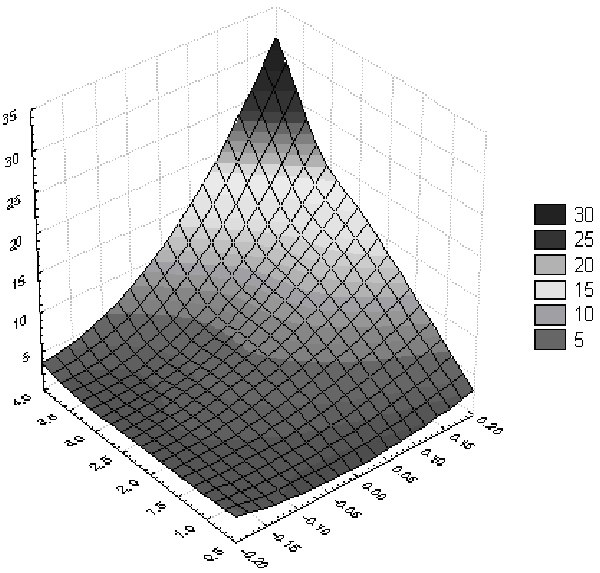

Рис. 1. Поверхность образованная функцией ρ(α, n) в координатах: α – уровень значимости; n – объём выборки

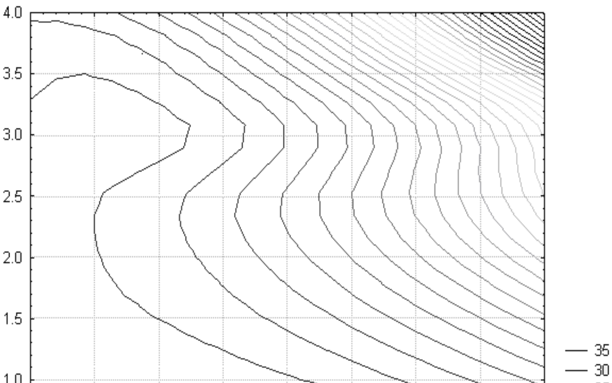

Рис. 2.Линии уровня функции ρ(α, n) в координатах: α – уровень значимости; n – объём выборки

Доверительный интервал для суммы страховых премий можно определить из условия:

$$P\left\{\underline{S_n}<S_n<\overline{S_n}\right\}=1-\varepsilon, \qquad (34)$$

где S_n – нижняя доверительная граница суммы;

S – верхняя граница.

В работе [2] приведены следующие выражения, реализующие условие (35):

$$\frac{R_n}{\left(1-\varepsilon^{1/n}\right)^{1/\alpha}}<S_n<\left(\frac{n}{\varepsilon}\right)^{1/\alpha}. \qquad (35)$$

Еще одно применение закона распределения Парето связано с моделированием оценки финансовых рисков. Постановка эта задача рассмотрена в работе [5]. Для ее решения предложено использовать обобщенное распределение Парето, имеющее вид:

$$F(x)=1-\left(1+\frac{\alpha x}{\beta}\right)^{-1/\alpha}. \qquad (36)$$

При условии, что $\beta>0;\ \ x\geq 0$, если $\alpha\geq 0$ и $0\leq x\leq-\dfrac{\beta}{\alpha}$, если $\alpha<0$.

Случай $\alpha=0$ в рамках данной работы не рассматривается.

Плотность распределения вида (36) примет вид:

$$f(x)=\left(\frac{\alpha x+\beta}{\beta}\right)^{-1/\alpha}\cdot(\alpha x+\beta)^{-1}. \qquad (37)$$

Функция правдоподобия для плотности вида (37) будет такой:

$$L=\prod_{i=1}^{n}\left(\frac{\alpha x_i+\beta}{\beta}\right)^{-1/\alpha}\cdot(\alpha x_i+\beta)^{-1}. \qquad (38)$$

Логарифм функции правдоподобия вида (38) будет следующий:

$$\ln L=-\frac{1}{\alpha}\sum_{i=1}^{n}\ln\left(\frac{\alpha x_i+\beta}{\beta}\right)^{-1/\alpha}-\sum_{i=1}^{n}\ln(\alpha x_i+\beta). \qquad (39)$$

Для определения оценок $\hat{\alpha}$ и $\hat{\beta}$ параметров α и β необходимо решить систему уравнений:

$$\begin{cases}\dfrac{\partial}{\partial\alpha}L=\dfrac{1}{\alpha^2}\times\\ \times\left(\displaystyle\sum_{i=1}^{n}\ln\left(\frac{\alpha x_i+\beta}{\beta}\right)-\frac{1}{\alpha}\sum_{i=1}^{n}\frac{x_i}{\alpha x_i+\beta}\right)-\displaystyle\sum_{i=1}^{n}\frac{x_i}{\alpha x_i+\beta}=0;\\ \dfrac{\partial L}{\partial\beta}=-\dfrac{1}{\alpha}\dfrac{\displaystyle\sum_{i=1}^{n}\left(\frac{1}{\beta}-\frac{\alpha x_{i+\beta}}{\beta^2}\right)}{\displaystyle\sum_{i=1}^{n}(\alpha x_i+\beta)}-\displaystyle\sum_{i=1}^{n}\frac{1}{\alpha x_i+\beta}.\end{cases} \qquad (40)$$

Для решения системы (40) необходимо определить начальные значения α_0 и β_0. С этой целью воспользуемся методом моментов:

$$M\left[x^2\right]=\int_0^\infty x\left(1+\frac{\alpha x}{\beta}\right)^{-1/\alpha}\cdot\left(\alpha x+\beta\right)^{-1}dx=\frac{\beta}{1-\alpha};\quad(41)$$

$$\alpha\neq 1;$$

$$M\left[x^2\right]=\int_0^\infty x^2\left(1+\frac{\alpha x}{\beta}\right)^{-1/\alpha}\cdot\left(\alpha x+\beta\right)^{-1}dx=$$
$$=\frac{2\beta^2}{(1-\alpha)(2\alpha-1)};\quad \alpha\neq 1.\quad(42)$$

Выражение (42) имеет смысл при $(1-\alpha)(2\alpha-1)\neq 0$. Следовательно, дисперсия σ_x^2 будет равна величине:

$$\sigma_x^2=M\left[x^2\right]-\left(M[x]\right)=\frac{\beta^2}{(\alpha-1)^2(1-2\alpha)}.\quad(43)$$

Заменив полученные характеристики $\left(M[x]\right)$ и σ_x^2 их выборочными оценками, получим систему вида:

$$\begin{cases} m_x=\dfrac{\beta}{1-\alpha}; \\ S=\dfrac{\beta^2}{(\alpha-1)^2(1-2\alpha)}. \end{cases}\quad(44)$$

Решив ее относительно α и β, получим:

$$\hat{\alpha}_0=\frac{S_x^2-m_x^2}{2S_x^2};\ \hat{\beta}_0=\frac{m_x\left(m_x^2+S_x^2\right)}{S_x^2};\quad(45)$$

при условии:

$$(\alpha-1)^2(2\alpha-1)\neq 0;\ \alpha\neq 1.\quad(46)$$

Полученные значения $\hat{\alpha}_0$ и $\hat{\beta}_0$ используем в качестве начальных приближений при решении системы (40).

Выводы

1. Получены методом максимального правдоподобия оценки параметров распределения Парето и обобщенного распределение Парето.
2. Показан пример применения распределения Парето для решения задачи об определении величины медианы максимального страхового вознаграждения.

Список литературы

1. *Кудрин Б.И. Унификация фенологических понятий и обозначений. / Б.И.Кудрин // Электрификация металлургических предприятий Сибири. – 2005. – Вып. 12. – С. 293-321.*
2. *Управление риском. Риск. Устойчивое развитие. Синергетика [Электронный ресурс] / Институт прикладной математики им. М.В. Келдыша Российской Академии наук. – Режим доступа к ресурсу: http://www.keldysh.ru/papers/2003/source/book/gmalin/titul.htm. – 04.01.2010. – Загл. с экрана.*
3. *Токарев Д.В. Оценка вероятности возникновения аварий на нефтеперерабатывающих, нефтехимических и химических предприятиях / Д.В.Токарев // [Электронный ресурс] / Электронный научный журнал "Нефтегазовое дело". – 2005. Режим доступа к журналу: http://www.ogbus.ru. – 04.01.2010 г. Загл. с экрана.*
4. *Вадзинский Р.Н. Справочник по вероятностным распределениям / Р.Н. Вадзинский. – СПб.: Наука, 2001. – 295 с.*
5. *Щетинин Е.Ю. Статистические методы и математические модели оценивания финансовых рисков / Е.Ю. Щетинин, С.А. Ланушкин // Математическое моделирование. – 2004. – Т. 16, №5. – С. 40-54.*

Поступила в редколлегию 25.01.2010

Рецензент: д-р техн. наук, проф. И.В. Кононенко, Национальный технический университет "ХПИ", Харьков.

ОЦІНКА ПАРАМЕТРІВ РАНГОВИХ РОЗПОДІЛІВ: РОЗПОДІЛ ПАРЕТО

В.Ю. Дубницький, О.І. Ходирєв

Методом максимуму правдоподібності отримані оцінки параметрів розподілу Парето і узагальненого розподілу Парето. Показані приклади застосування цих розподілів для вирішення завдань актуарної математики.

***Ключові слова**: рангові розподіли, розподіл Парето, узагальнений розподіл Парето, метод максимуму правдоподібності, оцінка параметрів розподілів.*

ESTIMATION OF RANK DISTRIBUTION PARAMETERS: PARETO DISTRIBUTION

V.Yu. Dubnitsky, A.I. Khodyrev

By soundness maximum method estimations were obtained for Pareto distribution and generalized Pareto distribution. Examples cited of application of those distributions for solution of actuarial mathematics problems.

***Keywords:** rank distributions, Pareto distribution, generalized Pareto distribution, soundness maximum method, distribution parameters estimation.*