Записки научных
семинаров ПОМИ
Том 486, 2019 г.

И. В. Волчёнкова, Л. Б. Клебанов

ХАРАКТЕРИЗАЦИИ РАСПРЕДЕНИЯ ПАРЕТО СВОЙСТВАМИ СОСЕДНИХ ПОРЯДКОВЫХ СТАТИСТИК

§1. Введение

Принято считать, что в случае распределения Парето максимальные наблюдения далеко отклоняются от меньших по величине. В предлагаемой заметке мы придадим точный смысл этому утверждению, а также покажем, что соответствующие свойства являются характеристическими для распределений Парето и сходных с ним. Некоторые характеристические свойства связаны с соседними порядковыми статистиками, точнее, с их отношениями, что позволяет получить результаты, не зависящие от выбора единицы масштаба.

§2. Об отношении соседних порядковых статистик

Пусть $X_1, \dots, X_n$ – повторная выборка из совокупности с функцией распределения (ф.р.) $F(x)$, причем $F(+0) = 0$. Обозначим $X_{1;n} \leqslant \dots \leqslant X_{n;n}$ соответствующие порядковые статистики. Первый интересующий нас вопрос: какова вероятность того, что r-тая порядковая статистика будет в фиксированное число раз меньшей $(r+1)$-ой порядковой статистики? Более точно, пусть $\kappa \in [0,1]$. Мы хотим изучить вероятность

$$H(\kappa) = \mathbf{P}\{X_{r;n} < \kappa X_{r+1;n}\} \tag{1}$$

как функцию κ. Ясно, что $H(0) = 0$ и $H(1) = 1$. При этом $H(\kappa)$ – монотонная функция κ. Таким образом, H представляет собой функцию распределения на промежутке $[0,1]$. Для того чтобы выразить

Ключевые слова: распределение Парето; вероятности больших значений; порядковые статистики; проблема моментов; характеризации распределений.

Работа была частично финансирована Грантовым Агентством Чешской Республики (Grant GAČR 19-04412S).

$H(k)$ через $F(x)$, воспользуемся хорошо известной формулой совместной плотности распределения статистик $X_{r;n}$ и $X_{s;n}$ $(1 \leqslant r < s \leqslant n)$:

$$p_{r,s;n}(x,y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} F^{r-1}(x)p(x)\cdot \\ \times \big(F(y) - F(x)\big)^{s-r-1} p(y)\big(1 - F(y)\big)^{n-s},\ x < y, \qquad (2)$$

которая может быть найдена, например, в [2]. Отсюда легко получаем, что $(s = r+1)$:

$$H(\kappa) = \mathbf{P}\{X_{r;n} < \kappa X_{r+1;n}\} \\ = \frac{n!}{r!(n-r-1)!} \int_0^\infty F(\kappa y)^r \big(1 - F(y)\big)^{n-r-1} dF(y). \qquad (3)$$

Легко также заметить, что последняя формула справедлива и без предположения об абсолютной непрерывности ф.р. $F(x)$. Если мы захотим подчеркнуть зависимость H также от n и r, то мы будем писать $H(\kappa; n, r)$ или $H(\kappa; n, r; F)$, если захотим указать и на зависимость от функции распределения F.

Первый интересующий нас вопрос касается асимптотического поведения вероятности (3) при $n \to \infty$ для фиксированных r и κ.

Теорема 1. *Предположим, что ф.р. $F(x)$ является правильно меняющейся в нуле функцией с показателем $a > 0$, а ее носитель представляет собой некоторый промежуток. Тогда*

$$\lim_{n\to\infty} H(\kappa; n, r) = \kappa^{ar} \qquad (4)$$

для любого $\kappa \in (0,1)$.

Доказательство. Имеем

$$H(\kappa; n, r) = \frac{n!}{r! \cdot (n-r-1)!} \int_0^\infty F(\kappa y)^r \big(1 - F(y)\big)^{n-r-1} dF(y)$$

$$= \frac{(m+r+1)!}{r! \cdot m!} \int_0^1 F^r(\kappa F^{-1}(u))(1-u)^m du$$

$$= \frac{(m+r+1)!}{r! \cdot m!} \cdot \frac{1}{m^{r+1}} \int_0^m \Big(\frac{F(\kappa F^{-1}(v/m))}{v/m}\Big)^r v^r (1 - v/m)^m dv.$$

Здесь обозначено $m = n-r-1$. Так как $F(x)$ – правильно меняющаяся в нуле функция с показателем $a > 0$, то

$$\lim_{m\to\infty} \frac{F(\kappa F^{-1}(v/m))}{v/m} = \lim_{m\to\infty} \frac{F(\kappa v/m)}{F(v/m)} = \kappa^a.$$

Ясно, что $(1-v/m)^m \to \exp(-v)$ при $m \to \infty$. Воспользовавшись этими соотношениями из выражения для вероятности $H(\kappa; n, r)$, находим

$$H(\kappa; n, r) \sim \frac{(m+r+1)!}{r! \cdot m!} \cdot \frac{1}{m^{r+1}} \cdot \kappa^{ar} \cdot \int_0^\infty v^r e^{-v} dv = \kappa^{ar}.$$

Здесь знак $\sim$ означает асимптотическую эквивалентность при $m \to \infty$. $\square$

Отметим, что результат теоремы 1 был получен при более жестких предположениях в работе [3]. Из доказанной теоремы вытекает, что с ростом r предельная вероятность $\lim_{n\to\infty} H(\kappa; n, r)$ убывает и становится сколь угодно малой при достаточно большом r. Внимательный анализ доказательства теоремы 1 показывает, что $H(\kappa; n, r)$ стремится к нулю при $r \to \infty$, лишь бы было $n > r$.

Перейдем теперь к изучению вопроса о том, насколько вероятность $H(\kappa; n, r)$, как функция n, определяет ф.р. $F(x)$.

Введем $\mathfrak{F}$ – класс ф.р. обратимых и таких, что для любой $F \in \mathfrak{F}$ существует неотрицательное число $a = a(F)$, удовлетворяющее условию: существует конечный ненулевой предел $\lim\limits_{u\to+0} F(u)/u^a$.

Теорема 2. *Пусть $F \in \mathfrak{F}$, $\kappa \in (0,1)$, а r – фиксированное целое положительное число. Тогда последовательность $\{H(\kappa; n, r; F),\ n = r+1, r+2 \ldots\}$ однозначно определяет ф.р. F в классе $\mathfrak{F}$ с точностью до масштаба.*

Доказательство. Допустим, что

$$H(\kappa; n, r; F) = H(\kappa; n, r; G), \quad n = r+1, r+2, \ldots \tag{5}$$

для некоторой $G \in \mathfrak{F}$. Это значит, что

$$\int_0^\infty F(\kappa y)^r \big(1 - F(y)\big)^{n-r-1} dF(y) = \int_0^\infty G(\kappa y)^r \big(1 - G(y)\big)^{n-r-1} dG(y).$$

Замена переменной $u = 1 - F(x)$ в первом интеграле и $u = 1 - G(x)$ во втором приводит нас к равенству

$$\int_0^1 F^r(\kappa F^{-1}(1-u))u^m du = \int_0^1 G^r(\kappa G^{-1}(1-u))u^m du;\ m = 0, 1, 2, \dots \quad (6)$$

Таким образом, мы имеем степенную проблему моментов для промежутка $(0, 1)$, которая, как известно (см., например, [1]), является определенной. Поэтому из (6) вытекает

$$F(\kappa F^{-1}(v)) = G(\kappa G^{-1}(v)) \quad (7)$$

при всех $v \in [0, 1]$. Уравнение (7) можно переписать в виде (замена $v = G(w)$)

$$F(\kappa F^{-1}(G(w))) = G(\kappa w), \quad w \geqslant 0,$$

или

$$\mathcal{K}(\kappa \mathcal{K}^{-1}(w)) = \kappa w,$$

где $\mathcal{K}(t) = G^{-1}(F(t))$, $t \geqslant 0$. Предыдущее уравнение может быть переписано как

$$\mathcal{K}(\kappa t) = \kappa \mathcal{K}(t), \quad t \geqslant 0. \quad (8)$$

Введем в рассмотрение функцию $\mathcal{L}$ положив $\mathcal{L}(t) = \mathcal{K}(t)/t$ для $t > 0$. Из (8) найдем

$$\mathcal{L}(\kappa t) = \mathcal{L}(t). \quad (9)$$

Последнее означает, что $\mathcal{L}$ – периодическая функция (с периодом $\log \kappa$) от логарифма. Таким образом, $\mathcal{L}(t)$ ограничена на положительной полуоси и, если не является константой, то не имеет предела при $t \to +0$. Однако, $\mathcal{K}(t) = G^{-1}(F(t))$ и, значит,

$$F(t) = G(t\mathcal{L}(t)), \quad t > 0. \quad (10)$$

Вспоминая определение класса $\mathfrak{F}$, запишем (10) в виде

$$\frac{F(t)}{t^{a(F)}} = \frac{G(t\mathcal{L}(t))}{(t\mathcal{L}(t))^{a(G)}} \cdot \frac{(t\mathcal{L}(t))^{a(G)}}{t^{a(F)}}. \quad (11)$$

При $t \to +0$ величина $t\mathcal{L}(t) \to 0$ в силу ограниченности $\mathcal{L}(t)$. Из определения класса $\mathfrak{F}$ следует, что отношения $F(t)/t^{a(F)}$ и $G(t\mathcal{L}(t))/(t\mathcal{L}(t))^{a(G)}$ имеют конечные ненулевые пределы при $t \to 0$. Из (11) теперь вытекает, что $(t\mathcal{L}(t))^{a(G)}/t^{a(F)}$ также имеет конечный ненулевой предел при $t \to 0$. Ясно, что если $a(F) \neq a(G)$, то последний предел либо равен нулю, либо не существует. Значит, $a(F) = a(G)$. А тогда должен существовать конечный предел $\mathcal{L}(t)$ при $t \to +0$, что возможно только для случая постоянной функции $\mathcal{L}(t)$. Требуемое следует теперь из равенства (10). $\square$

§3. Распределение отношения соседних порядковых статистик

Доказанная выше теорема 1 показывает, что предельное при $n \to \infty$ распределение соседних порядковых статистик представляет собой степенное распределение на интервале $(0, 1)$ (мы будем также называть его распределением Парето на $(0, 1)$). Естественным образом возникает вопрос о том, для каких ф.р. $F(x)$ распределение рассматриваемого отношения совпадает с распределением Парето на $(0, 1)$. Таким образом, речь идет о решении уравнения

$$\frac{n!}{r! \cdot (n-r-1)!} \int\limits_0^\infty F(\kappa y)^r \big(1 - F(y)\big)^{n-r-1} dF(y) = \kappa^{ar} \tag{12}$$

относительно ф.р. F при тех или иных предположениях. Одно семейство таких решений очень просто угадать. Именно, это само распределение Парето на $(0, 1)$. В самом деле, положим

$$F(x) = \begin{cases} 0 & \text{при} \quad x < 0, \\ x^a & \text{при} \quad 0 \leqslant x < 1, \\ 1 & \text{при} \quad x \geqslant 1 \end{cases}$$

и подставим эту функцию в левую часть (12). Получим

$$\frac{n!}{r! \cdot (n-r-1)!} \int\limits_0^1 \kappa^{ar} y^{ar} (1 - y^a)^{n-r-1} a y^{a-1} dy = \kappa^{ar},$$

что совпадает с (12).

Если мы рассматриваем (12) как уравнение, которое должно быть выполнено при всех $n = r+1, r+2, \ldots$, то мы можем применить теорему 2, предпологая, что искомая функция $F \in \mathfrak{F}$, чтобы убедиться, что другие решения отличаются от распределения Парето на $(0, 1)$ только параметром масштаба. Итак, нами получен следующий результат.

Теорема 3. *Пусть ф.р.* $F \in \mathfrak{F}$ *удовлетворяет уранению* (12) *при всех целых* $n \geqslant r+1$. *Тогда*

$$F(x) = \begin{cases} 0 & \text{при} \quad x < 0, \\ (Ax)^a & \text{при} \quad 0 \leqslant x < 1/A, \\ 1 & \text{при} \quad x \geqslant 1/A \end{cases}$$

при некотором $A > 0$.

Отметим, что для распределения Парето на $(0,1)$ правая часть (12) как функция κ совпадает с распределением статистики $X_{r;r}$. В связи с этим естественно возникает вопрос об условии одинаковой распределенности статистик $X_{r;n}/X_{r+1;n}$ и $X_{r;r}$. Следующая теорема как раз и дает такие условия.

Теорема 4. *Пусть ф.р.* $F(x)$ *принадлежит классу* $\mathfrak{F}$. *Статистики* $X_{r;n}/X_{r+1;n}$ *и* $X_{r;r}$ *одинаково распределены при всех* $n = r+1, r+2, \ldots$ *тогда и только тогда, когда* $F(x)$ *является ф.р. Парето на* $(0,1)$.

Доказательство. Свойство одинаковой распределенности статистик $X_{r;n}/X_{r+1;n}$ и $X_{r;r}$ может быть записано в виде

$$\frac{n!}{r! \cdot (n-r-1)!} \int_0^\infty F^r(\kappa y)\,(1 - F(y))^{n-r-1}\, dF(y) = F^r(\kappa). \tag{13}$$

Уравнение (13) можно переписать как

$$\int_0^1 \big(F^r(\kappa F^{-1}(u)) - F^r(\kappa) u^r\big)(1-u)^m du = 0, \quad m = 0, 1, \ldots.$$

В силу единственности решения проблемы моментов для промежутка $[0,1]$, мы заключаем, что

$$F^r(\kappa F^{-1}(u)) - F^r(\kappa) u^r = 0 \tag{14}$$

при всех $u \in [0,1]$. Уравнение (14) может быть решено тем же методом, что и уравнение (7). Мы опускаем детали. □

§4. Распределение Парето на интервале $[1, \infty)$

Приведенные выше результаты относились, в основном, к распределению Парето на интервале $(0,1)$. Оно, однако, тесно связано с классическим распределением Парето, которое сосредоточено на полуоси

$[1, \infty)$. В самом деле, если случайная величина X имеет распределение Парето на $(0, 1)$, т.е.

$$\mathbf{P}\{X < x\} = x^a, \quad x \in [0, 1],$$

то случайная величина $Y = 1/X$ имеет распределение Парето на полуоси $[1, \infty)$. Именно,

$$\mathbf{P}\{Y < x\} = 1 - 1/x^a, \quad x \in [1, \infty).$$

Эта связь позволяет или просто перенести предыдущие результаты на случай классического распределения Парето, либо легко найти их естественные аналоги. Ниже мы приведем некоторые факты, относящиеся к классическому распределению Парето.

Пусть $Y_1, \dots, Y_n$ – повторная выборка из совокупности с ф.р. $G(x)$, $G(+0) = 0$, а $Y_{1;n} \leqslant Y_{2;n} \leqslant \dots \leqslant Y_{n;n}$ – соответствующие порядковые статистики. Определим $X_j = 1/Y_j$, $j = 1, \dots, n$. Можно рассматривать величины X_j, $j = 1, \dots, n$, как повторную выборку из популяции с ф.р. $F(x) = 1 - G(1/x)$. Тогда порядковой статистике $Y_{r;n}$ соответствует $X_{n-r+1;n} = 1/Y_{r;n}$ или $X_{r;n} = 1/Y_{n-r+1;n}$. С учетом сказанного, мы можем переформулировать полученные ранее теоремы.

Положим

$$\mathcal{H}(\lambda) = \mathbf{P}\{Y_{n-r+1;n} > \lambda \cdot Y_{n-r;n}\}. \tag{15}$$

Теорема 5. *Предположим, что ф.р. $G(x)$ является правильно меняющейся на бесконечности функцией с показателем $-a, a > 0$, а ее носитель представляет собой некоторый промежуток вида $[A, \infty)$. Тогда*

$$\lim_{n\to\infty} \mathcal{H}(\lambda; n, r) = 1 - 1/\lambda^{ar} \tag{16}$$

для любого $\lambda > 1$.

Эта теорема является полным аналогом теоремы 1 и следует из нее.

Не представляет труда переформулировать остальные результаты, приведенные выше, для случая классического распределения Парето. Мы предоставляем это читателю.

Список литературы

1. Н. И. Ахиезер, *Классическая проблема моментов*, Наука, М., 1961.
2. H. A. David, H. N. Nagaraja, *Order Statistics*, John Wiley & Sons, INC., 2003.
3. L. B. Klebanov, J. Antoch, A. Karlova, A. V. Kakosyan, *Outliers and related problems*, arXiv:1701.06642v1, 2017.

Volchenkova I. V., Klebanov L. B. Characterizations of Pareto distribution by the properties of neighboring order statistics.

It is generally accepted that in the case of the Pareto distribution, the maximum observations far deviate from the smaller ones. In the proposed article, we give the exact meaning to this statement, and also show that the corresponding properties are characteristic for Pareto distributions and similar ones. Some characteristic properties are associated with neighboring ordinal statistics, more precisely, with their relationships, which allows us to obtain results that are independent of the choice of a unit of scale.

Чешский технический университет в Праге

Поступило 16 сентября 2019 г.

Кафедра теории вероятностей
и математической статистики,
Карлов университет,
Прага, Чешская республика
E-mail: levbkl@gmail.com