ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 25 МАТЕМАТИКИ Выпуск 3
2018

М. С. Т и х о в, В. Э. К а в а л е с а (Нижний Новгород, ННГУ). **Применение вероятностно-взвешенного метода оценивания параметров распределения экстремальных значений.**

Предсказание экстремальных значений, которые могут иметь место при последующих наблюдениях, является одной из важных задач математики чрезвычайных ситуаций. В докладе рассматривается вероятностно-взвешенный метод моментов (см. [1–3]) для оценивания параметров распределений экстремальных значений. Полученные оценки применены для оценивания практической ситуации экстремальных уровней снегопада в штате Северная Каролина (выборка взята из [4]).

Рассматривается обобщенное распределение экстремальных значений (GEV)

$$F(x)=F(x;\sigma,\gamma,\mu)=\begin{cases}\exp\left\{-\left(1+\gamma\dfrac{x-\mu}{\sigma}\right)^{-1/\gamma}\right\}, & \text{если } x>\mu-\dfrac{\sigma}{\gamma},\quad 0<\gamma<1,\\ \exp\left\{-\exp\left\{-\dfrac{x-\mu}{\sigma}\right\}\right\}, & \text{если } -\infty<x<+\infty,\quad \gamma=0,\end{cases}$$

где $-\infty<\mu<+\infty$.

Вероятностно-взвешенный метод (PWM) был введен в работах J. Greenwood, J. Landwehr, N. Matalas, J. Wallis [1–3] и состоит в следующем.

Обозначим $M_{q,r,s}$ математическое ожидание величины $X^q\cdot F^r(X)(1-F(X))^s$ и вычислим его для распределения GEV, когда $\gamma>0$, неизвестно, $q=1$, $s=0$ и $r=0,1,3$. Имеем:

$$\beta_r=M_{1,r,0}=\mathbf{E}(X\cdot F^r(X))=\frac{1}{r+1}\left(\mu-\frac{\sigma}{\gamma}+\frac{\sigma}{\gamma}(1+r)^{\gamma}(1-\gamma)\right),\qquad 0<\gamma<1.$$

Оценим β_r с помощью L-статистик вида

$$\widehat{\beta}_r=\frac{1}{n}\sum_{j=1}^{n}\left(\frac{j}{n}\right)^r X_n^{(j)},$$

где $X_n^{(1)}<X_n^{(2)}<\dots<X_n^{(n)}$ есть порядковые статистики, построенные по выборке $X_1,\dots,X_n$. Известно, что оценка $\widehat{\beta}_r$ асимптотически нормальна с математическим ожиданием $\alpha=\int xF^r(x)\,dF(x)$ и дисперсией

$$\sigma^2=2\iint_{s<t}F^{r+1}(s)F^r(t)\big(1-F(t)\big)\,ds\,dt.$$

Рассматривая систему уравнений $\left\{\beta_r=\widehat{\beta}_r,\ r=0,1,3,\right.$ найдем, что в качестве оценок параметров γ, σ, μ можно взять статистики

$$\widehat{\gamma}=\log_2\left(\frac{(4\widehat{\beta}_3-\widehat{\beta}_0)+\sqrt{(4\widehat{\beta}_3-\widehat{\beta}_0)^2+8(2\widehat{\beta}_1-\widehat{\beta}_0)(\widehat{\beta}_1-2\widehat{\beta}_3)}}{2(2\widehat{\beta}_1-\widehat{\beta}_0)}\right),$$

$$\widehat{\sigma}=\frac{(2\widehat{\beta}_1-\widehat{\beta}_0)\,\widehat{\gamma}}{(2^{\widehat{\gamma}}-1)(1-\widehat{\gamma})},\qquad \widehat{\mu}=\widehat{\beta}_0\widehat{\gamma}+\widehat{\sigma}\,(1-(1-\widehat{\gamma})).$$

В качестве альтернативных оценок для β_r можно использовать также статистики

$$\widetilde{\beta}_r = \frac{1}{n}\sum_{j=1}^{n}\left(\prod_{l=1}^{r}\frac{j-l+r}{n+l}\right) X_n^{(j)}, \quad \text{а также} \quad \beta_r^* = \frac{1}{n}\sum_{j=1}^{n}\left(\prod_{l=1}^{r}\frac{j-l}{n-l}\right) X_n^{(j)},$$

но вычисления показали, что их значения мало отличаются от значений $\widehat{\beta}_r$.

Применим PWM к следующему **примеру**, взятому из [4]: 25 января 2000 г. в аэропорту Роли-Дарем, Северная Каролина был зарегистрирован очень большой снегопад, высота снежного покрова достигла 20.3 дюймов. Это исключительно большой снегопад для этой части Соединенных Штатов, и он парализовал деятельность не только аэропорта. Является ли это событие довольно редким, например, оно происходит раз в 100-200 лет, или это не такое уж редкое событие и к нему надо готовиться?

Набор данных взят из U.S. National Statistics (приведен в [4]): 1.0, 2.5, 1.2, 1.2, 4.1, 9.0, 3.0, 1.0, 1.4, 2.0, 3.0, 1.7, 1.2, 1.2, 1.1, 1.5, 5.0, 1.6, 2.0, 0.1, 0.4, 0.8, 3.7, 1.3, 3.8, 0.1, 0.1, 0.2, 2.0, 7.6, 0.1, 1.8, 0.5, 0.5, 0.5, 1.1, 1.4, 1.0, 1.0, 0.7, 5.7, 0.4, 0.3, 1.8, 0.4, 1.0, 1.2, 2.6, 1.0, 5.0, 1.7, 2.4, 0.1, 0.5, 7.1, 0.2, 0.7, 0.1, 2.7, 2.9, 0.4, 2.0, — это ненулевые снегопады, которые были зарегистрированы в период 1948-1998 в аэропорту Роли-Дарем. По приведенным данным $\widehat{\gamma} = 0.33437$, $\widehat{\sigma} = 0.89525$, $\widehat{\mu} = 0.29397$. Таким образом, функция распределения равна $F_0(x) = \exp\{-1/(0.66+0.37x)^3\}$. Имеем: $1 - F_0(20.3) = 0.0018$, а $1 - F_0(9.0) = 0.0155$.

Можно рассмотреть и более общую модель $G(x) = \exp\{-\lambda(1+\xi(x-a)/\rho)^{-1/\xi}\}$,

$$G(x) = e^{-\lambda} + \sum_{k=1}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}\left(1 - \left(1 + \xi\,\frac{x-a}{\rho}\right)^{-1/\xi}\right)^k,$$

т. е. имеем смесь обобщенных распределений Парето, причем это есть модель GEV при $\rho = \sigma + \xi\,(a-\mu)$, $\lambda = \left(1 + \xi\,(a-\mu)/\rho\right)^{-1/\xi}$. Кроме того, подбор плотности распределения $f_1(x)$ с помощью кривых Пирсона ([5], с. 209) по тем же данным дает плотность $f_1(x) = 0.538\,(1 - x/64)^{30}(1 + x/38)^{-2}$ и функцию распределения $F_1(x) = \int_0^x f_1(t)\,dt$, $0 < x < 64$. Заметим также, что расстояние Колмогорова между эмпирической ф. р. $F_n(x)$ и ф. р. $F_0(x)$ равно 0.084, для $F_1(x)$ это расстояние равно 0.093. Без значений $x = 0.5$ и $x = 1.0$ в первом случае и $x = 1.0$ и $x = 2.0$ — во втором, эти значения сопоставимы и равны 0.059 и 0.057. Значения статистики χ^2 для интервалов с точками разбиения $1.1, 2.2, 3.1, 4.1, 5.1, 6.0, 8.0$ равны соответственно 7.56 (критическое значение равно $\chi_{0.95}(4) = 9.488$) и 5.52.

Предложенные модели рассматривались в предположении, что выборка является повторной. По-видимому, такое предположение не является универсальным, поэтому рассмотрим один случай неоднородной выборки. Именно, пусть $X_1, X_2, \ldots, X_{l(n)}, X_{l(n)+1}, \ldots, X_n$ — последовательность независимых случайных величин, причем $X_1, X_2, \ldots, X_{l(n)}$ распределены с функцией распределения $F(x)$, а $X_{l(n)+1}, \ldots, X_n$ — с функцией распределения $F_1(x)$. Пусть: 1) для всякого $\varepsilon > 0$ имеет место строгие неравенства $F(\varepsilon) < 1$, $F_1(\varepsilon) < 1$; 2) для всякого $\tau > 0$

$$\lim_{x\to\infty}\frac{1-F(x)}{1-F(\tau x)} = \tau^{\alpha}, \quad \lim_{x\to\infty}\frac{1-F_1(x)}{1-F_1(\tau x)} = \tau^{\alpha+\beta}, \qquad \alpha > 0, \quad \beta > 0.$$

Тогда $\mathbf{P}\{a_n X_n^{(n)} < x\} \to \exp\{-x^{-\alpha} - x^{-(\alpha+\beta)}\}$ при $n \to \infty, x > 0,$ если взять $a_n = n^{-1/(\alpha+\beta)}$, $l(n) = [n^{\alpha/(\alpha+\beta)}]$. При больших x с точностью до параметров сдвиг-масштаб имеем смесь распределений Парето.

СПИСОК ЛИТЕРАТУРЫ

1. *Greenwood J. A.*, *Landwehr J. M.*, *Wallis J. R.*, *Matalas N. C.* Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form. — Water Resources Research, 1979, v. 15, № 5, p. 1049–1054.

2. *Landwehr J. M.*, *Wallis J. R.*, *Matalas N. C.* Probability Weighted Moments Compared With Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles. — Water Resources Research, 1979, v. 15, № 5, p. 1055–1064.
3. *Hosking J. R..*, *Wallis J. R.*, *Wood E. F.* Estimation of the generalized Extreme-Value Distribution by the method of probability weighted moments. — Technometrics, 1985, v. 27, № 3, p. 251–261.
4. *Smith R. L.* Statistics of Extremes, with Applications in Environment, Insurance, and Finance. In: Extreme Values in Finance, Telecommunications and the Environment Ed. by B. Finkenstädt, H. Rootzén. London etc: Chapman and Hall/CRC, 2004, p. 1–78.
5. *Кендалл М.*, *Стьюарт А.* Теория распределений, М.: Наука, 1966, 588 с.