ROCZNIK NAUKOWO-DYDAKTYCZNY WSP W KRAKOWIE

Zeszyt 191 Prace z Rachunku Prawdopodobieństwa i jego Dydaktyki II 1998

Яков С. Бродский, Ольга К. Лысенко

О мощности статистических критериев для отбраковки резко выделяющихся элементов выборки

В работе [1] получены статистические критерии для исключения экстремальных значений выборок из экспоненциального распределения. В [2] приведены таблицы критических значений некоторых из этих критериев. В настоящей статье получены выражения для функции мощности рассмотренных ранее критериев.

Пусть $\xi_1, \ldots, \xi_n$ — независимые наблюдения за случайной величиной, имеющей показательное распределение, $\xi_{(1)} \leq \cdots \leq \xi_{(n)}$ — вариационный ряд, построенный по этим наблюдениям. Основная гипотеза H_0 утверждает, что все элементы выборки имеют одно и то же показательное распределение с параметром λ. В [1] рассмотрены статистики

$$Z_1 = \frac{\xi_{(n)} - \xi_{(n-1)}}{\xi_{(n)} - \xi_{(1)}}, \quad Z_2 = \frac{\xi_{(n)} - \xi_{(n-1)}}{\xi_{(n)} - \xi_{(2)}}, \quad Z_3 = \frac{\xi_{(n)} - \xi_{(n-2)}}{\xi_{(n)} - \xi_{(1)}},$$

с помощью которых построены статистические критерии для отбраковки резко выделяющегося максимального элемента выборки. Там же установлены распределения этих статистик. Аналогичные критерии получены для отбрасывания резко выделяющегося минимального элемента выборки:

$$Z_4 = \frac{\xi_{(2)} - \xi_{(1)}}{\xi_{(n)} - \xi_{(1)}}, \quad Z_5 = \frac{\xi_{(2)} - \xi_{(1)}}{\xi_{(n-1)} - \xi_{(1)}}, \quad Z_6 = \frac{\xi_{(3)} - \xi_{(1)}}{\xi_{(n)} - \xi_{(1)}}.$$

Критические области имеют вид $Z_k \geq t(\alpha, n)$, $k = 1, \ldots, 6$, где $t(\alpha, n)$ находится из условия $P(Z_k \geq t(\alpha, n) \mid H_0) = \alpha$, α — уровень значимости критерия.

Эта работа частично поддержана Международной Соросовской программой поддержки образования в области точных наук (JSSEP) Международного фонда „Возрождение", грант Nr APU061010.

Вычислим мощности критерия Z_k, $k = 1, \dots, 6$, т.е. вероятность попадания значения статистики Z_k критерия в критическую область $Z_k \ge t(\alpha, n)$, если верна альтернативная гипотеза. В качестве альтернативной гипотезы H_1 рассмотрим следующую: одно из наблюдений $\xi_1, \dots, \xi_n$ имеет показательное распределение с параметром λ_1, $\lambda_1 \ne \lambda$.

Для решения поставленной задачи нам понадобится совместные распределения трех порядковых статистик для n независимых наблюдений, когда одно из них $\tilde{\xi}$ имеет в качестве функции распределения $F_1(x)$, а остальнные $n-1$ — $F(x)$ $(F_1(x) \ne F(x))$. Предполагаем, что рассматриваемые случайные величины имеют плотности вероятностей. Формулы плотностей таких совместных распределений выводятся аналогично случаю одинаково распределенных независимых случайных величин (см. [3]).

Подсчитаем предварительно для $1 \le r < s < t \le n$, $x \le y \le z$,

$$P(x \le \xi_{(r)} < x + \Delta x, y \le \xi_{(s)}, y + \Delta y, z \le \xi_{(t)} < z + \Delta z).$$

Рассмотрим события:

$$\begin{aligned}
A_1 &= \{\tilde{\xi} < x\}, \\
A_2 &= \{x \le \tilde{\xi} < x + \Delta x\}, \\
A_3 &= \{x + \Delta x \le \tilde{\xi} < y\}, \\
A_4 &= \{y \le \tilde{\xi} < y + \Delta y\}, \\
A_5 &= \{y + \Delta y \le \tilde{\xi} < z\}, \\
A_6 &= \{z \le \tilde{\xi} < z + \Delta z\}, \\
A_7 &= \{\tilde{\xi} \ge z + \Delta z\}.
\end{aligned}$$

Они образуют полную группу несовместных событий. Поэтому

$$\begin{aligned}
&P(x \le \xi_{(r)} < x + \Delta x, y \le \xi_{(s)} < y + \Delta y, z \le \xi_{(t)} < z + \Delta z) \\
&\quad = \sum_{k=1}^{7} P((x \le \xi_{(r)} < x + \Delta x, y \le \xi_{(s)} < y + \Delta y, \\
&\qquad\qquad z \le \xi_{(t)} < z + \Delta z) \cap A_k).
\end{aligned} \tag{1}$$

Событие, стоящее под знаком вероятности в правой части последнего равенства при $k = 1$, реализуется (с точностью до членов, имеющих более высокий порядок малости для вероятности) в виде конфигурации:

$1 + (r-2)$	1	$s - r - 1$	1	$t - s - 1$	1	$n - t$
x		$x + \Delta x$ y		$y + \Delta y$ z		$z + \Delta z$

Это означает, что $r-2$ из одинаково распределенных наблюдений меньше x и $\xi < x$; одно из одинаково распределенных наблюдений попадает в промежуток $[x, x+\Delta x)$, $s-r-1$ из них — в промежуток $[x+\Delta x, y)$, одно — в $[y, y+\Delta y)$, $t-s-1$ — в $[y+\Delta y, z)$, одно — в $[z, z+\Delta z)$, остальные $n-t$ больше или равны $z+\Delta z$. Число способов, которыми $n-1$ наблюдений можно разбить на семь таких групп равно

$$\frac{(n-1)!}{(r-2)!\,1!\,(s-r-1)!\,1!\,(t-s-1)!\,1!\,(n-t)!}$$

и каждый из них имеет вероятность

$$\begin{aligned}&F^{r-2}(x)F_1(x)(F(x+\Delta x)-F(x))(F(y)-F(x+\Delta x))^{s-r-1}\\ &\times(F(y+\Delta y)-F(y))(F(z)-F(y+\Delta y))^{t-s-1}\\ &\times(F(z+\Delta z)-F(z))(1-F(z+\Delta z))^{n-t}.\end{aligned}$$

Поэтому при малых $\Delta x, \Delta y, \Delta z$ получим

$$\begin{aligned}&P((x \le \xi_{(r)} < x+\Delta x, y \le \xi_{(s)} < y+\Delta y, z \le \xi_{(t)} < z+\Delta z) \cap A_1)\\ &= \frac{(n-1)!}{(r-2)!\,(s-r-1)!\,(t-s-1)!\,(n-t)!}\\ &\quad\times F^{r-2}(x)F_1(x)(F(x+\Delta x)-F(x))\\ &\quad\times(F(y)-F(x+\Delta x))^{s-r-1}\\ &\quad\times(F(y+\Delta y)-F(y))(F(z)-F(y+\Delta y))^{t-s-1}\\ &\quad\times(F(z+\Delta z)-F(z))(1-F(z+\Delta z))^{n-t}+\Theta_\Delta,\end{aligned}$$

где Θ_Δ есть члены, содержащие множители вида $(\Delta x)^{n_1}(\Delta y)^{n_2}(\Delta z)^{n_3}$, прчем $n_j \ge 1$, j=1,2,3, а по крайней мере одно из n_j больше 1. Они включают в себя вероятности тех реализаций рассматриваемого события, где более, чем одно из ξ_j попадает хотя бы в один из промежутков $[x, x+\Delta x)$, $[y, y+\Delta y)$, $[z, z+\Delta z)$. Аналогично вычисляются остальные вероятности в правой части равенства (1). Подставляя их в равенство (1), деля обе части этого равенства на Δx, Δy, Δz и переходя к пределу при $\Delta x \to 0$; $\Delta y \to 0$; $\Delta z \to 0$, мы получим следующее выражение для совместной плотности $f_{rst}(x, y, z)$ порядковых статистик $\xi_{(r)}, \xi_{(s)}, \xi_{(t)}$ при $1 < r < s < t < n$, $s \ne r+1$, $t \ne s+1$, $x \le y \le z$:

$$
\begin{aligned}
&f_{rst}(x,y,z)\\
&= \frac{(n-1)!}{(r-1)!\,(s-r-1)!\,(t-s-1)!\,(n-t)!} F^{r-2}(x)(F(y)-F(x))^{s-r-2}\\
&\times(F(z)-F(y))^{t-s-2}(1-F(z))^{n-t-1}((r-1)F_1(x)f(x)(F(y)-F(x))\\
&\times f(y)(F(z)-F(y))f(z)(1-F(z))+F(x)f_1(x)(F(y)-F(x))f(y)\\
&\times(F(z)-F(y))f(z)(1-F(z))+(s-r-1)F(x)f(x)(F_1(y)-F_1(x))\\
&\times f(y)(F(z)-F(y))f(z)(1-F(z))+F(x)f(x)(F(y)-F(x))f_1(y)\\
&\times(F(z)-F(y))f(z)(1-F(z))+(t-s-1)F(x)f(x)(F(y)-F(x))f(y)\\
&\times(F_1(z)-F_1(y))f(z)(1-F(z))+F(x)f(x)(F(y)-F(x))f(y)\\
&\times(F(z)-F(y))f_1(z)(1-F(z))+(n-t)F(x)f(x)(F(y)-F(x))f(y)\\
&\times(F(z)-F(y)f(z)(1-F_1(z)),
\end{aligned}
$$

где $f(x)$ и $f_1(x)$ – плотности, соответствующие функциям распределения $F(x)$ и $F_1(x)$.

Аналогично вычисляются плотности $f_{rst}(x,y,z)$ при $r=1$, $t=n$, $s=r+1$, $t=s+1$.

Для вычисления функции мощности $P(Z_1 \geq t(\alpha,n)|H_1)$ нам понадобится совместная плотность $f_{1,n-1,n}(x,y,z)$ порядковых статистик $\xi_{(1)}$, $\xi_{(n-1)}$, $\xi_{(n)}$, построенных по n независимым наблюдениям, одно из которых имеет показательное распределение с параметром λ_1, а все остальные — показательное распределение с параметром λ. Она имеет вид:

$$
\begin{aligned}
&f_{1,n-1,n}(x,y,z)\\
&=(n-1)(n-2)\lambda_1 e^{-\lambda_1 x}(e^{-\lambda x}-e^{-\lambda y})^{n-3}\lambda e^{-\lambda y}\lambda e^{-\lambda z}\\
&\quad+(n-1)(n-2)(n-3)\lambda e^{-\lambda x}(e^{-\lambda x}-e^{-\lambda y})^{n-4}(e^{-\lambda_1 x}-e^{-\lambda_1 y})\lambda e^{-\lambda y}\lambda e^{-\lambda z}\\
&\quad+(n-1)(n-2)\lambda e^{-\lambda x}(e^{-\lambda x}-e^{-\lambda y})^{n-3}\lambda_1 e^{-\lambda_1 y}\lambda e^{-\lambda z}\\
&\quad+(n-1)(n-2)\lambda e^{-\lambda x}(e^{-\lambda x}-e^{-\lambda y})^{n-3}\lambda e^{-\lambda y}\lambda_1 e^{-\lambda_1 z}.
\end{aligned}
$$

Выполняя замену $t=\lambda x$, $u=\lambda(z-y)$, $v=\lambda(z-x)$ и учитывая, что модуль якобиана преобразования равен $\frac{1}{\lambda^3}$, получим совместную плотность случайных величин $\lambda\xi_{(1)}$, $\lambda(\xi_{(n)}-\xi_{n-1)})$, $\lambda(\xi_{(n)}-\xi_{(1)})$:

$$
\begin{aligned}
&\varphi(t,u,v)\\
&=(n-1)(n-2)\mu e^{-t(n-1+\mu)}e^{-(2v-u)}(1-e^{-(v-u)})^{n-3}\\
&\quad+(n-1)(n-2)(n-3)e^{-t(n-1+\mu)}e^{-(2v-u)}(1-e^{-(v-u)})^{n-4}(1-e^{-(v-u)\mu})\\
&\quad+(n-1)(n-2)\mu e^{-t(n-1+\mu)}e^{-v(\mu+1)}e^{\mu u}(1-e^{-(v-u)})^{n-3}\\
&\quad+(n-1)(n-2)\mu e^{-t(n-1+\mu)}e^{-v(\mu+1)}e^{u}(1-e^{-(v-u)})^{n-3},
\end{aligned}
$$

где $\mu = \frac{\lambda_1}{\lambda}$.

Зная $\varphi(t,u,v)$, можно вычислить совместную плотность $\psi(u,v)$ случайных величин $\lambda(\xi_{(n)} - \xi_{(n-1)})$, $\lambda(\xi_{(n)} - \xi_{(1)})$:

$$\psi(u,v) = \int_0^\infty \varphi(t,u,v)\,dt.$$

Так как $Z_1 = \frac{\xi_{(n)} - \xi_{(n-1)}}{\xi_{(n)} - \xi_{(1)}}$, то

$$P(Z_1 \geq x \mid H_1) = 1 - \int_0^\infty dv \int_0^{vx} \psi(u,v)\,du.$$

Окончательно функция мощности критерия Z_1 имеет вид:

$$1 - \beta_1(\mu) = \frac{(n-1)!\,(1-x)^{n-3}((n+\mu-2)(1-x)+1)}{n-1+\mu} \prod_{k=1}^{n-2} \frac{1}{1+k(1-x)}$$
$$-\frac{(n-1)!\,(1-x)^{n-3}}{n-1+\mu} \prod_{k=0}^{n-3} \frac{1}{(k+\mu)(1-x)+1}$$
$$+\frac{(n-1)!\,(1-x)^{n-2}}{n-1+\mu} \prod_{k=1}^{n-2} \frac{1}{\mu+k(1-x)}.$$

Здесь через x обозначено для краткости критическое значение $t(\alpha,n)$, $0< x <1$. В ходе преобразований использовалось тождество

$$\sum_{k=0}^{n} \frac{(-1)^k}{k+a} C_n^k = \frac{n!}{a(a+1)\cdot\ \cdots\ \cdot(a+n)}, \quad (a>0).$$

Аналогичным образом получены выражения для функций мощности критериев $Z_2, \ldots, Z_6$. Например, для Z_2 она имеет вид:

$$1 - \beta_2(\mu) = \frac{(\mu(1-x)(2n+\mu-3)+(n-1)(1+(n-3)(1-x)))}{(n-1+\mu)(n-2+\mu)}$$
$$\times(n-2)!\,(1-x)^{n-4} \prod_{k=1}^{n-3} \frac{1}{1+k(1-x)} + \frac{(n-1)!\,(1-x)^{n-3}}{(n-1+\mu)(n-2+\mu)}$$
$$\times \prod_{k=1}^{n-3} \frac{1}{\mu+k(1-x)}$$
$$+\frac{(n-1)!\,(1-x)^{n-4}}{(n-1+\mu)(n-2+\mu)} \prod_{k=0}^{n-4} \frac{1}{1+(\mu+k)(1-x)}, \quad 0<x<1.$$

Литература

[1] Бродский Я. С., Быцан Г. Н., Власенко В. М., *Об исключении экстремальных значений*, Заводская лаборатория, **7** (1975), 847-849.

[2] Барановский А. И, Бродский Я. С., *Некоторые статистические задачи, связанные с распределением Парето*, Университет Донецк, Деп УкрНИИНТИ, Nr 1574 Ук-Д84, 1984.

[3] Дейвид Г., *Порядковые статистики*, Наука, Москва 1979.

Кафедра теории вероятностей
Донецький державный університет
ул. Університетськая 24
340055 Донецьк
Ukraine
E-mail: oleg@kompas.donetsk.ua