ZASTOSOWANIA MATEMATYKI
V (1960)

E. FIDELIS i R. ZIELIŃSKI (Warszawa)

SZACOWANIE PARAMETRÓW ROZKŁADU ZA POMOCĄ STATYSTYK Z PRÓBKI UPORZĄDKOWANEJ

1. Zakładamy, że zmienna losowa Z w populacji generalnej, określona w przedziale (a, b) jest typu ciągłego i ma dwuparametrowy rozkład o dystrybuancie $F(z)$.

W niniejszym opracowaniu podamy pewne estymatory parametrów w populacji, opierając się na próbce uporządkowanej według następującego schematu:

1° Losujemy próbkę o liczności n i porządkujemy ją według rosnących wartości zmiennej losowej.

2° Znajdujemy medianę.

3° Dzielimy próbkę na dwie grupy zaliczając do I elementy o wartościach zmiennej losowej mniejszych od mediany, do II — elementy o wartościach większych.

4° Z grupy I wyłączamy $k-1$ najmniejszych elementów, z grupy II wyłączamy $k-1$ największych elementów.

5° Obliczamy średnią pozostałych elementów grupy I i oznaczamy ją przez x, oraz średnią pozostałych elementów grupy II i oznaczamy ją przez y (gdy $k = 1$, wartości x i y są średnimi wszystkich elementów odpowiednio grupy I i II).

Estymatorami parametrów w populacji generalnej przy tak określonym postępowaniu będą pewne funkcje x i y.

2. Uzasadnienie teoretyczne. Z populacji generalnej o rozkładzie określonym dystrybuantą $F(z)$ losujemy próbkę o liczności n i porządkujemy ją według rosnących wartości zmiennej losowej. Oznaczamy przez t_i $(i = 1, 2, \ldots, n)$ wartości zmiennej losowej w uporządkowanej próbce. Dzielimy próbkę na dwie równe grupy w następujący sposób:

1° Jeżeli liczność próbki n jest nieparzysta, do grupy I zaliczamy wartości t_i dla $i = 1, 2, \ldots, \frac{1}{2}(n-1)$, do grupy II wartości t_i dla $i = \frac{1}{2}(n+3), \ldots, n$. Mediana nie zostaje zaliczona do żadnej z grup.

2° Jeśli liczność próbki n jest parzysta, do grupy I zaliczamy wartości t_i dla $i = 1, 2, \ldots, \frac{1}{2}n$, do grupy II wartości t_i dla $i = \frac{1}{2}(n+2), \ldots, n$.

W obydwu przypadkach z grupy I wyłączamy $k-1$ najmniejszych elementów. Z pozostałych elementów grupy I obliczamy średnią, którą oznaczamy przez x.

Podobnie z grupy II wyłączamy $k-1$ największych elementów i z pozostałych obliczamy średnią, którą oznaczamy przez y.

Omówimy tylko przypadek 1°, to znaczy gdy n jest liczbą nieparzystą. W przypadku 2° sposób postępowania jest analogiczny.

Oznaczmy przez τ_k i τ_l k-tą i l-tą statystykę pozycyjną w próbce. Łączny rozkład tych statystyk jest dany znanym wzorem

$$(1)\qquad g(t_k, t_l) = \frac{n!}{(k-1)!(l-k)!(n-l-1)!}[F(t_k)]^{k-1}[F(t_l)-F(t_k)]^{l-k}\times$$
$$\times[1-F(t_l)]^{n-l-1}f(t_k)f(t_l).$$

Wprowadzimy dwie nowe zmienne losowe: ξ przyjmującą wartości x i η przyjmującą wartości y.

Łączna funkcja gęstości tych zmiennych jest

$$(2)\qquad h(x, y) = \frac{1}{\left(\frac{1}{2}(n+1)-k\right)^2}\sum_{i=k}^{\frac{1}{2}(n-1)}\sum_{j=\frac{1}{2}(n+3)}^{n-k+1} g(x_i, y_j).$$

Wartości oczekiwane zmiennych ξ i η, obliczone na podstawie rozkładu (1), są

$$(3)\qquad \begin{aligned} E(\xi) &= \frac{1}{\frac{1}{2}(n+1)-k}\sum_{i=k}^{\frac{1}{2}(n-1)} E(\tau_i), \\ E(\eta) &= \frac{1}{\frac{1}{2}(n+1)-k}\sum_{j=\frac{1}{2}(n+3)}^{n-k+1} E(\tau_j), \end{aligned}$$

co oznacza, że wartości oczekiwane zmiennych ξ i η są średnimi odpowiednich wartości oczekiwanych statystyk pozycyjnych. W szczególnym przypadku, gdy $k = 1$, $E(\xi)$ jest średnią wartości oczekiwanych statystyk pozycyjnych mniejszych od mediany, $E(\eta)$ — średnią wartości oczekiwanych statystyk pozycyjnych większych od mediany.

Rozpatrzmy następujące zmienne losowe:

$$(4)\qquad s = \frac{x+y}{2}, \qquad r = y-x.$$

Dokonując we wzorze (2) podstawienia zgodnie z (4) i całkując (2) względem x otrzymujemy funkcje gęstości zmiennych S i R w postaci;

$$\varphi(s) = 2a_n \sum_{ij} \lambda_{ij} \int_a^b [F(x)]^{i-1} [1-F(2s-x)]^{n-j-1} [F(2s-x)- \tag{5}$$

$$-F(x)]^{j-i} f(2s-x) f(x)\,dx$$

$$\text{dla} \quad x < s < \frac{b+x}{2},$$

$$\psi(r) = a_n \sum_{ij} \lambda_{ij} \int_a^b [F(x)]^{i-1} [1-F(R+x)]^{n-j-1} [F(R+x)- \tag{6}$$

$$-F(x)]^{j-i} f(R+x) - f(x)\,dx$$

$$\text{dla} \quad 0 < r < b-x,$$

gdzie

$$a_n = \frac{1}{\frac{1}{2}(n+1)-k},$$

$$\lambda_{ij} = \frac{n!}{(i-1)!(j-i)!(n-j-1)!};$$

$i = k, \ldots, \frac{1}{2}(n-1)$, $j = \frac{1}{2}(n+3), \ldots, n-k+1$.

Wartości oczekiwane zmiennych S i R obliczone na podstawie (5) i (6) są równe

$$E(S) = \frac{1}{2\left(\frac{1}{2}(n+1)-k\right)} \left[\sum_{i=k}^{\frac{1}{2}(n-1)} E(\tau_i) + \sum_{i=\frac{1}{2}(n+3)}^{n-k+1} E(\tau_i) \right], \tag{7}$$

$$E(R) = \frac{1}{\frac{1}{2}(n+1)-k} \left[\sum_{i=\frac{1}{2}(n+3)}^{n-k+1} E(\tau_i) - \sum_{i=k}^{\frac{1}{2}(n-1)} E(\tau_i) \right]. \tag{8}$$

W ten sposób wartości oczekiwane zmiennych S i R wyrażają się przez wartości oczekiwane statystyk pozycyjnych.

3. Estymacja parametrów rozkładu równomiernego. Załóżmy, że rozkład w populacji jest równomierny na odcinku (a, b):

$$F(x) = \begin{cases} 0 & \text{dla} \quad x \leqslant a, \\ \dfrac{x-a}{b-a} & \text{dla} \quad a < x < b, \\ 1 & \text{dla} \quad x \geqslant b. \end{cases} \tag{9}$$

Wartość oczekiwana k-tej statystyki pozycyjnej z próbki o liczności n dla tego rozkładu jest

$$E(\tau_k) = a+(b-a)\frac{k}{n+1}. \tag{10}$$

Wartości $E(S)$ i $E(R)$ są odpowiednio równe

$$\begin{aligned} E(S) &= a+\frac{b-a}{2}, \\ E(R) &= \frac{b-a}{2(n+1)}(n-2k+3). \end{aligned} \tag{11}$$

W ten sposób, przyjmując $\frac{1}{2}(x+y)$ i $y-x$ z próbki za oszacowanie wartości oczekiwanych zmiennych S i R, otrzymujemy estymatory parametrów a i b w populacji

$$a = c_1x+c_2y, \qquad b = c_2x+c_1y, \tag{12}$$

gdzie

$$c_1 = \frac{1}{2}\left(1+\frac{2(n+1)}{n-2k+3}\right); \qquad c_2 = \frac{1}{2}\left(1-\frac{2(n+1)}{n-2k+3}\right).$$

4. Estymacja parametrów rozkładu normalnego. Załóżmy, że rozkład w populacji jest $N(m, \sigma)$. Dokonując zamiany zmiennych

$$S' = \frac{2s-2m}{\sigma}, \qquad R' = \frac{R}{\sigma}, \tag{13}$$

na podstawie (7) i (8) otrzymujemy

$$\begin{aligned} E(S') &= \frac{1}{2\left(\frac{1}{2}(n+1)-k\right)}\left[\sum_{i=k}^{\frac{1}{2}(n-1)} E_i+\sum_{i=\frac{1}{2}(n+3)}^{n-k+1} E_i\right], \\ E(R') &= \frac{1}{\frac{1}{2}(n+1)-k}\left[\sum_{i=\frac{1}{2}(n+3)}^{n-k+1} E_i-\sum_{i=k}^{\frac{1}{2}(n-1)} E_i\right], \end{aligned} \tag{14}$$

gdzie E_i jest wartością oczekiwaną i-tej statystyki pozycyjnej z próbki o liczności n pochodzącej z populacji normalnej $N(0,1)$.

Ze względu na symetrię statystyk pozycyjnych mamy

$$\begin{aligned} E(S') &= 0, \\ E(R') &= \frac{2}{\frac{1}{2}(n+1)-k}\sum_{i=\frac{1}{2}(n+3)}^{n-k+1} E_i. \end{aligned} \tag{15}$$

Oznaczając $E(R')$ przez r_{nk} (jest to stała zależna tylko od n i k) otrzymujemy estymatory dla m i σ:

$$\hat{m} = \frac{x+y}{2}, \quad \hat{\sigma} = \frac{y-x}{r_{nk}}.$$

Wartość współczynnika r_{nk} dla niektórych n podano w tablicy 1.

TABLICA 1. Wartości r_{nk}

k \ n	3	5	7	9
1	1,6923688	1,6579835	1,6415064	1,6319037
2		0,9900380	1,1100813	1,1858628
3			0,7034140	0,8464967
4				0,5490518

5. W celu empirycznego sprawdzenia otrzymanych wyników wylosowano 50 próbek $n = 9$ z populacji $N(0,1)$ i dla $k = 3$ otrzymano

$$\hat{m} = 0{,}02, \quad \hat{\sigma} = 1{,}06.$$

INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK

Praca wpłynęła 18. 12. 1958

Е. ФИДЕЛИС и Р. ЗЕЛИНСКИ (Варшава)

ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ ПРИ ПОМОЩИ СТАТИСТИК ИЗ УПОРЯДОЧЕННОЙ ВЫБОРКИ

РЕЗЮМЕ

В статье предлагаются некоторые оценки параметров непрерывного двухпараметрового распределения, опирающиеся на упорядоченной выборке.

В качестве оценки принимаются некоторые функции средних, элементов меньших и больших медианы.

Теоретические рассмотрения иллюстрируются примером исчислений оценок для параметров прямоугольного и нормального распределений.

E. FIDELIS and R. ZIELIŃSKI (Warszawa)

ESTIMATION OF DISTRIBUTION PARAMETERS BY MEANS OF STATISTICS FROM AN ORDERED SAMPLE

SUMMARY

The authors propose certain estimators of parameters of a continuous two-parameter distribution based on an ordered sample.

They adopt as parameters certain functions of the means taken from the elements larger and smaller than the median.

Theoretical considerations are illustrated by an example where the estimators for the parameters of the uniform and the normal distribution are numerically computed.