Доклады Академии наук СССР
1975. Том 220, № 6

УДК 519.251.8 *МАТЕМАТИКА*

Ю. В. БОРОВСКИХ

О КРИТЕРИИ СОГЛАСИЯ КОЛМОГОРОВА ДЛЯ ВЫБОРОК ИЗ КОНЕЧНЫХ СОВОКУПНОСТЕЙ

(Представлено академиком А. Н. Колмогоровым 25 VII 1974)

Пусть $\mathfrak{X}$ – конечная совокупность действительных чисел X_i, $i=1, 2, \ldots, N$, таких, что $X_1<X_2<\ldots<X_N$. Свяжем с $\mathfrak{X}$ функцию распределения $F_N(x)=l/N$, где l – число $X_i\leqslant x$.

Путем случайного выбора без возвращения из $\mathfrak{X}$ последовательно извлекаем n чисел и располагаем их в порядке возрастания: $x_1<x_2<\ldots<x_n$. Стало быть, $x_1=X_{l_1}<x_2=X_{l_2}<\ldots<x_n=X_{l_n}$, при этом $1\leqslant l_1<l_2<\ldots<l_n\leqslant N$, $k\leqslant l_k\leqslant N-n+k$, $k=1, 2, \ldots, n$. По результатам выборки определим эмпирическую функцию распределения $F_n(x)=k/n$, где k – число $x_k\leqslant x$.

Далее, введем в рассмотрение случайные величины

$$\Delta_{n,N}^{+}=\sup_{-\infty<x<\infty}(F_n(x)-F_N(x)),\qquad \Delta_{n,N}^{-}=\sup_{-\infty<x<\infty}(F_N(x)-F_n(x)),\tag{1}$$

$$\Delta_{n,N}=\sup_{-\infty<x<\infty}|F_n(x)-F_N(x)|.\tag{2}$$

Статистика $\Delta_{n,N}$ недавно была предложена Ю. К. Беляевым [1] в качестве естественного аналога непараметрического критерия А. Н. Колмогорова [2–5] для выборок из конечных совокупностей. Относительно $\Delta_{n,N}$ Ю. К. Беляев и Л. В. Рыкова в совместной работе [6] показали, что при $N\to\infty$, $n\to\infty$, $\overline{\lim}\,(n/N)=\nu_0$, $0\leqslant\nu_0<1$ имеет место соотношение

$$\lim P\left(\sqrt{\frac{nN}{N-n}}\sup_{-\infty<x<\infty}|F_n(x)-F_N(x)|<\lambda\right)=\sum_{j=-\infty}^{\infty}(-1)^j e^{-2j^2\lambda^2}.\tag{3}$$

Положим

$$D^+(n_1, n_2)=\sup_{-\infty<x<\infty}(T_{n_1}(x)-G_{n_2}(x)),\ D^-(n_1, n_2)=\sup_{-\infty<x<\infty}(G_{n_2}(x)-T_{n_1}(x)),$$
$$D(n_1, n_2)=\sup_{-\infty<x<\infty}|T_{n_1}(x)-G_{n_2}(x)|,\tag{4}$$

где $T_{n_1}(x)$ и $G_{n_2}(x)$ – эмпирические функции распределения, построенные по совокупностям взаимно независимых случайных величин $(\xi_1, \xi_2, \ldots, \xi_{n_1})$ и $(\eta_1, \eta_2, \ldots, \eta_{n_2})$ соответственно с одинаковой непрерывной функцией распределения. Следовательно, (4) – классические двухвыборочные статистики Н. В. Смирнова [7, 8]. Аналоги статистик (4) для выборок из конечных совокупностей были предметом рассмотрений Л. В. Рыковой [9].

В нашей работе проводится исследование группы задач, связанных со статистиками (1) и (2). Сюда в первую очередь относятся теоремы о совместном распределении случайных величин $\Delta_{n,N}^{-}$ и $\Delta_{n,N}^{+}$ как при любых конечных значениях n и N, так и при $\min(n, N-n)\to\infty$, включая теоремы о полном асимптотическом разложении и большие уклонения.

Теорема 1. *Справедливы формулы*

$$\Delta_{n,N}^{+}=\max_{1\leqslant k\leqslant n}\left(\frac{k}{n}-\frac{l_k}{N}\right),\qquad \Delta_{n,N}^{-}=\max_{1\leqslant k\leqslant n}\left(\frac{l_k-1}{N}-\frac{k-1}{n}\right),$$
$$\Delta_{n,N}=\max(\Delta_{n,N}^{-}, \Delta_{n,N}^{+}).\tag{5}$$

Доказательство (5) основано на том замечании, что разность $F_n(x)-F_N(x)$ изменяется лишь при $x=X_j$, $j=1, 2, \ldots, n$.

Теорема 2. *При любых $a>0$ и $b>0$ справедлива формула*

$$P(\Delta^+_{n,N}<b;\ \Delta^-_{n,N}<a)=\frac{1}{C_N^{\,n}}\left|\,\left\|C^{\,j-k+1}_{\alpha_k-\beta_j+j-k-1}\right\|_{\substack{k=1,2,\ldots,n\\ j=1,2,\ldots,n}}\right|, \tag{6}$$

где

$$\alpha_k=\min\left\{-\left[-\frac{kN+nNa+n-N}{n}\right], N-n+k+1\right\},\qquad k=1,2,\ldots,n,$$

$$\beta_j=\max\left\{\left[\frac{jN-nNb}{n}\right], j-1\right\},\qquad j=1,2,\ldots,n,$$

символ $[z]$ обозначает наибольшее целое число, не превосходящее z, $\left|\,\|d_{kj}\|_{\substack{k=1,2,\ldots,n\\ j=1,2,\ldots,n}}\right|$ обозначает определитель матрицы $\|d_{kj}\|$ порядка n с общим элементом d_{kj}.

При $N=np$, p целое, формула (6) редуцируется к формуле

$$P(\Delta^+_{n,np}<b;\ \Delta^-_{n,p}<a)=1-$$

$$-\frac{1}{C_{np}^{\,n}}\sum_{j=1}^{\left[\frac{np+\beta_0}{\alpha_0+\beta_0}\right]}\ \sum\nolimits_1 \pi_{m_1}(\alpha_0+\beta_0)\ldots\pi_{m_{j-1}}(\alpha_0+\beta_0)\cdot\pi^*_{m_j}(\alpha_0)-$$

$$-\frac{1}{C_{np}^{\,n}}\sum_{j=1}^{\left[\frac{np+\beta_0}{\alpha_0+\beta_0}\right]}\ \sum\nolimits_1 \pi_{m_1}(\alpha_0+\beta_0)\ldots\pi_{m_{j-1}}(\alpha_0+\beta_0)\cdot\pi^*_{m_j}(\beta_0)+$$

$$+\frac{1}{C_{np}^{\,n}}\sum_{j=1}^{\left[\frac{np}{\alpha_0+\beta_0}\right]}\ \sum\nolimits_1 \pi_{m_1}(\alpha_0+\beta_0)\ldots\pi_{m_{j-1}}(\alpha_0+\beta_0)\cdot\pi^*_{m_j}(\alpha_0+\beta_0)+$$

$$+\frac{1}{C_{np}^{\,n}}\sum_{j=1}^{\left[\frac{np}{\alpha_0+\beta_0}\right]}\ \sum\nolimits_2 \pi_{m_1}(\alpha_0+\beta_0)\ldots\pi_{m_{j-1}}(\alpha_0+\beta_0)\cdot\pi_{m_j}(\alpha_0)\cdot\pi^*_{m_{j+1}}(\beta_0),$$

где, во-первых, $\alpha_0=-[-npa]$, $\beta_0=-[-npb]$; во-вторых, сумма $\sum_1$ содержит все слагаемые, отвечающие неотрицательным целым числам под условием $m_1+\ldots+m_j=n$, а сумма $\sum_2$ — все слагаемые, отвечающие неотрицательным целым числам под условием $m_1+\ldots+m_{j+1}=n$; в-третьих, при $\alpha>1$ и целом $m\geqslant 0$

$$\pi_m(\alpha)=\sum_{k=0}^{m-\left[\frac{\alpha-1}{p}\right]-1}\frac{(\alpha-1)\alpha-kp^2+kp}{(kp+\alpha-1)(k(p-1)+\alpha)}\binom{kp+\alpha-1}{k}\binom{(m-k)p-\alpha}{m-k}$$

$$\pi_m^*(\alpha)=\sum_{k=0}^{m-\left[\frac{\alpha-1}{p}\right]-1}\frac{\alpha}{k(p-1)+\alpha}\binom{kp+\alpha-1}{k}\cdot\binom{(m-k)p-\alpha}{m-k}$$

если $\alpha<mp-1$, и $\pi_m(\alpha)=\pi_m^*(\alpha)=0$, если $\alpha\geqslant mp-1$.

Теорема 3. *При любых $a>0$, $b>0$ и $N\geqslant n+1$ справедлива формула*

$$P(\Delta_{n,N}^{+}<b;\ \Delta_{n,N}^{-}<a)=P\left(D^{+}(n,N-n)<\frac{Nb}{N-n};\ D^{-}(n,N-n)<\frac{Na}{N-n}\right). \quad (7)$$

Теорема 4. *При любых фиксированных числах $\lambda_1>0$ и $\lambda_2>0$ справедлива формула*

$$\lim_{\min(n,\,N-n)\to\infty} P\left(\sqrt{\frac{nN}{N-n}}\,\Delta_{n,N}^{+}<\lambda_1;\ \sqrt{\frac{nN}{N-n}}\,\Delta_{n,N}^{-}<\lambda_2\right)=$$

$$=1-\sum_{k=1}^{\infty}e^{-2[k\lambda_1+(k-1)\lambda_2]^2}-\sum_{k=1}^{\infty}e^{-2[(k-1)\lambda_1+k\lambda_2]^2}+2\sum_{k=1}^{\infty}e^{-2[k(\lambda_1+\lambda_2)]^2}. \quad (8)$$

Доказательство (8) основано на формуле (7) и известных теоремах Гихмана ([10]).

Теорема 5. *При $p\geqslant 2$ имеет место формула (асимптотического характера)*

$$P\left(\sqrt{\frac{np}{p-1}}\,\Delta_{n,np}^{+}<\lambda_1;\ \sqrt{\frac{np}{p-1}}\,\Delta_{n,np}^{-}<\lambda_2\right)=1-\frac{(p-1)^{n(p-1)-1/2}p^{np+1/2}}{\sqrt{2\pi n}\,C_{np}^{\,n}}\times$$

$$\times\sum_{k=0}^{\infty}\left\{\sum_{j=1}^{\infty}e^{-2[j\lambda_1+(j-1)\lambda_2]^2}Q_k^{1}(2j-1,H)+\sum_{j=1}^{\infty}e^{-2[\lambda_1(j-1)+\lambda_2 j]^2}Q_k^{2}(2j-1,H)-\right.$$

$$\left.-\sum_{j=1}^{\infty}e^{-2[j(\lambda_1+\lambda_2)]^2}[P_k^{1}(2j,H)+P_k^{2}(2j,H)]\right\}\cdot\left(\frac{1}{\sqrt{n}}\right)^{k}, \quad (9)$$

где $Q_k^1(2j-1,H)$, $Q_k^2(2j-1,H)$, $P_k^1(2j,H)$, $P_k^2(2j,H)$ обозначают соответственно результат замены степени z^l, входящей в полином $Q_k(2j-1,z)\cdot z^k$, на $2^{-l/2}H_l(\sqrt{2}[j\lambda_1+(j-1)\lambda_2])$, $2^{-l/2}H_l(\sqrt{2}[(j-1)\lambda_1+j\lambda_2])$, степени z^l, входящей в $P_k^1(2j,z)\cdot z^k$, на $2^{-l/2}H_l(\sqrt{2j}(\lambda_1+\lambda_2))$ и степени z^l, входящей в $P_k^2(2j,z)\cdot z^k$, на $2^{-l/2}H_l(\sqrt{2j}(\lambda_1+\lambda_2))$.

Упомянутые полиномы $Q_k(2j-1,z)$, $P_k^1(2j,z)$ и $P_k^2(2j,z)$ вводятся цепочкой формул:

$$\ln\left[1-\sum_{k=1}^{\infty}\frac{2^{2k}B_{2k}}{(2k)!}\,\frac{(p-1)^{2k-1}-1}{(\sqrt{p-1})^{2k-1}(\sqrt{p})^{2k+1}}X^{2k-1}\right]=\sum_{k=1}^{\infty}M_kX^k; \quad (а)$$

$$\exp\left\{\sum_{k=1}^{\infty}S_k(2j-1,z)\cdot X^k\right\}=\sum_{k=0}^{\infty}Q_k(2j-1,z)\cdot X^k, \quad (б)$$

$$S_k(2j-1,z)=\begin{cases}(2j-1)\left[M_k+\dfrac{1-\omega_1-\omega_2}{\sqrt{p(p-1)}}\right]-\dfrac{1+\omega_1+\omega_2}{\sqrt{p(p-1)}}, & \text{если}\quad k=1,\\[2ex] (2j-1)\dfrac{1-(-1)^k}{2}M_k+\dfrac{1+(-1)^k}{2}\left[M_k+\dfrac{2^{k+2}B_{k+2}}{(k+2)(k+2)!}\times\right.\\ \left.\times\dfrac{p^{k+3}-(p-1)^{k-3}-1}{(\sqrt{p(p-1)})^{k+2}}z^2\right]; & \text{если}\quad k\geqslant 2;\end{cases}$$

$$\exp\left\{\sum_{k=1}^{\infty}R_k^{1}(2j,z)\cdot X^k\right\}=\sum_{k=0}^{\infty}P_k^{1}(2j,z)\cdot X^k, \quad (в)$$

$$R_k^1(2j,z)=$$

$$=\begin{cases} 2j\left[\dfrac{1-(-1)^k}{2}M_k+\dfrac{1-\omega_1-\omega_2}{\sqrt{p(p-1)}}+(-1)^kM_k-\dfrac{1}{\sqrt{p(p-1)}}\right], \text{ если } k=1, \\ 2j\dfrac{1+(-1)^k}{2}M_k+\dfrac{1+(-1)^k}{2}\dfrac{2^{k+2}B_{k+2}}{(k+2)(k+2)!}\dfrac{p^{k+3}-(p-1)^{k+3}-1}{(\sqrt{p(p-1)})^{k+2}}z^2, \\ \text{если } k\geqslant 2; \end{cases}$$

$$\exp\left\{\sum_{k=1}^{\infty}R_k^2(2j,z)X^k\right\}=\sum_{k=0}^{\infty}P_k^2(2j,z)\cdot X^k, \tag{г}$$

$$R_k^2(2j,z)=$$

$$=\begin{cases} 2j\left[\dfrac{1-(-1)^k}{2}M_k+\dfrac{1-\omega_1-\omega_2}{\sqrt{p(p-1)}}\right]+M_k, \text{ если } k=1, \\ 2j\dfrac{3-(-1)^k}{2}M_k+\dfrac{1+(-1)^k}{2}\dfrac{2^{k+2}B_{k+2}}{(k+2)(k+2)!}\dfrac{p^{k+3}-(p-1)^{k+3}-1}{(\sqrt{p(p-1)})^{k+2}}z^2 \\ \text{если } k\geqslant 2. \end{cases}$$

Наконец, B_k — число Бернулли, $H_l(t)$ — полином Эрмита,
$\omega_1=-[-\lambda_1\sqrt{np(p-1)}]-\lambda_1\sqrt{np(p-1)}$, $\omega_2=-[-\lambda_2\sqrt{np(p-1)}]-\lambda_2\sqrt{np(p-1)}$.

Доказательство (9) основано на представлении вероятности

$$P\left(\sqrt{\frac{np}{p-1}}\Delta_{n,np}^{+}<\lambda_1;\ \sqrt{\frac{np}{p-1}}\Delta_{n,np}^{-}<\lambda_2\right)$$

в виде контурного интеграла, установлении свойств подынтегральной функции и применении асимптотического метода Лапласа при большом значении параметра n, при этом применяется теорема 2.

Формулы (7) и (9) вместе дают полное асимптотическое разложение вероятности

$$P\left(\sqrt{\frac{np}{p+1}}D^+(n,np)<\lambda_1;\ \sqrt{\frac{np}{p+1}}D^-(n,np)<\lambda_2\right)$$

в ряд по степеням $1/\sqrt{n}$ при любом фиксированном целом $p\geqslant 1$. При $\lambda_1=\lambda_2=\lambda$ и $p=\infty$ (9) редуцируется к формуле, найденной ранее автором в работах ([11], [12]).

Заметим, что формула (7) и результаты Боровкова ([13]), Боровкова и Рогозина ([14]) дают ответ на вопрос о больших уклонениях для Δ-критериев (1) и (2).

Ленинградское отделение
Математического института им. В. А. Стеклова
Академии наук СССР

Поступило
16 VII 1974

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] *Ю. К. Беляев*, Тез. докл. Международн. конфер. по теории вероятностей и математической статистике, Вильнюс, 25–30 июня, т. 1, 1973, стр. 55. [2] *A. N. Kolmogorov*, Giorn. dill Inst. Ital. degli Attuari, v. 4, 83 (1933). [3] *A. N. Kolmogorov*, Ann. Math. Statist., v. 12, 461 (1941). [4] *Л. Н. Большев, Н. В. Смирнов*, Таблицы математической статистики, М., 1968. [5] *В. С. Королюк*, Теория вероятностей и ее применение, т. 4, № 4, 369 (1959). [6] *Ю. К. Беляев, Л. В. Рыкова*, ДАН, т. 210, № 6 (1973). [7] *Н. В. Смирнов*, Бюлл. Московск. унив., сер. А, 2, 3 (1939). [8] *Н. В. Смирнов*, Теория вероятностей и математическая статистика. Изб. тр., М., 1970. [9] *Л. В. Рыкова*, ДАН, т. 214, № 4 (1974). [10] *И. И. Гихман*, ДАН, т. 82, № 6 (1952). [11] *Ю. В. Боровских*, Тез. докл. Международн. конфер. по теории вероятностей и математической статистике, Вильнюс, 25–30 июня, т. 1, 1973, стр. 87. [12] *Ю. В. Боровских*, Зап. научн. семинаров Ленингр. отд. Матем. инст. им. В. А. Стеклова АН СССР, т. 41 (1974). [13] *А. А. Боровков*, Изв. АН СССР, сер. матем., т. 26, № 4, 605 (1962). [14] *А. А. Боровков, Б. А. Рагозин*, Теория вероятностей и ее применение, т. 9, № 3, 401 (1964).