

В. С. КОРОЛЮК

**АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ ДЛЯ КРИТЕРИЕВ СОГЛАСИЯ
А. Н. КОЛМОГОРОВА и Н. В. СМИРНОВА**

(Представлено академиком А. Н. Колмогоровым 25 I 1954)

Одной из задач математической статистики является задача проверки статистических гипотез, которая может быть сформулирована так: на основании некоторых соображений можно сказать, что функция распределения случайной величины ξ есть $F(x)$; спрашивается, совместимы ли наблюдаемые значения

$$x_1, x_2, \dots, x_n \quad (1)$$

с гипотезой, что ξ имеет действительно распределение $F(x)$?

Если имеются две серии независимых наблюдений

$$x_1, x_2, \dots, x_n \text{ и } y_1, y_2, \dots, y_m \quad (N = m + n), \quad (2)$$

то спрашивается, совместимы ли наблюдаемые значения с гипотезой, что обе серии получены в результате наблюдения над случайными величинами с одним и тем же распределением $F(x)$?

А. Н. Колмогоров⁽¹⁾ предложил критерий для проверки первой из указанных гипотез. Н. В. Смирновым был предложен аналогичный критерий для проверки второй гипотезы.

Критерий А. Н. Колмогорова состоит в следующем. По результатам наблюдений (1) строится эмпирическая функция распределения

$$S_n(t) = \frac{K_n(t)}{n},$$

где $K_n(t)$ — число x_k меньших, чем t .

Теорема А. Н. Колмогорова. Если $F(t)$ непрерывна, то

$$\Phi_n(z) = P \left\{ \sqrt{n} \sup_{-\infty < t < +\infty} |S_n(t) - F(t)| < z \right\} \xrightarrow{n \rightarrow \infty} \\ \xrightarrow{n \rightarrow \infty} K(z) = \begin{cases} \sum_{h=-\infty}^{+\infty} (-1)^h e^{-2h^2 z^2} & \text{при } z > 0; \\ 0 & \text{при } z \leq 0. \end{cases}$$

Для проверки второй гипотезы используется

Теорема Н. В. Смирнова. Если $F(t)$ непрерывна, то

$$\Phi_{nm}(z) = P \left\{ \sqrt{\frac{nm}{N}} \sup_{-\infty < t < +\infty} |T_m(t) - S_n(t)| < z \right\} \xrightarrow{m, n \rightarrow \infty} K(z),$$

где $T_m(t)$ — эмпирическая функция распределения второй серии наблюдений.

Кроме указанных теорем для проверки сформулированных гипотез могут быть использованы следующие теоремы (см. (2)):

Теорема 1. Если $F(t)$ непрерывна, то

$$\Phi_n^+(z) = P\left\{\sqrt{n} \sup_{-\infty < t < +\infty} [F(t) - S_n(t)] < z\right\} \xrightarrow{n \rightarrow \infty} 1 - e^{-2z^2} \quad (z > 0).$$

Теорема 2. Если $F(t)$ непрерывна, то

$$\Phi_{nm}^+(z) = P\left\{\sqrt{\frac{nm}{N}} \sup_{-\infty < t < +\infty} [T_m(t) - S_n(t)] < z\right\} \xrightarrow{n, m \rightarrow \infty} 1 - e^{-2z^2} \quad (z > 0).$$

Теоремы А. Н. Колмогорова и Н. В. Смирнова являются предельными и потому не учитывают влияния, которое оказывает на результат число наблюдений. Таким образом возникает задача оценки погрешностей, которые появляются при замене точных распределений предельными. Для распределения максимума односторонних уклонений между эмпирической и теоретической кривыми распределения эта задача была решена Н. В. Смирновым⁽²⁾.

В работе Б. В. Гнеденко⁽³⁾ даны асимптотические разложения для функций $\Phi_{nn}^+(z)$ и $\Phi_{nn}(z)$.

В общем случае асимптотические разложения для функций распределения максимальных уклонений определяются следующими теоремами (всюду p — целое положительное число ($p \leq 11$), $z > 0$):

Теорема 1. В точках скачков функции распределения максимума разности между двумя эмпирическими кривыми распределения при $z = O\left(\sqrt{\frac{nm}{N}}\right)$ имеет место формула:

$$\begin{aligned} \Phi_{mn}^+(z) &= P\left\{\sqrt{\frac{nm}{N}} \sup_{-\infty < t < +\infty} [T_m(t) - S_n(t)] < z\right\} = \\ &= 1 - e^{-2z^2} \left[1 - \sqrt{\frac{nm}{N}} \left(\frac{1}{m} - \frac{1}{n}\right) \frac{2z}{3} + \frac{m^2 + mn + n^2}{Nmn} \frac{2z^2}{3} \left(1 - \frac{2z^2}{3}\right)\right] + \\ &\quad + O\left[\left(\frac{m}{Nn}\right)^{3/2} z^p e^{-2z^2}\right]. \end{aligned} \quad (3)$$

Теорема 2. Функция распределения максимума разности между теоретической и эмпирической кривыми распределения при $z = O(\sqrt{n})$ представима в виде:

$$\begin{aligned} \Phi_n^+(z) &= P\left\{\sqrt{n} \sup_{-\infty < t < +\infty} [F(t) - S_n(t)] < z\right\} = \\ &= 1 - e^{-2z^2} \left[1 + \frac{2z}{3\sqrt{n}} + \frac{2z^2}{3n} \left(1 - \frac{2z^2}{3}\right)\right] + O\left[\frac{z^p}{n^{3/2}} e^{-2z^2}\right]. \end{aligned} \quad (4)$$

Теорема 3. В точках скачков функции распределения максимума модуля разности между двумя эмпирическими кривыми распределения при $z = O\left(\sqrt{\frac{nm}{N}}\right)$ имеет место формула:

$$\begin{aligned} \Phi_{nm}(z) &= P\left\{\sqrt{\frac{nm}{N}} \sup_{-\infty < t < +\infty} |T_m(t) - S_n(t)| < z\right\} = \\ &= \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2 z^2} + \frac{m^2 + mn + n^2}{Nmn} \sum_{k=-\infty}^{+\infty} (-1)^k \frac{2k^2 z^2}{3} \left(1 - \frac{2k^2 z^2}{3}\right) e^{-2k^2 z^2} + \\ &\quad + O\left[\left(\frac{m}{Nn}\right)^{3/2} z^p e^{-2z^2}\right]. \end{aligned} \quad (5)$$

Теорема 4. Функция распределения максимума модуля разности между теоретической и эмпирической кривыми распределения при $z = O(\sqrt{N})$ представима в виде

$$\Phi_n(z) = P\left\{\sqrt{N} \sup_{-\infty < t < +\infty} |F(t) - S_n(t)| < z\right\} = \quad (6)$$

$$= \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2 z^2} + \frac{1}{n} \sum_{k=-\infty}^{+\infty} (-1)^k \frac{2k^2 z^2}{3} \left(1 - \frac{2k^2 z^2}{3}\right) e^{-2k^2 z^2} + O\left(\frac{z^3}{n^{3/2}} e^{-2z^2}\right).$$

Для доказательства сформулированных теорем расположим результаты наблюдений обеих серий (2) по величине:

$$z_1 < z_2 < \dots < z_N$$

и рассмотрим суммы $S_k = \sum_{r=1}^k \varepsilon_r$ ($k = 1, 2, \dots, N$), $S_0 = 0$, случайных величин

$$\varepsilon_r = \alpha = \frac{1}{m} \sqrt{\frac{nm}{N}}, \text{ если } z_r = y_i; \quad \varepsilon_r = -\beta = -\frac{1}{n} \sqrt{\frac{nm}{N}}, \text{ если } z_r = x_j.$$

Легко показать, что

$$\max_{0 \leq k \leq N} S_k = \max_{-\infty < t < +\infty} \sqrt{\frac{nm}{N}} [T_m(t) - S_n(t)],$$

$$\max_{0 \leq k \leq N} |S_k| = \max_{-\infty < t < +\infty} \sqrt{\frac{nm}{N}} |T_m(t) - S_n(t)|.$$

Введем теперь распределения

$$p^+(x, t; z) = P\{\max_{r \leq k} S_r < z | S_k = x\}; \quad p(x, t; z) = P\{\max_{r \leq k} |S_r| < z | S_k = x\},$$

где t принимает значения вида k/N ($k = 1, 2, \dots, N$), а x — возможные значения суммы S_k .

Из определения функций $p^+(x, t; z)$ и $p(x, t; z)$ следует, что

$$\Phi_{nm}^+(z) = p^+(0, 1; z), \quad \Phi_{nm}(z) = p(0, 1; z).$$

В силу теоремы В. И. Гливенко ((⁴), стр. 282),

$$\Phi_n^+(z) = \lim_{m \rightarrow \infty} \Phi_{nm}^+(z) = \lim_{m \rightarrow \infty} p^+(0, 1; z),$$

$$\Phi_n(z) = \lim_{m \rightarrow \infty} \Phi_{nm}(z) = \lim_{m \rightarrow \infty} p(0, 1; z).$$

С другой стороны, функции $p^+(x, t; z)$ и $p(x, t; z)$ являются решениями разностного уравнения

$$B(p) = \frac{m}{N} \left(1 + \alpha \frac{x}{t}\right) p\left(x - \alpha, t - \frac{1}{N}\right) +$$

$$+ \frac{n}{N} \left(1 - \beta \frac{x}{t}\right) p\left(x + \beta, t - \frac{1}{N}\right) - p(x, t) = 0 \quad (7)$$

с граничными условиями для $p^+(x, t; z)$:

$$p^+(x, \beta x; z) = 1 \quad \text{при } 0 \leq x < z; \quad p^+(x, -\alpha x; z) = 1 \quad \text{при } x < 0;$$

$$p^+(x, t; z) = 0 \quad \text{при } x \geq z, \quad (8)$$

и для $p(x, t; z)$:

$$p(x, \beta x; z) = 1 \quad \text{при } 0 \leq x < z; \quad p(x, -\alpha x; z) = 1 \quad \text{при } -z < x < 0;$$

$$p(x, t; z) = 0 \quad \text{при } |x| \geq z. \quad (9)$$

Если теперь построить асимптотические разложения для функций $p^+(x, t; z)$ и $p(x, t; z)$ как решений разностного уравнения (7), то при $x=0$, $t=1$ получим асимптотические формулы для функций распределения максимальных уклонений $\Phi_{nm}^+(z)$ и $\Phi_{nm}(z)$, а их предельные значения при фиксированном n и $m \rightarrow \infty$ дадут асимптотические представления $\Phi_n^+(z)$ и $\Phi_n(z)$. Асимптотические разложения для $p^+(x, t; z)$ и $p(x, t; z)$ находятся как обычно, когда в уравнение входит малый параметр (здесь таковым служит $\alpha = \sqrt{n/Nm}$): представим разностный оператор B в виде ряда по степеням малого параметра α

$$B = A_0 + \alpha(1 - \nu)A_1 + \alpha^2 A_2 + \beta^3 A_3 \quad \left(\nu = \frac{m}{n}\right),$$

и решение уравнения (7) ищем в виде

$$p = p_0 + \alpha(1 - \nu)p_1 + \alpha^2 p_2 + \beta^3 p_3,$$

где $p_0(x, t)$, $p_1(x, t)$, $p_2(x, t)$ определяются последовательной системой дифференциальных уравнений:

$$\frac{1}{2} \frac{\partial^2 p_0}{\partial x^2} - \frac{x}{t} \frac{\partial p_0}{\partial x} - \frac{\partial p_0}{\partial t} = 0;$$

$$p_0^+(x, 0) = 1; \quad p_0^+(z, t) = 0; \quad p_0(x, 0) = 1; \quad p_0(z, t) = p_0(-z, t) = 0. \quad (10)$$

$$\frac{1}{2} \frac{\partial^2 p_1}{\partial x^2} - \frac{x}{t} \frac{\partial p_1}{\partial x} - \frac{\partial p_1}{\partial t} = \frac{1}{6} \frac{\partial^3 p_0}{\partial x^3} - \frac{x}{2t} \frac{\partial^2 p_0}{\partial x^2};$$

$$p_1^+(x, 0) = p_1^+(z, t) = 0; \quad p_1(x, 0) = p_1(z, t) = p_1(-z, t) = 0. \quad (11)$$

$$\frac{1}{2} \frac{\partial^2 p_2}{\partial x^2} - \frac{x}{t} \frac{\partial p_2}{\partial x} - \frac{\partial p_2}{\partial t} = \frac{\nu}{2} \frac{\partial^3 p_0}{\partial x^2 \partial t} + \frac{(\nu^2 - \nu + 1)x}{6t} \frac{\partial^3 p_0}{\partial x^3} - \frac{(\nu^2 - \nu + 1)}{24} \frac{\partial^4 p_0}{\partial x^4} - \frac{\nu x}{t} \frac{\partial^2 p_0}{\partial x \partial t} - \frac{\nu}{2} \frac{\partial^2 p_0}{\partial t^2} + (1 - \nu)^2 \left[\frac{1}{6} \frac{\partial^3 p_1}{\partial x^3} - \frac{x}{2t} \frac{\partial^2 p_1}{\partial x^2} \right];$$

$$p_2^+(x, 0) = p_2^+(z, t) = 0; \quad p_2(x, 0) = p_2(z, t) = p_2(-z, t) = 0. \quad (12)$$

Для остаточного члена $p_3(x, t)$ получается разностное уравнение (7) с правой частью:

$$B(p_3) = -\frac{1}{\beta^3} \left\{ \left[B - \frac{1}{N} (A_0 + \alpha(1 - \nu)A_1 + \alpha^2 A_2) \right] (p_0) + \alpha(1 - \nu) \left[B - \frac{1}{N} (A_0 + \alpha(1 - \nu)A_1) \right] (p_1) + \alpha^2 \left[B - \frac{1}{N} A_0 \right] (p_2) \right\}. \quad (13)$$

Граничные условия для $p_3(x, t)$ находятся из соотношения:

$$p_3(x, t) = \frac{1}{\beta^3} [p(x, t) - p_0(x, t) - \alpha(1 - \nu)p_1(x, t) - \alpha^2 p_2(x, t)]. \quad (14)$$

Выбор граничных значений для p_0 , p_1 и p_2 определяется условием, чтобы $p_3(x, t)$ на границе была ограниченной при любых n и m .

Асимптотические разложения обосновываются доказательством ограниченности остаточного члена $p_3(0, 1)$ при всех n и m .

Таким образом, получены формулы (3) и (5), из которых предельным переходом по $m \rightarrow \infty$ получаются асимптотические представления функций $\Phi_n^+(z)$ и $\Phi_n(z)$, т. е. формулы (4) и (6).

Любопытно отметить, что для двухсторонних уклонений первый остаточный член разложения имеет порядок $1/n$.

Поступило
25 I 1954

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

- ¹ А. Н. Колмогоров, *Ist. Ital. Attuari*, 4 (1933). ² Н. В. Смирнов, *Усп. матем. наук*, в. 10 (1944). ³ Б. В. Гнеденко, *ДАН*, 82, № 5 (1952). ⁴ Б. В. Гнеденко, *Курс теории вероятностей*, 1950.