Доклады Академии Наук СССР
1952. Том LXXXII, № 4

МАТЕМАТИКА

Действительный член Академии наук УССР Б. В. ГНЕДЕНКО и Е. Л. РВАЧЕВА

ОБ ОДНОЙ ЗАДАЧЕ СРАВНЕНИЯ ДВУХ ЭМПИРИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

В недавней статье одного из авторов и В. С. Королюка [1] был изложен метод решения задачи о расхождении эмпирических функций распределения. Настоящая заметка примыкает к ней как по методу, так и по постановке задачи.

Пусть имеются две серии результатов независимых испытаний, как внутри каждой серии, так и между сериями,

$$x_1, x_2, \ldots, x_n \quad \text{и} \quad y_1, y_2, \ldots, y_n,$$

произведенными над случайными величинами с одной и той же непрерывной функцией распределения $F(x)$. Введем обозначения

$$F_1(x) = \frac{k_1(x)}{n}, \quad F_2(x) = \frac{k_2(x)}{n},$$

где $k_1(x)$ — число x_k, меньших чем x, а $k_2(x)$ — число y_k, меньших чем x. Положим далее $\alpha = [x\sqrt{2n}]$, $\beta = [y\sqrt{2n}]$,

$$D_n^- = -\min_{-\infty<x<\infty}\{F_1(x) - F_2(x)\} = \max_{-\infty<x<\infty}\{F_2(x) - F_1(x)\},$$

$$D_n^+ = \max_{-\infty<x<\infty}\{F_1(x) - F_2(x)\}.$$

Задача настоящей статьи состоит в определении совместной функции распределения величин D_n^- и D_n^+.

Теорема 1. *В указанных предположениях имеют место равенства:*

$$\Phi_n(x,y) = P\left\{\sqrt{\frac{n}{2}}D_n^- < x, \sqrt{\frac{n}{2}}D_n^+ < y\right\} = \begin{cases} 0, & \text{если } \min(x,y) \leqslant \frac{1}{\sqrt{2n}}, \\ 1, & \text{если } \min(x,y) > \sqrt{\frac{n}{2}}; \end{cases}$$

$$\Phi_n(x,y) = \frac{1}{C_{2n}^n}\left[\sum_{s=-\left[\frac{n}{\alpha+\beta}\right]}^{\left[\frac{n}{\alpha+\beta}\right]} C_{2n}^{n-s(\alpha+\beta)} - \sum_{s=1}^{\left[\frac{n+\alpha}{\alpha+\beta}\right]} C_{2n}^{n+\alpha-s(\alpha+\beta)} - \sum_{s=1}^{\left[\frac{n+\beta}{\alpha+\beta}\right]} C_{2n}^{n+\beta-s(\alpha+\beta)}\right]$$

при остальных значениях x и y.

Доказательство в значительной мере повторяет рассуждения, намеченные в ([1]). Расположим результаты обеих серий наблюдений в порядке возрастания их величины $z_1 < z_2 < \cdots < z_{2n}$ и рассмотрим последовательность вспомогательных случайных величин $\xi_1, \xi_2, \ldots, \xi_{2n}$, определенных посредством условия

$$\xi_k = \begin{cases} +1, & \text{если } z_k \text{ принадлежит первой серии наблюдений,} \\ -1, & \text{если } z_k \text{ принадлежит второй серии наблюдений.} \end{cases}$$

Положив

$$S_k = \xi_1 + \xi_2 + \cdots + \xi_k,$$

находим, что

$$nD_n^+ = \sup_{1 \leqslant k \leqslant 2n} S_k, \quad nD_n^- = -\inf_{1 \leqslant k \leqslant 2n} S_k.$$

Изменение сумм S_k при изменении числа слагаемых k нам будет полезно проиллюстрировать геометрически следующим образом: рассмотрим частицу, подверженную случайным толчкам в моменты $t = 1, 2, \ldots, 2n$. В начальный момент частица находится в положении $S_0 = 0$, в результате каждого толчка она сдвигается на единицу вправо или единицу влево, в зависимости от значения, принятого ξ_k, и на единицу «вверх».

В нашей интерпретации вероятность $P\{nD_n^- < \alpha, nD_n^+ < \beta\}$ представляет собой вероятность того, что частица, двигающаяся указанным образом, во все время движения находится между прямыми $x = -\alpha$ (a) и $x = \beta$ (b), не достигая их. При этом следует учитывать, что каждая возможная траектория начинается в точке $(0, 0)$ и заканчивается в точке $(0, 2n)$ и все траектории равновероятны (общее число возможных траекторий равно C_{2n}^n). Остается подсчитать число траекторий, благоприятствующих событию $\{nD_n^- < \alpha, nD_n^+ < \beta\}$.

С этой целью разобьем множество всех путей $\mathfrak{M}$ на следующие непересекающиеся множества: $\mathfrak{A}_0$ — траектории, не пересекающие ни (a), ни (b); $\mathfrak{A}_1$ — траектории, пересекающие (a), но не пересекающие (b); $\mathfrak{B}_1$ — траектории, пересекающие (b), но не пересекающие (a); $\mathfrak{A}_2$ — траектории, сначала пересекающие (a), затем (b) и более не пересекающие (a); $\mathfrak{B}_2$ — траектории, пересекающие сначала (b), затем (a) и более не пересекающие (b); $\mathfrak{A}_3$ — траектории, сначала пересекающие (a), затем (b), затем снова (a) и более не пересекающие (b) и т. д.

Очевидно, что, начиная с некоторого момента, так образуемые множества траекторий окажутся пустыми. Кроме того, очевидно, что

$$\mathfrak{M} = \mathfrak{A}_0 + \sum \{\mathfrak{A}_i + \mathfrak{B}_i\}.$$

Наряду с только что построенными множествами, образуем еще следующие множества траекторий: A_1 — пересекающих (a) по меньшей мере один раз; B_1 — пересекающих (b) по меньшей мере один раз; A_2 — пересекающих по меньшей мере один раз (a), а затем (b); B_2 — пересекающих по меньшей мере один раз (b), а затем (a); A_3 — пересекающих прямые (a) и (b) по меньшей мере один раз в порядке aba. Этот процесс продолжаем до естественного исчерпания. Так как имеют место равенства

$$A_1 = \mathfrak{A}_1 + \sum_{i=2} (\mathfrak{A}_i + \mathfrak{B}_i); \quad B_1 = \mathfrak{B}_1 + \sum_{i=2} (\mathfrak{A}_i + \mathfrak{B}_i);$$

$$A_2 = \mathfrak{A}_2 + \sum_{i=3} (\mathfrak{A}_i + \mathfrak{B}_i); \quad B_2 = \mathfrak{B}_2 + \sum_{i=3} (\mathfrak{A}_i + \mathfrak{B}_i)$$

и т. д., то при любом $i \geqslant 1$

$$A_{2i-1} + B_{2i-1} - A_{2i} - B_{2i} = \mathfrak{A}_{2i-1} + \mathfrak{B}_{2i-1} + \mathfrak{A}_{2i} + \mathfrak{B}_{2i},$$

и, следовательно,

$$\mathfrak{A}_0 = \mathfrak{M} - \sum_{i=1} \{A_{2i-1} + B_{2i-1} - A_{2i} - B_{2i}\}.$$

Для решения нашей задачи остается подсчитать число путей в каждом из множеств A_{2i-1}, B_{2i-1}, A_{2i}, B_{2i} $(i = 1, 2, \ldots)$. Дадим этот подсчет для множеств A_1 и A_2. Каждой траектории, исходящей из точки $(0, 0)$ и достигающей прямой (a), соответствует траектория, исходящая из точки $(0, 0)$, идущая до прямой (a) по первоначальному пути, а от точки пересечения представляющая собой зеркальное отображение относительно прямой (a) первоначального пути. Эта новая траектория заканчивается в точке $(-2\alpha, 2n)$. Число таких траекторий равно $C_{2n}^{n-\alpha}$ (число отрицательных шагов равно $n+\alpha$, число положительных равно $n-\alpha$). Если траектория достигает прямой (a), а затем прямой (b), то построенная только что указанным приемом отраженная траектория достигает прямой $x = -2\alpha - \beta$. Для подсчета таких траекторий отразим отраженную траекторию еще раз, начиная от точки пересечения с прямой $x = -2\alpha - \beta$ около этой прямой. Отраженная траектория, начавшись в точке $(0, 0)$, закончится в точке $(-2\alpha - 2\beta, 2n)$. Число их равно $C_{2n}^{n-\alpha-\beta}$ (число положительных шагов равно $n - \alpha - \beta$, число отрицательных шагов равно $n + \alpha + \beta$). Подобными же рассуждениями находим, что число траекторий в множествах A_{2i-1}, B_{2i-1}, A_{2i}, B_{2i}, соответственно, равно $C_{2n}^{n-i\alpha-(i-1)\beta}$, $C_{2n}^{n-(i-1)\alpha-i\beta}$, $C_{2n}^{n-i(\alpha+\beta)}$, $C_{2n}^{n-i(\alpha+\beta)}$. Наша теорема доказана.

Результаты статьи ([1]) очевидным образом вытекают из только что сформулированной теоремы. Действительно, для получения вероятности $P\{D_n^+ < y\}$ достаточно найти $\Phi_n\left(\sqrt{\frac{n}{2}}, y\right)$:

$$P\{D_n^+ < y\} = \Phi_n\left(\sqrt{\frac{n}{2}}, y\right) = 1 - \frac{C_{2n}^{n-\beta}}{C_{2n}^n}.$$

Точно так же вероятность $P\{-y < D_n^+ < y\}$ равна $\Phi_n(y, y)$:

$$P\{-y < D_n < y\} = \Phi_n(y, y) = \sum_{s=-\left[\frac{n}{\beta}\right]}^{\left[\frac{n}{\beta}\right]} (-1)^s \frac{C_{2n}^{n-s\beta}}{C_{2n}^n}.$$

Полученная формула дает возможность получить ряд новых результатов и может быть использована для более детальной проверки правильности гипотезы, что в обеих сериях имелось одно и то же непрерывное распределение вероятностей.

Отметим в частности, что в условиях теоремы 1 вероятность осуществления неравенства

$$F_1(x) \geqslant F_2(x)$$

при всех значениях x равна $\frac{1}{n+1}$. Для доказательства достаточно в формуле для $\Phi_n(x, y)$ положить $\alpha = 1$, $\beta = n+1$.

Теорема 1 позволяет без труда доказать следующий предельный результат:

Теорема 2. *В условиях теоремы* 1 *при* $x \geqslant 0$ *и* $y \geqslant 0$ *имеет место следующее предельное равенство*:

$$\Phi(x, y) = \lim_{n\to\infty} \Phi_n(x, y) = \sum_{s=-\infty}^{\infty} e^{-2s^2(x+y)^2} - $$

$$- \sum_{s=1}^{\infty} \left[e^{-2[sx+(s-1)y]^2} + e^{-2[(s-1)x+sy]^2}\right].$$

Только что указанная функция $\Phi(x, y)$ была найдена Н. В. Смирновым как предельное совместное распределение для максимального и минимального уклонений эмпирического распределения от истинного.

Институт математики
Академии наук УССР

Поступило
13 XII 1951

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] Б. В. Гнеденко и В. С. Королюк, ДАН, 80, № 4 (1951).