Доклады Академии Наук СССР
1952. Том LXXXV, № 1

МАТЕМАТИКА

Действительный член Академии наук УССР Б. В. ГНЕДЕНКО
и В. С. МИХАЛЕВИЧ

ДВЕ ТЕОРЕМЫ О ПОВЕДЕНИИ ЭМПИРИЧЕСКИХ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ

Пусть имеются две последовательности

$$x_1, x_2, \ldots, x_n; \qquad y_1, y_2, \ldots, y_m \tag{1}$$

результатов независимых испытаний над случайными величинами с одной и той же непрерывной функцией распределения $F(x)$.

Обозначим через $S_n(x)$ и $T_m(x)$ соответствующие эмпирические функции распределения и назовем «положительными скачками» функции $S_n(x)$ все те точки x_k, в которых выполняется неравенство

$$S_n(x_k - 0) = \frac{k-1}{n} \geqslant T_m(x_k).$$

Обозначим далее через $C(n, m)$ число положительных скачков $S_n(x)$ относительно $T_m(x)$.

Теорема 1. *Если $n = mp$, где p — целое число, то случайная величина $C(n, m)$ распределена равномерно, т. е.*

$$P\{C(n, m) = k\} = \frac{1}{n+1} \quad при \ k = 0, 1, \ldots, n.$$

Только что сформулированная теорема несколько обобщает результат, полученный нами ранее [1] для случая $p = 1$.

Теорема 2. *Случайная величина Δ_m, равная мере проекции на ось ординат множества точек $(x, F(x))$, в которых $F(x) \geqslant T_m(x)$, при любом m распределена равномерно в интервале* $(0, 1)$.

Доказательство теоремы 1. Выпишем члены последовательностей (1) в одну общую последовательность в порядке возрастания их величины

$$z_1 \leqslant z_2 \leqslant \ldots \leqslant z_{n+m}. \tag{2}$$

Каждому числу z_k поставим в соответствие случайную величину

$$\xi_k = \begin{cases} +1, & \text{если} \quad z_k = \text{одному из } x_i, \\ -p, & \text{если} \quad z_k = \text{одному из } y_j. \end{cases}$$

Мы прибегнем далее к следующей геометрической иллюстрации: частица, находящаяся в момент $t = 0$ в положении $s = 0$, подвержена случайным толчкам в моменты $t = 0, 1, \ldots, n + m - 1$, в результате каждого из которых она сдвигается или на $+1$ или на $-p$. В плоскости (s, t) путь частицы при каждом толчке изобразится перемещением на единицу вверх и на единицу вправо или на p единиц влево. Звенья этого пути мы будем изображать отрезками, соединяющими каждую точку остановки движения с очередной точкой остановки. Легко проверить, что все траектории равновероятны. Для доказательства теоремы полезно произвести следующее преобразование: каждый временной промежуток, соответствующий движению справа налево, растянем в p раз. В результате такого преобразования конечная точка каждой траектории сдвинется из точки $(0,\ n+m)$ в точку $(0,\ 2n)$. Обозначим множество полученных таким образом траекторий через $\mathfrak{M}(n, m)$, а каждую из этих траекторий назовем p-траекторией. Для нас существенно отметить, что каждая p-траектория может рассматриваться теперь также в качестве траектории, входящей в состав множества $\mathfrak{M}(n, n)$.

Разобьем множество $\mathfrak{M}(n, n)$ на непересекающиеся подмножества A_k, включив в множество A_k те и только те траектории, которые имеют k положительных скачков. В статье ([1]) было доказано, что все множества A_k равночисленны. Докажем теперь, что p-траектории, входящие в множество A_k, можно поставить во взаимно-однозначное соответствие с p-траекториями, входящими в множество A_{k+1} $(k=0, 1, \ldots, n-1)$. Этим, очевидно, теорема будет доказана.

В заметке (1) было введено понятие 0-отрезка; напомним его. 0-отрезком называется часть траектории, состоящая из одинакового числа «положительных» и «отрицательных» шагов и обладающая вдобавок следующими свойствами: 1) до 0-отрезка число «плюсов» и «минусов» в траектории одинаково; 2) на всем протяжении 0-отрезка число «минусов» не меньше числа «плюсов».

Любая траектория, входящая в множество A_k, может быть разбита на следующие части: α_1) за первые $2c$ шагов происходят c положительных скачков $(c=0, 1, \ldots, k)$; α_2) от шага $2c$ до шага $2(c+s)$ идет 0-отрезок, не имеющий общих точек с осью ординат; α_3) от $2(c+s)$ до $2(c+s+t)$ идет 0-отрезок ($t \geqslant 0$; если $c=k$, то этот 0-отрезок заканчивается в точке $(0, 2n)$); α_4) если $c<k$, то остальная часть траектории начинается положительным скачком и от $2(c+s+t)$ до $2n$ происходят $k-c$ положительных скачков.

Каждой траектории a из A_k поставим в соответствие траекторию b из A_{k+1} следующим путем: отрезок α_3 траектории a помещаем в начале траектории, последний положительный шаг отрезка α_2 помещаем непосредственно за α_3, далее присоединяем отрезок α_1, за ним оставшуюся часть отрезка α_2 и за ней отрезок α_4. Легко видеть, что получившаяся траектория имеет $k+1$ положительный скачок, различным траекториям $a \in A_k$ соответствуют различные траектории $b \subset A_{k+1}$, каждая p-траектория переходит в p-траекторию.

Указанное соответствие взаимно-однозначно, так как обратным построением каждая траектория, принадлежащая A_{k+1}, может быть переведена единственным путем в некоторую траекторию можества A_k.

Доказательство теоремы 2. Доказательство теоремы мы осуществим таким путем: построим последовательность таких случайных величин $\zeta_m(p)$, что при $p \to \infty$

1. $P\{\xi_m(p) \to \Delta_m\} = 1.$

$$2.\ P\{\xi_m(p) < x\} \to \begin{cases} 0 & \text{при} \quad x \leqslant 0, \\ x & \text{при} \quad 0 < x \leqslant 1, \\ 1 & \text{при} \quad x > 1. \end{cases}$$

Очевидно, что этим теорема будет доказана.

Произведем над случайной величиной с функцией распределения $F(x)$ новую серию в mp независимых испытаний и обозначим через μ_p

число наблюдений, попавших в множество точек, в которых $F(x) \geqslant T_m(x)$. Согласно теореме Гливенко, при $p \to \infty$

$$P\left\{\frac{\mu_p}{mp} \to \Delta_m\right\} = 1. \tag{3}$$

В силу теоремы 1 при $p \to \infty$

$$P\left\{\frac{C(mp,m)}{mp} < x\right\} \to \begin{cases} 0 & \text{при} \quad x \leqslant 0, \\ x & \text{при} \quad 0 < x \leqslant 1, \\ 1 & \text{при} \quad 1 < x. \end{cases} \tag{4}$$

В силу теоремы Гливенко для каждого $\varepsilon > 0$ найдется такое p_0, что при всех $p > p_0$ с вероятностью, сколь угодно близкой к единице, будет выполняться неравенство

$$|S_{mp}(x) - F(x)| < \varepsilon.$$

Пусть $\varepsilon > 0$ достаточно мало ($\varepsilon < {}^1/_2\, m$); в этом случае очевидно, что $C(mp, m)$ может отличаться от μ_p самое большее на то число наблюдений, которые попадают в множество проекций на ось абсцисс тех частей функции $T_m(x)$, для которых выполняются неравенства

$$F(x) - \varepsilon < T_m(x) < F(x) + \varepsilon.$$

Так как это множество составлено не более чем из m интервалов и в каждый интервал вероятность попадания не более чем 2ε, то вероятность попадания в указанное множество при mp испытаниях не превосходит $2m^2p\varepsilon$. Отсюда мы заключаем, что при любом $\varepsilon > 0$ имеет место неравенство

$$|C(mp, m) - \mu_p| < 2m^2 p\varepsilon$$

с вероятностью, сколь угодно близкой к единице, если только $p > p_0$.

Это соотношение вместе с (3) доказывает, что при $p \to \infty$

$$P\left\{\frac{C(mp, m)}{mp} \to \Delta_m\right\} = 1.$$

Последнее соотношение и (4) показывают, что мы можем положить $\xi_m(p) = \frac{C(mp, m)}{mp}$. Теорема доказана.

Институт математики
Академии наук УССР и
Киевский государственный университет

Поступило
26 IV 1952

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] Б. В. Гнеденко и В. С. Михалевич, ДАН, **82**, № 6 (1952).