Доклады Академии Наук СССР
1947. Том LVI, № 1

МАТЕМАТИКА

Н. В. СМИРНОВ

О КРИТЕРИИ СИММЕТРИИ ЗАКОНА РАСПРЕДЕЛЕНИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

(Представлено академиком С. Н. Бернштейном 24 X 1946)

1. Пусть $F(x)=P(\xi\leqslant x)$ — непрерывная функция распределения случайной величины ξ. Если для любого x выполняется условие

$$F(a+x)-F(a)=F(a)-F(a-x), \tag{1}$$

то закон распределения называется симметричным относительно „центра" a. Из общих свойств функций распределения и (1) следует:

$$F(a)=\frac{1}{2}, \tag{2}$$

$$F(x+a)=1-F(a-x). \tag{3}$$

Центр симметрии совпадает с медианой кривой распределения.

Пусть над случайной величиной ξ произведены n независимых наблюдений. Если желательно по наблюденному материалу проверить предположение о симметричности неизвестного закона распределения относительно центра a, то целесообразно поступить следующим образом. Пусть для $x>0$ h_x^+ — число наблюдений, попавших в промежуток $(a, a+x)$, и h_x^- — число наблюдений в промежутке $(a-x, a)$. Рассмотрим величины:

$$\tau_n=\operatorname*{Max}_{(x>0)}\{h_x^+-h_x^-\}, \tag{4}$$

$$\tau_n^*=\operatorname*{Max}_{(x>0)}\{|h_x^+-h_x^-|\}. \tag{4*}$$

Если проверяемая гипотеза истинна, то величины τ_n и τ_n^*, в силу (1), должны быть невелики по сравнению со всем числом наблюдений. Мы покажем, что τ_n и τ_n^* следует законам распределения, не зависящим от вида функции $F(x)$, удовлетворяющей при сделанной гипотезе условию (1). С их помощью легко можно установить границы для вариации этих случайных величин, превышение которых ставит под сомнение правильность проверяемой гипотезы.

2. Величины τ_n (4) и τ_n^* (4*) при данном n могут принимать целые положительные значения, не бо́льшие n. Положим

$$P(\tau_n\geqslant\nu)=q_{\nu n}, \tag{5}$$

$$q_{0n}=1, \tag{5'}$$

$$P(\tau_n^*\geqslant\nu)=q_{\nu n}^*, \tag{6}$$

$$q_{1n}^*=1. \tag{6 }$$

Пусть, далее, $t_k=|x_k-a|$ $(k=1, 2, \ldots, n)$ — абсолютные величины уклонений наблюденных значений x_k от медианы a, расположенные в порядке возрастания, так что $0 < t_1 < t_2 < \ldots < t_n$ *. Пусть E_k — событие, заключающееся в осуществлении равенства $x_k - a = t_k > 0$ и $\overline{E}_k$ — противоположное событие, равносильное равенству $x_k - a = -t_k < 0$. Рассмотрим последовательность n испытаний относительно событий $E_1, E_2, \ldots, E_n$.

Легко усмотреть, что эти испытания при сделанной гипотезе независимы и вероятность $P(E_k) = P(\overline{E}_k) = \frac{1}{2}$ для $k = 1, 2, \ldots, n$. Если обозначить через m_s^+ и m_s^-, соответственно, число положительных и число отрицательных исходов первых s испытаний, то будем иметь

$$\tau_n = \operatorname{Max}(m_s^+ - m_s^-), \tag{7}$$

$$s = 1, 2, \ldots, n.$$

$$\tau_n^* = \operatorname{Max}|m_s^+ - m_s^-|, \tag{7*}$$

Вероятность $q_{\nu n}$ (5) есть, очевидно, вероятность того, что разность $m_s^+ - m_s^-$ для последовательных значений $s = 1, 2, \ldots, n$ хотя бы раз достигнет значения, равного ν. Эту вероятность можно поэтому интерпретировать, как вероятность разорения игрока, обладающего капиталом ν, при безобидной игре с бесконечно богатым партнером, в результате не более чем n партий.

Аналогичным образом $q^*_{\nu n}$ (6) можно интерпретировать, как вероятность разорения хотя бы одного из двух игроков в результате не более чем n партий, если каждый из них в начале игры имел капитал, равный ν единиц; в обоих случаях предполагается, что выигрыш или проигрыш за одну партию равен единице. Очевидно, что $q^*_{\nu n}$ равна удвоенной вероятности $z_{\nu n}$ разорения одного из двух игроков.

Применяя хорошо известные классические результаты, будем иметь ((1), § 29, стр. 219)

$$q_{\nu n} = \frac{1}{2} q_{\nu+1\ n-1} + \frac{1}{2} q_{\nu-1\ n-1} \quad (q_{0n} = 1,\ q_{\nu 0} = 0), \tag{8}$$

$$q_{\nu n} = \frac{1}{2^\nu}\Big[1 + \frac{\nu}{1!}\,\frac{1}{2^2} + \frac{\nu(\nu+3)}{2!}\,\frac{1}{2^4} + \frac{\nu(\nu+4)(\nu+5)}{3!}\,\frac{1}{2^6} + \ldots + {}$$
$$+ \frac{\nu(\nu+k+1)\ldots(\nu+2k-1)}{k!}\,\frac{1}{2^{2k}}\Big], \tag{9}$$

где $k = \frac{n-\nu}{2}$ или $k = \frac{n-\nu-1}{2}$, смотря по тому, будут ли ν и n одинаковой четности или нет.

Таким же образом**

$$\frac{2}{1} q^*_{\nu n} = z_{\nu n} = q_{\nu n} - q_{3\nu\ n} + q_{5\nu\ n} - \ldots + (-1)^{s-1} q_{(2s-1)\nu\ n}, \tag{10}$$

где

$$s = \frac{\left[\frac{n}{\nu} + 1\right]}{2}.$$

---

* Первоначальный вывод закона распределения для τ_n и τ_n^*, полученный автором, был значительно сложнее, чем приводимый в дальнейшем изложении, следующем в основных чертах идее, указанной академиком С. Н. Бернштейном.

** Эта формула следует из разложений, данных А. А. Марковым ((1), стр. 214—215, см. также (2), стр. 156—157).

Для вероятности $q_{\nu n}$ можно дать другое выражение, более удобное, чем (9) (([2], стр. 149):

$$q_{\nu n} = \sum_{l=0}^{l=\frac{n-\nu-2}{2}} \frac{C_n^l}{2^{n-1}} + \frac{C_n^{\frac{n-\nu}{2}}}{2^n}, \tag{11}$$

если n и ν одинаковой четности, и

$$q_{\nu n} = \sum_{l=0}^{l=\frac{n-\nu-1}{2}} \frac{C_n^l}{2^{n-1}}, \tag{11'}$$

если n и ν разной четности.

Обозначая через $\pi_{\nu n}$ вероятность равенства $\tau_n = \nu$, найдем

$$q_{\nu n} - q_{\nu+1\, n} = \pi_{\nu n}; \tag{12}$$

отсюда, пользуясь (11), (11'), получим:

$$\pi_{2k-1\; 2n} = \pi_{2k\; 2n} = \frac{C_{2n}^{n+k}}{2^{2n}}, \tag{13}$$

$$\pi_{2k\; 2n-1} = \pi_{2k+1\; 2n-1} = \frac{C_{2n-1}^{n+k}}{2^{2n-1}}. \tag{13'}$$

3. Для применения полученных результатов при больших значениях n удобнее пользоваться асимптотическими формулами. Для $n \geqslant 50$ имеем (([2], стр. 153):

$$q_{\nu n} = 1 - 2\Phi\left(\frac{\nu}{\sqrt{n + {}^2/_3}}\right) + \frac{\theta}{6n}, \tag{14}$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-u^2/2}\, du, \qquad |\theta| < 1. \tag{15}$$

Более точный результат можно получить на основании оценок, полученных акад. С. Н. Бернштейном ([3]): при $n \geqslant 128$ и $\nu \geqslant 3$ будем иметь

$$1 - \left[\Phi\left(\frac{\nu+1}{\sqrt{n}}\right) + \Phi\left(\frac{\nu+3}{\sqrt{n}}\right)\right] - \frac{e^{-\sqrt[3]{n/2}}}{\sqrt{\pi}\sqrt[6]{n/2}} < q_{\nu n} <$$
$$< 1 - \left[\Phi\left(\frac{\nu-1}{\sqrt{n}}\right) + \Phi\left(\frac{\nu-3}{\sqrt{n}}\right)\right] + 2e^{-3\sqrt{n/2}}, \tag{16}$$

если n и ν одинаковой четности, и

$$1 - 2\Phi\left(\frac{\nu+1}{\sqrt{n}}\right) - \frac{e^{-3\sqrt{n/2}}}{\sqrt{\pi}\sqrt[6]{n/2}} < q_{\nu n} < 1 - 2\Phi\left(\frac{\nu-1}{\sqrt{n}}\right) + 2e^{-\sqrt[3]{n/2}}, \tag{16'}$$

если ν и n разной четности.

Для вероятности $q^*_{\nu n}$ имеем, в силу (10), (16) и (16'):

$$q^*_{\nu n} \sim 2 \sum_{m=0}^{m=\infty} (-1)^m \left[1 - 2\Phi\left(\frac{(2m+1)\nu}{\sqrt{n}}\right)\right] \tag{17}$$

см. также ([4])).

Для оценки $q^*_{\nu n}$ можно использовать также неравенства

$$2q_{\nu n} - 2q_{3\nu n} < q^*_{\nu n} < 2q_{\nu n}, \tag{18}$$

где $q_{\nu n}$ и $q_{3\nu n}$ приближенно определяются из (16) или (16′).

В приводимой табличке для вероятностей α, стоящих в первой строке, указаны границы λ_α и λ^*_α, превышение которых для величин $\tau_n/\sqrt{n}$ и $\tau^*_n\sqrt{n}$ при большом n имеет вероятность α.

α	0,1	0,075	0,05	0,025	0,01
λ_α	1,64	1,78	1,96	2,24	2,62
λ^*_α	1,96	2,08	2,24	2,50	2,81

Математический институт
им. В. А. Стеклова Академии Наук СССР

Поступило
24 X 1946

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] А. А. Марков, Исчисление вероятностей, М., 1924. [2] Uspensky, Introduction to Mathematical Probability, N. Y., 1937. [3] С. Н. Бернштейн, Изв. АН СССР, сер. матем.. 7, 3 (1943). [4] Н. В. Смирнов, Изв. АН СССР, сер. матем., 3, 319 (1939).