Доклады Академии Наук СССР
1950. Том LXXI, № 2

МАТЕМАТИКА

И. Д. КВИТ

О ТЕОРЕМЕ Н. В. СМИРНОВА ОТНОСИТЕЛЬНО СРАВНЕНИЯ ДВУХ ВЫБОРОК

(Представлено академиком А. Н. Колмогоровым 12 I 1950)

В работе [1] Н. В. Смирновым была изучена следующая важная задача статистики: имеются две серии результатов независимых наблюдений над случайными величинами ξ_1 и ξ_2

$$x_1,\ x_2,\ \ldots,\ x_m$$

и

$$y_1,\ y_2,\ \ldots,\ y_n.$$

Спрашивается, при каких условиях можно считать, что функции распределения $F_1(x)=P\{\xi_1<x\}$ и $F_2(x)=P\{\xi_2<x\}$ равны, и когда расхождение опытных данных настолько существенно, что гипотеза $F_1(x)=F_2(x)$ должна быть отброшена.

В настоящей заметке мы показываем, что обобщение теорем А. Н. Колмогорова и Н. В. Смирнова, полученное Г. М. Мания [2] для одной выборки, переносится без труда и на задачу Н. В. Смирнова о двух выборках.

Рассмотрим эмпирические функции распределения

$$S_m(x)=\frac{k_1(x)}{m},$$

где $k_1(x)$ — число наблюденных значений ξ_1, меньших x, и

$$T_n(x)=\frac{k_2(x)}{n},$$

где $k_2(x)$ — число наблюденных значений ξ_2, меньших x.

В точках разрыва мы дополним эмпирические функции распределения вертикальными отрезками.

В указанной работе Н. В. Смирнов доказал, что если $F_1(x)=F_2(x)$, функция $F_1(x)$ непрерывна и всюду возрастает, $\frac{m}{n}=\text{const}$, то при $N=\frac{mn}{m+n}\to\infty$

$$P\left\{D(m,\ n)\leqslant\frac{z}{\sqrt{N}}\right\}\to\sum_{k=-\infty}^{\infty}(-1)^k e^{-2k^2z^2}, \tag{1}$$

где

$$D(m,\ n)=\sup_{-\infty<x<\infty}|S_m(x)-T_n(x)|.$$

Пусть θ_1 и θ_2 — произвольные числа, $0<\theta_1<\theta_2<1$. Предположим, что $F_1(x)=F_2(x)=F(x)$ и определим α и β посредством равенств

$$F(\alpha)=\theta_1, \quad F(\beta)=\theta_2.$$

Обозначим далее

$$\alpha_{mn}=\min_x\,[S_m(x)=\theta_1;\ T_n(x)=\theta_1], \quad \beta_{mn}=\max_x\,[S_m(x)=\theta_2;\ T_n(x)=\theta_2].$$

Тогда при $m\to\infty$, $n\to\infty$, в силу теоремы Гливенко, должно быть

$$\alpha_{mn}\to\alpha, \quad \beta_{mn}\to\beta.$$

Введем обозначения

$$D^+_{mn}(\theta_1,\ \theta_2)=\sup_{(\alpha_{mn},\,\beta_{mn})}\{S_m(x)-T_n(x)\},$$

$$D_{mn}(\theta_1,\ \theta_2)=\sup_{(\alpha_{mn},\,\beta_{mn})}|\,S_m(x)-T_n(x)\,|\,.$$

Полученные результаты могут быть сформулированы в виде следующих двух теорем.

Теорема 1. *Если при* $n\to\infty$ $\theta_1^{(n)}=\theta_1+o\left(\frac{1}{\sqrt{n}}\right)$ *и* $\theta_2^{(n)}=\theta_2+o\left(\frac{1}{\sqrt{n}}\right)$, $0<\theta_1<\theta_2<1$, *то*

$$P\left\{D^+_{mn}(\theta_1^{(n)},\ \theta_2^{(n)})\leqslant\frac{z}{\sqrt{N}}\right\}\xrightarrow[N\to\infty]{}\Phi^+(\theta_1,\ \theta_2;\ z)$$

где

$$\Phi^+(\theta_1,\ \theta_2;\ z)=\frac{1}{2\pi\sqrt{1-R^2}}\int\limits_{-\infty}^{a}\int\limits_{-\infty}^{b}e^{-\frac{1}{2}Q(z_1,\ z_2)}\,dz_1dz_2-$$

$$-\frac{e^{-2z^2}}{2\pi\sqrt{1-R^2}}\int\limits_{-\infty}^{a'}\int\limits_{-\infty}^{b'}e^{-\frac{1}{2}\overline{Q}(z_1,\ z_2)}dz_1dz_2,$$

$$a=\frac{z}{\sqrt{\theta_1(1-\theta_1)}}, \qquad b=\frac{z}{\sqrt{\theta_2(1-\theta_2)}},$$

$$a'=\frac{z-2z\theta_1}{\sqrt{\theta_1(1-\theta_1)}}, \quad b'=\frac{z-2z(1-\theta_2)}{\sqrt{\theta_2(1-\theta_2)}}, \quad R=\sqrt{\frac{\theta_1(1-\theta_2)}{\theta_2(1-\theta_1)}},$$

$$Q(z_1,\ z_2)=\frac{1}{1-R^2}[z_1^2+2Rz_1z_2+z_2^2],$$

$$\overline{Q}(z_1,\ z_2)=\frac{1}{1-R^2}[z_1^2-2Rz_1z_2+z_2^2].$$

Отсюда, в частности,

$$P\left\{D^+_{mn}(0,\ 1)\leqslant\frac{z}{\sqrt{N}}\right\}\xrightarrow[N\to\infty]{}1-e^{-2z^2}.$$

Мы видим, таким образом, что теорема Н. В. Смирнова об односторонних уклонениях эмпирической функции от теоретической переносится и на случай максимума односторонних уклонений двух эмпирических функций.

Теорема 2. *В предположениях теоремы* 1

$$P\left\{D_{mn}\left(\theta_1^{(n)},\ \theta_2^{(n)}\right) \leqslant \frac{z}{\sqrt{N}}\right\}\xrightarrow[N\to\infty]{}\Phi(\theta_1,\ \theta_2;\ z),$$

где

$$\Phi(\theta_1,\ \theta_2;\ z) = \frac{1}{2\pi\sqrt{1-R^2}}\int_{-a}^{a}\int_{-b}^{b} e^{-{}^1/_2 Q(z_1,\ z_2)}dz_1dz_2 -$$

$$-\frac{1}{\pi\sqrt{1-R^2}}\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2z^2}\int_{-a_k}^{a_k}\int_{-b_k}^{b_k} e^{-{}^1/_2 Q(z_1,\ z_2)}\,dz_1dz_2,$$

$$a_k = \frac{z-2kz\theta_1}{\sqrt{\theta_1(1-\theta_1)}},\qquad b_k = \frac{z-2kz(1-\theta_2)}{\sqrt{\theta_2(1-\theta_2)}}.$$

Отсюда при $\theta_1 = 0,\ \theta_2 = 1$ *получаем* (1).

Подобно теоремам Г. М. Мания, изложенные результаты позволяют использовать тот интервал наблюденных значений, в котором результаты наблюдений наиболее достоверны.

В работе мы воспользовались методом преобразований Лапласа, примененным В. Феллером [3] для доказательств теорем А. Н. Колмогорова и Н. В. Смирнова.

В заключение выражаю глубокую благодарность проф. Б. В. Гнеденко за постановку задачи и руководство при ее решении.

Львовский государственный университет
им. И. Франко

Поступило
12 I 1950

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] Н. В. Смирнов, Бюлл. НИИМ МГУ, **2**, в. 2 (1939). [2] Г. М. Мания, ДАН, **69**, № 4 (1949), [3] W. Feller, Ann. of Math. Statistics, **19**, 2, 177 (1948).