Доклады Академии Наук СССР
1952. Том LXXXIII, № 3

МАТЕМАТИКА

Ю. В. ЛИННИК

ЛИНЕЙНЫЕ СТАТИСТИКИ И НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ

(Представлено академиком А. Н. Колмогоровым 24 I 1952)

1. Пусть имеется $r \geqslant 1$ независимых наблюдений над некоторой случайной величиной X: $x_1, x_2, \ldots, x_r$ (повторная выборка объема r). Иногда приходится рассматривать линейные функции наблюдений (например, в методе наименьших квадратов).

Пусть

$$L_1(x) = a_1x_1 + \ldots + a_rx_r \quad \text{и} \quad L_2(x) = b_1x_1 + \ldots + b_rx_r \tag{1}$$

две такие функции (линейные статистики); мы допускаем при этом обращение в нуль некоторых коэффициентов a_i и b_j.

Введем целую функцию комплексного переменного z посредством равенства:

$$\sigma(z) = |a_1|^z + \ldots + |a_r|^z - |b_1|^z - \ldots - |b_r|^z.$$

Теорема. *Для эквивалентности двух утверждений*:

А. *Наблюдения x_i принадлежат к нормальному типу распределений.*

В. *Формы $L_1(x)$ и $L_2(x)$ одинаково распределены*

необходимо и достаточно выполнение следующих условий:

1) $a_1 + a_2 + \ldots + a_r = b_1 + b_2 + \ldots + b_r$;

2) $\sigma(2) = 0$;

3) *все положительные корни $\sigma(z)$, которые являются целыми числами, делящимися на 4, должны быть простыми*;

4) *все положительные корни $\sigma(z)$, которые являются целыми четными числами $\equiv 2 \pmod 4$, должны иметь кратность не свыше 2; притом, если есть такой двукратный корень, то он должен быть единственным и максимальным из положительных корней $\sigma(z)$*;

5) *если $\sigma(z)$ имеет положительный корень γ, который не является целым четным числом, то такой корень γ должен быть единственным и максимальным из положительных корней и $[\gamma/2]$ (целая часть $\gamma/2$) должна быть нечетной.*

2. Из формулировки теоремы явствует, что при нарушении хотя бы одного из условий 1), ..., 5) утверждения А и В неэквивалентны.

Например, если число 2 — трехкратный корень $\sigma(z)$, то при достаточно большом $C > 0$ функция $\varphi(t) = e^{-Ct^2 - t^2 \ln|t|}$ будет характеристической функцией такого закона распределения вероятностей для x_i, при котором $L_1(x)$ и $L_2(x)$ одинаково распределены. Если 5 — корень $\sigma(z)$, так что $[^5/_2] = 2$ четно, то такое же утверждение можно высказать о характеристической функции $\varphi_1(t) = e^{-Ct^2 - |t|^3}$ (при достаточно большом C).

Заметим, что в первом из приведенных примеров x_i не имеют дисперсии, а во втором — шестого момента. Такое свойство примеров не случайно: если x_i имеют все моменты и $\{a_i\} \neq \{\pm b_j\}$, то x_i необходимо нормальны ([1]).

Но из теоремы явствует, что наличие любого фиксированного количества моментов x_i не обеспечивает эквивалентности А и В. (Это дает ответ на вопрос, поставленный Марцинкевичем в конце работы ([1]).)

3. Сформулированную теорему можно употребить для построения статистического критерия нормальности типа наблюдений x_i, вернее, сведения такого критерия на критерий однородности выборки.

Пусть дана повторная выборка $x_1, x_2, \ldots, x_N$ объема $N = rn$. Составим формы (1) под условиями 1), ..., 5). Вычислим значение формы $L_1(x)$ для наблюдений $x_1, \ldots, x_r$ и обозначим его через $L_1^{(1)}$, затем значение $L_2(x)$ для следующих r наблюдений и обозначим через $L_2^{(1)}$, затем значение $L_1(x)$ для следующих r наблюдений обозначим через $L_1^{(2)}$ и т. д., пока не исчерпаем наблюдения. Получим n чисел $L_i^{(j)}$.

Для нормальности типа наблюдений x_i необходимо и достаточно, чтобы $L_i^{(j)}$ представляли выборку из независимых значений одной и той же величины (повторную однородную выборку объема n).

Таким образом, критерий нормальности типа x_i сводится на критерий однородности $L_i^{(j)}$. Из таких критериев удобнее всего известный критерий однородности Н. В. Смирнова, точное распределение основной статистики которого указано в недавней работе Б. В. Гнеденко и В. С. Королюка ([2]).

Повидимому, целесообразно при построении такого критерия выбирать формы $L_1(x)$ и $L_2(x)$ не только удовлетворяющими условиям 1), ..., 5), но еще и ортогональными (т. е. под условием $\sum_{i=1}^{r} a_i b_i = 0$; в этом случае они будут независимыми при нормальном типе x_i, хотя в них будут входить одинаковые переменные).

Это соображение приводит к построению матрицы:

$$L = \begin{Vmatrix} 46{,}683 & 0 & 14{,}0049 & -10{,}773 \\ 0 & -16{,}263 & 28{,}773 & 37{,}4049 \\ 14{,}0049 & 28{,}773 & -23{,}283 & 30{,}420 \\ -10{,}773 & 37{,}4049 & 30{,}420 & -7{,}137 \end{Vmatrix}.$$

Она ортогональна с точностью до скалярного множителя, и сумма элементов по строчкам постоянна. С помощью ее 4 строчек надо построить 4 линейные формы $L_1(x)$, $L_2(x)$, $L_3(x)$, $L_4(x)$. Если для каждой пары из них составим соответствующие $\sigma(z)$ (всего 6 функций), то окажется, что $\sigma(z)$ не имеют положительных нулей, кроме простого нуля 2.

Если дана выборка объема $N = 4n$, то мы можем разбить наблюдения на n четверок, и для каждой четверки вычислить 4 значения форм L_1, L_2, L_3, L_4. Так получим $N = 4n$ величин $L_i^{(j)}$ $(i = 1, 2, 3, 4)$, причем для нормальности типа x_i будет необходимо и достаточно, чтобы $L_i^{(j)}$ были независимыми и представляли однородную повторную выборку.

Мы можем разбить нашу систему величин $L_i^{(j)}$ на две равные части и применить критерий однородности Н. В. Смирнова.

Матрица L несколько громоздка. Если есть уверенность в том, что испытуемые наблюдения имеют конечную дисперсию, то матрицу L можно заменить более простой матрицей:

$$L_1 = \begin{Vmatrix} 12 & 3 & 6 & -6 \\ -1 & -4 & 12 & 8 \\ 8 & 2 & -6 & 11 \\ -4 & 14 & 3 & 2 \end{Vmatrix}.$$

Применение ортогональных матриц позволяет свести вопрос о нормальности типа N наблюдений x_i на вопрос об однородности выборки, состоящей из N новых независимых величин $L_i^{(j)}$. Однако вопрос о наивыгоднейшем в каком-либо смысле выборе матриц и о сравнительной силе данного критерия нормальности остается пока открытым.

Поступило
24 I 1952

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] J. Marcinkiewicz, Math. Zs., 44, 612 (1938). [2] Б. В. Гнеденко и В. С. Королюк, ДАН, 80, № 4 (1951).