Доклады Академии Наук СССР
1947. Том LVII, № 2

МАТЕМАТИКА

Г. И. ЕГУДИН

ОБ УСТОЙЧИВОСТИ ОДНОГО ВЕСЬМА ШИРОКОГО КЛАССА СТАТИСТИК

(Представлено академиком А. Н. Колмогоровым 17 I 1947)

Понятие устойчивости статистик данного эмпирического распределения, неявно присутствовавшее во всех статистических исследованиях, впервые более или менее отчетливо было сформулировано Фишером ([1–3]), но без выяснения общих условий, при которых оно реализуется.

Сделаем несколько общих замечаний, относящихся к выяснению понятия устойчивости. Легко видеть, что если t_n — случайная переменная, зависящая, вообще говоря, от n, то при $n \to \infty$ может существовать не больше одного числа θ такого, что при любом $\varepsilon > 0$

$$\lim_{n \to \infty} \text{Вер.}\ \{ |t_n - \theta| < \varepsilon \} = 1. \tag{1}$$

Пусть теперь статистика t_n, вычисленная для случайной выборки объема n, такова, что для нее существует число θ, определяемое (1). Если, кроме того, дисперсия t_n с возрастанием n стремится к нулю, то из неравенства Чебышева * и единственности числа θ следует, что

$$\theta = \lim_{n \to \infty} E(t_n).$$

Теперь можно сформулировать:

*Статистика t_n, определенная для повторной ** случайной выборки объема n, называется устойчивой ***, если*

I. *Для нее существует число θ, определяемое* (1).

II. *Предел ее дисперсия при $n \to \infty$ равен нулю.*

Очевидно, что при выполнении условия II условие I равносильно существованию $\lim\limits_{n \to \infty} E(t_n)$.

С некоторых точек зрения, существенных для статистики, представляет интерес следующая классификация статистик.

Если статистика

$$t_r = f(\dot{s}_1, \dot{s}_2, \ldots, \dot{s}_r, n),$$

где $\dot{s}_k = \sum\limits_{i=1}^{n} \dot{x}_i^k$, определенная для повторной выборки объема n $(\dot{x}_1, \dot{x}_2, \ldots, \dot{x}_n)$, такова, что ее математическое ожидание, определенное для всех возможных выборок одного и того же объема, не зависит от n,

* Следует заметить, что оценка вероятностей, даваемая неравенством Чебышева, в случае произвольного распределения не может быть улучшена.

** В случае бесконечной генеральной совокупности выборка может быть и повторной.

*** Т. е. стремящейся с увеличением объема выборки к «правильному», по терминологии Фишера, значению.

то для устойчивости t_r, очевидно, достаточно выполнения условия II. Назовем такие устойчивые статистики равномерно устойчивыми. К ним относятся выборочные начальные моменты, k-статистики Фишера, l-статистики, рассмотренные в ([4]).

Статистикой, согласной с параметром генеральной совокупности θ_r (согласной оценкой θ_r), назовем такую устойчивую статистику t_r', для которой $E(t_r')=\theta_r$. Очевидно, что статистика, согласная с любым параметром, будет равномерно устойчивой статистикой. И обратно, любая равномерно устойчивая статистика t_r есть статистика, согласная с некоторым параметром θ_r' — именно с тем, который равен $E(t_r)$.

Для большинства параметров $\theta_r=f(m_1, m_2, \ldots, m_r)$ $\left(m_k=\int\limits_{-\infty}^{\infty} x^k\, dF(x)\right)$

согласные с ними статистики будут зависеть от n:

$$t_r'=\varphi(\dot{m}_1, \dot{m}_2, \ldots, \dot{m}_r, n) \qquad (\dot{m}_k=s_k/n).$$

Статистику t_r, согласную с соответствующим * ей параметром θ_r (т. е., в частности, не зависящую от n), назовем равномерно согласной статистикой.

Заметим, что, используя формулы нашей работы ([5]), можно дать общий метод нахождения согласных оценок для параметров, выражающихся полиномом от начальных моментов.

Предполагая интегральный закон распределения $F(x)$ случайной переменной X в генеральной совокупности таким, что существует интеграл Стильтьеса

$$\int\limits_{-\infty}^{+\infty} x^r\, dF(x) \tag{2}$$

(r — целое положительное число), можно показать устойчивость любой статистики $t_r=f(\dot{m}_1, \dot{m}_2, \ldots \dot{m}_r)$ повторной выборки объема n $S(\dot{x}_i,\ n)$, где $\dot{m}_k=\frac{1}{n}\sum\limits_{i=1}^{n}\dot{x}_i^k$, выражающейся полиномом от выборочных начальных моментов **.

По условию теоремы

$$t_r=\sum A\dot{m}_a^{\alpha}\dot{m}_b^{\beta}\ldots\dot{m}_l^{\lambda}, \tag{3}$$

где A — произвольные постоянные независимые от $\dot{m}$ коэффициенты, а величины $a, b, \ldots, l$; $\alpha, \beta, \ldots, \lambda$, так же как и число сомножителей, могут меняться от слагаемого к слагаемому.

Применяя к (3) формулу (12) нашей работы ([4]), получим:

$$E(t_r)=\sum AE(\dot{m}_a^{\alpha}\dot{m}_b^{\beta}\cdots\dot{m}_l^{\lambda})=\sum AO_{a,\,b,\,\ldots,\,l}^{\alpha,\,\beta,\,\ldots,\,\lambda}(m_0^n)_*=$$

$$=\sum A\frac{n(n-1)\ldots(n-\alpha+1)(n-\alpha)(n-\alpha-1)\ldots(n-\alpha-\beta+1)\ldots(n-\alpha-\beta-\ldots-\lambda+1)}{n^{\alpha+\beta+\ldots+\lambda}}\times$$

$$\times m_a^{\alpha}\ldots m_l^{\lambda}+\text{член отрицательной степени относительно } n.$$

---

* Определение этого понятия дано в ([4]).

** Показанная в нашей работе ([5]) устойчивость выборочных центральных моментов является частным случаем сформулированной выше теоремы. Требуемое здесь существование интеграла (2), из чего вытекает существование того же интеграла для всех целых положительных $k<r$, не является, по существу, ограничением для статистик, выражающихся через выборочные начальные моменты. Это есть требование существования генеральных начальных моментов m_h $(h=\overline{1,r})$, а только в этом случае и можно говорить об устойчивости рассматриваемых статистик в установленном выше смысле.

Следовательно:

$$\lim_{n \to \infty} E(t_r) = \sum A m_a^{\alpha} m_b^{\beta} \ldots m_l^{\lambda} = \theta_r, \qquad (4)$$

где θ_r—параметр, соответствующий статистике t_r.

При t_r, определяемом (3), t_r^2 будет также определяться полиномом от m_h $(h = \overline{1, r})$, и, по доказанному,

$$\lim_{n \to \infty} E(t_r^2) = \theta_r^2.$$

Следовательно, предел при $n \to \infty$ дисперсии статистики t_r будет

$$\lim_{n \to \infty} \{E(t_r^2) - [E(t_r)]^2\} = \theta_r^2 - \theta_r^2 = 0.$$

Таким образом, условия I и II выполняются и, следовательно, статистика t_r, определяемая (3), будет устойчивой *.

Поступило
17 I 1947

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] R. Fisher, Phil. Trans., A, **222**, 309 (1921). [2] R. Fisher, Proc. Cambridge Phil. Soc., **22**, 5, 700 (1925). [3] R. Fisher, Statistical Methods for Research Workers, London, 1925, 1941. [4] Г. Егудин, ДАН, **48**, № 9 (1945). [5] Г. Егудин, ДАН, **53**, № 6 (1946).

* Приведенное доказательство, в отличие от имеющихся, очевидно, свободно от предположения существования определенной конечной вероятности для каждого из возможных значений случайной переменной в генеральной совокупности.