Доклады Академии Наук СССР
1947. Том LVIII, № 8

МАТЕМАТИКА

Г. И. ЕГУДИН

НЕКОТОРЫЕ СООТНОШЕНИЯ МЕЖДУ МОМЕНТАМИ РАСПРЕДЕЛЕНИЯ КРАЙНИХ ЗНАЧЕНИЙ В СЛУЧАЙНЫХ ВЫБОРКАХ

(Представлено академиком А. Н. Колмогоровым 29 VII 1947)

За последнее время появилось много работ, относящихся к изучению законов распределения крайних значений и размаха (ранга) в выборках данного объема, главным образом из нормальной совокупности ([1-6]). Содержащиеся в этих работах вычисления некоторых констант, между прочим, весьма существенны для целей статконтроля*.

Если установить и ввести в рассмотрение законы распределения не только крайних, но и всех промежуточных значений, то, помимо некоторых общих соотношений, можно установить рекуррентные формулы, связывающие между собой константы, о которых сказано выше, и значительно облегчающие их вычисление. Эти же формулы могут быть рассматриваемы и как соотношения между интегралами определенных типов.

Пусть из генеральной совокупности, интегральный закон распределения которой $F(x)$, производится случайная повторная, а в случае бесконечной генеральной совокупности и бесповторная, выборка объема n. Обозначим $X_{1,n}$, $X_{2,n}$, ..., $X_{n,n}$ случайные переменные, представляющие в возрастающем (не убывающем) порядке значения величины X, образующие каждую из производимых выборок объема n: $X_{1,n} < X_{2,n} < \ldots < X_{n,n}$.

Тогда легко вывести $F_{k,n}(x) =$ вер $\{X_{k,n} < x\}$—интегральный закон распределения величины $X_{k,n}$. Действительно, величина $X_{k,n}$ будет меньше x в одном из $n-k+1$ следующих несовместимых случаев: 1) когда все n значений X, образовавших выборку, будут меньше x; 2) когда $n-1$ значений X меньше x, а одно значение больше x; ...; $n-k+1$; 3) когда k значений X меньше x, а остальные $n-k$ его значений больше x. Вероятность того, что в выборке объема n ровно i определенных (скажем, первых) значений X будут меньше x, а остальные $n-i$ значений будут больше x, равна [вер $\{X<x\}]^i \times$ $\times$ [вер $\{X>x\}]^{n-i} = [F(x)]^i [1-F(x)]^{n-i}$.

Вероятность же того, что в такой выборке i любых значений X будут меньше x, а остальные $n-i$ значений будут больше x, очевидно, равна $C_n^i [F(x)]^i [1-F(x)]^{n-i}$.

Таким образом, искомая вероятность $F_{k,n}(x)$ будет:

$$F_{k,n}(x) = \sum_{i=k}^{n} C_n^i [F(x)]^i [1-F(x)]^{n-i}. \tag{1}$$

* Они, например, необходимы при «размаховом» методе статконтроля, рекомендуемом в ([7]).

Пусть, далее, для некоторых целых положительных значений r существуют интегралы

$$\overline{X}^r_{k,n} = \int_{-\infty}^{\infty} x^r \, dF_{k,n}(x)$$

и

$$\overline{X}^r_s = \int_{-\infty}^{\infty} x^r d[F(x)]^s = \int_{-\infty}^{\infty} x^r \, dF_{s,s}(x) \qquad (s = \overline{1, n}),$$

т. е. $\overline{X}^r_s$ — начальный момент порядка r наибольшего значения в выборке объема s. Тогда (1) дает

$$\overline{X}^r_{k,n} = \sum_{i=k}^{n} C^i_n \sum_{l=0}^{n-i} C^l_{n-i} (-1)^l \overline{X}^r_{i+l}. \tag{2}$$

Докажем теперь следующее предложение:

Средняя арифметическая начальных моментов порядка r величин $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$ равна начальному моменту того же порядка генеральной совокупности. Т. е.

$$\frac{1}{n} \sum_{k=1}^{n} \overline{X}^r_{k,n} = \frac{1}{n} \sum_{k=1}^{n} \int_{-\infty}^{\infty} x^r \, dF_{k,n}(x) = \int_{-\infty}^{\infty} x^r \, dF(x)^*. \tag{3}$$

Для доказательства (3), очевидно, достаточно показать, что

$$\frac{1}{n} \sum_{k=1}^{n} dF_{k,n}(x) = dF(x). \tag{4}$$

Из (1)

$$\sum_{k=1}^{n} dF_{k,n}(x) = \Big\{ \sum_{k=1}^{n} \sum_{i=k}^{n} [C^i_n \, i \, [F(x)]^{i-1} [1 - F(x)]^{n-i} -$$
$$- C^i_n (n-i) [F(x)]^i [1 - F(x)]^{n-i-1}] \Big\} \, dF(x).$$

Внутренняя сумма может быть преобразована:

$$C^k_n k [F(x)]^{k-1} [1 - F(x)] + \sum_{i=k+1}^{n} C^i_n \, i \, [F(x)]^{i-1} [1 - F(x)]^{n-i} -$$
$$- \sum_{i=k}^{n} C^i_n (n-i) [F(x)]^i [1 - F(x)]^{n-i-1} = C^k_n k [F(x)]^{k-1} [1 - F(x)]^{n-k}.$$

Значит,

$$\sum_{k=1}^{n} dF_{k,n}(x) = \sum_{k=1}^{n} C^k_n k [F(x)]^{k-1} [1 - F(x)]^{n-k} =$$
$$= \sum_{k=1}^{n-1} C^k_n k [F(x)]^{k-1} [1 - F(x)]^{n-k} + n [F(x)]^{n-1} =$$

---

* Теорема верна не только для начальных моментов, но, как следует из доказательства, для математических ожиданий произвольных функций.

$$=n\sum_{k=1}^{n-1}C_{n-1}^{k-1}[F(x)]^{k-1}[1-F(x)]^{n-k}+n[F(x)]^{n-1}=$$

$$=n[F(x)]^{x-1}+n[F(x)+[1-F(x)]]^{n-1}-n[F(x)]^{n-1}=n,$$

что и доказывает (4), а следовательно, и (3).

Заметим, что равенство (3) есть, конечно, распространение на рассматриваемый случай известного свойства ассоциативности средней арифметической n величин.

Используя (2), можно записать вместо (3)

$$\sum_{k=1}^{n}\sum_{i=k}^{n}C_n^i\sum_{l=0}^{n-i}C_{n-i}^l(-1)^l\overline{X}_{i+l}^r=nm_r(X), \tag{5}$$

где через $m_r(X)$ обозначен начальный момент порядка r генеральной совокупности.

Если, в частности, закон распределения генеральной совокупности нормален, то, считая центр распределения равным нулю:

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-x^2/2\sigma^2}dx. \tag{6}$$

Тогда

$$\overline{X}_s^r=\frac{s}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}x^r e^{-x^2/2\sigma^2}\left[\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-t^2/2\sigma^2}dt\right]^{s-1}dx=a_{s,r}\sigma^r, \tag{7}$$

где $a_{s,r}$, зависящие от s и r, и являются теми величинами, знание которых, во всяком случае для $r=1, 2, 3, 4$, для различных n необходимо в статистических приложениях.

Для генеральной совокупности (6) при всех нечетных r $m_r(X)=0$ и $\overline{X}^r_{\frac{n+1}{2}+h}=\overline{X}^r_{\frac{n+1}{2}-h}$. Поэтому (5) дает следующее рекуррентное соотношение между величинами $a_{n,r}$ при различных n и фиксированном r

$$\sum_{i=\frac{n+1}{2}}^{n}C_n^i\sum_{l=0}^{n-i}C_{n-i}^l(-1)^l a_{i+l,r}=0. \tag{8}$$

Очевидно, что это равенство может рассматриваться и как рекуррентное соотношение между интегралами типа (7) при нечетных r *

При четных r $\overline{X}^r_{\frac{n}{2}-h}=\overline{X}^r_{\frac{n}{2}+h}$ и (5) дает следующее рекуррентное соотношение между величинами $a_{n,r}$ при различных h и любом фиксированном r:

$$\sum_{k=\frac{n}{2}+1}^{n}\sum_{i=k}^{n}C_n^i\sum_{l=0}^{n-i}C_{n-i}^l(-1)^l a_{i+l,r}=\frac{n}{2}. \tag{9}$$

Очевидно, что и это равенство устанавливает соотношение между интегралами типа (7) при четных r.

Если обозначить $\sigma^2(X_n)=\overline{X}_n^2-(\overline{X}_n)^2=(a_{n,2}-a_{n,1}^2)\sigma^2=\beta_n^2\sigma^2$, то

$$a_{n,2}=\beta_n^2+a_{n,1}^2,$$

* В. И. Романовский [8], рассматривая интегралы (7) при $r=1$, установил между ними соотношение, содержащееся в (8).

и (9) вместе с (8) дает рекуррентные соотношения между дисперсиями распределений наибольших (наименьших) значений в выборках различных объемов из нормальной совокупности.

Статистическое значение констант $a_{n,\,r}$ очевидно. Пусть произведено m случайных выборок, объемом n каждая, из нормальной генеральной совокупности. Обозначим через $\dot{\overline{X}}_n$ и $\dot{\overline{X_n^2}}$ статистики, представляющие собой, соответственно, первый и второй выборочные начальные моменты наибольшего значения, т. е. средние арифметические, соответственно, первых и вторых степеней наблюденных наибольших значений в каждой из m произведенных выборок одного и того же объема n. Тогда, если известно, что центр генерального распределения равен нулю, то согласной оценкой σ^2 дисперсии генеральной совокупности будет, как это следует из (7), величина

$$\frac{1}{a_{n,\,r}}\dot{\overline{X_n^2}}. \tag{10}$$

Если нормальная генеральная совокупность имеет неизвестный центр распределения, равный $\overline{X}$, то согласной оценкой этого параметра будет, очевидно, величина $\frac{1}{2}(\dot{\overline{X}}_n + \dot{\overline{X}}_{1,\,n})$ *. Здесь $\dot{\overline{X}}_{1,\,n}$ — выборочная средняя наименьшего значения, т. е. средняя арифметическая m наблюденных в каждой из произведенных выборок наименьших значений. В этом случае (10) не может служить оценкой дисперсии генеральной совокупности σ^2. Помимо очевидной оценки σ^2 через размах выборки можно и теперь построить оценки параметров генеральной совокупности σ^2 и $\overline{X}$, не вводя в рассмотрение наименьших наблюденных значений в произведенных выборках, а оперируя только наибольшими выборочными значениями $\dot{\overline{X}}_n$. В этом случае в качестве оценки σ^2 следует принять величину $\frac{1}{\beta_n^2}\,\frac{m}{m-1}\left[\dot{\overline{X_n^2}} - (\dot{\overline{X}}_n)^2\right]$, математическое ожидание которой как раз равно σ^2.

Кроме того, если $\overline{X}$ — центр распределения генеральной совокупности, то $\overline{X}_n = a_{n,\,1}\sigma + \overline{X}$. Следовательно, в качестве оценки (не согласной) величины $\overline{X}$ можно принять

$$\dot{\overline{X}}_n - \frac{a_{n,\,1}}{\beta_n}\sqrt{\frac{m}{m-1}\left[\dot{\overline{X^2}} - (\dot{\overline{X}}_n)^2\right]}.$$

Сказанное в сноске в еще большей степени распространяется на эту оценку.

Поступило
29 VII 1947

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] L. H. Tippet, Biometrika, **17**, 364 (1925). [2] E. S. Pearson, Biometrika, **18**, 173 (1926). [3] E. S. Pearson, Biometrika, **24**, 404 (1932). [4] A. T. McKay and E. S. Pearson, Biometrika, **25**, 415 (1933). [5] D. Neuman, Biometrika, **31**, 10 (1939). [6] E. S. Pearson, Biometrika, **32**, 100 (1942). [7] B. P. Pudding and W. J. Jennet, Quality Control Charts B. S. 600 R, 1942. [8] V. Romanovsky, Biometrika, **25**, 195 (1933).

* Подсчет стандартной ошибки приближения, даваемого такой оценкой, показывает, что, например, в тех случаях, когда исследователь при определении генеральной средней по выборке лимитируется не столько объемом выборки, сколько трудностью производства замеров (что часто имеет место в индустриальных приложениях), можно, не уменьшая точности приближения, значительно уменьшить число таких замеров. Вместо измерений всех N единиц выборки, требуемых обычным способом измерения средней, следует, несколько увеличив N — объем всей выборки, подразделить его на m групп объемом n каждая и произвести всего $2m$ измерений крайних значений в каждой из групп. Отбор крайних образцов часто возможен без их измерения.