АСИМПТОТИЧЕСКОЕ ПОВЕДЕНИЕ *U* – СТАТИСТИКИ ПРИ СЛУЧАЙНОМ ОБЪЕМЕ ВЫБОРКИ. (ЧАСТЬ-I)

С.А.АҲМЕДОВ, А.Г.АБДУЛЛАЕВ

Мақолада танламалар сони N_n тасодифий миқдор бўлган ҳолда қурилган U – статистиканинг асимптотик нормаллиги ҳақидаги теорема исботланган. Теоремада танланма ҳажми N_n нинг кузатиш натижалари ξ_i лардан боғлиқсизлиги талаб этилмайди ($i = 1,2,...$), аммо $n \to \infty$ да шундай k_n-сонлар кетма-кетлиги ($k_n \to \infty$) ва N_0 мусбат тасодифий миқдор мавжуд бўлиб, $\frac{N_n}{K_n} \xrightarrow{p} N_0$ бўлишлиги талаб этилади.

Калит сўзлар: *тасодифий миқдор, U – статистика, асимптотик нормал, боғлиқсизлик, нотасодифий, тасодифий ҳажмли танланма.*

В работе доказана теорема об асимптотической нормальности U – статистики при случайном объеме выборки в случае, когда не предполагается независимость объема выборки N_n от самих результатов наблюдения ξ_i при любом i, причем предполагается, что $\frac{N_n}{K_n} \xrightarrow{p} N_0$ при $n \to \infty$, где k_n последовательность чисел такая, что $k_n \to \infty$ при $n \to \infty$ и N_0 - положительная с.в.

Ключевые слова: *случайная величина, U – статистики, асимптотически нормально, независимость, детерминированный, случайный объем выборки.*

In this article the theorem about asymptotic normality of statistics – U built when the number of sample is N_n random quantity. In case when the independence of the size of sample N_n is not assumed from results of observation ξ_i in any i, though $\frac{N_n}{K_n} \xrightarrow{p} N_0$ is assumed in $n \to \infty$, where k_n order of numbers $k_n \to \infty$ in $n \to \infty$ and N_0 is positive random variables.

Key words: *random quantity, statistics – U, asymptotic normality, independence, determined, random size of sample.*

Пусть $\{\xi_i\}$ – последовательность независимых, одинаково распределенных случайных величин (с.в.). Образуем статистику вида:

$U_n = (C_n^k)^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} f(\xi_{i_1}, \dots, \xi_{i_k}) = (C_n^k)^{-1} \sum_{(n,k)} f(\xi_{i_1}, \dots, \xi_{i_k})$, которые принято называть U – статистиками. Здесь $f(\xi_{i_1}, \dots, \xi_{i_k})$ – некоторая функция, симметричная относительно своих аргументов. Обозначим

$$\theta = Mf(\xi_1, \dots, \xi_k), \qquad f_m(x_1, \dots, x_m) = Mf(x_1, \dots, x_m, \xi_{m+1}, \dots, \xi_k)$$

$$(m \le k), \quad l_2 = Df_1(\xi_1).$$

U – статистики впервые ввел В. Хефдинг [1] и при условии

$$(\text{X}): \qquad Mf^2(\xi_1, \dots, \xi_k) < \infty, \quad l_2 \neq 0$$

доказал, что последовательность с.в.

$Z_n = \frac{\sqrt{n}}{k\sqrt{l_2}}(U_n - \theta)$ – асимптотически нормально с параметрами (0,1).

Далее интенсивно изучались различные свойства U – статистики при детерминированном объеме выборки. При случайном объеме выборки предельное поведение U – статистики рассматривалось в работе К. Атила [2], Р. Спроула [3] и др.

Пусть $\{N_n\}$ – последовательность положительных целочисленных с.в. заданные на вероятностном пространстве $\{\Omega, \mathcal{F}, P\}$, где и задана последовательность с.в. $\{\xi_n\}$.

В работе [2] доказано, что если при $n \to \infty$ последовательность с.в. $\left\{\frac{N_n}{n}\right\}$ стремится к единице по вероятности, то

$$P(Z_{N_n} < x) \to \Phi(x) = \int_{-\infty}^{x} e^{-u^2/2} du \qquad (1)$$

В этой работе также получена скорость сходимости в предельном соотношении (1). В [3] соотношение (1) доказано при условии

$$\frac{N_n}{n} \xrightarrow{p} N_0 \qquad (n \to \infty),$$

где $N_0 > 0$ – дискретная с.в.

Настоящая работа посвящена доказательству одной теоремы об асимтотическом поведении U – статистики при случайном объеме выборки. Доказанная теорема обобщает результаты работы [3] на случай, когда N_0 положительная с.в.

Теорема. Пусть выполнены условия (X) и

(A_0): Существуют последовательность чисел $\{k_n\}$, $k_n \to \infty$ при $n \to \infty$ и положительная с.в. N_0 такие, что $\frac{N_n}{k_n} \xrightarrow{p} N_0$ при $n \to \infty$. Тогда для любого $A \in \mathcal{F}$, $P(A) > 0$

$$P(Z_{N_n} < x/A) \to \Phi(x) \text{ при } n \to \infty.$$

Заметим, что при k = 1 и A = Ω теорема совпадает с теоремами Й. Модьороди [5] и Й Блюм, Д.Хансон, Й.Розенбблата [6].

Сначала докажем следующую лемму.

Лемма. Если $Mf^2(\xi_1, \dots, \xi_k) < \infty$, то для любого h $(h = 1,2, \dots, k)$

$$\delta_h = Dg^{(h)}(\xi_1, \dots, \xi_h) < \infty,$$

где $g^{(1)}(x_1) = f_1(x_1) - \theta$,

$$g^{(h)}(x_1, \dots, x_h) = f_h(x_1, \dots, x_h) - \theta - \sum_{j=1}^{h-1} \sum_{(h,j)} g^{(j)}(x_1, \dots, x_j), (h = 2, \dots, k).$$

Доказательство леммы. В силу теоремы В. Хефдинга [1] для каждого $i = \overline{1,k}$

$$M(f_i(\xi_1, \dots, \xi_k) - \theta)^2 = \zeta_i < \infty \qquad (2)$$

читывая симметричность функции $f(\xi_1, \dots, \xi_k)$ в [1] было доказано, что

$$M\left(f_i(\xi_{\alpha_1}, \dots, \xi_{\alpha_i}) - \theta\right)\left(f_j\left(\xi_{\beta_1}, \dots, \xi_{\beta_j}\right) - \theta\right) = \zeta_l, \qquad (3)$$

где l - число одинаковых индексов среди $(\alpha_1, \dots, \alpha_i)$ и $(\beta_1, \dots, \beta_j)$. . Так как

$Mg^{(h)}(\xi_1, \dots, \xi_h) = 0$, то из (3) вытекает, что δ_n можно представить в виде линейной комбинации величин $\zeta_1, \dots, \zeta_h$·. Отсюда и из (2) получим, что $\delta_n < \infty$.

Далее для удобства ссылки приведем одну теорему из [1] и некоторые известные леммы.

Теорема [1]. Пусть $Mf^2(\xi_1, \dots, \xi_k) < \infty$. Тогда

1) имеет место следующее разложение

$$U_n = \theta + \sum_{h=1}^{k} C_k^h V_n^{(h)} = \theta + kV_n^{(1)} + R_n,$$

где

$$V_n^{(h)} = (C_n^h)^{-1} \sum_{(n,h)} g^{(h)}(\xi_1, \dots, \xi_h), \; R_n = \sum_{h=2}^{k} C_k^h V_n^{(h)};$$

2) для каждого $h = \overline{1,k}$ последовательность с.в.

$S_n^{(h)} = C_n^h V_n^{(h)}$ является мартингалом.

Лемма 1 ([5]). Пусть $\{\mu_n\}$ – последовательность с.в. и μ_0 -некоторая с.в. такие,что

$$\mu_n \xrightarrow{p} \mu_0 \qquad \text{при} \qquad n \to \infty.$$

Пусть a, b $(a < b)$ есть точки непрерывности ф.р. $P(\mu_0 < x)$. . Положим

$$A_n = (a \le \mu_n < b), \qquad A_0 = (a \le \mu_0 < b).$$

Тогда при $n \to \infty$

$$P(\bar{A}_n A_0 + A_n \bar{A}_0) \to 0.$$

Лемма 2 ([6]**).** Пусть $\{\mu n\}$ – последовательность н.о.р.с.в. и A_n – некоторое событие, зависящее от

$$\text{с.в. } \mu_{k_n}, \mu_{k_n+1}, \dots, \mu_{m_n}, \qquad (m_n \ge k_n).$$

Положим $k_n \to \infty$, $m_n \to \infty$ при $n \to \infty$.

Тогда для любого $A \in \mathcal{F}$, $P(A) > 0$

$$P(A_n A) - P(A_n)P(A) \to 0 \qquad \text{при} \quad n \to \infty.$$

Лемма 3 ([7]**).** Пусть $\{\mu n\}$ - последовательность с.в. и μ_0 - некоторая положительная с.в. такие,что при $n \to \infty$

$$P(\mu_n < x) \to P(\mu_0 < x)$$

в каждой точке x, являющейся точкой непрерывности ф.р. $P(\mu_0 < x)$.. Тогда для любого $\varepsilon > 0$ существуют точки $0 < a < b < \infty$ и $n(\varepsilon)$ такие, что при всех $n > n(\varepsilon)$

$$P(a \le \mu_n < b) > 1 - \varepsilon.$$

Доказательство теоремы. Не ограничивая общности, предположим, что $k_n = n$. Для любого $\varepsilon > 0$ выберем числа

$$0 < a = b_1 < b_2 < \cdots < b_{m-1} < b_m = b < \infty$$

из множества точек непрерывности ф.р. $P(N_0 < x)$, так, чтобы

$$P(a \le N_0 < b) \ge 1 - \frac{\varepsilon}{3}$$

и

$$\max_{1 \le i \le m} |b_i - b_{i-1}| = \varepsilon_m \to 0 \text{ при } m \to \infty.$$

Обозначим

$$A_0 = (a \le N_0 < b), \qquad A_0^{(j)} = (b_{j-1} \le N_0 < b_j),$$

$$A_n = (an \le N < bn), \qquad A_n^{(j)} = (nb_{j-1} \le N < nb_j),$$

где $N = N_n, n_j = [nb_j]$, $[x]$ – целая часть числа x. После несложных преобразований получим

$$Z_N = Z_{n_{j-1}} + \sqrt{N}\left(U_N - U_{n_{j-1}}\right)\frac{1}{k\sqrt{l_2}} + Z_{n_{j-1}}\left(\sqrt{\frac{N}{n_{j-1}}} - 1\right) =$$
$$= Z_{n_{j-1}} + \eta_{Nj} + \gamma_{Nj}. \qquad (3)$$

Положим

$$B_{n1} = (\max_{n_{j-1}\le\alpha\le n_j}|\eta_{\alpha j}| \le \varepsilon), \quad B_{n2} = (\max_{n_{j-1}\le\alpha\le n_j}|\gamma_{\alpha j}| \le \varepsilon).$$

Тогда из равенства (4) и принятых обозначений, нетрудно убедиться в следующем

$$T_n^- \le P_A(Z_N < x) \le T_n^+ + P_A(\bar{A}_n), \qquad (5)$$

где

$$T_n^{\pm} = \sum_{j=1}^m P_A(Z_{n_{j-1}} < x \pm 2\varepsilon, A_n^{(j)}) \pm \sum_{j=1}^m P_A(\bar{B}_{n1}, A_n^{(j)}) \pm \sum_{j=1}^m P_A(\bar{B}_{n2}, A_n^{(j)}) =$$

$$= \Sigma^{\pm} \pm \Sigma_1 \pm \Sigma_2, \text{ и } \quad P_A(\cdot) = P(\cdot/A).$$

В силу леммы 1 существует число $n_0(\varepsilon)$ такое, что при $n > n_0(\varepsilon)$

$$\left|\Sigma^{\pm} - \sum_{j=1}^m P_A(Z_{n_{j-1}} < x \pm 2\varepsilon, A_0^{(j)})\right| \le \varepsilon \qquad (6)$$

Последовательность с.в. $\{Z_n\}$ обладает свойством R – перемешивания с предельной *ф.р.* $\Phi(x)$ [3]. Следовательно, существует число $n_1(\varepsilon)$ такое, что при $n > n_1(\varepsilon)$

$$\left|\sum_{j=1}^m P_A(Z_{n_{j-1}} < x \pm 2\varepsilon, A_0^{(j)}) - \Phi(x \pm 2\varepsilon)P_A(A_0)\right| \le \varepsilon. \qquad (7)$$

Так как, в силу леммы 3 $P(\bar{A}_n) \le \varepsilon$, то из (5) – (7) для достаточно больших n вытекает следующее неравенство:

$$\Phi(x - 2\varepsilon) - \Sigma_1 - \Sigma_2 - \frac{\varepsilon}{P(A)} - 2\varepsilon \le P_A(Z_N < x) \le$$
$$\le \Phi(x + 2\varepsilon) + \Sigma_1 + \Sigma_2 + \frac{\varepsilon}{P(A)} + 2\varepsilon \qquad (8)$$

Очевидно, что при $n_{j-1} \le N \le n_j$

$$0 < \sqrt{\frac{N}{n_{j-1}}} - 1 \le \varepsilon_m.$$

Теперь, обратившись опять к рассуждениям, использованным в (6) и (7), утверждаем: для заданного $\varepsilon > 0$ существуют числа $n(\varepsilon)$ и $m(\varepsilon)$ такие,что при $n > n(\varepsilon)$, $m > n(\varepsilon)$

$$\Sigma_2 \le \varepsilon \qquad (9)$$

Для оценивания сверху $\sum_1$, воспользуемся теоремой [1], вследствие чего получим:

$$\Sigma_1 \le \sum_{j=1}^m P_A\left(\max_{n_{j-1}\le\alpha\le n_j}\sqrt{\alpha}\left|V_\alpha^{(1)} - V_{n_{j-1}}^{(1)}\right| > \frac{\varepsilon k\sqrt{l_2}}{2}, A_n^{(j)}\right) +$$
$$+ \sum_{j=1}^m P_A\left(\max_{n_{j-1}\le\alpha\le n_j}\sqrt{\alpha}|R_\alpha| > \frac{\varepsilon k\sqrt{l_2}}{2}, A_n^{(j)}\right) = \Sigma_{11} + \Sigma_{12}. \qquad (10)$$

Нетрудно увидеть, что

$$\Sigma_{11} \le \sum_{j=1}^{m} P_A\left(\max_{n_{j-1}\le\alpha\le n_j} \frac{1}{\sqrt{\alpha}}\left|\sum_{i=n_{j-1}}^{\alpha}(f_1(\xi_i)-\theta)\right| \ge \frac{\varepsilon k\sqrt{l_2}}{4}, A_n^{(j)}\right) +$$

$$+ \sum_{j=1}^{m} P_A\left(\frac{n_j-n_{j-1}}{n_{j-1}} \cdot \frac{1}{\sqrt{n_{j-1}}}\left|\sum_{i=1}^{n_{j-1}}(f_1(\xi_i)-\theta)\right| \ge \frac{\varepsilon k\sqrt{l_2}}{4}, A_n^{(j)}\right) =$$

$$= \Sigma_{11}^{(1)} + \Sigma_{11}^{(2)} \qquad (11)$$

В силу леммы 1 в выражениях $\Sigma_{11}^{(1)} + \Sigma_{11}^{(2)}$ событие $A_n^{(j)}$ можно заменить событием $A_0^{(j)}$, т.е. для любого $\varepsilon > 0$ существует число $n_2(\varepsilon)$ такое, что при $n > n_2(\varepsilon)$

$$\left|\Sigma_{11}^{(1)} - \sum_{j=1}^{m} P_A\left(\max_{n_{j-1}\le\alpha\le n_j} \frac{1}{\sqrt{\alpha}}\left|\sum_{i=n_{j-1}}^{\alpha}(f_1(\xi_i)-\theta)\right| \ge \frac{\varepsilon k\sqrt{l_2}}{4}, A_0^{(j)}\right)\right| \le \varepsilon, \qquad (12)$$

$$\left|\Sigma_{11}^{(2)} - \sum_{j=1}^{m} P_A\left(\frac{n_j-n_{j-1}}{n_{j-1}} \cdot \frac{1}{\sqrt{n_{j-1}}}\left|\sum_{i=1}^{n_{j-1}}(f_1(\xi_i)-\theta)\right| \ge \frac{\varepsilon k\sqrt{l_2}}{4}, A_0^{(j)}\right)\right| \le \varepsilon. \qquad (13)$$

Воспользовавшись сначала леммой 2, а затем известным неравенством Колмогорова, приходим к тому, что существуют числа $n_3(\varepsilon)$ и $n_m(\varepsilon)$ такие, что при $n > n_3(\varepsilon)$, $m > n_3(\varepsilon)$

$$\sum_{j=1}^{m} P_A\left(\max_{n_{j-1}\le\alpha\le n_j} \frac{1}{\sqrt{\alpha}}\left|\sum_{i=n_{j-1}}^{\alpha}(f_1(\xi_i)-\theta)\right| \ge \frac{\varepsilon k\sqrt{l_2}}{4}, A_0^{(j)}\right) \le \varepsilon \qquad (14)$$

Так, как выражение $\frac{1}{\sqrt{n_{j-1}}}\sum_{i=1}^{n_{j-1}}(f_1(\xi_i)-\theta)$ обладает свойством R – перемешивания с предельным распределением Ф(х). Поэтому утверждаем, что существуют числа $n_4(\varepsilon)$ и $n_4(\varepsilon)$ такие, что при $n > n_4(\varepsilon)$, $m > m_4(\varepsilon)$

$$\sum_{j=1}^{m} P_A\left(\frac{n_j-n_{j-1}}{n_{j-1}} \cdot \frac{1}{\sqrt{n_{j-1}}}\left|\sum_{i=1}^{n_{j-1}}(f_1(\xi_i)-\theta)\right| \ge \frac{\varepsilon k\sqrt{l_2}}{4}, A_0^{(j)}\right) \le \varepsilon \qquad (15)$$

Из (11)-(15) для достаточно больших n и m получим

$$\Sigma_{11} \le c\varepsilon \qquad (16)$$

(здесь и далее c постоянная не зависящая от n, m и ε).

Нетрудно получить следующее неравенство

$$\Sigma_{12} \le \sum_{h=2}^{k} P_A\left(\max_{na\le\alpha\le nb}\left|S_\alpha^{(h)}\right| \ge \frac{\varepsilon k\sqrt{l_2}C_{[na]}^h}{4(k-1)\sqrt{nb}C_k^h}\right)$$

Так как $DS_n^h = C_n^h\delta_h$ и в силу теоремы [1] S_n^h является мартингалом, то, воспользовавшись неравенством Колмогорова для мартингалов, будем иметь:

$$\Sigma_{12} \le \sum_{h=2}^{k} \frac{16(k-1)^2(C_k^h)^2 nbC_{[nb]}^h\delta_h}{\varepsilon^2k^2l_2\left[C_{[na]}^h\right]^2} = \frac{c\delta_h}{\varepsilon^2 n},$$

Откуда для $n > \frac{c\delta_h}{\varepsilon^3}$ имеем

$$\Sigma_{12} \leq \varepsilon. \tag{17}$$

Из (10). (16) и (17) следует, что достаточно больших n и m

$$\Sigma_1 \leq c\varepsilon. \tag{18}$$

В силу произвольности $\varepsilon > 0$ из (8), (9) и (18) вытекает доказательство теоремы.

Литература:

1. Hoeffding W. The strong law of large numbers for U-statistics.–Inst. Of Stat. Univ. of N. Carolins, Mimeograph Series, N 302,1961.

2. Attila C. A theorem on departure of randomly indexed U – statistics from normality with an application in fixed-width sequential interval estimation.–Sankhya: The Indian J. Stat. v.43 1981, ser.A,1, p. 84-89.

3. Sproule R. Asymptotic properties of U- statistics.–Trans. Amer. Math. Soc.,v199,1974,p.55-64.

4. Robbins H. The asymptotic distribution of the sum of a random number of random variables. -- Bull. Amer. Math. Soc.,54,1948,p.1151-1161.

5. Mogyorodi J. A central limit theorem for the sum of random number of independent random variables.- Publ. Math. Inst. Hung. Acad. Sci., v.7, 1962,p. 409-424.

6. Blum J.R., Hanson D. L. and Rosenblatt J.I. On the central limit theorem for the sum of a random number of independent random variables. – Zeit. Wahr. Vew Gebiet. 1, 1963, p. 389-393.

7. Джамирзаев А.А. Обобщения теоремы А.Я. Хинчина об относительной устойчивости.- Научные труды ТашГУ, вып. 402, 1972, стр. 39-44.

Андижон давлат университети,
Андижон машинасозлик институти
2014 йил 27 февралда қабул қилинган
(Тақризчи: проф. Ғ.Мўминов)