LIMIT THEOREM FOR A STATISTIC PROPOSED BY V. HEFDLING

Abdullaev Abdumukhtor Ganievich
Associate Professor at the non-state higher educational institution "University of Economics and Pedagogy" (Andijan).
Phone: +99893-631-72-36
Email: abdullaev7591@gmail.com
Mamurov Egamnazar Narbaevich
Associate Professor of the Department of "Higher and Applied Mathematics" at the Tashkent University of Economics.
Phone: +99899-619-48-71
Email: imamurov58@gmail.com
https://doi.org/10.5281/zenodo.14049485

Annotation: The article first proves a limit theorem for a sequence of random variables called "U-statistics," introduced by V. Hefdling. The proven limit theorem generalizes the result of V. Hefdling's theorem to the case where the number of samples is a random variable. In the theorem, it is not required that the observation results of the sample size N_n are independent of the ξ_i where $(i = 1,2, \ldots,)$; however, as $n \to \infty$ there must exist a sequence of numbers - k_n with $(k_n \to \infty)$ and a positive random variable N_0 such that, $\frac{N_n}{k_n} \xrightarrow{p} N_0$ is required.

Keywords: Generalized statistics, sample size, normal distribution, independence, random variable, *U*-statistics, sequence of positive integer-valued random variables, distribution function, symmetry with respect to arguments.

Let us assume that $\{\xi_i\}$ – is a sequence of independent identically distributed random variables (r.v.). We form a statistic of the form:

$U_n = (C_n^k)^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} f(\xi_{i_1}, \ldots, \xi_{i_k}) = (C_n^k)^{-1} \sum_{(n,k)} f(\xi_{i_1}, \ldots, \xi_{i_k})$, which is commonly referred to as *U*-statistics. Here, $f(\xi_{i_1}, \ldots, \xi_{i_k})$ – is some function symmetric with respect to its arguments. We denote:

$$\theta = Mf(\xi_1, \ldots, \xi_k), \qquad f_m(x_1, \ldots, x_m) = Mf(x_1, \ldots, x_m, \xi_{m+1}, \ldots, \xi_k)$$
$$(m \le k), \quad l_2 = Df_1(\xi_1).$$

U-statistics were first introduced by W. Hoeffding [1], and under the condition

(H): $\qquad Mf^2(\xi_1, \ldots, \xi_k) < \infty, \quad l_2 \ne 0$

proved that the sequence of random variables

$Z_n = \frac{\sqrt{n}}{k\sqrt{l_2}}(U_n - \theta)$ - is asymptotically normal with parameters (0,1).

Subsequently, various properties of *U*-statistics were intensively studied with a deterministic sample size. The limiting behavior of *U*-statistics with a random sample size was examined in works [2], [3], and others.

Assuming $\{N_n\}$ – be a sequence of positive integer-valued random variables defined on the probability space $\{\Omega, \mathcal{F}, P\}$, where the sequence of random variables $\{\xi_n\}$ is also defined.

In work [2] it was proven that if as $n \to \infty$ the sequence of random variables. $\left\{\frac{N_n}{n}\right\}$ converges to one in probability, then

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

This work also obtained the rate of convergence in the limiting relation (1). In work [3], relation (1) was proven under the condition

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

where $N_0 > 0$ – is a discrete random variable.

The present work is dedicated to proving a theorem about the asymptotic behavior of U-statistics with a random sample size. The proven theorem generalizes the result of work [3] to the case when N_0 – is an arbitrary positive random variable.

Theorem: Let the conditions (H) and (A_0) be satisfied:

There exists a sequence of numbers such that $\{k_n\}$, $k_n \to \infty$ as $n \to \infty$ and a positive random variable N_0 such that $\frac{N_n}{k_n} \xrightarrow{p} N_0$ as $n \to \infty$. then for any $A \in \mathcal{F}$, $P(A) > 0$

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

Note that for $k = 1$ и $A = \Omega$ the theorem coincides with the theorems of Y. Modyorody [5] and Y. Blum, D. Hanson, Y. Rosenblatt [6].

First, we will prove the following lemma.

Lemma. If $Mf^2(\xi_1, \dots, \xi_k) < \infty$, then for any h $(h = 1,2, \dots, k)$

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

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

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

Proof of the lemma. By the theorem proven in work [1] for each $i = \overleftarrow{1,k}$ it holds that

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

Considering the symmetry of the function $f(\xi_1, \dots, \xi_k)$ it was proven in [1] that

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

where l - is the number of common indices among $(\alpha_1, \dots, \alpha_i)$ and $(\beta_1, \dots, \beta_j)$. Since $Mg^{(h)}(\xi_1, \dots, \xi_h) = 0$, it follows from (3) that δ_h can be expressed as a linear combination of the quantities $\zeta_1, \dots, \zeta_h$. Hence, from (2) we obtain that $\delta_h < \infty$.

Next, for convenience, we will state one theorem from [1] and some known lemmas.

Theorem [**1**]. Let's assume that $Mf^2(\xi_1, \dots, \xi_k) < \infty$. then

1) the following decomposition holds:

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

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

2) for each $h = \overleftarrow{1,k}$ the sequence of random variables $S_n^{(h)} = C_n^h V_n^{(h)}$ is a martingale.

Lemma 1 ([5]**).** Let $\{\mu_n\}$ - be a sequence of r.v. and μ_0 - be some r.v. such that

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

Let a, b $(a < b)$ be points of continuity of the c.d.f. $P(\mu_0 < x)$. Let's assume that

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

then as $n \to \infty$

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

IBAST | Volume 4, Issue 11, November

Lemma 2 ([6]). Let $\{\mu_n\}$ - be a sequence of non-degenerate r.v. and A_n– be some event depending on the r.v. $\mu_{k_n}, \mu_{k_n+1}, \dots, \mu_{m_n}, \quad (m_n \geq k_n)$.

Assume that $k_n \to \infty$, as $n \to \infty$.

Then for any $A \in \mathcal{F}$, $P(A) > 0$

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

Lemma 3 ([7]). Let $\{\mu_n\}$ - be a sequence of r.v. and μ_0 - be some positive r.v. such that as $n \to \infty$

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

at every point xxx that is a point of continuity of the c.d.f. $P(\mu_0 < x)$. Then for any $\varepsilon > 0$ there exist points $0 < a < b < \infty$ and $n(\varepsilon)$ such that for all $n > n(\varepsilon)$

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

Proof of the theorem: Without loss of generality, assume that $k_n = n$. For any $\varepsilon > 0$ we choose numbers

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

from the set of continuity points of the c.d.f. $P(N_0 < x)$, such that

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

and

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

Let us denote

$$A_0 = (a \leq N_0 < b), \qquad A_0^{(j)} = (b_{j-1} \leq N_0 < b_j),$$
$$A_n = (an \leq N < bn), \qquad A_n^{(j)} = (nb_{j-1} \leq N < nb_j),$$

where $N = N_n, n_j = [nb_j]$, $[x]$ – is the integer part of x. After some straightforward transformations, we obtain

$$Z_N = Z_{n_{j-1}} + \sqrt{N}\left(U_N - U_{n_{j-1}}\right)\frac{1}{k\sqrt{l_2}} + Z_{n_{j-1}}\left(\sqrt{\frac{N}{n_{j-1}}} - 1\right) =$$
$$= Z_{n_{j-1}} + \eta_{Nj} + \gamma_{Nj}. \qquad (4)$$

Let us assume that

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

Then from equation (4) and the definitions used, it is easy to verify the following:

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

where

$$T_n^{\pm} = \sum_{j=1}^m P_A(Z_{n_{j-1}} < x \pm 2\varepsilon, A_n^{(j)}) \pm \sum_{j=1}^m P_A(\bar{B}_{n1}, A_n^{(j)}) \pm \sum_{j=1}^m P_A(\bar{B}_{n2}, A_n^{(j)}) =$$
$$= \Sigma^{\pm} \pm \Sigma_1 \pm \Sigma_2, \text{ and } \quad P_A(\cdot) = P(\cdot / A).$$

By Lemma 1, there exists a number $n_0(\varepsilon)$ such that for $n > n_0(\varepsilon)$

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

The sequence of r.v. $\{Z_n\}$ possesses the property of R– mixing with a limiting distribution $\Phi(x)$ [3]. Subsequently, there exists a number $n_1(\varepsilon)$ such that for $n > n_1(\varepsilon)$

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

Since, due to Lemma 3 $(\bar{A}_n) \leq \varepsilon$, then from (5) – (7) for sufficiently large n we have the following inequality:

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

Clearly, for $n_{j-1}\le N\le n_j$

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

Now, returning to the reasoning used in (6) and (7), we assert that for a given $\varepsilon>0$ there exist numbers $n(\varepsilon)$ and $m(\varepsilon)$ such that for $n>n(\varepsilon)$, $m>m(\varepsilon)$

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

To estimate Σ_1 from above, we use Theorem [1], resulting in:

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

It is clearly visible that

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

By Lemma 1 in the expressions $\Sigma_{11}^{(1)}+\Sigma_{11}^{(2)}$ the event $A_n^{(j)}$ can be replaced with the event $A_0^{(j)}$, that is to say, for any $\varepsilon>0$ there exists a number $n_2(\varepsilon)$ such that when $n>n_2(\varepsilon)$

$$\left|\Sigma_{11}^{(1)}-\sum_{j=1}^m P_A\left(\max_{n_{j-1}\le\alpha\le n_j}\frac{1}{\sqrt{\alpha}}\left|\sum_{i=n_{j-1}}^{\alpha}(f_1(\xi_i)-\theta)\right|\ge\frac{\varepsilon k\sqrt{l_2}}{4},A_0^{(j)}\right)\right|\le\varepsilon, \quad (12)$$
$$\left|\Sigma_{11}^{(2)}-\sum_{j=1}^m P_A\left(\frac{n_j-n_{j-1}}{n_{j-1}}\cdot\frac{1}{\sqrt{n_{j-1}}}\left|\sum_{i=1}^{n_{j-1}}(f_1(\xi_i)-\theta)\right|\ge\frac{\varepsilon k\sqrt{l_2}}{4},A_0^{(j)}\right)\right|\le\varepsilon. \quad (13)$$

Using initially Lemma 2 and then Kolmogorov's inequality, we find numbers $n_3(\varepsilon)$ and $m_3(\varepsilon)$ such that for $n>n_3(\varepsilon)$, $m>m_3(\varepsilon)$

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

Since the expression $\frac{1}{\sqrt{n_{j-1}}}\sum_{i=1}^{n_{j-1}}(f_1(\xi_i)-\theta)$ has the property of $R-$ mixing with a limiting distribution (x). we assert that there exist numbers $n_4(\varepsilon)$ and $m_4(\varepsilon)$ such that for $n>n_4(\varepsilon)$, $m>m_4(\varepsilon)$

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

From (11)-(15) for sufficiently large n and m, we obtain

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

(Here and further, c is a constant that does not depend on n , m и ε).

It is not difficult to obtain the following inequality:

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

Since $DS_n^h=C_n^h\delta_h$ and due to Theorem [1] S_n^h is a martingale, applying Kolmogorov's inequality for martingales gives us:

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

whence for $n > \frac{c\delta_h}{\varepsilon^3}$ we have

$$\Sigma_{12} \leq \varepsilon \ . \qquad (17)$$

From (10), (16) and (17) it follows that for sufficiently large n and m

$$\Sigma_1 \leq c\varepsilon \ . \qquad (18)$$

Due to the arbitrariness of $\varepsilon > 0$ from (8), (9) and (18) we conclude the proof of the theorem.

Following the work of [8], theorems are stated in which the existence of a limiting distribution for a deterministic sequence is assumed, and under corresponding additional conditions, the existence of a limiting distribution for sequences with a random index is asserted. We will call these transfer theorems. Works [9] and [10] are devoted to the study of transfer theorems for sequences of terms in the variational series in cases where independence between the random index and the original sequence of random variables is not assumed (the so-called "dependent scheme"). The theorem we proved is a transfer theorem for U-statistics in the case of a "dependent scheme."

References:

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. Jamirzaev A.A. Generalizations of A.Ya. Khinchin's Theorem on Relative Stability. - Scientific Papers of Tashkent State University, vol. 402, 1972, pp. 39-44.
8. Gnedенко B.V., Fakham G. On a Transfer Theorem // Proceedings of the USSR Academy of Sciences. 1969. Vol. 187. No. 1. pp. 15-17.
9. Jamirzaev A.A., Mamurov I.N. Transfer Theorems. Monograph, Tashkent “Iqtisod-Moliya” 2019, 148 pp.
10. Mamurov I. Asymptotic distribution of the central variation members in the case of random sampling volume. Turkish Online Journal of Qualitative Inquiry (TOJQI). Volume 12, Issue 7, July 2021: 4626-4634.