

СИБИРСКИЕ ЭЛЕКТРОННЫЕ МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports

<http://semr.math.nsc.ru>

Том 17, стр. 672–682 (2020)

УДК 519.214

DOI 10.33048/semi.2020.17.045

MSC 60F05, 05C90, 94C15

NORMAL APPROXIMATION FOR U - AND V -STATISTICS OF A STATIONARY ABSOLUTELY REGULAR SEQUENCE

V.G. MIKHAILOV AND N.M. MEZHENNAYA

ABSTRACT. Let $(X_{n,t})_{t=1}^{\infty}$ be a stationary absolutely regular sequence of real random variables with the distribution dependent on the number n . The paper presents sufficient conditions for the asymptotic normality (for $n \rightarrow \infty$ and common centering and normalization) of the distribution of the nonhomogeneous U -statistic of order r which is given on the sequence $X_{n,1}, \dots, X_{n,n}$ with a kernel also dependent on n . The same results for V -statistics also hold. To analyze sums of dependent random variables with rare strong dependencies, the proof uses the approach that was proposed by S. Janson in 1988 and upgraded by V. Mikhailov in 1991 and M. Tikhomirova and V. Chistyakov in 2015.

Keywords: absolute regularity condition, characterizing graph, central limit theorem, dependency graph, U -statistic, V -statistic, stationary sequence.

INTRODUCTION

The study of a special class of functionals of a sequence of random variables X_1, \dots, X_n of the form

$$(1) \quad \hat{U}_n = \frac{1}{C_n^m} \sum_{1 \leq i_1 < \dots < i_m \leq n} f(X_{i_1}, \dots, X_{i_m}),$$

which were called U -statistics, began in the middle of the last century due to the investigation of the properties of sample characteristics (see [1] and the bibliography therein). The number m is the *order* and symmetrical function $f(x_1, \dots, x_m)$ is the

MIKHAILOV, V.G., MEZHENNAYA, N.M., NORMAL APPROXIMATION FOR U - AND V -STATISTICS OF A STATIONARY ABSOLUTELY REGULAR SEQUENCE.

© 2020 MIKHAILOV V.G., MEZHENNAYA N.M.

Received October, 16, 2019, published May, 8, 2020.

kernel of U -statistic. Examples of U -statistics are sample moments, Gini's mean difference, Spearman's rank correlation, etc.

It is known that such variables as the number of repetitions and the number of repetitions of tuples [2]–[5], the number of pairs of H -equivalent tuples [6]–[11], etc., in the random discrete sequence X_1, \dots, X_n belong (up to of the factor before the sum) to a class of quantities of the form (1).

In the following decades, a large number of research papers appeared devoted to the asymptotic properties of U -statistics of sequences of independent identically distributed random variables (see, e.g., the bibliography in [12]). In proving asymptotic normality in the case of increasing sums, W. Hoeffding [1] proposed a method for approximating the distribution of U -statistic by distribution of the sum of specially constructed independent random variables. This approach, in different forms, is also used to study U -statistics of sequences of random variables with conditions of weak or other dependence in the scheme of increasing sums (see, e.g., [13]–[16]).

The idea of the results of these papers is given by the following theorem of K. Yoshihara [13] which we present in a simplified form.

Let $(X_n)_{n=1}^\infty$ be a strictly stationary sequence of real random variables satisfying the *absolute regularity condition* [17, p. 3]

$$(2) \quad \beta(n) = \mathbf{E} \left\{ \sup_{A \in \mathcal{F}_n^\infty} |\mathbf{P}\{A|\mathcal{F}_{-\infty}^0\} - \mathbf{P}\{A\}| \right\} \downarrow 0, \quad n \rightarrow \infty,$$

where \mathcal{F}_a^b is the σ -algebra of events generated by the random variables X_a, \dots, X_b .

Let

$$\theta = \mathbf{E}f(\tilde{X}_1, \dots, \tilde{X}_m), \quad f_1(x) = \mathbf{E}f(x, \tilde{X}_2, \dots, \tilde{X}_m),$$

where $\tilde{X}_1, \dots, \tilde{X}_m$ are independent copies of X_1 ,

$$\sigma^2 = (\mathbf{E}f_1^2(X_1) - \theta^2) + 2 \sum_{t=1}^{\infty} (\mathbf{E}(f_1(X_1)f_1(X_{t+1})) - \theta^2).$$

Theorem 1 (Theorem 1 of [13]). *Let $n \rightarrow \infty$ and there is a number $\delta > 0$ such that*

$$\begin{aligned} \mathbf{E} \left| f(\tilde{X}_1, \dots, \tilde{X}_m) \right|^{2+\delta} &\leq M_0 < \infty, \\ \mathbf{E} |f(X_{i_1}, \dots, X_{i_m})|^{2+\delta} &\leq M_0 < \infty \quad \forall 1 \leq i_1 < \dots < i_m \leq n, \\ \beta(n) &= O \left(n^{-(2+\delta')/\delta'} \right) \quad \text{for some } \delta', 0 < \delta' < \delta. \end{aligned}$$

Then, if $\sigma^2 > 0$ holds, the distribution function of the random variable $\frac{\sqrt{n}}{m\sigma}(\hat{U}_n - \theta)$ converges to the distribution function of the standard normal law.

The research paper [18](see also [19, 20]) was devoted to adaptation the results of K. Yoshihara to triangular array schemes. Sh. Khashimov in [18] considered the case of second-order U -statistic whose kernel $f(x_1, x_2) = f_n(x_1, x_2)$ can change for $n \rightarrow \infty$. Again, the method of W. Hoeffding [1] was used.

The method of moments was no less promising for studying U -statistics in the triangular array scheme for dependent random variables. Back in 1975, V. Mikhailov [21], using the direct application of this method, derived sufficient conditions for asymptotic normality for a special case of U -statistics of a sequence

of finitely dependent random variables in a triangular array scheme (let's call it *the wide triangular array scheme*), where for $n \rightarrow \infty$ changes are allowed both to the kernel $f_n(x_1, \dots, x_m)$ and the distribution of the sequence $X_{n,1}, X_{n,2}, \dots$ (now in the notation we have to indicate dependence of the kernel and distribution on n).

A modern variation of the method of moments which was proposed by Svante Janson [22, 23] and upgraded by V. Mikhailov in [23] and by M. Tikhomirova and V. Chistyakov in [24] allows to obtain simpler and substantially more general sufficient conditions for the asymptotic normality of U - and V -statistics of any order of a sequence of random variables satisfying the absolute regularity condition in the wide triangular array scheme which we present in this paper. These results complement the results of K. Yoshihara [13] and V. Mikhailov [21].

It also should be noted that for problems related to tuples in a discrete random sequence (see, e.g., [8, 9, 25, 26]), the present results allow to consider the case of simultaneous consistent growth of the length of the random sequence n and the length of the tuple s to infinity. A separate work is supposed to be devoted to these applications.

1. LIMIT THEOREMS

Let us formulate the following conditions:

Condition A. Let $(X_{n,t})_{t=1}^\infty$ for every $n = 1, 2, \dots$ be a strictly stationary sequence of real random variables (e.g., the joint probability distribution function of $(X_{n,t_k+\tau})_{k=1}^m$ is equal to the joint probability distribution function of $(X_{n,t_k})_{k=1}^m$ for all $\tau, t_1, \dots, t_m = 1, 2, \dots, \infty$ and all $m \geq 1$) satisfying the absolute regularity condition (2).

Condition B. Let $f_{n;j_1, \dots, j_r} : \mathbb{R}^r \rightarrow \mathbb{R}$ be a bounded measurable function for every $n \geq 1$ and $1 \leq j_1, \dots, j_r \leq n$:

$$|f_{n;j_1, \dots, j_r}(x_{j_1}, \dots, x_{j_r})| \leq F_n < \infty.$$

The functionals called the *nonhomogeneous U-statistic* and *V-statistic* with the kernel $f_{n;j_1, \dots, j_r}$ are given by the formulas (the definitions are given in [1] or [12]):

$$(3) \quad U_n = U_n(X_{n,1}, \dots, X_{n,n}) = \sum_{1 \leq j_1 < \dots < j_r \leq n} f_{n;j_1, \dots, j_r}(X_{n,j_1}, \dots, X_{n,j_r}),$$

$$(4) \quad V_n = V_n(X_{n,1}, \dots, X_{n,n}) = \sum_{j_1, \dots, j_r=1}^n f_{n;j_1, \dots, j_r}(X_{n,j_1}, \dots, X_{n,j_r}),$$

respectively (in contrast to the traditional definition (1), the factors $1/C_n^r$ and $1/n^r$ are omitted before the sums).

Let

$$U_n^* = \frac{U_n - \mathbf{E}U_n}{\sqrt{\mathbf{D}U_n}}, \quad V_n^* = \frac{V_n - \mathbf{E}V_n}{\sqrt{\mathbf{D}V_n}}.$$

Theorem 2. *Let Condition A and Condition B be satisfied. For $n \rightarrow \infty$, let the distribution of $(X_{n,t})_{t=1}^\infty$ and the other parameters marked by index n vary so that there exist the sequence of natural numbers $(m_n)_{n=1}^\infty$ and the real number $b_0 \in (0, 2/3]$ such that for all $b \in (0, b_0]$*

$$(5) \quad \frac{F_n^2 m_n^{2-b} n^{2(r-1)+b}}{\mathbf{D}U_n} + (\beta_n(m_n))^b \frac{F_n^2 n^{2r}}{\mathbf{D}U_n} \rightarrow 0.$$

Then the moments and distribution function of the random variable U_n^* converge to the moments and distribution function of the standard normal law.

A similar statement holds for V -statistics.

Theorem 3. *Let Condition A and Condition B be satisfied. For $n \rightarrow \infty$, let the distribution of $(X_{n,t})_{t=1}^\infty$ and the other parameters marked by index n vary so that there exist the sequence of natural numbers $(m_n)_{n=1}^\infty$ and the real number $b_0 \in (0, 2/3]$ such that for all $b \in (0, b_0]$*

$$(6) \quad \frac{F_n^2 m_n^{2-b} n^{2(r-1)+b}}{\mathbf{D}V_n} + (\beta_n(m_n))^b \frac{F_n^2 n^{2r}}{\mathbf{D}V_n} \rightarrow 0.$$

Then the moments and distribution function of the random variable V_n^* converge to the moments and distribution function of the standard normal law.

We consider a special case in which the absolute regularity coefficient (2) decreases faster than any degree of t for $t \rightarrow \infty$.

Theorem 4. *Let Condition A and Condition B be satisfied. Let the number F_n be fixed starting at some value of n , $n \rightarrow \infty$, and the distribution of $(X_{n,t})_{t=1}^\infty$ and the other parameters marked by index n vary so that*

$$\beta_n(t) \leq t^{-h(t)}, \quad \mathbf{D}U_n \geq Cn^{2(r-1)+\varkappa}, \quad C, \varkappa > 0,$$

where the positive function $h(t) \rightarrow \infty$ for $t \rightarrow \infty$. Then the moments and distribution function of the random variable U_n^* converge to the moments and distribution function of the standard normal law.

Theorem 5. *Let Condition A and Condition B be satisfied. Let the number F_n be fixed starting at some value of n , $n \rightarrow \infty$, and the distribution of $(X_{n,t})_{t=1}^\infty$ and the other parameters marked by index n vary so that*

$$\beta_n(t) \leq t^{-h(t)}, \quad \mathbf{D}V_n \geq Cn^{2(r-1)+\varkappa}, \quad C, \varkappa > 0,$$

where the positive function $h(t) \rightarrow \infty$ for $t \rightarrow \infty$. Then the moments and distribution function of the random variable V_n^* converge to the moments and distribution function of the standard normal law.

2. ON THE METHOD OF PROVING OF LIMIT THEOREMS

In 1988, Svante Janson [22] proposed a simple technique for deriving sufficient conditions for the asymptotic normality of bounded random variables Y_1, \dots, Y_T with a joint distribution described by the *dependency graph*. Only one vertex in the dependency graph corresponds to each random variable Y_i , and these vertices are connected by a set of edges. The following condition is satisfied:

if \mathcal{V}_1 and \mathcal{V}_2 are two disjoint subsets of graph vertices such that no edge of the graph has one endpoint in \mathcal{V}_1 and the other in \mathcal{V}_2 , then the sets of random variables $\{Y_i, i \in \mathcal{V}_1\}$ and $\{Y_i, i \in \mathcal{V}_2\}$ are independent.

In [23], this approach was presented more generally and was subsequently used in numerous research studies devoted to the study of asymptotic distributions of functionals depending on a sequence of independent random variables ([4, 5, 10, 11]). Finally, in 2015, M. Tikhomirova and V. Chistyakov [24] proposed a modification of the method [22] and [23] which is applicable to the families of random variables with a complete dependency graph, but the majority of dependencies between the variables are weak.

Remark. A year after [24], the paper [27] appeared on the site arXiv.org and was devoted to transferring the approach by [22] and [23] to the case in which the dependency graph is a complete graph, but the majority of dependencies are weak. In [27], the joint distribution of a set of quantities is described by a weighted dependency graph in which each edge is assigned a numerical characteristic (weight), which describes the degree of dependence between adjacent variables in a certain way. The form of asymptotic normality conditions in [27] resembles similar conditions of [23], but the values included in the conditions are now determined by the weighted dependency graph. The results of applying the conditions by [27] to specific problems are presented in [27] and [28].

We present the main result of M. Tikhomirova and V. Chistyakov [24]. We assume that the joint distribution of the variables Y_1, \dots, Y_T , $|Y_i| \leq F < \infty$, is determined by the *characterizing* graph Γ . This is an undirected graph with the set of vertices $\mathcal{V} = \{1, \dots, T\}$ and the following properties:

- 1) the graph Γ contains the loop at every vertex;
- 2) for any natural number R , any subset $\mathcal{V}' \subset \mathcal{V}$, $|\mathcal{V}'| \leq R$, and any of its partition $\{\mathcal{V}_1, \mathcal{V}_2\}$ (i.e., $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$, $\mathcal{V}_1 \cup \mathcal{V}_2 = \mathcal{V}'$) such that there are no edges connecting vertices from \mathcal{V}_1 with vertices from \mathcal{V}_2 , there is a number $\gamma_R \in (0, 1)$ such that

$$(7) \quad \left| \mathbf{E} \left(\prod_{t \in \mathcal{V}'} Y_t \right) - \mathbf{E} \left(\prod_{t \in \mathcal{V}_1} Y_t \right) \mathbf{E} \left(\prod_{t \in \mathcal{V}_2} Y_t \right) \right| \leq \gamma_R F^{|\mathcal{V}'|}.$$

For any subset $\mathcal{V}' \subset \mathcal{V}$, we define its set of strong dependencies $L(\mathcal{V}')$ as the set of those vertices of Γ that are connected by the edges with vertices from the set \mathcal{V}' . Let $\mathcal{F}(\mathcal{V}') = \sigma\{Y_t, t \in \mathcal{V}'\}$ be σ -algebra generated by the random variables $\{Y_t, t \in \mathcal{V}'\}$, $M = \sum_{t=1}^T \mathbf{E}|Y_t|$, and

$$(8) \quad Q_R = \max_{\mathcal{V}' \subset \mathcal{V}: |\mathcal{V}'| \leq R} \sum_{t \in L(\mathcal{V}')} \mathbf{E}(|Y_t| | \mathcal{F}(\mathcal{V}')), \quad R = 1, 2, \dots$$

We put

$$S_T = \sum_{t=1}^T Y_t, \quad S_T^* = (S_T - \mathbf{E}S_T) / \sqrt{\mathbf{D}S_T}.$$

We suppose that the joint distribution of the random variables Y_1, \dots, Y_T depends on the natural number n assumed as a parameter. All characteristics mentioned above also depend on n : $T = T_n$, $F = F_n$, $\gamma_R = \gamma_{R,n}$, $M = M_n$, $Q_R = Q_{R,n}$, etc.

Theorem 6 (Theorem 1 from [24]). *For $n \rightarrow \infty$, let the numbers $T_n \rightarrow \infty$, the joint distribution of the random variables Y_1, \dots, Y_{T_n} , and the other parameters marked by index n vary so that $b_0 \in (0, 2/3]$ exists such that for all $b \in (0, b_0]$ and any natural number R*

$$(9) \quad \frac{M_n^b (Q_{R,n})^{2-b}}{\mathbf{D}S_{T_n}} + \gamma_{R,n}^b \frac{(F_n T_n)^2}{\mathbf{D}S_{T_n}} \rightarrow 0.$$

Then the moments and distribution function of the random variable $S_{T_n}^$ converge to the moments and distribution function of the standard normal law.*

3. PROOFS OF THEOREMS 2 AND 4 FOR U -STATISTICS

The number of summands $w_n(j_1, \dots, j_r) = f_{n;j_1, \dots, j_r}(X_{n,j_1}, \dots, X_{n,j_r})$ in (3) is $T_n = C_n^r$ (we recall that C_n^r denotes Binomial coefficient). We construct a characterizing graph for the family of random variables $w_n(j_1, \dots, j_r)$ as follows. We define some positive integer number m and denote the characterizing graph by $\Gamma_{n,m}$.

Each variable $w_n(j_1, \dots, j_r)$ is one-to-one assigned to the vertex $\alpha = (j_1, \dots, j_r)$ in the graph $\Gamma_{n,m}$. We denote the set of vertices by $\mathcal{V}(\Gamma_{n,m})$. The total number of vertices in the graph $\Gamma_{n,m}$ is $T_n = C_n^r$.

The set of edges is defined as follows:

- 1) the graph $\Gamma_{n,m}$ contains the loop at every vertex;
- 2) the vertices $\alpha = (j_1, \dots, j_r)$ and $\tilde{\alpha} = (\tilde{j}_1, \dots, \tilde{j}_r)$, $\alpha, \tilde{\alpha} \in \mathcal{V}(\Gamma_{n,m}), \alpha \neq \tilde{\alpha}$, are connected by the edge from $\Gamma_{n,m}$ if and only if at least one of inequalities holds:

$$(10) \quad |j_k - \tilde{j}_l| \leq m, \quad k, l = 1, \dots, r.$$

We show that $\Gamma_{n,m}$ satisfies the property (7).

Lemma 1. *Let conditions A and B be satisfied. Then for any set $\mathcal{V}' \subseteq \mathcal{V}(\Gamma_{n,m})$ and its partitions $\mathcal{V}_1 \cup \mathcal{V}_2 = \mathcal{V}'$ such that $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$, where there are no edges with one endpoint in \mathcal{V}_1 and the other endpoint in \mathcal{V}_2 ,*

$$(11) \quad \left| \mathbf{E} \left(\prod_{\alpha \in \mathcal{V}'} f_{n;j_1, \dots, j_r}(X_{n,j_1}, \dots, X_{n,j_r}) \right) - \mathbf{E} \left(\prod_{\alpha \in \mathcal{V}_1} f_{n;j_1, \dots, j_r}(X_{n,j_1}, \dots, X_{n,j_r}) \right) \mathbf{E} \left(\prod_{\alpha \in \mathcal{V}_2} f_{n;j_1, \dots, j_r}(X_{n,j_1}, \dots, X_{n,j_r}) \right) \right| \leq 8^{|\mathcal{V}'|} F_n^{|\mathcal{V}'|} \beta_n(m).$$

The inequality (11) shows that the condition (7) is satisfied for the graph $\Gamma_{n,m}$ for every n and

$$(12) \quad \gamma_R = \gamma_{R,n,m} = 8R\beta_n(m).$$

We need the following statement to derive Lemma 1.

Let $I = (i_1, \dots, i_{|I|}), 1 \leq i_1 < \dots < i_{|I|} \leq n$, and $I' = (i'_1, \dots, i'_{|I'|}), 1 \leq i'_1 < \dots < i'_{|I'|} \leq n$, be the sets of natural numbers, and g_1 and g_2 be measurable functions of $|I|$ and $|I'|$ real variables, respectively.

Lemma 2. *Let the strictly stationary sequence of random variables $(X_t)_{t=1}^\infty$ satisfy the absolute regularity condition (2) with the coefficient $\beta(m)$, $|g_1(X_i; i \in I)| \leq F'$, $|g_2(X_{i'}; i' \in I')| \leq F''$, and*

$$(13) \quad \min_{i \in I, i' \in I'} \{|i - i'| > m\}.$$

Then

$$(14) \quad \left| \mathbf{E} \{g_1(X_i; i \in I)g_2(X_{i'}; i' \in I')\} - \mathbf{E}g_1(X_i; i \in I)\mathbf{E}g_2(X_{i'}; i' \in I') \right| \leq 8(|I| + |I'|)F'F''\beta(m).$$

Proof of Lemma 2. The sets I and I' can be one-to-one split into disjoint subsets $I = I_1 \cup \dots \cup I_{s_1}$ and $I' = I'_1 \cup \dots \cup I'_{s_2}$, where $1 \leq s_1, s_2 \leq |I| + |I'|, |s_1 - s_2| \in \{0, 1\}$,

satisfying one of the following two conditions:

$$(15) \quad \max_{i' \in I'_k} i' < \min_{i \in I_k} i, \quad \max_{i \in I_k} i < \min_{i' \in I'_{k+1}} i', \quad k = 1, \dots, \max\{s_1, s_2\} - 1,$$

$$(16) \quad \max_{i \in I_k} i < \min_{i' \in I'_k} i', \quad \max_{i' \in I'_k} i' < \min_{i \in I_{k+1}} i, \quad k = 1, \dots, \max\{s_1, s_2\} - 1.$$

We put

$$X(I_k) = (X_i; i \in I_k), \quad k = 1, \dots, s_1,$$

$$X(I'_k) = (X_{i'}; i' \in I'_k), \quad k = 1, \dots, s_2.$$

Then, we can write

$$g_1(X_i; i \in I) = g_1(X(I_1), \dots, X(I_{s_1})),$$

$$g_2(X_{i'}; i' \in I') = g_2(X(I'_1), \dots, X(I'_{s_2})).$$

According to the order (15), (16), we denote I_k and I'_k by combined notation H_k . In the first case $H_1 = I'_1, H_2 = I_1, H_3 = I'_2, \dots$, and in the second case $H_1 = I_1, H_2 = I'_1, H_3 = I_2, \dots$. Therefore, we use the notation $X(H_k) = (X_i; i \in H_k)$, $k = 1, \dots, s_1 + s_2$. It follows from the definitions that

$$(17) \quad \min_{i \in H_{k+1}} i - \max_{i \in H_k} i > m, \quad k = 1, \dots, s_1 + s_2 - 1.$$

Let $\tilde{X}(I_k), \tilde{X}(I'_k), \tilde{X}(H_k)$ be mutually independent copies of the random variables $X(I_k), k = 1, \dots, s_1, X(I'_k), k = 1, \dots, s_2, X(H_k), k = 1, \dots, s_1 + s_2$.

We use the notation

$$(18) \quad h(x_i; i \in I \cup I') = g_1(x_i; i \in I)g_2(x_{i'}; i' \in I')$$

and

$$(19) \quad h_\sigma(x_i; i \in I \cup I') = h(x_{\sigma_i}; i \in I \cup I'),$$

where $\sigma = (\sigma_1, \dots, \sigma_{s_1+s_2})$ is a permutation of the numbers $1, 2, \dots, s_1 + s_2$ such that

$$(20) \quad h(X_i : i \in I \cup I') = h_\sigma(X(H_1), \dots, X(H_{s_1+s_2})).$$

We use one result of K. Yoshihara (see [13], Lemma 1). It refers to a more general case and, in particular, instead of the boundedness of the function h , it requires the property $\mathbf{E}|h(x_i; i \in I \cup I')|^{1+\delta} \leq B_\delta < \infty$ for some $\delta > 0$.

In our case, $B_\delta \leq (F'F'')^{1+\delta} < \infty$ for every $\delta > 0$. Thus, Yoshihara's lemma gives that

$$\begin{aligned} & \left| \mathbf{E}h_\sigma(X(H_1), X(H_2), \dots, X(H_{s_1+s_2})) - \right. \\ & \left. - \mathbf{E}h_\sigma(\tilde{X}(H_1), X(H_2), \dots, X(H_{s_1+s_2})) \right| \leq \\ & \leq \inf_{\delta > 0} \left\{ 4B_\delta^{1/(1+\delta)} (\beta(m))^{\delta/(1+\delta)} \right\} \leq 4F'F''\beta(m). \end{aligned}$$

Analogously, we derive the inequality

$$\begin{aligned} & \left| \mathbf{E}h_\sigma(x(H_1), X(H_2), X(H_3), \dots, X(H_{s_1+s_2})) - \right. \\ & \left. - \mathbf{E}h_\sigma(x(H_1), \tilde{X}(H_2), X(H_3), \dots, X(H_{s_1+s_2})) \right| \leq 4F'F''\beta(m). \end{aligned}$$

for the function $h_\sigma(x(H_1), X(H_2), X(H_3), \dots, X(H_{s_1+s_2}))$, where $x(H_1) = (x_i; i \in H_1)$. It follows that

$$\begin{aligned} & \left| \mathbf{E}h_\sigma(\tilde{X}(H_1), X(H_2), X(H_3), \dots, X(H_{s_1+s_2})) - \right. \\ & \left. - \mathbf{E}h_\sigma(\tilde{X}(H_1), \tilde{X}(H_2), X(H_3), \dots, X(H_{s_1+s_2})) \right| \leq 4F'F''\beta(m). \end{aligned}$$

We carry out similar estimates for other differences. The last among them is

$$\begin{aligned} & \left| \mathbf{E}h_\sigma(\tilde{X}(H_1), \tilde{X}(H_2), \dots, \tilde{X}(H_{s_1+s_2-1}), X(H_{s_1+s_2})) - \right. \\ & \left. - \mathbf{E}h_\sigma(\tilde{X}(H_1), \tilde{X}(H_2), \dots, \tilde{X}(H_{s_1+s_2-1}), \tilde{X}(H_{s_1+s_2})) \right| \leq 4F'F''\beta(m). \end{aligned}$$

Summarizing the obtained estimates and using the triangle inequality, we obtain

$$\begin{aligned} & \left| \mathbf{E}h_\sigma(X(H_1), X(H_2), \dots, X(H_{s_1+s_2})) - \right. \\ & \left. - \mathbf{E}h_\sigma(\tilde{X}(H_1), \tilde{X}(H_2), \dots, \tilde{X}(H_{s_1+s_2})) \right| \leq \\ (21) \quad & \leq 4(s_1 + s_2)F'F''\beta(m) = 4(|I| + |I'|)F'F''\beta(m). \end{aligned}$$

It follows from (18), (19), (20), (21) that

$$\begin{aligned} & \left| \mathbf{E}\{g_1(X(I_1), X(I_2), \dots, X(I_{s_1}))g_2(X(I'_1), X(I'_2), \dots, X(I'_{s_2}))\} - \right. \\ & \left. - \mathbf{E}\{g_1(\tilde{X}(I_1), \tilde{X}(I_2), \dots, \tilde{X}(I_{s_1}))g_2(\tilde{X}(I'_1), \tilde{X}(I'_2), \dots, \tilde{X}(I'_{s_2}))\} \right| \leq \\ (22) \quad & \leq 4(|I| + |I'|)F'F''\beta(m). \end{aligned}$$

Analogously (21), we derive the inequalities

$$\begin{aligned} & \left| \mathbf{E}g_1(X(I_1), \dots, X(I_{s_1})) - \mathbf{E}g_1(\tilde{X}(I_1), \dots, \tilde{X}(I_{s_1})) \right| \leq 4|I|F'\beta(m), \\ & \left| \mathbf{E}g_2(X(I'_1), \dots, X(I'_{s_2})) - \mathbf{E}g_2(\tilde{X}(I'_1), \dots, \tilde{X}(I'_{s_2})) \right| \leq 4|I'|F''\beta(m). \end{aligned}$$

It follows from these inequalities, the inequalities

$$\mathbf{E}|g_1(X(I_1), \dots, X(I_{s_1}))| \leq F', \quad \mathbf{E}|g_2(\tilde{X}(I'_1), \dots, \tilde{X}(I'_{s_2}))| \leq F''$$

and the triangle inequality that

$$\begin{aligned} & \left| \mathbf{E}g_1(X(I_1), \dots, X(I_{s_1}))\mathbf{E}g_2(X(I'_1), \dots, X(I'_{s_2})) - \right. \\ & \left. - \mathbf{E}g_1(\tilde{X}(I_1), \dots, \tilde{X}(I_{s_1}))\mathbf{E}g_2(\tilde{X}(I'_1), \dots, \tilde{X}(I'_{s_2})) \right| \leq \\ & \leq \mathbf{E}|g_1(X(I_1), \dots, X(I_{s_1}))| \left| \mathbf{E}g_2(X(I'_1), \dots, X(I'_{s_2})) - \mathbf{E}g_2(\tilde{X}(I'_1), \dots, \tilde{X}(I'_{s_2})) \right| + \\ & + \mathbf{E}|g_2(\tilde{X}(I'_1), \dots, \tilde{X}(I'_{s_2}))| \left| \mathbf{E}g_1(X(I_1), \dots, X(I_{s_1})) - \mathbf{E}g_1(\tilde{X}(I_1), \dots, \tilde{X}(I_{s_1})) \right| \leq \\ (23) \quad & \leq 4(|I| + |I'|)F'F''\beta(m). \end{aligned}$$

The formulas (22), (23) and the triangle inequality give (14). □

Proof of Lemma 1. Lemma 1 is immediate from Lemma 2, if we consider that $|f_{n;j_1, \dots, j_r}(x_1, \dots, x_r)| \leq F_n$ and $|I| + |I'| \leq |\mathcal{V}'|$. Thus, (14) leads to (11). □

Proof of Theorem 2. We estimate the quantities in the condition (9) for our case. We begin with

$$(24) \quad Q_{R,n,m} = \max_{\mathcal{V}' \subset \mathcal{V}(I_{n,m}): |\mathcal{V}'| \leq R} \sum_{\tilde{\alpha} \in L(\mathcal{V}')} \mathbf{E}(|w_n(\tilde{\alpha})| \sigma\{w_n(\alpha), \alpha \in \mathcal{V}'\}),$$

where we recall $w_n(\alpha) = f_{n;j_1, \dots, j_r}(X_{n,j_1}, \dots, X_{n,j_r})$.

The set of strong dependencies $L(\alpha) = L_{n,m}(\alpha)$ for the vertex $\alpha = (j_1, \dots, j_r)$ in the graph $\Gamma_{n,m}$ is the set of vertices $\tilde{\alpha} = (\tilde{j}_1, \dots, \tilde{j}_r) \in \mathcal{V}(\Gamma_{n,m})$ for which at least one of the inequalities (10) holds. The number of the elements of this set satisfies the inequality

$$(25) \quad |L(\alpha)| \leq r^2(2m + 1)C_n^{r-1}.$$

The set of strong dependencies $L(\mathcal{V}') = L_{n,m}(\mathcal{V}')$ for the set $\mathcal{V}' \subset \mathcal{V}(\Gamma_{n,m})$ is given by the formula

$$L(\mathcal{V}') = \bigcup_{\alpha \in \mathcal{V}'} L(\alpha)$$

and satisfies the inequality

$$(26) \quad |L(\mathcal{V}')| \leq r^2|\mathcal{V}'|(2m + 1)C_n^{r-1}.$$

From formulas (26), $|f_{n;j_1, \dots, j_r}(X_{n,j_1}, \dots, X_{n,j_r})| \leq F_n$ and (24), we have

$$(27) \quad Q_{R,n,m} \leq r^2RF_n(2m + 1)C_n^{r-1}.$$

We note that, in the case under consideration,

$$(28) \quad M_n = \sum_{1 \leq j_1 < \dots < j_r \leq n} \mathbf{E}|f_{n;j_1, \dots, j_r}(X_{n,j_1}, \dots, X_{n,j_r})| \leq F_n C_n^r.$$

The use of (11), (27), (28), equality (12), and the above cited Theorem 1 from [24] leads to the following condition of the asymptotic normality for U_n in the triangular array scheme: for every natural number R and all $b \leq 2/3$

$$\frac{F_n^2 m_n^{2-b} n^{2(r-1)+b}}{\mathbf{D}U_n} + (\beta_n(m_n))^b \frac{F_n^2 n^{2r}}{\mathbf{D}U_n} \rightarrow 0.$$

□

Proof of Theorem 4. Let the number F_n be independent of n , $\beta_n(m) \leq m^{-h(m)}$, where $h(m) \rightarrow \infty$ ($m \rightarrow \infty$), and $\mathbf{D}U_n \geq Cn^{2(r-1)+\varkappa}$, where $C, \varkappa > 0$. In this case (5) follows from the formula:

$$(29) \quad m_n^{2-b} n^{b-\varkappa} + m_n^{-h(m_n)b} n^{2(r-1)-\varkappa} \rightarrow 0.$$

Let us examine this. We put

$$m_n = \left[n^{(\varkappa-b_0)/4} \right], \quad 0 < b_0 < \min \{2/3, \varkappa\},$$

then, the expression in the left side (29) for $0 < b \leq b_0$ and $n \rightarrow \infty$ can be estimated by $O(n^{(b_0-\varkappa)/2}) \rightarrow 0$. □

4. REMARKS ABOUT PROOFS OF THEOREMS 3 AND 5 FOR V-STATISTICS

The proofs of Theorems 2 and 4 for V -statistics are completely the same, with the only difference that the graph $\Gamma_{n,m}$ contains n^r vertices, and in the formulas (25), (26), (27), and (28) the binomial coefficients C_n^s must be replaced by n^s , $s = r - 1, r$. The indicated replacement does not affect the form of the condition (5).

For example, consider the formula (25). Number of elements $\tilde{\alpha} = (\tilde{j}_1, \dots, \tilde{j}_r) : \tilde{j}_1, \dots, \tilde{j}_r = 1, \dots, n$, for which at least one of the inequalities (10) holds, can be estimated as follows. First, we select the indices k and l , there are $2m + 1$ elements satisfying (10) for each such pair, and the rest of the elements can be any. Thus,

$$|L(\alpha)| \leq r^2(2m + 1)n^{r-1}.$$

The remaining calculations are carried out similarly.

The authors are grateful to the reviewer for useful comments that improved the presentation of the results.

REFERENCES

- [1] W. Hoeffding, *A class of statistics with asymptotically normal distribution*, Ann. Math. Statist., **19**:3 (1948), 293–325. Zbl 0032.04101
- [2] A.M. Zubkov, V.G. Mikhailov, *Limit distributions of random variables connected with long duplications in a sequence of independent trials*, Theory Probab. Appl., **19**:1 (1974), 172–179. Zbl 0326.60024
- [3] A.M. Zubkov, V.G. Mikhailov, *On the repetitions of s -tuples in a sequence of independent trials*, Theory Probab. Appl., **24**:2 (1979), 269–282. Zbl 0434.60022
- [4] M.I. Tikhomirova, *Limit distributions of the number of absent chains of identical outcomes*, Discrete Math. Appl., **18**:3 (2008), 293–300. Zbl 1176.60016
- [5] M.I. Tikhomirova, *Asymptotic normality of the number of absent noncontinuous chains of independent trials*, Discrete Math. Appl., **19**:3 (2009), 293–308. Zbl 1236.60026
- [6] V.G. Mikhailov, *On the asymptotic properties of the distribution of the number of pairs of H -connected chains*, Discrete Math. Appl., **12**:4 (2002), 393–400. Zbl 1046.60007
- [7] A.M. Shoitov, *Limit distributions of the number of sets of H -equivalent segments in an equiprobable polynomial scheme of arrays*, Discrete Math. Appl., **12**:2 (2002), 165–181. MR1919858
- [8] V.G. Mikhailov, A.M. Shoitov, *Structural equivalence of s -tuples in random discrete sequences*, Discrete Math. Appl., **13**:6 (2003), 541–568. Zbl 1046.60052
- [9] A.M. Shoitov, *The Poisson approximation for the number of matches of values of a discrete function on segments of a sequence of random variables*, Discrete Math. Appl., **15**:3 (2005), 241–254. Zbl 1113.60029
- [10] A.M. Shoitov, *Normal approximation in a problem on equivalent tuples*, Tr. Diskr. Mat., **10** (2007), 326–349.
- [11] A.M. Shoitov, *Asymptotic distributional properties of the number of pairs of metrically close and functionally connected tuples in a polynomial scheme*, Tr. Diskr. Mat., **11**:1 (2008), 151–165.
- [12] V.S. Koroljuk, Yu.V. Borovskych, *Theory of U -statistics*, Springer, Dordrecht, Netherlands, 1994. MR1472486
- [13] K. Yoshihara, *Limiting behavior of U -statistics for stationary absolutely regular process*, Z. Wahrsch. Verw. Gebiete, **35** (1976), 237–252. Zbl 0314.60028
- [14] M. Denker, G. Keller, *On U -statistics and V . Mises' statistics for weakly dependent processes*, Z. Wahrsch. Verw. Gebiete, **64** (1983), 505–522. Zbl 0519.60028
- [15] I. Dewan, P. Rao, *Central limit theorem for U -statistics of associated random variables*, Stat. Probab. Lett., **57**:1 (2002), 9–15. Zbl 1004.60018
- [16] H. Dehling, *Limit theorems for dependent U -statistics*, in Bertail et al. *Depedence in Probability and Statistics*, Springer, New York, 2006, 65–86. Zbl 1105.62047
- [17] P. Doukhan, *Mixing: Properties and Examples*, Lecture Notes in Statistics **85**, Springer-Verlag, New-York, 1994. Zbl 0801.60027
- [18] Sh.A. Khashimov, *The central limit theorem for generalized U -statistics for weakly dependent vectors*, Theory Probab. Appl., **38**:3 (1993), 456–469. Zbl 0813.60019
- [19] Sh.A. Khashimov, *On the asymptotical distribution of the generalized U -statistic for depending values*, Theory Probab. Appl., **32**:2 (1987), 373–375. Zbl 0715.60027
- [20] Sh.A. Khashimov, *Limiting behavior of generalized U -statistics of weakly dependent stationary processes*, Theory Probab. Appl., **37**:1 (1993), 148–150. Zbl 0794.60028
- [21] V.G. Mikhailov, *The central limit theorem for nonhomogeneous U -statistics of finitely dependent random variables*, Math. USSR-Sb., **27**:4(1975) (1977), 554–563. Zbl 0373.62013
- [22] S. Janson, *Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs*, Ann. Probab., **16**:1 (1988), 306–312. Zbl 0639.60029
- [23] V.G. Mikhailov, *On a theorem of Janson*, Theory Probab. Appl., **36**:1 (1991), 173–176. Zbl 0751.60021

- [24] M.I. Tikhomirova, V.P. Chistyakov, *On the asymptotic normality of some sums of dependent random variables*, Discrete Math. Appl., **27**:2 (2017), 123–129. Zbl 1377.60043
- [25] V.G. Mikhailov, A.M. Shoitov, *On repetitions of long tuples in a Markov chain*, Discrete Math. Appl., **25**:5 (2015), 295–303. Zbl 1375.60117
- [26] V.G. Mikhailov, *Estimates of accuracy of the Poisson approximation for the distribution of number of runs of long string repetitions in a Markov chain*, Discrete Math. Appl., **26**:2 (2016), 105–113. Zbl 1375.60116
- [27] V. Féray, *Weighted dependence graph*, Electron. J. Probab., **23** (2018), paper no. 93, 65 pp.; arXiv:1605.0386v1 [math.PR] 12 May 2016.
- [28] V. Féray, *Central limit theorems for patterns in multiset permutations and set partitions*, arXiv:1811.03923v1 [math.CO] 9 Nov 2018.

VLADIMIR GAVRILOVICH MIKHAILOV
STEKLOV MATHEMATICAL INSTITUTE OF RUSSIAN ACADEMY OF SCIENCES,
8, GUBKINA STR.,
MOSCOW, 119991, RUSSIA
E-mail address: `mikhail@mi-ras.ru`

NATALIA MIKHAILOVNA MEZHENNAYA
BAUMAN MOSCOW STATE TECHNICAL UNIVERSITY,
5/1, 2-AYA BAUMANSKAYA STR.,
MOSCOW, 105005, RUSSIA
E-mail address: `natalia.mezhennaya@gmail.com`