

Limit theorems and exponential inequalities for canonical U - and V -statistics of dependent trials

Igor S. Borisov^{1,*} and Nadezhda V. Volodko^{1,†}

Sobolev Institute of Mathematics

Abstract: The limit behavior is studied for the distributions of normalized U - and V -statistics of an arbitrary order with *canonical* (degenerate) kernels, based on samples of increasing sizes from a stationary sequence of observations satisfying φ - or α -mixing. The case of m -dependent sequences is separately studied. The corresponding limit distributions are represented as infinite multilinear forms of a centered Gaussian sequence with a known covariance matrix. Moreover, under φ -mixing, exponential inequalities are obtained for the distribution tails of these statistics with bounded kernels.

1. Introduction. Preliminary results

In the present paper, we study the limit behavior of the distributions of normalized canonical U - and V -statistics based on stationary observations under φ - or α -mixing. Moreover, the Hoeffding-type exponential inequalities are obtained for the distribution tails of the statistics mentioned. The approach based on a kernel representation of the statistics under consideration as a multiple series, is quite similar to the approach in [16] where the analogous limit theorems were obtained for independent observations. First of all, introduce some definitions and notions (see [12, 13]).

Let X_1, X_2, \dots be a stationary sequence of random variables with values in an arbitrary measurable space $\{\mathfrak{X}, \mathcal{A}\}$ and distribution F . In addition to the stationary sequence introduced above, we need an auxiliary sequence $\{X_i^*\}$ consisting of independent copies of X_1 .

Denote by $L_2(\mathfrak{X}^r, F^r)$ the space of measurable functions $f(t_1, \dots, t_r)$ defined on the corresponding Cartesian power of the space $\{\mathfrak{X}, \mathcal{A}\}$ with the corresponding product-measure and satisfying the condition $\mathbf{E}f^2(X_1^*, \dots, X_r^*) < \infty$.

Definition 1. A function $f(t_1, \dots, t_r) \in L_2(\mathfrak{X}^r, F^r)$ is called *canonical* (or *degenerate*) if

$$(1.1) \quad \mathbf{E}f(t_1, \dots, t_{k-1}, X_k, t_{k+1}, \dots, t_r) = 0$$

*Supported in part by the Russian Foundation for Basic Research (grant 08-01-00962) and by the Ministry of Education and Science of the Russian Federation (grant RNP.2.1.1.346).

†Supported in part by the Ministry of Education and Science of the Russian Federation (grant RNP.2.1.1.346).

¹Sobolev Institute of Mathematics, Novosibirsk, 630090, Russia, e-mail: sibam@math.nsc.ru; nvolodko@gmail.com

AMS 2000 subject classifications: Primary 60H05, 60F05; secondary 62G20.

Keywords and phrases: stationary sequence of random variables, mixing, multiple orthogonal series, canonical U - and V -statistics.

for every $k \leq r$ and all $t_j \in \mathfrak{X}$.

Introduce one more notation:

$$I_n^r = \{(j_1, \dots, j_r) : j_k \leq n, j_k \neq j_l \text{ for } k \neq l\}.$$

Define a von Mises statistic (or a *V-statistic*) by the formula

$$(1.2) \quad V_n \equiv V_n(f) := n^{-r/2} \sum_{1 \leq j_1, \dots, j_r \leq n} f(X_{j_1}, \dots, X_{j_r}).$$

In the sequel, we consider only the statistics where the function $f(t_1, \dots, t_r)$ (the so-called *kernel* of the statistic) is canonical. In this case, the corresponding von Mises statistic is also called *canonical*. For independent $\{X_i\}$ (the i.i.d. case), such statistics called *canonical V-statistics* as well, are studied during last sixty years (see the reference and examples of such statistics in [13]). For the first time, some limit theorems in the bivariate case were obtained in [15, 11]. In addition to *V-statistics*, the so-called *U-statistics* were studied as well:

$$(1.3) \quad U_n \equiv U_n(f) := n^{-r/2} \sum_{(j_1, \dots, j_r) \in I_n^r} f(X_{j_1}, \dots, X_{j_r}).$$

Notice also that any *U-statistic* is represented as a finite linear combination of canonical *U-statistics* of orders from 1 to m (called the *Hoeffding decomposition*, see [13]). This fact allows us to reduce an asymptotic analysis of arbitrary *U-statistics* to that for canonical ones.

In this connection, recall some classic results connected with the expansion of a canonical function into a multiple orthogonal series with respect to an orthogonal basis of the Hilbert space $L_2(\mathfrak{X}, F)$. Let \mathfrak{X} be a separable metric space. Then the Hilbert space $L_2(\mathfrak{X}, F)$ is separable. It means that, in this space, there exists a countable orthonormal basis. Put $e_0(t) \equiv 1$. Using the Gram–Schmidt orthogonalization [12], one can construct an orthonormal basis in $L_2(\mathfrak{X}, F)$ containing the constant function $e_0(t) \equiv 1$. Denote by $\{e_i(t)\}_{i \geq 0}$ such basis. Then $\mathbf{E}e_i(X_1) = 0$ for every $i \geq 1$ due to the orthogonality of all the other basis elements to the function $e_0(t)$. The normalizing condition means that $\mathbf{E}e_i^2(X_1) = 1$ for all $i \geq 1$. Notice that the collection of functions

$$\{e_{i_1}(t_1)e_{i_2}(t_2) \cdots e_{i_r}(t_r); i_1, i_2, \dots, i_r = 0, 1, \dots\}$$

is an orthonormal basis in the Hilbert space $L_2(\mathfrak{X}^r, F^r)$ (for example, see [12]).

Thus, one can represent the kernel $f(t_1, \dots, t_r)$ of the statistics under consideration as a multiple orthogonal series in $L_2(\mathfrak{X}^r, F^r)$:

$$(1.4) \quad f(t_1, \dots, t_r) = \sum_{i_1, \dots, i_r=0}^{\infty} f_{i_1, \dots, i_r} e_{i_1}(t_1) \cdots e_{i_r}(t_r),$$

where the series on the right-hand side of equality (1.4) converges in the norm of $L_2(\mathfrak{X}^r, F^r)$. Moreover, if the coefficients $\{f_{i_1, \dots, i_r}\}$ are absolutely summable then, due to the B. Levi theorem and the simple estimate $\mathbf{E}|e_{i_1}(X_1^*) \cdots e_{i_r}(X_r^*)| \leq 1$, the series in (1.4) converges almost surely with respect to the distribution F^r of the vector (X_1^*, \dots, X_r^*) .

Consider the case $r = 2$ and the integral linear operator with a symmetric kernel $f \in L_2(\mathfrak{X}^2, F^2)$ mapping the space $L_2(\mathfrak{X}, F)$ into itself. Since this linear operator is

completely continuous and self-conjugate, in the separable Hilbert space $L_2(\mathfrak{X}, F)$, there exists an orthonormal basis consisting of eigenvectors of this integral operator and, for this basis, representation (1.4) for $r = 2$ is valid. Multiply by an arbitrary element $e_k(t_2)$ the both sides of (1.4) and integrate these modified parts with respect to the distribution $F(dt_2)$. Taking orthogonality of the basis elements into account we obtain the new identity

$$\lambda_k e_k(t_1) = \sum_{i=0}^{\infty} f_{i,k} e_i(t_1),$$

where λ_k is the corresponding eigenvalue. From here it immediately follows that $f_{k,k} = \lambda_k$ and $f_{i,k} = 0$ for $i \neq k$. Therefore, for this basis in the case $r = 2$, formula (1.4) has the form

$$(1.5) \quad f(t_1, t_2) = \sum_{k=0}^{\infty} \lambda_k e_k(t_1) e_k(t_2),$$

which was repeatedly employed by many authors.

Notice also the following property of canonical kernels.

Proposition 1 ([6]). *If $f(t_1, \dots, t_r)$ is a canonical kernel then $e_0(t)$ is absent in expansion (1.4), i. e., expansion (1.4) has the form*

$$(1.6) \quad f(t_1, \dots, t_r) = \sum_{i_1, \dots, i_r=1}^{\infty} f_{i_1, \dots, i_r} e_{i_1}(t_1) \cdots e_{i_r}(t_r).$$

Notice also that if the kernel $f(t_1, t_2)$ in (1.5) is canonical then the constant function $e_0(t)$ is the eigenfunction corresponding to the eigenvalue $\lambda_0 = 0$ of the integral operator. So, in this case, the summation in (1.5) starts with $k = 1$.

Thus, after replacement of the vector (t_1, \dots, t_r) by the independent observations (X_1^*, \dots, X_r^*) , the partial sums of the series on the right-hand side of (1.6) (or of (1.5) in the case $r = 2$) mean-square converge to the random variable $f(X_1^*, \dots, X_r^*)$ and hence they converge in distribution. However, in the present paper, we deal with dependent observations for which this property in general is not valid.

2. Limit theorems for U - and V -statistics under α - or φ -mixing

We study stationary sequences $\{X_j\}$ satisfying certain mixing conditions. Recall the definitions of the most popular mixing conditions. For $j \leq k$, denote by \mathfrak{M}_j^k the σ -field of all events generated by the random variables X_j, \dots, X_k .

Definition 2. A sequence X_1, X_2, \dots satisfies α -mixing (or strong mixing) if

$$\alpha(i) := \sup_{k \geq 1} \sup_{A \in \mathfrak{M}_1^k, B \in \mathfrak{M}_{k+i}^{\infty}} |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)| \rightarrow 0 \quad \text{as } i \rightarrow \infty.$$

Definition 3. A sequence X_1, X_2, \dots satisfies φ -mixing (or uniformly strong mixing) if

$$\varphi(i) := \sup_{k \geq 1} \sup_{A \in \mathfrak{M}_1^k, B \in \mathfrak{M}_{k+i}^{\infty}, \mathbb{P}(A) > 0} \frac{|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|}{\mathbb{P}(A)} \rightarrow 0 \quad \text{as } i \rightarrow \infty.$$

Definition 4. A sequence X_1, X_2, \dots satisfies ψ -mixing if

$$\psi(i) := \sup_{k \geq 1} \sup_{A \in \mathfrak{M}_1^k, B \in \mathfrak{M}_{k+i}^\infty, \mathbb{P}(A)\mathbb{P}(B) > 0} \frac{|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|}{\mathbb{P}(A)\mathbb{P}(B)} \rightarrow 0 \quad \text{as } i \rightarrow \infty$$

and $\psi(1) < \infty$.

It is clear that the sequences $\{\alpha(i)\}$, $\{\varphi(i)\}$, and $\{\psi(i)\}$ are nondecreasing and ψ -mixing is stronger than φ -mixing which in turn implies α -mixing.

In the sequel, in the case of φ -mixing, we assume that

$$(2.1) \quad \sum_{k=1}^{\infty} \varphi^{1/2}(k) < \infty.$$

Note that this known condition provides the central limit theorem for the corresponding stationary sequences (for example, see [3]).

Introduce also the following restriction on finite-dimensional distributions of the stationary sequence $\{X_i\}$.

(AC) For every collection of pairwise distinct subscripts (j_1, \dots, j_r) , the distribution of $(X_{j_1}, \dots, X_{j_r})$ is absolutely continuous with respect to the distribution of (X_1^*, \dots, X_r^*) .

Notice that this restriction will be nontrivial only for sequences under α - or φ -mixing because, in the case of ψ -mixing, by induction on r , from Definition 4 we can easily deduce the inequality

$$P(X_{j_1} \in A_1, \dots, X_{j_r} \in A_r) \leq (1 + \psi(1))^r \prod_{k=1}^r P(X_k \in A_k)$$

which is valid for every collection of Borel subsets (A_1, \dots, A_r) and for every pairwise distinct subscripts (j_1, \dots, j_r) . From here condition (AC) immediately follows.

Remark 1. As was mentioned before, the condition

$$(2.2) \quad \sum_{i_1, \dots, i_r=1}^{\infty} |f_{i_1, \dots, i_r}| < \infty$$

implies convergence of the series in (1.6) almost surely with respect to the distribution of the vector (X_1^*, \dots, X_r^*) . So, under condition (AC), this convergence is valid almost surely with respect to the distribution of the random vector $(X_{j_1}, \dots, X_{j_r})$. In other words, if condition (AC) is fulfilled then, for any pairwise distinct subscripts j_1, \dots, j_r , we can substitute the random variables X_{j_1}, \dots, X_{j_r} for t_1, \dots, t_r in (1.6).

Remark 2. Under restriction (2.2), one can sometimes obtain the above-mentioned multiple series expansions without any restrictions like (AC) on joint distributions of the initial stationary sequence. For example, if the kernel $f(t_1, \dots, t_r)$ is continuous in \mathbb{R}^r and all the basis functions $e_k(t)$ are continuous and bounded uniformly in k then, under condition (2.2), equality (1.6) is transformed into the *identity* in \mathbb{R}^r (see the proof of Theorem 2 below in [6]). Therefore, in this identity, one can substitute *arbitrarily dependent* random variables X_{j_1}, \dots, X_{j_r} (in particular, for coinciding subscripts j_k) for the arguments t_1, \dots, t_r .

Under φ -mixing for the sequence $\{X_i\}$ but without the restrictions mentioned in Remarks 1 and 2, in general one cannot use expansions (1.5) or (1.6). Such mistake is contained in [8] (see also [13]), where it is claimed that, in the case of φ -mixing stationary observations for $r = 2$, under condition (2.1) only but without any restrictions like (AC) and the regularity condition mentioned in Remark 2, the following analogue of von Mises' result [15] is valid:

$$(2.3) \quad U_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k (\tau_k^2 - 1),$$

where $\{\lambda_k\}$ are the eigenvalues of the integral operator with the symmetric kernel $f(t_1, t_2)$, which are assumed to be summable (i. e., under condition (2.2)), and $\{\tau_k\}$ is a Gaussian sequence of centered random variables with the covariances

$$(2.4) \quad \mathbf{E}\tau_k\tau_l = \mathbf{E}e_k(X_1)e_l(X_1) + \sum_{j=1}^{\infty} [\mathbf{E}e_k(X_1)e_l(X_{j+1}) + \mathbf{E}e_l(X_1)e_k(X_{j+1})],$$

where $\{e_k(t)\}$ are the eigenfunctions corresponding to the eigenvalues $\{\lambda_k\}$ and forming an orthonormal basis in $L_2(\mathfrak{X}, F)$ (actually, the first summand on the right-hand side (2.4) is the Kronecker symbol $\delta_{k,l}$). In (2.3) and in the sequel, we admit degeneracy of some random variables τ_i . In other words, we add to the class of Gaussian distributions the all weak limits when the variance tends to zero. Actually, the similar agreement is contained in [3]. To prove relation (2.3) in [8], the author use the expansion of the kernel $f(t_1, t_2)$ in series (1.5) with respect to the basis $\{e_k(t)\}$. Due to the above-mentioned arguments, we could now substitute the independent observations X_i^* and X_j^* for the variables t_1 and t_2 . The same is true for a pair X_i and X_j from a stationary sequence satisfying condition (AC). However, in [8], the author substituted a pair X_i and X_j from an arbitrary stationary sequence under φ -mixing with restriction (2.1) only, for the variables t_1 and t_2 . But under this replacement, the above-mentioned equalities may not be fulfilled with a nonzero probability. Moreover, in this case, the limit law in (2.3) may change the form. The idea of constructing examples of such a kind is very simple: We need to construct a stationary sequence $\{X_i\}$ with a non-atomic marginal distribution, such that its elements X_i and X_j coincide with nonzero probability for some subscripts $i \neq j$. We then can change the values of f on diagonal subspaces to break the above-mentioned identities with a nonzero probability when we replace the arguments X_i^* and X_j^* of the kernel with the dependent pair X_i and X_j . The corresponding construction is contained in the proof of the following assertion in [6].

Proposition 2. *There exist a stationary sequence $\{X_i\}$ and a canonical $f(t_1, t_2)$ satisfying all the restrictions in [8]. However, under substituting the observations X_1 and X_2 for t_1 and t_2 respectively, the series in (1.5) does not coincide with the kernel. Moreover, the weak limit for the distributions of the corresponding U -statistics differs from (2.3).*

The proof of Proposition 2 is contained in Section 4.

So, under certain conditions (say, conditions (2.2) and (AC)), U -statistic (1.3) can be represented as the following multiple series converging almost surely:

$$U_n = n^{-r/2} \sum_{i_1, \dots, i_r=1}^{\infty} f_{i_1, \dots, i_r} \sum_{(j_1, \dots, j_r) \in I_n^r} e_{i_1}(X_{j_1}) \cdots e_{i_r}(X_{j_r}).$$

Further analysis is similar to that in the i.i.d. case, i. e., it is reduced to extraction of V -statistics with splitting kernels from the multiple sum on the right-hand side of this identity. The main fragment of the proof in [16] is as follows: The value

$$U_n(e_{i_1} \cdots e_{i_r}) = n^{-r/2} \sum_{(j_1, \dots, j_r) \in I_n^r} e_{i_1}(X_{j_1}) \cdots e_{i_r}(X_{j_r})$$

is represented as a linear combination of products of the values

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n e_i(X_j), \frac{1}{n} \sum_{j=1}^n e_{i_1}(X_j) e_{i_2}(X_j), \dots, \frac{1}{n^{k/2}} \sum_{j=1}^n e_{i_1}(X_j) \cdots e_{i_k}(X_j).$$

The proof has a combinatorial character and does not depend of joint distributions of the random variables $\{X_j\}$. Further we apply the corresponding laws of large numbers as well as the central limit theorem and the following simple assertion.

Proposition 3. *Let $\Phi(x, y)$, $x \in \mathbb{R}^k$, $y \in \mathbb{R}^l$, be a continuous function. Let $\{\zeta_n\}$ be an arbitrary sequence of random vectors in \mathbb{R}^k weakly converging to some random vector ζ . Let $\{\eta_n\}$ be a sequence of random vectors in \mathbb{R}^l defined on a common probability space with $\{\zeta_n\}$, which converges in probability to a constant vector c_0 . Then the following weak convergence is valid:*

$$\Phi(\zeta_n, \eta_n) \xrightarrow{d} \Phi(\zeta, c_0).$$

In the present section, $\Phi(x, y)$ is a polynomial of components of vectors x and y , and the sequence ζ_n is defined by the formula

$$\zeta_n := \left\{ n^{-1/2} \sum_{j=1}^n e_1(X_j), \dots, n^{-1/2} \sum_{j=1}^n e_N(X_j) \right\},$$

and η_n is the finite collection (vector)

$$\left\{ n^{-k/2} \sum_{j=1}^n e_{i_1}(X_j) \cdots e_{i_k}(X_j); 2 \leq k \leq r, i_1, \dots, i_k \leq N \right\};$$

here N is an arbitrary natural number.

So, under the condition $f(t_1, \dots, t_r) \in L_2(\mathfrak{X}^r, F^r)$, in the case of i.i.d. random variables $\{X_i\}$, it was proved in [16] that

$$(2.5) \quad U_n \xrightarrow{d} \sum_{i_1, \dots, i_r=1}^{\infty} f_{i_1, \dots, i_r} \prod_{j=1}^{\infty} H_{\nu_j(i_1, \dots, i_r)}(\tau_j),$$

where $\{\tau_i\}$ is a sequence of independent random variables having the standard Gaussian distribution, $\nu_j(i_1, \dots, i_r) := \sum_{r=1}^r \delta_{j, i_r}$, and $H_k(x)$ are Hermite polynomials defined by the formula

$$H_k(x) := (-1)^k e^{x^2/2} \frac{d^k}{dx^k} (e^{-x^2/2})$$

or by the recurrent formula

$$H_0(x) \equiv 1, \quad H_1(x) = x, \\ H_{n+1}(x) = xH_n(x) - nH_{n-1}(x).$$

Thus, $H_k(x)$ is a polynomial of degree k and the product on the right-hand side of (2.5) can be represented in the form

$$(2.6) \quad \prod_{j=\min\{i_k\}}^{\max\{i_k\}} H_{\nu_j(i_1, \dots, i_r)}(\tau_j) = H_{r_1}(\tau_{j_1}) \cdots H_{r_s}(\tau_{j_s}),$$

where the natural numbers r_k and j_k are defined by the relation $r_k = \sum_{l=1}^r \delta_{j_k, i_l}$, and at that, $\sum_{k \leq s} r_k = r$ and $\min\{i_k\} \leq j_l \leq \max\{i_k\}$ for all $l \leq s$. Therefore, the right-hand of (2.6) is a polynomial of degree r of the variables $\tau_{j_1}, \dots, \tau_{j_s}$ and with coefficients having a universal upper bound depending on r only.

The goal of the present section is to formulate limit representations of the form (2.5) in the case of weakly dependent random variables $\{X_i\}$.

Introduce some additional restrictions on the mixing coefficients and the basis functions in the expansion in (2.5). In the sequel, we assume that the stationary sequence $\{X_i\}$ satisfies either α -mixing or φ -mixing, and moreover, the orthonormal basis $\{e_i(t)\}$ in $L_2(\mathfrak{X}, F)$ with the original element $e_0(t) \equiv 1$, satisfies the following additional restrictions:

1. In the case of φ -mixing, we assume condition (2.1) to be satisfied and

$$(2.7) \quad \sup_i \mathbf{E} |e_i(X_1)|^r < \infty;$$

2. In the case of α -mixing, we assume that, for some $\varepsilon > 0$ and an even number $c \geq r$,

$$(2.8) \quad \sup_i \mathbf{E} |e_i(X_1)|^{r+\varepsilon} < \infty,$$

$$(2.9) \quad \sum_{k=1}^{\infty} k^{c-2} \alpha^{\varepsilon/(c+\varepsilon)}(k) < \infty.$$

Further, introduce a sequence of centered Gaussian random variables $\{\tau_i\}$ with the covariances

$$(2.10) \quad \mathbf{E} \tau_k \tau_l = \mathbf{E} e_k(X_1) e_l(X_1) + \sum_{j=1}^{\infty} [\mathbf{E} e_k(X_1) e_l(X_{j+1}) + \mathbf{E} e_l(X_1) e_k(X_{j+1})].$$

The existence of the series in (2.10) follows from the above-mentioned restrictions on the mixing coefficients. In what follows, the Gaussian sequence $\{\tau_i\}$ will play a role of the weak limit as $n \rightarrow \infty$ for the sequence

$$\left\{ n^{-1/2} \sum_{j=1}^n e_1(X_j), n^{-1/2} \sum_{j=1}^n e_2(X_j), \dots \right\}.$$

We now formulate two limit theorems for the statistics under various mixing conditions.

Theorem 1 ([6]). *Let one of the following two conditions be fulfilled:*

1. *The stationary sequence $\{X_i\}$ satisfies φ -mixing, (2.1), and (2.7);*
2. *The stationary sequence $\{X_i\}$ satisfies α -mixing, (2.8), and (2.9).*

Then, for any canonical kernel $f(t_1, \dots, t_r) \in L_2(\mathfrak{X}^r, F^r)$, under conditions (2.2) and (AC), the following assertion holds:

$$(2.11) \quad U_n(f) \xrightarrow{d} \sum_{i_1, \dots, i_r=1}^{\infty} f_{i_1, \dots, i_r} \prod_{j=1}^{\infty} H_{\nu_j(i_1, \dots, i_r)}(\tau_j),$$

where $U_n(f)$ is a statistic of the form (1.3) and the centered Gaussian sequence $\{\tau_i\}$ has the covariance matrix defined in (2.10).

Theorem 2 ([6]). Let \mathfrak{X} be a separable metric space. Suppose that a canonical kernel $f(t_1, \dots, t_r)$ is continuous (in every argument) everywhere on \mathfrak{X}^r and condition (2.2) is fulfilled. Moreover, if all the basis functions $e_k(t)$ in (1.6) are continuous, satisfy (2.7), and one of the two conditions of Theorem 1 is valid then, as $n \rightarrow \infty$,

$$(2.12) \quad V_n(f) \xrightarrow{d} \sum_{i_1, \dots, i_r=1}^{\infty} f_{i_1, \dots, i_r} \tau_{i_1} \cdots \tau_{i_r},$$

where the Gaussian sequence $\{\tau_i\}$ is defined in Theorem 1.

Remark 3. It is known that, in the i.i.d. case, condition (2.2) of absolute summability of the coefficients in the series expansion of the kernel can be omitted. Recall that, in this case, summability of the coefficients squared is valid [16, 13]. In the same time, in limit theorems for the corresponding V -statistics, the latter condition does not describe the limit behavior since to define the weak limit, we need the existence of moments of the kernel on all the diagonal subspaces. For example, under the regularity conditions only (without (2.2)) of Theorem 2 for bivariate V -statistics, the assumption of finiteness of $\mathbb{E}|f(X_1, X_1)|$ is equivalent to summability of the sequence $\lambda_k \equiv f_{k,k}$ in representation (1.5), say, if all λ_k are positive. However, in the i.i.d. case, for the kernels of a bigger order, we need no summability of the coefficients f_{i_1, \dots, i_r} on the set of all pairwise distinct subscripts.

As is noted in Proposition 4 below, in the case of dependent observations, we cannot omit the above-mentioned restriction regarding summability of the coefficients f_{i_1, \dots, i_r} on the diagonal subspaces for U -statistics as well.

Proposition 4. There exist a stationary 1-dependent sequence $\{X_i\}$ satisfying condition (AC), and a canonical kernel $f(t_1, t_2) \in L_2(\mathfrak{X}^2, F^2)$ such that the weak limit of the corresponding U -statistics does not exist.

Remark 4. In [5], in the case of dependent observations, another approach was proposed for description of the limit distribution of canonical von Mises statistics as a multiple stochastic integral of the kernel under consideration, with respect to increments of a centered Gaussian process with a covariance function defined by joint distributions of the random variables $\{X_i\}$. In the i.i.d. case, such dual description of the limit law is well known (for example, see [13]).

However, the stationary sequence $\{X_i\}$ in [5] must satisfy a stronger ψ -mixing condition. In the same time, in contrast to the present results, condition (2.2) and the regularity conditions for the kernel and the basis functions of Theorem 2 were replaced in [5] with the more natural condition of integrability of the kernel on all the diagonal subspaces. Note that the above-mentioned regularity condition and (2.2) imply the boundedness of the kernel under consideration, i. e., the above-mentioned condition of the kernel integrability on the diagonal subspaces in [5] is fulfilled.

It is not clear importance of restriction (2.2) outside the diagonal subspaces (i. e., on the set of all pairwise distinct subscripts) to approximate U - and V -statistics of an arbitrary order for dependent trials.

Mention also the important particular case when the bivariate kernel of a V -statistic is represented as an inner product $f(x, y) = (x, y)$ in a separable Hilbert space. In this case, the corresponding von Mises statistic coincides with the Euclidian norm squared of a normalized sum of weakly dependent centered observations and we deal with the Central Limit Theorem (with respect to the class of all centered balls) for Hilbert-space-valued weakly dependent observations which was proved under various mixing conditions and the existence of the moment $\mathbb{E}(X_1, X_1)$ (for example, see [17]).

3. Limit theorems for m -dependent stationary sequences

Definition 5. A sequence $\{X_i\}$ is called a sequence of m -dependent random variables (or m -dependent sequence) if the two families of random variables $\{X_i; i \leq k\}$ and $\{X_i; i > k + m\}$ are independent for every natural k .

It is easy to see that a sequence of m -dependent random variables satisfies φ -mixing condition. But some sequences of this type do not satisfy the crucial restrictions of Theorem 1. Proposition 2 in Section 2 asserts that there exist φ -mixing stationary sequences satisfying all the conditions of Theorem 1 except (AC) such that the weak limit for the distributions of the U -statistics differs from that in Theorem 1. At the end of the present section, we will construct a sequence of 1-dependent random variables satisfying the conditions of Proposition 2.

The goal of the present section is to obtain the limit distributions of U - and V -statistics based on samples from a stationary sequence of m -dependent observations without additional non-classical restrictions from the previous section like (AC) or the regularity conditions on the kernels and the basis functions.

For m -dependent sequences, the random vector $\{X_{j_1}, \dots, X_{j_r}\}$ (in the case of V -statistics, multiplicities of the subscripts may exceed 1) can be divided into independent subvectors (or *blocks*) where the number of blocks can be varied from 1 to r . Let $T = \{j_{k_1}, \dots, j_{k_l}\} \subseteq \{j_1, \dots, j_r\}$ be a *well-ordered* set. We call the vector $\{X_{\pi(j_{k_1})}, \dots, X_{\pi(j_{k_l})}\}$ an *indivisible* block if $\max_{i < l} (j_{k_{i+1}} - j_{k_i}) \leq m$ and the dimension l (the *length* of the block) is the largest possible among the all subsets of $\{j_1, \dots, j_r\}$ which contain the set T ; here $\pi(\cdot)$ is an arbitrary permutation of the subscripts.

It is clear that the distribution of the vector $\{X_{j_1}, \dots, X_{j_r}\}$ (it is a sample from a stationary sequence) is defined by the distributions of its blocks. On the other hand, the distribution of a block is defined by the joint arrangement of its subscripts. Thus, the number of different distributions of $\{X_{j_1}, \dots, X_{j_r}\}$ depends only on r and m and we can divide all such subvectors into classes of identically distributed ones.

Let a random vector $\{X_{j_1}, \dots, X_{j_r}\}$ from a fixed class mentioned above can be divided into s indivisible blocks. For $s > r/2$, we assume the following restriction:

$$(3.1) \quad \mathbb{E}f^2(X_{j_1}, \dots, X_{j_r}) < \infty.$$

If $s \leq r/2$ then we assume that

$$(3.2) \quad \mathbb{E}|f(X_{j_1}, \dots, X_{j_r})| < \infty.$$

In the latter case, we actually may assume the existence of the moments of lower orders than 1 in dependence on s as in the i.i.d. case (see [13]).

As before, denote by $\{\tau_i\}$ a centered Gaussian sequence with covariance matrix (2.10). It is easy to see that, in the case of m -dependency, expression (2.10) transforms to the following:

$$(3.3) \quad M_{kl} = \mathbb{E}\tau_k\tau_l = \delta_{kl} + \sum_{j=1}^m [\mathbb{E}e_k(X_1)e_l(X_{j+1}) + \mathbb{E}e_l(X_1)e_k(X_{j+1})].$$

Recall that we admit degeneracy of some τ_i (the case $M_{ii} = 0$).

Introduce some notation.

Let $I(l_1, \dots, l_s, q)$ be the set of all ways to choose q various pairs from the elements $\{l_1, \dots, l_s\}$, where $q \leq [s/2]$ ($[\cdot]$ is the entier of a number). The number of elements of $I(l_1, \dots, l_s, q)$ equals

$$\frac{s!}{2^q(s-2q)!q!}.$$

Denote by p a concrete collection of the pairs described above:

$$p = \{\{l_{i_1}, l_{i_2}\}, \dots, \{l_{i_{2q-1}}, l_{i_{2q}}\}\} \in I(l_1, \dots, l_s, q);$$

Moreover, put

$$(3.4) \quad M_p = M_{l_{i_1}, l_{i_2}} \cdot \dots \cdot M_{l_{i_{2q-1}}, l_{i_{2q}}};$$

$$\bar{p} = \{l_1, \dots, l_s\} \setminus (\{l_{i_1}, l_{i_2}\} \cup \dots \cup \{l_{i_{2q-1}}, l_{i_{2q}}\}).$$

Fix some class K of vectors containing $s \geq [r/2] + 1$ indivisible blocks in which the length of each block does not exceed 2, and $\{X_{j_1}, \dots, X_{j_r}\} \in K$. Let $1 \leq k_1 < \dots < k_{s_1} \leq r$ be the numbers of one-element blocks. Now, define the coefficients

$$(3.5) \quad f_{i_1, \dots, i_{s_1}}^{(K)} = \mathbb{E}f(X_{j_1}, \dots, X_{j_r})e_{i_1}(X_{j_{k_1}})\dots e_{i_{s_1}}(X_{j_{k_{s_1}}}).$$

The main results of the present section are contained in the following two theorems.

Theorem 3. *Let all the classes of vectors divided into no more than $[r/2]$ indivisible blocks, satisfy (3.2). The rest of the classes satisfies (3.1). Then*

$$(3.6) \quad V_n \xrightarrow{d} \sum_{k=0}^{[(r-1)/2]} \sum_{i_1, \dots, i_{r-2k}=1}^{\infty} \sum_K f_{i_1, \dots, i_{r-2k}}^{(K)} \sum_{q=0}^{[r/2-k]} (-1)^q \sum_{p \in I(i_1, \dots, i_{r-2k}, q)} M_p \prod_{i \in \bar{p}} \tau_i + C,$$

where $\{\tau_i\}$ is a Gaussian centered sequence with covariance matrix (3.3), \sum_K is the sum over the classes in which the numbers of two-element and one-element indivisible blocks are equal to k and $r - 2k$ respectively; if r is odd then $C = 0$, otherwise,

$$C = \sum_{K'} \mathbb{E}f(X_{j'_1}, \dots, X_{j'_r}),$$

where $\sum_{K'}$ is the sum over the classes in which the number of two-element blocks equals $r/2$.

Remark 5. The sum \sum_K can be represented as follows:

$$(3.7) \quad \sum_K = \sum_{1 \leq k_{r-2k+1} < \dots < k_r \leq r} \sum_{p \in I(k_{r-2k+1}, \dots, k_r, k)} \sum_{-m \leq m_1, \dots, m_k \leq m} ;$$

and the sum $\sum_{K'}$, correspondingly, as

$$\sum_{K'} = \sum_{p' \in I(1, \dots, r, r/2)} \sum_{-m \leq m_1, \dots, m_{r/2} \leq m} .$$

Theorem 4. *Let the conditions of Theorem 3 be satisfied for the classes or blocks consisting of pairwise distinct elements. Then*

$$(3.8) \quad U_n \xrightarrow{d} \sum_{k=0}^{[(r-1)/2]} \sum_{i_1, \dots, i_{r-2k}=1}^{\infty} \sum_{K_1} f_{i_1 \dots i_{r-2k}}^{(K_1)} \sum_{q=0}^{[r/2-k]} (-1)^q \sum_{p \in I(i_1, \dots, i_{r-2k}, q)} M_p \prod_{i \in \bar{p}} \tau_i + C,$$

where τ_i are defined above, and

$$\sum_{K_1} = \sum_{1 \leq k_{r-2k+1} < \dots < k_r \leq r} \sum_{p \in I(k_{r-2k+1}, \dots, k_r, k)} \sum_{1 \leq |m_1|, \dots, |m_k| \leq m}$$

if r is odd then $C = 0$; otherwise,

$$C = \sum_{K'} \mathbb{E} f(X_{j'_1}, \dots, X_{j'_r}),$$

where

$$\sum_{K'_1} = \sum_{p' \in I(1, \dots, r, r/2)} \sum_{1 \leq |m_1|, \dots, |m_{r/2}| \leq m} .$$

4. Proofs of Theorems 3 and 4

Proof of Theorem 3. First, divide all the summands in (1.2) into classes of identically distributed ones (or the vectors $\{X_{j_1}, \dots, X_{j_r}\}$), where the number of these classes does not depend on n , and find the limit distribution of the normalized sum over the vectors from a fixed class.

Let the vectors $\{X_{j_1}, \dots, X_{j_r}\}$ from this class are divided into s indivisible blocks. It is worth noting, that the number of all vectors from this class does not exceed n^s and is equivalent to n^s as $n \rightarrow \infty$. In this case, we actually deal with statistics of the form

$$n^{-r/2} \sum_{*} \tilde{f}(\bar{X}_{i_1}^{(1)}, \dots, \bar{X}_{i_s}^{(s)}),$$

where the summation is taken over all the vectors of the chosen class, and for every fixed subscripts i_1, \dots, i_s , the random subvectors $\bar{X}_{i_1}^{(1)}, \dots, \bar{X}_{i_s}^{(s)}$ are independent identically distributed blocks defined by the class under consideration. In other words, we actually deal with the decoupled-type V -statistics based on a finite collection of independent stationary sequences each of them consists of k_l -dependent random subvectors, $l = 1, \dots, s$ (k may be greater than m), with a personal margin distribution.

Then, due to (3.2) and the law of large numbers for such statistics which is easily proved, we have for $s = r/2$

$$(4.1) \quad n^{-r/2} \sum_* f(X_{j_1}, \dots, X_{j_r}) \xrightarrow{P} \mathbb{E}f(X_{j_1^0}, \dots, X_{j_r^0})$$

as $n \rightarrow \infty$; here $\{X_{j_1^0}, \dots, X_{j_r^0}\}$ represents the class under consideration. If $s < r/2$ then, evidently,

$$(4.2) \quad n^{-r/2} \sum_* f(X_{j_1}, \dots, X_{j_r}) \xrightarrow{P} 0$$

as $n \rightarrow \infty$. Thus, we can see that the classes with $s < r/2$ are not essential for the statistic's weak limit. Therefore, we can add and subtract their summands in our calculations without changing the limit expression.

Consider the case of even r and $s = r/2$. If the length of some block is greater than 2 then there exists at least one another block with the length 1. Then, due to the independence of different blocks and the canonical property of the kernel,

$$\mathbb{E}f(X_{j_1^0}, \dots, X_{j_r^0}) = 0.$$

Therefore the non-zero limit above can be obtained if only the lengths of all indivisible blocks are equal to 2.

So, consider the case when it is possible to divide $\{X_{j_1}, \dots, X_{j_r}\}$ into $s > r/2$ indivisible blocks:

$$\{X_{j_1}, \dots, X_{j_r}\} = \{\overline{X}_1, \dots, \overline{X}_s\},$$

where $\{\overline{X}_1, \dots, \overline{X}_s\}$ are indivisible independent blocks with the relative lengths k_1, \dots, k_s . From condition (3.1) and the same argument as in the i.i.d. case, the kernel can be represented as a mean-square converging series

$$(4.3) \quad f(X_{j_1}, \dots, X_{j_r}) = \sum_{l_1, \dots, l_s=0}^{\infty} f_{l_1, \dots, l_s} e_{l_1}^{(1)}(\overline{X}_1) \dots e_{l_s}^{(s)}(\overline{X}_s),$$

where $\{e_{l_i}^{(i)}\}$ is an orthonormal basis of the corresponding space L_2 with $e_0^{(i)} \equiv 1$. Due to the orthogonality with e_0 , we have $\mathbb{E}e_{l_i}^{(i)}(\overline{X}_i) = 0$ for all $l_i > 0$. As was already noted, the number of different distributions of indivisible blocks depends only on r and m . So, for identically distributed blocks in different classes, choose the same basis $\{e_{l_i}^{(i)}\}$. In particular, if $k_i = 1$, let the corresponding basis coincide with $\{e_l(t)\}$ introduced before. Notice that, in this case, by the canonical property, we have $f_{l_1, \dots, l_s} = 0$ if $l_i = 0$ in (4.3) ([13], [6]).

If the distance between the closest subscripts of two or more blocks is less than $m + 1$ then they form another indivisible block. Let $\tilde{X}(j)$ be an indivisible block coinciding with one of \overline{X}_i , or formed by several ones, with the length k , $1 \leq k \leq rm$, and the minimal subscript equal j . Correspondingly, let \tilde{e} be a fixed basis element $e_{l_i}^{(i)}$ or a product of several ones. We now find the weak limit as $n \rightarrow \infty$ of the expression

$$(4.4) \quad n^{-k/2} \sum_{j=1}^{n-k} \tilde{e}(\tilde{X}(j)).$$

First, if $k = 1$, $\tilde{e}(\tilde{X}(j)) = e_i(X_j)$. Moreover, in view of the argument above, we need to consider only the case $i \geq 1$. In [6], it was shown that the sequence

$$(4.5) \quad \left\{ n^{-1/2} \sum_{j=1}^n e_1(X_j), n^{-1/2} \sum_{j=1}^n e_2(X_j), \dots \right\}$$

converges in distribution to a centered Gaussian sequence $\{\tau_i\}$ with covariance matrix (3.3).

If $k = 2$ then we can apply the law of large numbers. Notice that if $\tilde{e}(\tilde{X}(j)) = e_{l_i}^{(i)}(\overline{X}_i)$ with $k_i = 2$ then we obtain the non-zero limit in probability only in the case $l_i = 0$. Thus,

$$n^{-1} \sum_{j=1}^{n-k} \tilde{e}(\tilde{X}(j)) \xrightarrow{P} 1.$$

We now show that, in the case $k \geq 2$, the normalized sum in (4.4) converges to zero in probability. By definition, $\tilde{X}(j)$ is formed by one or several blocks \overline{X}_i . The convergence required can be proved by showing that, for any i , as $n \rightarrow \infty$, the following assertion is valid:

$$\max_{1 \leq j \leq n} n^{-1/2} |e_{l_i}^{(i)}(\overline{X}_i)| \xrightarrow{P} 0,$$

where j is the minimal index of the block. The last assertion holds if for an arbitrary $\varepsilon > 0$ $\mathbb{P}(|e_{l_i}^{(i)}(\overline{X}_i)| > \varepsilon n^{1/2}) = o(1/n)$ or, otherwise, $\mathbb{P}(|e_{l_i}^{(i)}(\overline{X}_i)| > x) = o(1/x^2)$ as $x \rightarrow \infty$. By construction $e_{l_i}^{(i)}(\overline{X}_i)$ has the finite second moment, and the property required is fulfilled.

Now, find the weak limit of the expression

$$(4.6) \quad n^{-r/2} \sum_* e_{l_1}^{(1)}(\overline{X}_1) \cdots e_{l_s}^{(s)}(\overline{X}_s),$$

where \sum_* , as before, means the summation over all the vectors from the fixed class. By adding and subtracting some summands, one can represent (4.6) as a linear combination of products of values (4.4). In view of the reasoning above, we consider only the classes where there are one- and two-element blocks only, and $l_i \geq 1$ in (4.6) if the length of \overline{X}_i equals 1, and $l_i = 0$ in the case $k_i = 2$. Assume that we have s_1 one-element and $(r - s_1)/2$ two-element blocks.

It was shown that expression (4.4) converges in probability to a constant or coincides with some element of sequence (4.5). Since expression (4.6) is a linear combination of products of sums (4.4), the aforesaid allows to obtain the limit in distribution of (4.6) changing in this linear combination all values (4.4) for their weak limits. We now find this limit.

As two-element blocks correspond only to zero basis element, they have effect only on the number of the same summands in (4.6), with one-element blocks fixed. It is easy to prove that this number is equivalent to $n^{(r-s_1)/2}$ as $n \rightarrow \infty$. Thus, taking into account finiteness of the expectations of the summands, the expression in (4.6) is equivalent to the following:

$$n^{-s_1/2} \sum_{|u_i - u_j| > m, 1 \leq i \neq j \leq s_1} e_{l'_1}(X_{u_1}) \cdots e_{l'_{s_1}}(X_{u_{s_1}}) =: g_{s_1}(l'_1, \dots, l'_{s_1}).$$

Lemma 1. As $n \rightarrow \infty$,

$$(4.7) \quad g_{s_1}(l_1, \dots, l_{s_1}) \xrightarrow{d} \sum_{q=0}^{\lfloor s_1/2 \rfloor} (-1)^q \sum_{p \in I(l_1, \dots, l_{s_1}, q)} M_p \prod_{i \in \bar{p}} \tau_i,$$

where $\{\tau_i\}$ is a centered Gaussian sequence with covariance matrix (3.3).

Proof. It is not difficult to obtain the next recurrent formula for $g_{s_1}(l_1, \dots, l_{s_1})$:

$$(4.8) \quad \begin{aligned} & g_{s_1}(l_1, \dots, l_{s_1}) \\ &= n^{-1/2} \sum_{j=1}^n e_{l_1}(X_j) g_{s_1-1}(l_2, \dots, l_{s_1}) \\ &\quad - \sum_{i=2}^{s_1} g_{s_1-2}(l_2, \dots, l_{i-1}, l_{i+1}, \dots, l_{s_1}) n^{-1} \\ &\quad \times \left(\sum_{j=1}^n e_{l_1}(X_j) e_{l_i}(X_j) + \sum_{k=1}^m \sum_{j=1}^{n-k} (e_{l_1}(X_j) e_{l_i}(X_{j+k}) + e_{l_1}(X_{j+k}) e_{l_i}(X_j)) \right) \\ &\quad + o_P(1). \end{aligned}$$

As $n \rightarrow \infty$, we obtain the following formula for the limits:

$$(4.9) \quad \begin{aligned} \lim g_{s_1}(l_1, \dots, l_{s_1}) &= \tau_{l_1} \lim g_{s_1-1}(l_2, \dots, l_{s_1}) \\ &\quad - \sum_{i=2}^{s_1} M_{l_1 l_i} \lim g_{s_1-2}(l_2, \dots, l_{i-1}, l_{i+1}, \dots, l_{s_1}). \end{aligned}$$

Prove (4.7) by induction on s_1 . It is easy to see that, for $s_1 = 1$ and $s_1 = 2$, the statement of the lemma holds (suppose that $g_0 \equiv 1$). Let it hold for all $s_2 < s_1$. Then

$$\begin{aligned} & \lim g_{s_1}(l_1, \dots, l_{s_1}) \\ &= \tau_{l_1} \sum_{q=0}^{\lfloor (s_1-1)/2 \rfloor} (-1)^q \sum_{p \in I(l_2, \dots, l_{s_1}, q)} M_p \prod_{i \in \bar{p}} \tau_i \\ &\quad - \sum_{i=2}^{s_1} M_{l_1 l_i} \sum_{q=0}^{\lfloor (s_1-2)/2 \rfloor} (-1)^q \sum_{p \in I(l_2, \dots, l_{i-1}, l_{i+1}, \dots, l_{s_1}, q)} M_p \prod_{j \in \bar{p}} \tau_j. \end{aligned}$$

For $q \leq \lfloor (s_1 - 1)/2 \rfloor$, the set $I(l_1, \dots, l_{s_1}, q)$ consists of elements of $I(l_2, \dots, l_{s_1}, q)$ and elements like $\{l_1, l_i\} \cup p$, where $p \in I(l_2, \dots, l_{i-1}, l_{i+1}, \dots, l_{s_1}, q-1)$; in the case of even s_1 and $q = s_1/2$, it consists of elements of the second type only. In view of the aforesaid, we proved the statement. \square

Lemma 2. Let all the vectors of the fixed class are divided into $s \geq \lceil r/2 \rceil + 1$ indivisible blocks in which there are s_1 one-element blocks. Then

1) if there exists a block with the length more than 2 then

$$(4.10) \quad n^{-r/2} \sum_{1 \leq j_1, \dots, j_r \leq n: \{X_{j_1}, \dots, X_{j_r}\} \in K} f(X_{j_1}, \dots, X_{j_r}) \xrightarrow{P} 0;$$

2) if the lengths of all blocks do not exceed 2 then

$$(4.11) \quad n^{-r/2} \sum_{1 \leq j_1, \dots, j_r \leq n: \{X_{j_1}, \dots, X_{j_r}\} \in K} f(X_{j_1}, \dots, X_{j_r}) \\ \xrightarrow{d} \sum_{i_1, \dots, i_{s_1}=1}^{\infty} f_{i_1, \dots, i_{s_1}}^{(K)} \sum_{q=0}^{\lfloor s_1/2 \rfloor} (-1)^q \sum_{p \in I(i_1, \dots, i_{s_1}, q)} M_p \prod_{i \in \bar{p}} \tau_i,$$

where the series on the right-hand side of (4.11) mean-square converges.

Proof. In view of (4.3),

$$(4.12) \quad n^{-r/2} \sum_{1 \leq j_1, \dots, j_r \leq n: \{X_{j_1}, \dots, X_{j_r}\} \in K} f(X_{j_1}, \dots, X_{j_r}) \\ = \sum_{l_1, \dots, l_s=0}^{\infty} f_{l_1, \dots, l_s} n^{-r/2} \sum e_{l_1}^{(1)}(\bar{X}_1) \cdots e_{l_s}^{(s)}(\bar{X}_s) \\ = \sum_{l_1, \dots, l_s=0}^N f_{l_1, \dots, l_s} n^{-r/2} \sum e_{l_1}^{(1)}(\bar{X}_1) \cdots e_{l_s}^{(s)}(\bar{X}_s) \\ + \sum_{\max(l_1, \dots, l_s) > N} f_{l_1, \dots, l_s} n^{-r/2} \sum e_{l_1}^{(1)}(\bar{X}_1) \cdots e_{l_s}^{(s)}(\bar{X}_s).$$

If there exists a block with the length greater than 2 then from the arguments above the first summand on the right-hand side of (4.12) converges to zero in probability. Otherwise, from Lemma 1 it follows that

$$\sum_{l_1, \dots, l_s=0}^N f_{l_1, \dots, l_s} n^{-r/2} \sum e_{l_1}^{(1)}(\bar{X}_1) \cdots e_{l_s}^{(s)}(\bar{X}_s) \\ \xrightarrow{d} \sum_{i_1, \dots, i_{s_1}=1}^N f_{i_1, \dots, i_{s_1}}^{(K)} \sum_{q=0}^{\lfloor s_1/2 \rfloor} (-1)^q \sum_{p \in I(i_1, \dots, i_{s_1}, q)} M_p \prod_{i \in \bar{p}} \tau_i.$$

We now prove that the second summand on the right-hand side of (4.12) mean-square converges to zero as $N \rightarrow \infty$. For the convenience, denote different vectors from the class K and the corresponding blocks by $\bar{X}^{(i)}$ and $\bar{X}_j^{(i)}$, respectively. We have

$$(4.13) \quad \mathbb{E}(n^{-r/2} \sum_{\max(l_1, \dots, l_s) > N} f_{l_1, \dots, l_s} e_{l_1}^{(1)}(\bar{X}_1) \cdots e_{l_s}^{(s)}(\bar{X}_s))^2 \\ = n^{-r} \sum_{\bar{X}^{(1)}, \bar{X}^{(2)} \in K} \sum_{\max\{l_k\}, \max\{i_k\} > N} f_{l_1, \dots, l_s} f_{i_1, \dots, i_s} \\ \times \mathbb{E} e_{l_1}^{(1)}(\bar{X}_1^{(1)}) \cdots e_{l_s}^{(s)}(\bar{X}_s^{(1)}) e_{i_1}^{(1)}(\bar{X}_1^{(2)}) \cdots e_{i_s}^{(s)}(\bar{X}_s^{(2)}).$$

Here we can permute the signs of the expectation and the sum due to the mean-square convergence of the corresponding series.

Note that, if there are $s \geq [r/2] + 1$ indivisible blocks in the class then there are among them at least $2s - r$ one-element blocks. Thus, among the blocks $\overline{X}_1^{(1)}, \dots, \overline{X}_s^{(2)}$, there are no more than $2r - 2s$ blocks with the lengths greater than 1. If there exists at least one such one-element block that the difference between its subscript and the closest one in the class under consideration is greater than m , then the expectation on the right-hand side of (4.13) vanishes (for all subscripts $l_1, \dots, l_s, i_1, \dots, i_s$). Then, for the non-zero summands, the distance between each one-element block and the neighbor one is not greater than m . The number of ways to choose $\overline{X}^{(1)}$ and $\overline{X}^{(2)}$ satisfying this condition, is equivalent to

$$C(r, m)n^{2r-2s}n^{2s-r} = C(r, m)n^r,$$

where $C(r, m)$ is a constant which depends only on r and m . And for fixed $\overline{X}^{(1)}$ and $\overline{X}^{(2)}$,

$$\begin{aligned} \mathbb{E} \sum_{\max\{l_k\}, \max\{i_k\} > N} f_{l_1, \dots, l_s} f_{i_1, \dots, i_s} e_{l_1}^{(1)}(\overline{X}_1^{(1)}) \cdots e_{l_s}^{(s)}(\overline{X}_s^{(1)}) e_{i_1}^{(1)}(\overline{X}_1^{(2)}) \cdots e_{i_s}^{(s)}(\overline{X}_s^{(2)}) \\ \leq \mathbb{E}^{1/2} \left(\sum_{\max\{l_k\} > N} f_{l_1, \dots, l_s} e_{l_1}^{(1)}(\overline{X}_1^{(1)}) \cdots e_{l_s}^{(s)}(\overline{X}_s^{(1)}) \right)^2 \\ \times \mathbb{E}^{1/2} \left(\sum_{\max\{i_k\} > N} f_{i_1, \dots, i_s} e_{i_1}^{(1)}(\overline{X}_1^{(2)}) \cdots e_{i_s}^{(s)}(\overline{X}_s^{(2)}) \right)^2 = \sum_{\max\{l_k\} > N} f_{l_1, \dots, l_s}^2. \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{E}(n^{-r/2} \sum_{\max\{l_1, \dots, l_s\} > N} f_{l_1, \dots, l_s} \sum e_{l_1}^{(1)}(\overline{X}_1) \cdots e_{l_s}^{(s)}(\overline{X}_s))^2 \\ \leq C(r, m) \sum_{\max\{l_k\} > N} f_{l_1, \dots, l_s}^2 \rightarrow 0 \end{aligned}$$

as $N \rightarrow \infty$, which required to be proved.

Now, show that the series in (4.11) mean-square converges.

Let $\{k_1, \dots, k_{s_1}\}$ be some permutation of $\{1, \dots, s_1\}$. First, prove that

$$\sum_{i_1, \dots, i_{s_1}, i'_1, \dots, i'_{s_1} = 1}^{\infty} |f_{i_1, \dots, i_{s_1}} f_{i'_1, \dots, i'_{s_1}}| |M_{i_1 i'_{k_1}} \cdots M_{i_{s_1} i'_{k_{s_1}}}| < \infty.$$

Due to the definition of M_{kl} , it is sufficient to show that, for arbitrary j_1, \dots, j_{s_1} ,

$$\begin{aligned} \sum_{i_1, \dots, i_{s_1}, i'_1, \dots, i'_{s_1} = 1}^{\infty} |f_{i_1, \dots, i_{s_1}} f_{i'_1, \dots, i'_{s_1}}| \\ \times |\mathbb{E} e_{i_1}(X_1) e_{i'_{k_1}}(X_{j_1+1})| \cdots |\mathbb{E} e_{i_{s_1}}(X_1) e_{i'_{k_{s_1}}}(X_{j_{s_1}+1})| < \infty. \end{aligned}$$

Fix all the subscripts except $i_{s_1}, i'_{k_{s_1}}$, and sum only over the last ones:

$$\begin{aligned} |\mathbb{E} e_{i_1}(X_1) e_{i'_{k_1}}(X_{j_1+1})| \cdots |\mathbb{E} e_{i_{s_1-1}}(X_1) e_{i'_{k_{s_1-1}}}(X_{j_{s_1-1}+1})| \\ \times \sum_{i_{s_1}, i'_{k_{s_1}} = 1}^{\infty} |f_{i_1, \dots, i_{s_1}} f_{i'_1, \dots, i'_{s_1}}| |\mathbb{E} e_{i_{s_1}}(X_1) e_{i'_{k_{s_1}}}(X_{j_{s_1}+1})| \end{aligned}$$

$$\begin{aligned} &\leq |\mathbb{E}e_{i_1}(X_1)e_{i'_{k_1}}(X_{j_1+1})| \dots |\mathbb{E}e_{i_{s_1-1}}(X_1)e_{i'_{k_{s_1-1}}}(X_{j_{s_1-1}+1})| \\ &\quad \times \left(\sum_{i_{s_1}=1}^{\infty} f_{i_1, \dots, i_{s_1}}^2 \right)^{1/2} \left(\sum_{i'_{k_{s_1}}=1}^{\infty} f_{i'_1, \dots, i'_{s_1}}^2 \right)^{1/2}. \end{aligned}$$

Similarly, summing over each pair of indices, we obtain

$$\begin{aligned} &\sum_{i_1, \dots, i_{s_1}, i'_1, \dots, i'_{s_1}=1}^{\infty} |f_{i_1, \dots, i_{s_1}} f_{i'_1, \dots, i'_{s_1}}| |M_{i_1 i'_{k_1}} \dots M_{i_{s_1} i'_{k_{s_1}}}| \\ &\leq (2m+1)^{s_1} \sum_{i_1, \dots, i_{s_1}=1}^{\infty} f_{i_1, \dots, i_{s_1}}^2. \end{aligned}$$

Now, consider the second moment of the difference between two partial sums of the series from (4.11):

$$\begin{aligned} &\sum_{\max\{i_1, \dots, i_{s_1}\}, \max\{i'_1, \dots, i'_{s_1}\}=N}^M f_{i_1, \dots, i_{s_1}}^{(K)} f_{i'_1, \dots, i'_{s_1}}^{(K)} \sum_{q, q'=0}^{\lfloor s_1/2 \rfloor} (-1)^{q+q'} \\ &\quad \times \sum_{p \in I(i_1, \dots, i_{s_1}, q)} \sum_{p' \in I(i'_1, \dots, i'_{s_1}, q')} M_p M_{p'} \mathbb{E} \prod_{i \in \bar{p}} \tau_i \prod_{i' \in \bar{p}'} \tau_{i'}. \end{aligned}$$

We now prove that

$$\begin{aligned} &\sum_{q, q'=0}^{\lfloor s_1/2 \rfloor} (-1)^{q+q'} \sum_{p \in I(i_1, \dots, i_{s_1}, q)} \sum_{p' \in I(i'_1, \dots, i'_{s_1}, q')} M_p M_{p'} \mathbb{E} \prod_{i \in \bar{p}} \tau_i \prod_{i' \in \bar{p}'} \tau_{i'} \\ (4.14) \quad &= \sum_{\{k_1, \dots, k_{s_1}\}} M_{i_1 i'_{k_1}} \dots M_{i_{s_1} i'_{k_{s_1}}}, \end{aligned}$$

where the summation is over all different permutations of $\{1, \dots, s_1\}$. Use the next property of the product of the Gaussian random variables:

$$\mathbb{E} \prod_{i \in \bar{p}} \tau_i \prod_{i' \in \bar{p}'} \tau_{i'} = \sum_{p_0 \in I(\bar{p}, \bar{p}', 2s_1 - 2q - 2q')} M_{p_0},$$

thus,

$$\begin{aligned} &\sum_{q, q'=0}^{\lfloor s_1/2 \rfloor} (-1)^{q+q'} \sum_{p \in I(i_1, \dots, i_{s_1}, q)} \sum_{p' \in I(i'_1, \dots, i'_{s_1}, q')} M_p M_{p'} \mathbb{E} \prod_{i \in \bar{p}} \tau_i \prod_{i' \in \bar{p}'} \tau_{i'} \\ &= \sum_{q, q'=0}^{\lfloor s_1/2 \rfloor} (-1)^{q+q'} \sum_{p \in I(i_1, \dots, i_{s_1}, q)} \sum_{p' \in I(i'_1, \dots, i'_{s_1}, q')} \sum_{p_0 \in I(\bar{p}, \bar{p}', 2s_1 - 2q - 2q')} M_p M_{p'} M_{p_0}. \end{aligned}$$

It is easy to see, that the right-hand side of (4.14) corresponds to the part of summands in the sum above with $q = 0$ and $q' = 0$ because the summands of this right-hand side consist of pairs with one subscript from $\{i_1, \dots, i_{s_1}\}$ and another from $\{i'_1, \dots, i'_{s_1}\}$. Consider the summands with $q = 0$ and $q' = 0$ which are not contained in the right-hand side of (4.14). Evidently, they correspond to different

$p_0 \in I(i_1, \dots, i_{s_1}, i'_1, \dots, i'_{s_1}, 2s_1)$ including at least one pair with both subscripts from $\{i_1, \dots, i_{s_1}\}$ or $\{i'_1, \dots, i'_{s_1}\}$. Denote all such pairs by p_1, \dots, p_N , where $N = 2C_{s_1}^2 = s_1(s_1 - 1)$. We have

$$\sum_{p_0 \in I(i_1, \dots, i'_{s_1}, 2s_1) : p_j \in p_0, 1 \leq j \leq N} M_{p_0}.$$

Note that if different pairs p_k and p_l have a common subscript then the fixed p_0 contains only one of them. Using the inclusion-exclusion formula and the formula for the expectation of the product of the Gaussian variables also, we can write the sum above as follows:

$$\begin{aligned} & \sum_{j=1}^N M_{p_j} \mathbb{E} \prod_{i \in \{i_1, \dots, i'_{s_1}\} / p_j} \tau_i - \sum_{k < l : p_k \cap p_l = \emptyset} M_{p_k} M_{p_l} \mathbb{E} \prod_{i \in \{i_1, \dots, i'_{s_1}\} / (p_k \cup p_l)} \tau_i \\ & + \dots - \sum_{k_1 < \dots < k_{\lfloor s_1/2 \rfloor} : p_{k_i} \cap p_{k_j} = \emptyset, i \neq j} M_{p_{k_1}} \dots M_{p_{k_{\lfloor s_1/2 \rfloor}}} \mathbb{E} \prod_{i \in \{i_1, \dots, i'_{s_1}\} / (p_{k_1} \cup \dots \cup p_{k_{\lfloor s_1/2 \rfloor})} \tau_i. \end{aligned}$$

Note that this decomposition coincides with that part of the left-hand side of (4.14) where at least one of q and q' is positive, with the opposite sign. Thus, equality (4.14) is proved. Then the sequence of partial sums of the series from (4.11) is a Cauchy sequence in L_2 , and, consequently, the series from (4.11) mean-square converges. \square

As was mentioned above, all the collections $\{X_{j_1}, \dots, X_{j_r}\}$ are divided into disjoint classes of identically distributed vectors where the number of classes depend only on r and m . So, every V -statistic can be divided into normalized sums of $f(X_{j_1}, \dots, X_{j_r})$ over the vectors from a certain class introduced above.

The limits in distribution of such normalized sums are obtained in Lemma 2, in (4.1), and (4.2). Due to the arguments at the beginning of the proof, to get the weak limit of the statistic we can sum the limit expressions on the right-hand sides of (4.11) and (4.1) over the corresponding classes. Theorem 3 is proved. \square

Proof of Theorem 4. The proof repeats the proof of Theorem 3. In the case of U -statistics, the only difference is that we consider only the classes of vectors with pairwise disjoint elements. Then, in two-element blocks, the distance between the observation subscripts cannot be equal to zero. The last fact explains the expressions for \sum_{K_1} and $\sum_{K'_1}$. Theorem 4 is proved. \square

Proof of Proposition 2. Let $\{Y_i; i \geq 1\}$ be a sequence of independent random variables uniformly distributed on $[0, 1]$, and let $\{\xi_i; i \geq 1\}$ be a sequence of independent symmetric Bernoulli random variables which are independent of $\{Y_i\}$ as well. Set $X_i = Y_{i+\xi_i}$. The random variables $\{X_i\}$ form a stationary 1-dependent sequence. Notice that, in this case, the random variables X_i are uniformly distributed on $[0, 1]$ as well. Thus, the stationary sequence $\{X_i\}$ satisfies φ -mixing condition and the restriction in (2.1). It is clear that, for the corresponding independent copies X_1^* and X_2^* , we have $P(X_1^* = X_2^*) = 0$, but for the originals, we obtain

$$P(X_1 = X_2) = P(\xi_1 = 1)P(\xi_2 = 0) = 1/4.$$

Notice that, in the example under consideration, the basis functions and the coefficients in the series expansion in (1.6) do not depend on the values of the kernel f on the diagonal due to continuity of the marginal distribution since, in this case, the characteristic equation has the form

$$\lambda_k e_k(t) = \int_0^1 f(t, s) e_k(s) ds.$$

For simplicity, let

$$\sup_{0 \leq t_1, t_2 \leq 1} |f(t_1, t_2)| \leq 1.$$

Now, change the diagonal values setting $f(t, t) \equiv 1 + \beta$ for all $t \in [0, 1]$, where $\beta > 0$. Then $P(f(X_1, X_2) = 1 + \beta) = 1/4$. In the same time, the series on the right-hand side of (1.6) does not depend on β . It is easy to show that, in the example under consideration, the limit law will essentially differs from (2.3). Indeed, using the series expansion (1.5) and the result proved above we obtain the form of the limit law:

$$\begin{aligned} U_n &\xrightarrow{d} 2\mathbb{E}f(X_1, X_2) + \sum_{k \geq 1} \lambda_k (\tau_k^2 - 3/2) \\ &= \mathbb{E}f(Y_1, Y_1)/2 + \sum_{k \geq 1} \lambda_k (\tau_k^2 - 3/2) \end{aligned}$$

due to the degeneracy of f and the relation $\mathbb{E}(f(X_1, X_2) | \xi_2 - \xi_1 = 0) = 0$. The right-hand side coincides with (2.3), say, under the conditions of Theorem 2, but does not coincide under the above-mentioned restriction on the diagonal values of f (for example, if $1 + \beta > \sum_{k \geq 1} \lambda_k$). Proposition 2 is proved. \square

5. Exponential inequalities for the distribution tails of U - and V -statistics

For independent observations $\{X_i\}$, we give below a brief review of results directly connected with the subject of the present paper. In this connection, we would like to mention the results in [4, Theorem 1], [2, Proposition 2.2], [1, Theorem 7, Corollary 3], and [9, Theorem 3.3].

One of the first papers where exponential inequalities for the distribution tails of U -statistics are obtained, is the article by W. Hoeffding [10] although he considered nondegenerated U -statistics only. In this case, the value $(n - r)!/n!$ equivalent to n^{-r} as $n \rightarrow \infty$, is used as the normalizing factor instead of $n^{-r/2}$. In [10], the following statement was proved:

$$(5.1) \quad \mathbb{P}(U - \mathbb{E}U \geq t) \leq e^{-2kt^2/(b-a)^2},$$

where

$$U = (n - r)!/n! \sum_{(j_1, \dots, j_r) \in I_n^r} f(X_{j_1}, \dots, X_{j_r}),$$

$a \leq f(t_1, \dots, t_r) \leq b$ and $k = [n/r]$. In the case $r = 1$, inequality (5.1) is usually called *Hoeffding's inequality* for sums of independent identically distributed bounded random variables. Notice that, in this case, the sums mentioned may be simultaneously considered as canonical or nondegenerate U -statistics.

In [4], an improvement of (5.1) was obtained for the case when there exists a splitting majorant of the canonical kernel under consideration:

$$(5.2) \quad |f(t_1, \dots, t_r)| \leq \prod_{i \leq r} g(t_i),$$

and the function $g(t)$ satisfies the condition

$$\mathbb{E}g(X_1)^k \leq \sigma^2 L^{k-2} k! / 2$$

for all $k \geq 2$. In this case, the following analogue of Bernstein's inequality holds:

$$(5.3) \quad \mathbb{P}(|V_n| \geq t) \leq c_1 \exp\left(-\frac{c_2 t^{2/r}}{\sigma^2 + Lt^{1/r} n^{-1/2}}\right),$$

where the constants c_1 and c_2 depend only on r . Moreover, as noted in [4], inequality (5.3) cannot be improved in a sense.

It is clear that if $\sup_{t_i} |f(t_1, \dots, t_r)| = B < \infty$ then one can set in (5.3) $\sigma = L = B^{1/r}$. Then it suffices to consider only the deviation zone $|t| \leq Bn^{r/2}$ (otherwise, the left-hand side of (5.3) vanishes). Therefore, for all $t \geq 0$, inequality (5.3) yields the upper bound

$$(5.4) \quad \mathbb{P}(|V_n| \geq t) \leq c_1 \exp\left(-\frac{c_2}{2}(t/B)^{2/r}\right)$$

which is an analogue of Hoeffding's inequality (5.1).

In [2], an inequality close to (5.3) is proved without condition (5.2), and relation (5.4) is given as a consequence. In [9], some refinement of (5.4) is obtained for $r = 2$, and in [1], the later result was extended to canonical U -statistics of an arbitrary order. The goal of the present section is to extend inequality (5.4) to the case of stationary random variables under φ -mixing. For dependent observations, we do not yet know how to get more precise inequalities close to Bernstein's inequality (5.3), for unbounded kernels under some moment restrictions only.

As in the previous sections, we denote by $\{e_i(t)\}$ an orthonormal basis of the separable Hilbert space $L_2(\mathfrak{X}, F)$, with $e_0(t) \equiv 1$. Consider only the spaces which have a bounded basis:

$$(5.5) \quad \sup_{i,t} |e_i(t)| \leq C.$$

Here, we consider only stationary sequences $\{X_j\}$ satisfying φ -mixing condition.

Introduce some additional restrictions on the mixing coefficient and the kernels of the statistics under consideration.

1. **(A)** $\sum_{i_1, \dots, i_r=1}^{\infty} |f_{i_1, \dots, i_r}| < \infty$ and $\varphi(k) \leq c_0 e^{-c_1 k^2}$, where $c_1 > 0$.
2. **(B)** There exists $\varepsilon > 0$ such that $\sum_{i_1, \dots, i_r=1}^{\infty} |f_{i_1, \dots, i_r}|^{1-\varepsilon} = c < \infty$, and

$$\varphi := \sum \varphi(k) < \infty.$$

The following two theorems were proved in [7].

Theorem 5. *Let a canonical kernel $f(t_1, \dots, t_r)$ be continuous (in every argument) everywhere on \mathfrak{X}^r and let condition (5.5) be fulfilled. Moreover, if $e_k(t)$ are continuous and one of conditions (A) or (B) is fulfilled then the following inequality holds:*

$$(5.6) \quad \mathbb{P}(|V_n| > x) \leq C_1 \exp\{-C_2 x^{2/r}/B(f)\},$$

where $C_2 > 0$ depends only on $\varphi(\cdot)$; in case (A),

$$B(f) := \left(C^r \sum_{i_1, \dots, i_r=1}^{\infty} |f_{i_1, \dots, i_r}| \right)^{2/r},$$

and in case (B),

$$B(f) := C^2 \left(\sum_{i_1, \dots, i_r=1}^{\infty} |f_{i_1, \dots, i_r}|^{1-\varepsilon} \right)^{\frac{2}{r(1-\varepsilon)}};$$

here the constant C is defined in (5.5).

Remark 6. Under condition (A), we may set $C_1 = 1$. Under condition (B), the value C_1 depends on the constants r, ε, c , and C . The dependence on the values c and C can be removed by considering “large enough” values of x , namely, satisfying the inequality

$$x^{2/r} \geq \varepsilon^{-1} 8r(1-\varepsilon) e C^2 \varphi c^{2/(r(1-\varepsilon))}.$$

The following theorem is an analogue of statement (5.6) for U -statistics.

Theorem 6. *Let the sequence X_1, X_2, \dots satisfy the following condition:*

(AC) *For every collection of pairwise distinct subscripts j_1, \dots, j_r , the distribution of $(X_{j_1}, \dots, X_{j_r})$ is absolutely continuous with respect to the distribution of (X_1^*, \dots, X_r^*) .*

Moreover, if the basis $\{e_i(t)\}_{i \geq 0}$ satisfies restriction (5.5) and one of conditions (A) or (B) is valid then

$$(5.7) \quad \mathbb{P}(|U_n| > x) \leq C_1 \exp \left\{ -C_2 x^{2/r} / B(f) \right\},$$

where, under condition (A), the constants C_1 and C_2 are the same as in Theorem 5, and under condition (B), the constant C_1 depends on r, ε, c , and C , and the constant C_2 depends on φ and r ; the value $B(f)$ is defined in Theorem 5.

We now show that, for m -dependent trials, it is possible to obtain inequality (5.4) without any additional restrictions like **(AC)**.

Theorem 7. *For any stationary m -dependent trials, inequality (5.4) is valid.*

Proof. We use the following Chebyshev-type inequality which is the standard tool to derive exponential inequalities for multilinear forms of random variables (independent or not) (see [4], [9], [14], etc.):

$$(5.8) \quad \mathbb{P}(|V_n| \geq x) \leq \min_N \mathbb{E} V_n^{2N} x^{-2N}.$$

Let $|f(t_1, \dots, t_r)| \leq B$. First we prove that

$$(5.9) \quad \mathbb{E} V_n^{2N} \leq (KB^{2/r} rN)^{rN}.$$

We start with the i.i.d. case. There are the following two cases:

(I) $n \leq rN$. We have

$$\mathbb{E} V_n^{2N} = n^{-rN} \sum_{1 \leq j_1, \dots, j_{2rN} \leq n} \mathbb{E} f(X_{j_1}, \dots, X_{j_r}) \dots f(X_{j_{2rN-r+1}}, \dots, X_{j_{2rN}})$$

$$(5.10) \quad \leq n^{-rN} B^{2N} n^{2rN} = B^{2N} n^{rN} \leq B^{2N} (rN)^{rN}.$$

(II) $n > rN$. If the multiplicity of certain subscript among j_1, \dots, j_{2rN} equals 1 then the corresponding mixed moment in (5.10) vanishes due to the canonical property. Then, in the non-zero expectations in (5.10), the multiplicity of each subscript is no less than 2.

Let among j_1, \dots, j_{2rN} there are s different ones. Due to the note above, $s \leq rN$. The number of ways to choose them from n variants is C_n^s . Given a fixed variant of this choice, the number of ways to distribute $2rN$ elements to s classes does not exceed the value s^{2rN} due to the polynomial scheme. Therefore,

$$(5.11) \quad \begin{aligned} \mathbb{E}V_n^{2N} &\leq B^{2N} n^{-rN} \sum_{s=1}^{rN} C_n^s s^{2rN} \leq B^{2N} \sum_{s=1}^{rN} n^{s-rN} s^{2rN} (s!)^{-1} \\ &\leq B^{2N} \sum_{s=1}^{rN} n^{s-rN} s^{2rN-s} e^s \leq B^{2N} \sum_{s=1}^{rN} s^{s-rN+2rN-s} e^s \leq B^{2N} e^{rN+1} (rN)^{rN}. \end{aligned}$$

Now, consider the case of m -dependent trials. It suffices to consider only the case $n > rN$. Denote by K_1, K_2, \dots, K_s the classes of identical subscripts mentioned above. Among them, choose 1-element ones: $K_1^0, K_2^0, \dots, K_l^0$. It is clear that $s-l \leq rN$.

Let $d(K_i, K_j)$ be the Hausdorff distance between sets. If there exists such a class K_i^0 that $\min_{K_j \neq K_i^0} d(K_i^0, K_j) > m$ then the corresponding mixed moments in (5.10) vanish. Otherwise, for each 1-element K_i^0 , there exists another set K_j (perhaps, 1-element set too) such that $d(K_i^0, K_j) \leq m$. Unite these two sets into one. Given a fixed K_j , there are no more than $2m$ variants for this union. Continuing this procedure, we conclude that the final upper bound in (5.11) may increase to no more than $(2m)^{rN}$ times. Inequality (5.9) is proved.

To estimate the minimum in (5.8), set $N = \varepsilon x^{2/r}$ for some $\varepsilon > 0$. Then we have

$$\mathbb{P}(|V_n| > x) \leq x^{-2N} B^{2N} (Kr\varepsilon)^{rN} x^{2N} = \exp \{ \varepsilon r \ln(KB^{2/r} r\varepsilon) x^{2/r} \}.$$

It is easy to verify that the multiplier $\varepsilon r \ln(KB^{2/r} r\varepsilon)$ reaches its minimum at the point $\varepsilon = (KB^{2/r} r e)^{-1}$, and this minimal value equals $-(KB^{2/r} e)^{-1}$. Then we finally obtain

$$\mathbb{P}(|V_n| > x) \leq \exp \{ -(KB^{2/r} e)^{-1} x^{2/r} \}$$

what required to be proved. □

References

- [1] ADAMCZAK, R. (2006). Moment inequalities for U -statistics. *Ann. Probab.* **34** (6) 2288–2314.
- [2] ARCONES, M. A. AND GINÉ, E. (1993). Limit theorems for U -processes. *Ann. Probab.* **21** (3) 1494–1542.
- [3] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley Series in Probability and Math. Statistics. John Wiley, New York.
- [4] BORISOV, I. S. (1991). Approximation of distributions of von Mises statistics with multidimensional kernels. *Siberian Math. J.* **32** (4) 554–566.

- [5] BORISOV, I. S. AND BYSTROV, A. A. (2006). Limit theorems for the canonical von Mises statistics with dependent data. *Siberian Math. J.* **47** (6) 980–989.
- [6] BORISOV, I. S. AND VOLODKO, N. V. (2008). Orthogonal series and limit theorems for canonical U - and V -statistics of stationary connected observations. *Siberian Adv. Math.* **18** (4) 244–259.
- [7] BORISOV, I. S. AND VOLODKO, N. V. (2009). Exponential inequalities for the distributions of canonical U - and V -statistics of dependent observations. *Siberian Adv. Math.* **19** (1) 1–12.
- [8] EAGLESON, G. K. (1979). Orthogonal expansions and U -statistics. *Austral. J. Statist.* **21** (3) 221–237.
- [9] GINÉ, E., LATAŁA, R. AND ZINN, J. “Exponential and moment inequalities for U -statistics,” in Proceedings of High Dimensional Probability II (Seattle, WA, 1999); *Progr. Probab.* **47** 13–38 (2000) (Boston: Birkhäuser).
- [10] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.
- [11] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.
- [12] KOLMOGOROV, A. N. AND FOMIN, S. V. (1961). *Elements of the Theory of Functions and of Functional Analysis*. Dover Publications, INC, Mineola, New-York.
- [13] KOROLYUK, V. S. AND BOROVSKIKH, YU. V. (1994). *Theory of U -Statistics*. Kluwer Academic Publ., Dordrecht.
- [14] MAJOR, P. (1981). *Multiple Wiener-Ito Integrals. Lecture Notes in Math.* **849**. Springer.
- [15] VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.
- [16] RUBIN, H. AND VITALE, R. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8** (1) 165–170.
- [17] TIKHOMIROV, A. N. (1991). On the accuracy of normal approximation of the probability of hitting a ball of sums of weakly dependent Hilbert-space-valued random variables I. *Theory Probab. Appl.* **36** (4) 738–751.