An estimate on the supremum of a nice class of stochastic integrals and U-statistics.

Péter Major
Alfréd Rényi Mathematical Institute of the Hungarian Academy of Sciences
Budapest, P.O.B. 127 H–1364, Hungary, e-mail: major@renyi.hu

Summary: Let a sequence of iid. random variables $\xi_1,\dots,\xi_n$ be given on a space $(X,\mathcal{X})$ with distribution μ together with a nice class $\mathcal{F}$ of functions $f(x_1,\dots,x_k)$ of k variables on the product space $(X^k,\mathcal{X}^k)$. For all $f\in\mathcal{F}$ we consider the random integral $J_{n,k}(f)$ of the function f with respect to the k-fold product of the normalized signed measure $\sqrt{n}(\mu_n-\mu)$, where μ_n denotes the empirical measure defined by the random variables $\xi_1,\dots,\xi_n$ and investigate the probabilities $P\left(\sup\limits_{f\in\mathcal{F}}|J_{n,k}(f)|>x\right)$ for all $x>0$. We show that for nice classes of functions, for instance if $\mathcal{F}$ is a Vapnik–Červonenkis class, an almost as good bound can be given for these probabilities as in the case when only the random integral of one function is considered. A similar result holds for degenerate U-statistics, too.

1. Introduction. Formulation of the main results

In some investigations about non-parametric maximum likelihood estimates (see [10] or [11]) I met the problem how to give a good estimate about the distribution of the supremum of appropriate classes of multiple integrals with respect to a normalized empirical measure. This problem is closely related to the study of the supremum of good classes of degenerate U-statistics. Hence, it is natural to study these two problems simultaneously. This will be done in the present paper. To formulate its main results first I introduce some notations and recall some definitions.

Let a probability measure μ be given on a measurable space $(X,\mathcal{X})$, take a sequence $\xi_1,\dots,\xi_n$ of independent, identically distributed $(X,\mathcal{X})$ valued random variables with distribution μ, and define the empirical measure μ_n,

$$\mu_n(A)=\frac{1}{n}\#\{j\colon \xi_j\in A,\ 1\le j\le n\},\quad A\in\mathcal{X}, \tag{1.1}$$

of the sample $\xi_1,\dots,\xi_n$. Let us take a nice set $\mathcal{F}$ of measurable functions $f(x_1,\dots,x_k)$ on the k-fold product space $(X^k,\mathcal{X}^k)$ and define the integrals $J_{n,k}(f)$ of the functions $f\in\mathcal{F}$ with respect to the k-fold product of the normalized empirical measure $\sqrt{n}(\mu_n-\mu)$ by the formula

$$J_{n,k}(f)=\frac{n^{k/2}}{k!}\int' f(x_1,\dots,x_k)(\mu_n(\,dx_1)-\mu(\,dx_1))\dots(\mu_n(\,dx_k)-\mu(\,dx_k)),$$

where the prime in $\int'$ means that the diagonals $x_j=x_l$, $1\le j<l\le k$, are omitted from the domain of integration. (1.2)

I try to give a good estimate on the probabilities $P\left(\sup_{f\in\mathcal{F}}|J_{n,k}(f)|>x\right)$ for all $x>0$. To formulate the result in this direction first I introduce the following definition.

Definition of L_p-dense classes of functions. *Let us have a measurable space $(Y,\mathcal{Y})$ and a set $\mathcal{G}$ of $\mathcal{Y}$-measurable functions on this space. We call $\mathcal{G}$ an L_p-dense class with parameter D and exponent L if for all numbers $1\ge\varepsilon>0$ and probability measures ν on the space $(Y,\mathcal{Y})$ there exists a finite ε-dense subset $\mathcal{G}_{\varepsilon,\nu}=\{g_1,\dots,g_m\}\subset\mathcal{G}$ in the space $L_p(Y,\mathcal{Y},\nu)$ consisting of $m\le D\varepsilon^{-L}$ elements, i.e. there is such a set $\mathcal{G}_{\varepsilon,\nu}\subset\mathcal{G}$ for which $\inf_{g_j\in\mathcal{G}_{\varepsilon,\nu}}\int|g-g_j|^p\,d\nu<\varepsilon^p$ for all functions $g\in\mathcal{G}$. (Here the set $\mathcal{G}_{\varepsilon,\nu}$ may depend on the measure ν, but its cardinality is bounded by a number depending only on ε.)*

In this paper we shall work with such classes of functions $\mathcal{F}$ which contain only functions with absolute value less than or equal to 1. In this case $\mathcal{F}$ is an L_p-dense class of functions for all $1\le p<\infty$ (with an exponent and a parameter depending on p) if there is a number $1\le p<\infty$ for which it is L_p-dense. We shall formulate our statements mainly for L_p-dense classes of functions with the parameter $p=2$, since this seems to be the most convenient choice. Our main result is the following

Theorem 1. *Let us have a non-atomic measure μ on the space $(X,\mathcal{X})$ together with an L_2-dense class $\mathcal{F}$ of functions $f=f(x_1,\dots,x_k)$ of k variables with some parameter $D>0$ and exponent $L\ge1$ on the product space $(X^k,\mathcal{X}^k)$ which consists of at most countably infinite functions, and satisfies the conditions*

$$\|f\|_\infty=\sup_{x_j\in X,\ 1\le j\le k}|f(x_1,\dots,x_k)|\le1,\qquad \textit{for all } f\in\mathcal{F} \tag{1.3}$$

and

$$\|f\|_2^2=Ef^2(\xi_1,\dots,\xi_k)=\int f^2(x_1,\dots,x_k)\mu(\,dx_1)\dots\mu(\,dx_k)\le\sigma^2\qquad \textit{for all } f\in\mathcal{F} \tag{1.4}$$

with some constant $0<\sigma\le1$. Then there exist some constants $C=C(k)>0$, $\alpha=\alpha(k)>0$ and $M=M(k)>0$ depending only on the parameter k such that the supremum of the random integrals $J_{n,k}(f)$, $f\in\mathcal{F}$, defined by formula (1.2) satisfies the inequality

$$P\left(\sup_{f\in\mathcal{F}}|J_{n,k}(f)|\ge x\right)\le CD\exp\left\{-\alpha\left(\frac{x}{\sigma}\right)^{2/k}\right\}$$
$$\textit{if}\quad n\sigma^2\ge\left(\frac{x}{\sigma}\right)^{2/k}\ge M(L+\beta)^{3/2}\log\frac{2}{\sigma}, \tag{1.5}$$

where $\beta=\max\left(\frac{\log D}{\log n},0\right)$, and the numbers D and L in formula (1.5) are the parameter and exponent of the L_2-dense class $\mathcal{F}$.

Theorem 1 has a natural counterpart about degenerate U-statistics formulated in Theorem 2 below. Before its formulation I recall the definition of U-statistics and degenerate U-statistics.

Let us have a sequence of independent and identically distributed random variables $\xi_1, \xi_2, \dots$ with distribution μ on a measurable space $(X, \mathcal{X})$ together with a function $f = f(x_1, \dots, x_k)$ on the k-th power $(X^k, \mathcal{X}^k)$ of the space $(X, \mathcal{X})$. We define with their help the U-statistic $I_{n,k}(f)$ of order k, as

$$I_{n,k}(f) = \frac{1}{k!} \sum_{\substack{1 \le j_s \le n,\ s=1,\dots,k \\ j_s \ne j_{s'} \text{ if } s \ne s'}} f\left(\xi_{j_1}, \dots, \xi_{j_k}\right). \tag{1.6}$$

(The function f in this formula will be called the kernel function of the U-statistic.)

A real valued function $f = f(x_1, \dots, x_k)$ on the k-th power $(X^k, \mathcal{X}^k)$ of a space $(X, \mathcal{X})$ is called a canonical kernel function (with respect to the probability measure μ on the space $(X, \mathcal{X})$) if

$$\int f(x_1, \dots, x_{j-1}, u, x_{j+1}, \dots, x_k) \mu(\,du) = 0 \quad \text{for all } 1 \le j \le k \quad \text{and} \quad x_s \in X,\ s \ne j.$$

I also introduce the notion of canonical functions in a more general case, because this notion appears later in Proposition 5 of this paper. We call a function $f(x_1, \dots, x_k)$ on the k-fold product $(X_1 \times \cdots \times X_k, \mathcal{X}_1 \times \cdots \times \mathcal{X}_k, \mu_1 \times \cdots \times \mu_k)$ of k not necessarily identical probability spaces $(X_j, \mathcal{X}_j, \mu_j)$, $1 \le j \le k$, canonical if

$$\int f(x_1, \dots, x_{j-1}, u, x_{j+1}, \dots, x_k) \mu_j(\,du) = 0 \quad \text{for all } 1 \le j \le k \quad \text{and} \quad x_s \in X_s,\ s \ne j.$$

A U-statistic with a canonical kernel function is called degenerate. Now I formulate Theorem 2.

Theorem 2. *Let us have a probability measure μ on a space $(X, \mathcal{X})$, a sequence of independent and μ distributed random variables $\xi_1, \dots, \xi_n$ together with an L_2-dense class $\mathcal{F}$ of canonical (with respect to the measure μ) kernel functions $f = f(x_1, \dots, x_k)$ with some parameter $D > 0$ and exponent $L \ge 1$ on the product space $(X^k, \mathcal{X}^k)$ which consists of at most countably infinite functions, and satisfies conditions (1.3) and (1.4) with some $0 < \sigma \le 1$. Then there exist some numbers $C = C(k) > 0$, $M = M(k) > 0$ $\alpha = \alpha(k) > 0$ depending only on the order k of the U-statistics we consider such that the degenerate U-statistics $I_{n,k}(f)$, $f \in \mathcal{F}$, defined in (1.6) satisfy the inequality*

$$\begin{aligned} P\left(\sup_{f \in \mathcal{F}} |n^{-k/2} I_{n,k}(f)| \ge x\right) &\le CD \exp\left\{-\alpha \left(\frac{x}{\sigma}\right)^{2/k}\right\} \\ &\qquad \textit{if} \quad n\sigma^2 \ge \left(\frac{x}{\sigma}\right)^{2/k} \ge M(L+\beta)^{3/2} \log \frac{2}{\sigma}, \end{aligned} \tag{1.7}$$

where $\beta = \max\left(\frac{\log D}{\log n}, 0\right)$.

To understand the relation between Theorems 1 and 2 let us observe that the definition of the U-statistics in (1.6) can be rewritten as

$$I_{n,k}(f) = \frac{n^k}{k!} \int' f(x_1, \dots, x_k)\mu_n(\,dx_1) \dots \mu_n(\,dx_k) \tag{1.8}$$

if the distribution μ of the random variables $\xi_1, \dots, \xi_n$ is non-atomic. (The non-atomic property is needed to guarantee that the random variables $\xi_1, \dots, \xi_n$ take different values with probability 1.) The difference between the random integrals $J_{n,k}(f)$ and the random integral representation (1.8) of U-statistics $I_{n,k}(f)$ is (beside the different norming constant) that in formula (1.8) we integrate with respect to the empirical measure μ_n and not with respect to its normalized version $\mu_n - \mu$. As a consequence, we can get a good estimate for U-statistics only under some restriction. In Theorem 2 we had to impose the condition that the functions of the class $\mathcal{F}$ are canonical, while no similar condition was needed in Theorem 1. Hence Theorem 1 can be better applied in statistical problems. On the other hand, the proof of Theorem 2 is simpler. But Theorem 1 can be deduced from it by means of a good representation of multiple random integrals $J_{n,k}(f)$ as a linear combination of degenerate U-statistics. In this work the following approach will be followed. In the main text Theorem 2 will be proved. The Appendix contains the proof of the above mentioned representation of random integrals which enables us to deduce Theorem 1 from Theorem 2.

Let us discuss the conditions of Theorems 1 and 2. We have assumed that $\mathcal{F}$ contains at most countably infinite functions. This condition, which is too restrictive for statistical applications can be weakened. The introduction of the following definition seems to be useful.

Definition of countable approximability. *A class of functions $\mathcal{F}$ is countably approximable in the space $(X^k, \mathcal{X}^k, \mu^k)$ if there exists a countable subset $\mathcal{F}' \subset \mathcal{F}$ such that for all numbers $x > 0$ the sets $A(x) = \{\omega\colon \sup\limits_{f \in \mathcal{F}} |J_{n,k}(f)(\omega)| \ge x\}$ and $B(x) = \{\omega\colon \sup\limits_{f \in \mathcal{F}'} |J_{n,k}(f)(\omega)| \ge x\}$ satisfy the identity $P(A(x) \setminus B(x)) = 0$.*

Clearly, $B(x) \subset A(x)$. In the above definition we demanded that for all $x > 0$ the set $B(x)$ is almost as large as $A(x)$. The following corollary of Theorems 1 and 2 holds.

Corollary of Theorem 1 or 2. *Let a class of functions $\mathcal{F}$ satisfy the conditions of Theorem 1 or 2 with the only exception that instead of the condition about the countable cardinality of $\mathcal{F}$ it is assumed that $\mathcal{F}$ is countably approximable in the space $(X^k, \mathcal{X}^k, \mu^k)$. Then $\mathcal{F}$ satisfies Theorem 1 or 2.*

In Theorems 1 and 2 we have imposed the condition that the class of functions $\mathcal{F}$ is countable to avoid some unpleasant measure theoretical difficulties. Otherwise we should have to work with possibly non-measurable sets. On the other hand, I have the impression that Corollary 1 can be applied in all statistical problems where we have to work with the supremum of multiple random integrals or U-statistics. It is not difficult to prove that Corollary 1 follows from Theorem 1 or 2. We have to show that if $\mathcal{F}$ is an L_2-dense class with some parameter D and exponent L, and $\mathcal{F}' \subset \mathcal{F}$, then $\mathcal{F}'$ is also an L_2-dense class with the same exponent L, only with a possibly different parameter D'.

To prove this statement let us choose for all numbers $1 \ge \varepsilon > 0$ and probability measures ν on $(Y, \mathcal{Y})$ some functions $f_1, \dots, f_m \in \mathcal{F}$ with $m \le D\left(\frac{\varepsilon}{2}\right)^{-L}$ elements, such that the sets $\mathcal{U}_j = \left\{f\colon \int |f - f_j|^2 \, d\nu \le \left(\frac{\varepsilon}{2}\right)^2\right\}$ satisfy the relation $\bigcup\limits_{j=1}^{m} \mathcal{U}_j = Y$. For all sets $\mathcal{U}_j$ for which $\mathcal{U}_j \cap \mathcal{F}'$ is non-empty choose a function $f_j' \in \mathcal{U}_j \cap \mathcal{F}'$. In such a way we get a collection of functions f_j' from the class $\mathcal{F}'$ containing at most $2^L D\varepsilon^{-L}$ elements which satisfy the condition imposed for L_2-dense classes with exponent L and parameter $2^L D$ for this number ε and measure ν.

In Theorems 1 and 2 we have considered the supremum of multiple random integrals and U-statistics of order k for a nice class of functions. It was shown that if the variances of the random integrals or U-statistics we have considered are less than some number $0 < \sigma^2 \le 1$, (formula (1.4) was a condition about these variances in an implicit way) then under some additional conditions this supremum takes a value larger than x with a probability less than $P(C\sigma\eta^k > x)$, where η is a standard normal random variable, and $C = C(k) > 0$ is a universal constant depending only on the multiplicity k of the random integrals. This is the sharpest estimate we can expect. Moreover, this estimate seems to be sharp also in that respect that the conditions imposed for its validity cannot be considerably weakened. If condition (1.3) does not hold or $n\sigma^2 < \left(\frac{x}{\sigma}\right)^{2/k}$, then the estimate of Theorem 1 or 2 may not hold any longer even if the class of functions $\mathcal{F}$ contains only one function. In such cases there exist examples for which the probability $P(J_{n,k}(f) > x)$ is too large. In [8] I gave such examples (Examples 3.2 and 8.6). Here I do not discuss them in detail.

If the other inequality is violated in the conditions of formula (1.5) or (1.7), i.e. if $\left(\frac{x}{\sigma}\right)^{2/k} < M \log \frac{2}{\sigma}$ with a not too large number $M > 0$, then the estimate of Theorem 1 or 2 may not hold for a different reason. The supremum of many small random variables may be large, and inequalities (1.5) or (1.7) may loose their validity for this reason. To understand this let us consider the following analogous problem. Take a Wiener process $W(t)$, $0 \le t \le 1$, and consider the supremum of the expressions $W(t) - W(s) = \int f_{s,t}(u) W(\,du) = \bar{J}(f_{s,t})$, with the functions $f_{s,t}(\cdot)$ on the interval $[0,1]$ defined by the formula $f_{s,t}(u) = 1$ if $s \le u \le t$, $f_{s,t}(u) = 0$ if $0 \le u < s$ or $t < u \le 1$. If we consider the class of functions $\mathcal{F}_\sigma = \{f_{s,t}\colon \int f_{s,t}^2(u)\,du = t - s \le \sigma^2\}$, then it is natural to expect that $P\left(\sup\limits_{f_{s,t} \in \mathcal{F}_\sigma} \bar{J}(f_{s,t}) > x\right) \le e^{-\text{const.}\,(x/\sigma)^2}$. However, this relation does not hold if $x = x(\sigma) < (1-\varepsilon)\sqrt{2 \log \frac{1}{\sigma}}\sigma$ with some $\varepsilon > 0$. In such cases $P\left(\sup\limits_{f_{s,t} \in \mathcal{F}_\sigma} \bar{J}(f_{s,t}) > x\right) \to 1$, as $\sigma \to 0$. This can be proved relatively simply with the help of the estimate $P(\bar{J}(f_{s,t}) > x(\sigma)) \ge \text{const.}\,\sigma^{1-\varepsilon}$ if $|t - s| = \sigma^2$ and the independence of the random integrals $\bar{J}(f_{s,t})$ if the functions $f_{s,t}$ are indexed by such pairs (s,t) for which the intervals (s,t) are disjoint.

Some additional work would show that a similar picture arises if we integrate with respect to the normalized empirical measure of a sample with uniform distribution on the interval $[0,1]$ instead of a Wiener process. This yields an example for an L_2-dense class of functions in the case $k=1$ for which the estimate of Theorem 1 does not hold any longer if $\left(\frac{x}{\sigma}\right)^{2/k} < M \log \frac{2}{\sigma}$ with some $M < \sqrt{2}$. Similar example can be constructed also in the case of Theorem 2. At a heuristic level it is clear that such an example can be given also for $k > 1$, and the number M in condition (1.5) or (1.7) has to be chosen larger if we want that Theorem 1 or Theorem 2 hold also for an L_2-dense class of functions $\mathcal{F}$ with a large exponent L. (In this paper I did not try to find the best possible condition of Theorem 1 or 2 in the right-hand side inequality of (1.5) or (1.7).)

One would like to see some interesting examples when Theorem 1 or 2 is applicable and to have some methods to check their conditions. It is useful to know that if $\mathcal{F}$ is a Vapnik–Červonenkis class of functions whose absolute values are bounded by 1, then $\mathcal{F}$ is an L_2-dense class.

To formulate the above statement more explicitly let us recall that a class of subsets $\mathcal{D}$ of a set S is a Vapnik–Červonenkis class if there exist some constants $B>0$ and $K>0$ such that for all integers n and sets $S_0(n)=\{x_1,\dots,x_n\}\subset S$ of cardinality n the collection of sets of the form $S_0(n)\cap D$, $D\in\mathcal{D}$, contains no more than Bn^K subsets of $S_0(n)$. A class of real valued functions $\mathcal{F}$ on a space $(Y,\mathcal{Y})$ is a Vapnik–Červonenkis class if the graphs of these functions is a Vapnik–Červonenkis class, i.e. if the sets $A(f)=\{(y,t)\colon y\in Y,\ \min(0,f(y))\le t\le \max(0,f(y))\}$, $f\in\mathcal{F}$, constitute a Vapnik–Červonenkis class of sets on the product space $Y\times R^1$.

An important result of Dudley states that a Vapnik–Červonenkis class of functions whose absolute values are bounded by 1 is an L_1-dense class. The parameter and exponent of this L_1-dense class can be bounded by means of the constants B and K appearing in the definition of Vapnik–Červonenkis classes. Beside this, an L_1-dense class of functions bounded by 1 is also an L_2-dense class (with possibly different exponent and parameter), since $\int |f-g|^2\,d\nu \le 2\int |f-g|\,d\nu$ in this case. Dudley's result, whose proof can be found e.g. in Chapter II of Pollard's book [9] (the 25° approximation lemma contains this result in a slightly more general form) is useful for us, because there are results which enable us to prove that certain classes of functions constitute a Vapnik–Červonenkis class.

I found some results similar to that of this paper in the work of Arcones and Gine [3], where the tail-behaviour of the supremum of degenerate U-statistics was investigated if the kernel functions of these U-statistics constitute a Vapnik–Červonenkis class. But the bounds of that paper do not give a better estimate if we have the additional information that the variances of the U-statistics we consider are small. The main goal of the present paper was to prove such estimates which take into account the bound we have on the variance of the random integrals $J_{n,k}(f)$ or U-statistics $I_{n,k}(f)$ we consider.

In the investigation of this work Alexander's paper [1] played an essential role. In Alexander's work a similar problem was considered in the special case $k=1$. It was interesting for me first of all, because I learned some ideas from it which I strongly needed in the present work. On the other hand, I also needed some new arguments, because in the study of multiple stochastic integrals or U-statistics some new difficulties had to be overcome.

This paper consists of six sections and an Appendix. In Section 2 Theorems 1 and 2 are reduced to two simpler statements formulated in Propositions 2 and 3. Section 3 contains some important results needed in the proof of Proposition 2, and the main ideas of its proof are explained there. It is shown that Proposition 2 follows from another statement formulated in Proposition 4. Proposition 4 is proved together with another result described in Proposition 5. To make the proof more transparent first I explain it in the special case $k=1$ in Section 4. Sections 5 and 6 contain the proof of Propositions 4 and 5 in the general case. In Section 5 it is shown how a symmetrization argument can be applied to prove Propositions 4 and 5, and finally the proof is completed in Section 6. The Appendix contains the proof about a result of an expansion of multiple random integrals in the form of a linear combination of degenerate U-statistics formulated in Proposition 3. This result enables us to deduce Theorem 1 from Theorem 2.

2. Reduction of Theorems 1 and 2 to some simpler results

First I prove with the help of a natural argument, called the Chaining argument in the literature, and the multi-dimensional generalization of Bernstein's inequality (see [2], Proposition 2.3(c)) a result that yields a reduction of Theorem 2 we shall need later. I shall apply the following consequence of this result (which is actually equivalent to it).

If $U_{n,k}(f)$ is a degenerate U-statistic of order k with a (canonical) kernel function f which satisfies relations (1.3) and (1.4) (formally the class of functions $\mathcal{F}$ consisting only of the function f satisfies these relations) with some number $0<\sigma\le 1$ and the distribution μ of the iid. sequence of the random variables $\xi_1,\dots,\xi_n$ taking part in the definition of the U-statistic $U_{n,k}(f)$, then there exist some constants $C=C(k)>0$ and $\alpha=\alpha(k)>0$ depending only on the order k of this U-statistic such that

$$P\left(n^{-k/2}|I_{n,k}(f)|>x\right)\le C\exp\left\{-\alpha\left(\frac{x}{\sigma}\right)^{2/k}\right\}\quad\text{for } 0\le x\le n^{k/2}\sigma^{k+1}. \tag{2.1}$$

Now I formulate the following result.

Proposition 1. *Let us fix some number $\bar A\ge 2^k$, and assume that a class of functions $\mathcal{F}$ satisfies the conditions of Theorem 2 with an appropriately chosen number M in these conditions which may depend also on $\bar A$. Then a number $\bar\sigma$, $0\le\bar\sigma\le\sigma\le 1$, and a collection of functions $\mathcal{F}_{\bar\sigma}=\{f_1,\dots,f_m\}\subset\mathcal{F}$ with $m\le D\bar\sigma^{-L}$ elements can be chosen in such a way that the sets $\mathcal{D}_j=\{f\colon f\in\mathcal{F}, \int|f-f_j|^2\,d\mu\le\bar\sigma^2\}$, $1\le j\le m$, satisfy the relation $\bigcup\limits_{j=1}^m\mathcal{D}_j=\mathcal{F}$, and*

$$P\left(\sup_{f\in\mathcal{F}_{\bar\sigma}} n^{-k/2}|I_{n,k}(f)|\ge\frac{x}{\bar A}\right)\le 2CD\exp\left\{-\alpha\left(\frac{x}{10\bar A\sigma}\right)^{2/k}\right\} \tag{2.2}$$
$$\textit{if}\quad n\sigma^2\ge\left(\frac{x}{\sigma}\right)^{2/k}\ge ML\log\frac{2}{\sigma}$$

with the constants $\alpha=\alpha(k)$, $C=C(k)$ *appearing in formula (2.1) and the exponent* L *and parameter* D *of the* L_2*-dense class* $\mathcal{F}$ *if the constant* $M=M(k,\bar A)$ *is chosen sufficiently large. Beside this, also the inequalities* $4\left(\frac{x}{\bar A\bar\sigma}\right)^{2/k}\ge n\bar\sigma^2\ge\frac{1}{64}\left(\frac{x}{\bar A\sigma}\right)^{2/k}$ *and* $n\bar\sigma^2\ge\frac{M^{2/3}(L+\beta)\log n}{1000\bar A^{4/3}}$ *hold, provided that* $n\sigma^2\ge\left(\frac{x}{\sigma}\right)^{2/k}\ge M(L+\beta)^{3/2}\log\frac{2}{\sigma}$ *with* $\beta=\max\left(\frac{\log D}{\log n},0\right)$.

Remark: The introduction of the number $\bar A\ge 2^k$ in Proposition 1 may seem a bit artificial. Its role is to guarantee that such a number $\bar\sigma$ could be defined in Proposition 1 which satisfies the inequality $\left(\frac{x}{\bar\sigma}\right)^{2/k}\ge An\bar\sigma^2$ with a sufficiently large previously fixed constant $A=A(k)$.

Proof of Proposition 1. For all $p=0,1,2,\dots$ choose a set $\mathcal{F}_p=\{f_{p,1},\dots,f_{p,m_p}\}\subset\mathcal{F}$ with $m_p\le D\,2^{2pL}\sigma^{-L}$ elements in such a way that $\inf\limits_{1\le j\le m_p}\int(f-f_{p,j})^2\,d\mu\le 2^{-4p}\sigma^2$ for all $f\in\mathcal{F}$. For all pairs (j,p), $p=1,2,\dots$, $1\le j\le m_p$, choose a predecessor $(j',p-1)$, $j'=j'(j,p)$, $1\le j'\le m_{p-1}$, in such a way that the functions $f_{j,p}$ and $f_{j',p-1}$ satisfy the relation $\int|f_{j,p}-f_{j',p-1}|^2\,d\mu\le\sigma^2 2^{-4(p-1)}$. Then we have $\int\left(\frac{f_{j,p}-f_{j',p-1}}{2}\right)^2\,d\mu\le 4\sigma^2 2^{-4p}$ and $\sup\limits_{x_j\in X,\,1\le j\le k}\left|\frac{f_{j,p}(x_1,\dots,x_k)-f_{j',p-1}(x_1,\dots,x_k)}{2}\right|\le 1$. Inequality (2.1) yields that

$$P(A(j,p))=P\left(n^{-k/2}|I_{n,k}(f_{j,p}-f_{j',p-1})|\ge\frac{2^{-(1+p)}x}{\bar A}\right)\le C\exp\left\{-\alpha\left(\frac{2^px}{8\bar A\sigma}\right)^{2/k}\right\}$$

$$\text{if}\quad n\sigma^2 2^{-4p}\ge\left(\frac{2^px}{8\bar A\sigma}\right)^{2/k},\quad 1\le j\le m_p,\ p=1,2,\dots,\tag{2.3}$$

and

$$P(B(s))=P\left(n^{-k/2}|I_{n,k}(f_{0,s})|\ge\frac{x}{2\bar A}\right)\le C\exp\left\{-\alpha\left(\frac{x}{2\bar A\sigma}\right)^{2/k}\right\},\quad 1\le s\le m,$$
$$\text{if}\quad n\sigma^2\ge\left(\frac{x}{2\bar A\sigma}\right)^{2/k}.\tag{2.4}$$

Choose an integer R, $R\ge 0$, in such a way that $2^{(4+2/k)(R+1)}\left(\frac{x}{\bar A\sigma}\right)^{2/k}\ge 2^{2+6/k}n\sigma^2\ge 2^{(4+2/k)R}\left(\frac{x}{\bar A\sigma}\right)^{2/k}$, and define $\bar\sigma^2=2^{-4R}\sigma^2$ and $\mathcal{F}_{\bar\sigma}=\mathcal{F}_R$. (As $n\sigma^2\ge\left(\frac{x}{\sigma}\right)^{2/k}$ and $\bar A\ge 2^k$ by our conditions, there exists such a non-negative number R.) Then the cardinality m of the set $\mathcal{F}_{\bar\sigma}$ is clearly not greater than $D\bar\sigma^{-L}$, and $\bigcup\limits_{j=1}^m\mathcal{D}_j=\mathcal{F}$. Beside this, the number R was chosen in such a way that inequalities (2.2) and (2.3) can be applied for $1\le p\le R$. Hence the definition of the predecessor of a pair (j,p) implies that

$$P\left(\sup_{f\in\mathcal{F}_{\bar\sigma}}n^{-k/2}|I_{n,k}(f)|\ge\frac{x}{\bar A}\right)\le P\left(\bigcup_{p=1}^R\bigcup_{j=1}^{m_p}A(j,p)\cup\bigcup_{s=1}^m B(s)\right)$$

$$\leq \sum_{p=1}^{R}\sum_{j=1}^{m_p} P(A(j,p)) + \sum_{s=1}^{m} P(B(s)) \leq \sum_{p=1}^{\infty} CD\, 2^{2pL}\sigma^{-L}\exp\left\{-\alpha\left(\frac{2^p x}{8\bar A\sigma}\right)^{2/k}\right\}$$
$$+ CD\sigma^{-L}\exp\left\{-\alpha\left(\frac{x}{2\bar A\sigma}\right)^{2/k}\right\}.$$

If the condition $\left(\frac{x}{\sigma}\right)^{2/k} \geq ML^{3/2}\log\frac{2}{\sigma}$ holds with a sufficiently large constant M (depending on $\bar A$), then the inequalities

$$2^{2pL}\sigma^{-L}\exp\left\{-\alpha\left(\frac{2^p x}{8\bar A\sigma}\right)^{2/k}\right\} \leq 2^{-p}\exp\left\{-\alpha\left(\frac{2^p x}{10\bar A\sigma}\right)^{2/k}\right\}$$

hold for all $p = 1, 2, \dots$, and

$$\sigma^{-L}\exp\left\{-\alpha\left(\frac{x}{2\bar A\sigma}\right)^{2/k}\right\} \leq \exp\left\{-\alpha\left(\frac{x}{10\bar A\sigma}\right)^{2/k}\right\}.$$

Hence the previous estimate implies that

$$P\left(\sup_{f\in\mathcal{F}_{\bar\sigma}} n^{-k/2}|I_{n,k}(f)| \geq \frac{x}{\bar A}\right) \leq \sum_{p=1}^{\infty} CD2^{-p}\exp\left\{-\alpha\left(\frac{2^p x}{10\bar A\sigma}\right)^{2/k}\right\}$$
$$+ CD\exp\left\{-\alpha\left(\frac{x}{10\bar A\sigma}\right)^{2/k}\right\} \leq 2CD\exp\left\{-\alpha\left(\frac{x}{10\bar A\sigma}\right)^{2/k}\right\},$$

and relation (2.2) holds. We have

$$n\bar\sigma^2 = 2^{-4R}n\sigma^2 \leq 2^{-4R}\cdot 2^{(4+2/k)(R+1)-2-6/k}\left(\frac{x}{\bar A\sigma}\right)^{2/k} = 2^{2-4/k}\cdot 2^{2R/k}\left(\frac{x}{\bar A\sigma}\right)^{2/k}$$
$$= 2^{2-4/k}\cdot\left(\frac{\sigma}{\bar\sigma}\right)^{1/k}\left(\frac{x}{\bar A\sigma}\right)^{2/k} = 2^{2-4/k}\cdot\left(\frac{\bar\sigma}{\sigma}\right)^{1/k}\left(\frac{x}{\bar A\bar\sigma}\right)^{2/k},$$

hence $n\bar\sigma^2 \leq 4\left(\frac{x}{\bar A\bar\sigma}\right)^{2/k}$. Beside this, as $n\sigma^2 \geq 2^{(4+2/k)R-2-6/k}\left(\frac{u}{\bar A\sigma}\right)^{2/k}$, $R\geq 1$,

$$n\bar\sigma^2 = 2^{-4R}n\sigma^2 \geq 2^{-2-6/k}\cdot 2^{2R/k}\left(\frac{x}{\bar A\sigma}\right)^{2/k} \geq \frac{1}{64}\left(\frac{x}{\bar A\sigma}\right)^{2/k}.$$

It remained to show that $n\bar\sigma^2 \geq \frac{M^{2/3}(L+\beta)\log n}{1000A^{4/3}}$.

This inequality clearly holds under the conditions of Proposition 1 if $\sigma \leq n^{-1/3}$, since in this case $\log\frac{2}{\sigma} \geq \frac{\log n}{3}$, and $n\bar\sigma^2 \geq \frac{1}{64}\left(\frac{x}{\bar A\sigma}\right)^{2/k} \geq \frac{1}{64}\bar A^{-2/k}M(L+\beta)^{3/2}\log\frac{2}{\sigma} \geq \frac{1}{192}\bar A^{-2/k}M(L+\beta)\log n \geq \frac{M^{2/3}(L+\beta)\log n}{1000A^{4/3}}$ if $M = M(\bar A, k)$ is chosen sufficiently large.

If $\sigma \geq n^{-1/3}$, then the inequality $2^{(4+2/k)R}\left(\frac{x}{\bar A\sigma}\right)^{2/k} \leq 2^{2+6/k}n\sigma^2$ holds. Hence $2^{-4R} \geq 2^{-4(2+6/k))/(4+2/k)}\left[\frac{\left(\frac{x}{\bar A\sigma}\right)^{2/k}}{n\sigma^2}\right]^{4/(4+2/k)}$, and

$$n\bar\sigma^2 = 2^{-4R}n\sigma^2 \geq \frac{2^{-16/3}}{\bar A^{4/3}}(n\sigma^2)^{1-\gamma}\left[\left(\frac{x}{\sigma}\right)^{2/k}\right]^{\gamma} \text{ with } \gamma = \frac{4}{4+\frac{2}{k}} \geq \frac{2}{3}.$$

Since $n\sigma^2 \ge (\frac{x}{\sigma})^{2/k} \ge \frac{M}{3}(L+\beta)^{3/2}$, and $n\sigma^2 \ge n^{1/3}$, the above estimates yield that $n\bar\sigma^2 \ge \frac{\bar A^{-4/3}}{50}(n\sigma^2)^{1/3}\left[\left(\frac{x}{\sigma}\right)^{2/k}\right]^{2/3} \ge \frac{\bar A^{-4/3}}{50}n^{1/9}\left(\frac{M}{3}\right)^{2/3}(L+\beta) \ge \frac{M^{2/3}(L+\beta)\log n}{1000\bar A^{4/3}}$.

Now I formulate Proposition 2 and show that Theorem 2 follows from Propositions 1 and 2.

Proposition 2. *Let us have a probability measure μ on a space $(X,\mathcal{X})$ together with a sequence of independent and μ distributed random variables $\xi_1,\dots,\xi_n$ and an L_2-dense class $\mathcal{F}$ of canonical kernel functions $f=f(x_1,\dots,x_k)$ (with respect to the measure μ) with some parameter $D>0$ and exponent $L\ge 1$ on the product space $(X^k,\mathcal{X}^k)$ which consists of at most countably many functions, and satisfies conditions (1.3) and (1.4) with some $0<\sigma\le 1$. Let $n\sigma^2 > K(L+\beta)\log n$ with $\beta=\max\left(\frac{\log D}{\log n},0\right)$ and a sufficiently large constant $K=K(k)$. Then there exist some numbers $\bar C=\bar C(k)>0$, $\gamma=\gamma(k)>0$ and threshold index $A_0=A_0(k)>0$ depending only on the order k of the U-statistics we consider such that the degenerate U-statistics $I_{n,k}(f)$, $f\in\mathcal{F}$, defined in (1.6) satisfy the inequality*

$$P\left(\sup_{f\in\mathcal{F}}|n^{-k/2}I_{n,k}(f)|\ge An^{k/2}\sigma^{k+1}\right)\le \bar Ce^{-\gamma A^{1/2k}n\sigma^2}\quad \textit{if } A\ge A_0. \tag{2.5}$$

In the proof of Theorem 2 with the help of Propositions 1 and 2 we exploit our freedom in the choice of the parameters in these results. Let us choose a number $\bar A_0$ such that $\bar A_0\ge A_0$ and $\gamma\bar A_0^{1/2k}\ge\frac{1}{K}$ with the numbers A_0, K and γ in Proposition 2. We shall apply Proposition 1 with the choice $\bar A=\max(2^{k+2}\bar A_0,2^k)$. Then by Proposition 1 and the choice of the numbers $\bar A$ and $\bar A_0$ also the inequality $\left(\frac{x}{\bar\sigma}\right)^{2/k}\ge\frac{\bar A^{2/k}}{4}n\bar\sigma^2\ge(4\bar A_0)^{2/k}n\bar\sigma^2$ holds, hence $x\ge 4\bar An^{k/2}\bar\sigma^{k+1}$ with the number $\bar\sigma$ in Proposition 1. This implies that $\left(\frac12-\frac{1}{2\bar A}\right)x\ge\frac{x}{4}\ge\bar An^{k/2}\bar\sigma^{k+1}$, and $\bar A\ge A_0$. The numbers x considered in these estimations satisfy the condition $n\sigma^{2/k}\ge\left(\frac{x}{\sigma}\right)^{2/k}\ge M(L+\beta)^{3/2}\log\frac{2}{\sigma}$ imposed in Proposition 1 with some appropriately chosen constant M. Choose the number $M\ge M(\bar A,k)$ in Proposition 1 (which also can be chosen as the number M in formula (1.7) of Theorem 2) in such a way that it also satisfies the inequality $\frac{M^{2/3}(L+\beta)\log n}{1000\bar A^{4/3}}\ge K(L+\beta)\log n$ with the number K appearing in the conditions of Proposition 2. With such a choice the inequality $n\bar\sigma^2\ge\frac{M^{2/3}(L+\beta)\log n}{1000\bar A^{4/3}}\ge K(L+\beta)\log n$ holds, and Proposition 2 can be applied with the choice $\bar\sigma$ defined in Proposition 1 for the parameter σ, the number $\left(\frac12-\frac{1}{2\bar A}\right)x$ as the number A in this result, together with the classes of functions $\bar{\mathcal{D}}_j=\left\{g=\frac{f-f_j}{2},\ f\in\mathcal{D}_j\right\}$, $1\le j\le m$, where the classes of functions $\mathcal{D}_j$ and functions f_j, $1\le j\le m$, are defined in Proposition 1.

Then Propositions 1 and 2 together with the above observations yield that

$$P\left(\sup_{f\in\mathcal{F}}n^{-k/2}|I_{n,k}(f)|\ge x\right)\le P\left(\sup_{f\in\mathcal{F}_{\bar\sigma}}n^{-k/2}|I_{n,k}(f)|\ge\frac{x}{\bar A}\right)$$

$$+\sum_{j=1}^{m} P\left(\sup_{g\in\mathcal{D}_j} n^{-k/2}\left|I_{n,k}\left(\frac{f_j-g}{2}\right)\right|\ge\left(\frac12-\frac{1}{2\bar A}\right)x\right) \tag{2.6}$$
$$\le 2CD\exp\left\{-\alpha\left(\frac{x}{10\bar A\sigma}\right)^{2/k}\right\}+\bar CD\bar\sigma^{-L}e^{-\gamma\bar A^{1/2k}n\bar\sigma^2}.$$

To get the result of Theorem 2 from inequality (2.6) we have to replace its second term at the right-hand side with a more appropriate expression where, in particular, we get rid of the coefficient $\bar\sigma^{-L}$. The condition $n\bar\sigma^2\ge K(L+\beta)\log n$ implies that $\bar\sigma\ge n^{-1/2}$, and by our choice of $\bar A_0$ we have $\gamma\bar A_0^{1/2k}n\bar\sigma^2\ge\frac1K n\bar\sigma^2\ge L\log n\ge 2L\log\frac1{\bar\sigma}$, i.e. $\bar\sigma^{-L}\le e^{\gamma\bar A_0^{1/2k}n\bar\sigma^2/2}$. By the estimates of Proposition 1 $n\bar\sigma^2\ge\frac1{64}\left(\frac{x}{\bar A\sigma}\right)^{2/k}$. The above relations imply that $\bar\sigma^{-L}e^{-\gamma\bar A_0^{1/2k}n\bar\sigma^2}\le e^{-\gamma\bar A_0^{1/2k}n\bar\sigma^2/2}\le\exp\left\{-\frac{\gamma}{128}\bar A_0^{1/2k}\bar A^{-2/k}\left(\frac u\sigma\right)^{2/k}\right\}$. Hence relation (2.6) yields that

$$P\left(\sup_{f\in\mathcal F} n^{-k/2}|I_{n,k}(f)|\ge x\right)$$
$$\le 2CD\exp\left\{-\frac{\alpha}{(10\bar A)^2}\left(\frac{x}{\sigma}\right)^{2/k}\right\}+\bar CD\exp\left\{-\frac{\gamma}{128}\bar A_0^{1/2k}\bar A^{-2/k}\left(\frac x\sigma\right)^{2/k}\right\},$$

and this estimate implies Theorem 2 with some new appropriately defined constants $\alpha>0$ and $C>0$.

Thus I have reduced the proof of Theorem 2 to that of Proposition 2. I also show in this section that the proof of Theorem 1 can be reduced to that of Theorem 2 and a decomposition result of random integrals $J_{n,k}(f)$ formulated in Proposition 3 below whose proof will be given in the Appendix. Proposition 3 gives the representation of a random integral $J_{n,k}(f)$ in the form of a linear combination of degenerate U-statistics. To get this representation we can observe that a random integral $J_{n,k}(f)$ can be rewritten in the form of a sum of U-statistics. By applying an important result, called Hoeffding's decomposition, we can write a general U-statistic in the form of a sum of degenerate U-statistics of different order. Proposition 3 contains the result we get by carrying out this procedure. Let us recall that we have integrated with respect to the signed measure $\mu_n-\mu$ in the definition (1.3) of the random integrals $J_{n,k}(f)$. This has a very strong cancellation effect, and the main content of Proposition 3 is that this implies that the representation of $J_{n,k}(f)$ in the form of a linear combination of degenerate U-statistics contains small coefficients.

Beside Proposition 3 we need another result to deduce Theorem 1 from Theorem 2. We must have some control on the exponent and parameter of the classes of functions appearing in the Hoeffding decomposition of the class of functions we consider together with a good L_2-norm of these functions. Hoeffding's decomposition is made with the help of certain projections introduced in formulas (2.7) and (2.8) below. In Lemma 1 I prove the properties of these projections I shall need later. I shall need Lemma 1 also in the proof of Proposition 2, since Hoeffding's decomposition is applied in it.

Let some measurable spaces $(Y_1,\mathcal{Y}_1)$, $(Y_2,\mathcal{Y}_2)$ and $(Z,\mathcal{Z})$ be given together with a probability measure μ on the space $(Z,\mathcal{Z})$. Consider a function $f(y_1,z,y_2)$ on the product space $(Y_1\times Z\times Y_2,\mathcal{Y}_1\times\mathcal{Z}\times\mathcal{Y}_2)$, $y_1\in\mathcal{Y}_1$, $z\in\mathcal{Z}$, $y_2\in\mathcal{Y}_2$, and define their projections

$$P_\mu f(y_1,y_2)=\int f(y_1,z,y_2)\mu(\,dz),\quad y_1\in Y_1,\ y_2\in Y_2, \tag{2.7}$$

and

$$\begin{aligned}Q_\mu f(y_1,z,y_2)&=(I-P_\mu)f(y_1,z,y_2)\\&=f(y_1,z,y_2)-P_\mu f(y_1,z,y_2),\quad y_1\in Y_1,\ z\in Z,\ y_2\in Y_2,\end{aligned} \tag{2.8}$$

where $P_\mu f(y_1,z,y_2)=P_\mu f(y_1,y_2)$, i.e. I have introduced a fictive argument z of the function $\bar P_\mu f$ in formula (2.8) to make it meaningful. Now I formulate the following

Lemma 1. *Let us have some measurable spaces $(Y_1,\mathcal{Y}_1)$, $(Y_2,\mathcal{Y}_2)$ and $(Z,\mathcal{Z})$, a probability measure μ on the space $(Z,\mathcal{Z})$ and a probability measure ρ on the product space $(Y_1\times Y_2,\mathcal{Y}_1\times\mathcal{Y}_2)$. The transformations P_μ and Q_μ defined in (2.7) and (2.8) are contractions from the space $L_2(Y_1\times Z\times Y_2,\rho\times\mu)$ to the spaces $L_2(Y_1\times Y_2,\rho)$ and $L_2(Y_1\times Z\times Y_2,\rho\times\mu)$ respectively, i.e.*

$$\begin{aligned}\|P_\mu f\|^2_{L_2,\rho}&=\int P_\mu f(y_1,z,y_2)^2\rho(\,dy_1,\,dy_2)\\&\le\|f\|^2_{L_2,\rho\times\mu}=\int f(y_1,z,y_2)^2\rho(\,dy_1,\,dy_2)\mu(\,dz),\end{aligned} \tag{2.9}$$

and

$$\begin{aligned}\|Q_\mu f\|^2_{L_2,\rho}&=\int Q_\mu f(y_1,z,y_2)^2\rho(\,dy_1,\,dy_2)\\&=\int\left(f(y_1,z,y_2)-P_\mu f(y_1,z,y_2)\right)^2\rho(\,dy_1,\,dy_2)\mu(\,dz)\\&\le\|f\|^2_{L_2,\rho\times\mu}=\int f(y_1,z,y_2)^2\rho(\,dy_1,\,dy_2)\mu(\,dz).\end{aligned} \tag{2.9$'$}$$

Also the inequalities

$$\sup_{y_1,y_2}|P_\mu f(y_1,y_2)|\le\sup_{y_1,z,y_2}|f(y_1,z,y_2)| \tag{2.10}$$

$$\sup_{y_1,z,y_2}|Q_\mu f(y_1,z,y_2)|\le 2\sup_{y_1,z,y_2}|f(y_1,z,y_2)| \tag{2.10$'$}$$

hold. If $\mathcal{F}$ is an L_2-dense class of functions $f(y_1,z,y_2)$ on the product space $(Y_1\times Z\times Y_2,\mathcal{Y}_1\times\mathcal{Z}\times Y_2)$, $y_1\in\mathcal{Y}_1$, $z\in\mathcal{Z}$, $y_2\in\mathcal{Y}_2$ with parameter D and exponent L, then also the classes $\mathcal{F}_\mu=\{P_\mu f,:f\in\mathcal{F}\}$ with the functions $P_\mu f$ defined in formulas (2.7) are L_2-dense classes with parameter D and exponent L in the space $(Y_1\times Y_2,\mathcal{Y}_1\times\mathcal{Y}_2)$. Beside this, the class of functions $\mathcal{G}_\mu=\{\frac12 Q_\mu f=\frac12(f-P_\mu f),\ f\in\mathcal{F}\}$ is also an L_2-dense class with exponent L and parameter D.

Proof of Lemma 1. The Schwarz inequality yields that $P_\mu(f)^2 \le \int f(y_1,z,y_2)^2\mu(\,dz)$, and the inequality $\int[f(y_1,z,y_2)-P_\mu f(y_1,z,y_2)]^2\mu(dz) \le \int f(y_1,z,y_2)^2\mu(\,dz)$ also holds. Integrating these inequalities with respect to the probability measure $\rho(\,dy_1,\,dy_2)$ we get formulas (2.9) and (2.9′). The proof of relations (2.10) and (2.10′) is self-evident.

Let us consider an arbitrary probability measure ρ on the space $(Y_1\times Y_2,\mathcal{Y}_1\times\mathcal{Y}_2)$. To prove that $\mathcal{F}_\mu$ is an L_2-dense class with exponent L and parameter D we have to find $m\le D\varepsilon^L$ functions $f_j\in\mathcal{F}_\mu$, $1\le j\le m$, such that $\inf\limits_{1\le j\le m}\int(f_j-f)^2\,d\rho\le\varepsilon^2$ for all $f\in\mathcal{F}_\mu$. We can find such a sequence, since a similar statement holds for the class of functions $\mathcal{F}$ in the space $Y_1\times Z\times Y_2$ with the probability measure $\rho\times\mu$. This fact together with the L_2 contraction property of P_μ formulated in (2.9) imply that $\mathcal{F}_\mu$ is an L_2-dense class.

The L_2-density property of the set $\mathcal{G}_\mu$ under the appropriate conditions can be deduced from the following observation. For any probability measure ρ on the space $Y_1\times Z\times Y_2$ and pair of functions f and g such that $\int(f-g)^2\frac12\,(\,d\rho+\,d\bar\rho\times\,du)\le\varepsilon^2$, where $\bar\rho$ is the projection of the measure ρ to the space $Y_1\times Y_2$, i.e. $\bar\rho(A)=\rho(A\times Z)$ for all $A\in\mathcal{Y}_1\times\mathcal{Y}_2$, the inequality $\int((f-P_\mu f)-(g-P_\mu g))^2\,d\rho\le 2\int(f-g)^2\,d\rho+2\int(P_\mu f-P_\mu g)^2\,d\rho\le 2\int(f-g)^2\,d\rho+2\int(f-g)^2\,d\bar\rho\times d\mu\le 4\varepsilon^2$ holds. This means that if $\{f_1,\dots,f_m\}$ is an ε-dense subset of $\mathcal{F}$ in the space $L_2(Y_1\times Z\times Y_2,\mathcal{Y}_1\times\mathcal{Z}\times\mathcal{Y}_2,\tilde\rho)$ with $\tilde\rho=\frac12(\rho+\bar\rho\times\mu)$, then $\{Q_\mu f_1,\dots,Q_\mu f_m\}$ is a 2ε-dense subset of $2\mathcal{G}_\mu=\{f-P_\mu f\colon f\in\mathcal{F}\}$ in the space $L_2(Y_1\times Z\times Y_2,\mathcal{Y}_1\times\mathcal{Z}\times\mathcal{Y}_2,\rho)$. Hence, if $\{f_1,\dots,f_m\}$ is an ε-dense subset with respect to the measure $\tilde\rho=\frac12\,(\rho+\bar\rho\times\mu)$, then $\{\frac12 Q_\mu f_1,\dots,\frac12 Q_\mu f_m\}$ is an ε-dense subset of $\mathcal{G}_\mu$ in the space $L_2(Y_1\times Z\times Y_2,\mathcal{Y}_1\times\mathcal{Z}\times\mathcal{Y}_2,\rho)$ space.

To formulate Proposition 3 first I introduce the following notation. Given a function $f(x_1,\dots,x_k)$ of k variables on $(X^k,\mathcal{X}^k)$ together with some probability measure μ let us introduce for all sets $V\subset\{1,\dots,k\}$ the function f_V depending on the arguments x_j, $j\in V$ by the formulas

$$f_V(x_s,\,s\in V)=\left(\prod_{s\in\{1,\dots,k\}\setminus V}P_{\mu,s}\prod_{s\in V}Q_{\mu,s}\right)f(x_1,\dots,x_k), \tag{2.11}$$

where $P_{\mu,s}$ and $Q_{\mu,s}$ denote the operators P_μ and Q_μ defined in formulas (2.7) and (2.8) in the space $(Y_1\times Z\times Y_2,\mathcal{Y}_1\times\mathcal{Z}\times\mathcal{Y}_2)$, where $(Y_1,\mathcal{Y}_1)$ the product of the first $s-1$, $(Y_2,\mathcal{Y}_2)$ the product of the last $k-s$ coordinates, and $(Z,\mathcal{Z})$ is the s-th coordinate of the product space $(X^k,\mathcal{X}^k)$. The function f_V depends only on the coordinates x_s, $s\in V$, because at the application of the operator $P_{\mu,s}$ the s-th coordinate disappears. It can be shown that the function f_V is canonical. To see this we have to observe that the canonical property of the function f_V can be reformulated as $P_{\mu,s}f_V\equiv0$ for all $s\in V$. Beside this, the operator $P_{\mu,s}$ or $Q_{\mu,s}$ is exchangeable with $P_{\mu,s'}$ or $Q_{\mu,s'}$ if $s\ne s'$, and $P_{\mu,s}Q_{\mu,s}=P_{\mu,s}-P_{\mu,s}^2=0$. The functions f_V defined in (2.11) appear in the Hoeffding decomposition of a U-statistic with kernel function f.

Now I formulate Proposition 3 which will be proved in the Appendix.

Proposition 3. *Let us have a non-atomic measure μ on a measurable space $(X,\mathcal{X})$ together with a sequence of independent, μ-distributed random variables $\xi_1,\dots,\xi_n$, and* *take a function $f(x_1,\dots,x_k)$ of k variables on the space $(X^k,\mathcal{X}^k)$ such that*

$$\int f^2(x_1,\dots,x_k)\mu(\,dx_1)\dots\mu(\,dx_k)<\infty.$$

Let us consider the empirical distribution function μ_n of the sequence $\xi_1,\dots,\xi_n$ introduced in (1.1) together with the k-fold random integral $J_{n,k}(f)$ of the function f defined in (1.2). The identity

$$J_{n,k}(f)=\sum_{V\subset\{1,\dots,k\}} C(n,k,V)n^{-|V|/2}I_{n,|V|}(f_V), \tag{2.12}$$

holds with the canonical (with respect to the measure μ) functions $f_V(x_j,\ j\in V)$ defined in (2.11) and appropriate real numbers $C(n,k,V)$, $V\subset\{1,\dots,k\}$, where $I_{n,|V|}(f_V)$ is the (degenerate) U-statistic with kernel function f_V and random sequence $\xi_1,\dots,\xi_n$ defined in (1.6). The constants $C(n,k,V)$ in (2.12) satisfy the relations $|C(n,k,V)|\le C(k)$ with some constant $C(k)$ depending only on the order k of the integral $J_{n,k}(f)$, $\lim\limits_{n\to\infty} C(n,k,V)=C(k,V)$ with some constant $C(k,V)<\infty$ for all $V\subset\{1,\dots,k\}$, and $C(n,k,\{1,\dots,k\})=1$ for $V=\{1,\dots,k\}$.

Theorem 1 can be simply deduced from Theorems 2, Proposition 3 and Lemma 1. Indeed, Lemma 1 together with formula (2.11) imply that if $\mathcal{F}$ is an L_2-dense class of functions with exponent L and parameter D, and the elements of $\mathcal{F}$ satisfy relations (1.3) and (1.4) with some $\sigma>0$, then for all $V\subset\{1,\dots,k\}$ the class of functions $\mathcal{F}_V=\{2^{-|V|}f_V\colon f\in\mathcal{F}\}$, where f_V is defined in (2.11), and $|V|$ denotes the cardinality of the set V is again L_2-dense with exponent L and parameter D, whose elements satisfy relations (1.3) and (1.4) with parameter $2^{-|V|}\sigma$. Beside this, the elements of $\mathcal{F}_V$ are canonical functions. Hence, by Proposition 3 we can write

$$P\left(\sup_{f\in\mathcal{F}}|J_{n,k}(f)|>x\right)\le\sum_{V\subset\{1,\dots,k\}} P\left(\sup_{f\in\mathcal{F}} n^{-|V|/2}|I_{n,|V|}(f_V)|>\frac{x}{2^kC(k)}\right) \tag{2.13}$$

with a constant $C(k)$ satisfying the inequality $C(n,k,|V|)\le C(k)$ for all coefficients $C(n,k,|V|)$ in (2.12), and each term at the right-hand side of (2.13) can be estimated by means of Theorem 2 if $\mathcal{F}$ satisfies the conditions of Theorem 1.

Theorem 1 with appropriate universal constants $M>0$, $C>0$ and $\alpha>0$ can be proved with the help of some calculation if we bound each probability on the right-hand side of (2.13) by means of Theorem 2. Let me remark that Theorem 1 implicitly contains the condition that $n\sigma^2\ge M(L+\beta)^{3/2}\log\frac{2}{\sigma}$, which means that the set of numbers x which satisfy the condition in relation (1.5) is not empty. Hence we may assume that $n\sigma^2\ge1$. We need this observation to check that under the conditions of Theorem 1 $n\sigma^2\ge\left(\frac{x}{\sigma}\right)^{2/l}$ for all $l\le k$, and we can apply Theorem 1 for each term $V\subset\{1,\dots,k\}$ in the estimation of the right-hand side of (2.13).

3. Some basic tools of the proof of Proposition 2

I shall prove Proposition 2 by means of some symmetrization procedure. The proof becomes simpler with the help of a decoupling argument. This means the introduction of decoupled U-statistics and the proof of a version of Proposition 2 about decoupled U-statistics. It can be shown with the help of some known results that this statement implies Proposition 2 in its original form. To carry out such a program first I recall the definition of decoupled U-statistics.

Definition of decoupled U-statistics. *Let k independent copies $\xi_{1,s}, \dots, \xi_{n,s}$, $1 \le s \le k$, of a sequence of independent and identically distributed random variables $\xi_1, \dots, \xi_n$ with distribution μ be given on a measurable space $(X, \mathcal{X})$ together with a function $f = f(x_1, \dots, x_k)$ on the k-th power $(X^k, \mathcal{X}^k)$ of the space $(X, \mathcal{X})$. We define with their help the decoupled U-statistic $\bar{I}_{n,k}(f)$ of order k with kernel function f by the formula*

$$\bar{I}_{n,k}(f) = \frac{1}{k!} \sum_{\substack{1 \le j_s \le n,\ s=1,\dots,k \\ j_s \ne j_{s'} \text{ if } s \ne s'}} f(\xi_{j_1,1}, \dots, \xi_{j_k,k}). \tag{3.1}$$

A decoupled U-statistic is called degenerate if its kernel function is canonical.

I shall prove the following version of Proposition 2.

Proposition $2'$. *Let us have a probability measure μ on a space $(X, \mathcal{X})$ together with k independent copies $\xi_{1,s}, \dots, \xi_{n,s}$, $1 \le s \le k$, of a sequence of independent and μ distributed random variables $\xi_1, \dots, \xi_n$ and a countable L_2-dense class $\mathcal{F}$ of canonical kernel functions $f = f(x_1, \dots, x_k)$ (with respect to the measure μ) with some parameter $D > 0$ and exponent $L \ge 1$ on the product space $(X^k, \mathcal{X}^k)$ which satisfies conditions (1.3) and (1.4) with some $0 < \sigma \le 1$. Let $n\sigma^2 > K(L+\beta)\log n$ with $\beta = \max\left(\frac{\log D}{\log n}, 0\right)$ and a sufficiently large constant $K = K(k)$. There exists some threshold index $A_0 = A_0(k) > 0$ such that the decoupled U-statistics $\bar{I}_{n,k}(f)$, $f \in \mathcal{F}$, defined in (3.6) satisfy the inequality*

$$P\left(\sup_{f \in \mathcal{F}} |n^{-k/2} \bar{I}_{n,k}(f)| \ge A n^{k/2} \sigma^{k+1}\right) \le e^{-A^{1/2k} n\sigma^2} \quad \textit{if } A \ge A_0. \tag{3.2}$$

Proposition 2 follows from Proposition $2'$ and the following Proposition A.

Proposition A. *Let us consider a countable sequence $f_l(x_1, \dots, x_k)$, $l = 1, 2, \dots$, of functions on the k-fold product $(X^k, \mathcal{X}^k)$ of some space $(X, \mathcal{X})$ together with some probability measure μ on the space $(X, \mathcal{X})$. Given a sequence of independent and identically distributed random variables $\xi_1, \xi_2, \dots$ with distribution μ on $(X, \mathcal{X})$ together with k independent copies $\xi_{1,s}, \xi_{2,s}, \dots$, $1 \le s \le k$, of it we can define the U-statistics $I_{n,k}(f_l)$ and decoupled U-statistics $\bar{I}_{n,k}(f_l)$ for all $l = 1, 2, \dots$ and $n = 1, 2, \dots$. They satisfy the inequality*

$$P\left(\sup_{1 \le l < \infty} |I_{n,k}(f_l)| > x\right) \le A P\left(\sup_{1 \le l < \infty} \left|\bar{I}_{n,k}(f_l)\right| > \gamma x\right) \tag{3.3}$$

for all $x \ge 0$ with some constants $A = A(k) > 0$ and $\gamma = \gamma(k) > 0$ depending only on the order k of the U-statistics.

Proposition A can be deduced from Theorem 1 in paper [6] of de la Peña and Montgomery–Smith which compares the distribution of a single U-statistic with its decoupled U-statistic counterpart. It holds for U-statistics with a kernel function taking values in a general separable Banach space, and it compares the distribution of the norm of a U-statistic with its decoupled counterpart. This result states that formula (3.3) remains valid if we fix a function f of k-variables taking values in a separable Banach space and replace $\sup |I_{n,k}(f_l)|$ by $\|I_{n,k}(f)\|$ and $\sup |\bar I_{n,k}(f_l)|$ by $\|\bar I_{n,k}(f)\|$. Moreover, the universal constants A and γ do not depend on the Banach space, where the function f takes its values. In the proof of Proposition A we exploit our freedom to work in an arbitrary separable Banach space.

The proof of Proposition A (with the help of paper [6].) Let us fix an arbitrary positive integer N, and apply the first part of Theorem 1 of [6] in the Banach space ℓ_∞^N consisting of the sequences $x = (x_1, \dots, x_N)$ of length N of real numbers with norm $\|x\| = \sup\limits_{1 \le l < N} |x_l|$ for the U-statistic and degenerate U-statistic with kernel functions $f_{j_1,\dots,j_k}(x_1,\dots,x_k) = \bar f(x_1,\dots,x_k)$, $\bar f = (f_1,\dots,f_N)$, with the functions f_l, $1 \le l \le N$, in Proposition A, which maps the space $(X^k, \mathcal{X}^k)$ into the space ℓ_∞^N. (Here we do not exploit that in the result of [6] the kernel functions may depend on the indices $(j_1,\dots,j_k)$.) The first part of Theorem 1 in [6] states that

$$
\begin{aligned}
&P\left(\left\| \sum_{\substack{1\le j_s\le n,\ s=1,\dots,k\\ j_s\ne j_{s'} \text{ if } s\ne s'}} \bar f\left(\xi_{j_1},\dots,\xi_{j_k}\right)\right\| > x\right)\\
&\qquad \le AP\left(\left\| \sum_{\substack{1\le j_s\le n,\ s=1,\dots,k\\ j_s\ne j_{s'} \text{ if } s\ne s'}} \bar f\left(\xi_{j_1,1},\dots,\xi_{j_k,k}\right)\right\| > \gamma x\right)
\end{aligned}
\tag{3.4}
$$

with some universal constants $A = A(k) > 0$ and $\gamma = \gamma(k) > 0$, and this statement is equivalent to a weaker version of relation (3.3), where $\sup\limits_{1 \le l < \infty}$ is replaced by $\sup\limits_{1 \le l \le N}$. We get relation (3.3) from relation (3.4) by letting $N \to \infty$ (and exploiting that the constants A and γ in formula (3.4) do not depend on the number N.)

Remark: I have introduced the number N in the above proof instead of working in the space of infinite sequences with L_∞ norm to avoid the difficulty which would arise if we had to work in non-separable Banach spaces.

Thus I have reduced the proof of Theorem 2 to that of Proposition $2'$. It will be proved by means of a symmetrization argument. To apply this argument I shall need two auxiliary results, the multi-dimensional version of Hoeffding's inequality and an appropriate generalization of a well-known symmetrization lemma. First I discuss the multi-dimensional version of Hoeffding's inequality.

Lemma 2. (The multi-dimensional version of Hoeffding's inequality.) *Let $\varepsilon_1,\dots,\varepsilon_n$ be independent random variables, $P(\varepsilon_j=1)=P(\varepsilon_j=-1)=\frac12$, $1\le j\le n$. Fix a positive integer k, and define the random variable*

$$Z=\sum_{\substack{(j_1,\dots,j_k)\colon 1\le j_l\le n \text{ for all } 1\le l\le k\\ j_l\ne j_{l'} \text{ if } l\ne l'}} a(j_1,\dots,j_k)\varepsilon_{j_1}\cdots\varepsilon_{j_k} \tag{3.5}$$

with the help of some real numbers $a(j_1,\dots,j_k)$ which are given for all sets of indices such that $1\le j_l\le n$, $1\le l\le k$, and $j_l\ne j_{l'}$ if $l\ne l'$. Put

$$S^2=\sum_{\substack{(j_1,\dots,j_k)\colon 1\le j_l\le n \text{ for all } 1\le l\le k\\ j_l\ne j_{l'} \text{ if } l\ne l'}} a^2(j_1,\dots,j_k). \tag{3.6}$$

Then

$$P(|Z|>x)\le C\exp\left\{-B\left(\frac{x}{S}\right)^{2/k}\right\}\quad \text{for all } x\ge 0 \tag{3.7}$$

with some constants $B>0$ and $C>0$ depending only on the parameter k. Relation (3.5) holds for instance with the choice $B=\frac{k}{2e(k!)^{1/k}}$ and $C=e^k$.

Lemma 2 is a relatively simple consequence of an important result of the probability theory, the hypercontractive inequality for Rademacher functions (see e.g. [4] or [5]). It yields some moment inequalities that imply Lemma 2. Such an inequality is formulated e.g. in Theorem 3.2.2 of [5]. It states (with the choice $p=2$ in this result and the observation $EZ^2\le k!S^2$) that

$$E|Z|^q\le (q-1)^{kq/2}(k!S^2)^{q/2}\qquad \text{for } q\ge 2. \tag{3.8}$$

Here I used the notation of Lemma 2.

The Markov inequality and inequality (3.8) imply that

$$P(|Z|>x)\le\left(q^{k/2}\frac{\sqrt{k!}S}{x}\right)^q\quad \text{for all } x>0 \quad \text{and } q\ge 2.$$

Choose the number q as the solution of the equation $q\left(\frac{\sqrt{k!}S}{x}\right)^{2/k}=\frac1e$. Then we get that $P(|Z|>x)\le \exp\left\{-B\left(\frac{x}{S}\right)^{2/k}\right\}$ with $B=\frac{k}{2e(k!)^{1/k}}$, provided that $q=\frac{1}{ek!^{1/k}}\left(\frac{x}{S}\right)^{2/k}\ge 2$, i.e. $B\left(\frac{x}{S}\right)^{2/k}\ge k$. By multiplying the above upper bound with $C=e^k$ we get such an estimate for $P(|Z|>x)$ which holds for all $x>0$. In such a way we get the proof of Lemma 2.

Remark: The parameter B given in Lemma 2 is not sharp. In paper [9] I have shown that the right choice of B in formula (3.7) is $B = \frac{1}{2}$.

The second result I need is a slight generalization of a simple lemma that can be found for instance in Pollard's book [12] (8° Symmetrization Lemma) or Lemma 2.5 in [7]. In this paper I need the result given in Lemma 3 below to carry out my arguments. Its proof consists of a slight modification of the method in [7] or [12].

Lemma 3. (Symmetrization Lemma) *Let $Z(n)$ and $\bar{Z}(n)$, $n = 1, 2, \dots$, be two sequences of random variables on a probability space $(\Omega, \mathcal{A}, P)$. Let a σ-algebra $\mathcal{B} \subset \mathcal{A}$ be given on the probability space $(\Omega, \mathcal{A}, P)$ together with a $\mathcal{B}$ measurable set B and two numbers $\alpha > 0$ and $\beta > 0$ such that the random variables Z_n, $n = 1, 2, \dots$, are $\mathcal{B}$ measurable, and the inequality*

$$P(|\bar{Z}_n| \le \alpha | \mathcal{B})(\omega) \ge \beta \quad \text{for all } n = 1, 2, \dots \text{ if } \omega \in B \tag{3.9}$$

holds. Then

$$P\left(\sup_{1 \le n < \infty} |Z_n| > \alpha + x\right) \le \frac{1}{\beta} P\left(\sup_{1 \le n < \infty} |Z_n - \bar{Z}_n| > x\right) + (1 - P(B)) \quad \text{for all } x > 0. \tag{3.10}$$

In particular, if the sequences Z_n, $n = 1, 2, \dots$, and $\bar{Z}_n$, $n = 1, 2, \dots$, are two independent sequences of random variables, and $P(|\bar{Z}_n| \le \alpha) \ge \beta$ for all $n = 1, 2, \dots$, then

$$P\left(\sup_{1 \le n < \infty} |Z_n| > \alpha + x\right) \le \frac{1}{\beta} P\left(\sup_{1 \le n < \infty} |Z_n - \bar{Z}_n| > x\right). \tag{3.10$'$}$$

Proof of Lemma 3. Put $\tau = \min\{n \colon |Z_n| > \alpha + x\}$ if there exists such an n, and $\tau = 0$ otherwise. Then

$$\begin{aligned}
P(\{\tau = n\} \cap B) &\le \frac{1}{\beta} \int_{\{\tau = n\} \cap B} P(|\bar{Z}_n| \le \alpha | \mathcal{B}) \, dP = \frac{1}{\beta} P(\{\tau = n\} \cap \{|\bar{Z}_n| \le \alpha\} \cap B) \\
&\le \frac{1}{\beta} P(\{\tau = n\} \cap \{|Z_n - \bar{Z}_n| > x\}) \quad \text{for all } n = 1, 2, \dots.
\end{aligned}$$

Hence

$$\begin{aligned}
P\left(\sup_{1 \le n < \infty} |Z_n| > \alpha + x\right) - (1 - P(B)) &\le P\left(\left\{\sup_{1 \le n < \infty} |Z_n| > \alpha + x\right\} \cap B\right) \\
&= \sum_{n=1}^{\infty} P(\{\tau = n\} \cap B) \le \frac{1}{\beta} \sum_{n=1}^{\infty} P(\{\tau = n\} \cap \{|Z_n - \bar{Z}_n| > x\}) \\
&\le \frac{1}{\beta} P\left(\sup_{1 \le n < \infty} |Z_n - \bar{Z}_n| > x\right).
\end{aligned}$$

Thus formula 3.10 is proved. If Z_n and $\bar{Z}_n$ are two independent sequences, and $P(|\bar{Z}_n| \le \alpha) \ge \beta$ for all $n = 1, 2, \dots$, and we define $\mathcal{B}$ as the σ-algebra generated by the random variables Z_n, $n=1,2,\dots$, then condition (3.9) is satisfied also with $B=\Omega$. Hence relation (3.10′) holds in this case. Lemma 3 is proved.

Before turning to the proof of Proposition 2′ I explain the main ideas of its proof. These ideas are taken from the paper [1] of Alexander.

Let us restrict our attention to the case $k=1$. In this case a probability of the form $P\left(n^{-1/2}\sup\limits_{f\in\mathcal F}\left|\sum\limits_{j=1}^n f(\xi_j)\right|>x\right)$ has to be estimated. By taking an independent copy of the sequence ξ_n (which disappears at the end of the of the calculation) a symmetrization argument can be applied which reduces the problem to the estimation of the probability $P\left(n^{-1/2}\sup\limits_{f\in\mathcal F}\left|\sum\limits_{j=1}^n \varepsilon_j f(\xi_j)\right|>\bar x\right)$, where the random variables ε_j, $P(\varepsilon_j=1)=P(\varepsilon_j=-1)=\frac12$, $j=1,\dots,n$, are independent, and they are independent also of the random variables ξ_j. Beside this, the number $\bar x$ is only slightly smaller than the number $x/2$. Let us bound the conditional probability of the event we have just introduced if the values random variables ξ_j are prescribed in it. This conditional probability can be bounded by means of the one-dimensional version of Lemma 2, and the estimate we get in such a way is useful if the conditional variance of the random variable we have to handle has a good upper bound. Such a bound exists, and some calculation reduces the original problem to the estimation of the probability $P\left(n^{-1/2}\sup\limits_{f\in\mathcal F'}\left|\sum\limits_{j=1}^n f(\xi_j)\right|>x^{1+\alpha}\right)$ with some new nice class of functions $\mathcal F'$ and number $\alpha>0$. This problem is very similar to the original one, but it is simpler, since the number x is replaced by a larger number $x^{1+\alpha}$ in it. By repeating this argument successively, in finitely many steps we get to an inequality that clearly holds.

The above sketched argument suggests a backward induction procedure to prove Proposition 2′. To carry out such a program I shall prove a result formulated in Proposition 4. First I introduce the following notion.

Definition of good tail behaviour for a class of U-statistics. *Let us have some measurable space $(X,\mathcal X)$ and a probability measure μ on it. Let us consider some class $\mathcal F$ of functions $f(x_1,\dots,x_k)$ on the k-fold product $(X^k,\mathcal X^k)$ of the space $(X,\mathcal X)$. Fix some positive integer n and positive number $\sigma>0$, and take k independent copies $\xi_{1,s},\dots,\xi_{n,s}$, $1\le s\le k$, of a sequence of independent μ-distributed random variables $\xi_1,\dots,\xi_n$. Let us introduce with the help of these random variables the decoupled U-statistics $\bar I_{n,k}(f)$, $f\in\mathcal F$. Given some real number $T>0$ we say that the set of decoupled U-statistics determined by the class of functions $\mathcal F$ has a good tail behaviour at level T if the following inequality holds:*

$$P\left(\sup_{f\in\mathcal F}|n^{-k/2}\bar I_{n,k}(f)|\ge An^{k/2}\sigma^{k+1}\right)\le\exp\left\{-A^{1/2k}n\sigma^2\right\}\quad \textit{for all } A\ge T. \tag{3.11}$$

Now I formulate Proposition 4 which enables us to make the inductive procedure leading to the proof of Proposition $2'$.

Proposition 4. *Let us fix a positive integer n, real number $0 < \sigma \le 2^{-(k+1)}$ and a probability measure μ on a measurable space $(X, \mathcal{X})$ together with a countable L_2-dense class $\mathcal{F}$ of canonical kernel functions $f = f(x_1, \dots, x_k)$ (with respect to the measure μ) on the k-fold product space $(X^k, \mathcal{X}^k)$ with some exponent $L \ge 1$ and parameter $D > 0$. Let us also assume that all functions $f \in \mathcal{F}$ satisfy the conditions $\sup\limits_{x_j \in X, 1 \le j \le k} |f(x_1, \dots, x_k)| \le 2^{-(k+1)}$, $\int f^2(x_1, \dots, x_k)\mu(\,dx_1)\dots\mu(\,dx_k) \le \sigma^2$, and $n\sigma^2 > K(L+\beta)\log n$ with a sufficiently large fixed number $K = K(k)$ and $\beta = \max\left(\frac{\log D}{\log n}, 0\right)$.*

There exists some real number $A_0 = A_0(k) > 1$ such that for all classes of functions $\mathcal{F}$ which satisfy the above conditions of Proposition 4 the set of decoupled U-statistics determined by the functions $f \in \mathcal{F}$ have a good tail behaviour at level T for some $T \ge A_0$, provided that for all classes of functions $\mathcal{F}$ with such properties the set of decoupled U-statistics with kernel functions $f \in \mathcal{F}$ have a good tail behaviour at level $T^{4/3}$.

It is not difficult to deduce Proposition $2'$ from Proposition 4. Indeed, let us observe that the set of (decoupled) U-statistics determined by a class of functions $\mathcal{F}$ satisfying the conditions of Proposition 4 has a good tail-behaviour at level $T_0 = \sigma^{-(k+1)}$, since the probability at the left-hand side of (3.11) equals zero for $x > \sigma^{-(k+1)}$. Then we get from Proposition 4 by induction with respect to the number j that all sets of U-statistics $\bar{I}_{n,k}(f)$, $f \in \mathcal{F}$, with a class of functions $\mathcal{F}$ satisfying the conditions of Proposition 4 have a good tail-behaviour also for $T \ge T_0^{(3/4)^j} = \sigma^{-(3/4)^j(k+1)}$ for all $j = 1, 2, \dots$ such that $\sigma^{-(3/4)^j(k+1)} \ge A_0$. This implies that if a class of functions $\mathcal{F}$ satisfies the conditions of Proposition 4, then the set of U-statistics determined by this class of functions has a good tail-behaviour at level $T = A_0^{4/3}$, i.e. at a level which depends only on the order k of the (decoupled) U-statistics. This result implies Proposition $2'$, only we have to apply it not directly for the class of functions $\mathcal{F}$ appearing in Proposition $2'$, but these functions have to be multiplied by a sufficiently small positive number depending only on k.

Thus to complete the proof of Theorem 2 it is enough to prove Proposition 4. I describe its proof in the special case $k = 1$ in the next section. This case is considered separately, because it may help to understand the ideas of the proof in the general case.

The main difficulty in the proof of Proposition 4 is related to a symmetrization procedure which is an essential part of the proof. I want to apply some randomization with the help of a symmetrization argument, and this requires a special justification. It is not difficult to justify the right for this randomization in the case $k = 1$, when it simply follows from Lemma 3 and a (simple) estimation of the variance of an appropriate U-statistic, but it becomes hard for $k \ge 2$. In this case we have to give a good estimate on certain conditional variances of some (decoupled) U-statistics with respect to some appropriate conditions. To overcome this difficulty I formulate a result in Proposition 5 and prove Propositions 4 and 5 simultaneously. Their proof follows the following line.

First Proposition 4 and Proposition 5 will be proved for $k=1$. Then, if Propositions 4 and 5 are already proven for all $k'<k$, then first I prove Proposition 4 for k, and then Proposition 5 for the same k. Proposition 5 has a structure similar to that of Proposition 4. Before its formulation I introduce the following definition.

Definition of good tail behaviour for a class of integrals of U-statistics. *Let us have a product space $(X^k\times Y, \mathcal{X}^k\times\mathcal{Y})$ with some product measure $\mu^k\times\rho$, where $(X^k,\mathcal{X}^k,\mu^k)$ is the k-fold product of some probability space $(X,\mathcal{X},\mu)$, and $(Y,\mathcal{Y},\rho)$ is some other probability space. Fix some positive integer n and positive number $\sigma>0$, and consider some class $\mathcal{F}$ of functions $f(x_1,\dots,x_k,y)$ on the product space $(X^k\times Y,\mathcal{X}^k\times\mathcal{Y},\mu^k\times\rho)$. Take k independent copies $\xi_{1,s},\dots,\xi_{n,s}$, $1\le s\le k$, of a sequence of independent, μ-distributed random variables $\xi_1,\dots,\xi_n$. For all $f\in\mathcal{F}$ and $y\in Y$ let us define the decoupled U-statistics $\bar I_{n,k}(f,y)$ by means of these random variables $\xi_{1,s},\dots,\xi_{n,s}$, $1\le s\le k$, and the kernel function $f_y(x_1,\dots,x_k)=f(x_1,\dots,x_k,y)$ in formula (3.1). Define with the help of these U-statistics $\bar I_{n,k}(f,y)$ the random integrals*

$$H_{n,k}(f)=\int \bar I_{n,k}(f,y)^2\rho(\,dy),\quad f\in\mathcal{F}. \tag{3.12}$$

Choose some real number $T>0$. We say that the set of random integrals $H_{n,k}(f)$, $f\in\mathcal{F}$, have a good tail behaviour at level T if

$$P\left(\sup_{f\in\mathcal{F}} n^{-k}H_{n,k}(f)\ge A^2n^k\sigma^{2k+2}\right)\le\exp\left\{-A^{1/(2k+1)}n\sigma^2\right\}\quad \textit{for } A\ge T. \tag{3.13}$$

Proposition 5. *Fix some positive integer n and real number $0<\sigma\le 2^{-(k+1)}$, and let us have a product space $(X^k\times Y,\mathcal{X}^k\times\mathcal{Y})$ with some product measure $\mu^k\times\rho$, where $(X^k,\mathcal{X}^k,\mu^k)$ is the k-fold product of some probability space $(X,\mathcal{X},\mu)$, and $(Y,\mathcal{Y},\rho)$ is some other probability space. Let us have a countable L_2-dense class $\mathcal{F}$ of canonical functions $f(x_1,\dots,x_k,y)$ on the product space $(X^k\times Y,\mathcal{X}^k\times\mathcal{Y},\mu^k\times\rho)$ with some exponent $L\ge1$ and parameter $D>0$. Let us also assume that the functions $f\in\mathcal{F}$ satisfy the conditions*

$$\sup_{x_j\in X, 1\le j\le k, y\in Y}|f(x_1,\dots,x_k,y)|\le 2^{-(k+1)}$$

and

$$\int f^2(x_1,\dots,x_k,y)\mu(\,dx_1)\dots\mu(\,dx_k)\rho(\,dy)\le\sigma^2\quad\textit{for all } f\in\mathcal{F}.$$

Let the inequality $n\sigma^2>K(L+\beta)\log n$ hold with a sufficiently large fixed number $K=K(k)$.

There exists some number $A_0=A_0(k)>1$ such that for all classes of functions $\mathcal{F}$ which satisfy the conditions of Proposition 5 the random integrals $H_{n,k}(f)$, $f\in\mathcal{F}$, defined in (3.9) have a good tail behaviour at level T, provided that the random integrals

$H_{n,k}(f)$, $f \in \mathcal{F}$, of all classes $\mathcal{F}$ with such properties have a good tail behaviour at level $T^{(2k+1)/2k}$.

Similarly to the argument formulated after Proposition 4 an inductive procedure yields the following corollary of Proposition 5.

Corollary of Proposition 5. *If the class of functions $\mathcal{F}$ satisfies the conditions of Proposition 5, then there exists a constant $\bar{A}_0 = \bar{A}_0(k) > 0$ depending only on k such that the integrals $H_{n,k}(f)$ determined by the class of functions $\mathcal{F}$ have a good tail behaviour at level $\bar{A}_0$.*

4. The proof of Proposition 4 in the case $k = 1$

In this section Proposition 4 is proved in the special case $k = 1$. In this case we have to show that

$$P\left(\frac{1}{\sqrt{n}}\sup_{f\in\mathcal{F}}\left|\sum_{j=1}^n f(\xi_j)\right| \ge An^{1/2}\sigma^2\right) \le e^{-A^{1/2}n\sigma^2} \quad \text{if } A \ge T \tag{4.1}$$

if we know the same estimate for $A > T^{4/3}$ and all classes of functions satisfying the conditions of Proposition 4. This statement will be proved by means of the following symmetrization argument.

Lemma 4. *Let the class of functions $\mathcal{F}$ satisfy the conditions of Proposition 4 for $k = 1$. Let $\varepsilon_1, \dots, \varepsilon_n$ be a sequence of independent random variables, $P(\varepsilon_j = 1) = P(\varepsilon_j = -1) = \frac{1}{2}$, independent also of the μ distributed random variables $\xi_1, \dots, \xi_n$. Then*

$$\begin{aligned} &P\left(\frac{1}{\sqrt{n}}\sup_{f\in\mathcal{F}}\left|\sum_{j=1}^n f(\xi_j)\right| \ge An^{1/2}\sigma^2\right) \\ &\qquad \le 4P\left(\frac{1}{\sqrt{n}}\sup_{f\in\mathcal{F}}\left|\sum_{j=1}^n \varepsilon_j f(\xi_j)\right| \ge \frac{A}{3}n^{1/2}\sigma^2\right) \quad \textit{if } A \ge T. \end{aligned} \tag{4.2}$$

There are several similar results in the literature. Lemma 4 follows simply from Part b) of Lemma 2.7 in [7] with the choice $t = An\sigma^2$. (The quantity α in this result agrees with our σ.) It is enough to check that $t \ge 2^{1/2}n^{1/2}\sigma$ and $(t - 2^{1/2}n^{1/2}\sigma)/2 \ge \frac{A}{3}n\sigma^2$ if $A \ge T \ge A_0$ is chosen sufficiently large, since under the conditions of Proposition 4 (if the parameter K is sufficiently large in Proposition 4) $n\sigma^2 \ge 1$.

To prove Proposition 4 for $k = 1$ let us investigate the conditional probability

$$P(f, A|\xi_1, \dots, \xi_n) = P\left(\frac{1}{\sqrt{n}}\left|\sum_{j=1}^n \varepsilon_j f(\xi_j)\right| \ge \frac{A}{6}\sqrt{n}\sigma^2 \,\middle|\, \xi_1, \dots, \xi_n\right)$$

for all functions $f\in\mathcal{F}$, $A\ge T$ and values $(\xi_1,\dots,\xi_n)$. By Lemma 2 (with $k=1$) we can write

$$P(f,A|\xi_1,\dots,\xi_n)\le C\exp\left\{-\frac{\frac{B}{36}A^2n\sigma^4}{S^2(f,\xi_1,\dots,\xi_n)}\right\} \tag{4.3}$$

with

$$S^2(f,x_1,\dots,x_n)=\frac1n\sum_{j=1}^n f^2(x_j),\qquad f\in\mathcal{F}.$$

Let us introduce the set

$$H=H(A)=\left\{(x_1,\dots,x_n)\colon \sup_{f\in\mathcal{F}} S^2(f,x_1,\dots,x_n)\ge\left(1+A^{4/3}\right)\sigma^2\right\}. \tag{4.4}$$

I claim that

$$P((\xi_1,\dots,\xi_n)\in H)\le e^{-A^{2/3}n\sigma^2}\qquad\text{if } A\ge T. \tag{4.5}$$

To prove relation (4.5) let us consider the functions $\bar f=\bar f(f)$ for all $f\in\mathcal{F}$ defined by the formula $\bar f(x)=f^2(x)-\int f^2(x)\mu(\,dx)$, and introduce the class of functions $\mathcal{F}'=\{\bar f(f)\colon f\in\mathcal{F}\}$. Let us show that the class of functions $\mathcal{F}'$ satisfies the conditions of Proposition 4. By the assumption of Proposition 4 this implies that the estimate (3.11) with $k=1$, i.e. the estimate (4.1) holds for the class of functions $\mathcal{F}'$ if the condition $A\ge T$ is replaced by $A\ge T^{4/3}$ in it.

Relation $\int\bar f(x)\mu(\,dx)=0$ clearly holds. (In the case $k=1$ this means that $\bar f$ is a canonical function.) The condition $\sup|\bar f(x)|\le\frac18<\frac14$ also holds if $\sup|f(x)|\le\frac14$, and $\int\bar f^2(x)\mu(\,dx)\le\int f^4(x)\mu(\,dx)\le\frac14\int f^2(x)\,\mu(\,dx)\le\frac{\sigma^2}{4}<\sigma^2$ if $f\in\mathcal{F}$. It remained to show that $\mathcal{F}'$ is an L_2-dense class with exponent L and parameter D.

To show this observe that $\int(\bar f(x)-\bar g(x))^2\rho(\,dx)\le 2\int(f^2(x)-g^2(x))^2\rho(\,dx)+2\int(f^2(x)-g^2(x))^2\mu(\,dx)\le 2(\sup(|f(x)|+|g(x)|)^2\left(\int(f(x)-g(x))^2(\rho(\,dx)+\mu(\,dx)\right)\le\int(f(x)-g(x))^2\bar\rho(\,dx)$ for all $f,g\in\mathcal{F}$, $\bar f=\bar f(f)$, $\bar g=\bar g(g)$ and probability measure ρ, where $\bar\rho=\frac{\rho+\mu}{2}$. This means that if $\{f_1,\dots,f_m\}$ is an ε-dense subset of $\mathcal{F}$ in the space $L_2(X,\mathcal{X},\bar\rho)$, then $\{\bar f_1,\dots,\bar f_m\}$ is an ε-dense subset of $\mathcal{F}'$ in the space $L_2(X,\mathcal{X},\rho)$, and not only $\mathcal{F}$, but also $\mathcal{F}'$ is an L_2-dense class with exponent L and parameter D.

We get, by applying formula (4.1) for the number $A^{4/3}\ge T^{4/3}$ and the class of functions $\mathcal{F}'$ that

$$\begin{aligned}P((\xi_1,\dots,\xi_n)\in H)&=P\left(\sup_{f\in\mathcal{F}}\left(\frac1n\sum_{j=1}^n\bar f(\xi_j)+\frac1n\sum_{j=1}^n Ef^2(\xi_j)\right)\ge\left(1+A^{4/3}\right)\sigma^2\right)\\&\le P\left(\sup_{f\in\mathcal{F}}\frac{1}{\sqrt n}\sum_{j=1}^n\bar f(\xi_j)\ge A^{4/3}n^{1/2}\sigma^2\right)\le e^{-A^{2/3}n\sigma^2},\end{aligned}$$

i.e. relation (4.5) holds.

Formula (4.3) and the definition (4.4) of the set H yield the estimate

$$P(f,A|\xi_1,\dots,\xi_n)\le Ce^{-BA^{2/3}n\sigma^2/40} \quad \text{if } (\xi_1,\dots,\xi_n)\notin H \tag{4.6}$$

for all $f\in\mathcal{F}$ and $A\ge T$ for the conditional probability $P(f,A|\xi_1,\dots,\xi_n)$. Let us introduce the conditional probability

$$P(\mathcal{F},A|\xi_1,\dots,\xi_n)=P\left(\sup_{f\in\mathcal{F}}\frac{1}{\sqrt{n}}\left|\sum_{j=1}^n\varepsilon_jf(\xi_j)\right|\ge\frac{A}{3}\sqrt{n}\sigma^2\,\middle|\,\xi_1,\dots,\xi_n\right)$$

for all $(\xi_1,\dots,\xi_n)$ and $A\ge T$. We shall estimate this conditional probability with the help of relation (4.6) if $(\xi_1,\dots,\xi_n)\notin H$. Given some set of n points $(x_1,\dots,x_n)$ in the space $(X,\mathcal{X})$ let us introduce the measure $\nu=\nu(x_1,\dots,x_n)$ on $(X,\mathcal{X})$ in such a way that ν is concentrated in the points $x_1,\dots,x_n$, and $\nu(\{x_j\})=\frac{1}{n}$. If $\int f^2(x)\nu(\,dx)\le\delta^2$ for a function f, then $\left|\frac{1}{\sqrt{n}}\sum\limits_{j=1}^n\varepsilon_jf(x_j)\right|\le n^{1/2}\int|f(x)|\nu(\,dx)\le n^{1/2}\delta$. Since we have assumed that $n\sigma^2\ge1$, this estimate implies that if f and g are two functions such that $\int(f-g)^2\nu(\,dx)\le\delta^2$ with $\delta=\frac{A}{6n}$, then $\left|\frac{1}{\sqrt{n}}\sum\limits_{j=1}^n\varepsilon_jf(x_j)-\frac{1}{\sqrt{n}}\sum\limits_{j=1}^n\varepsilon_jg(x_j)\right|\le\frac{A}{6\sqrt{n}}\le\frac{A}{6}\sqrt{n}\sigma^2$.

Given some (random) point $(\xi_1,\dots,\xi_n)\in H$ let us consider the measure $\nu=\nu(\xi_1,\dots,\xi_n)$ corresponding to it, and choose a $\bar\delta$-dense subset $\{f_1,\dots,f_m\}$ of $\mathcal{F}$ in the space $L_2(X,\mathcal{X},\nu)$ with $\bar\delta=\frac{1}{6n}\le\delta=\frac{A}{6n}$, whose cardinality m satisfies the inequality $m\le D\bar\delta^{-L}$. This is possible because of the L_2-dense property of the class $\mathcal{F}$. (This is the point where the L_2-dense property of the class of functions $\mathcal{F}$ is exploited in its full strength.) The above facts imply that $P(\mathcal{F},A|\xi_1,\dots,\xi_n)\le\sum\limits_{l=1}^mP(f_l,A|\xi_1,\dots,\xi_n)$ with these functions $f_1,\dots,f_m$. Hence relation (4.6) yields that

$$P(\mathcal{F},A|\xi_1,\dots,\xi_n)\le CD(6n)^Le^{-BA^{2/3}n\sigma^2/40}\quad\text{if } (\xi_1,\dots,\xi_n)\notin H \text{ and } A\ge T.$$

This inequality together with Lemma 4 and estimate (4.5) imply that

$$\begin{aligned}P\left(\frac{1}{\sqrt{n}}\sup_{f\in\mathcal{F}}\left|\sum_{j=1}^nf(\xi_j)\right|\ge An^{1/2}\sigma^2\right)&\le4P\left(\frac{1}{\sqrt{n}}\sup_{f\in\mathcal{F}}\left|\sum_{j=1}^n\varepsilon_jf(\xi_j)\right|\ge\frac{A}{3}n^{1/2}\sigma^2\right)\\&\le4CD(6n)^Le^{-BA^{2/3}n\sigma^2/40}+4e^{-A^{2/3}n\sigma^2}\quad\text{if } A\ge T.\end{aligned}\tag{4.7}$$

Since we have a better power of A in the exponent at the right-hand side of formula (4.7) than we need, the relation $n\sigma^2\ge K(L+\beta)\log n$ holds, and we have the right to choose the constants K and A_0, $A\ge A_0$, sufficiently large, it is not difficult to deduce relation (4.1) from relation (4.7). Indeed, the expression in the exponent at the right-hand side of (4.7) satisfies the inequality $\frac{B}{40}A^{2/3}n\sigma^2 \ge A^{1/2}n\sigma^2 + K(L+\beta)\log n$ if A_0 is sufficiently large, and

$$P\left(\frac{1}{\sqrt{n}}\sup_{f\in\mathcal{F}}\left|\sum_{j=1}^{n} f(\xi_j)\right| \ge An^{1/2}\sigma^2\right)$$
$$\le 4C(6n)^{\beta+L}e^{-K}n^{-K(L+\beta)}e^{-A^{1/2}n\sigma^2} + 4e^{-A^{2/3}n\sigma^2} \le e^{-A^{1/2}n\sigma^2}$$

if $A \ge T$, and the constants A_0 and K are chosen sufficiently large.

5. The symmetrization argument

In the proof of Propositions 4 and 5 we need two symmetrization results for all $k \ge 1$ which play the same role as Lemma 4 in the case $k=1$. These results are described in Lemmas 5A and 5B. In this section these results are formulated and proved. The proofs go by induction with respect to k. During the proof of Propositions 4 and 5 for k we may assume that they hold for $k' < k$.

Lemma 5A. *Let $\mathcal{F}$ be a class of functions on the space $(X^k, \mathcal{X}^k)$ which satisfies the conditions of Proposition 4 with some probability measure μ. Let us have k independent copies $\xi_{1,s},\dots,\xi_{n,s}$, $1 \le s \le k$, of a sequence of independent μ distributed random variables $\xi_1,\dots,\xi_n$, and a sequence of independent random variables $\varepsilon = (\varepsilon_1,\dots,\varepsilon_n)$, $P(\varepsilon_s = 1) = P(\varepsilon_s = -1) = \frac{1}{2}$, which is independent also of the random variables $\xi_{j,s}$, $1 \le j \le n$, $1 \le s \le k$. Consider the decoupled U-statistics $\bar{I}_{n,k}(f)$, $f \in \mathcal{F}$, defined from these random variables by formula (3.1) and their randomized version*

$$\bar{I}^{\varepsilon}_{n,k}(f) = \frac{1}{k!}\sum_{\substack{1\le j_s\le n,\ s=1,\dots,k\\ j_s\ne j_{s'} \text{ if } s\ne s'}} \varepsilon_{j_1}\cdots\varepsilon_{j_k} f\left(\xi_{j_1,1},\dots,\xi_{j_k,k}\right), \quad f\in\mathcal{F}. \tag{5.1}$$

There exists some constant $A_0 = A_0(k)$ such that the inequality

$$P\left(\sup_{f\in\mathcal{F}} n^{-k/2}\left|\bar{I}_{n,k}(f)\right| > An^{k/2}\sigma^{k+1}\right) < 2^{k+1}P\left(\sup_{f\in\mathcal{F}}\left|\bar{I}^{\varepsilon}_{n,k}(f)\right| > 2^{-(k+1)}An^k\sigma^{k+1}\right) + 2^k n^{k-1} e^{-A^{1/(2k-1)}n\sigma^2/k} \tag{5.2}$$

holds for all $A \ge A_0$.

Before formulating Lemma 5B needed in the proof of Proposition 5 I introduce some notations. Some of them will be needed later.

Let us consider a set $\mathcal{F}$ of functions $f(x_1,\dots,x_k,y) \in \mathcal{F}$ on a space $(X^k\times Y, \mathcal{X}^k\times\mathcal{Y}, \mu^k\times\rho)$ which satisfies the conditions of Proposition 5. Let us choose $2k$ independent copies $\xi^{(1)}_{1,s},\dots,\xi^{(1)}_{n,s}$, $\xi^{(-1)}_{1,s},\dots,\xi^{(-1)}_{n,s}$, $1\le s\le k$, of a sequence of independent μ distributed random variables $\xi_1,\dots,\xi_n$ together with a sequence of independent random variables $(\varepsilon_1,\dots,\varepsilon_n)$, $P(\varepsilon_s=1)=P(\varepsilon_s=-1)=\frac12$, $1\le s\le n$, which are independent of them. For all subsets $V\subset\{1,\dots,k\}$ of the set $\{1,\dots,k\}$ let $|V|$ denote the cardinality of this set, and define for all functions $f(x_1,\dots,x_k,y)\in\mathcal F$ and $V\subset\{1,\dots,k\}$ the decoupled U-statistics

$$\bar I^V_{n,k}(f,y)=\frac1{k!}\sum_{\substack{1\le j_s\le n,\ s=1,\dots,k\\ j_s\ne j_{s'} \text{ if } s\ne s'}} f\left(\xi^{(\delta_1)}_{j_1,1},\dots,\xi^{(\delta_k)}_{j_k,k},y\right),\quad f\in\mathcal F, \tag{5.3}$$

where $\delta_s=\pm1$, $1\le s\le k$, $\delta_s=1$ if $s\in V$, and $\delta_s=-1$ if $s\notin V$, together with the random variables

$$H^V_{n,k}(f)=\int \bar I^V_{n,k}(f,y)^2\rho(\,dy),\quad f\in\mathcal F. \tag{5.3$'$}$$

Put

$$\bar I_{n,k}(f,y)=\bar I^{\{1,\dots,k\}}_{n,k}(f,y),\quad H_{n,k}(f)=H^{\{1,\dots,k\}}_{n,k}(f), \tag{5.3$''$}$$

i.e. these random variables agree with those defined in (5.3) and (5.3$'$) with the choice $V=\{1,\dots,k\}$.

Let us also define the 'randomized version' of the random variables $\bar I^V_{n,k}(f,y)$ and $H^V_{n,k}(f)$ as

$$\bar I^{(V,\varepsilon)}_{n,k}(f,y)=\frac1{k!}\sum_{\substack{1\le j_s\le n,\ s=1,\dots,k\\ j_s\ne j_{s'} \text{ if } s\ne s'}} \varepsilon_{j_1}\cdots\varepsilon_{j_k} f\left(\xi^{(\delta_1)}_{j_1,1},\dots,\xi^{(\delta_k)}_{j_k,k},y\right),\quad f\in\mathcal F, \tag{5.4}$$

where $\delta_s=1$ if $s\in V$, and $\delta_s=-1$ if $s\notin V$, and

$$H^{(V,\varepsilon)}_{n,k}(f)=\int \bar I^{(V,\varepsilon)}_{n,k}(f,y)^2\rho(\,dy),\quad f\in\mathcal F. \tag{5.4$'$}$$

Let us also introduce the random variables

$$\bar W(f)=\int\left[\sum_{V\subset\{1,\dots,k\}}(-1)^{|V|}\bar I^{(V,\varepsilon)}_{n,k}(f,y)\right]^2\rho(\,dy),\quad f\in\mathcal F. \tag{5.5}$$

Now I formulate the symmetrization result applied in the proof of Proposition 5.

Lemma 5B. *Let $\mathcal F$ be a set of functions on $(X^k\times Y,\mathcal X^k\times\mathcal Y)$ which satisfies the conditions of Proposition 5 with some probability measure $\mu^k\times\rho$. Let us have $2k$ independent copies $\xi^{\pm1}_{1,s},\dots,\xi^{\pm1}_{n,s}$, $1\le s\le k$, of a sequence of independent μ distributed random variables together with a sequence of independent random variables $\varepsilon_1,\dots,\varepsilon_n$, $P(\varepsilon_s=1)=P(\varepsilon_s=-1)=\frac12$, $1\le s\le n$, which is independent of them.*

There exists some $A_0=A_0(k)$ such that if the integrals $H_{n,k}(f)$, $f\in\mathcal{F}$, determined by this class of functions $\mathcal{F}$ have a good tail behaviour at level $T^{(2k+1)/2k}$ for some $T\ge A_0$, (this property was defined at the end of Section 3), then the inequality

$$P\left(\sup_{f\in\mathcal{F}} H_{n,k}(f)>A^2n^{2k}\sigma^{2(k+1)}\right)<2P\left(\sup_{f\in\mathcal{F}}\left|\bar W(f)\right|>\frac{A^2}{2}n^{2k}\sigma^{2(k+1)}\right)+2^{2k+1}n^{k-1}e^{-A^{1/2k}n\sigma^2/k} \tag{5.6}$$

holds with the random variables $H_{n,k}(f)$ and $\bar W(f)$ defined in formulas (5.3″) *and (5.5) for all $A\ge T$.*

Let us observe that in the symmetrization argument of Lemma 5B we have applied the randomization $\bar I_{n,k}^{(V,\varepsilon)}(f,y)$ of $\bar I_{n,k}^{(V)}(f,y)$, (compare formulas (5.3) and (5.4)), and compared the integral of the square of the random function $\bar I_{n,k}(f,y)$ with the integral of the square of a linear combination of the random functions $\bar I_{n,k}^{(V,\varepsilon)}(f,y)$. After this integration the effect of the 'randomizing factors' ε_j will be weaker. Nevertheless, also such an estimate will be sufficient for us. But the effect of this symmetrization procedure has to be followed more carefully. Hence a corollary of Lemma 5B will be presented which can be better applied than the original lemma. We get it by rewriting the random variable $\bar W(f)$ defined in (5.5) in another form with the help of some diagrams introduced below.

Let $\mathcal{G}=\mathcal{G}(k)$ denote the set of all diagrams consisting of two rows such that both rows are the set $\{1,\dots,k\}$ and the diagrams of $\mathcal{G}$ contain some edges $(l_1,l_1'),\dots,(l_s,l_s')$, $0\le s\le k$ connecting some points (vertices) of the first row with some points of the second row. The vertices $l_1,\dots,l_s$ in the first row are all different, and the same relation holds also for the vertices $l_1',\dots,l_s'$ in the second row. For each diagram $G\in\mathcal{G}$ let us define $e(G)=\{(l_1,l_1')\dots,(l_s,l_s')\}$, the set of its edges, $v_1(G)=\{l_1,\dots,l_s\}$, the set of vertices in the first row and $v_2(G)=\{l_1',\dots,l_s'\}$, the set of vertices in the second row of $\mathcal{G}$ from which an edge starts.

Given some diagram $G\in\mathcal{G}$ and two sets $V_1,V_2\subset\{1,\dots,k\}$, we define with the help of the random variables $\xi_{s,1}^{(1)},\dots,\xi_{s,n}^{(1)}$, $\xi_{s,1}^{(-1)},\dots,\xi_{s,n}^{(-1)}$, $1\le s\le k$, and $\varepsilon=(\varepsilon_1,\dots,\varepsilon_n)$ taking part in the definition of the expressions $\bar W(f)$, $f\in\mathcal{F}$, the random variables $H_{n,k}(f|G,V_1,V_2)$:

$$\begin{aligned}H_{n,k}(f|G,V_1,V_2)=&\sum_{\substack{(j_1,\dots,j_k,\ j_1',\dots,j_k')\\ 1\le j_s\le n,\, j_s\ne j_{s'} \text{ if } s\ne s',\, 1\le s,s'\le k,\\ 1\le j_s'\le n,\, j_s'\ne j_{s'}' \text{ if } s\ne s',\, 1\le s,s'\le k,\\ j_s=j'_{s'} \text{ if } (s,s')\in e(G),\ j_s\ne j'_{s'} \text{ if } (s,s')\notin e(G)}}\ \prod_{s\in\{1,\dots,k\}\setminus v_1(G)}\varepsilon_{j_s}\prod_{s\in\{1,\dots,k\}\setminus v_2(G)}\varepsilon_{j_s'}\\ &\frac{1}{k!^2}\int f(\xi_{j_1,1}^{(\delta_1)},\dots,\xi_{j_k,k}^{(\delta_k)},y)f(\xi_{j_1',1}^{(\bar\delta_1)},\dots,\xi_{j_k',k}^{(\bar\delta_k)},y)\rho(\,dy),\quad f\in\mathcal{F},\end{aligned} \tag{5.7}$$

where $\delta_s=1$ if $s\in V_1$, $\delta_s=-1$ if $s\notin V_1$, and $\bar\delta_s=1$ if $s\in V_2$, $\bar\delta_s=-1$ if $s\notin V_2$.

With the help of these random variables we can write that

$$\bar W(f)=\sum_{G\in\mathcal{G},\,V_1,V_2\subset\{1,\dots,k\}}(-1)^{|V_1|+|V_2|}H_{n,k}(f|G,V_1,V_2)\quad\text{for all } f\in\mathcal{F},$$

because

$$\int \bar I_{n,k}^{(V_1,\varepsilon)}(f,y)\bar I_{n,k}^{(V_2,\varepsilon)}(f,y)\rho(\,dy)=\sum_{G\in\mathcal{G}}H_{n,k}(f|G,V_1,V_2),\quad \text{for all } V_1,V_2\subset\{1,\dots,k\}.$$

Since the number of terms in this sum is less than $2^{4k}k!$, it implies that Lemma 5B has the following corollary:

Corollary of Lemma 5B. *Let a set of functions $\mathcal{F}$ satisfy the conditions of Proposition 5. Then there exists some $A_0=A_0(k)$ such that if the integrals $H_{n,k}(f)$, $f\in\mathcal{F}$, determined by this class of functions $\mathcal{F}$ have a good tail behaviour at level $T^{(2k+1)/2k}$ for some $T\ge A_0$, then the inequality*

$$\begin{aligned}P&\left(\sup_{f\in\mathcal{F}}H_{n,k}(f)>A^2n^{2k}\sigma^{2(k+1)}\right)\\ &\qquad\le 2\sum_{G\in\mathcal{G},\,V_1,V_2\in\{1,\dots,k\}}P\left(\sup_{f\in\mathcal{F}}|H_{n,k}(f|G,V_1,V_2)|>\frac{A^2}{2^{4k+1}k!}n^{2k}\sigma^{2(k+1)}\right)\\ &\qquad\quad+2^{2k+1}n^{k-1}e^{-A^{1/2k}n\sigma^2/k}\end{aligned}\tag{5.8}$$

holds with the random variables $H_{n,k}(f)$ and $H_{n,k}(f|G,V_1,V_2)$ defined in formulas (5.3″) *and (5.7) for all $A\ge T$.*

The proof of Lemmas 5A and 5B uses the result of the following Lemma 6 which states that certain random vectors have the same distribution.

Lemma 6. *Let $\varepsilon=(\varepsilon_1,\dots,\varepsilon_n)$ be a sequence of independent random variables, $P(\varepsilon_s=1)=P(\varepsilon_s=-1)=\frac12$, $1\le s\le n$, which is independent also of $2k$ fixed independent copies $\xi_{1,s}^{(1)},\dots,\xi_{n,s}^{(1)}$ and $\xi_{1,s}^{(-1)},\dots,\xi_{n,s}^{(-1)}$, $1\le s\le k$, of a sequence $\xi_1,\dots,\xi_n$ of independent μ distributed random variables.*

a) Let $\mathcal{F}$ be a class of functions which satisfies the conditions of Proposition 4. With the help of the above random variables introduce the decoupled U-statistic

$$\bar I_{n,k}^{V}(f)=\frac{1}{k!}\sum_{\substack{1\le j_s\le n,\ s=1,\dots,k\\ j_s\ne j_{s'}\text{ if } s\ne s'}}f\left(\xi_{j_1,1}^{(\delta_1)},\dots,\xi_{j_k,k}^{(\delta_k)}\right),\quad f\in\mathcal{F},\tag{5.9}$$

for all sets $V\subset\{1,\dots,k\}$ and functions $f\in\mathcal{F}$ together with its 'randomized version'

$$\bar I_{n,k}^{(V,\varepsilon)}(f)=\frac{1}{k!}\sum_{\substack{1\le j_s\le n,\ s=1,\dots,k\\ j_s\ne j_{s'}\text{ if } s\ne s'}}\varepsilon_{j_1}\cdots\varepsilon_{j_k}f\left(\xi_{j_1,1}^{(\delta_1)},\dots,\xi_{j_k,k}^{(\delta_k)}\right),\quad f\in\mathcal{F},\tag{5.9'}$$

where $\delta_s=\pm1$, $1\le s\le k$, $\delta_s=1$ if $s\in V$, and $\delta_s=-1$ if $s\notin V$.

Then the sets of random variables

$$S(f)=\sum_{V\subset\{1,\dots,k\}}(-1)^{|V|}\bar I_{n,k}^V(f),\quad f\in\mathcal{F},\tag{5.10}$$

and sets of random variables

$$\bar S(f)=\sum_{V\subset\{1,\dots,k\}}(-1)^{|V|}\bar I_{n,k}^{(V,\varepsilon)}(f),\quad f\in\mathcal{F},\tag{5.10$'$}$$

have the same joint distribution.

b) Let $\mathcal{F}$ be a class of functions satisfying Proposition 5. For all functions $f\in\mathcal{F}$ and $V\subset\{1,\dots,k\}$ consider the decoupled U-statistics $\bar I_{n,k}^V(f,y)$ determined by the random variables $\xi_{1,s}^{(1)},\dots,\xi_{n,s}^{(1)}$ and $\xi_{1,s}^{(-1)},\dots,\xi_{n,s}^{(-1)}$, $1\le s\le k$, by formula (5.3), and define with their help the random variables

$$W(f)=\int\left[\sum_{V\subset\{1,\dots,k\}}(-1)^{|V|}\bar I_{n,k}^V(f,y)\right]^2\rho(\,dy),\quad f\in\mathcal{F}.\tag{5.11}$$

Then the random vectors $\{W(f)\colon f\in\mathcal{F}\}$ defined in (5.11) and $\{\bar W(f)\colon f\in\mathcal{F}\}$ defined in (5.5) have the same distribution.

Proof of Lemma 6. Let us consider Part a) of Lemma 6. I claim that for all $M\in\{1,\dots,n\}$ the conditional distribution of the random vector in (5.10$'$) under the condition that $\varepsilon_j=1$ if $j\in M$ and $\varepsilon_j=-1$ if $\varepsilon_j\in\{1,\dots,n\}\setminus M$ agrees with the distribution of the vector in (5.10). Since the distribution of the vector in (5.10) does not change if we exchange the random variables $\xi_{j,s}^{(1)}$ and $\xi_{j,s}^{(-1)}$ in it for $j\notin M$, $1\le s\le k$, and do not exchange them otherwise, it is enough to understand that the random vector we get from the vector in (5.10) after this transformation agrees with the random vector in (5.10$'$) if we write $\varepsilon_j=1$ for $j\in M$ and $\varepsilon_j=-1$ for $j\notin M$ in it. These random vectors really agree (not only in distribution) since for all functions $f\in\mathcal{F}$ both vectors have a component which is the sum of terms of the form $f(\xi_{j_1,1}^{(\delta_{j_1})},\dots,\xi_{j_k,k}^{(\delta_{j_k})})$, $\delta_{j_s}=\pm1$, $1\le s\le k$, multiplied with an appropriate power of -1, and this power equals the number of -1 components in the sequence $\delta_{j_1},\dots,\delta_{j_k}$ plus the cardinality of the set $\{j_1,\dots,j_k\}\cap M$.

Part b) of Lemma 6 can be proved in the same way, hence it is omitted.

Lemma 5A will be proved with the help of part a) of Lemma 6 and the following Lemma 7A.

Lemma 7A. *Let us consider a class of functions $\mathcal{F}$ satisfying the conditions of Proposition 4 and the random variables $\bar I_{n,k}^V(f)$, $f\in\mathcal{F}$, $V\subset\{1,\dots,k\}$, defined in formula (5.1). Let $\mathcal{B}=\mathcal{B}(\xi_{1,s}^{(1)},\dots,\xi_{n,s}^{(1)};\,1\le s\le k)$ denote the σ-algebra generated by the* *random variables* $\xi^{(1)}_{1,s},\dots,\xi^{(1)}_{n,s}$ *,* $1\le s\le k$*, taking part in the definition of the random variables* $\bar I^V_{n,k}(f)$*. For all* $V\subset\{1,\dots,k\}$*,* $V\ne\{1,\dots,k\}$*, there exists a number* $A_0=A_0(k)>0$ *such that the inequality*

$$P\left(\sup_{f\in\mathcal F} E\left(\bar I^V_{n,k}(f)^2\big|\,\mathcal B\right)>2^{-(3k+3)}A^2n^{2k}\sigma^{2k+2}\right)<n^{k-1}e^{-A^{1/(2k-1)}n\sigma^2/k} \tag{5.12}$$

holds for all $A\ge A_0$*.*

Proof of Lemma 7A. Let us first consider the case $V=\emptyset$. We have $E\left(\bar I^{\emptyset}_{n,k}(f)^2\Big|\,\mathcal B\right)=E\left(\bar I^{\emptyset}_{n,k}(f)^2\right)\le\frac{n!}{k!}\sigma^2\le n^{2k}\sigma^{2k+2}$ for all $f\in\mathcal F$. In the above calculation we exploited that the functions $f\in\mathcal F$ are canonical, and this implies certain orthogonalities, and beside this the inequality $n\sigma^2\ge1$ holds. The above relation implies inequality (5.12) for $V=\emptyset$ for all $\omega\in\Omega$ if the number A_0 is chosen sufficiently large.

To avoid some complications in the notation let us restrict our attention to the sets $V=\{1,\dots,u\}$, $1\le u<k$, and prove relation (5.12) for such sets. For this goal let us introduce the random variables

$$\bar I^V_{n,k}(f,j_{u+1},\dots,j_k)=\frac1{k!}\sum_{\substack{1\le j_s\le n,\ s=1,\dots,u\\ j_s\ne j_{s'}\text{ if } s\ne s',\ 1\le s,s'\le k}} f\left(\xi^{(1)}_{j_1,1},\dots,\xi^{(1)}_{j_u,u},\xi^{(-1)}_{j_{u+1},u+1},\dots,\xi^{(-1)}_{j_k,k}\right),$$

for all $f\in\mathcal F$, i.e. we fix some indices $j_{u+1},\dots,j_k$, $1\le j_s\le n$, $u+1\le s\le k$, $j_s\ne j_{s'}$ if $s\ne s'$, and sum up only those terms in the sum defining $\bar I^V_{n,k}(f)$ which contain $\xi^{(-1)}_{j_{u+1},u+1},\dots,\xi^{(-1)}_{j_k,k}$ in their last $k-u$ coordinates. Then we can write

$$\begin{aligned}E\left(\bar I^V_{n,k}(f)^2\big|\,\mathcal B\right)&=E\left(\left(\sum_{\substack{1\le j_s\le n\ s=u+1,\dots,k\\ j_s\ne j_{s'}\text{ if } s\ne s'}}\bar I^V_{n,k}(f,j_{u+1},\dots,j_{u_k})\right)^2\,\Bigg|\,\mathcal B\right)\\ &=\sum_{\substack{1\le j_s\le n\ s=u+1,\dots,k\\ j_s\ne j_{s'}\text{ if } s\ne s'}}E\left(\bar I^V_{n,k}(f,j_{u+1},\dots,j_{u_k})^2\big|\,\mathcal B\right).\end{aligned}\tag{5.13}$$

The last relation follows from the identity

$$E\left(\bar I^V_{n,k}(f,j_{u+1},\dots,j_{u_k})\bar I^V_{n,k}(f,j'_{u+1},\dots,j'_{u_k})\big|\,\mathcal B\right)=0$$

if $(j_{u+1},\dots,j_k)\ne(j'_{u+1},\dots,j'_k)$, which relation holds, since f is a canonical function.

It follows from relation (5.13) that

$$\begin{aligned}&\left\{\omega\colon\sup_{f\in\mathcal F}E\left(\bar I^V_{n,k}(f)^2\big|\,\mathcal B\right)(\omega)>2^{-(3k+3)}A^2n^{2k}\sigma^{2k+2}\right\}\\ &\qquad\subset\bigcup_{\substack{1\le j_s\le n\ s=u+1,\dots,k\\ j_s\ne j_{s'}\text{ if } s\ne s'}}\left\{\omega\colon\sup_{f\in\mathcal F}E\left(\bar I^V_{n,k}(f,j_{u+1},\dots,j_{u_k})^2\big|\,\mathcal B\right)(\omega)>\frac{A^2n^{2k}\sigma^{2k+2}}{2^{(3k+3)}n^{k-u}}\right\}.\end{aligned}\tag{5.14}$$

The probability of the events in the union at the right-hand side of (5.14) can be estimated with the help of the corollary of Proposition 5 with parameter $u<k$ instead of k. (We may assume that Proposition 5 holds for $u<k$.) This corollary yields that

$$P\left(\sup_{f\in\mathcal{F}} E\left(\bar I^V_{n,k}(f,j_{u+1},\dots,j_{u_k})^2\big|\,\mathcal{B}\right)>\frac{A^2\sigma^{2k+2}n^{k+u}}{2^{(3k+3)}}\right)\le e^{-A^{-1/(2u+1)}(n-u)\sigma^2}. \quad (5.15)$$

Indeed, the expression $E\left(\bar I^V_{n,k}(f,j_{u+1},\dots,j_{u_k})^2\Big|\,\mathcal{B}\right)$ can be calculated in the following way: Take the decoupled U-statistic

$$\bar I^V_{n,k}(f,x_{u+1},\dots,x_k)=\frac{1}{k!}\sum_{\substack{j_s\in\{1,\dots,n\}\setminus\{j_{u+1},\dots,j_k\},\\ s=1,\dots,u,\ j_s\ne j_{s'}\text{ if } s\ne s'}} f\left(\xi^{(1)}_{j_1,1},\dots,\xi^{(1)}_{j_u,u},x_{u+1},\dots,x_k\right) \quad (5.16)$$

of order u with sample size $n-k+u$, and integrate the square of this function with respect to the variables $x_{u+1},\dots,x_k$ by the measure μ^{k-u}. The expression at the left-hand side of (5.15) can be bounded by means of Proposition 5 if we apply it for our class of functions $\mathcal{F}$ considering them as functions on $(X^u\times Y,\mathcal{X}^u\times\mathcal{Y},\mu^u\times\rho)$ with $(Y,\mathcal{Y},\rho)=(X^{k-u},\mathcal{X}^{k-u},\mu^{k-u})$. (A small inaccuracy was committed in the above statement because to define the expression in (5.16) as a U-statistic we should have divided by $u!$ instead of $k!$. But this causes no real problem.)

We get inequality (5.15) from Proposition 5 by replacing the level $\frac{A^2\sigma^{2k+2}n^{k+u}}{2^{(3k+3)}}$ in the probability at the left-hand side by $A^2(n-u)^{2u}\sigma^{2u+2}<\frac{A^2\sigma^{2k+2}n^{k+u}}{2^{(2k+2)}}$. The last inequality really holds if the constant K is chosen sufficiently large in the condition $n\sigma^2>K\log n$ of Proposition 4.

Relations (5.14) and (5.15) imply that

$$P\left(\sup_{f\in\mathcal{F}} E\left(\bar I^V_{n,k}(f)^2\big|\,\mathcal{B}\right)(\omega)>2^{-(3k+3)}A^2n^{2k}\sigma^{2k+2}\right)\le n^{k-u}e^{-A^{-1/(2u+1)}(n-u)\sigma^2},$$

and $u\le k-1$. Hence also inequality (5.12) holds.

Now I prove Lemma 5A.

Proof of Lemma 5A. I show with the help of Lemma 3 and Lemma 7A that

$$P\left(\sup_{f\in\mathcal{F}} n^{-k/2}\left|\bar I_{n,k}(f)\right|>An^{k/2}\sigma^{k+1}\right)<2P\left(\sup_{f\in\mathcal{F}}|S(f)|>\frac{A}{2}n^k\sigma^{k+1}\right)+2^kn^{k-1}e^{-A^{1/(2k-1)}n\sigma^2/k} \quad (5.17)$$

with the function $S(f)$ defined in (5.10). To prove relation (5.17) introduce the random variables $Z(f)=(-1)^k\bar I^{\{1,\dots,k\}}_{n,k}(f)$ and $\bar Z(f)=\sum\limits_{V\subset\{1,\dots,k\},\,V\ne\{1,\dots,k\}}(-1)^{|V|+1}\bar I^V_{n,k}(f)$ for all $f\in\mathcal{F}$, the σ-algebra $\mathcal{B}$ considered in Lemma 7A and the set

$$B=\bigcap_{\substack{V\subset\{1,\dots,k\}\\ V\neq\{1,\dots,k\}}}\left\{\omega\colon \sup_{f\in\mathcal{F}} E\left(\bar I^V_{n,k}(f)^2\big|\,\mathcal{B}\right)(\omega)\le 2^{-(3k+3)}A^2n^{2k}\sigma^{2k+2}\right\}.$$

Observe that $S(f)=Z(f)-\bar Z(f)$, $f\in\mathcal{F}$, $B\in\mathcal{B}$, and by Lemma 7A the inequality $1-P(B)\le 2^kn^{k-1}e^{-A^{1/(2k-1)}n\sigma^2/k}$ holds. Hence to prove relation (5.17) as a consequence of Lemma 3 it is enough to show that

$$P\left(|\bar Z(f)|>\frac{A}{2}n^k\sigma^{k+1}\bigg|\,\mathcal{B}\right)(\omega)\le\frac12\quad\text{for all } f\in\mathcal{F}\quad\text{if } \omega\in\mathcal{B}. \tag{5.18}$$

But $P\left(|\bar I^V_{n,k}(f)|>2^{-(k+1)}An^k\sigma^{k+1}|\mathcal{F}\right)(\omega)\le 2^{-(k+1)}$ for all $f\in\mathcal{F}$, $V\subset\{1,\dots,k\}$, $V\neq\{1,\dots,k\}$ if $\omega\in B$ by the 'conditional Chebishev inequality', hence relation (5.18) holds.

Lemma 5A follows from relation (5.17), part a) of Lemma 6 and the observation that the random vectors $\{\bar I^{(V,\varepsilon)}_{n,k}(f),\ f\in\mathcal{F}\}$, defined in (5.9′) have the same distribution for all $V\subset\{1,\dots,k\}$ as the random vector $\{\bar I^{\varepsilon}_{n,k}(f),\ f\in\mathcal{F}\}$, considered in the formulation of Lemma 5A. Hence

$$P\left(\sup_{f\in\mathcal{F}}|S(f)|>\frac{A}{2}n^k\sigma^{k+1}\right)\le 2^kP\left(\sup_{f\in\mathcal{F}}\left|\bar I^{\varepsilon}_{n,k}(f)\right|>2^{-(k+1)}An^k\sigma^{k+1}\right).$$

In the proof of Lemma 5B I apply the following Lemma 7B which is a version of Lemma 7A.

Lemma 7B. *Let us consider a class of functions $\mathcal{F}$ satisfying the conditions of Proposition 5 and the random variables $\bar I^V_{n,k}(f,y)$, $f\in\mathcal{F}$, $V\subset\{1,\dots,k\}$, defined in formula (5.3). Let $\mathcal{B}=\mathcal{B}(\xi^{(1)}_{1,s},\dots,\xi^{(1)}_{n,s};\ 1\le s\le k)$ denote the σ-algebra generated by the random variables $\xi^{(1)}_{1,s},\dots,\xi^{(1)}_{n,s}$, $1\le s\le k$, taking part in the definition of the random variables $\bar I^V_{n,k}(f,y)$ and $H^V_{n,k}(f)$.*

a) *For all $V\subset\{1,\dots,k\}$, $V\neq\{1,\dots,k\}$, there exists a number $A_0=A_0(k)>0$ such that the inequality*

$$P\left(\sup_{f\in\mathcal{F}}E(H^V_{n,k}(f)|\mathcal{B})>2^{-(4k+4)}A^{(2k-1)/k}n^{2k}\sigma^{2k+2}\right)<n^{k-1}e^{-A^{1/2k}n\sigma^2/k} \tag{5.19}$$

holds for all $A\ge A_0$.

b) *Given two subsets $V_1,V_2\subset\{1,\dots,k\}$ of the set $\{1,\dots,k\}$ define the random integrals*

$$H^{(V_1,V_2)}_{n,k}(f)=\int|\bar I^{V_1}_{n,k}(f,y)\bar I^{V_2}_{n,k}(f,y)|\rho(\,dy),\quad f\in\mathcal{F},$$

with the help of the functions $\bar I^V_{n,k}(f,y)$ defined in (5.3). If at least one of the sets V_1 and V_2 is not the set $\{1,\dots,k\}$, then there exists some number $A_0=A_0(k)>0$ such that if the integrals $H_{n,k}(f)$, $f\in\mathcal F$, determined by this class of functions $\mathcal F$ have a good tail behaviour at level $T^{(2k+1)/2k}$ for some $T\ge A_0$, then the inequality

$$P\left(\sup_{f\in\mathcal F}E(H^{(V_1,V_2)}_{n,k}(f)|\mathcal B)>2^{-(2k+2)}A^2n^{2k}\sigma^{2k+2}\right)<2n^{k-1}e^{-A^{1/2k}n\sigma^2/k} \tag{5.20}$$

holds for all $A\ge T$.

Proof of Lemma 7B. The proof of part a) of Lemma 7B is similar to that of Lemma 7A, only the formulas applied in it become a little bit more complicated. Hence I omit it. (The difference between the power of the parameter A at the right-hand side of formulas (5.19) and (5.12) appear, since the left-hand side of (5.19) contains the term $A^{(2k-1)/2k}$ and not A^2.) Part b) will be proved with the help of Part a) and the inequality

$$\sup_{f\in\mathcal F}E(H^{(V_1,V_2)}_{n,k}(f)|\mathcal B)\le\left(\sup_{f\in\mathcal F}E(H^{V_1}_{n,k}(f)|\mathcal B)\right)^{1/2}\left(\sup_{f\in\mathcal F}E(H^{V_2}_{n,k}(f)|\mathcal B)\right)^{1/2}$$

which follows from the Schwarz inequality applied for integrals with respect to conditional distributions. Let us assume that $V_1\ne\{1,\dots,k\}$. Then the last inequality implies that

$$\begin{aligned}P&\left(\sup_{f\in\mathcal F}E(H^{(V_1,V_2)}_{n,k}(f)|\mathcal B)>2^{-(2k+2)}A^2n^{2k}\sigma^{2k+2}\right)\\&\le P\left(\sup_{f\in\mathcal F}E(H^{V_1}_{n,k}(f)|\mathcal B)>2^{-(4k+4)}A^{(2k-1)/k}n^{2k}\sigma^{2k+2}\right)\\&\qquad+P\left(\sup_{f\in\mathcal F}E(H^{V_2}_{n,k}(f)|\mathcal B)>A^{(2k+1)/k}n^{2k}\sigma^{2k+2}\right).\end{aligned}$$

Hence the estimate (5.19) for $V=V_1$ together with the inequality

$$P\left(\sup_{f\in\mathcal F}E(H^{V_2}_{n,k}(f)|\mathcal B)>A^{(2k+1)/k}n^{2k}\sigma^{2k+2}\right)\le n^{k-1}e^{-A^{1/2k}n\sigma^2/k}$$

which follows from Part a) if $V_2\ne\{1,\dots,n\}$ (in this case the level $A^{(2k+1)/k}n^{2k}\sigma^{2k+2}$ can be replaced by $2^{-(4k+4)}A^{(2k-1)/k}n^{2k}\sigma^{2k+2}$ in the probability we consider) and from the conditions of Part b) if $V_2=\{1,\dots,k\}$ imply relation (5.20).

Now I prove Lemma 5B.

Proof of Lemma 5B. By Part b) of Lemma 6 it is enough to prove that relation (5.6) holds if the random variables $\bar W(f)$ are replaced in it by the random variables $W(f)$ defined in formula (5.11). We shall prove this by applying Lemma 3 with the choice of $Z(f)=H_{n,k}^{(\bar V,\bar V)}(f)$, $\bar V=\{1,\dots,k\}$, $\bar Z(f)=W(f)-Z(f)$, $f\in\mathcal F$, $\mathcal B=\mathcal B(\xi_{1,s}^{(1)},\dots,\xi_{n,s}^{(1)};\ 1\le s\le k)$, and the set

$$B=\bigcap_{\substack{(V_1,V_2)\colon V_j\subset\{1,\dots,k\},\ j=1,2\\ V_1\ne\{1,\dots,k\}\text{ or }V_2\ne\{1,\dots,k\}}}\left\{\omega\colon \sup_{f\in\mathcal F} E(H_{n,k}^{(V_1,V_2)}(f)|\mathcal B)(\omega)\le 2^{-(2k+2)}A^2n^{2k}\sigma^{2k+2}\right\}.$$

By Lemma 7B $1-P(B))\le 2^{2k+1}n^{k-1}e^{-A^{1/2k}n\sigma^2/k}$, and to prove Lemma 5B with the help of Lemma 3 it is enough to show that

$$P\left(|\bar Z(f)|>\frac{A^2}{2}n^{2k}\sigma^{2(k+1)}\,\middle|\,\mathcal B\right)(\omega)\le\frac12\quad\text{for all } f\in\mathcal F \text{ if } \omega\in B.$$

To prove this relation observe that

$$E(|\bar Z(f)||\mathcal B)\le\sum_{\substack{(V_1,V_2)\colon V_j\subset\{1,\dots,k\},\ j=1,2\\ V_1\ne\{1,\dots,k\}\text{ or }V_2\ne\{1,\dots,k\}}} E(H_{n,k}^{(V_1,V_2)}(f)|\mathcal B)\le\frac{A^2}{4}n^{2k}\sigma^{2k+2}\quad\text{if } \omega\in B$$

for all $f\in\mathcal F$. Hence the 'conditional Markov inequality' implies that

$$P\left(|\bar Z(f)|>\frac{A^2}{2}n^{2k}\sigma^{2k+2}\,\middle|\,\mathcal B\right)\le\frac12\quad\text{if } \omega\in B\quad\text{and } f\in\mathcal F.$$

Lemma 5B is proved.

6. The proof of Propositions 4 and 5

The proof of Propositions 4 and 5 for general $k\ge1$ with the help of the symmetrization lemmas 5A and 5B is similar to the proof of Proposition 4 in the case $k=1$ presented in Section 4. The proof applies an induction procedure with respect to the parameter k. In the proof of Proposition 4 for parameter k we may assume that Propositions 4 and 5 hold for $k'<k$. In the proof of Proposition 5 we may also assume that Proposition 4 holds for the parameter k.

In the proof of Proposition 4 let us introduce (with the notation of this proposition) the functions

$$S^2_{n,k}(f)(x_{j,s},\,1\le j\le n,\,1\le s\le k)=\frac{1}{k!}\sum_{\substack{1\le j_s\le n,\ s=1,\dots,k\\ j_s\ne j_{s'} \text{ if } s\ne s'}} f^2\left(x_{j_1,1},\dots,x_{j_k,k}\right),\quad f\in\mathcal{F}, \tag{6.1}$$

where $x_{j,s}\in X$, $1\le j\le n$, $1\le s\le k$. Fix some number $A>T$, and define the set H

$$\begin{aligned}H=H(A)=\biggl\{&(x_{j,s},\,1\le j\le n,\,1\le s\le k),\\ &\sup_{f\in\mathcal{F}} S^2_{n,k}(f)(x_{j,s},\,1\le j\le n,\,1\le s\le k)>2^kA^{4/3}n^k\sigma^2\biggr\}.\end{aligned} \tag{6.2}$$

We want to show that

$$P(\{\omega\colon(\xi_{j,s}(\omega),\,1\le j\le n,\,1\le s\le k)\in H\})\le 2^ke^{-A^{2/3k}n\sigma^2}\quad\text{if } A\ge T. \tag{6.3}$$

Relation (6.3) will be proved by means of the Hoeffding decomposition of the U-statistics with kernel functions $f^2(x_1,\dots,x_k)$, $f\in\mathcal{F}$, and by the estimation of the sum this decomposition yields. More explicitly, write

$$f^2(x_1,\dots,x_k)=\sum_{V\subset\{1,\dots,k\}} f_V(x_s,s\in V) \tag{6.4}$$

with

$$f_V(x_s,s\in V)=\prod_{s\notin V}P_{\mu,s}\prod_{s\in V}Q_{\mu,s}f^2(x_1,\dots,x_k), \tag{6.5}$$

where $P_{\mu,s}$ and $Q_{\mu,s}$ are the operators P_μ and Q_μ defined in formulas (2.7) and (2.8) if $(Y_1\times Z\times Y_2,\mathcal{Y}_1\times\mathcal{Z}\times Y_2)$ is the k-fold product $(X^k,\mathcal{X}^k)$ of the measurable space $(X,\mathcal{X})$ in these definitions, Y_1 is the product of the first $s-1$ components, Z is the s-th component and Y_2 is the product of the last $k-s$ components in this product space. (Relation (6.4) follows from the identity $f^2=\prod\limits_{s=1}^{k}(P_{\mu,s}+Q_{\mu,s})f^2$ if the multiplications are carried out in this formula. In the calculation we exploit that the operators $P_{\mu,s}$ and $P_{\mu,s'}$ are commutative if $s\ne s'$, and the same relation holds for the pairs $P_{\mu,s}$ and $Q_{\mu,s'}$ or $Q_{\mu,s}$ and $P_{\mu,s'}$ or $Q_{\mu,s}$ and $Q_{\mu,s'}$.)

The identity $S^2_{n,k}(f)(\xi_{j,r}\, 1\le j\le n, 1\le r\le k)=k!I_{n,k}(f^2)$ holds for all $f\in\mathcal{F}$, and by writing the (Hoeffding type) decomposition (6.4) for each term $f^2(\xi_{j_1,1}\dots,\xi_{j_k,k})$ in the expression $I_{n,k}(f^2)$ we get that

$$\begin{aligned} &P\left(\sup_{f\in\mathcal{F}} S^2_{n,k}(f)(\xi_{j,s},\, 1\le j\le n,\, 1\le s\le k)>2^kA^{4/3}n^k\sigma^2\right)\\ &\qquad\le \sum_{V\subset\{1,\dots,k\}} P\left(\sup_{f\in\mathcal{F}} n^{k-|V|}|\bar I_{n,|V|}(f_V)|>A^{4/3}n^k\sigma^2\right) \end{aligned} \tag{6.6}$$

with the functions f_V defined in (6.5). We want to give a good estimate for all terms in the sum at the right-hand side in (6.6). For this goal we show that the classes of functions $\mathcal{F}_V=\{f_V\colon f\in\mathcal{F}\}$ satisfy the conditions of Proposition 4 for all $V\subset\{1,\dots,k\}$.

The functions f_V are canonical for all $V\subset\{1,\dots,k\}$. (This follows from the commutativity relations between the operators $P_{\mu,j}$ and $Q_{\mu,j}$ mentioned before, the identity $P_{\mu,j}Q_{\mu,j}=0$ and the fact that the canonical property of the function can be expressed in the form $P_{\mu,j}f_V=0$ for all $j\in V$.) We have $|f^2(x_1,\dots,x_k)|\le 2^{-2(k+1)}$. The norm of $Q_{\mu,j}$ as a map from the L_∞ space to L_∞ space is less than 2, the norm of $P_{\mu,j}$ is less than 1, hence $\left|\sup\limits_{x_j\in X, j\in V} f_V(x_j,j\in V)\right|\le 2^{-(k+2)}\le 2^{-(k+1)}$ for all $V\subset\{1,\dots,k\}$. We have $\int f^4(x_1,\dots,x_k)\mu(\,dx_1)\dots\mu(\,dx_k)\le 2^{-(k+1)}\sigma^2$, hence $\int f_V^2(x_j,j\in V)\prod\limits_{j\in V}\mu(\,dx_j)\le 2^{-(k+1)}\sigma^2\le\sigma^2$ for all $V\subset\{1,\dots,k\}$ by Lemma 1. Finally, to check that the class of functions $\mathcal{F}_V=\{f_V\colon f\in\mathcal{F}\}$ is L_2-dense with exponent L and parameter D observe that for all probability measures ρ on $(X^k,\mathcal{X}^k)$ and pairs of functions $f,g\in\mathcal{F}$ $\int(f^2-g^2)^2\,d\rho\le 2^{-2k}\int(f-g)^2\,d\rho$. This implies that if $\{f_1,\dots,f_m\}$, $m\le D\varepsilon^{-L}$, is an ε-dense subset of $\mathcal{F}$ in the space $L_2(X^k,\mathcal{X}^k,\rho)$, then the set of functions $\{2^kf_1^2,\dots,2^kf_m^2\}$ is an ε-dense subset of the class of functions $\mathcal{F}'=\{2^kf^2\colon f\in\mathcal{F}\}$ in the same space. Then by Lemma 1 and formula (6.5) the set of functions $\{(f_1)_V,\dots,(f_m)_V\}$ is an ε-dense subset of the class of functions $\mathcal{F}_V$ in the space $L_2(X^k,\mathcal{X}^k,\rho)$ for all $V\subset\{1,\dots,k\}$. This means that $\mathcal{F}_V$ is also L_2-dense with exponent L and parameter D.

For $V=\emptyset$ the relation $f_V=\int f^2(x_1,\dots,x_k)\mu(\,dx_1)\dots\mu(\,dx_k)\le\sigma^2$ holds, and $I_{|V|}(f_{|V|})|=f_V\le\sigma^2$. Therefore the term corresponding to $V=\emptyset$ in the sum at the right-hand side of (6.6) equals zero if $A_0\ge1$ in the conditions of Proposition 4. The terms corresponding to sets V, $1\le|V|\le k$, in these sums satisfy the inequality

$$\begin{aligned} &P\left(\sup_{f\in\mathcal{F}}|\bar I_{n,|V|}(f_V)|>A^{4/3}n^{|V|}\sigma^2\right)\\ &\qquad\le P\left(\sup_{f\in\mathcal{F}}|\bar I_{n,|V|}(f_V)|>A^{4/3}n^{|V|}\sigma^{|V|+1}\right)\le e^{-A^{2/3k}n\sigma^2}\quad\text{if } 1\le|V|\le k. \end{aligned}$$

This inequality follows from the inductive hypothesis if $|V|<k$, and in the case $V=\{1,\dots,k\}$ from the inequality $A\ge T$ and the assumption that U-statistics determined by a class of functions satisfying the conditions of Proposition 4 have a good

tail behaviour at level $T^{4/3}$. The last relation together with formula (6.6) imply relation (6.3).

By conditioning the probability $P\left(\left|\bar I^\varepsilon_{n,k}(f)\right|>2^{-(k+2)}An^{k/2}\sigma^{k+1}\right)$ with respect to the random variables $\xi_{j,s}$, $1\le j\le n$, $1\le s\le k$, we get with the help of Lemma 2 that

$$\begin{aligned}P&\left(\left|\bar I^\varepsilon_{n,k}(f)\right|>2^{-(k+2)}An^k\sigma^{k+1}\right|\left.\xi_{j,s}(\omega)=x_{j,s},1\le j\le n,1\le s\le k\right)\\&\le C\exp\left\{-B\left(\frac{A^2n^{2k}\sigma^{2(k+1)}}{2^{2k+4}S^2_{n,k}(x_{j,s},1\le j\le n,1\le s\le k)}\right)^{1/k}\right\}\\&\le Ce^{-2^{-3-4/k}BA^{2/3k}n\sigma^2}\quad\text{for all } f\in\mathcal F\quad\text{if } \{x_{j,s},\,1\le j\le n,\,1\le s\le k\}\notin H.\end{aligned}\tag{6.7}$$

Given some points $x_{j,s}$, $1\le j\le n$, $1\le s\le k$, define the probability measures ρ_s, $1\le s\le k$, uniformly distributed on the set $x_{j,s}$, $1\le j\le s$, i.e. $\rho_s(x_{j,s})=\frac1n$, $1\le j\le n$, and their product $\rho=\rho_1\times\cdots\times\rho_k$. If f is a function on $(X^k,\mathcal X^k)$ such that $\int f^2\,d\rho\le\delta^2$ with some $\delta>0$, then $|f(x_{j,s})|\le\delta n^{k/2}$ for all $1\le s\le k$, $1\le j\le n$, and $P\left(\left|\bar I^\varepsilon_{n,k}(f)\right|>\delta n^{3k/2}\right|\left.\xi_{j,s}=x_{j,s},1\le j\le n,\,1\le s\le k\right)=0$. Choose the numbers $\bar\delta=An^{-k/2}2^{-(k+2)}\sigma^{k+1}$ and $\delta=2^{-(k+2)}n^{-k-1/2}\le\bar\delta$. (The inequality $\delta\le\bar\delta$ holds, since $A\ge A_0\ge1$, and $\sigma\ge n^{-1/2}$.) Choose a δ-dense set $\{f_1,\dots,f_m\}$ in the $L_2(X^k,\mathcal X^k,\rho)$ space with $m\le D\delta^{-L}\le 2^{(k+2)L}n^{\beta+(k+1/2)L}$ elements. Then formula (6.7) implies that

$$\begin{aligned}P&\left(\sup_{f\in\mathcal F}\left|\bar I^\varepsilon_{n,k}(f)\right|>2^{-(k+1)}An^k\sigma^{k+1}\right|\left.\xi_{j,s}(\omega)=x_{j,s},1\le j\le n,1\le s\le k\right)\\&\le\sum_{j=1}^m P\left(\left|\bar I^\varepsilon_{n,k}(f_j)\right|>2^{-(k+2)}An^k\sigma^{k+1}\right|\left.\xi_{j,s}(\omega)=x_{j,s},1\le j\le n,1\le s\le k\right)\\&\le C2^{(k+2)L}n^{\beta+(k+1/2)L}e^{-2^{-3-4/k}BA^{2/3k}n\sigma^2}\quad\text{if } \{x_{j,s},\,1\le j\le n,\,1\le s\le k\}\notin H.\end{aligned}\tag{6.8}$$

Relations (6.3) and (6.8) imply that

$$\begin{aligned}P&\left(\sup_{f\in\mathcal F}\left|\bar I^\varepsilon_{n,k}(f)\right|>2^{-(k+1)}An^k\sigma^{k+1}\right)\\&\le C2^{(k+2)L}n^{\beta+(k+1/2)L}e^{-2^{-3-4/k}BA^{2/3k}n\sigma^2}+2^ke^{-A^{2/3k}n\sigma^2}\quad\text{if } A\ge T.\end{aligned}\tag{6.9}$$

Proposition 4 follows from the estimates (5.2) and (6.9) if the constants A_0 and K in the condition $n\sigma^2\ge K(L+\beta)\log n$ are chosen sufficiently large. In this case the upper bound these estimates yield for the probability at the left-hand side of (3.11) is smaller than $e^{-A^{2/k}n\sigma^2}$.

Let us turn to the proof of Proposition 5. By formula (5.8) it is enough to show that

$$\begin{aligned}&P\left(\sup_{f\in\mathcal F}|H_{n,k}(f|G,V_1,V_2)|>\frac{A^2}{2^{4k+1}k!}n^{2k}\sigma^{2(k+1)}\right)\le e^{-A^{1/2k}n\sigma^2}\\&\qquad\text{for all } G\in\mathcal G\quad\text{and}\quad V_1,V_2\in\{1,\dots,k\}\quad\text{if } A\ge A_0\end{aligned}\tag{6.10}$$

with the random variables $H_{n,k}(f|G,V_1,V_2)$ defined in formula (5.7). Let us first prove (6.10) in the case when $|e(G)|=k$, i.e. if all vertices of the diagram G are an end-point of some edge, and the expression $H_{n,k}(f|G,V_1,V_2)$ contains no 'symmetryzing term' ε_j. By the Schwarz inequality

$$|H_{n,k}(f|G,V_1,V_2)| \le \left(\sum_{\substack{j_1,\dots,j_k,1\le j_s\le n,\\ j_s\ne j_{s'} \text{ if } s\ne s'}} \int f^2(\xi_{j_1,1}^{(\delta_1)},\dots,\xi_{j_k,k}^{(\delta_k)},y)\rho(\,dy)\right)^{1/2} \left(\sum_{\substack{j_1,\dots,j_k,1\le j_s\le n,\\ j_s\ne j_{s'} \text{ if } s\ne s'}} \int f^2(\xi_{j_1,1}^{(\bar\delta_1)},\dots,\xi_{j_k,k}^{(\bar\delta_k)},y)\rho(\,dy)\right)^{1/2}, \tag{6.11}$$

for such diagrams G, where $\delta_s=1$ if $s\in V_1$, $\delta_s=-1$ if $s\notin V_1$, and $\bar\delta_s=1$ if $s\in V_2$, $\bar\delta_s=-1$ if $s\notin V_2$. Hence

$$\begin{aligned}
&\left\{\omega\colon \sup_{f\in\mathcal F}|H_{n,k}(f|G,V_1,V_2)(\omega)|>\frac{A^2}{2^{4k+1}k!}n^{2k}\sigma^{2(k+1)}\right\}\\
&\quad\subset\left\{\omega\colon \sup_{f\in\mathcal F}\sum_{\substack{j_1,\dots,j_k,1\le j_s\le n,\\ j_s\ne j_{s'} \text{ if } s\ne s'}}\int f^2(\xi_{j_1,1}^{(\delta_1)}(\omega),\dots,\xi_{j_k,k}^{(\delta_k)}(\omega),y)\rho(\,dy)>\frac{A^2n^{2k}\sigma^{2(k+1)}}{2^{4k+1}k!}\right\}\\
&\quad\cup\left\{\omega\colon \sup_{f\in\mathcal F}\sum_{\substack{j_1,\dots,j_k,1\le j_s\le n,\\ j_s\ne j_{s'} \text{ if } s\ne s'}}\int f^2(\xi_{j_1,1}^{(\bar\delta_1)}(\omega),\dots,\xi_{j_k,k}^{(\bar\delta_k)}(\omega),y)\rho(\,dy)>\frac{A^2n^{2k}\sigma^{2(k+1)}}{2^{4k+1}k!}\right\}.
\end{aligned}$$

The last relation implies that

$$\begin{aligned}
&P\left(\sup_{f\in\mathcal F}|H_{n,k}(f|G,V_1,V_2)|>\frac{A^2}{2^{4k+1}k!}n^{2k}\sigma^{2(k+1)}\right)\\
&\qquad\le 2P\left(\sup_{f\in\mathcal F}\sum_{\substack{j_1,\dots,j_k,1\le j_s\le n,\\ j_s\ne j_{s'} \text{ if } s\ne s'}}h_f(\xi_{j_1,1},\dots,\xi_{j_k,k})>\frac{A^2n^{2k}\sigma^{2(k+1)}}{2^{4k+1}k!}\right)
\end{aligned}\tag{6.12}$$

with the function $h_f(x_1,\dots,x_k)=\int f^2(x_1,\dots,x_k,y)\rho(\,dy)$, $f\in\mathcal F$. (In this upper bound we could get rid of the terms δ_j and $\bar\delta_j$, i.e. on the dependence of the expression $H_{n,k}(f|G,V_1,V_2)$ on the sets V_1 and V_2, since the probability of the events in the previous formula do not depend on these terms.)

I claim that

$$P\left(\sup_{f\in\mathcal F}|\bar I_{n,k}(h_f)|\ge An^k\sigma^2\right)\le 2^k e^{-A^{1/2k}n\sigma^2}\quad\text{for } A\ge A_0 \tag{6.13}$$

if the constant A_0 and K are chosen sufficiently large in Proposition 5. Relation (6.13) together with the relation $\frac{n^{2k}\sigma^{2(k+1)}}{2^{4k+1}k!}\ge n^k\sigma^2$ imply that the probability at the right-hand side of (6.12) can be bounded by $2^{k+1}e^{-A^{1/k}n\sigma^2}$, and the estimate (6.10) holds in the case $|e(G)|=k$. Relation (6.13) can be proved similarly to formula (6.3) in the proof of Proposition 4. It is not difficult to check that $0\le\int h_f(x_1,\dots,x_k)\mu(\,dx_1)\dots\mu(\,dx_k)\le\sigma^2$, $\sup|h_f(x_1,\dots,x_k)|\le 2^{-2(k+1)}$, and the class of functions $\mathcal H=\{2^kh_f,\ f\in\mathcal F\}$ is an L_2-dense class with exponent L and parameter D. This means that by applying the (Hoeffding type) decomposition of the functions h_f, $f\in\mathcal F$, similarly to formula (6.4) we get such sets of functions $(h_f)_V$, $f\in\mathcal F$ for all $V\subset\{1,\dots,k\}$ which satisfy the conditions of Proposition 4. Hence a natural adaptation of the estimate given for the expression at the right-hand side of (6.6) yields the proof of formula (6.13). Let us observe that by our inductive hypothesis the result of Proposition 4 holds also for k, and this allows us to carry out the estimates we need also for the class of functions $(h_f)_V$, $f\in\mathcal F$, with $V=\{1,\dots,k\}$ if $A\ge A_0$.

In the case $e(G)<k$ formula (6.10) will be proved with the help of Lemma 2. To carry out this proof first an appropriate expression will be introduced and bounded for all sets $V_1,V_2\subset\{1,\dots,k\}$ and diagrams $G\in\mathcal G$ such that $|e(G)|<k$. To define the expression we shall bound first some notations will be introduced.

Let us consider the set $J_0(G)=J_0(G,k,n)$,

$$\begin{aligned}J_0(G)=\{(j_1,\dots,j_k,j_1',\dots,j_k')\colon &1\le j_s,j_s'\le n,\ 1\le s\le k,\ j_s\ne j_{s'}\text{ if } s\ne s',\\ &j_s'\ne j_{s'}'\text{ if } s\ne s',\ j_s=j_{s'}'\text{ if } (s,s')\in G,\ j_s\ne j_{s'}'\text{ if } (s,s')\notin G\},\end{aligned}$$

the set of those sequences $(j_1,\dots,j_k,j_1',\dots,j_k')$ which appear as indices in the summation in formula (5.7). I introduce a partition of $J_0(G)$ appropriate for our purposes.

To do this first I define the sets $M_1=M_1(G)=\{s(1),\dots,s(k-|e(G)|)\}=\{1,\dots,k\}\setminus v_1(G)$, $s(1)<\cdots<s(k-|e(G)|)$, and $M_2=M_2(G)=\{\bar s(1),\dots,\bar s(k-|e(G|)\}=\{1,\dots,k\}\setminus v_2(G)$, $\bar s(1)<\cdots<\bar s(k-|e(G|)$, the sets of those vertices of the first and second row of the diagram G in increasing order from which no edges start. Let us also introduce the set $V(G)=V(G,n,k)$,

$$\begin{aligned}V(G)=\{(j_{s(1)},\dots,j_{s(k-|e(G)|)},j'_{\bar s(1)},\dots,j'_{\bar s(k-|e(G)|)})\colon 1\le j_{s(p)},j'_{\bar s(p)}\le n,\\ 1\le p\le k-|e(G)|,\ j_{s(p)}\ne j_{s(p')},\ j'_{\bar s(p)}\ne j'_{\bar s(p')}\text{ if } p\ne p',\ 1\le p,p'\le k-|e(G)|\\ j_{s(p)}\ne j'_{\bar s(p')},\ 1\le p,p'\le k-|e(G)|\},\end{aligned}$$

which is the set consisting from the restriction of the coordinates of the vectors

$$(j_1,\dots,,j_k,j_1',\dots,j_k')\in J_0(G)$$

to $M_1\cup M_2$. Given a vector $v\in V(G)$ let $v(r)$, $1\le r\le k-|e(G)|$, and $\bar v(r)$, $1\le r\le k-|e(G)|$, denote its coordinates corresponding to the set M_1 and M_2 respectively. Put

$$\begin{aligned}E_G(v)=\{(j_1,\dots,j_k,j'_1,\dots,j'_k)\colon & 1\le j_s\le n,\ 1\le j'_{s'}\le n,\quad 1\le s,s'\le k,\\ & j_s\ne j_{s'} \text{ if } s\ne s',\ j'_{\bar s}\ne j'_{\bar s'} \text{ if } \bar s\ne \bar s',\\ & j_s=j'_{s'} \text{ if } (s,s')\in e(G) \text{ and } j_s\ne j'_{s'} \text{ if } (s,s')\notin e(G)\\ & j_{s(r)}=v(r),\ j'_{\bar s(r)}=\bar v(r),\ 1\le r\le k-|e(G)|\},\quad v\in V(G),\end{aligned}$$

where $\{s(1),\dots,s(k-|e(G)|)\}=M_1$, $\{\bar s(1),\dots,\bar s(k-|e(G)|)\}=M_2$ in the last line of this definition. The set $E_G(v)$ contains those vectors in $J_0(G)$ whose coordinates in $M_1\cup M_2$ are prescribed by the vector $v\in V(G)$, the remaining coordinates can be put into pairs (s,s') such that $(s,s')\in e(G)$, and the values $j_s=j'_{s'}$, $(s,s')\in e(G)$, must be different for different pairs, otherwise they can be chosen freely from the set $\{1,\dots,n\}\setminus\{v(1),\dots,v(k-|e(G)|),\bar v(1),\dots,\bar v(k-|e(G)|)\}$.

Now we define the partition

$$J_0(G)=\bigcup_{v\in V(G)}E_G(v)$$

of the set $J_0(G)$.

The inequality

$$P\left(S(\mathcal F|G,V_1,V_2))>A^{8/3}n^{2k}\sigma^4\right)\le 2^{k+1}e^{-A^{2/3k}n\sigma^2}\quad \text{if } A\ge A_0 \text{ and } e(G)<k \tag{6.14}$$

will be proved for the random variable

$$\begin{aligned}S(\mathcal F|G,V_1,V_2)=\sup_{f\in\mathcal F}\sum_{v\in V(G)}\Biggl(\sum_{(j_1,\dots,j_k,j'_1,\dots,j'_k)\in E_G(v)}&\int f(\xi^{(\delta_1)}_{j_1,1},\dots,\xi^{(\delta_k)}_{j_k,k},y)\\ &f(\xi^{(\bar\delta_1)}_{j'_1,1},\dots,\xi^{(\bar\delta_k)}_{j'_k,k},y)\rho(\,dy)\Biggr)^2,\end{aligned}$$

where $\delta_s=1$ if $s\in V_1$, $\delta_s=-1$ if $s\notin V_1$, and $\bar\delta_s=1$ if $s\in V_2$, $\bar\delta_s=-1$ if $s\notin V_2$.

To prove formula (6.14) let us first fix some $v\in V(G)$ and apply the Schwarz inequality. It yields that

$$\begin{aligned}&\left(\sum_{(j_1,\dots,j_k,j'_1,\dots,j'_k)\in E_G(v)}\int f(\xi^{(\delta_1)}_{j_1,1},\dots,\xi^{(\delta_k)}_{j_k,k},y)f(\xi^{(\bar\delta_1)}_{j'_1,1},\dots,\xi^{(\bar\delta_k)}_{j'_k,k},y)\rho(\,dy)\right)^2\\ &\quad\le\left(\sum_{(j_1,\dots,j_k,j'_1,\dots,j'_k)\in E_G(v)}\int f^2(\xi^{(\delta_1)}_{j_1,1},\dots,\xi^{(\delta_k)}_{j_k,k},y)\rho(\,dy)\right)\\ &\qquad\left(\sum_{(j_1,\dots,j_k,j'_1,\dots,j'_k)\in E_G(v)}\int f^2(\xi^{(\bar\delta_1)}_{j'_1,1},\dots,\xi^{(\bar\delta_k)}_{j'_k,k},y)\rho(\,dy)\right).\end{aligned}$$

for all $v\in V(G)$. Summing up these inequalities for all $v\in V(G)$ we get that

$$
\begin{aligned}
S(\mathcal{F}|G,V_1,V_2)\le \sup_{f\in\mathcal{F}}\sum_{v\in V(G)}&\left(\sum_{(j_1,\dots,j_k,j'_1,\dots,j'_k)\in E_G(v)}\int f^2(\xi^{(\delta_1)}_{j_1,1},\dots,\xi^{(\delta_k)}_{j_k,k},y)\rho(\,dy)\right)\\
&\left(\sum_{(j_1,\dots,j_k,j'_1,\dots,j'_k)\in E_G(v)}\int f^2(\xi^{(\bar\delta_1)}_{j'_1,1},\dots,\xi^{(\bar\delta_k)}_{j'_k,k},y)\rho(\,dy)\right)\\
\le \sup_{f\in\mathcal{F}}&\left(\sum_{(j_1,\dots,j_k,j'_1,\dots,j'_k)\in J_0(G)}\int f^2(\xi^{(\delta_1)}_{j_1,1},\dots,\xi^{(\delta_k)}_{j_k,k},y)\rho(\,dy)\right)\\
&\sup_{f\in\mathcal{F}}\left(\sum_{(j_1,\dots,j_k,j'_1,\dots,j'_k)\in J_0(G)}\int f^2(\xi^{(\bar\delta_1)}_{j'_1,1},\dots,\xi^{(\bar\delta_k)}_{j'_k,k},y)\rho(\,dy)\right)
\end{aligned}
\tag{6.15}
$$

To check formula (6.15) we have to observe that by multiplying the inner sum at the left-hand side of this inequality each term $f^2(\xi^{(\delta_1)}_{j_1,1},\dots,\xi^{(\delta_k)}_{j_k,k},y)f^2(\xi^{(\bar\delta_1)}_{j'_1,1},\dots,\xi^{(\bar\delta_k)}_{j'_k,k},y)$ appears only once. (In particular, it is determined which index $v\in V(G)$ has to be taken in the outer sum to get this term. The coordinates of this vector v agree with the coordinates of the vector $j=(j_1,\dots,j_k,j'_1,\dots,j'_k)$ in $M_1\cup M_2$, with the coordinates of the vector j which correspond to those vertices from which no edges of the diagram G start.) Beside this, all these products appear if the multiplications at the right-hand side expression are carried out.

Relation (6.15) implies that

$$
P(S(\mathcal{F}|G,V_1,V_2)>A^{8/3}n^{2k}\sigma^4)\le 2P\left(\sup_{f\in\mathcal{F}}\bar I_{n,k}(h_f)>A^{4/3}n^k\sigma^2\right)
$$

with $h_f(x_1,\dots,x_k)=\int f^2(x_1,\dots,x_k,y)\rho(\,dy)$. (Here we exploited that in the last formula $S(\mathcal{F}|G,V_1,V_2)$ is bounded by the product of two random variables whose distributions do not depend on the sets V_1 and V_2.) Thus to prove inequality (6.14) it is enough to show that

$$
2P\left(\sup_{f\in\mathcal{F}}\bar I_{n,k}(h_f)>A^{4/3}n^k\sigma^2\right)\le 2^{k+1}e^{-A^{2/3k}}\quad\text{if } A\ge A_0. \tag{6.16}
$$

Actually formula (6.16) has been already proved, only formula (6.13) has to be applied, and the parameter A has to be replaced by $A^{4/3}$ in it.

The proof of Proposition 5 can be completed similarly to Proposition 4. It follows from Lemma 2 that

$$P\left(|H_{n,k}(f|G,V_1,V_2)| > \frac{A^2}{2^{4k+2}k!}n^{2k}\sigma^{2(k+1)}\right| \left.\xi_{j,s}^{\pm 1},\ 1\le j\le n,\ 1\le s\le k\right)(\omega)$$
$$\le Ce^{-B2^{-(4+2/k)}(k!)^{-1/k}A^{2/3k}n\sigma^2} \quad \text{if} \quad S(\mathcal{F}|G,V_1,V_2)(\omega)\le A^{8/3}n^{2k}\sigma^4$$
$$\text{for all } f\in\mathcal{F},\ G\in\mathcal{G},\ |e(G)|<k, \quad \text{and} \quad V_1,V_2\in\{1,\dots,k\} \quad \text{if } A\ge A_0. \tag{6.17}$$

Indeed, in this case the conditional probability considered in (6.17) can be bounded by $C\exp\left\{-B\left(\frac{A^4n^{4k}\sigma^{4(k+1)}}{2^{8k+4}(k!)^2S(\mathcal{F}|G,V_1,V_2)}\right)^{1/2j}\right\}\le C\exp\left\{-B\left(\frac{A^{4/3}n^{2k}\sigma^{4k}}{2^{8k+4}(k!)^2}\right)^{1/2j}\right\}$, where $2j=2k-2|e(G)|$, the number of vertices of the diagram G from which no edges start. Since $j\le k$, $n\sigma^2\ge 1$, and also $\frac{A^{4/3}}{2^{8k+4}(k!)^2}\ge 1$ if A_0 is chosen sufficiently large the above calculation implies relation (6.17).

Let us show that also the inequality

$$P\left(\sup_{f\in\mathcal{F}}|H_{n,k}(f|G,V_1,V_2)| > \frac{A^2}{2^{4k+1}k!}n^{2k}\sigma^{2(k+1)}\right| \left.\xi_{j,s}^{\pm 1},\ 1\le j\le n,\ 1\le s\le k\right)(\omega)$$
$$\le Cn^{(3k+1)L/2+\beta}e^{-BA^{2/3k}n\sigma^2/2^{(4+2/k)}(k!)^{1/k}} \quad \text{if } S(\mathcal{F}|G,V_1,V_2))(\omega)\le A^{8/3}n^{2k}\sigma^4$$
$$\text{for all } G\in\mathcal{G},\ |e(G)|<1, \quad \text{and} \quad V_1,V_2\in\{1,\dots,k\} \quad \text{if } A\ge A_0 \tag{6.18}$$

holds.

To prove formula (6.18) let us fix an elementary event $\omega\in\Omega$ which satisfies the relation $S(\mathcal{F}|G,V_1,V_2))(\omega)\le A^{8/3}n^{2k}\sigma^4$, two sets $V_1,V_2\subset\{1,\dots,k\}$, a diagram G, consider the points $x_{j,s}^{(\pm1)}=x_{j,s}^{(\pm1)}(\omega)=\xi_{j,s}^{(\pm1)}(\omega)$, $1\le j\le n$, $1\le s\le k$, and introduce with their help the following probability measures: For all $1\le s\le k$ define the probability measures $\nu_s^{(1)}$ which are uniformly distributed on the points $x_{j,s}^{(\delta_s)}$, $1\le j\le n$, and $\nu_s^{(2)}$ which are uniformly distributed on the points $x_{j,s}^{(\bar\delta_s)}$, $1\le j\le n$, where $\delta_s=1$ if $s\in V_1$, $\delta_s=-1$ if $s\notin V_1$, and similarly $\bar\delta_s=1$ if $s\in V_2$ and $\bar\delta_s=-1$ if $s\notin V_2$. Let us consider the product measures $\alpha_1=\nu_1^{(1)}\times\cdots\times\nu_k^{(1)}\times\rho$, $\alpha_2=\nu_1^{(2)}\times\cdots\times\nu_k^{(2)}\times\rho$ on the product space $(X^k\times Y,\mathcal{X}^k\times\mathcal{Y})$, where ρ is that probability measure on $(Y,\mathcal{Y})$ which appears in Proposition 5, together with the measure $\alpha=\frac{\alpha_1+\alpha_2}{2}$. Given two functions $f\in\mathcal{F}$ and $g\in\mathcal{F}$ we give an upper bound for $|H_{n,k}(f|G,V_1,V_2)(\omega)-H_{n,k}(g|G,V_1,V_2)(\omega)|$ if $\int(f-g)^2\,d\alpha\le\delta$ with some $\delta>0$. (This bound does not depend on the 'randomizing terms' $\varepsilon_j(\omega)$ in the definition of the random variable $H_{n,k}(\cdot|G,V_1,V_2)$.)

In this case $\int(f-g)^2\,d\alpha_j\le 2\delta^2$, and

$$\int\left|f(x_{1,j_1}^{(\delta_1)},\dots,x_{k,j_k}^{(\delta_k)},y)-g(x_{1,j_1}^{(\delta_1)},\dots,x_{k,j_k}^{(\delta_k)},y)\right|^2\rho(\,dy)\le 2\delta^2n^k,$$
$$\int\left|f(x_{1,j_1}^{(\delta_1)},\dots,x_{k,j_k}^{(\delta_k)},y)-g(x_{1,j_1}^{(\delta_1)},\dots,x_{k,j_k}^{(\delta_k)},y)\right|\rho(\,dy)\le\sqrt{2}\delta n^{k/2}$$

for all $1 \le s \le k$, and $1 \le j_s \le n$, and the same result holds if all δ_s is replaced by $\bar\delta_s$, $1 \le s \le k$. Since $|f| \le 1$ for $f \in \mathcal{F}$, the condition $\int (f-g)^2\, d\alpha \le \delta^2$ implies that

$$\int \left| f(\xi^{(\delta_1)}_{j_1,1}(\omega),\dots,\xi^{(\delta_k)}_{j_k,k}(\omega),y) f(\xi^{(\bar\delta_1)}_{j'_1,1}(\omega),\dots,\xi^{(\bar\delta_k)}_{j'_k,k}(\omega),y)) \right.$$
$$\left. - g(\xi^{(\delta_1)}_{j_1,1}(\omega),\dots,\xi^{(\delta_k)}_{j_k,k}(\omega),y) g(\xi^{(\bar\delta_1)}_{j'_1,1}(\omega),\dots,\xi^{(\bar\delta_k)}_{j'_k,k}(\omega),y) \right| \rho(\,dy) \le 2\sqrt{2}\delta n^{k/2}$$

for all vectors $(j_1,\dots,j_k,j'_1,\dots,j'_k)$ which appear as an index in the summation in (5.7), and

$$|H_{n,k}(f|G,V_1,V_2)(\omega) - H_{n,k}(g|G,V_1,V_2)(\omega)| \le 2\sqrt{2}\delta n^{5k/2}$$

if the originally fixed $\omega \in \Omega$ is considered.

Put $\bar\delta = \frac{A^2 n^{-k/2} \sigma^{2(k+1)}}{2^{(4k+7/2)} k!}$, and $\delta = n^{-(3k+1)/2} \le \bar\delta$ (since $\sigma \ge \frac{1}{\sqrt{n}}$ and we may assume that $A \ge A_0$ is sufficiently large), choose a δ-dense subset $\{f_1,\dots,f_m\}$ in the $L_2(X^k \times Y, \mathcal{X}^k \times Y, \alpha)$ space with $m \le D\delta^{-L} \le n^{(3k+1)L/2+\beta}$ elements. Relation (6.17) for these functions together with the above estimates yield formula (6.18).

It follows from relations (6.14) and (6.18) that

$$P\left(\sup_{f\in\mathcal{F}} |H_{n,k}(f|G,V_1,V_2)| > \frac{A^2}{2^{4k+1}k!} n^{2k} \sigma^{2(k+1)} \right) \le 2^{k+1} e^{-A^{2/3k} n\sigma^2}$$
$$+ C n^{(3k+1)L/2+\beta} e^{-BA^{2/3k} n\sigma^2 / 2^{(4+2/k)} (k!)^{1/k}} \quad \text{if } A \ge A_0$$

for all $V_1, V_2 \subset \{1,\dots,k\}$ also in the case $|e(G)| \le k-1$. This means that relation (6.10) holds also in this case if the constants A_0 and K are chosen sufficiently large in Proposition 5. Proposition 5 is proved.

Appendix. The proof of Proposition 3

I shall explain the proof of Proposition 3 in a concise form. A more detailed explanation can be found in [8].

The proof of Proposition 3. Let us first introduce the (random) probability measures δ_{ξ_j}, $1 \le j \le n$, concentrated in the sample points ξ_j. We can write $\mu_n - \mu = \frac{1}{n}\left(\sum\limits_{j=1}^{n} \left(\delta_{\xi_j} - \mu\right)\right)$, and formula (1.2) can be rewritten as

$$J_{n,k}(f) = \frac{1}{n^{k/2}k!} \sum_{\substack{(j_1,\dots,j_k) \\ 1\le j_s \le n \text{ for all } 1\le s\le k}} \int' f(x_1,\dots,x_k) \left(\delta_{\xi_{j_1}}(\,dx_1) - \mu(\,dx_1)\right) \dots \left(\delta_{\xi_{j_k}}(\,dx_k) - \mu(\,dx_k)\right). \tag{A1}$$

To rearrange the above sum in a way more appropriate for us let us introduce the following notations: Let $\mathcal{P} = \mathcal{P}_k$ denote the set of all partitions of the set $\{1, 2, \dots, k\}$, and given a sequence $(j_1, \dots, j_k)$, $1 \le j_s \le n$, $1 \le s \le k$, of length k let $H(j_1, \dots, j_k)$ denote that partition of $\mathcal{P}_k$ in which two points s and t, $1 \le s, t \le k$, belong the same element of the partition if and only if $j_s = j_t$. Given a set A, let $|A|$ denote its cardinality.

Let us rewrite the expression (A1) for $J_{n,k}(f)$ in the form

$$J_{n,k}(f) = \frac{1}{n^{k/2}k!} \sum_{P\in\mathcal{P}} \sum_{\substack{(j_1,\dots,j_k), \\ 1\le j_s\le n,\, 1\le s\le k \\ H(j_1,\dots,j_k)=P}} \int' f(x_1,\dots,x_k) \left(\delta_{\xi_{j_1}}(\,dx_1) - \mu(\,dx_1)\right) \dots \left(\delta_{\xi_{j_k}}(\,dx_k) - \mu(\,dx_k)\right). \tag{A2}$$

Let us remember that the diagonals $x_s = x_t$, $s \ne t$, were omitted from the domain of integration in the formula defining $J_{n,k}(f)$. This implies that in the case $j_s = j_t$ the measure $\delta_{\xi_{j_s}}(\,dx_s)\delta_{\xi_{j_t}}(\,dx_t)$ has zero measure in the domain of integration. We have to understand the cancellation effects caused by this relation. I want to show that because of these cancellations the expression in formula (A2) can be rewritten as a linear combination of the degenerate U-statistics $I_{n,|V|}(f_V)$ defined in (2.11) with not too large coefficients. This seems to be a natural approach, but the detailed proof demands some rather unpleasant calculations.

Let us fix some $P \in \mathcal{P}$ and investigate the inner sum at the right-hand side of (A2) corresponding to this partition P. For the sake of simplicity let us first consider such an inner sum that corresponds to a partition $P \in \mathcal{P}$ which contains a set of the form $\{1, \dots, s\}$ with some $s \ge 2$. The products of measures corresponding to the terms in the sum determined by such a partition contain a part of length s which has the form $\big(\delta_{\xi_j}(dx_1)-\mu(dx_1)\big)\dots\big(\delta_{\xi_j}(dx_s)-\mu(dx_s)\big)$ with some $1\le j\le n$. This part of the product can be rewritten in the domain of integration as

$$\sum_{l=1}^{s}(-1)^{s-1}\mu(\,dx_1)\dots\mu(\,dx_{l-1})(\delta_{\xi_j}(\,dx_l)-\mu(\,dx_l))\mu(\,dx_{l+1})\dots\mu(\,dx_s)$$
$$+(-1)^{s-1}(s-1)\mu(dx_1)\dots\mu(dx_s).$$

Here we have exploited that all other terms of this product disappear in the domain of integration. Let us also observe that the term $(-1)^{s-1}(s-1)\mu(\,dx_1)\dots\mu(\,dx_l)$ appears n-times as we sum up for $1\le j\le n$. Similar calculation can be made for all partitions $P\in\mathcal{P}$ and all sets contained in the partitions, only the notation of the indices will be more complicated.

Let us fix a general partition $P=\{R_1,\dots,R_u\}\in\mathcal{P}$, and let us rewrite the inner sum in formula (A2) in a more appropriate form. We can get the proof of Proposition 3 by means of summing up the identities we get in such a way for all $P\in\mathcal{P}$. To get the desired formula fix some vector $(j_1,\dots,j_k)$ such that $H(j_1,\dots,j_k)=P$, and let us rewrite the multiple integral in the inner sum of (A2) corresponding to this index.

We can get by working out the above mentioned calculation in the general case that for a vector $(j_1,\dots,j_k)$ such that $H(j_1,\dots,j_k)=P$ the relation

$$\int' f(x_1,\dots,x_k)\left(\delta_{\xi_{j_1}}(\,dx_1)-\mu(\,dx_1)\right)\dots\left(\delta_{\xi_{j_k}}(\,dx_k)-\mu(\,dx_k)\right)$$
$$=\sum_{V\in\mathcal{T}(P)}\alpha(V,P)\int f(x_1,\dots,x_k)\prod_{s\in V}\left(\delta_{\xi_{j_s}}(\,dx_s)-\mu(\,dx_s)\right)\prod_{s'\in\{1,\dots,k\}\setminus V}\mu(\,dx_{s'}) \tag{A3}$$

holds with some appropriate constants $\alpha(V,P)$, where the class $\mathcal{T}(P)$ which consists of subsets of $\{1,\dots,k\}$ and depends on the partition $P=\{R_1,\dots,R_u\}$ is defined in the following way. For all elements R_t, $1\le t\le u$, of the partition P a set $V\in\mathcal{T}(P)$ contains zero or 1 elements of the set R_t. If $R_t=\{b_t\}$ consists of one elements, then the set V contains this point b_t. $\mathcal{T}(P)$ consists of all subsets of $\{1,\dots,k\}$ which satisfy the two above properties.

The coefficients $\alpha(V,P)$ at the right-hand side of (A3) could be calculated explicitly, but we do not have to do this. It is enough to know that it depends only on the partition P and the set $V\in\mathcal{T}(P)$. Let us also observe that at the right-hand side of (A3) the prime is missing in the integral, i.e. here integration is taken on the whole space X^k, the diagonals are not taken out from the domain of integration. Indeed, it can be seen that because of the non-atomic property of the measure μ we do not change the value of the integrals at the right-hand side of (A3) by inserting the diagonals to the domain of integration.

Formula (A3) can be rewritten in the following way.

$$\int' f(x_1,\dots,x_k)\left(\delta_{\xi_{j_1}}(\,dx_1)-\mu(\,dx_1)\right)\dots\left(\delta_{\xi_{j_k}}(\,dx_k)-\mu(\,dx_k)\right)$$

$$= \sum_{V \in \mathcal{T}(P)} \alpha(V,P) \left(\left(\prod_{s' \in \{1,\dots,k\} \setminus V} P_{\mu,s'} \prod_{s \in V} Q_{\mu,s} \right) \right) f(\xi_{j_s}, s \in V). \tag{A4}$$

Here $Q_{\mu,s} = I - P_{\mu,s}$ is the operator Q_μ defined in (2.8) if Y_1 is the product of the first $s-1$ components of the product space X^k, Z is its s-th component and Y_2 is the product of the last $k-s$ components. The operator $P_{\mu,s'}$ is the operator P_μ defined in (2.7) with the choice of Y_1 as the first $s'-1$ components, Z as the s'-th component and Y_2 as the product of the last $k-s'$ components of the space X^k. To see why formula (A4) holds we have to understand that integration with respect to $\left(\delta_{\xi_{j_s}}(\,dx_s) - \mu(\,dx_s)\right)$ means the application of the operator $Q_{\mu,s}$ and then putting the value ξ_{j_s} in the argument x_s, while integration with respect to $\mu(\,dx_{s'})$ means the application of the operator $P_{\mu,s'}$. Beside this, the operators $Q_{\mu,s}$ and $P_{\mu,s'}$ are exchangeable.

By fixing some $V \in \mathcal{T}(P)$ and summing up the term corresponding to it at the right-hand side of formula (A4) for all $(j_1,\dots,j_k)$ such that $H(j_1,\dots,j_k) = P$ we get that

$$\alpha(V,P) \sum_{\substack{(j_1,\dots,j_k) \\ 1 \le j_s \le n,\ 1 \le s \le k \\ H(j_1,\dots,j_k)=P}} \left(\prod_{s' \in \{1,\dots,k\} \setminus V} P_{\mu,s'} \prod_{s \in V} Q_{\mu,s} \right) f(\xi_{j_s},\ s \in V) = \bar{\alpha}(V,P,k,n) I_{n,|V|}(f_V) \tag{A5}$$

where $I_{n,|V|}(f_V)$ is the U-statistic of order $|V|$ with the kernel function

$$f_V(x_s, s \in V) = \left(\prod_{s' \in \{1,\dots,k\} \setminus V} P_{\mu,s'} \prod_{s \in V} Q_{\mu,s} \right) f(x_1,\dots,x_k), \tag{A6}$$

and the coefficients $\bar{\alpha}(V,P,k,n)$ at the right-hand side of (A5) are appropriate coefficients which could be calculated explicitly. But we do not need such a formula. It can be shown with some work that they satisfy the inequality $|\bar{\alpha}(V,P,k,n)| \le D(k) n^{\beta(P,V)}$, where $\beta(P,V) = u - |V|$ is the number of those components R_s, $1 \le s \le u$, of the partition P for which $R_s \cap V = \emptyset$, and the constant $D(k) < \infty$ depends only on the multiplicity k of the integral $J_{n,k}(f)$. Such an estimate is sufficient for us.

We get from relations (A2), (A4) and (A5) by summing up identity (A5) for all $P \in \mathcal{P}$ and $V \in \mathcal{T}(P)$ that

$$J_{n,k}(f) = \sum_{V \subset \{1,2,\dots,k\}} C(n,k,V) n^{-|V|/2} I_{n,|V|}(f_V) \tag{A7}$$

with some coefficients $C(n,k,V)$. Moreover, a careful analysis shows that the above coefficients satisfy the inequality $|C(n,k,V)| \le G(k)$ with some constant $G(k) > 0$. The explicit expression for the coefficients $C(n,k,V)$ has a rather complicated form, but the above estimate about their magnitude is sufficient for our purposes. This estimate for $C(k,n,V)$ is sharp, because for a fixed set V those partitions $P \in \mathcal{P}$ which contain the

$|V|$ one-point subsets of V and $(k-|V|)/2$ subsets of cardinality 2 of $\{1,\dots,k\}\setminus V$ yield a contribution of order $n^{-k/2}n^{k/2-|V|/2}$ to the coefficient $C(n,k,V)n^{-|V|/2}$. A more careful analysis also shows that for a fixed set $V\subset\{1,\dots,k\}$ the sequence $C(n,k,V)$ has a finite limit as $n\to\infty$. It is not difficult to see that $C(n,k,\{1,\dots,k\})=1$ for $V=\{1,\dots,k\}$.

The definition of the function f_V in formula (A6) agrees with the definition of f_V in formula (2.11). Hence formulas (A6), (A7) and the estimates on the coefficients $C(n,k,V)$ in formula (A7) imply Proposition 3.

References:

1.) Alexander, K. (1987) The central limit theorem for empirical processes over Vapnik–Červonenkis classes. *Ann. Probab.* **15**, 178–203

2.) Arcones, M. A. and Giné, E. (1993) Limit theorems for U-processes. *Ann. Probab.* **21**, 1494–1542

3.) Arcones, M. A. and Giné, E. (1994) U-processes indexed by Vapnik–Červonenkis classes of functions with application to asymptotics and bootstrap of U-statistics with estimated parameters. *Stoch. Proc. Appl.* **52**, 17–38

4.) Beckner, W. (1975) Inequalities in Fourier Analysis. Ann. Math. **102**, 159–182

5.) de la Peña, V. H. and Giné, E. (1999) Decoupling From dependence to independence. Springer series in statistics. Probability and its application. Springer Verlag, New York, Berlin, Heidelberg

6.) de la Peña, V. H. and Montgomery–Smith, S. (1995) Decoupling inequalities for the tail-probabilities of multivariate U-statistics. *Ann. Probab.*, **25**, 806–816

7.) Giné, E. and Zinn, J. (1984) Some limit theorems for empirical processes. *Ann. Probab.*, **12**, 929–989

8.) Major, P. (2004) On the tail behaviour of multiple random integrals and degenerate U-statistics. (manuscript for a future Lecture Note)

9.) Major, P. (2005) A multivariate generalization of Hoeffding's inequality. Submitted to *Ann. Probab.*

10.) Major, P. and Rejtő, L. (1988) Strong embedding of the distribution function under random censorship. *Annals of Statistics*, **16**, 1113–1132

11.) Major, P. and Rejtő, L. (1998) A note on nonparametric estimations. In the conference volume to the 65. birthday of Miklós Csörgő. 759–774

12.) Pollard, D. (1984) Convergence of Stochastic Processes. Springer–Verlag, New York

Supported by the OTKA foundation Nr. 037886