Electronic Journal of Probability

Vol. 12 (207), Paper no. 34, pages 966–988.

Journal URL
http://www.math.washington.edu/~ejpecp/

On a multivariate version of Bernstein's inequality *

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
homepage http://www.renyi.hu/~major

Abstract

We prove such a multivariate version of Bernstein's inequality about the tail distribution of degenerate U-statistics which is an improvement of some former results. This estimate will be compared with an analogous bound about the tail distribution of multiple Wiener-Itô integrals. Their comparison shows that our estimate is sharp. The proof is based on good estimates about high moments of degenerate U-statistics. They are obtained by means of a diagram formula which enables us to express the product of degenerate U-statistics as the sum of such expressions.

Key words: Bernstein inequality, (degenerate) U-statistics, Wiener–Itô integrals, diagram formula, moment estimates.

AMS 2000 Subject Classification: Primary 60E15, 60F10; Secondary: 60H99.

Submitted to EJP on March 6, 2006, final version accepted April 30, 2007.

*Research supported by the Hungarian OTKA Foundation Nr. K61052

1 Introduction.

Let us consider a sequence of iid. random variables $\xi_1, \xi_2, \dots$, on a measurable space $(X, \mathcal{X})$ with some distribution μ together with a real valued function $f = f(x_1, \dots, x_k)$ of k variables defined on the k-th power $(X^k, \mathcal{X}^k)$ of the space $(X, \mathcal{X})$ and define with their help the U-statistics $I_{n,k}(f)$, $n = k, k+1, \dots$,

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

We want to get good estimates on the probabilities $P\left(n^{-k/2} k! |I_{n,k}(f)| > u\right)$ for $u > 0$ under appropriate conditions.

Let me first recall a result of Arcones and Giné (2) in this direction. They have proved the inequality

$$\begin{aligned} &P\left(k! n^{-k/2} |I_{n,k}(f)| > u\right) \\ &\qquad \le c_1 \exp\left\{ - \frac{c_2 u^{2/k}}{\sigma^{2/k} \left(1 + c_3 \left(u n^{-k/2} \sigma^{-(k+1)}\right)^{2/k(k+1)}\right)} \right\} \end{aligned} \tag{2}$$

for all $u > 0$ with some universal constants c_1, c_2 and c_3 depending only on the order k of the U-statistic $I_{n,k}(f)$ defined in (1) if the function f satisfies the conditions

$$\|f\|_\infty = \sup_{x_j \in X,\, 1 \le j \le k} |f(x_1, \dots, x_k)| \le 1, \tag{3}$$

$$\|f\|_2^2 = \int f^2(x_1, \dots, x_k) \mu(\,dx_1) \dots \mu(\,dx_k) \le \sigma^2, \tag{4}$$

and it is canonical with respect to the probability measure μ, i.e.

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

A U-statistic defined in (1) with the help of a canonical function f is called degenerate in the literature. Degenerate U-statistics are the natural multivariate versions of sums of iid. random variables with expectation zero.

Actually Arcones and Giné formulated their result in a slightly different but equivalent form. They called their estimate (2) a new Bernstein-type inequality. The reason for such a name is that the original Bernstein inequality (see e.g. (3), 1.3.2 Bernstein inequality) states relation (2) in the special case $k = 1$ with constants $c_1 = 2$, $c_2 = \frac{1}{2}$ and $c_3 = \frac{1}{3}$ if the function $f(x)$ satisfies the conditions $\sup_x |f(x)| \le 1$, $\int f(x)\mu(\,dx) = 0$ and $\int f^2(x)\mu(\,dx) \le \sigma^2$. (Bernstein's inequality states a slightly stronger estimate in the case $k = 1$. It states this inequality with constants $c_1 = 1$, $c_2 = \frac{1}{2}$ and $c_3 = \frac{1}{3}$ if there is no absolute value inside the probability at the left-hand side of (2).)

Our goal is to prove such an improvement of this result which gives the right value of the parameter c_2 in formula (2), and we also want to explain the probabilistic content of such an improvement. For this goal let us first make a more detailed comparison between Bernstein's inequality and estimate (2).

Let us consider the sum $S_n = \sum\limits_{j=1}^{n} \xi_j$ of iid. random variables $\xi_1, \dots, \xi_n$ such that $E\xi_1 = 0$, $P(|\xi_1| \le 1) = 1$, and consider the probability $p_n(u) = P\left(\frac{1}{\sqrt{n}} S_n > u\right)$ for all $u > 0$. Put $\sigma^2 = E\xi_1^2$. Bernstein's inequality implies that

$$p_n(u) \le \exp\left\{-\left(1 - \frac{Ku}{\sqrt{n}\sigma^2}\right)\frac{u^2}{2\sigma^2}\right\} \quad \text{for all } 0 \le u \le \sqrt{n}\sigma^2$$

with some number $K < 1$. A similar estimate holds for $0 \le u \le C\sqrt{n}\sigma^2$ for any number $C > 0$, but in the case $u \gg \sqrt{n}\sigma^2$ only a much weaker inequality holds. (See Example 2.4 in (10) for an example where only a very weak estimate holds if $u \gg \sqrt{n}\sigma^2$.) This means that Bernstein's inequality has the following perturbation type character. For small numbers u (if $0 < u < \varepsilon\sqrt{n}\sigma^2$ with some small $\varepsilon > 0$) the expression in the exponent of the upper bound given for $p_n(u)$ is a small perturbation of $-\frac{\sigma^2}{2}$, of the expression suggested by the central limit theorem. For $u \le \text{const.}\sqrt{n}\sigma^2$ a similar bound holds, only with a worse constant in the exponent. If $u \gg \sqrt{n}\sigma^2$, then no good Gaussian type estimate holds for the probability $p_n(u)$.

Next I formulate the main result of this paper, Theorem 1, which is an estimate similar to that of (2). But, as I will show, it is sharper, and it has a perturbation type character, similar to Bernstein's inequality.

Theorem 1. *Let $\xi_1, \dots, \xi_n$ be a sequence of iid. random variables on a space $(X, \mathcal{X})$ with some distribution μ. Let us consider a function $f(x_1, \dots, x_k)$, canonical with respect to the measure μ on the space $(X^k, \mathcal{X}^k)$ which satisfies conditions (3) and (4) with some $0 < \sigma^2 \le 1$ together with the degenerate U-statistic $I_{n,k}(f)$ with this kernel function f. There exist some constants $A = A(k) > 0$ and $B = B(k) > 0$ depending only on the order k of the U-statistic $I_{n,k}(f)$ such that*

$$P(k!n^{-k/2}|I_{n,k}(f)| > u) \le A\exp\left\{-\frac{u^{2/k}}{2\sigma^{2/k}\left(1 + B\left(un^{-k/2}\sigma^{-(k+1)}\right)^{1/k}\right)}\right\} \tag{5}$$

for all $0 \le u \le n^{k/2}\sigma^{k+1}$.

Remark: Actually, the universal constant $B > 0$ can be chosen independently of the order k of the degenerate U-statistic $I_{n,k}(f)$ in inequality (5).

To understand the content of Theorem 1 better let us recall the following limit distribution result about degenerate U-statistics, (see e.g. (4)). If the canonical function f of k variables satisfies condition (4), then the degenerate U-statistics $n^{-k/2}I_{n,k}(f)$ converge in distribution to the k-fold Wiener–Itô integral $J_{\mu,k}(f)$,

$$J_{\mu,k}(f) = \frac{1}{k!}\int f(x_1, \dots, x_k)\mu_W(\,dx_1)\dots\mu_W(dx_k), \tag{6}$$

of the function f with respect to a white noise μ_W with reference measure μ. Here μ is the distribution of the random variables ξ_j, $j = 1, 2, \dots$ appearing in the U-statistics $I_{n,k}(f)$. Let me recall that a white noise μ_W with reference measure μ on $(X, \mathcal{X})$ is a set of jointly Gaussian random variables $\mu_W(A)$, $A \in \mathcal{X}$, $\mu(A) < \infty$, such that $E\mu_W(A) = 0$, $E\mu_W(A)\mu_W(B) = \mu(A \cap B)$ for all $A \in \mathcal{X}$ and $B \in \mathcal{X}$. The definition of Wiener–Itô integrals can be found for instance in (6) or (8).

The above result suggests to describe the tail-distribution of the Wiener–Itô integral $J_{\mu,k}(f)$ and to show that Theorem 1 gives such an estimate which the above mentioned limit theorem and the tail distribution of $J_{\mu,k}(f)$ suggests. At this moment there appears an essential difference between the problem discussed in Bernstein's inequality and its multivariate version.

We want to estimate both the U-statistic $I_{n,k}(f)$ and the Wiener–Itô integral $J_{\mu,k}(f)$ by means of their variance. (Let me remark that the integral in formula (4) equals the variance of $(k!)^{1/2}J_{\mu,k}(f)$, and it is asymptotically equal to the variance of $(k!)^{1/2}n^{-k/2}I_{n,k}(f)$ for large n. At least, this is the case if f is a symmetric function of its variables. But, since $I_{n,k}(f) = I_{n,k}(\operatorname{Sym} f)$, $J_{\mu,k}(f) = J_{\mu,k}(\operatorname{Sym} f)$, and $\|\operatorname{Sym} f\|_2^2 \le \|f\|_2^2$ we may restrict our attention to this case.) But while the variance and expectation determines the distribution of a Gaussian random variable, the distribution of a Wiener–Itô integral is not determined by its variance and (zero) expectation. Hence if we want to compare the estimation of degenerate U-statistics by means of their variance with a natural Gaussian counterpart of this problem, then it is natural to consider first the following problem.

Find such an upper estimate for the tail distribution of Wiener–Itô integrals which holds for all of them with a prescribed bound on their variances, and which is sharp in the following sense. There is a Wiener–Itô integral whose variance is not larger than the prescribed bound, and which satisfies a very similar lower estimate. Then the estimate for degenerate U-statistics has to be compared with such an estimate for Wiener–Itô integrals. The following Theorem 2 and Example 3 give an estimate for Wiener–Itô integrals with the desired properties. (These results were proven in (11).) They suggest to compare the upper bound in Theorem 1 with the function const. $\exp\left\{-\frac{1}{2}\left(\frac{u}{\sigma}\right)^{2/k}\right\}$ with some appropriate constant.

Theorem 2. *Let us consider a σ-finite measure μ on a measurable space together with a white noise μ_W with reference measure μ. Let us have a real-valued function $f(x_1, \dots, x_k)$ on the space $(X^k, \mathcal{X}^k)$ which satisfies relation (4) with some $\sigma^2 < \infty$. Take the random integral $J_{\mu,k}(f)$ introduced in formula (6). This random integral satisfies the inequality*

$$P(k!|J_{\mu,k}(f)| > u) \le C \exp\left\{-\frac{1}{2}\left(\frac{u}{\sigma}\right)^{2/k}\right\} \quad \text{for all } u > 0 \tag{7}$$

with an appropriate constant $C = C(k) > 0$ depending only on the multiplicity k of the integral.

Example 3. *Let us have a σ-finite measure μ on some measure space $(X, \mathcal{X})$ together with a white noise μ_W on $(X, \mathcal{X})$ with reference measure μ. Let $f_0(x)$ be a real valued function on $(X, \mathcal{X})$ such that $\int f_0(x)^2\mu(\,dx) = 1$, and take the function $f(x_1, \dots, x_k) = \sigma f_0(x_1)\cdots f_0(x_k)$ with some number $\sigma > 0$ and the Wiener–Itô integral $J_{\mu,k}(f)$ introduced in formula (6).*

Then the relation $\int f(x_1, \dots, x_k)^2\,\mu(\,dx_1)\dots\,\mu(\,dx_k) = \sigma^2$ holds, and the random integral $J_{\mu,k}(f)$ satisfies the inequality

$$P(k!|J_{\mu,k}(f)| > u) \ge \frac{\bar{C}}{\left(\frac{u}{\sigma}\right)^{1/k} + 1} \exp\left\{-\frac{1}{2}\left(\frac{u}{\sigma}\right)^{2/k}\right\} \quad \text{for all } u > 0 \tag{8}$$

with some constant $\bar{C} > 0$.

By Theorem 1 there are some constants $\alpha > 0$, $C_1 > 0$, $1 > C_2 > 0$, $C_1\alpha < 1$ such that under the conditions of this result

$$P(k!n^{-k/2}|I_{n,k}(f)| > u) \le A\exp\left\{-\frac{1}{2}\left(\frac{u}{\sigma}\right)^{2/k}\left(1 - C_1\left(\frac{u}{n^{k/2}\sigma^{k+1}}\right)^{1/k}\right)\right\}$$
$$\text{if } 0 < u \le \alpha n^{k/2}\sigma^{k+1}$$

and

$$P(k!n^{-k/2}|I_{n,k}(f)| > u) \le A\exp\left\{-C_2\left(\frac{u}{\sigma}\right)^{2/k}\right\}$$
$$\text{if } \alpha n^{k/2}\sigma^{k+1} < u \le n^{k/2}\sigma^{k+1}.$$

A comparison of these estimates with Theorem 2 and Example 3 shows that Theorem 1 has a behaviour similar to that of Bernstein's inequality. For relatively small numbers $u > 0$, more precisely if $0 < u < \varepsilon n^{k/2}\sigma^{k+1}$ with some $\varepsilon > 0$, the expression in the exponent at the right-hand side of this estimate is very close to $-\frac{1}{2}\left(\frac{u}{\sigma}\right)^{2/k}$, the term suggested by Theorem 2 and Example 3. In the more general case $u \le n^{k/2}\sigma^{k+1}$ a similar, but somewhat worse estimate holds. The term $-\left(\frac{u}{\sigma}\right)^{2/k}$ in the upper estimate is multiplied by a constant $C_2 > 0$ in this case which may be much smaller than $\frac{1}{2}$. So the estimate of Theorem 1 has such a perturbation type character which is missing from the estimate (2).

On the other hand it may seem that the estimate (2) has the advantage that it yields a bound for the tail-distribution of a degenerate U-statistic for all numbers $u > 0$, while formula (5) holds only under the condition $0 \le u \le n^{k/2}\sigma^{k+1}$. Nevertheless, formula (5) implies such an estimate also for $u > n^{k/2}\sigma^{k+1}$ which is not weaker than the inequality (2) (at least if we do not bother about the value of the universal constants in these estimates). To see this observe that relation (3) remains valid if σ^2 is replaced by any $\bar{\sigma}^2 \ge \sigma^2$. As a consequence, for $n^{k/2} \ge u > n^{k/2}\sigma^{k+1}$ relation (5) holds with the replacement of σ by $\bar{\sigma} = \left(un^{-k/2}\right)^{1/(k+1)}$, since all conditions of Theorem 1 are satisfied with such a choice. It yields that $P(k!n^{-k/2}|I_{n,k}(f)| > u) \le A\exp\left\{-\frac{1}{2(1+B)^{1/k}}\left(\frac{u}{\bar{\sigma}}\right)^{2/k}\right\} = Ae^{-(u^2n)^{1/(k+1)}/2(1+B)^{1/k}}$. On the other hand, $\sigma^{2/k}\left(1 + c_3\left(un^{-k/2}\sigma^{-(k+1)}\right)^{2/k(k+1)}\right) \ge c_3u^{2/k(k+1)}n^{-1/(k+1)}$, hence the right-hand side of (2) can be bounded from below by $c_1e^{-c_2(u^2n)^{1/(k+1)}/c_3}$. Thus relation (5) implies relation (1.2) if $n^{k/2} \ge u > n^{k/2}\sigma^{k+1}$ with possibly worse constants $\bar{c}_1 = A$, c_2 and $\bar{c}_3 = 2c_2(1+B)^{1/k}$. If $u > n^{k/2}$, then the left-hand side of (2) equals zero because of the boundedness of the function f, and relation (2) clearly holds.

Actually the condition $u \le n^{k/2}\sigma^{k+1}$ was rather natural in Theorem 1. It can be shown that in the case $u \gg n^{k/2}\sigma^{k+1}$ there are such degenerate U-statistics satisfying the conditions of Theorem 1 for which the probability $P(n^{-k/2}k!I_{n,k}(f) > u)$ is much greater than the expression suggested by the limit theorem for degenerate U-statistics together with Theorem 2 and Example 3. Such an example is presented in Examples 4.5 in (10) for $k = 2$. With some extra work similar examples of degenerate U-statistics of order k could also be constructed for any $k = 2, 3, \ldots$.

Let me say some words about the method of proofs. Theorem 1 will be proved by means of good estimates on high moments of degenerate U-statistics. These moment estimates will be obtained with the help of a new type of diagram formula which enables us to write the product of degenerate U-statistics as the sum of degenerate U-statistics. Such a formula may be interesting in itself. It is a version of an important result about the representation of a product of Wiener–Itô integrals in the form of sums of Wiener–Itô integrals. It makes possible to adapt the methods in the theory of Wiener–Itô integrals to the study of degenerate U-statistics. It also gives some insight why the tail distributions of degenerate U-statistics and Wiener–Itô integral satisfy similar estimates.

This approach is essentially different from that of earlier papers in this field, e.g. from the proof of paper (2). I had to choose a different method, because the technique of previous papers was not strong enough to prove Theorem 1. They give only such weaker estimates for high moments of degenerate U-statistics which are not sufficient for our purposes. This weakness has different causes. First, previous proofs apply an estimate called Borell's inequality in the literature, which does not supply a sharp estimate in certain cases. This has the consequence that we can get only a relatively weak estimate about high moments of degenerate U-statistics in such a way. (See the end of my paper (11) for a more detailed discussion of this problem.) Beside this, earlier papers in this field apply a method called the decoupling technique in the literature, and this method has some properties which enable only the proof a weaker version of Theorem 1.

The decoupling technique contains some randomization procedure, and as a more careful analysis shows, its application allows us to prove only relatively weak estimates. The randomization procedure applied in the decoupling technique makes possible to reduce the estimation of the degenerate U-statistic we want to bound to the estimation of another degenerate U-statistic which can be better handled. But this new U-statistic has a larger variance than the original one. As a consequence, this method cannot give such a good estimate which 'resembles' to the limit distribution of the original U-statistic. Hence for relatively small numbers u it supplies a weaker estimate for the distribution of degenerate U-statistics than formula 1.5.

Let me still remark that at recent time some new estimates are proved about the tail distribution of degenerate U-statistics. (See (1), (5), (7).) They may supply a better bound in certain cases with the help of some additional quantities related to the properties of the kernel function of the U-statistic. Such problems will be not discussed in this paper, but I would remark that the method of this paper may work also in such investigations. The diagram formula supplies a better estimate for the moments of a degenerate U-statistic if its kernel function has some nice properties. There is some hope that the recent results about the tail distribution of degenerate U-statistics can be proved in such a way.

This paper consists of six sections. In Section 2 the proof of Theorem 1 is reduced to a moment estimate for degenerate U-statistics formulated in that section. To understand the content of this moment estimate better I also present its Wiener–Itô integral counterpart. Theorem 2 follows from this moment estimate for Wiener–Itô integrals in a standard way. This proof will be omitted, since it can be found in (11). The proof of Example 3 can also be found in (11), hence its proof will be also omitted. Sections 3, 4 and 5 contain the proof of the diagram formula for the product of degenerate U-statistics needed in the proof of the moment estimate in Section 2. The diagram formula about the product of two degenerate U-statistics is formulated in Section 3, and its proof is given in Section 4. Section 5 contains the formulation and proof of the diagram formula for the products of degenerate U-statistics in the general case. In Section 6 the moment estimate given is Section 2 is proved by means of the diagram formula. In such a way the proof of Theorem 1 is completed.

2 The reduction of the proof of Theorem 1 to a moment estimate.

Theorem 1 will be proved by means of the following

Proposition A. *Let us consider a degenerate U-statistic $I_{n,k}(f)$ of order k with sample size n and with a kernel function f satisfying relations (3) and (4) with some $0 < \sigma^2 \le 1$. Fix a positive number $\eta > 0$. There exist some universal constants $A = A(k) > \sqrt{2}$, $C = C(k) > 0$ and $M_0 = M_0(k) \ge 1$ depending only on the order k of the U-statistic $I_{n,k}(f)$ such that*

$$E\left(n^{-k/2}k!I_{n,k}(f)\right)^{2M} \le A\,(1+C\sqrt{\eta})^{2kM}\left(\frac{2}{e}\right)^{kM}(kM)^{kM}\,\sigma^{2M}$$
$$\text{for all integers } M \text{ such that } kM_0 \le kM \le \eta n\sigma^2. \tag{9}$$

The constant $C = C(k)$ in formula (9) can be chosen e.g. as $C = 2\sqrt{2}$ which does not depend on the order k of the U-statistic $I_{n,k}(f)$.

To understand the content of Proposition A better I formulate its Wiener–Itô integral counterpart in the following

Proposition B. *Let the conditions of Theorem 2 be satisfied for a multiple Wiener–Itô integral $J_{\mu,k}(f)$ of order k. Then, with the notations of Theorem 2, the inequality*

$$E\,(k!|J_{\mu,k}(f)|)^{2M} \le 1\cdot 3\cdot 5\cdots(2kM-1)\sigma^{2M} \quad \text{for all } M = 1,2,\dots \tag{10}$$

holds.

By the Stirling formula Proposition B implies that

$$E(k!|J_{\mu,k}(f)|)^{2M} \le \frac{(2kM)!}{2^{kM}(kM)!}\sigma^{2M} \le A\left(\frac{2}{e}\right)^{kM}(kM)^{kM}\sigma^{2M} \tag{11}$$

for any $A > \sqrt{2}$ if $M \ge M_0 = M_0(A)$. The right-hand side of formula (10) is almost as large as the right-hand side of formula (11). Hence the estimate (11) gives an almost as good estimate as Proposition B. We shall use this estimate in the sequel because of its simpler form.

Proposition B can be considered as a corollary of a most important result about Wiener–Itô integrals called the diagram formula. This result enables us to rewrite the product of Wiener–Itô integrals as a sum of Wiener–Itô integrals of different order. It got the name 'diagram formula' because the kernel functions of the Wiener–Itô integrals appearing in the sum representation of the product of Wiener–Itô integrals are defined with the help of certain diagrams. As the expectation of a Wiener–Itô integral of order k equals zero for all $k \ge 1$, the expectation of the product is equal to the sum of the constant terms (i.e. of the integrals of order zero) in the diagram formula. In such a way the diagram formula yields an explicit (although somewhat complicated) formula about the moments of Wiener–Itô integrals. Proposition B can be proved relatively simply by means of this relation. Since it is written down in paper (11), I omit the details.

We shall see that there is such a version of the diagram formula which expresses the product of degenerate U-statistics as a sum of degenerate U-statistics of different order by means of some appropriately defined diagrams. Proposition A can be proved by means of this version of the diagram formula similarly to Proposition B. The proof of Proposition A with the help of this version of the diagram formula will be given in Section 6. The main difference between the proof of Propositions A and B with the help of the corresponding diagram formula is that in the case of degenerate U-statistics the diagram formula contains some additional new diagrams, and their contribution also has to be estimated. It will be shown that if not too high moments of U-statistics are calculated by means of the diagram formula, then the contribution of the new diagrams is not too large.

To understand better the content of Proposition A let us compare formulas (9) and (11). These estimates are very similar. The upper bound given for the $2M$-th moment of a degenerate U-statistic in formula (9) is less than A^M-times the upper bound given for the $2M$-th moment of the corresponding Wiener–Itô integral in formula (11) with some universal constant $A>1$. Moreover, the constant A is very close to 1 if the parameter M is relatively small, if $M\le\varepsilon n\sigma^2$ with some small number $\varepsilon>0$. But the estimate (9) holds only for not too large parameters M, because of the condition $kM<\eta n\sigma^2$ in it. Because of this condition Proposition A gives a much worse bound for the $2M$-th moment of a degenerate U-statistic if $M\gg n\sigma^2$ than inequality (11) yields for the $2M$-th moment of the corresponding Wiener–Itô integral. These properties of the moment estimates in Proposition A are closely related to the behaviour of the estimate in Theorem 1, in particular to the condition $u\le n^{k/2}\sigma^{k+1}$ in it.

Theorem 2 can be proved by means of Proposition B and the Markov inequality $P(|J_{\mu,k}(f)|>u)\le\frac{EJ_{\mu,k}(f)^{2M}}{u^{2M}}$ with a good choice of the parameter M. This is a rather standard approach, and this proof is written down in (11). Hence I omit it. Theorem 1 can be proved similarly with the help of Proposition A and the Markov inequality, but in this case a more careful analysis is needed to find the good choice of the parameter M with which the Markov inequality should be applied. I work out the details.

Proof of Theorem 1 by means of Proposition A. We can write by the Markov inequality and Proposition A with the choice $\eta=\frac{kM}{n\sigma^2}$ that

$$\begin{aligned}P(k!n^{-k/2}|I_{n,k}(f)|>u) &\le \frac{E\left(k!n^{-k/2}I_{n,k}(f)\right)^{2M}}{u^{2M}} \\ &\le A\left(\frac{1}{e}\cdot 2kM\left(1+C\frac{\sqrt{kM}}{\sqrt{n}\sigma}\right)^2\left(\frac{\sigma}{u}\right)^{2/k}\right)^{kM}\end{aligned}\tag{12}$$

for all integers $M\ge M_0$ with some $M_0=M_0(k)$ and $A=A(k)$.

We shall prove relation (5) with the help of estimate (12) first in the case $D\le\frac{u}{\sigma}\le n^{k/2}\sigma^k$ with a sufficiently large constant $D=D(k,C)>0$ depending on k and the constant C in (12). To this end let us introduce the numbers $\bar M$,

$$k\bar M=\frac12\left(\frac u\sigma\right)^{2/k}\frac{1}{1+B\frac{\left(\frac u\sigma\right)^{1/k}}{\sqrt n\sigma}}=\frac12\left(\frac u\sigma\right)^{2/k}\frac{1}{1+B\left(un^{-k/2}\sigma^{-(k+1)}\right)^{1/k}} \tag{13}$$

with a sufficiently large number $B=B(C)>0$ and $M=[\bar M]$, where $[x]$ means the integer part of the number x.
Observe that $\sqrt{k\bar M}\le\left(\frac u\sigma\right)^{1/k}$, $\frac{\sqrt{k\bar M}}{\sqrt n\sigma}\le\left(un^{-k/2}\sigma^{-(k+1)}\right)^{1/k}\le1$, and

$$\left(1+C\frac{\sqrt{k\bar M}}{\sqrt n\sigma}\right)^2\le1+B\frac{\sqrt{k\bar M}}{\sqrt n\sigma}\le1+B\left(un^{-k/2}\sigma^{-(k+1)}\right)^{1/k}$$

with a sufficiently large $B=B(C)>0$ if $\frac u\sigma\le n^{k/2}\sigma^k$. Hence

$$\begin{aligned}\frac1e\cdot2kM\left(1+C\frac{\sqrt{kM}}{\sqrt n\sigma}\right)^2\left(\frac\sigma u\right)^{2/k}&\le\frac1e\cdot2k\bar M\left(1+C\frac{\sqrt{k\bar M}}{\sqrt n\sigma}\right)^2\left(\frac\sigma u\right)^{2/k}\\&\le\frac1e\cdot\frac{\left(1+C\frac{\sqrt{k\bar M}}{\sqrt n\sigma}\right)^2}{1+B\left(un^{-k/2}\sigma^{-(k+1)}\right)^{1/k}}\le\frac1e\end{aligned}\tag{14}$$

if $\frac u\sigma\le n^{k/2}\sigma^k$. If the inequality $D\le\frac u\sigma$ also holds with a sufficiently large $D=D(B,k)>0$, then $M=[\bar M]\ge M_0$ because of the definition of $[\bar M]$ in formula (13) and the relation $un^{-k/2}\sigma^{k+1}\le1$. With such a choice the conditions of inequality (12) hold. By applying it together with inequality (14) we get that

$$P(k!n^{-k/2}|I_{n,k}(f)|>u)\le Ae^{-kM}\le Ae^ke^{-k\bar M}$$

if $D\le\frac u\sigma\le n^{k/2}\sigma^k$. This means that inequality (5) holds in this case with a pre-exponential constant Ae^k. Since $e^{-k\bar M}$ is bounded from below for $\frac u\sigma\le D$ relation (5) holds for all $0\le\frac u\sigma\le n^{k/2}\sigma^k$ with a possible increase of the pre-exponential coefficient Ae^k in it. Theorem 1 is proved.

Let us observe that the above calculations show that the constant B in formula (8) can be chosen independently of the order k of the U-statistics $I_{n,k}(f)$.

3 The diagram formula for the product of two degenerate U-statistics.

To prove Proposition A we need a good identity which expresses the expectation of the product of degenerate U-statistics in a form that can be better handled. Such an identity can be proved by means of a version of the diagram formula for Wiener–Itô integrals where the product of degenerate U-statistics is represented as the sum of degenerate U-statistics with appropriate kernel functions. In such a formula the kernel functions of the sum representation are defined with the help of some diagrams, and to get a useful result we also need a good estimate on their L_2-norm.

We shall prove such a result. First we prove its special case about the product of two degenerate U-statistics together with a good estimate on the L_2-norm of the kernel functions in the sum representation. Then the result in the general case can be obtained by induction.

In the case of the product of two degenerate U-statistics the result we want to prove can be obtained with the help of the following observation. Let us have a sequence of iid. random variables $\xi_1, \xi_2, \dots$ with some distribution μ on a measurable space $(X, \mathcal{X})$ together with two functions $f(x_1, \dots, x_{k_1})$ and $g(x_1, \dots, x_{k_2})$ on $(X^{k_1}, \mathcal{X}^{k_1})$ and on $(X^{k_2}, \mathcal{X}^{k_2})$ respectively which are canonical with respect to the probability measure μ. We consider the degenerate U-statistics $I_{n,k_1}(f)$ and $I_{n,k_2}(g)$ and want to express their normalized product $k_1!k_2!n^{-(k_1+k_2)/2} I_{n,k_1}(f) I_{n,k_2}(g)$ as a sum of (normalized) degenerate U-statistics. This product can be presented as a sum of U-statistics in a natural way. Then by writing each term of this sum as a sum of degenerate U-statistics by means of the Hoeffding decomposition we get the desired representation of the product. This result will be formulated in Theorem A.

In this Section Theorem A will be described together with the introduction of the notations needed for its formulation. Its proof will be given in the next Section.

To define the kernel functions of the U-statistics appearing in the diagram formula for the product of two U-statistics first we introduce a class of objects $\Gamma(k_1, k_2)$ we shall call coloured diagrams. We define graphs $\gamma \in \Gamma(k_1, k_2)$ that contain the vertices $(1,1), (1,2), \dots, (1, k_1)$ which we shall call the first row and $(2,1) \dots, (2, k_2)$ which we shall call the second row of these graphs. From each vertex there starts zero or one edge, and each edge connects vertices from different rows. Each edge will get a colour $+1$ or -1. $\Gamma(k_1, k_2)$ consists of all γ obtained in such a way. These objects γ will be called coloured diagrams.

Given a coloured diagram $\gamma \in \Gamma(k_1, k_2)$ let $B_u(\gamma)$ denote the set of upper end-points $(1, j)$ of the edges of the graph γ, $B_{(b,1)}(\gamma)$ the set of lower end-points $(2, j)$ of the edges of γ with colour 1, and $B_{(b,-1)}(\gamma)$ the set of lower end-points $(2, j)$ of the edges of γ with colour -1. (The letter 'b' in the index was chosen because of the word below.) Finally, let $Z(\gamma)$ denote the set of edges with colour 1, $W(\gamma)$ the set of edges with colour -1 of a coloured graph $\gamma \in \Gamma(k_1, k_2)$, and let $|Z(\gamma)|$ and $|W(\gamma)|$ denote their cardinality.

Given two functions $f(x_1, \dots, x_{k_1})$ and $g(x_1, \dots, x_{k_2})$ let us define the function

$$
\begin{aligned}
&(f \circ g)(x_{(1,1)}, \dots, x_{(1,k_1)}, x_{(2,1)}, \dots, x_{(2,k_2)}) \\
&\qquad = f(x_{(1,1)}, \dots, x_{(1,k_1)}) g(x_{(2,1)}, \dots, x_{(2,k_2)})
\end{aligned}
\tag{15}
$$

Given a function $h(x_{u_1}, \dots, x_{u_r})$ with coordinates in the space $(X, \mathcal{X})$ (the indices $u_1, \dots, u_r$ are all different) let us introduce its transforms $P_{u_j} h$ and $Q_{u_j} h$ by the formulas

$$
(P_{u_j} h)(x_{u_l} \colon\ u_l \in \{u_1, \dots, u_r\} \setminus \{u_j\}) = \int h(x_{u_1}, \dots, x_{u_r}) \mu(\,dx_{u_j}), \quad 1 \le j \le r, \tag{16}
$$

and

$$
(Q_{u_j} h)(x_{u_1}, \dots, x_{u_r}) = h(x_{u_1}, \dots, x_{u_r}) - \int h(x_{u_1}, \dots, x_{u_r}) \mu(\,dx_{u_j}), \;\; 1 \le j \le r. \tag{17}
$$

At this point I started to apply a notation which may seem to be too complicated, but I think that it is more appropriate in the further discussion. Namely, I started to apply a rather general enumeration $u_1,\dots,u_r$ of the arguments of the functions we are working with instead of their simpler enumeration with indices $1,\dots,r$. But in the further discussion there will appear an enumeration of the arguments by pairs of integers (l,j) in a natural way, and I found it simpler to work with such an enumeration than to reindex our variables all the time. Let me remark in particular that this means that the definition of the U-statistic with a kernel function $f(x_1,\dots,x_k)$ given in formula (1) will appear sometimes in the following more complicated, but actually equivalent form: We shall work with kernel function $f(x_{u_1},\dots,x_{u_k})$ instead of $f(x_1,\dots,x_k)$, the random variables ξ_j will be indexed by u_s, i.e. to the coordinate x_{u_s} we shall put the random variables $\xi_{j_{u_s}}$ with indices $1\le j_{u_s}\le n$, and in the new notation formula (1) will look like

$$I_{n,k}(f)=\frac{1}{k!}\sum_{\substack{1\le j_{u_s}\le n,\ s=1,\dots,k\\ j_{u_s}\neq j_{u'_s} \text{ if } u_s\neq u'_s}} f\left(\xi_{j_{u_1}},\dots,\xi_{j_{u_k}}\right).$$

Let us define for all coloured diagrams $\gamma\in\Gamma(k_1,k_2)$ the function $\alpha_\gamma(1,j)$, $1\le j\le k_1$, on the vertices of the first row of γ as $\alpha_\gamma(1,j)=(1,j)$ if no edge starts from $(1,j)$, and $\alpha_\gamma(1,j)=(2,j')$ if an edge of γ connects the vertices $(1,j)$ and $(2,j')$. Given two functions $f(x_1,\dots,x_{k_1})$ and $g(x_1,\dots,x_{k_2})$ together with a coloured diagram $\gamma\in\Gamma(k_1,k_2)$ let us introduce, with the help of the above defined function $\alpha_\gamma(\cdot)$ and $(f\circ g)$ introduced in (15) the function

$$\begin{aligned}&\overline{(f\circ g)}_\gamma(x_{(1,j)},x_{(2,j')},\ j\in\{1,\dots,k_1\}\setminus B_u(\gamma),\ 1\le j'\le k_2)\\ &\qquad=(f\circ g)(x_{\alpha_\gamma(1,1)},\dots,x_{\alpha_\gamma(1,k_1)},x_{(2,1)},\dots,x_{(2,k_2)}).\end{aligned}\tag{18}$$

(In words, we take the function $(f\circ g)$, and if there is an edge of γ starting from a vertex $(1,j)$, and it connects this vertex with the vertex $(2,j')$, then the argument $x_{(1,j)}$ is replaced by the argument $x_{(2,j')}$ in this function.) Let us also introduce the function

$$\begin{aligned}&(f\circ g)_\gamma\left(x_{(1,j)},x_{(2,j')},\ j\in\{1,\dots,k_1\}\setminus B_u(\gamma),\ j'\in\{1,\dots,k_2\}\setminus B_{(b,1)}\right)\\ &\quad=\prod_{(2,j')\in B_{(b,1)}(\gamma)}P_{(2,j')}\prod_{(2,j')\in B_{(b,-1)}(\gamma)}Q_{(2,j')}\\ &\qquad\overline{(f\circ g)}_\gamma\left(x_{(j,1)},x_{(j',2)},\ j\in\{1,\dots,k_1\}\setminus B_u(\gamma),\ 1\le j'\le k_2\right).\end{aligned}\tag{19}$$

(In words, we take the function $\overline{(f\circ g)}_\gamma$ and for such indices $(2,j')$ of the graph γ from which an edge with colour 1 starts we apply the operator $P_{(2,j')}$ introduced in formula (16) and for those indices $(2,j')$ from which an edge with colour -1 starts we apply the operator $Q_{(2,j')}$ defined in formula (17).) Let us also remark that the operators $P_{(2,j')}$ and $Q_{(2,j')}$ are exchangeable for different indices j', hence it is not important in which order we apply the operators $P_{(2,j')}$ and $Q_{(2,j')}$ in formula (19).

In the definition of the function $(f\circ g)_\gamma$ those arguments $x_{(2,j')}$ of the function $\overline{(f\circ g)}_\gamma$ which are indexed by such a pair $(2,j')$ from which an edge of colour 1 of the coloured diagram γ starts will disappear, while the arguments indexed by such a pair $(2,j')$ from which an edge of colour -1 of the coloured diagram γ starts will be preserved. Hence the number of arguments in the function $(f\circ g)_\gamma$ equals $k_1+k_2-2|B_{(b,1)}(\gamma)|-|B_{(b,-1)}(\gamma)|$, where $|B_{(b,1)}(\gamma)|$ and $|B_{(b,-1)}(\gamma)|$ denote the cardinality of the lower end-points of the edges of the coloured diagram γ with colour 1 and -1 respectively, In an equivalent form we can say that the number of arguments of $(f\circ g)_\gamma$ equals $k_1+k_2-(2|Z(\gamma)|+|W(\gamma)|)$.

Now we are in the position to formulate the diagram formula for the product of two degenerate U-statistics.

Theorem A. *Let us have a sequence of iid. random variables $\xi_1, \xi_2, \dots$ with some distribution μ on some measurable space $(X, \mathcal{X})$ together with two bounded, canonical functions $f(x_1, \dots, x_{k_1})$ and $g(x_1, \dots, x_{k_2})$ with respect to the probability measure μ on the spaces $(X^{k_1}, \mathcal{X}^{k_1})$ and $(X^{k_2}, \mathcal{X}^{k_2})$. Let us introduce the class of coloured diagrams $\Gamma(k_1, k_2)$ defined above together with the functions $(f \circ g)_\gamma$ defined in formulas (15)—(19).*

For all $\gamma \in \Gamma$ the function $(f \circ g)_\gamma$ is canonical with respect to the measure μ with $k(\gamma) = k_1 + k_2 - (2|Z(\gamma)| + |W(\gamma)|)$ arguments, where $|Z(\gamma)|$ denotes the number of edges with colour 1 and $|W(\gamma)|$ the number of edges with colour -1 of the coloured diagram γ. The product of the degenerate U-statistics $I_{n,k_1}(f)$ and $I_{n,k_2}(g)$, $n \ge \max(k_1, k_2)$, defined in (1) satisfies the identity

$$\begin{aligned} &k_1! n^{-k_1/2} I_{n,k_1}(f) k_2! n^{-k_2/2} I_{n,k_2}(g) \\ &\quad = \sideset{}{'^{(n)}}\sum_{\gamma \in \Gamma(k_1,k_2)} \frac{\prod\limits_{j=1}^{|Z(\gamma)|} (n - (k_1 + k_2) + |W(\gamma)| + |Z(\gamma)| + j)}{n^{|Z(\gamma)|}} \\ &\qquad\qquad n^{-|W(\gamma)|/2} \cdot k(\gamma)! n^{-k(\gamma)/2} I_{n,k(\gamma)}((f \circ g)_\gamma), \end{aligned} \tag{20}$$

where $\sum'^{(n)}$ means that summation is taken only for such coloured diagrams $\gamma \in \Gamma(k_1, k_2)$ which satisfy the inequality $k_1 + k_2 - (|Z(\gamma)| + |W(\gamma)|) \le n$, and $\prod\limits_{j=1}^{|Z(\gamma)|}$ equals 1 in the case $|Z(\gamma)| = 0$.

The L_2-norm of the functions $(f \circ g)_\gamma$ is defined by the formula

$$\begin{aligned} &\|(f \circ g)_\gamma\|_2^2 \\ &\quad = \int (f \circ g)_\gamma^2 (x_{(1,j)}, x_{(2,j')},\ j \in \{1, \dots, k_1\} \setminus B_u(\gamma),\ j' \in \{1, \dots, k_2\} \setminus B_{(b,1)}) \\ &\qquad\qquad \prod_{(1,j):\ j \in \{1,\dots,k_1\} \setminus B_u(\gamma)} \mu(\,dx_{(1,j)}) \prod_{(2,j'):\ j' \in \{1,\dots,k_2\} \setminus B_{(b,1)}} \mu(\,dx_{(2,j')}). \end{aligned}$$

If $W(\gamma) = 0$, then the inequality

$$\|(f \circ g)_\gamma\|_2 \le \|f\|_2 \|g\|_2 \tag{21}$$

holds. In the general case we can say that if the functions f and g satisfy formula (3), then the inequality

$$\|(f \circ g)_\gamma\|_2 \le 2^{|W(\gamma)|} \min(\|f\|_2, \|g\|_2) \tag{22}$$

holds. Relations (21) and (22) remain valid if we drop the condition that the functions f and g are canonical.

Relations (21) and (22) mean in particular, that we have a better estimate for $\|(f \circ g)_\gamma\|_2$ in the case when the coloured diagram γ contains no edge with colour -1, i.e. if $|W(\gamma)| = 0$, than in the case when it contains at least one edge with colour -1.

Let us understand how we define those terms at the right-hand side of (20) for which $k(\gamma) = 0$. In this case $(f \circ g)_\gamma$ is a constant, and to make formula (20) meaningful we have to define the term $I_{n,k(\gamma)}((f \circ g)_\gamma)$ also in this case. The following convention will be used. A constant c will be called a degenerate U-statistic of order zero, and we define $I_{n,0}(c) = c$.

Theorem A can be considered as a version of the result of paper (9), where a similar diagram formula was proved about multiple random integrals with respect to normalized empirical measures. Degenerate U-statistics can also be presented as such integrals with special, canonical kernel functions. Hence there is a close relation between the results of this paper and (9). But there are also some essential differences. For one part, the diagram formula for multiple random integrals with respect to normalized empirical measures is simpler than the analogous result about the product of degenerate U-statistics, because the kernel functions in these integrals need not be special, canonical functions. On the other hand, the diagram formula for degenerate U-statistics yields a simpler formula about the expected value of the product of degenerate U-statistics, because the expected value of a degenerate U-statistic of order $k \geq 1$ equals zero, while the analogous result about multiple random integrals with respect to normalized empirical measures may not hold. Another difference between this paper and (9) is that here I worked out a new notation which, I hope, is more transparent.

4 The proof of Theorem A.

The proof of Theorem A. Let us consider all possible sets $\{(u_1, u'_1), \dots, (u_l, u'_l)\}$, $1 \leq l \leq \min(k_1, k_2)$ containing such pairs of integers for which $u_s \in \{1, \dots, k_1\}$, $u'_s \in \{1, \dots, k_2\}$, $1 \leq s \leq l$, all points $u_1, \dots, u_l$ are different, and the same relation holds for the points $u'_1, \dots, u'_l$, too. Let us correspond the diagram containing two rows $(1,1), \dots, (1,k_1)$ and $(2,1), \dots, (2,k_2)$ and the edges connecting the vertices $(1, u_s)$ and $(2, u'_s)$, $1 \leq s \leq l$ to the set of pairs $\{(u_1, u'_1), \dots, (u_l, u'_l)\}$, and let $\bar\Gamma(k_1, k_2)$ denote the set of all (non-coloured) diagrams we can obtain in such a way. Let us consider the product $k_1! I_{n,k_1}(f) k_2! I_{n,k_2}(g)$, and rewrite it in the form of the sum we get by carrying out a term by term multiplication in this expression. Let us put the terms of this sum into disjoint classes indexed by the elements of the diagrams $\bar\gamma \in \bar\Gamma(k_1, k_2)$ in the following way: A product $f(\xi_{j_1}, \dots, \xi_{j_{k_1}}) g(\xi_{j'_1}, \dots, \xi_{j'_{k_2}})$ belongs to the class indexed by the graph $\bar\gamma \in \bar\Gamma(k_1, k_2)$ with edges $\{((1, u_1), (2, u'_1)), \dots, ((1, u_l), (2, u'_l))\}$ if $j_{u_s} = j'_{u'_s}$, $1 \leq s \leq l$, for the indices of the random variables appearing in the above product, and no more coincidence may exist between the indices $j_1, \dots, j_{k_1}, j'_1, \dots, j'_{k_2}$. With such a notation we can write

$$n^{-k_1/2} k_1! I_{n,k_1}(f) n^{-k_2/2} k_2! I_{n,k_2}(g) = \sum_{\bar\gamma \in \bar\Gamma}{}^{\prime(\mathrm{n})} n^{-(k_1+k_2)/2} \bar k(\bar\gamma)! I_{n,\bar k(\bar\gamma)}(\overline{f \circ g})_{\bar\gamma}), \tag{23}$$

where the functions $(\overline{f \circ g})_{\bar\gamma}$ are defined in formulas (15) and (18). (Observe that a lthough formula (18) was defined by means of coloured diagrams, the colours played no role in it. The formula remains meaningful, and does not change if we replace the coloured diagram γ by the diagram $\bar\gamma$ we get by omitting the colours of its edges.) The quantity $\bar k(\bar\gamma)$ equals the number of such vertices of $\bar\gamma$ from the first row from which no edge starts plus the number of vertices in the second row, and the notation $\sum'^{(n)}$ means that summation is taken only for such diagrams $\bar\gamma \in \bar\Gamma$ for which $n \geq \bar k(\bar\gamma)$.

Let the set $V_1 = V_1(\bar\gamma)$ consist of those vertices $(1,u_1) = (1,u_1)_\gamma, \dots, (1,u_{s_1}) = (1,u_{s_1})_\gamma$ of the first row $\{(1,1),\dots,(1,k_1)\}$ of the diagram $\bar\gamma$ from which no edge starts, and let $V_2 = V_2(\bar\gamma)$ contain those vertices $(2,v_1) = (2,v_1)_\gamma, \dots, (2,v_{s_1}) = (2,v_{s_2})_\gamma$ from the second row $\{(2,1),\dots,(2,k_2)\}$ of γ from which no edges start. Then $\bar k(\bar\gamma) = s_1 + k_2$, and the function $(\overline{f \circ g})_{\bar\gamma}$ has arguments of the form $x_{(1,u_p)}$, $(1,u_p) \in V_1$ and $x_{(2,v)}$, $1 \le v \le k_2$.

Relation (23) is not appropriate for our goal, since the functions $(\overline{f \circ g})_{\bar\gamma}$ in it may be non-canonical. Hence we apply Hoeffding's decomposition for the U-statistics $I_{n,\bar k(\bar\gamma)}(\overline{f \circ g})_{\bar\gamma}$ in formula (23) to get the desired representation for the product of degenerate U-statistics. Actually some special properties of the function $(\overline{f \circ g})_{\bar\gamma}$ enable us to simplify a little bit this decomposition. (The Hoeffding decomposition is a simple but important result which gives an explicit method to rewrite a general U-statistic in the form of sums of degenerate U-statistics. It has an equivalent reformulation by which an arbitrary (kernel) function of several variables can be rewritten as the sum of canonical functions with different number of variables. It has a concise explanation for instance in the Appendix of (4). In the subsequent considerations I write down what this result yields in the present situation.)

To carry out this procedure let us observe that a function $f(x_{u_1},\dots,u_{u_k})$ is canonical if and only if $P_{u_l} f(x_{u_1},\dots,x_{u_k}) = 0$ with the operator P_{u_l} defined in (16) for all indices u_l. Beside this, the condition that the functions f and g are canonical implies the relations $P_{(1,u)}(\overline{f \circ g})_{\bar\gamma} = 0$ for $(1,u) \in V_1$ and $P_{(2,v)}(\overline{f \circ g})_{\bar\gamma} = 0$ for $(2,v) \in V_2$. Moreover, these relations remain valid if we replace the functions $(\overline{f \circ g})_{\bar\gamma}$ by such functions which we get by applying the product of some transforms $P_{(2,v)}$ and $Q_{(2,v)}$, $(2,v) \in \{(2,1),\dots,(2,k_2)\} \setminus V_2$ for them with the transforms P and Q defined in formulas (16) and (17). (Here we applied such transforms P and Q which are indexed by those vertices of the second row of $\bar\gamma$ from which some edge starts.)

Beside this, the transforms $P_{(2,v)}$ or $Q_{(2,v)}$ are exchangeable with the operators $P_{(2,v')}$ or $Q_{(2,v')}$ if $v \ne v'$, $P_{(2,v)} + Q_{(2,v)} = I$, where I denotes the identity operator, and $P_{(2,v)}Q_{(2,v)} = 0$, since $P_{(2,v)}Q_{(2,v)} = P_{(2,v)} - P^2_{(2,v)} = 0$. The above relations enable us to make the following decomposition of the function $(\overline{f \circ g})_{\bar\gamma}$ to the sum of canonical functions (just as it is done in the Hoeffding decomposition): Let us introduce the class of those coloured diagram $\Gamma(\bar\gamma)$ which we can get by colouring all edges of the diagram γ either with colour 1 or colour -1. Some calculation shows that

$$(\overline{f \circ g})_{\bar\gamma} = \left(\prod_{(2,v)\in\{(2,1),\dots,(2,k_2)\}\setminus V_2} (P_{(2,v)} + Q_{(2,v)}) \right) (\overline{f \circ g})_{\bar\gamma} = \sum_{\gamma\in\Gamma(\bar\gamma)} (f \circ g)_\gamma, \tag{24}$$

where the function $(f \circ g)_\gamma$ is defined in formula (19). We get the right-hand side of relation (24) by carrying out the multiplications for the middle term of this expression, and exploiting the properties of the operators $P_{(2,v)}$ and $Q_{(2,v)}$. Moreover, these properties also imply that the functions $(f \circ g)_\gamma$ are canonical functions of their variables $x_{(1,u)}$, $(1,u) \in V_1$ and $x_{(2,v)}$, $(2,v) \in B_{(b,-1)}(\gamma) \cup V_2$. Indeed, the above properties of the operators $P_{(2,v)}$ and $Q_{(2,v)}$ imply that $P_{(1,u)}(f \circ g)_\gamma = 0$ if $(1,u) \in V_1$, and $P_{(2,v)}(f \circ g)_\gamma = 0$ if $(2,v) \in B_{(b,-1)}(\gamma) \cup V_2$.

Let $Z(\gamma)$ denote the set of edges of colour 1, $W(\gamma)$ the set of edges of colour -1 in the coloured diagram γ, and let $|Z(\gamma)|$ and $W(\gamma)|$ be their cardinality. Then $(f \circ g)_\gamma$ is a (canonical) function with $k(\gamma) = k_1 + k_2 - (|W(\gamma)| + 2|Z(\gamma)|)$ variables, and formula (24) implies the following representation of the U-statistic $I_{n,\bar k(\bar\gamma)}\left(\overline{f\circ g})_{\bar\gamma}\right)$ in the form of a sum of degenerate U-statistics:

$$
\begin{aligned}
&n^{-(k_1+k_2)/2}\bar k(\bar\gamma)!I_{n,\bar k(\bar\gamma)}\left((\overline{f\circ g})_{\bar\gamma}\right)\\
&\qquad=n^{-(k_1+k_2)/2}\sum_{\gamma\in\Gamma(\bar\gamma)}J_n(\gamma)n^{|Z(\gamma)|}k(\gamma)!I_{n,k(\gamma)}\left((f\circ g)_\gamma\right)
\end{aligned}
\tag{25}
$$

with $J_n(\gamma)=1$ if $|Z(\gamma)|=0$, and

$$
J_n(\gamma)=\frac{\prod\limits_{j=1}^{|Z(\gamma)|}(n-(k_1+k_2)+|W(\gamma)|+|Z(\gamma)|+j)}{n^{|Z(\gamma)|}}\quad\text{if }\ |Z(\gamma)|>0.
$$

The coefficient $J_n(\gamma)n^{|Z(\gamma)|}$ appeared in formula (25), since if we apply the decomposition (24) for all terms $(\overline{f\circ g})_{\bar\gamma}(\xi_{j_{(1,u)}},\xi_{j_{(2,v)}},\,(1,u)\in V_1,\,(2,v)\in\{1,\dots k_2\})$ of the U-statistic $\bar k(\bar\gamma)!I_{n,k(\bar\gamma)}\left((\overline{f\circ g})_{\bar\gamma}\right)$, then each term

$$
(f\circ g)_\gamma(\xi_{j_{(1,u)}},\xi_{j_{(2,v)}},\,(1,u)\in V_1,\,(2,v)\in V_2\cup V_1)
$$

of the U-statistic $I_{n,k(\gamma)}\left((f\circ g)_\gamma\right)$ appears $J_n(\gamma)n^{|Z(\gamma)|}$ times. (This is so, because $\bar k(\gamma)=k_1+k_2-(|W(\gamma)|+2|Z(\gamma)|)$ variables are fixed in the term $(f\circ g)_\gamma$ from the $k(\bar\gamma)=k_1+k_2-(|W(\gamma)|+|Z(\gamma)|)$ variables in the term $(\overline{f\circ g})_{\bar\gamma}$, and to get formula (25) from formula (24) the indices of the remaining $|Z(\gamma)|$ variables can be freely chosen from the indices $1,\dots,n$, with the only restriction that all indices must be different.)

Formula (20) follows from relations (23) and (25). To see that we wrote the right power of n in this formula observe that

$$
n^{-(k_1+k_2)/2}n^{|Z(\gamma)|}=n^{-k(\gamma)/2}n^{-|W(\gamma)|/2}.
$$

To prove inequality (21) in the case $|W(\gamma)|=0$ let us estimate first the value of the function $(f\circ g)^2_\gamma(x_{(1,u)},x_{(2,v)},\ (1,u)\in V_1,\ (2,v)\in V_2)$ by means of the Schwarz inequality. We get that

$$
\begin{aligned}
&(f\circ g)^2_\gamma(x_{(1,u)},x_{(2,v)},\ (1,u)\in V_1,\ (2,v)\in V_2)\\
&\quad\le\int f^2(x_{(1,u)},x_{(2,v)},\ (1,u)\in V_1,\ (2,v)\in B_{(b,1)}(\gamma))\prod_{(2,v)\in B_{(b,1)}(\gamma)}\mu(\,dx_{(2,v)})\\
&\qquad\qquad\int g^2(x_{(2,v)},\ (2,v)\in V_2\cup B_{(b,1)}(\gamma),)\prod_{(2,v)\in B_{(b,1)}(\gamma)}\mu(\,dx_{(2,v)})\\
&\quad=\prod_{(2,v)\in B_{(b,1)}(\gamma)}P_{(2,v)}f^2(x_{(1,u)},x_{(2,v)},\ (1,u)\in V_1,\ (2,v)\in B_{(b,1)}(\gamma))\\
&\qquad\qquad\prod_{(2,v)\in B_{(b,1)}(\gamma)}P_{(2,v)}g^2(x_{(2,v)},\ (2,v)\in V_2\cup B_{(b,1)}(\gamma))
\end{aligned}
\tag{26}
$$

with the operators P defined in formula (16).

Let us observe that the two functions at the right-hand side of (26) are functions of different arguments. The first of them depends on the arguments $x_{(1,u)}$, $(1,u)\in V_1$, the second one on the arguments $x_{(2,v)}$, $(2,v)\in V_2$. Beside this, as the operators P appearing in their definition are contraction in L_1-norm, these functions are bounded in L_1 norm by $\|f\|_2^2$ and $\|g\|_2^2$ respectively. Because of the above relations we get formula (21) by integrating inequality (26) and applying Fubini's theorem.

To prove inequality (22) let us introduce, similarly to formula (17), the operators

$$\tilde Q_{u_j} h(x_{u_1},\dots,x_{u_r}) = h(x_{u_1},\dots,x_{u_r}) + \int h(x_{u_1},\dots,x_{u_r})\mu(\,dx_{u_j}), \quad 1\le j\le r,$$

in the space of functions $h(x_{u_1},\dots,x_{u_r})$ with coordinates in the space $(X,\mathcal{X})$. (The indices $u_1,\dots,u_r$ are all different.) Observe that both the operators $\tilde Q_{u_j}$ and the operators P_{u_j} defined in (16) are positive, i.e. these operators map a non-negative function to a non-negative function. Beside this, $Q_{u_j}\le \tilde Q_{u_j}$, i.e. $\tilde Q_{u_j}-Q_{u_j}$ is a non-negative operator, and the norms of the operators $\frac{\tilde Q_{u_j}}{2}$ and P_{u_j} are bounded by 1 both in the $L_1(\mu)$, the $L_2(\mu)$ and the supremum norm.

Let us define the function

$$\begin{aligned}
&\widetilde{(f\circ g)}_\gamma\left(x_{(1,j)},x_{(2,j')},\,j\in\{1,\dots,k_1\}\setminus B_u(\gamma),\,j'\in\{1,\dots,k_2\}\setminus B_{(b,1)}\right)\\
&\quad=\prod_{(2,j')\in B_{(b,1)}(\gamma)} P_{(2,j')} \prod_{(2,j')\in B_{(b,-1)}(\gamma)} \tilde Q_{(2,j')}\\
&\qquad\qquad \overline{(f\circ g)}_\gamma\left(x_{(j,1)},x_{(j',2)},\; j\in\{1,\dots,k_1\}\setminus B_u(\gamma),\,1\le j'\le k_2\right)
\end{aligned}$$

with the notation of Section 3. We have defined the function $\widetilde{(f\circ g)}_\gamma$ with the help of $\overline{(f\circ g)}_\gamma$ similarly to the definition of $(f\circ g)_\gamma$ in (19), only we have replaced the operators $Q_{(2,j')}$ by $\tilde Q_{(2,j')}$ in it.

We may assume that $\|g\|_2\le\|f\|_2$. We can write because of the properties of the operators P_{u_j} and $\tilde Q_{u_j}$ listed above and the condition $\sup|f(x_1,\dots,x_k)|\le1$ that

$$|(f\circ g)_\gamma| \le \widetilde{(|f|\circ|g|)}_\gamma \le \widetilde{(1\circ|g|)}_\gamma, \tag{27}$$

where '$\le$' means that the function at the right-hand side is greater than or equal to the function at the left-hand side in all points, and 1 denotes the function which equals identically 1. Because of relation (27) to prove relation (22) it is enough to show that

$$\begin{aligned}
\|\widetilde{(1\circ|g|)}_\gamma\|_2 &= \left\|\prod_{(2,j)\in B_{(b,1)}(\gamma)} P_{(2,j)} \prod_{(2,j)\in B_{(b,-1)}(\gamma)} \tilde Q_{(2,j)}\,|g(x_{(2,1)},\dots,x_{(2,k_2)})|\right\|_2\\
&\le 2^{|W(\gamma)|}\|g\|_2.
\end{aligned}\tag{28}$$

But this inequality trivially holds, since the norm of all operators $P_{(2,j)}$ in formula (28) is bounded by 1, the norm of all operators $\tilde Q_{(2,j)}$ is bounded by 2 in the $L_2(\mu)$ norm, and $|B_{(b,-1)}|=|W(\gamma)|$.

5 The diagram formula for the product of several degenerate U-statistics.

The product of more than two degenerate U-statistics can also be expressed in the form of a sum of degenerate U-statistics by means of a recursive application of Theorem A. We shall present this result in Theorem B and prove it together with an estimate about the L_2-norm of the kernel functions of the degenerate U-statistics appearing in Theorem B. This estimate will be given in Theorem C. Since the expected value of all degenerate U-statistics of order $k \geq 1$ equals zero, the representation of the product of U-statistics in the form of a sum of degenerate U-statistics implies that the expected value of this product equals the sum of the constant terms in this representation. In such a way we get a formula for the expected value of a product of degenerate U-statistics which together with Theorem C will be sufficient to prove Proposition A. But the formula we get in this way is more complicated than the analogous diagram formula for products of Wiener–Itô integrals. To overcome this difficulty it is useful to work out a good "book-keeping procedure".

Let us have a sequence of iid. random variables $\xi_1, \xi_2, \dots$ taking values on a measurable space $(X, \mathcal{X})$ with some distribution μ, and consider L functions $f_l(x_1, \dots, x_{k_l})$ on the measure spaces $(X^{k_l}, \mathcal{X}^{k_l})$, $1 \leq l \leq L$, canonical with respect to the measure μ. We want to represent the product of $L \geq 2$ normalized degenerate U-statistics $n^{-k_l/2} k_l! I_{n,k_l}(f_{k_l})$ in the form of a sum of degenerate U-statistics similarly to Theorem A. For this goal I define a class of coloured diagrams $\Gamma(k_1, \dots, k_L)$ together with some canonical functions $F_\gamma = F_\gamma(f_{k_1}, \dots, f_{k_L})$ depending on the diagrams $\gamma \in \Gamma(k_1, \dots, k_L)$ and the functions $f_l(x_1, \dots, x_{k_l})$, $1 \leq l \leq L$.

The coloured diagrams will be graphs with vertices (l, j) and (l, j, C), $1 \leq l \leq L$, $1 \leq j \leq k_l$, and edges between some of these vertices which will get either colour 1 or colour -1. The set of vertices $\{(l, j), (l, j, C), 1 \leq j \leq k_l\}$ will be called the l-th row of the diagrams. (The vertices (l, j, C) are introduced, because it turned out to be useful to take a copy (l, j, C) of some vertices (l, j). The letter C was chosen to indicate that it is a copy.) From all vertices there starts either zero or one edge, and edges may connect only vertices in different rows. We shall call all vertices of the form (l, j) permissible, and beside this some of the vertices (l, j, C) will also be called permissible. Those vertices will be called permissible from which some edge may start.

We shall say that an edge connecting two vertices (l_1, j_1) with (l_2, j_2) or (a permissible) vertex (l_1, j_1, C) with another vertex (l_2, j_2) such that $l_2 > l_1$ is of level l_2, and (l_2, j) will be called the lower end-point of such an edge. (The coloured diagrams we shall define contain only edges with lower end-points of the form (l, j).) We shall call the restriction $\gamma(l)$ of the diagram γ to level l that part of a diagram γ which contains all of its vertices together with those edges (together with their colours) whose levels are less than or equal to l, and tells which of the vertices (l', j, C) are permissible for $1 \leq l' \leq l$. We shall define the diagrams $\gamma \in \Gamma(k_1, \dots, k_L)$ inductively by defining their restrictions $\gamma(l)$ to level l for all $l = 1, 2, \dots, L$. Those diagrams γ will belong to $\Gamma(k_1, \dots, k_L)$ whose restrictions $\gamma(l)$ can be defined through the following procedure for all $l = 1, 2, \dots, L$.

The restriction $\gamma(1)$ of a diagram γ to level 1 contains no edges, and no vertex of the form $(1, j, C)$, $1 \leq j \leq k_1$, is permissible. If we have defined the restrictions $\gamma(l-1)$ for some $2 \leq l \leq L$, then those diagrams will be called restrictions $\gamma(l)$ at level l which can be obtained from a restriction $\gamma(l-1)$ in the following way: Take the vertices (l, j), $1 \leq j \leq k_l$, from the l-th row. From each of these vertices there starts either zero or one edge, and they get either colour 1 or colour -1. The other end-point of these edges must be such a vertex (l', j') or a permissible vertex (l', j', C) with some $1 < l' < l$ which is not an end-point of a vertex in $\gamma(l-1)$. We define $\gamma(l)$ first by adjusting the coloured edges constructed in the above way to the (coloured) edges of $\gamma(l-1)$, and then defining the set of permissible vertices in $\gamma(l)$. It contains beside the permissible vertices of $\gamma(l-1)$ and the vertices (l, j), $1 \leq j \leq k_l$, those vertices (l, j, C) for which (l,j) is the lower end-point of an edge with colour -1 in $\gamma(l)$. $\Gamma(k_1,\dots,k_L)$ will consist of all coloured diagrams $\gamma=\gamma(L)$ obtained in such a way.

Given a coloured diagram $\gamma\in\Gamma(k_1,\dots,k_L)$ we shall define recursively some (canonical) functions $F_{l,\gamma}$ with the help of the functions $f_1,\dots,f_l$ for all $1\le l\le L$ in the way suggested by Theorem A. Then we put $F_\gamma=F_{L,\gamma}$ and give the desired representation of the product of the degenerate U-statistics with the help of U-statistics with kernel functions F_γ and constants $J_n(l,\gamma)$, $\gamma\in\Gamma(k_1,\dots,k_L)$, $1\le l\le L$.

Let us fix some coloured diagram $\gamma\in\Gamma(k_1,\dots,k_L)$ and introduce the following notations: Let $B_{(b,-1)}(l,\gamma)$ denote the set of lower end-points of the form (l,j) of edges with colour -1 and $B_{(b,1)}(l,\gamma)$ the set of lower end-points of the form (l,j) with colour 1. Let $U(l,\gamma)$ denote the set of those permissible vertices (l',j) and (l',j,C) with $l'\le l$ from which no edge starts in the restriction $\gamma(l)$ of the diagram γ to level l, i.e. either no edge starts from this vertex, or if some edge starts from it, then its other end-point is a vertex (l',j) with $l'>l$. Beside this, given some integer $1\le l_1<l$ let $U(l,l_1,\gamma)$ denote the restriction of $U(l,\gamma)$ to its first l_1 rows, i.e. $U(l,l_1,\gamma)$ consists of those vertices (l',j) and (l',j,C) which are contained in $U(l,\gamma)$, and $l'\le l_1$. We shall define the functions $F_l(\gamma)$ with arguments of the form $x_{(l',j)}$ and $x_{(l',j,C)}$ with $(l',j)\in U(l,\gamma)$ and $(l',j,C)\in U(l,\gamma)$. For this end put first

$$F_{1,\gamma}(x_{(1,1)},\dots,x_{(k_1,1)})=f_1(x_{(1,1)},\dots,x_{(k_1,1)}). \tag{29}$$

To define the function $F_{l,\gamma}$ for $l\ge2$ first we introduce a function $\alpha_{l,\gamma}(\cdot)$ on the set of vertices in $U(l-1,\gamma)$ in the following way. If a vertex (l',j') or (l',j',C) in $U(\gamma,l-1)$ is such that it is connected to no vertex (l,j), $1\le j\le k_l$, then $\alpha_{l,\gamma}(l',j')=(l',j')$, $\alpha_{l,\gamma}(l',j',C)=(l',j',C)$ and if (l',j') is connected to a vertex (l,j), then $\alpha_{l,\gamma}(l',j')=(l,j)$, if (l',j',C) is connected with a vertex (l,j), then $\alpha_{l,\gamma}(l',j',C)=(l,j)$. We define, similarly to the formula (18) the functions

$$\begin{aligned}&\bar F_{l,\gamma}(x_{(l',j')},x_{(l',j',C)},\ (l',j')\text{ and }(l',j',C)\in U(l,l-1,\gamma),\ x_{(l,j)},\ 1\le j\le k_l)\\&\quad=F_{l-1,\gamma}(x_{\alpha_{l,\gamma}(l',j')},x_{\alpha_{l,\gamma}(l',j',C)},\ (l',j')\text{ and }(l',j',C)\in U(l-1,\gamma))\\&\qquad\qquad f_l(x_{(l,1)},\dots,x_{(l,k_l)}),\end{aligned} \tag{30}$$

i.e. we take the function $F_{l-1,\gamma}\circ f_l$ and replace the arguments of this function indexed by such a vertex of γ which is connected by an edge with a vertex in the l-th row of γ by the argument indexed with the lower end-point of this edge.

Then we define with the help of the operators P_{u_j} and Q_{u_j} introduced in (16) and (17) the functions

$$\begin{aligned}&\bar{\bar F}_{l,\gamma}(x_{(l',j')},x_{(l',j',C)},\ (l',j')\text{ and }(l',j',C)\in U(l,l-1,\gamma),\\&\qquad x_{(l,j)},\ j\in\{1,\dots,k_l\}\setminus B_{(l,1)}(l,\gamma))\\&\quad=\prod_{(l,j)\in B_{(b,1)}(l,\gamma)}P_{(l,j)}\prod_{(l,j)\in B_{(b,-1)}(l,\gamma)}Q_{(l,j)}\bar F_{l,\gamma}(x_{(l',j')},x_{(l',j',C)},\\&\qquad(l',j')\text{ and }(l',j',C)\in U(l,l-1,\gamma),\ x_{(l,j)},\ 1\le j\le k_l),\end{aligned} \tag{31}$$

similarly to the formula (19), i.e. we apply for the function $\bar F_l(\gamma)$ the operators $P_{(l,j)}$ for those indices (l,j) which are the lower end-points of an edge with colour 1 and the operators $Q_{(l,j)}$ for those indices (l,j) which are the lower end-points of an edge with colour -1.

Finally we define the function $F_{l,\gamma}$ simply by reindexing some arguments of the function $\bar{\bar{F}}_{l,\gamma}$ to get a function which is indexed by the vertices in $U(l,\gamma)$. To this end we define the function $A_{l,\gamma}(\cdot)$ on the set of vertices $\{(l,j)\colon\ (l,j)\in\{(l,1),\dots,(l,k_l)\}\setminus B_{(b,1)}(l,\gamma)$ as $A_{l,\gamma}(l,j)=(l,j,C)$ if $(l,j)\in B_{(b,-1)}(l,\gamma)$, and $A_{l,\gamma}(l,j)=(l,j)$ if $(l,j)\in\{(l,1),\dots,(l,k_l)\}\setminus(B_{(b,1)}(l,\gamma)\cup B_{(b,-1)}(l,\gamma))$. Then we put

$$\begin{aligned}
&F_{l,\gamma}(x_{(l',j')},x_{(l',j',C)},\ (l',j')\text{ and }(l',j',C)\in U(l,\gamma))\\
&\qquad=\bar{\bar{F}}_{l,\gamma}(x_{(l',j')},x_{(l',j',C)},\ (l',j')\text{ and }(l',j',C)\in U(l,l-1,\gamma),\\
&\qquad\qquad x_{A_{l,\gamma}(l,j)},\ (l,j)\in\{(l,1),\dots,(l,k_l)\}\setminus B_{(b,1)}(l,\gamma)).
\end{aligned}\tag{32}$$

Now we can formulate the following generalization of Theorem A.

Theorem B. *Let us have a sequence of iid. random variables $\xi_1,\xi_2,\dots$ with some distribution μ on a measurable space $(X,\mathcal{X})$ together with $L\ge 2$ bounded functions $f_l(x_1,\dots,x_{k_l})$ on the spaces $(X^{k_l},\mathcal{X}^{k_l})$, $1\le l\le L$, canonical with respect to the probability measure μ. Let us introduce the class of coloured diagrams $\Gamma(k_1,\dots,k_L)$ defined above together with the functions $F_\gamma=F_{L,\gamma}(f_1,\dots,f_L)$ defined in formulas (29)—(32).*

Put $k(\gamma(l))=\sum\limits_{p=1}^{l}k_p-\sum\limits_{p=2}^{l}(2|B_{(b,1)}(p,\gamma)|+|B_{(b,-1)}(p,\gamma)|)$, where $|B_{(b,1)}(p,\gamma)|$ denotes the number of lower end-points in the p-th row of γ with colour 1 and $|B_{(b,-1)}(p,\gamma)|$ is the number of lower end-points in the p-th row of γ with colour -1, $1\le l\le L$, and define $k(\gamma)=k(\gamma(L))$. Then $k(\gamma(l))$ is the number of variables of the function $F_{l,\gamma}$, $1\le l\le L$.

The functions F_γ are canonical with respect to the measure μ with $k(\gamma)$ variables, and the product of the degenerate U-statistics $I_{n,k_l}(f)$, $n\ge\max\limits_{1\le l\le L}k_l$, defined in (1) satisfies the identity

$$\begin{aligned}
&\prod_{l=1}^{L}k_l!n^{-k_l/2}I_{n,k_l}(f_{k_l})\\
&\qquad=\sum_{\gamma\in\Gamma(k_1,\dots,k_L)}{}^{\prime(n,\ L)}\left(\prod_{l=1}^{L}J_n(l,\gamma)\right)n^{-|W(\gamma)|/2}\cdot k(\gamma)!n^{-k(\gamma)/2}I_{n,k(\gamma)}(F_\gamma),
\end{aligned}\tag{33}$$

where $|W(\gamma)|=\sum\limits_{l=2}^{L}|B_{(b,-1)}(l,\gamma)|$ is the number of edges with colour -1 in the coloured diagram γ, and $\sum^{\prime(n,L)}$ means that summation is taken for those $\gamma\in\Gamma(k_1,\dots,k_L)$ which satisfy the relation $k(\gamma(l-1))+k_l-(|B_{(b,1)}(l,\gamma)|+|B_{(b,-1)}(l,\gamma)|)\le n$ for all $2\le l\le L$. Beside this, the constants $J_n(l,\gamma)$, $1\le l\le L$, in formula (33) are defined by the relations $J_n(1,\gamma)=1$, and

$$J_n(l,\gamma)=\frac{\prod\limits_{j=1}^{|B_{(b,1)}(l,\gamma)|}(n-(k(\gamma(l-1))+k_l)+|B_{(b,-1)}(l,\gamma)|+|B_{(b,1)}(l,\gamma)|+j)}{n^{|B_{(b,1)}(l,\gamma)|}},\tag{34}$$

$2\le l\le L$, if $|B_{(b,1)}(l,\gamma)|\ge 1$, and $J_n(l,\gamma)=1$ if $|B_{(b,1)}(l,\gamma)|=0$, where $|B_{(b,1)}(l,\gamma)|$ and $|B_{(b,-1)}(l,\gamma)|$ denote the number of those edges in γ with colour 1 and with colour -1 respectively whose lower end-points are in the l-th row of γ.

Let $\bar\Gamma(k_1,\dots,k_L)$ denote the class of those coloured diagrams of $\Gamma(k_1,\dots,k_L)$ for which every permissible vertex is the end-point of some edge. A coloured diagram $\gamma\in\Gamma(k_1,\dots,k_L)$ satisfies the relation $\gamma\in\bar\Gamma(k_1,\dots,k_L)$ if and only if $k(\gamma)=0$. In this case F_γ is constant, and $I_{n,k(\gamma)}(F_\gamma)=F_\gamma$. For all other coloured diagrams $\gamma\in\Gamma(k_1,\dots,k_L)$ $k(\gamma)\ge 0$. The identity

$$E\left(\prod_{l=1}^L k_l!n^{-k_l/2}I_{n,k_l}(f_{k_l})\right)=\sum_{\gamma\in\bar\Gamma(k_1,\dots,k_L)}^{\prime(n,\ L)}\left(\prod_{l=1}^L J_n(l,\gamma)\right)n^{-|W(\gamma)|/2}\cdot F_\gamma \tag{35}$$

holds.

Theorem B can be deduced relatively simply from Theorem A by induction with respect to the number L of the functions. Theorem A contains the results of Theorem B in the case $L=2$. A simple induction argument together with the formulas describing the functions $F_{l,\gamma}$ by means of the functions $F_{l-1,\gamma}$ and f_l and Theorem A imply that all functions F_γ in Theorem B are canonical. Finally, an inductive procedure with respect to the number L of the functions f_l shows that relation (33) holds. Indeed, by exploiting that formula (33) holds for the product of the first $L-1$ degenerate U-statistics, then multiplying this identity with the last U-statistic and applying for each term at the right-hand side Theorem A we get that relation (33) also holds for the product L degenerate U-statistics.

A simple inductive procedure with respect to l shows that for all $2\le l\le L$ the diagram $\gamma(l)$ contains $k(\gamma(l))=\sum\limits_{p=1}^{l}k_l-\sum\limits_{p=2}^{l}(2|B_{(b,1)}(p,\gamma)|+|B_{(b,-1)}(p,\gamma)|)$ permissible vertices in its first l rows which are not an end-point of an edge in $\gamma(l)$. In particular, $k(\gamma)=0$ if and only if $\gamma\in\bar\Gamma(k_1,\dots,k_L)$ with the class of coloured diagrams $\bar\Gamma(k_1,\dots,k_L)$ introduced at the end of Theorem B. Since $EI_{n,k}(f)=0$ for all degenerate U-statistics of order $k\ge1$, this property together with relation (33) imply identity (35).

In the proof of Proposition A we shall also need an estimate formulated in Theorem C. It is a simple consequence of inequalities (21) and (22) in Theorem A.

Theorem C. *Let us have L functions $f_l(x_1,\dots,x_{k_l})$ on the spaces $(X^{k_l},\mathcal X^{k_l})$, $1\le l\le L$, which satisfy formulas (3) and (4) (if we replace the index k by index k_l in these formulas), but these functions need not be canonical. Let us take a coloured diagram $\gamma\in\Gamma(k_1,\dots,k_L)$ and consider the function $F_\gamma=F_{L,\gamma}(f_1,\dots,f_L)$ defined by formulas (29)—(33). The L_2-norm of the function F_γ (with respect to a power of the measure μ to the space, where F_γ is defined) satisfies the inequality $\|F_\gamma\|_2\le 2^{|W(\gamma)|}\sigma^{(L-U(\gamma))}$, where $|W(\gamma)|$ denotes the number of edges of colour -1, and $U(\gamma)$ the number of rows which contain a lower vertex of colour -1 in the coloured diagram γ.*

Proof of Theorem C. We shall prove the inequality

$$\|F_{l,\gamma}\|_2\le 2^{|W(l,\gamma)|}\sigma^{(l-U(l,\gamma))}\quad\text{for all } 1\le l\le L, \tag{36}$$

where $|W(l,\gamma)|$ denotes the number of edges with colour 1, and $U(l,\gamma)$ is the number of rows containing a lower point of an edge with colour -1 in the coloured diagram $\gamma(l)$. Formula (36) will be proved by means of induction with respect to l. It implies Theorem C with the choice $l=L$.

Relation (36) clearly holds for $l = 1$. To prove this relation by induction with respect to l for all $1 \le l \le L$ let us first observe that $\sup 2^{-|W(l,\gamma)|}|F_{l,\gamma}| \le 1$ for all $1 \le l \le L$. This relation can be simply checked by induction with respect to l.

If we know relation (36) for $l-1$, then it follows for l from relation (21) if $|B_{(b,-1)}(l,\gamma)| = 0$, that is if there is no edge of colour -1 with lower end-point in the l-th row. Indeed, in this case $\|F_{l,\gamma}(f_1,\dots,f_l)\|_2 \le \|F_{l-1,\gamma}\|_2\|f_l\|_2 \le \|F_{l-1,\gamma}(f_1,\dots,f_{l-1})\|_2 \cdot \sigma$, $|W(l,\gamma)| = |W(l-1,\gamma)|$, and $U(l,\gamma) = U(l-1,\gamma)$. Hence relation (36) holds in this case.

If $|B_{(b,-1)}(l,\gamma)| \ge 1$, then we can apply formula (22) for the expression $\|F_{l,\gamma}\|_2 = \|\bar{\bar{F}}_{l,\gamma}\|_2 = \|(F_{l-1,\gamma} \circ f_l)_{\tilde\gamma(l)}\|_2$, where $\tilde\gamma(l)$ is that coloured diagram with two rows whose first row consists of the indices of the variables of the function $F_{l-1,\gamma}$, its second row consists of the vertices (l,j), $1 \le j \le k_l$, and $\tilde\gamma(l)$ contains the edges of γ between these vertices together with their colour. Then relation (22) implies that

$$\|F_{l,\gamma}\|_2 \le 2^{|B_{(b,-1)}|}\|F_{l-1,\gamma}\|_2 \le 2^{(|W(l-1,\gamma)|+|B_{(b,-1)}(l,\gamma)|)}\sigma^{(l-1-U(l-1,\gamma))}$$

if $|B_{(b,-1)}(l,\gamma)| \ge 1$. Beside this, $|W(l-1,\gamma)|+|B_{(b,-1)}(l,\gamma)| = |W(l,\gamma)|$, and $l-1-U(l-1,\gamma) = l-U(l,\gamma)$ in this case. Hence relation (36) holds in this case, too.

6 The proof of Proposition A.

Proof of Proposition A. We shall prove relation (9) by means of identity (35) and Theorem C with the choice $L = 2M$ and $f_l(x_1,\dots,x_{k_l}) = f(x_1,\dots,x_k)$ for all $1 \le l \le 2M$. We shall partition the class of coloured diagrams $\gamma \in \Gamma(k,M) = \bar\Gamma(\underbrace{k,\dots,k}_{2M \text{ times}})$ with the property that all permissible vertices are the end-points of some edge to classes $\Gamma(k,M,p)$, $1 \le p \le M$, in the following way: $\gamma \in \Gamma(k,M,p)$ for a coloured diagram $\gamma \in \Gamma(k,M)$ if and only if it has $2p$ permissible vertices of the form (l,j,C). (A coloured diagram $\gamma \in \Gamma(k,M)$ has even number of such vertices.) First we prove the following estimate:

There exists some constant $A = A(k) > 0$ and threshold index $M_0 = M_0(k)$ such that for all $M \ge M_0$ and $0 \le p \le kM$ the cardinality $|\Gamma(k,M,p)|$ of the set $\Gamma(k,M,p)$ can be bounded from above by $A2^{2p}\binom{2kM}{2p}\left(\frac{2}{e}\right)^{kM}(kM)^{kM+p}$.

We can bound the number of coloured diagrams in $\Gamma(k,M,p)$ by calculating first the number of choices of the $2p$ permissible vertices from the $2kM$ vertices of the form (l,j,C) which we adjust to the $2kM$ permissible vertices (l,j) and then by calculating the number of such graphs whose vertices are the above chosen permissible vertices, and from all vertices there starts exactly one edge. (Here we allow to connect vertices from the same row. Observe that by defining the set of permissible vertices (l,j,C) in a coloured diagram γ we also determine the colouring of its edges.) Thus we get that $|\Gamma(k,M,p)|$ can be bounded from above by $\binom{2kM}{2p}1\cdot3\cdot5\cdots(2kM+2p-1) = \binom{2kM}{2p}\frac{(2kM+2p)!}{2^{kM+p}(kM+p)!}$. (The appearance of the factor $1\cdot3\cdot5\cdots(2kM+2p-1)$ in this estimate can be explained in a standard way. Let us list the $2kM+2p$ vertices in some order. The first vertex can be connected with $2kM+2p-1$ vertices by an edge. Then the first vertex from which no edge starts can be connected with $2kM+2p-3$ vertices. Continuing this procedure we get the above product for the number of possible system of edges between the already fixed vertices.) We can write by the Stirling formula, similarly to the estimation of the right-hand side of formula (10) that $\frac{(2kM+2p)!}{2^{kM+p}(kM+p)!} \le A\left(\frac{2}{e}\right)^{kM+p}(kM+p)^{kM+p}$ with some constant $A > \sqrt{2}$ if $M \ge M_0$ with some $M_0 = M_0(A)$. Since $p \le kM$ we can write $(kM+p)^{kM+p} \le (kM)^{kM}\left(1+\frac{p}{kM}\right)^{kM}(2kM)^p \le (kM)^{kM+p}e^p 2^p$. The above inequalities imply that

$$|\Gamma(k,M,p)| \le A\binom{2kM}{2p}\left(\frac{2}{e}\right)^{kM}(kM)^{kM+p}2^{2p} \quad \text{if } M \ge M_0, \tag{37}$$

as we have claimed.

Observe that for $\gamma \in \Gamma(k,M,p)$ the quantities introduced in the formulation of Theorems B and C satisfy the relations $|W(\gamma)| = 2p$, $|F_\gamma| = \|F_\gamma\|_2$ and $U(\gamma) \le |W(\gamma)| = 2p$. Hence by Theorem C we have $n^{-|W(\gamma)|/2}|F_\gamma| \le 2^p n^{-p}\sigma^{2M-U(\gamma)} \le 2^p\left(n\sigma^2\right)^{-p}\sigma^{2M} \le \eta^p 2^p (kM)^{-p}\sigma^{2M}$ if $kM \le \eta n\sigma^2$ and $\sigma^2 \le 1$.

This estimate together with relation (35) and the fact that the constants $J_n(l,\gamma)$ defined in (34) are bounded by 1 imply that for $kM \le \eta n\sigma^2$

$$\begin{aligned} E\left(n^{-k/2}k!I_{n,k}(f_k)\right)^{2M} &\le \sum_{\gamma\in\Gamma(k,M)} n^{-|W(\gamma)|/2}\cdot|F_\gamma| \\ &\le \sum_{p=0}^{kM}|\Gamma(k,M,p)|\eta^p 2^p (kM)^{-p}\sigma^{2M}. \end{aligned}$$

Hence by formula (37)

$$\begin{aligned} E\left(n^{-k/2}k!I_{n,k}(f_k)\right)^{2M} &\le A\left(\frac{2}{e}\right)^{kM}(kM)^{kM}\sigma^{2M}\sum_{p=0}^{kM}\binom{2kM}{2p}\left(2\sqrt{2\eta}\right)^{2p} \\ &\le A\left(\frac{2}{e}\right)^{kM}(kM)^{kM}\sigma^{2M}\left(1+2\sqrt{2\eta}\right)^{2kM} \end{aligned}$$

if $kM \le \eta n\sigma^2$. Thus we have proved Proposition A with $C = 2\sqrt{2}$.

References

[1] Adamczak, R. (2006) Moment inequalities for U-statistics. *Annals of Probability* **34** 2288–2314

[2] Arcones, M. A. and Giné, E. (1993) Limit theorems for U-processes. *Annals of Probability* **21**, 1494–1542

[3] Dudley, R. M. (1998) *Uniform Central Limit Theorems.* Cambridge University Press, Cambridge U.K.

[4] Dynkin, E. B. and Mandelbaum, A. (1983) Symmetric statistics, Poisson processes and multiple Wiener integrals. *Annals of Statistics* **11**, 739–745

[5] Giné, E., Latała, R. and Zinn, J. (2000) Exponential and moment inequalities for U-statistics in *High dimensional probability II.* Progress in Probability 47. 13–38. Birkhäuser Boston, Boston, MA.

[6] Itô, K. (1951) Multiple Wiener integral. *J. Math. Soc. Japan* **3**. 157–164

[7] Latała, R. (2006) Estimates of moments and tails of Gaussian chaoses. *Annals of Probability* **34** 2315–2331

[8] Major, P. (1981) Multiple Wiener–Itô integrals. *Lecture Notes in Mathematics* **849**, Springer Verlag, Berlin Heidelberg, New York,

[9] Major, P. (2005) An estimate about multiple stochastic integrals with respect to a normalized empirical measure. *Studia Scientarum Mathematicarum Hungarica.* **42** 3. 295–341

[10] Major, P. (2005) Tail behaviour of random multiple random integrals and U-statistics. Probability Reviews, Vol. 2, 448–505

[11] Major, P. (2006) A multivariate generalization of Hoeffding's inequality. *Electron. Comm. Probab.* **2** 220–229