Elect. Comm. in Probab. **11** (2006), 220–229

ELECTRONIC COMMUNICATIONS in PROBABILITY

A MULTIVARIATE VERSION OF HOEFFDING'S INEQUALITY

PÉTER MAJOR[1]
Alfréd Rényi Mathematical Institute of the Hungarian Academy of Sciences
email: major@renyi.hu

Submitted 10 March 2006, *accepted in final form* 14 September 2006

AMS 2000 Subject classification: Primary 60E15, Secondary 60H05.
Keywords: Hoeffding's inequality, Borell's inequality, multiple Wiener–Itô integrals, diagram formula.

Abstract
In this paper a multivariate version of Hoeffding's inequality is proved about the tail distribution of homogeneous polynomials of Rademacher functions with an optimal constant in the exponent of the upper bound. The proof is based on an estimate about the moments of homogeneous polynomials of Rademacher functions which can be considered as an improvement of Borell's inequality in a most important special case.

1 Introduction. Formulation of the main result.

This paper contains a multivariate version of Hoeffding's inequality. Hoeffding's inequality has the following form. (see e. g. [3], Proposition 1.3.5.)

Theorem A. (Hoeffding's inequality). *Let $\varepsilon_1,\dots,\varepsilon_n$ be independent random variables, $P(\varepsilon_j=1)=P(\varepsilon_j=-1)=\frac12$, $1\le j\le n$, and let $a_1,\dots,a_n$ be arbitrary real numbers. Put $Z=\sum\limits_{j=1}^n a_j\varepsilon_j$ and $V^2=\sum\limits_{j=1}^n a_j^2$. Then*

$$P(Z>u)\le \exp\left\{-\frac{u^2}{2V^2}\right\}\quad \textit{for all } u>0. \tag{1}$$

To formulate its multivariate version first some notations will be introduced.
Let us fix two positive integers k and n, $n\ge k$, and some real numbers $a(j_1,\dots,j_k)$ for all sets of arguments $\{j_1,\dots,j_k\}$ such that $1\le j_l\le n$, $1\le l\le k$, and $j_l\ne j_{l'}$ if $l\ne l'$ in such a way that the numbers $a(j_1,\dots,j_k)$ are symmetric functions of their arguments, i.e. $a(j_1,\dots,j_k)=a(j_{\pi(1)},\dots,j_{\pi(k)})$ for all permutations $\pi\in\Pi_k$ of the set $\{1,\dots,k\}$.

[1]RESEARCH SUPPORTED BY OTKA NR. T061052

Let us define with the help of the above real numbers and a sequence of independent random variables $\varepsilon_1,\dots,\varepsilon_n$, $P(\varepsilon_j=1)=P(\varepsilon_j=-1)=\frac12$, $1\le j\le n$, the random variable

$$Z=\sum_{\substack{(j_1,\dots,j_k)\colon\ 1\le j_l\le n \text{ for all } 1\le l\le k\\ j_l\neq j_{l'} \text{ if } l\neq l'}} a(j_1,\dots,j_k)\varepsilon_{j_1}\cdots\varepsilon_{j_k} \tag{2}$$

and the number

$$V^2=\sum_{\substack{(j_1,\dots,j_k)\colon\ 1\le j_l\le n \text{ for all } 1\le l\le k\\ j_l\neq j_{l'} \text{ if } l\neq l'}} a^2(j_1,\dots,j_k). \tag{3}$$

The following result will be proved.

Theorem 1. (The multivariate version of Hoeffding's inequality). *The random variable Z defined in formula (2) satisfies the inequality*

$$P(|Z|>u)\le A\exp\left\{-\frac12\left(\frac{u}{V}\right)^{2/k}\right\}\quad \textit{for all } u\ge 0 \tag{4}$$

with the constant V defined in (3) and some constants $A>0$ depending only on the parameter k in the expression Z.

I make some comments about this result.
The condition that the coefficients $a(j_1,\dots,j_k)$ are symmetric functions of their variables does not mean a real restriction, since by replacing all coefficients $a(j_1,\dots,j_k)$ by $a_{\rm Sym}(j_1,\dots,j_k)=\frac{1}{k!}\sum\limits_{\pi\in\Pi_k} a(j_{\pi(1)},\dots,j_{\pi(k)})$ in formula (2), where Π_k denotes the set of all permutations of the set $\{1,\dots,k\}$ we do not change the random variable Z. Beside this, the above symmetrization of the coefficients in formula (2) decreases the number V introduced in formula (3).
The identities $EZ=0$, $EZ^2=k!V^2$ hold. Thus Theorem 1 yields an estimate on the tail behaviour of a homogeneous polynomial of order k of independent random variables $\varepsilon_1,\dots,\varepsilon_j$, $P(\varepsilon_j=1)=P(\varepsilon_j=-1)=\frac12$, $1\le j\le n$, with the help of the variance of this polynomial. Such an estimate may be useful in the study of degenerate U-statistics. Thus for instance in paper [10] a weaker form of Theorem 1 played an important role. In Lemma 2 of that paper such a weaker version of the estimate (4) was proved, where the constant $\frac12$ in the exponent at its right-hand side was replaced by the number $\frac{k}{2e(k!)^{1/k}}$. This estimate, which is a fairly simple consequence of Borell's inequality was satisfactory in that paper. (Borell's inequality together with its relation to the problem of this paper will be discussed in Section 3.) However, the question arose whether it can be improved. In particular, I was interested in the question whether such an estimate holds which a comparison with the Gaussian case suggests. In the case $k=1$ it is natural to compare the tail behaviour of Z with that of $V\eta$, where η is a random variable with standard normal distribution. Theorem A gives an estimate suggested by such a comparison.

If Z is a homogeneous random polynomial of order k defined in (2), then it is natural to compare its tail distribution with that of $VH_k(\eta)$, where η has standard normal distribution, and $H_k(\cdot)$ is the k-th Hermite polynomial with leading coefficient 1. Theorem 1 yields an estimate suggested by such a comparison. The next example shows that this estimate is sharp. It also explains, why it is natural to compare the random variable Z with $VH_k(\eta)$.

Define for all $n=k,k+1,\dots$ the random variables $Z=Z_n$ by means of formula (2) with the coefficients

$$a(j_1,\dots,j_k)=a_n(j_1,\dots,j_k)=\frac{V}{\sqrt{n(n-1)\cdots(n-k+1)}}.$$

For the sake of simplicity let us assume that the random variables ε_j, $j=1,\dots,n$, in formula (2) are given in the form $\varepsilon_j=h(\zeta_j)$, $1\le j\le n$, where $\zeta_1,\dots,\zeta_n$ are independent random variables, uniformly distributed in the interval $[0,1]$, and $h(x)=-1$ if $0\le x<\frac12$, $h(x)=1$ if $\frac12\le x\le 1$. (Such a representation of the random variables ε_j is useful for us, because it enables us to apply the subsequent limit theorem about degenerate U-statistics of iid. random variables with non-atomic distribution.)

In this example $\frac{\sqrt{n(n-1)\cdots(n-k+1)}}{k!}Z_n$ are degenerate U-statistics with kernel function

$$f(x_1,\dots,x_k)=h(x_1)\dots h(x_k)$$

and a sequence $\zeta_1,\dots,\zeta_n$ of iid. random variables with uniform distribution on the interval $[0,1]$. $EZ_n^2=k!V^2$, and a limit theorem about degenerate U-statistics (see e.g. [4]) implies that the random variables Z_n converge in distribution to the k-fold Wiener–Itô integral

$$Z^{(0)}=V\int h(x_1)\dots h(x_k)W(\,dx_1)\dots W(dx_k)$$

as $n\to\infty$, where $W(\cdot)$ is a Wiener process on the interval $[0,1]$.

Moreover, the random variable $Z^{(0)}$ has a simpler representation. Namely, by Itô's formula for multiple Wiener–Itô integrals (see e.g. [6]) it can be written in the form $Z^{(0)}=VH_k(\eta)$, where $H_k(\cdot)$ is the k-th Hermite polynomial with leading coefficient 1, and $\eta=\int h(x)W(\,dx)$ is a random variable with standard normal distribution. Simple calculation shows that there are some constants $C>0$ and $D>0$ such that $P(H_k(\eta)>u)\ge Cu^{-1/k}e^{-u^{2/k}/2}$ if $u>D$. (Actually, this estimate is proved in [11].) Hence

$$\lim_{n\to\infty}P(Z_n>u)=P\left(H_k(\eta)>\frac{u}{V}\right)\ge C\left(\frac{V}{u}\right)^{1/k}\exp\left\{-\frac12\left(\frac{u}{V}\right)^{2/k}\right\}\quad\text{if }\frac{u}{V}\ge D$$

with some appropriate constants $C>0$ and $D>0$. This inequality implies that the estimate (4) is essentially sharp. It does not hold with a smaller constant in the exponent at its right-hand side; this upper bound can be improved at least with a pre-exponential factor.

Theorem 1 will be proved in Section 2. It is a fairly simple consequence of a good estimate on the moments of the random variable Z formulated in Theorem 2. These moments will be estimated by means of two lemmas. The first of them, Lemma 1, enables us to bound the moments of Z by those of an appropriate polynomial of independent standard Gaussian random variables. There is a diagram formula to calculate the moments of polynomials of Gaussian random variables. This makes the estimation of the moments of Gausian random variables relatively simple. This is done in Lemma 2. Actually it turned out that it is simpler to rewrite these polynomials in the form of a multiple Wiener–Itô integral and to apply the diagram formula for multiple Wiener–Itô integrals. To make the explanation complete I give a more detailed description of the diagram formula at the end of Section 2. In the final part of this work, in Section 3, I try to explain the background of the proof of Theorem 1 in more detail. In particular, I make some comments about the role of the Gaussian bounding of moments in Lemma 1 and compare the moment estimates obtained by means of the method of this paper with the estimates supplied by Borell's inequality.

2 The proof of Theorem 1.

Theorem 1 will be obtained as a consequence of the following Theorem 2.

Theorem 2. *The random variable Z defined in formula (2) satisfies the inequality*

$$EZ^{2M} \le 1 \cdot 3 \cdot 5 \cdots (2kM-1)V^{2M} \quad \text{for all } M = 1, 2, \dots \tag{5}$$

with the constant V defined in formula (3).

Theorem 2 will be proved with the help of two lemmas. To formulate them, first the following random variable $\bar{Z}$ will be introduced.

$$\bar{Z} = \sum_{\substack{(j_1,\dots,j_k)\colon\ 1\le j_l\le n \text{ for all } 1\le l\le k\\ j_l\ne j_{l'} \text{ if } l\ne l'}} |a(j_1,\dots,j_k)|\eta_{j_1}\cdots\eta_{j_k}, \tag{6}$$

where $\eta_1, \dots, \eta_n$ are iid. random variables with standard normal distribution, and the numbers $a(j_1,\dots,j_k)$ agree with those in formula (2). Now we state

Lemma 1. *The random variables Z and $\bar{Z}$ defined in formulas (2) and (6) satisfy the inequality*

$$EZ^{2M} \le E\bar{Z}^{2M} \quad \text{for all } M = 1, 2, \dots. \tag{7}$$

and

Lemma 2. *The random variable $\bar{Z}$ defined in formula (6) satisfies the inequality*

$$E\bar{Z}^{2M} \le 1 \cdot 3 \cdot 5 \cdots (2kM-1)V^{2M} \quad \text{for all } M = 1, 2, \dots \tag{8}$$

with the constant V defined in formula (3).

Theorem 2 is a straightforward consequence of Lemmas 1 and 2. So to get this result it is enough to prove Lemmas 1 and 2.

Proof of Lemma 1. We can write, by carrying out the multiplications in the expressions EZ^{2M} and $E\bar{Z}^{2M}$, by exploiting the additive and multiplicative properties of the expectation for sums and products of independent random variables together with the identities $E\varepsilon_j^{2p+1} = 0$ and $E\eta_j^{2p+1} = 0$ for all $p = 0, 1, \dots$ that

$$EZ^{2M} = \sum_{\substack{(j_1,\dots,j_l,\,m_1,\dots,m_l)\colon\\ 1\le j_s\le n,\ m_s\ge 1 \text{ for all } 1\le s\le l \text{ with some } k\le l\le kM,\\ m_1+\cdots+m_l=kM}} A(j_1,\dots,j_l,m_1,\dots,m_l)E\varepsilon_{j_1}^{2m_1}\cdots E\varepsilon_{j_l}^{2m_l} \tag{9}$$

and

$$E\bar{Z}^{2M} = \sum_{\substack{(j_1,\dots,j_l,\,m_1,\dots,m_l)\colon\\ 1\le j_s\le n,\ m_s\ge 1 \text{ for all } 1\le s\le l \text{ with some } k\le l\le kM,\\ m_1+\cdots+m_l=kM}} B(j_1,\dots,j_l,m_1,\dots,m_l)E\eta_{j_1}^{2m_1}\cdots E\eta_{j_l}^{2m_l} \tag{10}$$

with some coefficients $A(j_1,\dots,j_l,m_1,\dots,m_l)$ and $B(j_1,\dots,j_l,m_1,\dots,m_l)$ such that

$$|A(j_1,\dots,j_l,m_1,\dots,m_l)| \le B(j_1,\dots,j_l,m_1,\dots,m_l). \tag{11}$$

The coefficients $A(\cdot,\cdot,\cdot)$ and $B(\cdot,\cdot,\cdot)$ could have been expressed in an explicit form, but we do not need such a formula. What is important for us is that $A(\cdot,\cdot,\cdot)$ can be expressed as the sum of certain terms, and $B(\cdot,\cdot,\cdot)$ as the sum of the absolute value of the same terms, hence relation (11) holds. (There may be such indices $(j_1,\dots,j_l,m_1,\dots,m_l)$ for which the sum defining $A(\cdot,\cdot,\cdot)$ and $B(\cdot,\cdot,\cdot)$ with these indices is empty. The value of an empty sum will be defined as zero. As empty sums appear for some index in (9) and (10) simultaneously, their appearance causes no problem.) Since $E\varepsilon_j^{2m}\le E\eta_j^{2m}$ for all parameters j and m, formulas (9), (10) and (11) imply Lemma 1.

Proof of Lemma 2. I found simpler to construct an appropriate multiple Wiener–Itô integral $\tilde Z$ whose distribution agrees with that of the random variable $\bar Z$ defined in (6) and to estimate its moment. To do this, let us consider a white noise $W(\cdot)$ on the unit interval $[0,1]$, i.e. let us take a set of (jointly) Gaussian random variables $W(A)$ indexed by the measurable sets $A\subset[0,1]$ such that $EW(A)=0$, $EW(A)W(B)=\lambda(A\cap B)$ with the Lebesgue measure λ on the real line. (We also need the relation $W(A\cup B)=W(A)+W(B)$ with probability 1 if $A\cap B=\emptyset$, but this relation is the consequence of the previous ones. Indeed, they yield that $E(W(A\cup B)-W(A)-W(B))^2=0$ if $A\cap B=\emptyset$, and this implies the desired identity.) Let us introduce the random variables $\eta_j=n^{1/2}W\left(\left[\frac{j-1}{n},\frac{j}{n}\right)\right)$, $1\le j\le n$, together with the function $f(t_1,\dots,t_k)$, with arguments $0\le t_s<1$ for all indices $1\le s\le k$, defined as

$$f(t_1,\dots,t_k)=\begin{cases} n^{k/2}|a(j_1,\dots,j_k)| \quad \text{if } t_s\in\left[\frac{j_s-1}{n},\frac{j_s}{n}\right),\ \text{and } j_s\ne j_{s'} \text{ for } s\ne s', \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad 1\le j_s\le n,\ 1\le s\le k \\ 0 \quad \text{if } t_s\in\left[\frac{j_s-1}{n},\frac{j_s}{n}\right),\ \text{and } j_s=j_{s'} \text{ for some } s\ne s', \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad 1\le j_s\le n,\ 1\le s\le k \end{cases}$$

and the k-fold Wiener–Itô integral

$$\tilde Z=\int f(t_1,\dots,t_k)W(\,dt_1)\dots W(\,dt_k) \tag{12}$$

of this (elementary) function f. (For the definition of Wiener–Itô integrals see e.g. [6] or [8].) Observe that the above defined random variables $\eta_1,\dots,\eta_n$ are independent with standard normal distribution. Hence the definition of the Wiener–Itô integral of elementary functions and the definition of the function f imply that the distributions of the random integral $\tilde Z$ and of the random variable $\bar Z$ introduced in (6) agree. Beside this, the identity

$$\int f^2(t_1,\dots,t_k)\,dt_1\dots\,dt_k=V^2 \tag{13}$$

also holds with the number V defined in formula (3). Since the distribution of the random variables $\bar Z$ and $\tilde Z$ agree, formulas (12), (13) together with the following estimate about the moments of Wiener–Itô integrals complete the proof of Lemma 2.

In this estimate a function f of k variables and a σ-finite measure μ on some measurable space $(X,\mathcal{X})$ are considered which satisfy the inequality

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

with some $\sigma^2<\infty$. The moments of the k-fold Wiener–Itô integral

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

of the function f with respect to a white-noise μ_W with reference measure μ satisfy the inequality

$$E\left(k!J_{\mu,k}(f)\right)^{2M} \le 1\cdot 3\cdots(2kM-1)\sigma^{2M} \tag{14}$$

for all $M=1,2,\dots$. This result can be got relatively simply from the diagram formula for the product of Wiener–Itô integrals, and it is actually proven in Proposition A of paper [11]. It can be obtained as a straightforward consequence of the results in Lemma 7.31 and Theorem 7.33 of the book [7]. For the sake of completeness I explain this result at the end of this section.

After the proof of Theorem 2 with the help of the diagram formula it remained to derive Theorem 1 from it.

Proof of Theorem 1. By the Stirling formula we get from the estimate of Theorem 2 that

$$EZ^{2M} \le \frac{(2kM)!}{2^{kM}(kM)!}V^{2M} \le K\left(\frac{2}{e}\right)^{kM}(kM)^{kM}V^{2M} \tag{15}$$

for any $K>\sqrt{2}$ if $M\ge M_0(K)$. Hence the Markov inequality yields the estimate

$$P(Z>u) \le \frac{EZ^{2M}}{u^{2M}} \le K\left(\frac{2kM}{e}\left(\frac{V}{u}\right)^{2/k}\right)^{kM} \tag{16}$$

for all $K>\sqrt{2}$ if $M\ge M_0(K)$. Put $k\bar{M}=k\bar{M}(u)=\frac{1}{2}\left(\frac{u}{V}\right)^{2/k}$, and $M=M(u)=[\bar{M}]$, where $[x]$ denotes the integer part of the number x. Let us choose the number u_0 as the solution of the identity $\frac{1}{2k}\left(\frac{u_0}{V}\right)^{2/k}=M_0(K)+1$. Formula (16) can be applied with $M=M(u)$ for $u\ge u_0$, and it yields that

$$P(Z>u)\le Ke^{-kM}\le Ke^{k}e^{-k\bar{M}}=Ke^{k}\exp\left\{-\frac{1}{2}\left(\frac{u}{V}\right)^{2/k}\right\} \qquad \text{if } u\ge u_0. \tag{17}$$

Formula (17) means that relation (4) holds for $u\ge u_0$ with the constant $A=Ke^k$. Hence relation (4) holds with a sufficiently large constant $A>0$ for all $u\ge 0$.

Estimation of the moments of a Wiener–Itô integral by means of the diagram formula.

Let us have m real-valued functions $f_j(x_1,\dots,x_{k_j})$, $1\le j\le m$, on a measurable space $(X,\mathcal{X},\mu)$ with some σ-finite non-atomic measure μ such that

$$\int f_j^2(x_1,\dots,x_{k_j})\mu(\,dx_1)\dots\mu(\,dx_{k_j})<\infty \qquad \text{for all } 1\le j\le m. \tag{18}$$

A white noise μ_W with reference measure μ can be introduced on $(X,\mathcal{X})$. It is an ensemble of jointly Gaussian random variables $\mu_W(A)$ indexed by the measurable sets $A\in\mathcal{X}$ such that $\mu(A)<\infty$ with the property $E\mu_W(A)=0$ and $E\mu_W(A)\mu_W(B)=\mu(A\cap B)$. Also the Wiener–Itô integrals

$$k_j!J_{\mu,k}(f_j)=\int f_j(x_1,\dots,x_{k_j})\mu_W(\,dx_1)\dots\mu_W(\,dx_{k_j})$$

of these functions with respect to the white noise μ_W can be defined if they satisfy relation (18). The definition of these integrals is rather standard, (see e.g [6] or [8]). First they are defined with respect to simple, so-called elementary functions in a natural way, and then they are extended to general functions by means of an L_2-isomorphism. I omit the details.

A most important result in the theory of multiple Wiener–Itô integrals is the so-called diagram formula, which expresses the product of multiple Wiener–Itô integrals in the form of a sum of Wiener–Itô integrals of different order. (The number of variables of the kernel function in a Wiener–Itô integral is called the order of this integral.) The kernel functions of the integrals in the sum representation of a product of Wiener–Itô integrals are defined by means of diagrams. This is the reason for the name 'diagram formula'.

All Wiener–Itô integrals of order $k \geq 1$ have expectation zero, hence if the product of Wiener–Itô integrals is written in the form of a sum of Wiener–Itô integrals, then the expectation of the product can be calculated as the sum of the constant terms in its sum representation. In the present paper only this consequence of the diagram formula will be needed, hence only this result will be described.

This result will be formulated by means of the notion of (closed) diagrams. The class of closed diagrams will be denoted by $\Gamma = \Gamma(k_1, \dots, k_m)$. A diagram $\gamma \in \Gamma(k_1, \dots, k_m)$ consists of vertices of the form (j, l), $1 \leq j \leq m$, $1 \leq l \leq k_j$, and edges $((j, l), (j', l'))$, $1 \leq j, j' \leq m$, $1 \leq l \leq k_j$, $1 \leq l' \leq k'_j$. The set of vertices of the form (j, l) with a fixed number j is called the j-th row of the diagram. All edges $((j, l), (j', l'))$ of a diagram $\gamma \in \Gamma$ connect vertices from different rows, i.e. $j \neq j'$. It is also demanded that from all vertices of a diagram γ there starts exactly one edge. The class $\Gamma(k_1, \dots, k_m)$ of (closed) diagrams contains the diagrams γ with the above properties. If $j < j'$ for an edge $((j, l), (j', l')) \in \gamma$, then (j, l) is called the upper and (j', l') the lower end point of this edge. Let $U(\gamma)$ denote the upper and $L(\gamma)$ the lower end points of a diagram $\gamma \in \Gamma(k_1, \dots, k_m)$. Define the function $\alpha_\gamma(j, l) = (j, l)$ if (j, l) is the upper end point and $\alpha_\gamma(j, l) = (j', l')$ if (j, l) is the lower end point of an edge $((j, l), (j'l'))$ of a diagram $\gamma \in \Gamma(k_1, \dots, k_m)$.

For the sake of simpler notations let us rewrite the functions f_j with reindexed variables in the form $f_j(x_{j,1}, \dots, x_{j,k_j})$, $1 \leq j \leq m$, and define the function

$$F(x_{j,l},\, 1 \leq j \leq m,\, 1 \leq l \leq k_j) = \prod_{j=1}^{m} f_j(x_{j,1}, \dots, x_{j,k_j}).$$

Define with the help of the functions F and α_γ the constants

$$F_\gamma = \int F(x_{\alpha_\gamma(j,l)},\, 1 \leq j \leq m,\, 1 \leq l \leq k_j) \prod_{(j,l) \in U(\gamma)} \mu(\,dx_{j,l}) \tag{19}$$

for all $\gamma \in \Gamma(k_1, \dots, k_m)$.

The expected value of the product of Wiener–Itô integrals $k_j! J_{\mu,k}(f_j)$, $1 \leq j \leq m$, can be expressed with the help of the above quantities F_γ. The following result holds.

Formula about the expected value of products of Wiener–Itô integrals. *Let us consider the Wiener–Itô integrals $k_j! J_{\mu,k}(f_j)$ of some functions f_j, $1 \leq j \leq m$, satisfying relation (18). The expected value of this product satisfies the identity*

$$E\left(\prod_{j=1}^{m} k_j! J_{\mu,k}(f_j) \right) = \sum_{\gamma \in \Gamma(k_1, \dots, k_m)} F_\gamma$$

with the numbers F_γ defined in (19). These numbers satisfy the inequality

$$F_\gamma^2 \le \prod_{j=1}^m \|f_j\|^2 \quad \textit{for all } \gamma \in \Gamma(k_1,\dots,k_m)$$

with the square of the L_2-norm $\|f_j\|^2 = \int f_j^2(x_1,\dots,x_{k_j})\mu(\,dx_1)\dots\mu(\,dx_{k_j})$ of the functions f_j, $1 \le j \le m$.

Let us consider the above result in the special case $m = 2M$ and $f_j = f$ for all $1 \le j \le m$ with a square integrable function f of k variables. Let $\Gamma(k, M)$ denote the class of diagrams $\Gamma(k_1,\dots,k_m)$ in this case, and $|\Gamma(k, M)|$ the number of diagrams it contains. The above result yields the estimate

$$E\,(k!J_{\mu,k}(f))^{2M} \le |\Gamma(k,M)|\|f\|^{2M}. \tag{20}$$

It is not difficult to see that $|\Gamma(k, M)| \le 1\cdot3\cdot5\cdots(2kM-1)$. Indeed, if we omit the restriction that the edges of a diagram can connect only vertices from different rows, then the number of diagrams with $2M$ rows and k vertices in each row equals $1\cdot3\cdot5\cdots(2kM-1)$. Relation (20) together with this observation imply (14).

It is also worth mentioning that the estimate (20) is sharp in the following sense. If

$$f(x_1,\dots,x_k) = f(x_1)\cdots f(x_k)$$

with some square integrable function f, then relation (20) holds with identity. In this case $k!J_{\mu,k}(f)$ equals $\text{const.}\,H_k(\eta)$ with some standard normal random variable η and the k-th Hermite polynomial $H_k(\cdot)$ because of Itô's formula for multiple Wiener–Itô integrals.

3 Some remarks about the results.

The proof of Theorem 1 was based on an estimate of the (high) moments of the homogeneous random polynomial Z of Rademacher functions defined in (2). Although bounds on the tail distribution of sums of independent random variables are generally proved by means of a good estimate on the moment generating function, in the present problem it was more natural to estimate the moments because of the following reason.

As the example discussed in Section 1 shows, if Z is a random polynomial of order k, then the tail distribution $P(Z > u)$ should behave for large numbers u as $e^{-\text{const.}\,u^{-\alpha(k)}}$ with $\alpha(k) = \frac{2}{k}$. In the case $k \ge 3$ a random variable with such a tail distribution has no finite moment generating function. Hence the estimation of the moment generating function does not work in such cases. On the other hand, a good estimate of the (high) moments of the random variable Z is sufficient to prove Theorem 1. It has to be shown that the high moments of Z are not greater than constant times the appropriate moments of a random variable with tail distribution $e^{-\text{const.}\,u^{-\alpha(k)}}$. Here the same constant is in the exponent as in the exponent of the upper bound in Theorem 1.

Theorem 2 contains a good estimate on all even moments of a homogeneous polynomial of Rademacher functions of order k, and it can be considered as a Gaussian type estimate. (It has the same order as the moments of a k-order Hermite polynomial of a standard normal random variable multiplied with a constant.) The moments of degenerate U-statistics were also studied. Proposition B of paper [11] contains a result in this direction. It turned out that high moments of degenerate U-statistics show a worse behaviour. Only their not too high moments satisfy a good 'Gaussian type' estimate. This difference has a deeper cause. There are degenerate U-statistics which have a relatively bad tail behaviour at high levels. Such examples can be found in Example 2.4 for sums of independent random variables and in Example 4.5 for degenerate U-statistics of order 2 in paper [9]. In such cases much worse moment estimates hold than in Theorem 2.

Lemma 1 made possible to reduce the estimation of the moments (and as a consequence the tail of distribution) of a homogeneous polynomial of Rademacher functions to the estimation of the moments of a homogeneous polynomial of Gaussian random variables. This result provided a good tail distribution estimate at all high levels. It can be generalized to other polynomials of independent random variables with good moment behaviour. On the other hand, general U-statistics may have a much worse tail behaviour at high levels than the behaviour suggested by a Gaussian comparison. It would be interesting to get a better understanding about the question when a U-statistic has such a good tail behaviour at all levels which a Gaussian comparison suggests, and when it has a relatively bad tail behaviour at very high level. At any rate, the fact that homogeneous polynomials of Rademacher functions satisfy a good 'Gaussian type' estimate at all levels $u>0$ has an important consequence. This property was needed for the application of an important symmetrization argument in paper [10]. This symmetrization argument made possible to get a good estimate on the supremum of degenerate U-statistics also in such cases when other methods do not work.

There is another result, called Borell's inequality, which makes possible to bound the high moments, and as a consequence the tail distribution of a homogeneous polynomial of Rademacher functions. Actually, this estimate is a simple consequence of the hypercontractive inequality for Rademacher functions proved by A. Bonami [1] and L. Gross [5] independently of each other. It may be interesting to compare the estimates provided by Borell's inequality with those of the present paper.

Borell's inequality, (see e.g. [2]) states the following estimate.

Theorem B. (Borell's inequality). *The moments of the random variable Z defined in formula (2) satisfy the inequality*

$$E|Z|^p \le \left(\frac{p-1}{q-1}\right)^{kp/2} (E|Z|^q)^{p/q} \qquad \textit{if} \quad 1<q\le p<\infty.$$

Let us apply Borell's inequality with the choice $p=2M$ and $q=2$ for the random variable Z defined in (2). It gives the bound $EZ^{2M} \le (2M-1)^{kM}(EZ^2)^M \le A(k)(2M)^{kM}(k!)^M V^{2M}$ with the constant $A(k)=e^{-k/2}$. (The expression in the last part of this inequality is slightly larger than the middle term, but this has no importance in the subsequent consideration.) On the other hand, Theorem 2, more precisely its consequence relation (15), yields the bound $EZ^{2M} \le K(2M)^{kM}\left(\frac{k}{e}\right)^{kM} V^{2M}$ with some appropriate constant $K=K(k)>0$ not depending on M. It can be seen that the inequality $\left(\frac{k}{e}\right)^k < k!$ holds for all integers $k\ge 1$. This means that the estimate of the present paper yields a const. $\cdot\,\alpha^M$-times smaller bound for the moment EZ^{2M} than the estimate given by Borell's inequality, where $\alpha=\frac{1}{k!}\left(\frac{k}{e}\right)^k<1$. As a consequence, Borell's inequality can give the right type of estimate for the tail distribution of the random variable Z, but it cannot give the optimal constant in the exponent. In such large deviation type estimates the moment estimates based on the diagram formula seem to work better.

Acknowledgement. The author would thank the referee for his careful work.

References

[1] Bonami, A. (1970) Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)* **20** 335–402

[2] Borell, C. (1979) On the integrability of Banach space valued Walsh polynomials. *Séminaire de Probabilités XIII, Lecture Notes in Math. 721* 1–3. Springer, Berlin.

[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] Gross, L. (1975) Logarithmic Sobolev inequalities. Amer. J. Math. **97**, 1061–1083

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

[7] Janson, S. (1997) Gaussian Hilbert spaces. Cambridge Tracts in Mathematics, 129. Cambridge University Press, Cambridge

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

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

[10] Major, P. (2006) An estimate on the maximum of a nice class of stochastic integrals. *Probability Theory and Related Fields.*, **134**, 489–537

[11] Major, P. (2006) On a multivariate version of Bernstein's inequality. Submitted to *Journal of Electronic Probability*