

EXPONENTIAL AND MOMENT INEQUALITIES FOR U -STATISTICS

EVARIST GINÉ*, RAFAL LATAŁA† AND JOEL ZINN

ABSTRACT A Bernstein-type exponential inequality for (generalized) canonical U -statistics of order 2 is obtained and the Rosenthal and Hoffmann-Jørgensen inequalities for sums of independent random variables are extended to (generalized) U -statistics of any order whose kernels are either nonnegative or canonical.

1. Introduction

Exponential inequalities, such as Bernstein's and Prohorov's, and moment inequalities, such as Rosenthal's and Hoffmann-Jørgensen's, are among the most basic tools for the analysis of sums of independent random variables. Our object here consists in developing analogues of such inequalities for generalized U -statistics, in particular, for U -statistics and for multilinear forms in independent random variables.

Hoffmann-Jørgensen type moment inequalities for canonical (that is, completely degenerate) U -statistics of any order m were first considered by Giné and Zinn (1992), and their version for U -statistics with nonnegative kernels turned out to be useful for obtaining best possible necessary integrability conditions in limit theorems for U -statistics. (By Khinchin's inequality it is irrelevant whether one considers canonical or nonnegative kernels in moment inequalities, at least if multiplicative constants are not at issue). Klass and Nowicki (1997) also obtained moment inequalities for nonnegative generalized U -statistics, but only for order $m = 2$, and their decomposition of the moments is more complete than that in Giné and Zinn (1992). Ibragimov and Sharakhmetov (1998, 1999) recently obtained analogues of Rosenthal's inequality for nonnegative and for canonical U -statistics. The moment inequalities we present in the first part of this article, valid for canonical and for nonnegative generalized U -statistics of any order m , when specialized to $m = 2$, represent the same level of moment decomposition as the Klass-Nowicki inequalities, coincide with theirs for

*Research partially supported by NSF Grant No. DMS-96-25457.

†Research partially supported by Polish Grant KBN 2 PO3A 043 15.

powers $p > 1$ (except for constants) and are expressed in terms of different, simpler quantities for powers $p < 1$. Proposition 2.1 below, which constitutes the first step towards more elaborate bounds such as those in Theorem 2.3 below, has also been obtained, up to constants, by Ibragimov and Sharakhmetov. Our proofs consist of simple iterations of the classical moment inequalities for sums of independent random variables.

The moment inequalities in the first part of this article do imply exponential bounds for canonical U -statistics of any order and with bounded kernels which are sharper than those in Arcones and Giné (1993); however, they are not of the best kind as they do not exhibit Gaussian behavior for part of the tail, which they should in view of the tail behavior of Gaussian chaos.

In the second part of this article we improve the moment inequalities from the first part in the case of generalized canonical U -statistics of order 2, and for moments of order $p \geq 2$ (Theorem 3.2). The bounds not only involve moments but also the L_2 operator norm of the matrix of kernels. Then we show how these improved moment inequalities imply what we believe is the correct analogue (up to constants) of Bernstein's exponential inequality for generalized canonical U -statistics of order 2 (Theorem 3.3). This exponential inequality, which does exhibit Gaussian behavior for small values of t , is strong enough to imply the law of the iterated logarithm for canonical U -statistics under conditions which are also necessary. The main new ingredient in this part of the paper is Talagrand's (1996) exponential bound for empirical processes, which gives a Rosenthal-Pinelis type inequality for moments of empirical processes (Proposition 3.1) basic for the derivation of the moment inequality for U -statistics of order 2.

Because of the decoupling results of de la Peña and Montgomery-Smith (1995), we can work with decoupled U -statistics, and this allows us to proceed by conditioning and iteration.

2. Moment inequalities

We consider estimation of moments of *generalized decoupled U -statistics*, defined as

$$(2.1) \quad \sum_{1 \leq i_1, \dots, i_m \leq n} h_{i_1, \dots, i_m}(X_{i_1}^{(1)}, \dots, X_{i_m}^{(m)}),$$

where the random variables $X_i^{(j)} : 1 \leq i \leq n, 1 \leq j \leq m, m \leq n$, are independent (not necessarily with the same distribution) and take values in a measurable space (S, \mathcal{S}) , and h_{i_1, \dots, i_m} are real valued measurable functions on S^m . For short, this sum is denoted by $\sum_{\mathbf{i}} h_{\mathbf{i}}$.

Given $J \subseteq \{1, \dots, m\}$ ($J = \emptyset$ is not excluded), and $\mathbf{i} = (i_1, \dots, i_m) \in \{1, \dots, n\}^m$ we set \mathbf{i}_J to be the point of $\{1, \dots, n\}^{|J|}$ obtained from \mathbf{i} by

deleting the coordinates in the places not in J (e.g., if $\mathbf{i} = (3, 4, 2, 1)$ then $\mathbf{i}_{\{1,3\}} = (3, 2)$). Also, $\sum_{\mathbf{i}_J}$ indicates sum over $1 \leq i_j \leq n$, $j \in J$ (for instance, if $m = 4$ and $J = \{1, 3\}$, then

$$\sum_{\mathbf{i}_J} h_{\mathbf{i}} = \sum_{\mathbf{i}_{\{1,3\}}} h_{i_1, i_2, i_3, i_4} = \sum_{1 \leq i_1, i_3 \leq n} h_{i_1, i_2, i_3, i_4}(X_{i_1}^{(1)}, \dots, X_{i_4}^{(4)}).$$

By convention, $\sum_{\mathbf{i}_0} a = a$.

Likewise, while E will denote expected value with respect to all the variables, E_J will denote expected value only with respect to the variables $X_i^{(j)}$ with $j \in J$ and $i \in \{1, \dots, n\}$. By convention, $E_0 a = a$.

Rosenthal's inequality is easiest to extend to U -statistics because it involves only moments of sums (as opposed to moments of maxima and quantiles for Hoffmann-Jørgensen's inequality). So, we will first obtain analogues of Rosenthal's inequality, and then we will transform these inequalities into analogues of Hoffmann-Jørgensen's by first showing that some moments of sums can be replaced by moments of maxima, and then, that the lowest moment can in fact be replaced by a quantile. We will illustrate this three-steps procedure first in the case of nonnegative kernels and moments of order $p \geq 1$. Then we will see that this also solves, via Khinchin's inequality, the case of canonical kernels and moments of order $p \geq 2$. Finally, we will consider the case of moments of order $p < 1$ for positive kernels and $p < 2$ for canonical, cases in which the inequalities are less neat, but still useful. We will pay some attention to the behavior of the constants as $p \rightarrow \infty$ in these inequalities since such behavior translates into (exponential) integrability properties.

2.1. Nonnegative kernels, moments of order $p \geq 1$. For nonnegative independent random variables ξ_i , we have the following two improvements of Rosenthal's inequalities, valid for $p \geq 1$:

1) Latała's, 1997:

$$(R_1) \quad E\left(\sum \xi_i\right)^p \leq (2e)^p \max\left[\frac{e}{p} p^p \sum E \xi_i^p, e^p \left(\sum E \xi_i\right)^p\right], \quad p > 1,$$

(see Pinelis (1994) for the corresponding inequality when the random variables are centered);

2) Johnson, Schechtman and Zinn's, 1985:

$$(R_2) \quad E\left(\sum \xi_i\right)^p \leq K^p \left(\frac{p}{\log p}\right)^p \max\left[\sum E \xi_i^p, \left(\sum E \xi_i\right)^p\right], \quad p > 1,$$

where K is a universal constant. See Utev (1985) and Figiel, Hitczenko, Johnson, Schechtman and Zinn (1997) for more precise inequalities of the same type.

And for general $p > 0$, we have the following improved Hoffmann-Jørgensen inequality, that follows from Kwapien and Woyczyński (1992) and which

can be obtained as in the proof of Theorem 1.2.3 in de la Peña and Giné (1999):

3)

$$(H) \quad E \left\| \sum \xi_i \right\|^p \leq 2^{p-2} \cdot 2^{(p-1)\vee 0} \cdot (p+1)^{p+1} \left[t_0^p + E \max \|\xi_i\|^p \right], \quad p > 0,$$

where

$$t_0 := \inf \left[t > 0 : \Pr \left\{ \left\| \sum \xi_i \right\| > t \right\} \leq \frac{1}{2} \right],$$

and where we write norm for absolute value in order to include not only independent nonnegative real random variables, but also independent nonnegative random functions ξ_i taking values in certain ‘rearrangement invariant spaces’ such as $L_s(\Omega, \Sigma, \mu)$, $0 < s < \infty$, with $\|\xi\| := (\int |\xi|^s d\mu)^{1/(s\vee 1)}$, or $\ell_\infty(L_s)$. Note that, by Markov,

$$t_0 \leq 2^{1/r} \left(E \left\| \sum \xi_i \right\|^r \right)^{1/r},$$

so that, (H) becomes:

4) for $0 < r < p < \infty$,

$$(H_r) \quad E \left\| \sum \xi_i \right\|^p \leq 2^{p-2} \cdot 2^{(p-1)\vee 0} \cdot (p+1)^{p+1} \left[2^{p/r} \left(E \left\| \sum \xi_i \right\|^r \right)^{p/r} + E \max \|\xi_i\|^p \right]$$

Inequalities (H) and (H_r) hold for spaces of functions which are quasi-normed measurable linear spaces whose quasinorm $\|\cdot\|$ has the property that $\|x\| \leq \|y\|$ whenever $0 \leq x \leq y$.

In the following proposition we extend inequalities (R₁) and (R₂) by means of an easy induction.

Proposition 2.1. *Let $m \in \mathbf{N}$, $p > 1$, and, for all $\mathbf{i} \in \{1, \dots, n\}^m$, let $h_{\mathbf{i}}$ be a nonnegative function of m variables whose p -th power is integrable for the law of $\mathbf{X}_{\mathbf{i}} = (X_{i_1}^{(1)}, \dots, X_{i_m}^{(m)})$. Then,*

$$(2.2) \quad \begin{aligned} & \max_{J \subseteq \{1, \dots, m\}} \left[\sum_{\mathbf{i}_J} E_J \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}} \right)^p \right] \leq E \left(\sum_{\mathbf{i}} h_{\mathbf{i}} \right)^p \\ & \leq (2e^2)^{mp} \sum_{J \subseteq \{1, \dots, m\}} \left[p^{|J|p} \sum_{\mathbf{i}_J} E_J \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}} \right)^p \right], \end{aligned}$$

and also, there exists a universal constant $K < \infty$ such that

$$(2.2') \quad E \left(\sum_{\mathbf{i}} h_{\mathbf{i}} \right)^p \leq K^{mp} \left(\frac{p}{\log p} \right)^{mp} \max_{J \subseteq \{1, \dots, m\}} \left[\sum_{\mathbf{i}_J} E_J \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}} \right)^p \right].$$

Proof. The proof of (2.2') with sum over the subsets J instead of maximum differs from that of (2.2) only in the starting point ((R₂) instead of (R₁));

then, replacing sum by maximum simply increases the constant by a factor of 2^m . The left side inequality in (2.2) follows by Hölder since $p \geq 1$. Consider the right hand side inequality. For $m = 1$ this is just inequality (R_1) and we can proceed by induction. Suppose the result holds for $m - 1$. By applying the induction hypothesis to

$$E\left(\sum_{\mathbf{i}} h_{\mathbf{i}}\right)^p = E_m E_{\{1, \dots, m-1\}} \left[\sum_{\mathbf{i}_{\{1, \dots, m-1\}}} \left(\sum_{i_m} h_{\mathbf{i}}\right) \right]^p,$$

we only have to consider the generic term in the decomposition (2.2) for the new kernels $\left(\sum_{i_m} h_{\mathbf{i}}\right)$ with the $X_i^{(m)}$ variables fixed. In other words, letting J_{m-1} be any subset of $\{1, \dots, m-1\}$ and J_{m-1}^c its complement with respect to $\{1, \dots, m-1\}$, we must estimate

$$\begin{aligned} & E_m \sum_{\mathbf{i}_{J_{m-1}}} E_{J_{m-1}} \left(\sum_{\mathbf{i}_{J_{m-1}^c}} E_{J_{m-1}^c} \left(\sum_{i_m} h_{\mathbf{i}} \right) \right)^p \\ &= \sum_{\mathbf{i}_{J_{m-1}}} E_{J_{m-1}} E_m \left(\sum_{i_m} \left(E_{J_{m-1}^c} \sum_{\mathbf{i}_{J_{m-1}^c}} h_{\mathbf{i}} \right) \right)^p. \end{aligned}$$

Rosenthal's inequality (R_1) applied to the kernels $E_{J_{m-1}^c} \sum_{\mathbf{i}_{J_{m-1}^c}} h_{\mathbf{i}}$ with the variables in J_{m-1} fixed, gives

$$\begin{aligned} E_m \left(\sum_{i_m} \left(E_{J_{m-1}^c} \sum_{\mathbf{i}_{J_{m-1}^c}} h_{\mathbf{i}} \right) \right)^p &\leq (2e^2)^p \left[\left(\sum_{i_m, \mathbf{i}_{J_{m-1}^c}} E_m E_{J_{m-1}^c} h_{\mathbf{i}} \right)^p \right. \\ &\quad \left. + p^p \sum_{i_m} E_m \left(E_{J_{m-1}^c} \sum_{\mathbf{i}_{J_{m-1}^c}} h_{\mathbf{i}} \right)^p \right]. \end{aligned}$$

Upon integrating each term with respect to $E_{J_{m-1}}$ and summing over $\mathbf{i}_{J_{m-1}}$, we then obtain

$$\begin{aligned} & E_m \sum_{\mathbf{i}_{J_{m-1}}} E_{J_{m-1}} \left(\sum_{\mathbf{i}_{J_{m-1}^c}} E_{J_{m-1}^c} \left(\sum_{i_m} h_{\mathbf{i}} \right) \right)^p \\ &\leq (2e^2)^p \left[\sum_{\mathbf{i}_{J_{m-1}}} E_{J_{m-1}} \left(\sum_{\mathbf{i}_{J_{m-1}^c \cup \{m\}}} E_{J_{m-1}^c \cup \{m\}} h_{\mathbf{i}} \right)^p \right. \\ &\quad \left. + p^p \sum_{\mathbf{i}_{J_{m-1} \cup \{m\}}} E_{J_{m-1} \cup \{m\}} \left(E_{J_{m-1}^c} \sum_{\mathbf{i}_{J_{m-1}^c}} h_{\mathbf{i}} \right)^p \right]. \end{aligned}$$

Multiplying by $(2e^2)^{(m-1)p} p^{|J_{m-1}|}$, this is the sum of two terms of the form $(2e)^{mp} p^{|J|p} \sum_{\mathbf{i}_J} E_J \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}} \right)^p$ (for $J = J_{m-1}$ and for $J = J_{m-1} \cup \{m\}$), proving the proposition. \square

This proposition solves the problem of estimating, up to constants, the moments of a decoupled U -statistic by ‘computable’ expressions. For instance, if the functions $h_{\mathbf{i}}$ are all equal and if the variables $X_i^{(j)}$ are i.i.d., then the typical term at the right of (2.1) just becomes $n^{|J|+p|J^c|} E_J(E_{J^c} h)^p$, a ‘mixed moment’ of h . For $m = 2$ the right hand side of inequality (2.2) is just:

$$(2.2'') \quad E \left(\sum_{i,j} h_{i,j}(X_i^{(1)}, X_j^{(2)}) \right)^p \leq (2e^2)^{2p} \left[\left(\sum_{i,j} E h_{i,j}(X_i^{(1)}, X_j^{(2)}) \right)^p \right. \\ \left. + p^p \sum_i E_1 \left(\sum_j E_2 h_{i,j}(X_i^{(1)}, X_j^{(2)}) \right)^p \right. \\ \left. + p^p \sum_j E_2 \left(\sum_i E_1 h_{i,j}(X_i^{(1)}, X_j^{(2)}) \right)^p \right. \\ \left. + p^{2p} \sum_{i,j} E h_{i,j}^p(X_i^{(1)}, X_j^{(2)}) \right].$$

We have been careful with the dependence on p of the constants because it is of some interest to obtain constants of the best order as $p \rightarrow \infty$. In fact, (2.2') exhibits constants of the best order as can be seen by taking the product of two independent copies of the example in Johnson, Schechtman and Zinn (1985), Proposition 2.9.

Next we replace the external sums of expected values at the right side of the above inequalities by expectations of maxima without significantly altering the order of the multiplicative constants. If ξ_i are independent nonnegative random variables, then,

$$(2.3) \quad \frac{1}{2} \left[\delta_0^p \vee \sum E \xi_i^p I_{\xi_i > \delta_0} \right] \leq E \max \xi_i^p \leq \delta_0^p + \sum E \xi_i^p I_{\xi_i > \delta_0}, \quad 0 < p < \infty,$$

where

$$(2.4) \quad \delta_0 = \inf \left[t > 0 : \sum \Pr \{ \xi_i > t \} \leq 1 \right]$$

(Giné and Zinn (1983); see also de la Peña and Giné (1999), page 22). The left hand side of (2.3) gives that, for $0 < r < p$ and ξ_i independent,

$$(2.5) \quad \sum E |\xi_i|^p \leq 2E \max |\xi_i|^p + 2 \left(\sum E |\xi_i|^r \right) \left(E \max |\xi_i|^p \right)^{(p-r)/p}$$

(e.g., de la Peña and Giné (1999), page 48). This inequality, applied with $r = 1 < p$, yields

$$(2.6) \quad p^{\alpha p} \sum E |\xi_i|^p \leq 2(1 + p^\alpha) \max \left[p^{\alpha p} E \max |\xi_i|^p, \left(\sum E |\xi_i| \right)^p \right]$$

for all $\alpha \geq 0$. There are similar inequalities for other values of r ; $r = 1$ is adequate for $\xi_i \geq 0$, but $r = 2$ is better for centered variables. If we use

inequality (2.6) in (2.2''), iteratively for the last term, we obtain that, for a universal constant K (easy but cumbersome to compute), $h_{i,j} \geq 0$, $p > 1$,

$$(2.7) \quad E\left(\sum_{i,j} h_{i,j}\right)^p \leq K^p (2e^2)^p p^4 \left[\left(\sum_{i,j} E h_{i,j}\right)^p + p^p E_1 \max_i \left(\sum_j E_2 h_{i,j}\right)^p \right. \\ \left. + p^p E_2 \max_j \left(\sum_i E_1 h_{i,j}\right)^p + p^{2p} E \max_{i,j} h_{i,j}^p \right].$$

Inequality (2.7) was obtained, up to constants, by Klass and Nowicki (1997) (it is their inequality (4.14)). Our proof is different, and it is contained in the proof of the next corollary, which extends inequality (2.7) to any m .

Corollary 2.2. *Under the same hypotheses as in Proposition 2.1, there exist universal constants K_m such that*

$$(2.8) \quad \max_{J \subseteq \{1, \dots, m\}} \left[E_J \max_{\mathbf{i}_J} \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}}\right)^p \right] \leq E\left(\sum_{\mathbf{i}} h_{\mathbf{i}}\right)^p \\ \leq K_m^p \sum_{J \subseteq \{1, \dots, m\}} \left[p^{|J|p} E_J \max_{\mathbf{i}_J} \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}}\right)^p \right],$$

and

$$(2.8') \quad E\left(\sum_{\mathbf{i}} h_{\mathbf{i}}\right)^p \leq K_m^p \left(\frac{p}{\log p}\right)^{mp} \max_{J \subseteq \{1, \dots, m\}} \left[E_J \max_{\mathbf{i}_J} \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}}\right)^p \right].$$

Proof. The left side of (2.8) follows by Hölder. Inequality (2.8') has a proof similar to that of the right hand side of (2.8), and therefore we only prove the latter. We will prove it by induction over m simultaneously with the inequality

$$(2.9) \quad p^{mp} \sum_{\mathbf{i}} E h_{\mathbf{i}}^p \leq \tilde{K}_m^p \sum_{J \subseteq \{1, \dots, m\}} \left[p^{|J|p} E_J \max_{\mathbf{i}_J} \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}}\right)^p \right].$$

Let us first note that the inequalities (2.9) for $1, \dots, m-1$ together with (2.2) imply (2.8). It is therefore enough to show that if (2.8) and (2.9) hold for $1, \dots, m-1$ then (2.8) is satisfied for m . We will follow the notation of the proof of Proposition 2.1. Inequality (2.9) for $m=1$ is just (2.6), and (2.8) for $m=1$ is (H_1) (which also follows from (R_1) and (2.6)). By the induction assumptions we have

$$(2.10) \quad p^{mp} \sum_{\mathbf{i}} E h_{\mathbf{i}}^p = p^p \sum_{i_m} E_m p^{(m-1)p} \sum_{\mathbf{i}_{\{1, \dots, m-1\}}} E_{\{1, \dots, m-1\}} h_{\mathbf{i}}^p \\ \leq \tilde{K}_{m-1}^p \\ \times \sum_{J_{m-1} \subseteq \{1, \dots, m-1\}} \left[p^{(|J_{m-1}|+1)p} E_{J_{m-1}} E_m \sum_{i_m} \max_{\mathbf{i}_{J_{m-1}}} \left(\sum_{\mathbf{i}_{J_{m-1}^c}} E_{J_{m-1}^c} h_{\mathbf{i}}\right)^p \right].$$

Now, by (2.6), for any $J_{m-1} \subset \{1, \dots, m-1\}$ we have

$$\begin{aligned}
& p^{(|J_{m-1}|+1)p} E_{J_{m-1}} E_m \sum_{i_m} \max_{i_{J_{m-1}}} \left(\sum_{i_{J_{m-1}^c}} E_{J_{m-1}^c} h_i \right)^p \\
& \leq 2(1+p) \left[p^{(|J_{m-1}|+1)p} E_{J_{m-1} \cup \{m\}} \max_{i_{J_{m-1} \cup \{m\}}} \left(\sum_{i_{J_{m-1}^c}} E_{J_{m-1}^c} h_i \right)^p \right. \\
(2.11) \quad & \left. + p^{|J_{m-1}|p} E_{J_{m-1}} \left(\sum_{i_m} E_m \max_{i_{J_{m-1}}} \sum_{i_{J_{m-1}^c}} E_{J_{m-1}^c} h_i \right)^p \right].
\end{aligned}$$

To estimate the last term we note that

$$\begin{aligned}
& p^{|J_{m-1}|p} E_{J_{m-1}} \left(\sum_{i_m} E_m \max_{i_{J_{m-1}}} \sum_{i_{J_{m-1}^c}} E_{J_{m-1}^c} h_i \right)^p \\
(2.12) \quad & \leq p^{|J_{m-1}|p} E_{J_{m-1}} \left(\sum_{i} E_{J_{m-1}^c \cup \{m\}} h_i \right)^p \\
& \leq \tilde{K}_{|J_{m-1}|}^p \sum_{J \subset J_{m-1}} p^{|J|p} E_J \max_{i_J} \left(\sum_{i_{(J_{m-1} \setminus J) \cup J_{m-1}^c \cup \{m\}}} E_{(J_{m-1} \setminus J) \cup J_{m-1}^c \cup \{m\}} h_i \right)^p,
\end{aligned}$$

where in the last line we use the induction assumption (2.8) for $|J_{m-1}| < m$. Finally (??), (??) and (??) imply (??) and complete the proof. \square

Remark. The proof of Proposition 2.6 below will use a version of Corollary 2.2 for nonnegative random functions taking values in L_r . The inequality is as follows: for $p > 1$ there exists $K_{m,p,r} < \infty$ such that

$$(2.8'') \quad E \left\| \sum_{i} h_i \right\|^p \leq K_{m,p,r} \max_{J \subset \{1, \dots, m\}} \left[E_J \max_{i_J} \left(E_{J^c} \left\| \sum_{i_{J^c}} h_i \right\| \right)^p \right].$$

The proof is similar to the previous ones and is omitted: one takes (H_p) as the starting point of the induction.

Finally we come to the third step, which will extend Hoffmann-Jørgensen's inequality (H) for $p \geq 1$. If we want to use the inequalities from Corollary 2.2 to obtain boundedness of moments from stochastic boundedness of a sequence of U -statistics, we need to replace the term corresponding to $J = \emptyset$ by the p -th power of a quantile of $\sum_i h_i$. For this we use Paley-Zygmund's inequality (e.g., Kahane (1968) or de la Peña and Giné (1999)): if A is a nonnegative random variable and $0 < r < p < \infty$, then, for all $0 < \lambda < 1$,

$$(2.13) \quad \Pr \left\{ A > \lambda \|A\|_r \right\} \geq \left[(1 - \lambda^r) \frac{\|A\|_r}{\|A\|_p} \right]^{p/(p-r)},$$

where $\|A\|_r = (E|A|^r)^{1/r}$ for $0 < r < \infty$. Consider for instance inequality (2.8). It has the form

$$EA^p \leq B + K_m^p (EA)^p, \quad p > 1,$$

with $A = \sum_{\mathbf{i}} h_{\mathbf{i}}$. Then, either $B \geq K_m^p (EA)^p$, in which case we have $EA^p \leq 2B$, or $B < K_m^p (EA)^p$, in which case we have $EA^p \leq 2K_m^p (EA)^p$ and we can apply Paley-Zygmund's (2.13) with $\lambda = 1/2$ and $r = 1$. It gives

$$\Pr\left\{A > \frac{1}{2}EA\right\} \geq \frac{1}{2^{(p+1)/(p-1)} K_m^{p/(p-1)}}.$$

Hence, if we define

$$(2.14) \quad t_0 = \inf\left[t \geq 0 : \Pr\{A > t\} \leq \frac{1}{2^{(p+1)/(p-1)} K_m^{p/(p-1)}}\right],$$

we obtain $EA \leq 2t_0$. So, in either case,

$$EA^p \leq 2B + 2^{1+p} K_m^p t_0^p.$$

Also, by Markov's inequality,

$$\frac{1}{2^{(p+1)/(p-1)} K_m^{p/(p-1)}} t_0^p \leq EA^p.$$

We then have:

Theorem 2.3. *Under the hypotheses of Proposition 2.1, there exist a universal constants $K_m < \infty$ such that, if t_0 is as defined by (2.14) for $A = \sum_{\mathbf{i}} h_{\mathbf{i}}$, then*

$$(2.15) \quad \begin{aligned} & \frac{1}{(4K_m)^{p/(p-1)}} t_0^p \vee \max_{\substack{J \subseteq \{1, \dots, m\} \\ J \neq \emptyset}} \left[E_J \max_{\mathbf{i}_J} \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}} \right)^p \right] \\ & \leq E \left(\sum_{\mathbf{i}} h_{\mathbf{i}} \right)^p \\ & \leq (4K_m)^p \left\{ 2^{1+p} t_0^p + \sum_{\substack{J \subseteq \{1, \dots, m\} \\ J \neq \emptyset}} \left[p^{|J|} E_J \max_{\mathbf{i}_J} \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}} \right)^p \right] \right\}. \end{aligned}$$

A similar inequality with different constants can be obtained from (2.8'). This is the most elaborate form we will give to our bounds for $h \geq 0$ and $p > 1$.

The right hand side of (2.15) for $m = 2$ becomes, disregarding constants,

$$(2.15') \quad \begin{aligned} E \left(\sum_{i,j} h_{i,j} \right)^p & \leq C \max \left[E_1 \max_i \left(\sum_j E_2 h_{i,j} \right)^p, E_2 \max_j \left(\sum_i E_1 h_{i,j} \right)^p, \right. \\ & \left. E \max_{i,j} h_{i,j}^p, t_0^p \right]. \end{aligned}$$

So, we get the p -th moment of the double sum controlled by moments of partial maxima of conditional expectations plus a quantile. The Giné-Zinn (1992) inequality (for $m = 2$),

$$E\left(\sum_{i,j} h_{i,j}\right)^p \leq C \max\left[E \max_i \left(\sum_j h_{i,j}\right)^p, t_0^p\right], \quad p \geq 1,$$

is slightly weaker in appearance than (2.15') (actually, we only published the result for canonical U -statistics, but we applied it as well to nonnegative variables, for which the proof is the same: see, e.g., Giné and Zhang (1996)). For applications of this inequality in the asymptotic theory of U -statistics see Giné and Zhang (1996), Giné, Kwapien, Latała and Zinn (1999) and de la Peña and Giné (1999).

Remark. The constants in the definition of t_0 in (2.15) depend on p , hence, so does t_0 . This is not the case when $m = 1$ (as a consequence of the improved Hoffmann-Jørgensen's inequality of Kwapien and Woyczyński -see, de la Peña and Giné (1999) p. 11-). But in most applications it does not matter whether the definition of the quantile depends on p .

2.2. Canonical kernels, moments of order $p \geq 2$. If ξ_i are centered and independent and $p \geq 2$, then, by convexity and the Khinchin-Bonami inequality (e.g., de la Peña and Giné, 1999, p. 113), we have

$$\begin{aligned} 2^{-p} E\left(\sum \xi_i^2\right)^{p/2} &\leq 2^{-p} E\left|\sum \varepsilon_i \xi_i\right|^p \leq E\left|\sum \xi_i\right|^p \\ (2.16) \qquad \qquad \qquad &\leq 2^p E\left|\sum \varepsilon_i \xi_i\right|^p \leq 2^p (p-1)^{p/2} E\left(\sum \xi_i^2\right)^{p/2}, \end{aligned}$$

where ε_i are independent identically distributed Rademacher random variables, independent from $\{\xi_i\}$. Suppose $h_{\mathbf{i}}$ is canonical for the variables $\{X_i^{(j)}\}$ given in the previous subsection, that is, suppose

$$(2.17) \quad E_j h(X_{i_1}^{(1)}, \dots, X_{i_m}^{(m)}) = 0 \text{ a.s. for all } j = 1, \dots, m, \quad 1 \leq i_1, \dots, i_m \leq n.$$

Let $\varepsilon_i^{(j)}$ be an independent Rademacher array independent of $\{X_i^{(j)}\}$, and set

$$\varepsilon_{\mathbf{i}} := \varepsilon_{i_1}^{(1)} \cdots \varepsilon_{i_m}^{(m)}.$$

Then, recursive application of inequality (2.16) gives

$$\begin{aligned} 2^{-mp} E\left(\sum_{\mathbf{i}} h_{\mathbf{i}}^2\right)^{p/2} &\leq 2^{-mp} E\left|\sum_{\mathbf{i}} \varepsilon_{\mathbf{i}} h_{\mathbf{i}}\right|^p \leq E\left|\sum_{\mathbf{i}} h_{\mathbf{i}}\right|^p \\ (2.18) \qquad \qquad \qquad &\leq 2^{mp} E\left|\sum_{\mathbf{i}} \varepsilon_{\mathbf{i}} h_{\mathbf{i}}\right|^p \leq 2^{mp} (p-1)^{mp/2} E\left(\sum_{\mathbf{i}} h_{\mathbf{i}}^2\right)^{p/2}. \end{aligned}$$

This inequality reduces estimation of moments of canonical U -statistics to estimation of moments of nonnegative ones (and conversely), at least if constants are not an issue. Combined with Proposition 2.1, it gives the

analogue of Rosenthal's inequality for centered variables and $p > 2$, and if we apply it in conjunction with Corollary 2.2, we obtain the following inequality:

Proposition 2.4. *If, for $p > 2$ and all $\mathbf{i} \in \{1, \dots, n\}^m$, $h_{\mathbf{i}}(X_{i_1}^{(1)}, \dots, X_{i_m}^{(m)})$ is p -integrable and $E_j h_{\mathbf{i}}(X_{i_1}^{(1)}, \dots, X_{i_m}^{(m)}) = 0$ a.s. for all $j = 1, \dots, m$, then*

$$(2.19) \quad \begin{aligned} & 2^{-mp} \max_{J \subseteq \{1, \dots, m\}} \left[E_J \max_{\mathbf{i}_J} \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}}^2 \right)^{p/2} \right] \leq E \left| \sum_{\mathbf{i}} h_{\mathbf{i}} \right|^p \\ & \leq K_m^p \sum_{J \subseteq \{1, \dots, m\}} \left[p^{(m+|J|)p/2} E_J \max_{\mathbf{i}_J} \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}}^2 \right)^{p/2} \right] \end{aligned}$$

for universal constant $K_m < \infty$.

And, applying Paley-Zygmund with $r = 2$, we finally have:

Theorem 2.5. *Let $h_{\mathbf{i}}$ be as in Proposition 2.4, and let $p > 2$. Then, there exist universal constants $K_m < \infty$ such that, if t_0 is defined as*

$$t_0 = \inf \left[t \geq 0 : \Pr \left\{ \left| \sum_{\mathbf{i}} h_{\mathbf{i}} \right| > t \right\} \leq \left(\frac{3}{4} \right)^{p/(p-2)} \frac{1}{(2K_m^p p^{mp/2})^{1/(p-2)}} \right],$$

then

$$(2.20) \quad \begin{aligned} & \frac{1}{(4K_m^p p^{m/2})^{p/(p-2)}} t_0^p \vee \max_{\substack{J \subseteq \{1, \dots, m\} \\ J \neq \emptyset}} \left[2^{-mp} \max_{\mathbf{i}_J} \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}}^2 \right)^{p/2} \right] \\ & \leq E \left| \sum_{\mathbf{i}} h_{\mathbf{i}} \right|^p \\ & \leq 2K_m^p \left\{ (2p^{m/2})^p t_0^p + \sum_{\substack{J \subseteq \{1, \dots, m\} \\ J \neq \emptyset}} \left[p^{(m+|J|)p/2} E_J \max_{\mathbf{i}_J} \left(\sum_{\mathbf{i}_{J^c}} E_{J^c} h_{\mathbf{i}}^2 \right)^{p/2} \right] \right\}. \end{aligned}$$

If, instead of inequality (2.2), we wish to obtain an analogue of inequality (2.2'), that is, if we want to replace the constants at the right hand side of (2.19) by $(Kp/\log p)^{mp}$, then we cannot use Khinchin's inequality and must proceed directly with an induction as in Proposition 2.1 with the following change: we must consider the variables $\sum_{\mathbf{i}_{J^c}} h_{\mathbf{i}}$ as taking values in $L_2(J_{m-1}^c)$ and apply inequality (1.5) in Kwapien and Szulga (1991), which gives Rosenthal's inequality with best constants for centered independent random variables in Banach spaces. We skip the details.

2.3. Nonnegative kernels, moments of order $p \leq 1$. It seems impossible to obtain inequalities as simple as in the previous section for this case. However, one can still obtain inequalities that may become useful when combined with Paley-Zygmund. Here is an analogue of Corollary 2.2 for

$h \geq 0$ and $p \leq 1$. The method of proof is inefficient regarding constants as Hoffmann-Jørgensen is applied twice at each step. Hence, constants will not be specified.

Proposition 2.6. *Let $0 < r < p \leq 1$, $m < \infty$ and assume that the kernels $h_{\mathbf{i}} \geq 0$ have integrable p -th powers. Then*

$$(2.21) \quad \begin{aligned} & \max_{J \subseteq \{1, \dots, m\}} \left[E_J \max_{\mathbf{i}_J} \left(E_{J^c} \left(\sum_{\mathbf{i}_{J^c}} h_{\mathbf{i}} \right)^r \right)^{p/r} \right] \leq E \left(\sum_{\mathbf{i}} h_{\mathbf{i}} \right)^p \\ & \leq K_{r,p,m} \max_{J \subseteq \{1, \dots, m\}} \left[E_J \max_{\mathbf{i}_J} \left(E_{J^c} \left(\sum_{\mathbf{i}_{J^c}} h_{\mathbf{i}} \right)^r \right)^{p/r} \right], \end{aligned}$$

where $K_{r,p,m}$ depends only on the parameters r , p , m .

Note that all the terms in this bound represent a reduction in the number of sums *except* for the term corresponding to $J = \emptyset$, which consists of a power of the r -th moment of a U -statistic of order m . We will deal later with this term by means of the Paley-Zygmund argument.

Proof. The inequality at the left side of (2.21) follows from Hölder. Inequality (H_r) is just the right hand side of inequality (2.21) for $m = 1$ and we can proceed by induction. We still use the notation from Proposition 2.1. By the induction hypothesis we have

$$(2.22) \quad \begin{aligned} E \left(\sum_{\mathbf{i}} h_{\mathbf{i}} \right)^p &= E_m E_{\{1, \dots, m-1\}} \left(\sum_{\mathbf{i}_{\{1, \dots, m-1\}}} \sum_{i_m} h_{\mathbf{i}} \right)^p \\ &\leq K_{r,p,m-1} \sum_{J_{m-1} \subset \{1, \dots, m-1\}} E_{J_{m-1}} E_m \max_{\mathbf{i}_{J_{m-1}}} \left[E_{J_{m-1}^c} \left(\sum_{\mathbf{i}_{J_{m-1}^c}} \sum_{i_m} h_{\mathbf{i}} \right)^r \right]^{p/r}. \end{aligned}$$

Let us fix $J_{m-1} \subset \{1, \dots, m-1\}$ and note that, for fixed $(X_j^{(i)})_{i \in J_{m-1}}$, we have

$$\max_{\mathbf{i}_{J_{m-1}}} E_{J_{m-1}^c} \left(\sum_{\mathbf{i}_{J_{m-1}^c}} \sum_{i_m} h_{\mathbf{i}} \right)^r := \left\| \sum_{i_m} \tilde{h}_{i_m} \right\|^r$$

for suitably chosen independent r.v.'s \tilde{h}_{i_m} in $l^\infty(L^r)$. Therefore by (H_r) , which still holds in this space (as the norm, restricted to nonnegative vectors, is monotone increasing), we have

$$E_{J_{m-1}} E_m \max_{\mathbf{i}_{J_{m-1}}} \left[E_{J_{m-1}^c} \left(\sum_{\mathbf{i}_{J_{m-1}^c}} \sum_{i_m} h_{\mathbf{i}} \right)^r \right]^{p/r} = E_{J_{m-1}} E_m \left\| \sum_{i_m} \tilde{h}_{i_m} \right\|^{p/r}$$

$$\begin{aligned}
&\leq C_{p,r} E_{J_{m-1}} \left[E_m \max_{i_m} \|\tilde{h}_{i_m}\|^{p/r} + \left(E_m \left\| \sum_{i_m} \tilde{h}_{i_m} \right\| \right)^{p/r} \right] \\
&= C_{p,r} \left[E_{J_{m-1} \cup \{m\}} \max_{\mathbf{i}_{J_{m-1} \cup \{m\}}} \left(E_{J_{m-1}^c} \left(\sum_{\mathbf{i}_{J_{m-1}^c}} h_{\mathbf{i}} \right)^r \right)^{p/r} \right. \\
(2.23) \quad &\quad \left. + E_{J_{m-1}} \left(E_m \max_{\mathbf{i}_{J_{m-1}}} E_{J_{m-1}^c} \left(\sum_{\mathbf{i}_{J_{m-1}^c}} \sum_{i_m} h_{\mathbf{i}} \right)^r \right)^{p/r} \right].
\end{aligned}$$

Now, to estimate the last term, we note that

$$\begin{aligned}
&E_{J_{m-1}} \left[E_m \max_{\mathbf{i}_{J_{m-1}}} E_{J_{m-1}^c} \left(\sum_{\mathbf{i}_{J_{m-1}^c}} \sum_{i_m} h_{\mathbf{i}} \right)^r \right]^{p/r} \\
(2.24) \quad &\leq E_{J_{m-1}} \left[E_{J_{m-1}^c \cup \{m\}} \left(\sum_{\mathbf{i}} h_{\mathbf{i}} \right)^r \right]^{p/r} \\
&\leq K_{p/r,1,|J_{m-1}|} \\
&\quad \times \sum_{J \subset J_{m-1}} E_J \max_{\mathbf{i}_J} \left[E_{J_{m-1} \setminus J \cup J_{m-1}^c \cup \{m\}} \left(\sum_{\mathbf{i}_{J_{m-1} \setminus J \cup J_{m-1}^c \cup \{m\}}} h_{\mathbf{i}} \right)^r \right]^{p/r},
\end{aligned}$$

which follows by the version of Corollary 2.2 for L^r ((2.8'') for $p/r > 1$). Now (??), (??) and (??) complete the induction step. \square

To deal with the term corresponding to $J = \emptyset$ in Proposition 2.6 we apply Paley-Zygmund as above, but now with $r < p$ replacing $1 < p$. The conclusion is:

Theorem 2.7. *There is a constant $K_{r,p,m}$ such that for $0 < r < p \leq 1$, $m < \infty$, and $h_{\mathbf{i}} \geq 0$ with integrable p -th powers, we have*

$$\begin{aligned}
&\frac{1}{(2^{p+1} K_{r,p,m})^{1/(p-r)}} t_0^p \vee \sum_{\substack{J \subseteq \{1, \dots, m\} \\ J \neq \emptyset}} \left[E_J \max_{\mathbf{i}_J} \left(E_{J^c} \left(\sum_{\mathbf{i}_{J^c}} h_{\mathbf{i}} \right)^r \right)^{p/r} \right] \\
(2.25) \quad &\leq E \left(\sum_{\mathbf{i}} h_{\mathbf{i}} \right)^p \\
&\leq 2K_{r,p,m} \left\{ 2^{p/r} t_0^p + \sum_{\substack{J \subseteq \{1, \dots, m\} \\ J \neq \emptyset}} \left[E_J \max_{\mathbf{i}_J} \left(E_{J^c} \left(\sum_{\mathbf{i}_{J^c}} h_{\mathbf{i}} \right)^r \right)^{p/r} \right] \right\},
\end{aligned}$$

where

$$t_0 = \inf \left[t : \Pr \left\{ \sum_{\mathbf{i}} h_{\mathbf{i}} > t \right\} \leq \frac{1}{2} (2^{p+1} K_{r,p,m})^{-1/(p-r)} \right].$$

Hence, the p -th moment of a U -statistic of order m can be estimated by partial moments of maxima (or sums) of conditional moments of U -statistics of lower order plus the p -power of a quantile of the original U -statistic.

2.4. Canonical kernels and moments of order $1 \leq p \leq 2$, or kernels h separately symmetric in each of the coordinates and $0 < p < 1$. The canonical case reduces to the positive case by means of inequality (2.18), as before. The convexity part of inequality (2.18) fails for $p < 1$, but in this case, if h is symmetric separately in each of the coordinates, we can still randomize by products of independent Rademacher variables and recursive application of Khinchin's inequality still reduces this case to nonnegative h . We leave the resulting statements to the reader in order to avoid repetition.

2.5. Regular (undecoupled) general U -statistics. If $h_{\mathbf{i}}(\mathbf{x}) = h_{\mathbf{i} \circ s}(\mathbf{x} \circ s)$ for any permutation s of $\{1, \dots, m\}$ and $h_{\mathbf{i}} = 0$ if \mathbf{i} has repeated indices, and if the sequences $\{X_i^{(j)} : i = 1, \dots, n\}$ are independent copies of each other, then the decoupling inequalities of de la Peña and Montgomery-Smith (1995), together with the decoupling inequality for maxima in Hitczenko (1988) in combination with the previous inequalities give moment inequalities for the generalized U -statistics

$$\sum_{\mathbf{i}} h_{i_1, \dots, i_m}(X_{i_1}, \dots, X_{i_m})$$

where $\{X_i\}$ is a sequence of independent random variables, at the cost of vastly increasing the numerical constants (see e.g. Giné and Zinn (1992) for a similar application of the decoupling inequalities). We omit the resulting statements.

2.6. Comparison with previous results. We have already noted, below the statement of Theorem 2.3, that the inequalities there are better than the Hoffmann-Jørgensen type inequalities for U -statistics in Giné and Zinn (1992) in that they represent a decomposition into simpler quantities. Also, as mentioned in the Introduction, Ibrahimov and Sharakhmetov (1998, 1999) obtained, except for constants, Proposition 2.1 and its analogue for canonical kernels for $m = 2$ and announced the result for general m ; the final results in the present article for $p > 1$ in the nonnegative case (Theorem 2.3) and for $p > 2$ in the canonical case (Theorem 2.5), replacing some sums by maxima and lower moments by quantiles, seem to be more useful. As mentioned above, Corollary 2.2 restricted to $m = 2$ recovers inequalities (4.14) in Klass and Nowicki (1997). The inequalities in the last mentioned article for nonnegative kernels, $p < 1$ and $m = 2$ (the nonconvex case, inequalities (4.13) there) are different from our inequalities in Theorem 2.7 for $m = 2$, although they represent a similar level of decomposition of the p -th moment of the U -statistic. Basically, the difference is that they use inverses of truncated conditional moments whereas we use

inverses of tail probabilities together with partial moments. This can be better seen by comparing Hoffmann-Jørgensen, which is Theorem 2.7 for $m = 1$, with their inequality for $m = 1$. The result of Klass and Nowicki (1997) can be described as the iteration of an inequality that follows from Hoffmann-Jørgensen, Paley-Zygmund ((2.13)) and (2.3), as follows. Given ξ_i , $i = 1, \dots, n$, nonnegative, define v_0 as

$$(2.26) \quad v_0 = \sup \left\{ v \geq 0 : \sum E \left(\frac{\xi_i}{v} \wedge 1 \right) \geq 1 \right\}$$

or, what is the same, v_0 is the largest number satisfying

$$(2.27) \quad v_0 = \sum E(\xi_i \wedge v_0).$$

Then, the inequality in question is:

Corollary 2.8. (Klass and Nowicki, 1997, Cor. 2.7) *Let ξ_i , $i = 1, \dots, n$, be independent nonnegative random variables. Then, for all $p > 0$,*

$$(2.28) \quad E \left(\sum \xi_i \right)^p \simeq E \max \xi_i^p + v_0^p.$$

Proof. Since

$$\sum E(\xi_i \wedge \delta_0) = \sum E \xi_i I_{\xi_i < \delta_0} + \delta_0 \sum \Pr\{\xi_i \geq \delta_0\} \geq \delta_0,$$

it follows that $\delta_0 \leq v_0$. Therefore, if $p \leq 1$, inequality (2.3) and the definition of v_0 give

$$E \left(\sum \xi_i \right)^p \leq \left(\sum E(\xi_i \wedge v_0) \right)^p + \sum_j E \xi_j^p I_{\xi_j > v_0} \leq v_0^p + 2E \max_i \xi_i^p.$$

And if $p > 1$, Hoffmann-Jørgensen ((H)) and the previous inequality (with $p = 1$) give

$$E \left(\sum \xi_i \right)^p \lesssim \left(E \sum \xi_i \right)^p + E \max \xi_i^p \lesssim v_0^p + E \max \xi_i^p.$$

For the reverse inequality, if $p > 1$,

$$v_0^p = \left(\sum E(\xi_i \wedge v_0) \right)^p \leq E \left(\sum \xi_i \right)^p.$$

And if $p < 1$, following the proof of Lemma 2.2 in Klass and Nowicki (1997), we first observe that Paley-Zygmund and the first part of this proof give that for some universal constant C ,

$$\begin{aligned} \Pr \left\{ \sum \xi_i \wedge v_0 > \frac{v_0}{2} \right\} &\geq \frac{1}{4} \frac{\left(E \sum (\xi_i \wedge v_0) \right)^2}{E \left(\sum (\xi_i \wedge v_0) \right)^2} = \frac{1}{4} \frac{v_0^2}{E \left(\sum (\xi_i \wedge v_0) \right)^2} \\ &\geq \frac{C}{4} \frac{v_0^2}{E \max (\xi_i \wedge v_0)^2 + v_0^2} \geq \frac{C}{8}; \end{aligned}$$

therefore,

$$\begin{aligned} E\left(\sum \xi_i\right)^p &\geq E\left(\sum(\xi_i \wedge v_0)\right)^p \\ &\geq E\left[\left(\sum(\xi_i \wedge v_0)\right)^p I_{\sum(\xi_i \wedge v_0) > v_0/2}\right] \geq \frac{C}{8} \frac{v_0^p}{2^p}. \end{aligned}$$

□

In fact, if we bound t_0 by $t_0^p \leq 2E\left(\sum(\xi_i \wedge t_0)\right)^p$ and apply the above proof to the variables $\xi_i \wedge t_0$, Hoffmann-Jørgensen gives the following seemingly weaker inequality: letting \tilde{v}_0 be the parameter v_0 for the smaller variables $\xi_i \wedge t_0$ (note $\tilde{v}_0 \leq v_0$), then

$$(2.22') \quad E\left(\sum \xi_i\right)^p \simeq E \max \xi_i^p + \tilde{v}_0^p.$$

3. Improved moment inequalities and exponential inequalities for $m = 2$

The right hand side of inequality (2.19) for $m = 1$ is just

$$(3.1) \quad E\left|\sum \xi_i\right|^p \leq K^p \max\left[p^p E \max \xi_i^p, p^{p/2} \left(\sum E \xi_i^2\right)^{p/2}\right], \quad p \geq 2,$$

where ξ_i are independent mean zero random variables. These inequalities were first obtained by Pinelis (1994). Part of their interest lie on the fact that they are basically equivalent to Bernstein's inequality up to constants. Here is how (3.1) (for all $p \geq 2$) implies Bernstein's inequality up to constants. Assume $\|\xi_i\|_\infty \leq A < \infty$ for all i , and set $C^2 = \sum E \xi_i^2$. Then, (3.1) has the form

$$E\left|\sum \xi_i\right|^p \leq K^p \max[p^p A^p, p^{p/2} C^p], \quad p \geq 2.$$

Let

$$p = \frac{x}{KeA} \wedge \left(\frac{x}{KeC}\right)^2$$

for any x for which $p \geq 2$. Then, by Markov's inequality, (3.1) gives, for these values of x ,

$$\Pr\left\{\left|\sum \xi_i\right| > x\right\} \leq \begin{cases} \frac{K^p p^p A^p}{x^p} \leq e^{-p} & \text{if } p^p A^p \geq p^{p/2} C^p \\ \frac{K^p p^{p/2} C^p}{x^p} \leq e^{-p} & \text{otherwise.} \end{cases}$$

Hence,

$$(3.2) \quad \Pr\left\{\left|\sum \xi_i\right| > x\right\} \leq e^2 e^{-p} = e^2 \exp\left\{-\frac{x}{KeA} \wedge \left(\frac{x}{KeC}\right)^2\right\}$$

for all $x > 0$. Similarly, from the iteration (2.19) of the inequalities (3.1) we can obtain exponential inequalities for generalized decoupled U -statistics

of any order. However, the inequalities we obtain, while better than the existing ones, are not of the best kind, as we will see below. We illustrate this comment by considering the case $m = 2$. In this case, inequality (2.19) is as follows:

$$(3.3) \quad E \left| \sum_{i,j} h_{i,j} \right|^p \leq K^p \max \left[p^p \left(\sum_{i,j} E h_{i,j}^2 \right)^{p/2}, p^{3p/2} E_1 \max_i \left(\sum_j E_2 h_{i,j}^2 \right)^{p/2}, \right. \\ \left. p^{3p/2} E_2 \max_j \left(\sum_i E_1 h_{i,j}^2 \right)^{p/2}, p^{2p} E \max_{i,j} |h_{i,j}|^p \right].$$

For bounded canonical kernels $h_{i,j}$ we define

$$(3.4) \quad A = \max_{i,j} \|h_{i,j}\|_\infty, \quad C^2 = \sum_{i,j} E h_{i,j}^2, \\ B^2 = \max_{i,j} \left[\left\| \sum_i E_1 h_{i,j}^2(X_i^{(1)}, y) \right\|_\infty, \left\| \sum_j E_2 h_{i,j}^2(x, X_j^{(2)}) \right\|_\infty \right].$$

Then, we can proceed as in the deduction of (3.2) from (3.1), and easily obtain from (3.3) that there is a universal constant K such that

$$(3.5) \quad \Pr \left\{ \left| \sum_{i,j} h_{i,j} \right| > x \right\} \leq K \exp \left\{ -\frac{1}{K} \min \left[\frac{x}{C}, \left(\frac{x}{B} \right)^{2/3}, \left(\frac{x}{A} \right)^{1/2} \right] \right\}.$$

This inequality also holds for regular canonical U -statistics by the decoupling inequalities of de la Peña and Montgomery-Smith (1995).

Inequality (3.5) is better than the Bernstein type inequality in Arcones and Giné (1993) as it is better for $x \leq n^2 A$ and the probability is zero for $x \geq n^2 A$. Inequality (3.5) is suboptimal for small values of x , for which the exponent should be a constant times $-x^2$, just as for chaos variables of order 2 (see Hanson and Wright (1971), Ledoux and Talagrand (1991) and Latała (1999)). This suggests that inequality (2.9) is not of the best kind, and can be improved.

Next we improve the Rosenthal type inequality (2.9) for $m = 2$ (that is, (3.3)) and deduce from it an exponential inequality for canonical U -statistics of order two which does detect the Gaussian portion of the tail probability.

First we show how Talagrand's (1996) extension of Prohorov's inequality to empirical processes, actually in Massart's (1999) version, produces an improved Rosenthal's inequality for empirical processes. Then, we will use this inequality to estimate the terms resulting from conditionally applying inequality (3.1) to the U -statistic.

To describe Massart's version of Talagrand's inequality we must establish the setting and define some parameters. Let Z_i be independent random variables with values in some measurable space (T, \mathcal{T}) , let \mathcal{F} be a countable class of measurable real functions on T , and define

$$S := \sup_{f \in \mathcal{F}} \sum f(Z_i), \quad \sigma^2 = \sup_{f \in \mathcal{F}} \sum E(f(Z_i))^2, \quad a := \max_i \sup_{f \in \mathcal{F}} \|f(Z_i)\|_\infty.$$

Then,

$$(3.6) \quad \Pr\left\{|S| \geq 2E|S| + \sigma\sqrt{8x} + 34.5ax\right\} \leq e^{-x}$$

for all $x > 0$. It follows easily from inequality (3.6) that

$$(3.7) \quad E|S|^p \leq K^p \left[(E|S|)^p + p^{p/2}\sigma^p + p^p a^p \right]$$

for some universal constant $K < \infty$ and all $p \geq 1$, in fact, inequality (3.7) for all p large enough and inequality (3.6) for all $x > 0$ are equivalent up to constants. (We do not plan to keep track of constants in the derivation below and, therefore, we refrain from specifying a value for K in (3.7).)

Proposition 3.1. *Let $\{Z_i\}$ be as above, let \mathcal{F} be a countable class of functions such that $Ef^2(Z_i) < \infty$ and $Ef(Z_i) = 0$ for all i . Then, in the notation from the previous paragraph,*

$$(3.8) \quad E|S|^p \leq K^p \left[(E|S|)^p + p^{p/2}\sigma^p + p^p E \max_i \sup_{f \in \mathcal{F}} |f(Z_i)|^p \right]$$

for all $p \geq 1$, where K is a universal constant.

Proof. Set $F := \sup_{f \in \mathcal{F}} |f|$ and $M^p := 8 \cdot 3^p E \max_i |F(Z_i)|^p$. Since the variables $f(Z_i)$ are centered, we can randomize by independent Rademacher variables ε_i independent of the Z variables (at the price of increasing the value of the constant K). Set $\tilde{S} := \sup_f \left| \sum \varepsilon_i f(Z_i) \right|$. Then,

$$|\tilde{S}| \leq \sup_f \left| \sum \varepsilon_i f(Z_i) I_{F(Z_i) \leq M} \right| + \sup_f \left| \sum \varepsilon_i f(Z_i) I_{F(Z_i) > M} \right| := S_1 + S_2,$$

and notice that, since $ES_1^p \leq 2^{p+1}E|S|^p$ (e.g., Lemmas 1.2.6 and 1.4.3 in de la Peña and Giné, 1999), inequality (3.7) gives

$$ES_1^p \leq K^p \left[(E|S|)^p + p^{p/2}\sigma^p + p^p M^p \right].$$

To estimate ES_2^p we apply the original Hoffmann-Jørgensen inequality (from e.g., Ledoux and Talagrand (1991), (6.9) in page 156) to get

$$ES_2^p \leq 2 \cdot 3^p (t_0^p + E \max_i F(Z_i)^p),$$

where t_0 is any number such that $\Pr\{S_2 > t_0\} \leq (8 \cdot 3^p)^{-1}$. But the choice of M implies that we can take $t_0 = 0$ because

$$\Pr\{S_2 > 0\} = \Pr\{\max_i F(Z_i) > M\} \leq \frac{1}{8 \cdot 3^p},$$

proving the proposition. \square

In what follows we will assume, just as above, that the kernels $h_{i,j}$, $i, j \leq n$, are completely degenerate and define

$$(3.9) \quad D = \|(h_{i,j})\|_{L^2 \rightarrow L^2} := \sup \left\{ E \sum_{i,j} h_{i,j}(X_i^{(1)}, X_j^{(2)}) f_i(X_i^{(1)}) g_j(X_j^{(2)}) \right\}$$

$$: E \sum_i f_i^2(X_i^{(1)}) \leq 1, E \sum_j g_j^2(X_j^{(2)}) \leq 1 \}.$$

Theorem 3.2. *There exists a universal constant $K < \infty$ such that, if $h_{i,j}$ are bounded canonical kernels of two variables for the independent random variables $X_i^{(1)}, X_j^{(2)}$, $i, j = 1, \dots, n$, $n \in \mathbf{N}$, then*

$$(3.10) \quad E \left| \sum_{1 \leq i, j \leq n} h_{i,j}(X_i^{(1)}, X_j^{(2)}) \right|^p \leq K^p \left[p^{p/2} \left(\sum_{i,j} E h_{i,j}^2 \right)^{p/2} + p^p \| (h_{i,j}) \|_{L^2 \rightarrow L^2}^p \right. \\ \left. + p^{3p/2} \left[E_1 \max_i \left(\sum_j E_2 h_{i,j}^2 \right)^{p/2} + E_2 \max_j \left(\sum_i E_1 h_{i,j}^2 \right)^{p/2} \right] \right. \\ \left. + p^{2p} E \max_{i,j} |h_{i,j}|^p \right]$$

for all $p \geq 2$.

Inequality (3.10) is strictly better than the right hand side inequality in (2.9) for $m = 2$, that is, than (3.3).

Proof. Inequality (3.1) applied conditionally on the variables $X_i^{(1)}$ gives

$$(3.11) \quad E \left| \sum_{i,j} h_{i,j} \right|^p \leq K^p E_1 \left(p^{p/2} \left[\sum_j E_2 \left(\sum_i h_{i,j} \right)^2 \right]^{p/2} + p^p E_2 \sum_j \left| \sum_i h_{i,j} \right|^p \right).$$

To bound the first summand at the right hand side of (3.11) we first notice that

$$\left[\sum_j E_2 \left(\sum_i h_{i,j} \right)^2 \right]^{1/2} \\ = \sup \left[\sum_i E_2 \sum_j h_{i,j}(X_i^{(1)}, X_j^{(2)}) f_j(X_j^{(2)}) : E \sum_j f_j^2(X_j^{(2)}) \leq 1 \right],$$

where in fact, the sup is taken only over a countable subset of mean zero vector functions (f_1, \dots, f_n) dense in the unit ball of $L_2(\mathcal{L}(X_1^{(2)})) \times \dots \times L_2(\mathcal{L}(X_n^{(2)}))$ for the seminorm $|(f_j)_{j \leq n}| = \left(\sum E f_j^2(X_j^{(2)}) \right)^{1/2}$. [To see this, first apply duality in ℓ_2^n and then in $L_2(\mathcal{L}(X_j^{(2)}))$ for each j .] So we can apply (3.8) to $Z_i = (h_{i,j})_{j=1}^n$ with $f(Z_i) = E_2 \sum_j h_{i,j}(X_i^{(1)}, X_j^{(2)}) f_j(X_j^{(2)})$. In this case, the right hand side terms in (3.8) can be estimated as follows. The first term:

$$(E|S|)^2 \leq E|S|^2 = E \left[\sum_j E_2 \left(\sum_i h_{i,j} \right)^2 \right] = E \sum_{i,j} h_{i,j}^2 = C^2.$$

For the second we see that, since, by the previous duality argument,

$$\sum_i E_1 \left(E_2 \sum_j h_{i,j}(X_i^{(1)}, X_j^{(2)}) f_j(X_j^{(2)}) \right)^2 \leq \| (h_{i,j}) \|_{L^2 \rightarrow L^2}^2 = D^2,$$

it follows that $\sigma \leq D$. The third term:

$$\begin{aligned} E \max_i \sup_f |f(Z_i)|^p &= E_1 \max_i \sup_{E \sum f_j^2 \leq 1} \left[E_2 \sum_j h_{i,j}(X_i^{(1)}, X_j^{(2)}) f_j(X_j^{(2)}) \right]^p \\ &\leq E_1 \max_i \sup_{E \sum f_j^2 \leq 1} \left[\left(E_2 \sum_j h_{i,j}^2 \right)^{1/2} \left(E \sum_j f_j^2 \right)^{1/2} \right]^p \\ &= E_1 \max_i \left(E_2 \sum_j h_{i,j}^2 \right)^{p/2}. \end{aligned}$$

Thus, inequality (3.8) gives

$$\begin{aligned} (3.12) \quad p^{p/2} E \left(\sum_j E_2 \left(\sum_i h_{i,j} \right)^2 \right)^{p/2} \\ \leq K^p \left[p^{p/2} C^p + p^p D^p + p^{3p/2} E_1 \max_i \left(E_2 \sum_j h_{i,j}^2 \right)^{p/2} \right]. \end{aligned}$$

To estimate the second summand at the right hand side of (3.11), we apply (3.1) once more and obtain

$$\begin{aligned} (3.13) \quad p^p E_2 \sum_j E_1 \left| \sum_i h_{i,j} \right|^p \\ \leq K^p \left[p^{3p/2} E_2 \sum_j \left(\sum_i E_1 h_{i,j}^2 \right)^{p/2} + p^{2p} E \sum_{i,j} |h_{i,j}|^p \right]. \end{aligned}$$

Thus, to complete the proof of the theorem it suffices to replace the sum in j and the sum in i, j respectively by maxima in j and in i, j on the terms at the right hand side of this inequality. But this is an easy exercise of application of inequality (2.6). For completeness sake, here it is. Applying (2.6) with $\alpha = 3$ and $p/2$ instead of p , the first term at the right of (3.13) bounds as:

$$\begin{aligned} p^{3p/2} E_2 \sum_j \left(\sum_i E_1 h_{i,j}^2 \right)^{p/2} \\ \leq 2^{1+3p/2} (1 + (p/2)^3) \left[\left(\frac{p}{2} \right)^{3p/2} E_2 \max_j \left(\sum_i E_1 h_{i,j}^2 \right)^{p/2} + C^p \right], \end{aligned}$$

which produces the conversion of the sum into a maximum without increasing the order of the multiplicative constant in front of C^p . The second term in (3.13) requires two steps. First, we apply (2.6) for $p/2$ and $\alpha = 4$,

conditionally on $\{X_i^{(1)}\}$:

$$(3.14) \quad p^{2p} E \sum_{i,j} |h_{i,j}|^p \\ \leq 2^{2p+1} (1 + (p/2)^4) E_1 \sum_i \left[\left(\frac{p}{2}\right)^{2p} E_2 \max_j |h_{i,j}|^p + \left(\sum_j E_2 h_{i,j}^2\right)^{p/2} \right].$$

We apply (2.6) with respect to E_1 , for $p/2$ and $\alpha = 0$, to the second term at the right hand side of (3.14) and we obtain the bound

$$2^{2p+3} (1 + (p/2)^4) \left[E_1 \max_i \left(\sum_j E_2 h_{i,j}^2\right)^{p/2} + C^p \right],$$

which is in terms of some of the quantities appearing at the right hand side of (3.10) and with coefficients of lower order. As for the first term at the right of (3.14), we apply (2.6) with respect to E_1 , again for $p/2$ and $\alpha = 4$, and get it bounded by

$$2^{4p+2} (1 + (p/2)^4)^2 \left[\left(\frac{p}{2}\right)^{2p} E \max_{i,j} |h_{i,j}|^p + E_2 \left(\sum_i E_1 \max_j h_{i,j}^2\right)^{p/2} \right].$$

Here the first term coincides with the last one in (3.10), and the second is dominated by

$$K^p E_2 \left[\sum_j \left(\sum_i E_1 h_{i,j}^2\right) \right]^{p/2}.$$

Applying inequality (R_1) with respect to E_2 this is in turn dominated by

$$K^p \left(\frac{p}{2}\right)^{p/2} E_2 \sum_j \left(\sum_i E_1 h_{i,j}^2\right)^{p/2} + K^p C^p,$$

and the first summand has already been handled above (first term at the right of (3.13)). Collecting terms we obtain inequality (3.10). \square

Theorem 3.2 implies the following moment inequality and exponential bound for bounded kernels.

Theorem 3.3. *There exist universal constants $K < \infty$ and $L < \infty$ such that, if $h_{i,j}$ are bounded canonical kernels of two variables for the independent random variables $X_i^{(1)}, X_j^{(2)}$, $i, j = 1, \dots, n$, and if A, B, C, D are as defined in (3.4) and (3.9), then*

$$(3.15) \quad E \left| \sum_{1 \leq i, j \leq n} h_{i,j}(X_i^{(1)}, X_j^{(2)}) \right|^p \leq K^p \left[p^{p/2} C^p + p^p D^p + p^{3p/2} B^p + p^{2p} A^p \right]$$

for all $p \geq 2$ and, equivalently,

$$(3.16) \quad \Pr \left\{ \left| \sum_{i,j \leq n} h_{i,j}(X_i^{(1)}, X_j^{(2)}) \right| \geq x \right\} \leq L \exp \left[-\frac{1}{L} \min \left(\frac{x^2}{C^2}, \frac{x}{D}, \frac{x^{2/3}}{B^{2/3}}, \frac{x^{1/2}}{A^{1/2}} \right) \right]$$

for all $x > 0$.

The moment inequality is immediate from Theorem 3.2 and the equivalence with the exponential inequality follows just like (3.2) follows from (3.1) in one direction, and, in the other, by integration of tail probabilities.

Next we comment on the exponential inequality. For comparison purposes, let $h_{i,j}(X_i^{(1)}, X_j^{(2)}) = g_i g_j' x_{i,j}$ with g_i, g_j' independent standard normal. In this case,

$$C^2 = \sum_{i,j} x_{i,j}^2 \text{ and } D = \sup \left\{ \sum_{i,j} u_i v_j x_{i,j} : \sum u_i^2 \leq 1, \sum v_j^2 \leq 1 \right\}$$

and the Gaussian chaos inequalities of Hanson and Wright (1971) and Latała (1999) yield the existence of universal constants $0 < k < K < \infty$ such that

$$\Pr \left\{ \left| \sum_{i,j} h_{i,j} \right| \geq K(Cx^{1/2} + Dx) \right\} \leq e^{-x}$$

and

$$\Pr \left\{ \left| \sum_{i,j} h_{i,j} \right| \geq k(Cx^{1/2} + Dx) \right\} \geq k \wedge e^{-x}.$$

By the central limit theorem for canonical U -statistics, this implies that the coefficients of x^2 and x in (3.16) are correct (except for K). It is natural to have terms in smaller powers of x in (3.16) e.g., by comparison with Bernstein's inequality for sums of independent random variables. In fact, the term in $x^{1/2}$ cannot be avoided, at least up to logarithmic factors. To see this, consider the product V of two independent centered Poisson variables with parameter 1, which is the limit in law of $V_n = \sum_{i,j \leq n} X_i^{(n)} Y_j^{(n)}$ where $X_i^{(n)}$ and $Y_j^{(n)}$ are centered Bernoulli random variables with parameter $p = 1/n$; then, for large x , the tail probabilities of V are of the order of $\exp(-x^{1/2} \log x)$, and therefore, so are those of V_n for large n . Also, note that the term in $x^{2/3}$ in the exponent corresponds, up to logarithmic factors, to the tail probabilities of the product of two independent random variables, one normal and the other centered Poisson.

If $X, Y, X_i^{(1)}, X_j^{(2)}$ are i.i.d., $h_{i,j} = h$ for all i, j and h is completely degenerate, then the parameters defined by (3.4) and (3.8) become:

$$A = \|h\|_\infty, \quad B^2 = n(\|E_Y h^2(x, Y)\|_\infty + \|E_X h^2(X, y)\|_\infty), \quad C^2 = n^2 E h^2$$

and

$$\begin{aligned} D &= n \sup \left\{ Eh(X, Y) f(X) g(Y) : Ef^2(X) \leq 1, Eg^2(Y) \leq 1 \right\} \\ &:= n \|h\|_{L_2 \mapsto L_2}, \end{aligned}$$

where $\|h\|_{L_2 \mapsto L_2}$ is the norm of the operator of $L_2(\mathcal{L}(X))$ with kernel h . Then, inequalities (3.15) and (3.16) become:

Corollary 3.4. *Under the above assumptions, there exist universal constants $K < \infty$, $L < \infty$ such that, for all $n \in \mathbf{N}$ and $p \geq 2$,*

$$(3.17) \quad E \left| \sum_{i, j \leq n} h(X_i^{(1)}, X_j^{(2)}) \right|^p \leq K^p \left[p^{p/2} n^p (Eh^2)^{p/2} + p^p n^p \|h\|_{L_2 \mapsto L_2}^p \right. \\ \left. + p^{3p/2} n^{p/2} (\|E_Y h^2\|_\infty + \|E_X h^2\|_\infty)^{p/2} + p^{2p} \|h\|_\infty^p \right]$$

and

$$(3.18) \quad \Pr \left\{ \left| \sum_{i, j \leq n} h(X_i^{(1)}, X_j^{(2)}) \right| \geq x \right\} \leq K \exp \left[-\frac{1}{K} \min \left(\frac{x^2}{n^2 E h^2}, \right. \right. \\ \left. \left. \frac{x}{n \|h\|_{L_2 \mapsto L_2}}, \frac{x^{2/3}}{[n(\|E_Y h^2\|_\infty + \|E_X h^2\|_\infty)]^{1/3}}, \frac{x^{1/2}}{\|h\|_\infty^{1/2}} \right) \right].$$

Inequality (3.18) provides an analogue of Bernstein's inequality for degenerate U -statistics of order 2: note that inequalities (3.15), (3.16), (3.17) and (3.18) can all be 'undecoupled' using the result of de la Peña and Montgomery-Smith's (1995). It should also be noted that this exponential inequality for canonical U -statistics is strong enough to imply the sufficiency part of the law of the iterated logarithm for these objects: this can be seen by applying it to the kernels h_n in Steps 7 and 8 of the proof of Theorem 3.1 in Giné, Kwapien, Latała and Zinn (1999) (and using some of the computations there for the parameters C to D). Neither inequality (3.5) nor any of the previously published inequalities for U -statistics can do this.

Acknowledgement. We thank Stanislaw Kwapien for several useful conversations.

References

- ARCONES, M. AND GINÉ, E. (1993). Limit theorems for U -processes. *Ann. Probab.* **21** 1494-1542.
- DE LA PEÑA, V. AND GINÉ, E. (1999). *Decoupling: From Dependence to Independence*. Springer-Verlag, New York.
- DE LA PEÑA, V. AND MONTGOMERY-SMITH, S. (1995). Decoupling inequalities for the tail probabilities of multivariate U -statistics. *Ann. Probab.* **23** 806-816.

- FIGIEL, T.; HITCZENKO, P.; JOHNSON, W.B.; SCHECHTMAN, G.; AND ZINN, J. (1997). Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. *Trans. Amer. Math. Soc.* **349** 997-1027.
- GINÉ, E.; KWAPIEŃ, S.; LATAŁA, R.; AND ZINN, J. (1999). The LIL for canonical U -statistics of order two. To appear.
- GINÉ, E. AND ZHANG, C.-H. (1996). On integrability in the LIL for degenerate U -statistics. *J. Theoret. Probab* **9** 385-412.
- GINÉ, E. AND ZINN, J. (1983). Central limit theorems and weak laws of large numbers in certain Banach spaces. *Zeits. Wahrsch. v. Geb.* **62** 323-354.
- GINÉ, E. AND ZINN, J. (1992). On Hoffmann-Jørgensen's inequality for U -processes. *Probability in Banach Spaces 8* 80-91. Birkhäuser, Boston.
- HANSON, D. L. AND WRIGHT, F. T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.* **42** 52-61.
- HITCZENKO, P. (1988). Comparison of moments for tangent sequences of random variables. *Probab. Th. Rel. Fields* **78** 223-230.
- JOHNSON, W. B.; SCHECHTMAN, G.; AND ZINN, J. (1985). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.* **13** 234-253.
- IBRAGIMOV, R. AND SHARAKHMETOV, SH. (1998). Exact bounds on the moments of symmetric statistics. In: Abstracts of the 7-th Vilnius Conference on Probability Theory and Mathematical Statistics/ 22nd European Meeting of Statisticians, pp. 243-244. Vilnius.
- IBRAGIMOV, R. AND SHARAKHMETOV, SH. (1999). Analogues of Khintchine, Marcinkiewicz-Zygmund and Rosenthal inequalities for symmetric statistics. *Scand. J. Statist.* **26** 621-623.
- KAHANE, J.-P. (1968). *Some Random Series of Functions*. Heath, Lexington, Massachusetts.
- KLASS, M. AND NOWICKI, K. (1997). Order of magnitude bounds for expectations of Δ_2 functions of nonnegative random bilinear forms and generalized U -statistics. *Ann. Probab.* **25** 1471-1501.
- KWAPIEŃ, S. AND SZULGA, J. (1991). Hypercontraction methods in moment inequalities for series of independent random variables in normed spaces. *Ann. Probab.* **19** 369-379.
- KWAPIEŃ, S. AND WOYCZYŃSKI, W. (1992). *Random Series and Stochastic Integrals: Single and Multiple*. Birkhäuser, Boston.
- LATAŁA, R. (1997). Estimation of moments of sums of independent random variables. *Ann. Probab.* **25** 1502-1513.
- LATAŁA, R. (1999). Tails and moment estimates for some type of chaos. *Studia Math.* **135** 39-53.
- LATAŁA, R. AND ZINN, J. (1999). Necessary and sufficient conditions for the strong law of large numbers for U -statistics. *Ann. Probab.*, to appear.

LEDoux, M. AND TALAGRANd, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, New York.

MASSART, P. (1999). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, to appear.

PINELIS, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** 1679-1706.

TALAGRANd, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505-563.

UTEV, S. A. (1985). Extremal problems in moment inequalities. In: *Limit Theorems in Probability Theory*, Trudy Inst. Math., Novosibirsk, 56-75 (in Russian).

Evarist Giné
Department of Mathematics
and Department of Statistics
University of Connecticut
Storrs, CT 06269
USA
gine@uconnvm.uconn.edu

Rafał Latała
Institute of Mathematics
Warsaw University
Banacha 2
02-097 Warszawa
Poland
rlatala@mimuw.edu.pl

Joel Zinn
Department of Mathematics
Texas A&M University
College Station, TX 77843
jzinn@math.tamu.edu