BOUNDS ON THE TAIL PROBABILITY OF 
U-STATISTICS AND QUADRATIC FORMS

VICTOR H. DE LA PEÑA AND S. J. MONTGOMERY-SMITH

It is very common for expressions of the form:

$$\sum_{1 \leq i_1 \neq \cdots \neq i_k \leq n} f_{i_1 \ldots i_k}(X_{i_1}, \ldots, X_{i_k})$$

to appear in Probability theory. Here \( \{X_i\} \) is a sequence of independent random variables taking values in a measurable space \((S, \mathcal{S})\), and \( \{f_{i_1 \ldots i_k}\} \) is a sequence of measurable functions from \( S^k \) into a Banach Space \((B, \|\cdot\|)\). Special cases of this type of random variable appear, for example, in Statistics in the form of U-statistics and quadratic forms. Throughout we will refer to them as generalized U-statistics.

There is great interest in decoupling such quantities, that is, by replacing the above quantity by the expression

$$\sum_{1 \leq i_1 \neq \cdots \neq i_k \leq n} f_{i_1 \ldots i_k}(X^{(1)}_{i_1}, \ldots, X^{(k)}_{i_k}),$$

where \( \{X^{(1)}_i\}, \{X^{(2)}_i\}, \ldots, \{X^{(k)}_i\} \) are \( k \) independent copies of \( \{X_i\} \).

Decoupling inequalities allow one to compare expressions of the first kind with expressions of the second kind. Such results permit the almost direct transfer of results for sums of independent random variables to the case of generalized U-statistics. The reason being that conditionally on \( \{X^{(2)}_i\}, \ldots, \{X^{(k)}_i\} \), the second sum above is a sum independent random variables. It is important to remark that such results have led to the development of several optimal results in the functional theory of U-Statistics (cf. [1] and [7]) and various other areas including the study of the invertibility of large matrices (cf. [2]), stochastic integration (cf. [10]), the study of integral operators on Lebesgue-Bochner spaces (cf. a result of T. R. McConnell and D. Burkholder found in [3]). Aside from those directly cited in this paper, other important contributors to the area of decoupling inequalities include A. de Acosta, P. Hitczenko, J. Jacod, A. Jakubowski, O. Kallenberg, M. Klass, W. Krakowiak, G. Pisier, J. Rosinski, and J. Szulga. Due to space restrictions, we refer the reader to [10] for a more complete account.

In this paper, we announce a result, which allows one to compare the tail probabilities of the above quantities. In particular, this inequality represents the definitive

1991 Mathematics Subject Classification. AMS 1991 subject classifications: Primary 60E15. Secondary 60D05, 60E05.

Key words and phrases. U-statistics, Quadratic Forms, Decoupling.

Both authors were supported by NSF grants.

**Theorem 1.** There is a constant $C_k > 0$, depending only on $k$, such that for all $n \geq k$,

$$P\left( \left\| \sum_{1 \leq i_1 \neq \cdots \neq i_k \leq n} f_{i_1 \cdots i_k} (X_{i_1}, \ldots, X_{i_k}) \right\| \geq t \right) \leq C_k P\left( C_k \left\| \sum_{1 \leq i_1 \neq \cdots \neq i_k \leq n} f_{i_1 \cdots i_k} (X_{i_1}^{(1)}, \ldots, X_{i_k}^{(k)}) \right\| \geq t \right), \text{ for all } t > 0. \tag{1}$$

Moreover, the reverse inequality holds if the functions satisfy the condition

$$f_{i_1 \cdots i_k} (X_{i_1}, \ldots, X_{i_k}) = f_{\pi_{i_1(1)} \cdots \pi_{i_k(k)}} (X_{i_1^{\pi(1)}}, \ldots, X_{i_k^{\pi(k)}})$$

for all permutations $\pi$ of $\{1, \ldots, k\}$. That is,

$$P\left( \left\| \sum_{1 \leq i_1 \neq \cdots \neq i_k \leq n} f_{i_1 \cdots i_k} (X_{i_1}, \ldots, X_{i_k}) \right\| \geq t \right) \geq \frac{1}{C_k} P\left( \frac{1}{C_k} \left\| \sum_{1 \leq i_1 \neq \cdots \neq i_k \leq n} f_{i_1 \cdots i_k} (X_{i_1}^{(1)}, \ldots, X_{i_k}^{(k)}) \right\| \geq t \right), \text{ for all } t > 0. \tag{2}$$

Note that the expression $i_1 \neq \cdots \neq i_k$ means that $i_r \neq i_s$ for all $1 \leq r \neq s \leq k$.

An example to illustrate this result can be found in the study of random graphs (see also [5]). Given a sequence of independent random points $\{X_i\}$ in $\mathbb{R}^N$, we might consider a measure of clustering

$$D_1 = \sum_{1 \leq i \neq j \leq n} d(X_i, X_j),$$

where $d(x, y)$ denotes the distance between $x$ and $y$. The above result allows us to compare $D_1$, which measures the distance ‘within’ the graph formed by the random cluster of points $\{X_i\}$, to a quantity $D_2$, which is a measure of the distance ‘between’ the two independent clusters $\{X_i\}$ and $\{\tilde{X}_i\}$,

$$D_2 = \sum_{1 \leq i \neq j \leq n} d(\tilde{X}_i, X_j),$$

where $\{\tilde{X}_i\}$ is an independent copy of $\{X_i\}$. Then we have for all $t > 0$ that,

$$C_2^{-1} P(|D_1| \geq C_2 t) \leq P(|D_2| \geq t) \leq C_2 P(|D_1| \geq C_2^{-1} t).$$

Other examples where U-statistics are used in graph theory may be found in [8].
We will prove the Theorem in the special case that $k = 2$. For ease of notation, let us suppose that $X_i = X_i^{(1)}$, and denote $X_i = X_i^{(2)}$. The proof of the more general result will appear elsewhere. We will use a sequence of lemmas. Following [6], our point of departure is equation (4) below which provides a partial decoupling result and focuses attention on a polarized version of the U-statistic kernel as the key element in the development of a solution of the problem at hand. Let

$$T_n = \sum_{1 \leq i \neq j \leq n} \{ f_{ij}(X_i, X_j) + f_{ij}(X_i, \tilde{X}_j) + f_{ij}(\tilde{X}_i, X_j) + f_{ij}(\tilde{X}_i, \tilde{X}_j) \},$$

then by using the triangle inequality one obtains that,

$$P(\| T_n \| \geq \frac{t}{3}) \leq 3P(\| X + Y \| \geq \frac{2t}{3}).$$

This observation reduces the proof (1) to the problem of obtaining the bounds

$$P(\| \sum_{1 \leq i \neq j \leq n} f_{ij}(X_i, X_j) \| \geq t) \leq cP(\| \sum_{1 \leq i \neq j \leq n} f_{ij}(X_i, \tilde{X}_j) + f_{ij}(\tilde{X}_i, X_j) \| \geq t),$$

and

$$P(\| T_n \| \geq t) \leq cP(\| \sum_{1 \leq i \neq j \leq n} f_{ij}(\tilde{X}_i, X_j) \| \geq t).$$

We obtain (5) by means of Lemma 1 (possibly of independent interest). The proof of (6) is somewhat involved. In obtaining it, we used (conditionally) an extension of the Paley-Zygmund inequality found in [10] in combination with a symmetrization identity similar to the one introduced in [12].

**Lemma 1.** Let $X, Y$ be two i.i.d. random variables. Then

$$P(\| X \| \geq t) \leq 3P(\| X + Y \| \geq \frac{2t}{3}).$$

**Proof.** Let $X, Y$ and $Z$ be i.i.d. random variables. Then

$$P(\| X \| \geq t)$$
$$= P(\| (X + Y) + (X + Z) - (Y + Z) \| \geq 2t)$$
$$\leq P(\| X + Y \| \geq \frac{2t}{3}) + P(\| X + Z \| \geq \frac{2t}{3}) + P(\| Y + Z \| \geq \frac{2t}{3})$$
$$= 3P(\| X + Y \| \geq \frac{2t}{3}).$$

It is to be remarked that the very desirable 'Universal Symmetrization Lemma', $P(\| X \| \geq t) \leq cP(\| X - Y \| \geq t)$ is not true. This makes the above result all the more surprising.

The following is an observation found in Section 6.2 of [10] will be used in combination with Lemma 2 in proving Theorem 1.
Proposition 1. Let \( Y \) be any mean zero random variable with values in a Banach space \((B, \|\cdot\|)\). Then, for all \( a \in B \), \( P(\|a + Y\| \geq \|a\|) \geq \frac{1}{2} \), where, let \( \kappa = \inf_{x \in B^*} \frac{(E|x(x)|^2)^2}{E|x(x)|^2} \).

As a consequence of the above we obtain the following lemma.

Lemma 2. Let \( x, a_i, b_{ij} \) all belong to a Banach space \((B, \|\cdot\|)\), with \( b_{ii} = 0 \). Let \( \{\epsilon_i\} \) be a sequence of independent and symmetric Bernoulli random variables, that is, \( P(\epsilon_i = 1) = P(\epsilon_i = -1) = \frac{1}{2} \). Then, for a universal constant \( c > 0 \),

\[
P(\|x + \sum_{i=1}^n a_i \epsilon_i + \sum_{1 \leq i \neq j \leq n} b_{ij} \epsilon_i \epsilon_j \| \geq \|x\|) \geq c^{-1}.
\]

Proof. Suppose that \( a_i, b_{ij} \) are in \( R \), then it follows easily from (1.4) of [9] (see also sections 6.2 and 6.5 of [10]) that

\[
(\mathcal{E} \sum_{i=1}^n a_i \epsilon_i + \sum_{1 \leq i \neq j \leq n} b_{ij} \epsilon_i \epsilon_j)^4 \leq c(\mathcal{E} \sum_{i=1}^n a_i \epsilon_i + \sum_{1 \leq i \neq j \leq n} b_{ij} \epsilon_i \epsilon_j)^2,\]

for some constant \( c > 0 \). Next, observe that \( \|\xi\|_4 \leq c\|\xi\|_2 \) implies that \( \|\xi\|_2 \leq c^2\|\xi\|_4 \) (since \( E(\xi)^2 \leq (E(\xi))^{2/3} \cdot (E(\xi)^4)^{1/3} \)). The result then follows by Proposition 1.

Proof of Theorem 1. We first transform the problem of proving (6) into a problem dealing (conditionally) with a non-homogeneous binomial in Bernoulli random variables. Let \( \{\epsilon_i\} \) be a sequence of independent and symmetric Bernoulli random variables independent of \( \{X_i\} \), \( \{\tilde{X}_i\} \). Let \( (Z_i, \tilde{Z}_i) = (X_i, \tilde{X}_i) \) if \( \epsilon_i = 1 \) and \( (Z_i, \tilde{Z}_i) = (\tilde{X}_i, X_i) \) if \( \epsilon_i = -1 \). Then,

\[
4f_{ij}(\tilde{Z}_i, Z_j) = \{(1 - \epsilon_i)(1 + \epsilon_j)f_{ij}(X_i, X_j) + (1 + \epsilon_i)(1 + \epsilon_j)f_{ij}({\tilde{X}}_i, X_j) + (1 - \epsilon_i)(1 + \epsilon_j)f_{ij}(X_i, {\tilde{X}}_j) + (1 + \epsilon_i)(1 - \epsilon_j)f_{ij}({\tilde{X}}_i, {\tilde{X}}_j)\}.
\]

Setting \( \mathcal{G} = \sigma(X_i, \tilde{X}_i; i = 1, ..., n) \) we get

\[
4E(f_{ij}(\tilde{Z}_i, Z_j) | \mathcal{G}) = \{f_{ij}(X_i, X_j) + f_{ij}(X_i, \tilde{X}_j) + f_{ij}({\tilde{X}}_i, X_j) + f_{ij}({\tilde{X}}_i, {\tilde{X}}_j)\}.
\]

From Lemma 2, (3), (8) and (9), and letting \( x = T_n \), it follows that for some \( c > 0 \),

\[
P(4\| \sum_{1 \leq i \neq j \leq n} f_{ij}(\tilde{Z}_i, Z_j)\| \geq \|T_n\| | \mathcal{G}) \geq c^{-1}.
\]

Integrating over the set \( \{\|T_n\| \geq t\} \) we get

\[
\frac{1}{c} P(\|T_n\| \geq t) \leq P(4\| \sum_{1 \leq i \neq j \leq n} f_{ij}(\tilde{Z}_i, Z_j)\| \geq t) = P(4\| \sum_{1 \leq i \neq j \leq n} f_{ij}(\tilde{X}_i, X_j)\| \geq t),
\]

since the sequence \( \{(X_i, \tilde{X}_i), i = 1, ..., n\} \) has the same distribution as \( \{(Z_i, \tilde{Z}_i), i = 1, ..., n\} \). The proof is completed by using this inequality along with (4) and (5).
The proof of (2) is similar and uses an analogue of (8) concerning \( 4f(Z_i, Z_j) \). In obtaining this bound, one does not need to use Lemma 1. Instead one uses the symmetry condition on the functions \( f_{ij} \), introduced after (1) and equation (3) to get,

\[
P(\| \sum_{1 \leq i \neq j \leq n} f_{ij}(\tilde{X}_i, X_j) \| \geq t) = P(\| \sum_{1 \leq i \neq j \leq n} f_{ij}(X_i, \tilde{X}_j) + f_{ij}(\tilde{X}_i, X_j) \| \geq 2t) \\
\leq P(\| T_n \| \geq \frac{2}{3} t) + 2P(\| \sum_{1 \leq i \neq j \leq n} f_{ij}(X_i, X_j) \| \geq \frac{2}{3} t).
\]

References

Department of Statistics, Columbia University, New York, New York 10027
E-mail address: vp@wald.stat.columbia.edu

Department of Mathematics, University of Missouri, Columbia, Missouri 65211
E-mail address: stephen@mont.cs.missouri.edu