The Annals of Probability
1997, Vol. 25, No. 4, 2055–2081

MOMENTS OF RANDOMLY STOPPED *U*-STATISTICS

By Victor H. de la Peña[1] and Tze Leung Lai[2]

Columbia University and Stanford University

In this paper we provide sharp bounds on the L_p-norms of randomly stopped U-statistics. These bounds consist mainly of decoupling inequalities designed to reduce the level of dependence between the U-statistics and the stopping time involved. We apply our results to obtain Wald's equation for U-statistics, moment convergence theorems and asymptotic expansions for the moments of randomly stopped U-statistics. The proofs are based on decoupling inequalities, symmetrization techniques, the use of subsequences and induction arguments.

1. Introduction. Let $X_1, X_2, \dots$ be i.i.d. random variables and let T be a stopping time adapted to $\{F_n\}$, where F_n is the σ-algebra generated by $X_1, \dots, X_n$. Wald's (1945) equation says that if $EX_1 = 0$ and $ET < \infty$, then $E(\sum_{i=1}^T X_i) = 0$. Chow, de la Peña and Teicher (1993) recently generalized this fundamental result in sequential analysis to multilinear U-statistics of the form

$$S_n = \sum_{1 \le i_1 < \cdots < i_k \le n} X_{i_1} \cdots X_{i_k}. \tag{1.1}$$

In the case $k = 2$, de la Peña and Lai (1994) further extended the result to general U-statistics of order $k(=2)$, showing that

$$E \sum_{1 \le i < j \le T} f(X_i, X_j) = 0, \tag{1.2}$$

if for some $1 < p \le 2$,

$$\begin{gathered} E(T^{1/(p-1)}) < \infty, \qquad E|f(X_1, X_2)|^p < \infty \quad \text{and} \\ E(f(X_1, X_2)|X_1) = E(f(X_1, X_2)|X_2) = 0. \end{gathered} \tag{1.3}$$

The sharpness of this result can be checked by taking $f(x, y) = xy$ and $p = 2$, in which case (1.2) along with Wald's first equation gives Wald's second equation.

In Sections 2 and 5 we extend (1.2) to general U-statistics of order $k \ge 2$ under a similar assumption which reduces to (1.3) when $k = 2$. In this connection we also prove a sharp bound for the absolute pth moment of a randomly stopped U-statistic for $p \le 2$. This bound has the property of

---

Received November 1995.

[1]Research supported by NSF Grant DMS-93-10682.
[2]Research supported by NSF Grant DMS-94-03794.
AMS 1991 *subject classifications.* Primary 60G40, 60F25; secondary 62L12.
Key words and phrases. Decoupling inequalities, martingales, stopping times, U-statistics, Wald's equation, uniform integrability.

reducing the level of dependence between the sequence of U-statistics and the stopping time. In Section 3 we use the decoupling method to obtain a bound for the absolute pth moment, with $p \geq 2$, of randomly stopped U-statistics. We provide several applications of the bounds obtained in Sections 3 and 4. In particular, they are used to establish moment convergence results in Anscombe's theorem for U-statistics, generalizing corresponding results of Chow, Hsiung and Lai (1979) for stopped random walks. They are also used to obtain asymptotic expansions of the moments of randomly stopped normalized U-statistics, generalizing recent results of Aras and Woodroofe (1993) for stopped sample means to stopped U-statistics and addressing a problem of Aras (1988) related to sequential estimation based on U-statistics.

2. A sharp decoupling inequality and Wald's equation for U-statistics with kernels in L_p, $1 < p \leq 2$. The proof of Wald's equation for the multilinear U-statistics (1.1) given in Chow, de la Peña and Teicher (1993) does not extend to U-statistics of the general form

$$U_n = \sum_{1 \leq i_1 < \cdots < i_k \leq n} f(X_{i_1}, \ldots, X_{i_k}), \tag{2.1}$$

since the proof given there depends heavily on the multilinear features of the problem. In the case $k = 2$, de la Peña and Lai (1997) proved Wald's equation for (2.1) by using the bivariate structure and decoupling inequalities to bound $E(\max_{n \leq T} |U_n|)$. To generalize this result to general k, we need some new decoupling inequalities which will in turn enable us to exend and integrate the two different approaches in Chow, de la Peña and Teicher (1993) and in de la Peña and Lai (1997).

Throughout the sequel we let $X_1, X_2, \ldots$ be i.i.d random variables with values in a measurable space $(S, \mathcal{S})$ and let $\mathcal{F}_n$ be the σ-algebra generated by $X_1, \ldots, X_n$. Let $k \geq 2$ and let $f\colon S^k \to \mathbf{R}$ be a Borel measurable function such that

$$E\big(f(X_1, \ldots, X_k) | X_{i_1}, \ldots, X_{i_h}\big) = 0 \quad \text{for every } \{i_1, \ldots, i_h\} \subset \{1, \ldots, k\} \text{ with } h < k. \tag{2.2}$$

Define U_n by (2.1) and let T be a stopping time adapted to $\{\mathcal{F}_n\}$. The following is a sharp decoupling inequality for $E(\max_{k \leq n \leq T} |U_n|^p)$ and will be proved in Section 5.

THEOREM 1. *Suppose that* (2.2) *holds and that* $E|f(X_1, \ldots, X_k)|^p < \infty$ *for some* $1 \leq p \leq 2$. *Moreover, assume that f is symmetric in its arguments, that is,*

$$f(X_1, \ldots, X_k) = f\big(X_{\pi(1)}, \ldots, X_{\pi(k)}\big) \quad \textit{for any permutation } \pi \textit{ of } \{1, \ldots, k\}. \tag{2.3}$$

Let $\{X_n^{(1)}\},\dots,\{X_n^{(k)}\}$ *be* k *independent copies of* $\{X_n\}$. *Then there exists a universal constant* $C_{k,p}$, *depending only on* k *and* p, *such that*

$$E \max_{k\le n\le T}\Big|\sum_{1\le i_1<\cdots<i_k\le n} f(X_{i_1},\dots,X_{i_k})\Big|^p$$

$$\le C_{k,p}\sum_{h=0}^{k-1}\sum_{t=1}^{k-h} E\Big|\sum_{i_{h+1}=1}^{T}\cdots\sum_{i_{h+t}=1}^{T} f_{T,k,h}\big(X_{i_{h+1}}^{(h+1)},\dots,X_{i_{h+t}}^{(h+t)},\dots,X_{i_{h+t}}^{(k)}\big)\Big|^p, \tag{2.4}$$

where $f_{T,k,0}=f$ *and for* $h\ge 1$, $f_{T,k,h}\colon S^{k-h}\to\mathbf{R}$ *is given by*

$$f_{T,k,h}(y_{h+1},\dots,y_k)=\sum_{1\le i_1<\cdots<i_h\le T} f(X_{i_1},\dots,X_{i_h},y_{h+1},\dots,y_k). \tag{2.5}$$

This bound generalizes that in de la Peña (1992b) dealing with multilinear forms of i.i.d. random variables. In particular, since the bound obtained in that work is known to be sharp, this bound is also sharp. This fact can also be verified by observing that the results in de la Peña (1992a) imply that if the stopping time is assumed to be independent of all the variables involved, then both sides of (2.4) are of the same order of magnitude. A variant of the bound (2.4) is also proved in Section 5 and is used to prove the following extension of Wald's equation to the case of multivariate U-statistics.

THEOREM 2. *Suppose* (2.2) *and* (2.3) *hold and that for some* $1<p\le 2$,

$$E|f(X_1,\dots,X_k)|^p<\infty \quad \text{and} \quad ET^{\rho(k,p)}<\infty, \tag{2.6}$$

where $\rho(k,p)=(p^{1/(k-1)}-1)^{-1}$. *Then* $E(\max_{k\le n\le T}|U_n|)<\infty$ *and consequently* $EU_T=0$.

When $k=2$, condition (2.6) reduces to (1.3) since $\rho(2,p)=1/(p-1)$. Theorem 2 will be proved in Section 5 where it will be explained in the remarks following the proof how $\rho(k,p)$ arises naturally in the induction argument which modifies and generalizes that of Chow, de la Peña and Teicher (1993) for the multilinear case (1.1). Denoting the right-hand side of (2.1) by $U_{n,k}$ to show that the kernel of the U-statistic has k arguments (or order k), a key idea of the proof is to use decoupling inequalities to bound $E(\max_{k\le n\le T}|U_{n,k}|)$ in terms of $E(\max_{k\le n\le T}|U_{n,h+1}|)^{\alpha_h}$, where $1\le\alpha_h\le p$ and the $U_{n,h}$ represent U-statistics corresponding to certain kernels of lower order.

3. L_p-norms of randomly stopped U-statistics with $p\ge 2$. It is well known that for $p\ge 2$, $E|U_n|^p=O(n^{kp/2})$ if $E|f(X_1,\dots,X_k)|^p<\infty$. In this section we first prove an upper bound of $E(\max_{k\le n\le T}|U_n|^p)$ in terms of $ET^{kp/2}$ and then apply the result to prove uniform integrability and moment convergence. We assume that (2.2) holds as before, but the symmetry condition (2.3) is not assumed here.

THEOREM 3. *For any $p \ge 2$, there exists a universal constant $C_{k,p}$, depending only on k and p, such that*

$$E \max_{k \le n \le T} \left| \sum_{1 \le i_1 < \cdots < i_k \le n} f(X_{i_1}, \dots, X_{i_k}) \right|^p \le C_{k,p} (E T^{kp/2}) \left(E |f(X_1, \dots, X_k)|^{kp} \right)^{1/k}. \tag{3.1}$$

PROOF. Let $\{X_i^{(1)}\}, \dots, \{X_i^{(k)}\}$ be k independent copies of $\{X_i\}$. Let G_n be the σ-algebra generated by $\{X_1, \dots, X_n\} \cup \{X_i^{(1)}: i \ge 1\} \cup \cdots \cup \{X_i^{(k)}: i \ge 1\}$. In what follows, C_p and $C_{k,p}$ denote universal constants that may change from one bound to another. The explanation of different steps below is given in the remarks, labelled (A), (B), and so on, following the proof:

$$\begin{aligned}
&E\left\{ \max_{n \le T} \left| \sum_{1 \le i_1 < \cdots < i_k \le n} f(X_{i_1}, \dots, X_{i_k}) \right|^p \right\} \\
&\quad = E \sup_n \left| \sum_{i_k = k}^{n} \left\{ \sum_{i_{k-1} = k-1}^{i_k - 1} \cdots \sum_{i_1 = 1}^{i_2 - 1} f(X_{i_1}, \dots, X_{i_k}) \right\} I(T \ge i_k) \right|^p \\
&\quad \le C_p E \left\{ \sum_{i_k = k}^{\infty} \left[\sum_{i_{k-1} = k-1}^{i_k - 1} \cdots \sum_{i_1 = 1}^{i_2 - 1} f(X_{i_1}, \dots, X_{i_{k-1}}, X_{i_k}) \times I(T \ge i_k) \right]^2 \right\}^{p/2} \quad [\text{see (A)}] \\
&\quad \le C_p E \left\{ \sum_{i_k = k}^{\infty} \left[\sum_{i_{k-1} = k-1}^{i_k - 1} \cdots \sum_{i_1 = 1}^{i_2 - 1} f(X_{i_1}, \dots, X_{i_{k-1}}, X_{i_k}^{(k)}) \times I(T \ge i_k) \right]^2 \right\}^{p/2} \quad [\text{see (B)}] \\
&\quad = C_p E \left\{ \sum_{i_k = k}^{\infty} \left(E\left[\sum_{i_{k-1} = k-1}^{T} \cdots \sum_{i_1 = 1}^{i_2 - 1} f(X_{i_1}, \dots, X_{i_{k-1}}, X_{i_k}^{(k)}) \times I(T \ge i_k) \,\Big|\, G_{i_k - 1} \right] \right)^2 \right\}^{p/2} \quad [\text{see (C)}]
\end{aligned}$$

$$\le C_p E\left\{\sum_{i_k=k}^{\infty} E\left[\left(\sum_{i_{k-1}=k-1}^{T}\cdots\sum_{i_1=1}^{i_2-1} f\left(X_{i_1},\ldots,X_{i_{k-1}},X_{i_k}^{(k)}\right)\right.\right.\right.$$

$$\left.\left.\left.\times I(T\ge i_k)\right)^2\middle|G_{i_k-1}\right]\right\}^{p/2}$$

$$\le C_p E\left\{\sum_{i_k=k}^{\infty}\left[\sum_{i_{k-1}=k-1}^{T}\cdots\sum_{i_1=1}^{i_2-1} f\left(X_{i_1},\ldots,X_{i_{k-1}},X_{i_k}^{(k)}\right)\right.\right.$$

$$\left.\left.\times I(T\ge i_k)\right]^2\right\}^{p/2}\quad[\text{see (D)}]$$

$$\le C_p E\left\{\left|\sum_{i_k=k}^{T}\sum_{i_{k-1}=k-1}^{T}\sum_{i_{k-2}=k-2}^{i_{k-1}-1}\cdots\sum_{i_1=1}^{i_2-1} f\left(X_{i_1},\ldots,X_{i_{k-1}},X_{i_k}^{(k)}\right)\right|^p\right\}$$

$$[\text{see (E)}]$$

$$\le C_p E\left\{T^{p/2}\left|\sum_{i_{k-1}=k-1}^{T}\sum_{i_{k-2}=k-2}^{i_{k-1}-1}\cdots\sum_{i_1=1}^{i_2-1} f\left(X_{i_1},\ldots,X_{i_{k-1}},X_1^{(k)}\right)\right|^p\right\}$$

$$[\text{see (F)}]$$

$$\le C_p\{ET^{kp/2}\}^{1/k}$$

$$\times\left\{E\left|\sum_{i_{k-1}=k-1}^{T}\sum_{i_{k-2}=k-2}^{i_{k-1}-1}\cdots\sum_{i_1=1}^{i_2-1} f\left(X_{i_1},\ldots,X_{i_{k-2}},X_{i_{k-1}},X_1^{(k)}\right)\right|^{kp/(k-1)}\right\}^{(k-1)/k}$$

$$\le C_p\{ET^{kp/2}\}^{1/k}$$

$$\times\left\{ET^{kp/2(k-1)}\right.$$

$$\left.\times\left|\sum_{i_{k-2}=k-2}^{T}\sum_{i_{k-3}=k-3}^{i_{k-2}-1}\cdots\sum_{i_1=1}^{i_2-1} f\left(X_{i_1},\ldots,X_{i_{k-2}},X_{i_{k-1}}^{(k-1)},X_1^{(k)}\right)\right|^{kp/(k-1)}\right\}^{(k-1)/k}$$

$$[\text{see (G)}]$$

$$\le\cdots\le C_{k,p}(ET^{kp/2})\left(E\left|f\left(X_1^{(1)},X_1^{(2)},\ldots,X_1^{(k)}\right)\right|^{kp}\right)^{1/k}\quad[\text{see (H)}].\qquad\square$$

REMARKS. We explain here various inequalities and an equality that has been marked in the preceding proof.

(A) The bound follows from Burkholder's inequality [cf. Chow and Teicher (1988)]: there exists a constant B_p such that for all martingale difference sequences (d_n),

$$E \sup_n \left| \sum_{i=1}^n d_i \right|^p \le B_p E \left(\sum_{i=1}^\infty d_i^2 \right)^{p/2}.$$

(B) Apply Theorem 2 of Hitczenko (1988): for $r > 0$, there exists a constant B_r such that for all tangent sequences (d_n) and (e_n) of nonnegative random variables, $E(\sum_{i=1}^n d_i)^r \le B_r E(\sum_{i=1}^n e_i)^r$. The sequences (d_n) and (e_n) of random variables adapted to a filtration (F_n) are said to be tangent to each other if $P\{d_n \le x \mid F_{n-1}\} = P\{e_n \le x \mid F_{n-1}\}$ a.s. for all $n \ge 1$ and all real numbers x.

(C) Note that $E[\sum_{i_{k-1}=i_k}^T \cdots \sum_{i_1=1}^{i_2-1} f(X_{i_1}, \dots, X_{i_{k-1}}, X_{i_k}^{(k)}) I(T \ge i_k) \mid G_{i_k-1}] = 0$, assuming T to be bounded in (3.1), where we can first replace T by $\min(T, m)$ and then let $m \to \infty$. Hence

$$\begin{aligned} E&\left[\sum_{i_{k-1}=k-1}^T \cdots \sum_{i_1=1}^{i_2-1} f\left(X_{i_1}, \dots, X_{i_{k-1}}, X_{i_k}^{(k)} \right) I(T \ge i_k) \middle| G_{i_k-1} \right] \\ &= \sum_{i_{k-1}=k-1}^{i_k-1} \cdots \sum_{i_1=1}^{i_2-1} f\left(X_{i_1}, \dots, X_{i_{k-1}}, X_{i_k}^{(k)} \right) I(T \ge i_k). \end{aligned}$$

(D) Apply Lemma 3 in Section 11.3 of Chow and Teicher (1988), recalling that $p \ge 2$: for $r \ge 1$, there exists a constant B_r such that for every sequence of nonnegative random variables (z_n) and every filtration (F_n), $E\{\sum_{i=1}^\infty E(z_i \mid F_{i-1})\}^r \le B_r E(\sum_{i=1}^\infty z_i)^r$.

(E) The bound follows from Burkholder's inequality, assuming the stopping time T to be bounded in (3.1): there exists a constant A_p such that for all martingale differences (d_n), $E|\sum_{i=1}^n d_i|^p \ge A_p E(\sum_{i=1}^n d_i^2)^{p/2}$. See Chow and Teicher (1988), page 396.

(F) Condition on T and $\{X_i\}$ and use the bound $E|\sum_{i=1}^n Z_i|^p \le A_p n^{p/2} E|Z_1|^p$ for i.i.d. zero-mean random variables $Z_1, Z_2, \dots (p \ge 2)$.

(G) Apply the same argument as in (F) but with p replaced by $kp/(k-1)$.

(H) Proceed by induction. First note that by Hölder's inequality relating the expectation of the product of two nonnegative random variables to their L_{k-1} and $L_{(k-1)/(k-2)}$ norms,

$$\left\{ E\left(T^{kp/2(k-1)} |Z|^{kp/(k-1)} \right) \right\}^{(k-1)/k} \le \left(ET^{kp/2} \right)^{1/k} \left(E|Z|^{kp/(k-2)} \right)^{(k-2)/k},$$

in which

$$Z = \sum_{i_{k-2}=k-2}^T \cdots \sum_{i_1=1}^{i_2-1} f\left(X_{i_1}, \dots, X_{i_{k-2}}, X_{i_{k-1}}^{(k-1)}, X_1^{(k)} \right).$$

Then apply the same argument as before to $E|Z|^{kp/(k-2)}$. Repeated use of this argument gives the final result.

Let $m \ge k$. If we replace $\sum_{i_k=k}^n$ in the chain of inequalities in the proof of Theorem 3 by $\sum_{i_k=m}^n$, then we can replace $T^{p/2}$ by $T^{p/2} I(T \ge m)$ in the inequality marked by reference to Remark (F) and $(ET^{kp/2})^{1/k}$ by

$(ET^{kp/2}I(T \geq m))^{1/k}$ in the subsequent inequality, so the preceding proof also gives the following variant of (3.1):

$$E \sup_{n \geq m} \left| \sum_{i_k = m}^{n} \left\{ \sum_{1 \leq i_1 < \cdots < i_{k-1} < i_k} f(X_{i_1}, \dots, Xi_{i_k}) \right\} I(T \geq i_k) \right|^p \\ \leq C_{k,p} \{ET^{kp/2} I(T \geq m)\}^{1/k} (ET^{kp/2})^{(k-1)/k} \\ \times \left(E|f(X_1, \dots, X_k)|^{kp}\right)^{1/k}. \tag{3.2}$$

If $E|f(X_1, \dots, X_k)|^r < \infty$ for some $r \geq 2$, then by Lemma 1 of Lai and Wang (1993) together with Doob's inequality for martingales,

$$E \max_{n \leq m} \left| \sum_{1 \leq i_1 < \cdots < i_k \leq n} f(X_{i_1}, \dots, X_{i_k}) \right|^r \leq C_{k,r} m^{kr/2} E|f(X_1, \dots, X_k)|^r. \tag{3.3}$$

We now give a corollary of (3.2) and (3.3) that will enable us to prove moment convergence in Anscombe's (1952) theorem for U-statistics. Let $p \geq 2$. For $A \subset (0, \infty)$ and i.i.d. random variables $Z_1, Z_2, \dots$ with $EZ_1 = 0$ and $E|Z_1|^p < \infty$, Chow and Yu [(1981), Lemma 5] have shown that if $\{T(a),\ a \in A\}$ is a family of stopping times such that $\{(a^{-1}T(a))^{p/2},\ a \in A\}$ is uniformly integrable, then $\{|\sum_{i=1}^{T(a)} Z_i/\sqrt{a}|^p,\ a \in A\}$ is uniformly integrable. This is extended to the U-statistics (2.1) in the following.

COROLLARY 1. *Let $p \geq 2$ and $A \subset (0, \infty)$. Suppose that $E|f(X_1, \dots, X_k)|^{kp} < \infty$ and that $\{T(a),\ a \in A\}$ is a family of stopping times adapted to $\{F_n\}$. If $\{(a^{-1}T(a))^{kp/2}a,\ a \in A\}$ is uniformly integrable, then $\{|a^{-k/2}U_{T(a)}|^p,\ a \in A\}$ is uniformly integrable.*

PROOF. Let $m_{a,\lambda} = [\lambda a]$ and note that

$$|U_{T(a)}| \leq \max_{n \leq m_{a,\lambda}} \left| \sum_{1 \leq i_1 < \cdots < i_k \leq n} f(X_{i_1}, \dots, X_{i_k}) \right| \\ + \left| \sum_{i_k = m_{a,\lambda}}^{\infty} I(T(a) \geq i_k) \sum_{1 \leq i_1 < \cdots < i_{k-1} < i_k} f(X_{i_1}, \dots, X_{i_k}) \right|. \tag{3.4}$$

Given $\varepsilon > 0$, we can choose λ sufficiently large so that

$$\left\{E\left(a^{-1}T(a)\right)^{kp/2} I\left(T(a) \geq m_{a,\lambda}\right)\right\}^{1/k} \leq \varepsilon \quad \text{for all } a \in A.$$

Moreover, $\sup_{a \in A} E(a^{-1}T(a))^{kp/2} = B < \infty$. Hence by (3.2), for all $a \in A$,

$$E\left| \sum_{i_k = m_{a,\lambda}}^{\infty} I(T(a) \geq i_k) \sum_{1 \leq i_1 < \cdots < i_{k-1} < i_k} f(X_{i_1}, \dots, X_{i_k}) \right|^p \\ \leq C_{k,p} B^{(k-1)/k} \left(E|f(X_1, \dots, X_k)|^{kp}\right)^{1/k} \varepsilon a^{kp/2}.$$

Therefore in view of (3.4), the uniform integrability of $|a^{-k/2}U_{T(a)}|^p$ will follow if it can be shown that for every $\lambda > 0$,

$$\left\{ \left(\max_{n \le m_{a,\lambda}} \left| \sum_{1 \le i_1 < \cdots < i_k \le n} f(X_{i_1}, \dots, X_{i_k}) \right| \Big/ m_{a,\lambda}^{k/2} \right)^p,\ a \in A \right\} \tag{3.5}$$

is uniformly integrable. The bound (3.3) with $r = kp(> p)$ implies the uniform integrability of (3.5), completing the proof of Corollary 1. □

Suppose $Ef^2(X_1, \dots, X_k) < \infty$ and f is symmetric in its arguments. Then $\{k!n^{-k/2}U_{[nt]},\ 0 \le t \le T\}$ converges weakly to a multiple Wiener integral $\{W_k(t),\ 0 \le t \le T\}$ with respect to some Gaussian random measure [cf. Mandelbaum and Taqqu (1984)]. Let $\{T(a),\ a \ge a_0\}$ be a family of $\{F_n\}$-adapted stopping times such that $T(a)/a \to_P c$ for some constant $c > 0$. Then by Anscombe's (1952) theorem, $a^{-k/2}U_{T(a)}$ converges in distribution to $c^{k/2}W_k(1)/k!$. Corollary 1 yields the following result on moment convergence of $a^{-k/2}U_{T(a)}$.

Corollary 2. *Let* $p \ge 2$. *Suppose that* f *satisfies* (2.3) *and* $E|f(X_1, \dots, X_k)|^{kp} < \infty$. *Let* $T(a)$ *be* $\{F_n\}$*-adapted stopping times such that* $\lim_{a\to\infty} E|a^{-1}T(a) - c|^{kp/2} = 0$ *for some constant c. Then for every continuous function* f: $\mathbf{R} \to \mathbf{R}$ *satisfying* $f(x) = O(|x|^p)$ *as* $|x| \to \infty$,

$$\lim_{a\to\infty} Ef\left(a^{-k/2}U_{T(a)}\right) = Ef\left(c^{k/2}W_k(1)/k!\right). \tag{3.6}$$

4. Moments of normalized *U*-statistics in sequential analysis. Let g: $S^k \to \mathbf{R}$ be a Borel function of k variables such that g is symmetric in its arguments [i.e., (2.3) holds with g in place of f] and

$$E\left|g(X_1, \dots, X_k)\right|^\beta < \infty \quad \text{for some } \beta > 2. \tag{4.1}$$

Let $\theta = Eg(X_1, \dots, X_k)$. For $n \ge k$, the normalized U-statistic with kernel g is

$$\hat{\theta}_n = \left\{ \sum_{1 \le i_1 < \cdots < i_k \le n} g(X_{i_1}, \dots, X_{i_k}) \right\} \Big/ \binom{n}{k}. \tag{4.2}$$

Normalized U-statistics were introduced by Halmos (1946) to provide unbiased estimates of θ. Hoeffding (1948, 1961) subsequently proved the asymptotic normality of $\hat{\theta}_n$ by using the decomposition

$$\begin{aligned} \hat{\theta}_n - \theta = {} & kn^{-1} \sum_{i=1}^n g_1(X_i) + k(k-1)\{n(n-1)\}^{-1} \sum_{1 \le i < j \le n} g_2(X_i, X_j) \\ & + \cdots + k!\{n \cdots (n-k+1)\}^{-1} \sum_{1 \le i_1 < \cdots < i_k \le n} g_k(X_{i_1}, \dots, X_{i_k}), \end{aligned} \tag{4.3}$$

where

$$\begin{aligned} g_1(x) &= Eg(x, X_2, \dots, X_k) - \theta, \\ g_2(x, y) &= Eg(x, y, X_3, \dots, X_k) - g_1(x) - g_1(y) + \theta \quad \text{and so on.} \end{aligned} \tag{4.4}$$

Thus, g_j is symmetric in its arguments and

$$(4.5)\qquad \begin{gathered} E\left|g_j(X_1,\dots,X_j)\right|^\beta<\infty,\qquad Eg_1(X_1)=0,\\ E\left\{g_j(X_1,\dots,X_j)\mid X_{i_1},\dots,X_{i_h}\right\}=0 \end{gathered}$$

for every proper subset $\{i_1,\dots,i_h\}$ of $\{1,\dots,j\}$.

Although $\hat{\theta}_n$ is an unbiased estimate of θ, the corresponding normalized U-statistic $\hat{\theta}_T$ based on a sample $\{X_1,\dots,X_T\}$ from a sequential experiment, in which the sample size T is not fixed in advance but is sequentially determined from the current and past data, is biased. In addition to the bias, the mean squared error $E(\hat{\theta}_T-\theta)^2$ is also of fundamental interest in the theory of sequential estimation.

In this section we consider sequential experiments whose stopping rules are of the form

$$(4.6)\qquad T=\inf\left\{n\ge k:\ \sum_{i=1}^n Y_i+\xi_n\ge a\right\},$$

where $(X_1,Y_1),(X_2,Y_2),\dots$ are i.i.d. random vectors and for some $c>0$,

$$(4.7)\qquad EY_1=\mu>0,\qquad EY_1^2<\infty,\qquad \sum_{n=1}^\infty nP\{\xi_n+\mu n<cn\}<\infty,$$

$$(4.8)\qquad \left\{\left[\left(\sum_{i=1}^n Y_i+\xi_n-n/c\right)^+\right]^p,\ n\ge 1\right\}$$

is uniformly integrable for some $p\ge 3$,

$$(4.9)\qquad \lim_{\delta\to 0}\sup_{n\ge 1}P\left\{\max_{k\le n\delta}|\xi_{n+k}-\xi_n|>\varepsilon\right\}=0\quad\text{for all }\varepsilon>0.$$

In addition, it is assumed that there are events A_n such that for some $\alpha\ge 3/2$,

$$(4.10)\qquad \begin{gathered} \sum_{n=1}^\infty nP\left(\bigcup_{k=n}^\infty A_k'\right)<\infty\quad\text{and}\\ \left\{\max_{k\le n}|\xi_{n+k}|^\alpha I(A_{n+k}),\ n\ge 1\right\}\text{ is uniformly integrable,} \end{gathered}$$

where A_k' denotes the complement of A_k.

Stopping rules of the form (4.6) arise naturally in many sequential testing and estimation problems [cf. Woodroofe (1982), Siegmund (1985)], in which stopping occurs whenever some statistic Z_n exceeds a threshold a. Typically Z_n can be written in terms of a random walk $\sum_{i=1}^n Y_i$ plus a remainder term ξ_n. Regularity conditions on ξ_n like those above were first introduced by Lai and Siegmund (1977) in the development of a renewal theory for the perturbed random walk $\sum_1^n Y_i+\xi_n$. The conditions (4.7)–(4.10) were formulated by Aras and Woodroofe (1993) in their work on asymptotic expansions for the first four moments of randomly stopped sample means in sequential analysis. The following theorem extends their results from sample means to normalized U-statistics.

THEOREM 4. *Suppose* $g: S^k \to \mathbf{R}$ *satisfies the symmetry condition* (2.3) (*with* g *in place of* f). *Assume* (4.1) *and* (4.7)–(4.10). *Define* $\hat{\theta}_n$ *by* (4.2), $T(=T_a)$ *by* (4.6), g_1 *and* g_2 *by* (4.4).

(i) *If* $\beta \geq \max\{4, 2\alpha/(\alpha-1), 2p/(p-2)\}$, *then as* $a \to \infty$,

$$E\hat{\theta}_T = \theta + a^{-1}k\{EY_1 g_1(X_1) + o(1)\}, \tag{4.11}$$

$$E\left(\hat{\theta}_T - \theta\right)^2 = a^{-1}\mu k^2 E g_1^2(X_1) + O(a^{-2}). \tag{4.12}$$

(ii) *If* $\beta \geq \max\{4, 6\alpha/(2\alpha-1), 3p/(p-2)\}$, *then*

$$\lim_{a\to\infty} a^2 E\left(\hat{\theta}_T - \theta\right)^3/k^3 = 6\mu\left(Eg_1^2(X_1)\right)\left(EY_1 g_1(X_1)\right) + \mu^2 E g_1^3(X_1). \tag{4.13}$$

If $\beta \geq \max\{4, 4p/(p-2)\}$, *then*

$$\lim_{a\to\infty} a^2 E\left(\hat{\theta}_T - \theta\right)^4/k^4 = 3\mu^2\left(Eg_1^2(X_1)\right)^2. \tag{4.14}$$

(iii) *Let* $g_{12}(x) = E\{g_1(X_1)g_2(x, X_1)\}$, $S_n = \sum_{i=1}^n Y_i$ *and* $\tau_+ = \inf\{n: S_n > 0\}$. *Suppose* $\beta \geq \max\{4, 2\alpha/(\alpha-1), 2p/(p-2)\}$, Y_1 *is nonlattice and* $(n^{-1/2}\sum_{i=1}^n (g(X_i), Y_i), \xi_n)$ *converges weakly to* (Z_X, Z_Y, ξ). *Then* (Z_X, Z_Y) *is normal and as* $a \to \infty$,

$$\left(T^{-1/2}\sum_{i=1}^T \left(g(X_i), Y_i\right), \xi_T, S_T + \xi_T - a\right) \tag{4.15}$$

converges weakly to (Z_X, Z_Y, ξ, ρ_+),

where ρ_+ *is independent of* (Z_X, Z_Y, ξ) *and has distribution function*

$$P\{\rho_+ \leq u\} = \int_0^u P\{S_{\tau+} > x\}\, dx/ES_{\tau+} \quad \text{for } u > 0. \tag{4.16}$$

Moreover, $ET = a/\mu + (E\rho_+ - E\xi)/\mu + o(1)$ *and* (4.12) *can be strengthened to*

$$\begin{aligned}
&\lim_{a\to\infty}\left\{a^2 E\left(\hat{\theta}_T - \theta\right)^2 - a\mu k^2 E g_1^2(X_1)\right\} \\
&\quad = \mu^2 k^2 (k-1)^2 E g_2^2(X_1, X_2)/2 + 4\mu k^2(k-1) EY_1 g_{12}(X_1) \\
&\qquad + 2k^2\left\{\mu E(Y_1 - \mu) g_1^2(X_1) + \left(\mu EY_1 g_1(X_1)\right)^2\right\} \\
&\qquad + k^2\left\{\operatorname{Var} Y_1 + 2E\xi Z_X^2 + \mu(E\rho_+ - E\xi) - 2E\rho_+\right\} E g_1^2(X_1).
\end{aligned} \tag{4.17}$$

The second-order expansion (4.17) of the mean squared error $E(\hat{\theta}_T - \theta)^2$ is of fundamental interest in the theory of sequential estimation based on U-statistics, extending corresponding results in the sample mean case [cf. Aras and Woodroofe (1993) and the references therein]. In his study of sequential estimation based on U-statistics, Aras (1988) was able to show that $E(\hat{\theta}_T - \theta)^2 = O(a^{-1})$ and pointed out the main obstacles in his method to the derivation of a more precise asymptotic formula for $E(\hat{\theta}_T - \theta)^2$.

The proof of Theorem 4 uses Corollary 1 together with arguments similar to those of Aras and Woodroofe (1993) (AW henceforth). The following four lemmas, which use the same notation and assumptions as in Theorem 4, provide the modifications needed in extending AW to U-statistics.

LEMMA 1. *Let*

$$\bar{g}_{j,n} = k(k-1)\cdots(k-j+1) \sum_{1 \le i_1 < \cdots < i_j \le n} \frac{g_j(X_{i_1}, \dots, X_{i_j})}{n(n-1)\cdots(n-j+1)}.$$

Then for every $\eta > 0$,

$$E\Big(\sup_{n \ge \eta a} |\bar{g}_{j,n}|^r\Big) = o(a^{-2}) \quad \text{for } j \ge 2 \text{ and } \beta \ge r \ge 3,$$

$$E\Big(\sup_{n \ge \eta a} \bar{g}_{j,n}^2\Big) = O(a^{-j}) \quad \text{for } j \ge 1.$$

PROOF. Since $\{\bar{g}_{j,n},\ n \ge j\}$ is a reverse martingale [cf. Gram and Serfling (1973)], it follows from Doob's inequality that

$$\begin{aligned} E\Big(\sup_{n \ge m} |\bar{g}_{j,n}|^r\Big) &= O\big(E|\bar{g}_{j,m}|^r\big) \\ &= O\Bigg(m^{-jr} E\Bigg|\sum_{1 \le i_1 < \cdots < i_j \le m} g_j(X_{i_1}, \dots, X_{i_j})\Bigg|^r\Bigg) \qquad (4.18) \\ &= O(m^{-jr/2}), \end{aligned}$$

where the last relation above follows from Lemma 1 of Lai and Wang (1993). Putting $m = [\eta a]$ in (4.18) gives the desired conclusions. □

LEMMA 2. *Let* $\phi(x, y, z) = g_1(x)g_2(y, z) + g_1(y)g_2(x, z) + g_1(z)g_2(x, y)$, $\psi(x, y) = g_1(x)g_2(x, y) + g_1(y)g_2(x, y) - g_{12}(x) - g_{12}(y)$. *Then* $Eg_{12}(X_1) = 0$ *and for* $n \ge 2$,

$$\begin{aligned} &\Big(\sum_{l=1}^n g_1(X_l)\Big)\Big(\sum_{1 \le i < j \le n} g_2(X_i, X_j)\Big) \\ &\quad = \sum_{1 \le i < j < l \le n} \phi(X_i, X_j, X_l) + \sum_{1 \le i < j \le n} \psi(X_i, X_j) + (n-1)\sum_{i=1}^n g_{12}(X_i). \end{aligned}$$

LEMMA 3. *Define* ϕ *as in Lemma* 2 *and let*

$$U_n = \sum_{1 \le i < j < l \le n} \phi(X_i, X_j, X_l).$$

Then $EU_T/\{T^2(T-1)\} = o(a^{-2})$.

PROOF. By Proposition 2(iii) of AW, $P\{T \geq 5a/c\} = o(a^{-2})$, where c is given by (4.7) and (4.8) and will be assumed to be less than $\min(\mu, 1/\mu)$ without loss of generality. Let $\overline{U}_n = U_n/\{n^2(n-1)\}$. Then $\{E(\sup_{n \geq 5a/c} \overline{U}_n^2)\}^{1/2} = O(a^{-3/2})$ as in (4.18), and therefore by the Schwarz inequality,

$$E\big|U_T/\big(T^2(T-1)\big)\big|I(T \geq 5a/c) \leq \left\{E\left(\sup_{n \geq 5a/c} \overline{U}_n^2\right)P(T \geq 5a/c)\right\}^{1/2}$$
$$= o(a^{-5/2}).$$

Let $\tau = \min\{T, [5a/c]\}$. Noting that $\overline{U}_T - \overline{U}_\tau = (\overline{U}_T - \overline{U}_{[5a/c]})I(T > 5a/c)$, the preceding argument also shows that

$$E|\overline{U}_T - \overline{U}_\tau| = o(a^{-5/2}). \tag{4.19}$$

Since the stopping time τ is bounded by $5a/c$, $E\tau^2(\tau-1)\overline{U}_\tau = EU_\tau = 0$ and therefore

$$a^2 E\overline{U}_\tau = aE\left\{\left(a - \frac{\mu^3\tau^2(\tau-1)}{a^2}\right)\overline{U}_\tau\right\}. \tag{4.20}$$

Using the property $E\sup_{n \geq m}|\overline{U}_n|^\beta = O(m^{-3\beta/2})$ with $\beta \geq 2p/(p-2)$ and an argument similar to the proof of Proposition 6 of AW, it can be shown that

$$\begin{aligned} &aE\big\{\big|a - \mu^3\tau^2(\tau-1)/a^2\big||\overline{U}_\tau|I(\tau \leq ca/5)\big\} \\ &\quad \leq a^2(1+\mu^3c^3)E\big\{|\overline{U}_\tau|I(\tau \leq ca/5)\big\} \to 0. \end{aligned} \tag{4.21}$$

We next show that

$$\big\{a\big|a - \mu^3\tau^2(\tau-1)/a^2\big||\overline{U}_\tau|I(ca/5 \leq \tau),\ a \geq 1\big\} \quad \text{is uniformly integrable.} \tag{4.22}$$

First note that on $\{ca/5 \leq \tau(\leq 5a/c)\}$,

$$\begin{aligned} \big|a - \mu^3\tau^2(\tau-1)/a^2\big| &\leq |a - \mu\tau| + \mu\tau|a^2 - \mu^2\tau^2|/a^2 + \mu^3\tau^2/a^2 \\ &\leq |a-\mu\tau|\big\{1 + 5\mu c^{-1}\big(1 + 5\mu c^{-1}\big)\big\} + 25\mu^3c^{-2}, \end{aligned} \tag{4.23}$$

$$|\overline{U}_\tau| \leq \sup_{n \geq ca/5}|\overline{U}_n|. \tag{4.24}$$

Since $\tau \leq 5a/c$, it follows from Proposition 8 of AW that $\{(|a-\mu\tau|/\sqrt{a})^{\beta'}, a \geq 1\}$ is uniformly integrable, where $\beta' = \min(4, 2\alpha, p)$. Moreover, $E^{1/r}\sup_{n \geq ca/5}|\overline{U}_n|^r = O(a^{-3/2})$ for all $2 \leq r \leq \beta$. Hence by Hölder's inequality,

$$E^{1/s}\left\{\big(|(a-\mu\tau)/\sqrt{a}| + 1\big)\sup_{n \geq ca/5}|\overline{U}_n|\right\}^s = O(a^{-3/2})$$

for some $s > 1$, implying that $\{a(\sqrt{a} + |a - \mu\tau|)\sup_{n \geq ca/5}|\overline{U}_n|, a \geq 1\}$ is uniformly integrable. This together with (4.23) yields (4.22).

Let $Y_i' = Y_i - \mu$ and $S_n' = \sum_{i=1}^n Y_i'$. Then as shown in AW, $(a-\mu\tau)/\sqrt{a} = S_\tau'/\sqrt{a} + o_p(1)$ and $\tau/a \to 1/\mu$ a.s. Moreover, $a^{3/2}\overline{U}_\tau = (a^{-3/2}U_\tau)(a/\tau)^2 a/$

$(\tau - 1)$. Hence as $a \to \infty$, $((a - \mu\tau)/\sqrt{a},\ a^{3/2}\overline{U}_\tau)$ has the same limiting distribution as that of $(S_n'/\sqrt{n\mu},\ (\mu/n)^{3/2}\sum_{1 \le i < j < l \le n} \phi(X_i, X_j, X_l))$ (as $n \to \infty$), whose first component has a limiting normal distribution and whose second component converges in distribution to a random variable that can be expressed as a multiple Wiener integral (see Section 3). Therefore by (4.22),

$$
\begin{aligned}
&\lim_{a\to\infty} aE\left\{\left(a - \mu(\tau-1)\frac{\mu^2\tau^2}{a^2}\right)\overline{U}_\tau\right\} \\
&\quad = \lim_{a\to\infty} E\left\{\left(a^{3/2}\overline{U}_\tau\right)\left(\frac{a-\mu\tau}{\sqrt{a}}\right)\right\} \qquad (4.25)\\
&\quad = \lim_{n\to\infty} \mu n^{-2} E\left\{\left(\sum_{h=1}^n Y_h'\right)\left(\sum_{1\le i<j<l\le n}\phi(X_i, X_j, X_l)\right)\right\} = 0,
\end{aligned}
$$

since $E\{Y_h' g_1(X_i) g_2(X_j, X_l)\} = 0$ for $i < j < l$. Combining (4.25) with (4.20) yields $\lim_{a\to\infty} a^2 E\overline{U}_\tau = 0$, which implies the desired conclusion in view of (4.19). □

LEMMA 4. *For* $2 \le r \le \beta$, $E(\hat\theta_T - \theta)^r I(T \le ca/5) = o(a^{-p(1-r/\beta)})$.

PROOF. In view of (4.3) and (4.18), $E(\sup_{n\ge m}|\hat\theta_n - \theta|^r) = O(m^{-r/2})$, as in the sample mean case considered by AW. Hence the desired conclusion can be proved by using the same argument as that used to prove Proposition 6 of AW. □

PROOF OF THEOREM 4. The proof relies on the decomposition (4.3); in fact, the results involve g_1 and g_2 in the decomposition. We again assume that $c < \min(\mu, 1/\mu)$ without loss of generality.

The argument of de la Peña and Lai (1997) for proving (4.11) for the case $k = 2$ can be easily extended to general k by making use of Lemma 1. In view of Lemmas 1 and 4, (4.13) and (4.14) follow from corresponding results for sample means in AW. To prove (4.12) and (4.17), application of Lemmas 4 and 1 to (4.3) shows that

$$
\begin{aligned}
E\left(\hat\theta_T - \theta\right)^2 &= E\left\{\left(\bar g_{1,T}^2 + \bar g_{2,T}^2 + 2\bar g_{1,T}\bar g_{2,T} + 2\bar g_{1,T}\bar g_{3,T}\right)\right\} \\
&\qquad \times I(T - 1 \ge [ca/5]_*) + o(a^{-2}) \qquad (4.26)\\
&= E\bar g_{1,\tilde T}^2 + E\bar g_{2,\tilde T}^2 + 2E\bar g_{1,\tilde T}\bar g_{2,\tilde T} + 2E\bar g_{1,\tilde T}\bar g_{3,\tilde T} + o(a^{-2}),
\end{aligned}
$$

where $\tilde T = \max\{T, 1 + [ca/5]_*\}$ and $[x]_*$ denotes the smallest integer greater than or equal to $ca/5$. We shall show that as $a \to \infty$,

$$a^2 E\bar g_{2,\tilde T}^2 \to \mu^2 k^2 (k-1)^2 E g_2^2(X_1, X_2)/2, \qquad (4.27)$$

$$a^2 E\bar g_{1,\tilde T}\bar g_{2,\tilde T} \to 2\mu k^2 (k-1) E Y_1 g_{12}(X_1), \qquad (4.28)$$

$$a^2 E\bar g_{1,\tilde T}\tilde g_{3,\tilde T} \to 0. \qquad (4.29)$$

From (8) and Proposition 6 of AW applied to $E\bar{g}^2_{1,\tilde{T}}$ [with the necessary modifications, since AW considers $\mu = 1$ and more restricted forms of $(Y_i, g_1(X_i))$] together with (4.26)–(4.29), (4.12) follows. Likewise (4.17) follows from (9) and Proposition 6 of AW together with (4.26)–(4.29).

To prove (4.27), we first show that

$$\{a^2\bar{g}^2_{2,\tilde{T}},\ a \ge 1\} \text{ is uniformly integrable.} \tag{4.30}$$

This follows from Corollary 1, noting that $\tilde{T} - 1 \ge ca/5$ implies

$$a^2\bar{g}^2_{2,\tilde{T}} \le (5/c)^4 k^2(k-1)^2 \left\{ a^{-1} \sum_{1 \le i < j \le \tilde{T}} g_2(X_i, X_j) \right\}^2.$$

Since $\tilde{T}/a \to 1/\mu$ a.s. as $a \to \infty$, $\mu^2(a/\mu)^2\bar{g}^2_{2,\tilde{T}}$ has the same limiting distribution as $\{k(k-1)\}^2\mu^2 n^{-2}\sum_{1 \le i < j \le n} g_2(X_i, X_j)$ (as $n \to \infty$), and the limiting distribution can be expressed in terms of a multiple Wiener integral. Therefore by (4.30),

$$\begin{aligned}\lim_{a\to\infty} a^2 E\bar{g}^2_{2,\tilde{T}} &= \{k(k-1)\}^2\mu^2 \lim_{n\to\infty} n^{-2}E\left(\sum_{1\le i<j\le n} g_2(X_i, X_j)\right)^2 \\ &= \tfrac{1}{2}\mu^2 k^2(k-1)^2 E g_2^2(X_1, X_2).\end{aligned}$$

To prove (4.28), we apply the representation in Lemma 2 to

$$\bar{g}_{1,\tilde{T}}\bar{g}_{2,\tilde{T}} = k^2(k-1)\left\{\sum_{l=1}^{\tilde{T}} g_1(X_l)\right\}\left\{\sum_{1\le i<j\le\tilde{T}} g_2(X_i, X_j)\right\} \Big/ \{\tilde{T}^2(\tilde{T}-1)\}.$$

An argument similar to the proof of (4.30) shows that

$$\left\{\left(a^{-1}\sum_{1\le i<j\le\tilde{T}} \psi(X_i, X_j)\right) a^3/\left(\tilde{T}^2(\tilde{T}-1)\right),\ a \ge 1\right\}$$

is uniformly integrable, from which it follows that

$$\lim_{a\to\infty} a^2 E\left\{\sum_{1\le i<j\le\tilde{T}} \psi(X_i, X_j)/\left(\tilde{T}^2(\tilde{T}-1)\right)\right\} = 0, \tag{4.31}$$

noting that $E\psi(X_1, X_2) = 0$. By Proposition 2(iii) of AW, $P\{T \ge 5a/c\} = o(a^{-2})$, and therefore by the Schwarz inequality,

$$\begin{aligned} &E\tilde{T}^{-2}\left|\sum_{i=1}^{\tilde{T}} g_{12}(X_i)\right| I(T \ge 5a/c) \\ &\quad \le E\left\{T^{-1}I(T \ge 5a/c) \sup_{n\ge 5a/c} n^{-1}\left|\sum_{i=1}^{n} g_{12}(X_i)\right|\right\} \\ &\quad \le ca^{-1}P^{1/2}(T \ge 5a/c)\, E^{1/2}\left(\sup_{n\ge 5a/c}\left|n^{-1}\sum_{i=1}^{n} g_{12}(X_i)\right|^2\right) = o(a^{-2}).\end{aligned}$$

Hence, letting $\tilde{\tau} = \min\{\tilde{T}, [5a/c]\}$, we have

$$a^2 E\left\{\tilde{T}^{-2}\sum_{i=1}^{\tilde{T}} g_{12}(X_i)\right\} = a^2 E\left\{\tilde{\tau}^{-2}\sum_{i=1}^{\tilde{\tau}} g_{12}(X_i)\right\} + o(1)$$
$$(4.32)\qquad = E\left\{\left(a^2/\tilde{\tau}^2 - \mu^2\right)\sum_{i=1}^{\tilde{\tau}} g_{12}(X_i)\right\} + o(1),$$

where the last equality follows from Wald's equation $E\sum_{i=1}^{\tilde{\tau}} g_{12}(X_i) = 0$. Since $ca/5 < \tilde{\tau} \le 5a/c$,

$$|a^2/\tilde{\tau}^2 - \mu^2|\left|\sum_{i=1}^{\tilde{\tau}} g_{12}(X_i)\right|$$
$$< 5c^{-1}\left(5c^{-1} + \mu\right)\left|a^{-1/2}(a - \mu\tilde{\tau})\right| a^{-1/2}\max_{m \le 5a/c}\left|\sum_{i=1}^{m} g_{12}(X_i)\right|.$$

Since $\{(a^{-1/2}(a - \mu\tilde{\tau}))^2,\ a \ge 1\}$ is uniformly integrable by Proposition 8 of AW (recalling that $\tilde{\tau} \le 5a/c$) and since $\{n^{-1}\max_{m \le n}(\sum_{i=1}^m g_{12}(X_i))^2,\ n \ge 1\}$ is uniformly integrable, it then follows that $\{|a^2/\tilde{\tau}^2 - \mu^2|\,|\sum_{i=1}^{\tilde{\tau}} g_{12}(X_i)|,\ a \ge 1\}$ is uniformly integrable. Since $\tilde{\tau}/a \to 1/\mu$ a.s. and $a - \mu\tilde{\tau} = \sum_{i=1}^{\tilde{\tau}}(Y_i - \mu) + o_p(\sqrt{a})$, the preceding uniform integrability yields

$$\lim_{a\to\infty} E\left\{\left(a^2/\tilde{\tau}^2 - \mu^2\right)\sum_{i=1}^{\tilde{\tau}} g_{12}(X_i)\right\}$$
$$(4.33)\qquad = 2\mu \lim_{n\to\infty} E\left\{n^{-1/2}\sum_{i=1}^{n}(Y_i - \mu)\right\}\left\{n^{-1/2}\sum_{j=1}^{n} g_{12}(X_j)\right\}$$
$$= 2\mu E(Y_1 - \mu) g_{12}(X_1) = 2\mu E Y_1 g_{12}(X_1).$$

From Lemmas 2, 3 and (4.31)–(4.23), (4.28) follows.

The proof of (4.29) is similar, replacing $g_{12}(x)$ by $g_{13}(x, y) = Eg_1(X_1)g_3(x, y, X_1)$ and $\psi(x, y)$ in Lemma 2 by

$$\tilde{\psi}(x, y, z) = g_1(x)g_3(x, y, z) + g_1(y)g_3(x, y, z) + g_1(z)g_3(x, y, z)$$
$$- g_{13}(y, z) - g_{13}(x, z) - g_{13}(x, y),$$

and noting that $E(g_{13}(X_1, X_2) \mid X_1) = 0 = E(g_{13}(X_1, X_2) \mid X_2)$. □

5. A sharp decoupling inequality and Wald's equation for randomly stopped *U*-statistics. This section is divided into three subsections. In Section 5.1 we introduce some background results needed in the proofs. Section 5.2 contains the proof of Theorem 1 and Section 5.3 the proof of Theorem 2.

5.1. *The K-function and a symmetrization lemma.* The proof of Theorem 1 uses the following lemmas. Lemma 5 states certain results of Klass (1976, 1981) on his K-function while Lemma 6 provides a symmetrization inequality. In Lemma 6 and elsewhere we shall use the notation $i_1 \neq \cdots \neq i_k$ to denote that $i_r \neq i_s$ whenever $r \neq s$.

LEMMA 5. *Let $Z_1, Z_2, \dots$ be i.i.d. random variables with $EZ_1 = 0$. Let $1 \le p \le 2$. There exist universal constants c_p and C_p, depending only on p, and an increasing function $K_p(n)$ depending on p and the distribution of Z_1 such that*

$$c_p K_p(n) \le E\left|\sum_{i=1}^n Z_i\right|^p \le C_p K_p(n). \tag{5.1}$$

Moreover, $K_p(n)/n^{p/2}$ increases with n while $K_p(n)/n$ decreases with n. Consequently, for all $m \le n$,

$$E\left|\sum_{i=1}^m Z_i\right|^p \le (C_p/c_p)(m/n)^{p/2} E\left|\sum_{i=1}^n Z_i\right|^p, \tag{5.2}$$

$$E\left|\sum_{i=1}^n Z_i\right|^p \le (C_p/c_p) nE|Z_1|^p. \tag{5.3}$$

LEMMA 6. *With the same notation and assumptions as in Theorem 1, there exist universal constants $A_{k,p}$ and $B_{k,p}$, depending only on k and p, such that*

$$\begin{aligned} &E\left|\sum_{1\le i_1\neq\cdots\neq i_k\le n} f(X_{i_1},\dots,X_{i_k})\right|^p \\ &\quad\le A_{k,p} E\left|\sum_{1\le i_1\neq\cdots\neq i_k\le n} f\left(X_{i_1}^{(1)},\dots,X_{i_k}^{(k)}\right)\right|^p \\ &\quad\le B_{k,p}\sum_{t=1}^k E\left|\sum_{i_1=1}^n\cdots\sum_{i_t=1}^n f\left(X_{i_1}^{(1)},\dots,X_{i_t}^{(t)},\dots,X_{i_t}^{(k)}\right)\right|^p. \end{aligned} \tag{5.4}$$

PROOF. The first inequality is from de la Peña (1992a). As for the second inequality, the proof follows easily by using the i.i.d. property of the random variables, together with the triangle inequality for L_p norms. We only show how to do this in the case $k = 3$ since the general case follows similarly:

$$\begin{aligned} &E\left|\sum_{1\le i_1\neq i_2\neq i_3\le n} f\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}, X_{i_3}^{(3)}\right)\right|^p \\ &\quad\le C_{3,p}\left\{E\left|\sum_{i_1=1}^n\sum_{i_2=1}^n\sum_{i_3=1}^n f\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}, X_{i_3}^{(3)}\right)\right|^p\right. \\ &\qquad\qquad + E\left|\sum_{i_1=1}^n\sum_{1\le i_2\le n:\, i_2\neq i_1} f\left(X_{i_1}^{(1)}, X_{i_1}^{(2)}, X_{i_2}^{(3)}\right)\right|^p \end{aligned}$$

$$+E\left|\sum_{i_1=1}^{n}\sum_{1\le i_2\le n:\ i_2\ne i_1} f\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}, X_{i_1}^{(3)}\right)\right|^p$$

$$+E\left|\sum_{i_1=1}^{n}\sum_{1\le i_2\le n:\ i_2\ne i_1} f\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}, X_{i_2}^{(3)}\right)\right|^p$$

$$\left.+E\left|\sum_{i_1=1}^{n} f\left(X_{i_1}^{(1)}, X_{i_1}^{(2)}, X_{i_1}^{(3)}\right)\right|^p\right\}.$$

Since the variables are i.i.d. and f is symmetric, the three terms involving double sums above are equal. Moreover, we can remove the restriction $i_2 \ne i_1$ by adding and subtracting terms. □

5.2. *A sharp decoupling inequality for U-statistics and the proof of Theorem* 1. We begin by introducing some notation. Recall that $f\colon S^k \to \mathbf{R}$ in Theorem 1 is assumed to be symmetric in its arguments; that is, (2.3) holds. Define $X_i(f)\colon S^{k-1} \to \mathbf{R}$ by $X_i(f)(y_1, \dots, y_{k-1}) = f(y_1, \dots, y_{k-1}, X_i)$. More generally, for $1 \le r \le k$, let G_r denote the class of functions $g\colon S^r \to \mathbf{R}$ such that g is symmetric in its arguments. For $g \in G_r$, define $X_i(g) \in G_{r-1}$ by $X_i(g)(y_1, \dots, y_{r-1}) = g(y_1, \dots, y_{r-1}, X_i)$, where we set $G_0 = \mathbf{R}$. Moreover, for $g \in G_r$ and $1 \le h < r$, $0 \le m < n$, define $S_{h,n}^{(m)}(g) \in G_{r-h}$ by

$$S_{h,n}^{(m)}(g)(y_1, \dots, y_{r-h}) = \sum_{m < i_1 < \cdots < i_h \le n} g\left(X_{i_1}, \dots, X_{i_h}, y_1, \dots, y_{r-h}\right). \tag{5.5}$$

For notational simplicity, let $S_{h,n} = S_{h,n}^{(0)}$. We can also extend the preceding definition to the case $h = r$ and $h = 0$ by defining

$$\begin{aligned} S_{r,n}(g) = S_{r,n}^{(0)}(g) &= \sum_{0 < i_1 < \cdots < i_r \le n} g\left(X_{i_1}, \dots, X_{i_r}\right), \\ S_{0,n}^{(m)}(g)(y_1, \dots, y_r) &= g(y_1, \dots, y_r). \end{aligned} \tag{5.6}$$

Since the function $S_{k-h,n}^{(m)}(f)$ has h arguments, applying $S_{h,n}$ to this function yields a constant by (5.6); that is,

$$\begin{aligned} &S_{h,n} S_{k-h,n}^{(m)}(f) \\ &\quad = \sum_{1 \le i_1 < \cdots < i_h \le n;\ m < i_{h+1} < i_{h+2} < \cdots < i_{k-1} < i_k \le n} f\left(X_{i_1}, \dots, X_{i_k}\right). \end{aligned} \tag{5.7}$$

In the same spirit, we have

$$\begin{aligned} S_{k,n}(f) &= \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} f\left(X_{i_1}, \dots, X_{i_k}\right) \\ &= \sum_{j=k}^{n} S_{k-1,j-1} X_j(f), \qquad n \ge k. \end{aligned} \tag{5.8}$$

Moreover, for $k \ge r \ge 1$ and $j > 2^i$, we have the following extension of the fundamental relation introduced in Chow, de la Peña and Teicher (1993):

$$S_{r,j-1} X_j(f) = \sum_{h=0}^{r} S_{h,2^i} S_{r-h,j-1}^{(2^i)} X_j(f). \tag{5.9}$$

To see this, note that analagous to (5.8), we have for $g = X_j(f)$,

$$\begin{aligned} S_{r,j-1}X_j(f) &= S_{r,2^i}(g) + \sum_{h_1=2^i+1}^{j-1} S_{r-1,h_1-1}X_{h_1}(g) \\ &= S_{r,2^i}(g) + S_{r-1,2^i}S_{1,j-1}^{(2^i)}(g) \\ &\quad + \sum_{h_1>2^i}^{j-1}\sum_{h_2>2^i}^{h_1-1} S_{r-2,h_2-1}X_{h_2}X_{h_1}(g) \\ &= \cdots = \sum_{h=0}^{r} S_{h,2^i}S_{r-h,j-1}^{(2^i)}(g). \end{aligned}$$

Let $\{\tilde{X}_n,\ n \geq 1\}$ be an independent copy of $\{X_n,\ n \geq 1\}$ and define $\tilde{X}_i(f)$, $\tilde{S}_{h,n}^{(m)}(f)$ in terms of the $\tilde{X}_i$ instead of X_i, for example,

$$\tilde{S}_{k,n}(f) = \sum_{1 \leq i_1 < \cdots < i_k \leq n} f\left(\tilde{X}_{i_1}, \ldots, \tilde{X}_{i_k}\right).$$

Composition of the operators $S_{h,n}$ and $\tilde{S}_{k-h,n}^{(m)}$ therefore leads to

$$\begin{aligned} &S_{h,n}\tilde{S}_{k-h,n}^{(m)}(f) \\ &\quad = \sum_{1 \leq i_1 < \cdots < i_h \leq n}\ \sum_{m < i_{h+1} < \cdots < i_k \leq n} f\left(X_{i_1}, \ldots, X_{i_h}, \tilde{X}_{i_{h+1}}, \ldots, \tilde{X}_{i_k}\right). \end{aligned} \tag{5.10}$$

Throughout the sequel, we use C_p, $C_{k,p}$ and so on, to denote absolute constants which may change from one line to another. Moreover, for two non-negative functions g and h defined on S^r, we write $g \leq h$ if $g(y_1, \ldots, y_r) \leq h(y_1, \ldots, y_r)$ for all $(y_1, \ldots, y_r) \in S^r$.

We now proceed to prove a basic decoupling inequality given by (5.12) and (5.19) below. First, note from (5.9) that

$$\begin{aligned} &\sum_{j=r+1}^{2^{n+1}} \left(S_{r,j-1}X_j(f)\right)^2 I(T \geq j) \\ &\quad \leq C_r \sum_{i=[\log_2(r+1)]}^{n} I(T \geq 2^i) \sum_{j=2^i+1}^{2^{i+1}} \sum_{h=0}^{r} \left(S_{h,2^i}S_{r-h,j-1}^{(2^i)}X_j(f)\right)^2 \\ &\quad = C_r \sum_{i=[\log_2(r+1)]}^{n} I(T \geq 2^i) \sum_{h=0}^{r} \sum_{j=2^i+1}^{2^{i+1}} \left(S_{h,2^i}S_{r-h,j-1}^{(2^i)}X_j(f)\right)^2, \end{aligned} \tag{5.11}$$

where $C_r = 2^{r+1}$. Let $1 \leq r < k$ and $1 \leq \alpha \leq p$. Since

$$S_{r+1,n}(f)(y_{r+2,\ldots,}y_k) = \sum_{1 \leq i_1 < \cdots < i_{r+1} \leq m} f\left(X_{i_1}, \ldots, X_{i_{r+1}}, y_{r+2}, \ldots, y_n\right),$$

application of the inequality in Remark (A) of Section 3, due to Burkholder in the case $p > 1$ and to Davis in the case $p = 1$ [cf. Corollary 11.2.1 and Theorem 11.2.2 of Chow and Teicher (1988)], gives the bound

$$E \max_{r+1 \leq m \leq T} \left|S_{r+1,m}(f)\right|^{\alpha} \leq C_\alpha J_{r,\alpha,\infty}(f), \tag{5.12}$$

where we define $J_{r,\alpha,n}(f) \in G_{k-r-1}$ by

$$J_{r,\alpha,n}(f) = E\left\{\sum_{j=r+1}^{2^{n+1}} \left(S_{r,j-1}X_j(f)\right)^2 I(T \geq j)\right\}^{\alpha/2}. \tag{5.13}$$

We next bound $J_{r,\alpha,n}(f)$ as follows:

$$\begin{aligned}
&J_{r,\alpha,n}(f)\\
&\quad \leq C_{r,\alpha} E\left\{\sum_{i=[\log_2(r+1)]}^{n} I(T \geq 2^i)\right.\\
&\qquad\qquad \left.\times \sum_{h=0}^{r}\left[\sum_{j=2^i+1}^{2^{i+1}}\left(S_{h,2^i}S^{(2^i)}_{r-h,j-1}X_j(f)\right)^2\right]^{\alpha/2}\right\}\\
&\quad = C_{r,\alpha}\sum_{i=[\log_2(r+1)]}^{n}\\
&\qquad\qquad \times \sum_{h=0}^{r} E\left\{I(T \geq 2^i)\left[\sum_{j=2^i+1}^{2^{i+1}}\left(S_{h,2^i}S^{(2^i)}_{r-h,j-1}X_j(f)\right)^2\right]^{\alpha/2}\right\}\\
&\quad = C_{r,\alpha}\sum_{i=[\log_2(r+1)]}^{n}\\
&\qquad \times \sum_{h=0}^{r} E\left\{I(T \geq 2^i)\, E\left(\left[\sum_{j=2^i+1}^{2^{i+1}}\left(S_{h,2^i}\tilde{S}^{(2^i)}_{r-h,j-1}\tilde{X}_j(f)\right)^2\right]^{\alpha/2}\middle| F_{2^i}\right)\right\}\\
&\quad \leq C_{r,\alpha}\sum_{i=[\log_2(r+1)]}^{n}\sum_{h=0}^{r}\\
&\qquad \times\left\{I(T \geq 2^i)\, E\left(\max_{2^i+1\leq m\leq 2^{i+1}}\left|\sum_{j=2^i+1}^{m} S_{h,2^i}\tilde{S}^{(2^i)}_{r-h,j-1}\tilde{X}_j(f)\right|^{\alpha}\middle| F_{2^i}\right)\right\}\\
&\quad \leq C_{r,\alpha}\sum_{i=[\log_2(r+1)]}^{n}\\
&\qquad\qquad \times \sum_{h=0}^{r} E\left\{I(T \geq 2^i)\, E\left(\left|\sum_{j=2^i+1}^{2^{i+1}} S_{h,2^i}\tilde{S}^{(2^i)}_{r-h,j-1}\tilde{X}_j(f)\right|^{\alpha}\middle| F_{2^i}\right)\right\}.
\end{aligned} \tag{5.14}$$

Line 1 in (5.14) follows from (5.9). To derive line 3, condition on F_{2^i}, and use the independence of the X_j from F_{2^i} for $j > 2^i$. Line 4 follows by applying (conditionally) the reverse inequality of Burkholder and Davis [$E \sup_n|\sum_{i=1}^n d_i|^p \geq A_p E(\sum_{i=1}^{\infty} d_i^2)^{p/2}$ in the notation of Remarks (A) and (E) of Section 3], and the last line follows from an easy extension of Corollary 2 in de la Peña (1992a). The next step is to bound

$$EI(T \geq 2^i)\left|\sum_{j=2^i+1}^{2^{i+1}} S_{h,2^i}\tilde{S}^{(2^i)}_{r-h,j-1}\tilde{X}_j(f)\right|^{\alpha}. \tag{5.15}$$

Conditional Jensen's inequality and Corollary 7.4.6 of Chow and Teicher (1988) give that with probability 1

$$E\{|S_{h,T}(g)|^{\alpha} \mid F_m\} \geq |S_{h,m}(g)|^{\alpha} I(T \geq m) \quad \text{for } g \in G_{h'} \text{ with } h' \geq h.$$

Hence, working conditionally on the $\tilde{X}$'s, we have

$$\begin{aligned} &E\left\{I(T \geq 2^i)\left|\sum_{j=2^i+1}^{2^{i+1}} S_{h,2^i}\tilde{S}^{(2^i)}_{r-h,j-1}\tilde{X}_j(f)\right|^{\alpha}\right\} \\ &\qquad \leq E\left\{I(T \geq 2^i)\left|\sum_{j=2^i+1}^{2^{i+1}} S_{h,T}\tilde{S}^{(2^i)}_{r-h,j-1}\tilde{X}_j(f)\right|^{\alpha}\right\}. \end{aligned} \tag{5.16}$$

In view of the symmetry condition (2.3), we can replace the right-hand side of (5.16) evaluated at $(y_{r+2},\dots,y_k)$ by

$$\begin{aligned} &E\left\{\frac{I(T \geq 2^i)}{(r+1-h)!}\left|\sum_{1\leq i_1<\cdots<i_h\leq T}\ \sum_{2^i+1\leq i_{h+1}\neq i_{h+2}\neq\cdots\neq i_{r+1}\leq 2^{i+1}}\right.\right. \\ &\qquad \left.\left. f\left(X_{i_1},\dots,X_{i_h},\tilde{X}_{i_{h+1}},\dots,\tilde{X}_{i_{r+1}},y_{r+2},\dots,y_k\right)\right|^{\alpha}\right\}. \end{aligned} \tag{5.17}$$

Using de la Peña (1992a) conditionally on $\{X_i\}$, this quantity can be decoupled. That is, if we take $\{X_i^{(h+1)}\},\{X_i^{(h+2)}\},\dots,\{X_i^{(r+1)}\}$ to be i.i.d. copies of $\{X_i\}$, then (5.17) is less than or equal to

$$\begin{aligned} &C_{r,h,\alpha}E\left\{I(T \geq 2^i)\left|\sum_{1\leq i_1<\cdots<i_h\leq T}\ \sum_{2^i+1\leq i_{h+1}\neq i_{h+2}\neq\cdots\neq i_{r+1}\leq 2^{i+1}}\right.\right. \\ &\qquad \left.\left. f\left(X_{i_1},\dots,X_{i_h},X^{(h+1)}_{i_{h+1}},\dots,X^{(r+1)}_{i_{r+1}},y_{r+2},\dots,y_k\right)\right|^{\alpha}\right\}. \end{aligned} \tag{5.18}$$

Hence (5.15) is less than or equal to (5.18).

Putting the upper bound (5.18) for (5.15) into (5.14) shows that $J_{r,\alpha,n}(f)(y_{r+2},\dots,y_k)$ is bounded above by

$$\begin{aligned} &C_{r,\alpha}\sum_{i=1}^{n}\sum_{h=0}^{r}E\left\{I(T \geq 2^i)\left|\sum_{2^i+1\leq i_{h+1}\neq i_{h+2}\neq\cdots\neq i_{r+1}\leq 2^{i+1}}\right.\right. \\ &\qquad \left.\left. S_{h,T}(f)\left(X^{(h+1)}_{i_{h+1}},\dots,X^{(r+1)}_{i_{r+1}},y_{r+2},\dots,y_k\right)\right|^{\alpha}\right\}. \end{aligned}$$

Hence, conditional on $\{X_i\}$, applying Lemma 6 to $S_{h,T}(f)$ (instead of f) then yields the following bound for $J_{r,\alpha,n}(f)$:

$$
\begin{aligned}
&J_{r,\alpha,n}(f)(y_{r+2},\dots,y_k)\\
&\quad\le C_{r,\alpha}\sum_{i=1}^{n}\sum_{h=0}^{r}\sum_{t=1}^{r+1-h}E\Bigg\{I(T\ge 2^i)\\
&\qquad\qquad\times\Bigg|\sum_{i_{h+1}=1}^{2^i}\cdots\sum_{i_{h+t}=1}^{2^i}S_{h,T}(f)\Big(X_{i_{h+1}}^{(h+1)},\dots,X_{i_{h+t}}^{(h+t)},\\
&\qquad\qquad\qquad\dots,X_{i_{h+t}}^{(r+1)},y_{r+2},\dots,y_k\Big)\Bigg|^{\alpha}\Bigg\},
\end{aligned}
\tag{5.19}
$$

noting that $2^{i+1}-2^i=2^i$ and that conditional on $\{X_n\}$, the $X_{i_{h+1}}^{(h+1)},\dots,X_{i_h}^{(k)}$ are i.i.d. random variables.

PROOF OF THEOREM 1. Set $r+1=k$ and $\alpha=p$ in (5.19). Then the summands in the upper bound of (5.19) reduce to

$$
\begin{aligned}
&E\Bigg\{I(T\ge 2^i)\Bigg|\sum_{i_{h+1}=1}^{2^i}\cdots\sum_{i_{h+t}=1}^{2^i}S_{h,T}(f)\Big(X_{i_{h+1}}^{(h+1)},\dots,X_{i_{h+t}}^{(h+t)},\dots,X_{i_{h+t}}^{(k)}\Big)\Bigg|^{p}\Bigg\}\\
&\quad\le C_{k,p}E\Bigg\{I(T\ge 2^i)\big(2^i/T\big)^{p/2}\\
&\qquad\qquad\times\Bigg|\sum_{i_{h+1}=1}^{2^i}\cdots\sum_{i_{h+t-1}=1}^{2^i}\sum_{i_{h+t}=1}^{T}S_{h,T}(f)\Big(X_{i_{h+1}}^{(h+1)},\dots,X_{i_{h+t}}^{(h+t)},\\
&\qquad\qquad\qquad\dots,X_{i_{h+t}}^{(k)}\Big)\Bigg|^{p}\Bigg\}\\
&\quad\le\cdots\\
&\quad\le C_{k,p}E\Bigg\{I(T\ge 2^i)\big(2^{it}/T^t\big)^{p/2}\\
&\qquad\qquad\times\Bigg|\sum_{i_{h+1}=1}^{T}\cdots\sum_{i_{h+t-1}=1}^{T}\sum_{i_{h+t}=1}^{T}S_{h,T}(f)\Big(X_{i_{h+1}}^{(h+1)},\dots,X_{i_{h+t}}^{(h+t)},\\
&\qquad\qquad\qquad\dots,X_{i_{h+t}}^{(k)}\Big)\Bigg|^{p}\Bigg\},
\end{aligned}
$$

where the inequalities follow by repeated use of Lemma 5 [see (5.2)] conditionally. Since $S_{h,T}(f)=f_{T,k,h}$, putting this in (5.19) and combining the result with (5.12) (in which $\alpha = p$ and $r + 1 = k$) yield

$$
\begin{aligned}
E \max_{k \le n \le T} & \left| \sum_{1 \le i_1 < \cdots < i_k \le n} f(X_{i_1}, \dots, X_{i_k}) \right|^p \\
& \le C_{k,p} \sum_{h=0}^{k-1} \sum_{t=1}^{k-h} E \Bigg\{ \sum_{i=1}^{\infty} 2^{ip/2} \sum_{j=i}^{\infty} 2^{-tjp/2} I(2^j \le T \le 2^{j+1}) \\
& \qquad \times \left| \sum_{i_{h+1}=1}^{T} \cdots \sum_{i_{h+t}=1}^{T} f_{T,k,h}\big(X_{i_{h+1}}^{(h+1)}, \dots, X_{i_{h+t}}^{(h+t)}, \dots, X_{i_{h+t}}^{(k)}\big) \right|^p \Bigg\} \\
& \le C_{k,p} \sum_{h=0}^{k-1} \sum_{t=1}^{k-h} E \left| \sum_{i_{h+1}=1}^{T} \cdots \sum_{i_{h+t}=1}^{T} f_{T,k,h}\big(X_{i_{h+1}}^{(h+1)}, \dots, X_{i_{h+t}}^{(h+t)}, \dots, X_{i_{h+t}}^{(k)}\big) \right|^p .
\end{aligned}
$$

□

5.3. *Proof of Theorem* 2. The preceding proof of Theorem 1 uses the bounds (5.12) and (5.19) in the special case $\alpha = p$ and $r + 1 = k$. We can consider more general (α, r) and the same argument as in the proof of Theorem 1 gives the following variant of Theorem 1: for $1 \le \alpha \le p$ and $1 \le r < k$,

$$
\begin{aligned}
E & \max_{r+1 \le n \le T} \left| S_{r+1,n}(f)(y_{r+2}, \dots, y_k) \right|^\alpha \\
& \le C_{r,\alpha} \sum_{h=0}^{r} \sum_{t=1}^{r+1-h} E \Bigg| \sum_{i_{h+1}=1}^{T} \cdots \sum_{i_{h+t}=1}^{T} S_{h,T}(f)\big(X_{i_{h+1}}^{(h+1)}, \dots, X_{i_{h+t}}^{(h+t)}, \\
& \qquad \dots, X_{i_{h+t}}^{(r+1)}, \\
& \qquad y_{r+2}, \dots, y_k\big) \Bigg|^\alpha .
\end{aligned} \tag{5.20}
$$

Let $q = \rho(k, p) (= (p^{1/(k-1)} - 1)^{-1})$ and

$$\alpha_r = p\big(q/(q+1)\big)^r = p^{1 - r/(k-1)} \text{ for } r = 0, 1, \dots, \tag{5.21}$$

so $\alpha_0 = p$ and $\alpha_{k-1} = 1$. For $0 \le h \le r$ and $1 \le t \le r + 1 - h$, let

$$
\begin{aligned}
Y_{h,t} = & \sum_{i_{h+1}=1}^{T} \cdots \sum_{i_{h+t}=1}^{T} S_{h,T}(f) \\
& \times \big(X_{i_{h+1}}^{(h+1)}, \dots, X_{i_{h+1}}^{(h+t)}, \dots, X_{i_{h+t}}^{(r+1)}, y_{r+2}, \dots, y_k\big).
\end{aligned} \tag{5.22}
$$

Conditional on $\{X_n\}, \{X_n^{(h+1)}\}, \dots, \{X_n^{(h+t-1)}\}$, $Y_{h,t}$ is a sum of i.i.d. zero-mean random variables. For $1 \le h \le r$, since $1 \le \alpha_r < \alpha_{h-1} \le 2$, it follows from

Lemma 5 [see (5.3)] that

$$
\begin{aligned}
&E\Big\{E\Big[|Y_{h,t}|^{\alpha_r}\Big|\{X_n\},\{X_n^{(h+1)}\},\dots,\{X_n^{(h+t-1)}\}\Big]\Big\}\\
&\quad\le C_h E\Big(E\Big[|Y_{h,t}|^{\alpha_{h-1}}\Big|\{X_n\},\{X_n^{(h+1)}\},\dots,\{X_n^{(h+t-1)}\}\Big]\Big)^{\alpha_r/\alpha_{h-1}}\\
&\quad\le C_h E\Bigg\{T^{\alpha_r/\alpha_{h-1}}\Bigg(E\Bigg[\Bigg|\sum_{i_{h+1}=1}^{T}\cdots\sum_{i_{h+t-1}=1}^{T}S_{h,T}(f)\Big(X_{i_{h+1}}^{(h+1)},\dots,X_{i_{h+t-1}}^{(h+t-1)},\\
&\qquad\qquad X_1^{(h+t)},\dots,X_1^{(r+1)},y_{r+2},\dots,y_k\Big)\Bigg|^{\alpha_{h-1}}\Bigg|\\
&\qquad\qquad \{X_n\},\{X_n^{(h+1)}\},\dots,\{X_n^{(h+t-1)}\}\Bigg]\Bigg)^{\alpha_r/\alpha_{h-1}}\Bigg\}.
\end{aligned}
\tag{5.23}
$$

Repeating this argument then shows that $E(E[|Y_{h,t}|^{\alpha_r}|\{X_n\}])$ is bounded above by

$$
\begin{aligned}
&C_h E\Big\{T^{t\alpha_r/\alpha_{h-1}}\Big(E\big[|S_{h,T}(f)\big(X_1^{(h+1)},\dots,X_1^{(r+1)},\\
&\qquad\qquad y_{r+2},\dots,y_k\big)|^{\alpha_{h-1}}|\{X_n\}\big]\Big)^{\alpha_r/\alpha_{h-1}}\Big\}
\end{aligned}
\tag{5.24}
$$

for $h=1,\dots,r$ and $1\le t\le r+1-h$. In the case $h=0$, since $S_{0,T}(f)=f$ by (5.6), a similar argument again gives (5.24) with $\alpha_{-1}=p$ as an upper bound for $E(E[|Y_{0,t}|^{\alpha_r}|\{X_n\}])$. We can also replace $t\alpha_r/\alpha_{h-1}$ by $(r+1-h)\alpha_r/\alpha_{h-1}$ in the upper bound (5.24). Putting this upper bound for $E(E[|Y_{h,t}|^{\alpha_r}|\{X_n\}])$ into (5.20) yields

$$
\begin{aligned}
&E\max_{r+1\le n\le T}|S_{r+1,n}(f)(y_{r+2},\dots,y_k)|^{\alpha_r}\\
&\quad\le C_r\sum_{h=0}^{r}E\Big\{T^{(r+1-h)\alpha_r/\alpha_{h-1}}\\
&\qquad\qquad\times\Big(E\Big[\big|S_{h,T}(f)\big(X_1^{(h+1)},\dots,X_1^{(r+1)},\\
&\qquad\qquad\qquad y_{r+2},\dots,y_k\big)\big|^{\alpha_{h-1}}|\{X_n\}\Big]\Big)^{\alpha_r/\alpha_{h-1}}\Big\}\\
&\quad\le C_r ET^{\rho(k,p)}\Big\{\sum_{h=1}^{r}\Big(E\big|S_{h,T}(f)\big(X_1^{(h+1)},\dots,X_1^{(r+1)},\\
&\qquad\qquad\qquad y_{r+2},\dots,y_k\big)\big|^{\alpha_{h-1}}\Big)^{\alpha_r/\alpha_{h-1}}\\
&\qquad\qquad+\Big(E\big|f\big(\tilde{X}_1^{(1)},\dots,\tilde{X}_1^{(r+1)},y_{r+1},\dots,y_k\big)\big|^{p}\Big)^{\alpha_r/p}\Big\},
\end{aligned}
\tag{5.25}
$$

where the last relation follows from Hölder's inequality, noting that by (5.21), for $1 \le h \le r$,

$$(r+1-h)(\alpha_r/\alpha_{h-1})(1-\alpha_r/\alpha_{h-1})^{-1} = (r+1-h)/\big(p^{(r+1-h)/(k-1)}-1\big) \le \big(p^{1/(k-1)}-1\big)^{-1} = \rho(k,p),$$

since the function $x/(e^{cx}-1)$ is decreasing in $x>0$ for any $c>0$. In this connection, also note that $(r+1)\alpha_r/p = (r+1)p^{-r/(k-1)} \le \rho(k,p)$ since $(1+\varepsilon)^r \ge (r+1)\varepsilon$ for $0<\varepsilon(=p^{1/(k-1)}-1)\le 1$ and $r=1,2,\dots$.

Since $E|f(X_1^{(1)},\dots,X_1^{(k)})|^p<\infty$, it follows from (5.25) that for $1\le r<k$,

$$(5.26)\qquad \begin{aligned} &E\Big\{\max_{r+1\le n\le T}\big|S_{r+1,n}(f)\big(X_1^{(r+2)},\dots,X_1^{(k)}\big)\big|^{\alpha_r}\Big\}<\infty \quad \text{if}\\ &\max_{1\le h\le r} E\Big\{\big|S_{h,T}(f)\big(X_1^{(h+1)},\dots,X_1^{(k)}\big)\big|^{\alpha_{h-1}}\Big\}<\infty. \end{aligned}$$

For $r=h=1$, $\alpha_{h-1}=\alpha_0=p(\in(1,2])$ and it follows from the Marcinkiewicz–Zygmund inequality that

$$\begin{aligned} E\big|S_{1,T}(f)\big(X_1^{(2)},\dots,X_1^{(k)}\big)\big|^p &= E\left|\sum_{i=1}^T f\big(X_i,X_1^{(2)},\dots,X_1^{(k)}\big)\right|^p\\ &\le C_p E\left\{\sum_{i=1}^\infty f^2\big(X_i,X_1^{(2)},\dots,X_1^{(k)}\big)I(T\ge i)\right\}^{p/2}\\ &\le C_p E\left\{\sum_{i=1}^\infty \big|f\big(X_i,X_1^{(2)},\dots,X_1^{(k)}\big)\big|^p I(T\ge i)\right\}\\ &= C_p(ET)E\big|f(X_1,\dots,X_k)\big|^p. \end{aligned}$$

Hence, (5.26) gives

$$E\Big\{\max_{2\le n\le T}\big|S_{2,n}(f)\big(X_1^{(3)},\dots,X_1^{(k)}\big)\big|^{\alpha_1}\Big\}<\infty.$$

Proceeding inductively, (5.26) yields for $r=k-1$ the desired conclusion since $\alpha_{k-1}=1$ and $S_{k,n}(f)=\sum_{1\le i_1<\dots<i_k\le n} f(X_{i_1},\dots,X_{i_k})$ by (5.6). □

REMARK. For the special case $f(x_1,\dots,x_k)=x_1\dots x_k$,

$$(5.27)\qquad S_{r,n}(f)(y_{r+1},\dots,y_k)=\left(\sum_{1\le i_1<\cdots<i_r\le n} X_{i_1}\cdots X_{i_r}\right)y_{r+1}\cdots y_k,$$

and therefore we can simply identify $S_{r,n}(f)$ as $\sum_{1\le i_1<\cdots<i_r\le n} X_{i_1}\cdots X_{i_r}$. The factorization (5.27) also implies that for $1 \le \alpha \le p$,

$$E\left|\sum_{i_{h+1}=1}^{T}\cdots\sum_{i_{h+t}=1}^{T} S_{h,T}(f)\left(X_{i_{h+1}}^{(h+1)},\dots,X_{h+t}^{(h+t)},\dots,X_{h+t}^{(r+1)},y_{r+2},\dots,y_k\right)\right|^{\alpha}$$

$$= E\left\{|S_{h,T}(f)|^{\alpha}\left|\sum_{i_{h+1}=1}^{T} X_{i_{h+1}}^{(h+1)}\right|^{\alpha}\cdots \times\left|\sum_{i_{h+t-1}=1}^{T} X_{i_{h+t-1}}^{(h+t-1)}\right|^{\alpha}\left|\sum_{i_{h+t}=1}^{T} X_{i_{h+t}}^{(h+t)},\dots,X_{i_{h+t}}^{(r+1)}\right|^{\alpha}\right\}|y_{r+2}\cdots y_k|^{\alpha}. \tag{5.28}$$

Using this identity and the independence of the X_i and the $X_n^{(j)}$, we get

$$\begin{aligned}
&E\left\{|S_{h,T}(f)|^{\alpha}E\left(\left|\sum_{i_{h+1}=1}^{T} X_{i_{h+1}}^{(h+1)}\right|^{\alpha}\middle|\{X_n\}\right)\cdots\right.\\
&\qquad\left.\times E\left(\left|\sum_{i_{h+t}=1}^{T} X_{i_{h+t}}^{(h+t)},\dots,X_{i_{h+t}}^{(r+1)}\right|^{\alpha}\middle|\{X_n\}\right)\right\}\\
&\le E\left\{|S_{h,T}(f)|^{\alpha}\left(E\left(\left|\sum_{i_{h+1}=1}^{T} X_{i_{h+1}}^{(h+1)}\right|^{p}\middle|\{X_n\}\right)\right)^{\alpha/p}\cdots\right.\\
&\qquad\left.\times\left(E\left(\left|\sum_{i_{h+t}=1}^{T} X_{i_{h+t}}^{(h+t)},\dots,X_{i_{h+t}}^{(r+1)}\right|^{p}\middle|\{X_n\}\right)\right)^{\alpha/p}\right\}\\
&\le C_p\left(E|X_1|^p\right)^{(r+1-h)\alpha/p}ET^{t\alpha/p}|S_{h,T}(f)|^{\alpha}\quad\text{by (5.3)}.
\end{aligned}\tag{5.29}$$

Hence in this case, we can replace α_{h-1} in (5.23) by p, which will be shown to lead to the weaker moment condition $ET^{(k-1)/(p-1)} < \infty$ as assumed by Chow, de la Peña and Teicher (1993).

Consider the moment condition $ET^q < \infty$ for some $q(\ge 1)$ not necessarily equal to the $\rho(k,p)$ of Theorem 2. The bounds in (5.25) obtained via Hölder's inequality requires that $(r+1-h)(\alpha_r/\alpha_{h-1})(1-\alpha_r/\alpha_{h-1})^{-1} \le q$ for $1 \le h \le r$. In particular, the case $h = r$ entails that $(\alpha_r/\alpha_{r-1})(1-\alpha_r/\alpha_{r-1})^{-1} \le q$, so replacing "$\le q$" by "$= q$" leads to $\alpha_r/\alpha_{r-1} = q/(q+1)$. Since $\alpha_0 = p$, this yields $\alpha_r = p(q/(q+1))^r$ as in (5.21). The boundary condition $\alpha_{k-1} = 1$ then implies $q = \rho(k,p)$. In the multilinear case $f(x_1,\dots,x_k) = x_1\cdots x_k$, (5.29) enables us to replace (5.25) by the sharper bound

$$\begin{aligned}
&E\max_{r+1\le n\le T}|S_{r+1,n}(f)(y_{r+2},\dots,y_k)|^{\beta_r}\\
&\quad\le C_{p,r}\left(E|X_1|^p\right)^{(r+1-h)\beta_r/p}\sum_{h=0}^{r}E\left\{T^{(r+1-h)\beta_r/p}|S_{h,T}(f)|^{\beta_r}\right\}|y_{r+2}\cdots y_k|^{\beta_r}
\end{aligned}$$

$$\leq C_{p,r}\big(E|X_1|^p\big)^{(r+1-h)}|y_{r+2}\cdots y_k|^{\beta_r}$$
$$\times\sum_{h=0}^{r}\Big(ET^{(r+1-h)(\beta_r/p)(1-\beta_r/\beta_{h-1})^{-1}}\Big)^{1-\beta_r/\beta_{h-1}}\Big(E\big|S_{h,T}(f)\big|^{\beta_{h-1}}\Big)^{\beta_r/\beta_{h-1}}.$$

Hence in this case, instead of $(r+1-h)(\alpha_r/\alpha_{h-1})(1-\alpha_r/\alpha_{h-1})^{-1} \leq q$, we require that

$$(5.30)\qquad (r+1-h)(\beta_r/p)(1-\beta_r/\beta_{h-1})^{-1} \leq q \quad \text{for } 1 \leq h \leq r.$$

Replacing "$\leq q$" by "$= q$" in the case $h = r$ leads to $(\beta_r/p)(1-\beta_r/\beta_{r-1})^{-1} = q$, or equivalently, $\beta_r = q/(p^{-1} + q\beta_{r-1}^{-1})$. The initial condition $\beta_0 = p$ and the terminal condition $\beta_{k-1} = 1$ then yield $\beta_r = pq/(q+r)$, $q = (k-1)\cdot(p-1)^{-1}$. Hence in the multilinear case, the weaker moment condition $ET^{(k-1)/(p-1)} < \infty$ suffices for the conclusions of Theorem 1. Moreover, the factorization (5.27) simplifies the result considerably.

REFERENCES

ANSCOMBE, F. J. (1952). Large-sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* **48** 600–607.

ARAS, G. (1988). Sequential point estimation based on *U*-statistics. *Sequential Anal.* **7** 203–226.

ARAS, G. and WOODROOFE, M. (1993). Asymptotic expansions for the moments of a randomly stopped average. *Ann. Statist.* **21** 503–519.

CHOW, Y. S., DE LA PEÑA, V. H. and TEICHER, H. (1993). Wald's lemma for a class of denormalized *U*-statistics. *Ann. Probab.* **21** 1151–1158.

CHOW, Y. S., HSIUNG, C. and LAI, T. L. (1979). Extended renewal theory and moment convergence in Anscombe's theorem. *Ann. Probab.* **7** 304–318.

CHOW, Y. S. and TEICHER, H. (1988). *Probability Theory*: *Independence*, *Interchangeability*, *Martingales*, 2nd ed. Springer, New York.

CHOW, Y. S. and YU, K. F. (1981). The performance of a sequential procedure for estimating the mean. *Ann. Statist.* **9** 184–188.

DE LA PEÑA, V. H. (1992a). Decoupling and Khintchine's inequalities for *U*-statistics. *Ann. Probab.* **20** 1877–1892.

DE LA PEÑA, V. H. (1992b). Sharp bounds on the L_p norm of a randomly stopped multilinear form with an application to Wald's equation. In *Probability in Banach Spaces* **8** (R. M. Dudley, M. G. Hahn and J. D. Kuelbs, eds.) 69–79. Birkhäuser, New York.

DE LA PEÑA, V. H. and LAI, T. L. (1997). Wald's equation and asymptotic bias of randomly stopped *U*-statistics. *Proc. Amer. Math. Soc.* **125** 917–925.

GRAMS, W. F. and SERFLING, R. J. (1973). Convergence rates for *U*-statistics and related statistics. *Ann. Statist.* **1** 153–160.

HALMOS, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statist.* **17** 34–43.

HITCZENKO, P. (1988). Comparison of moments of tangent sequences of random variables. *Probab. Theory Related Fields* **78** 223–230.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

HOEFFDING, W. (1961). The strong law of large numbers for *U*-statistics. Mimeograph series no. 302, Inst. Statistics, Univ. North Carolina.

KLASS, M. J. (1976). Toward a universal law of the iterated logarithm I. *Z. Wahrsch. Verw. Gebiete* **36** 165–178.

KLASS, M. J. (1981). A method of approximating expectations of functions of sums of independent random variables. *Ann. Probab.* **9** 413–428.

LAI, T. L. and SIEGMUND, D. (1977). A nonlinear renewal theory with applications to sequential analysis I. *Ann. Statist.* **5** 946–954.

LAI, T. L. and WANG, J. Q. (1993). Edgeworth expansions for symmetric statistics with applications to bootstrap method. *Statist. Sinica* **3** 517–542.

MANDELBAUM, A. and TAQQU, M. S. (1984). Invariance principle for symmetric statistics. *Ann. Statist.* **12** 483–496.

SIEGMUND, D. (1985). *Sequential Analysis*: *Tests and Confidence Intervals*. Springer, New York.

WALD, A. (1945). Some generalizations of the theory of cumulative sums of random variables. *Ann. Math. Statist.* **16** 287–293.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
BOX 10 MATHEMATICS
NEW YORK, NEW YORK 10027

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
SEQUOIA HALL
STANFORD, CALIFORNIA 94305
E-MAIL: karola@stat.stanford.edu