

## WALD'S EQUATION AND ASYMPTOTIC BIAS OF RANDOMLY STOPPED $U$ -STATISTICS

VICTOR H. DE LA PEÑA AND TZE LEUNG LAI

(Communicated by Wei Y. Loh)

ABSTRACT. In this paper we make use of decoupling arguments and martingale inequalities to extend Wald's equation for sample sums to randomly stopped de-normalized  $U$ -statistics. We also apply this result in conjunction with nonlinear renewal theory to obtain asymptotic expansions for the means of normalized  $U$ -statistics from sequential samples.

### 1. INTRODUCTION

Let  $X_1, \dots, X_n$  be i.i.d. random variables taking values in a measurable space  $(S, \mathcal{S})$  and having a common distribution  $F$ . Let  $g : S^k \rightarrow \mathbb{R}$  be a Borel measurable function that is symmetric, i.e.,  $g(x_1, \dots, x_k) = g(x_{\pi(1)}, \dots, x_{\pi(k)})$  for any permutation  $\pi$ . For  $n \geq k$ , the normalized  $U$ -statistic with kernel  $g$  is

$$U_n = \left\{ \sum_{1 \leq i_1 < \dots < i_k \leq n} g(X_{i_1}, \dots, X_{i_k}) \right\} / \binom{n}{k},$$

and  $k$  is called the *degree* of the  $U$ -statistic.  $U$ -statistics were introduced by Halmos [Ha] to provide unbiased estimates of functionals of the form

$$\theta (= \theta(F)) = \int \dots \int g(x_1, \dots, x_k) dF(x_1) \dots dF(x_k)$$

of the population distribution  $F$ . Hoeffding [H] subsequently developed a comprehensive theory of  $U$ -statistics. In particular, he introduced the basic decomposition

$$(1) \quad \begin{aligned} U_n - \theta &= kn^{-1} \sum_{i=1}^n f_1(X_i) + k(k-1)\{n(n-1)\}^{-1} \sum_{1 \leq i < j \leq n} f_2(X_i, X_j) + \dots \\ &\quad + k!\{n \dots (n-k+1)\}^{-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} f_k(X_{i_1}, \dots, X_{i_k}), \end{aligned}$$

---

Received by the editors October 15, 1994 and, in revised form, July 28, 1995.

1991 *Mathematics Subject Classification*. Primary 60G40, 62L12; Secondary 62L10.

*Key words and phrases*. Hoeffding decomposition, decoupling, martingales, Wald's equation, stopping times.

The first author's research was supported by the National Science Foundation under DMS-9310682.

The second author's research was supported by the National Science Foundation under DMS-9403794.

in which  $Ef_1(X_1) = 0$  and  $f_j$  is symmetric with  $E\{f_j(X_1, \dots, X_j)|X_1, \dots, X_{j-1}\} = 0$  for  $2 \leq j \leq k$ . In fact,

$$f_1(x) = Eg(x, X_2, \dots, X_k) - \theta,$$

$$f_2(x, y) = Eg(x, y, X_3, \dots, X_k) - f_1(x) - f_1(y) + \theta,$$

etc. Using this decomposition, he proved the asymptotic normality of  $U_n$ . The theory of  $U$ -statistics has played a fundamental role in the development of non-parametric statistical methodology.

The unbiased property  $EU_n = \theta$  no longer holds if we replace the fixed sample size  $n$  by a stopping time  $T$ . To begin with, consider the case  $k = 1$ , for which  $U_n$  reduces to the sample mean  $\bar{W}_n = n^{-1} \sum_{i=1}^n W_i$ , where  $W_i = g(X_i)$ . Following the pioneering work of Cox [Co], much progress has been made in the development of asymptotic approximations for the bias of  $\bar{W}_T$ , and in bias-corrected modifications of  $\bar{W}_T$  for stopping times  $T$  that are commonly used in sequential hypothesis testing and sequential estimation problems, cf. [S1], [S2], [W2], [AW]. In Section 3 we derive similar asymptotic approximations for the bias when  $\bar{W}_T$  is replaced by  $U_T$ , where  $U_n$  is a normalized  $U$ -statistic of general degree  $k$ .

De-normalizing the sample mean  $\bar{W}_T$  yields the sample sum  $\sum_{i=1}^T W_i$ , whose expected value is given by the following theorem, often referred to as “Wald’s equation” or “Wald’s lemma”, which has provided a basic tool in the asymptotic analysis of the bias of  $\bar{W}_T$  (cf. [AW], [S1]).

**Theorem 1.** *Let  $\{W_i\}$  be a sequence of i.i.d. random variables with  $EW_i = \mu$ , and let  $S_n = \sum_{i=1}^n (W_i - \mu)$ . Let  $T$  be a stopping time adapted to  $\{W_i\}$  such that  $ET < \infty$ . Then  $ES_T = 0$ .*

In Section 2 we generalize Theorem 1 to the (de-normalized)  $U$ -statistics  $S_n = \sum_{1 \leq i_1 < \dots < i_j \leq n} f_j(X_{i_1}, \dots, X_{i_j})$ . The generalization involves much deeper probabilistic ideas, including recent results from decoupling theory. By using a different method, Chow, de la Peña and Teicher [CdlPT] recently generalized Theorem 1 to the multilinear  $U$ -statistics  $S_n = \sum_{1 \leq i_1 < \dots < i_k \leq n} X_{i_1} \dots X_{i_k}$ . Their method, however, depends heavily on the multilinear form of  $S_n$  and cannot be directly extended to the general  $U$ -statistics treated in Section 2.

## 2. WALD’S EQUATION FOR DE-NORMALIZED $U$ -STATISTICS

To highlight the main ideas and to simplify the arguments, we focus here on  $U$ -statistics of degree 2, for which we shall prove the following extension of Wald’s equation.

**Theorem 2.** *Let  $\{X_i\}$  be a sequence of i.i.d. random variables with values in a measurable space  $(S, \mathcal{S})$ . Let  $T$  be a stopping time adapted to  $\{X_i\}$  such that*

$$E(T^{1/(p-1)}) < \infty \quad \text{for some } 1 < p \leq 2.$$

*Let  $f : S \times S \rightarrow \mathbb{R}$  be a Borel measurable function such that*

$$E|f(X_1, X_2)|^p < \infty \quad \text{and} \quad E(f(X_1, X_2)|X_1) = E(f(X_1, X_2)|X_2) = 0 \text{ a.s.}$$

*Then  $E(\sum_{1 \leq i < j \leq T} f(X_i, X_j)) = 0$ .*

Before starting the proof of this result, let us recall an easy consequence of Levy’s inequality which will be used repeatedly.

**Lemma 1.** Let  $\{X_i\}$  be a sequence of independent mean zero random variables. Then, for all  $p \geq 1$ ,  $E \max_{j \leq n} |\sum_{i=1}^j X_i|^p \leq C_p E |\sum_{i=1}^n X_i|^p$ , where the constant  $C_p > 0$  depends on  $p$  only.

*Proof.* The proof of this result follows easily by using a symmetrization argument along with Levy's inequality for the tail probabilities of sums of symmetric random variables. See [K] for a proof.  $\square$

*Proof of Theorem 2.* Since  $\{\sum_{1 \leq i < j \leq n} f(X_i, X_j)\}$  is a martingale, the desired conclusion holds (using the optional sampling theorem) if we replace  $T$  by the bounded stopping time  $\min(T, n)$ . Hence by the dominated convergence theorem, the desired conclusion follows, for general stopping times, if it can be shown that

$$(2) \quad E \max_{n \leq T} \left| \sum_{1 \leq i < j \leq n} f(X_i, X_j) \right| < \infty.$$

We will prove (2) by using a simple decoupling argument to obtain a partial decoupling between the  $X$ 's and  $T$ . A centering argument is used next to neutralize the remaining effect the stopping time has on the original variables. This neutralizing effect is realized by means of an averaging device coupled with Hölder's inequality. Throughout the proof we will also use the Burkholder-Gundy (B-G) and Davis' inequalities for martingales, cf. [CT], p. 395.

Let  $\{\tilde{X}_i\}$  and  $\{\tilde{X}_i\}$  be two independent copies of  $\{X_i\}$ . Set

$$d_j = \left( \sum_{i=1}^{j-1} f(X_i, X_j) \right)^2 \mathbf{1}(T \geq j), \quad e_j = \left( \sum_{i=1}^{j-1} f(X_i, \tilde{X}_j) \right)^2 \mathbf{1}(T \geq j).$$

Then  $\{e_j\}$  and  $\{d_j\}$  are tangent with respect to  $\{\mathcal{F}_n\}$ , i.e.,  $\mathcal{L}(d_j | \mathcal{F}_{j-1}) = \mathcal{L}(e_j | \mathcal{F}_{j-1})$  for all  $j \geq 2$ , where  $\mathcal{F}_n = \sigma(X_1, \dots, X_n; \tilde{X}_1, \dots, \tilde{X}_n)$ . Recall from [Z] that if  $\{d_i\}$ ,  $\{e_i\}$  are two tangent sequences of nonnegative random variables, then for all  $0 < r < 1$ ,

$$(3) \quad E \left( \sum_{i=1}^n d_i \right)^r \leq C E \left( \sum_{i=1}^n e_i \right)^r.$$

This observation along with a double use of Davis' inequality yields

$$\begin{aligned} E \max_{n \leq T} \left| \sum_{1 \leq i < j \leq n} f(X_i, X_j) \right| &\leq C E \left\{ \sum_{j=2}^{\infty} \left( \sum_{i=1}^{j-1} f(X_i, X_j) \right)^2 \mathbf{1}(T \geq j) \right\}^{\frac{1}{2}} \\ &\leq C E \left\{ \sum_{j=2}^{\infty} \left( \sum_{i=1}^{j-1} f(X_i, \tilde{X}_j) \right)^2 \mathbf{1}(T \geq j) \right\}^{\frac{1}{2}} \\ &\leq C E \sup_n \left| \sum_{j=2}^n \left( \sum_{i=1}^{j-1} f(X_i, \tilde{X}_j) \right) \mathbf{1}(T \geq j) \right|. \end{aligned}$$

Here and in the sequel,  $C$  denotes a generic constant which may change from one bound to another. Observe that  $\sum_{j=2}^n \left( \sum_{i=1}^{j-1} f(X_i, \tilde{X}_j) \right) \mathbf{1}(T \geq j)$  is a sum of independent mean-zero random variables (with the sum converging a.s. to a limit as  $n \rightarrow \infty$ ) once we condition on  $(\{X_i\}, T)$ . Therefore, using Lemma 1 conditionally,

we can bound the last expectation above by

$$\begin{aligned}
 CE \left| \sum_{j=2}^{\infty} \left( \sum_{i=1}^{j-1} f(X_i, \tilde{X}_j) \right) 1(T \geq j) \right| &= CE \left| \sum_{1 \leq i < j \leq T} f(X_i, \tilde{X}_j) \right| \\
 &\leq CE \left| \sum_{j=1}^T \sum_{i=1}^T f(X_i, \tilde{X}_j) \right| + CE \left| \sum_{1 \leq i < j \leq T} f(X_j, \tilde{X}_i) \right| + CE \left| \sum_{i=1}^T f(X_i, \tilde{X}_i) \right| \\
 &= \text{(I)} + \text{(II)} + \text{(III)}.
 \end{aligned}$$

Of the three quantities above, **(II)** and **(III)** are easier to bound, and one can decouple **(II)** further by using (3).

As for **(I)**, note that both  $\{X_i\}$  and  $\{\tilde{X}_i\}$  appear in  $\sum_{j=1}^T \sum_{i=1}^T f(X_i, \tilde{X}_j)$  in a symmetric fashion. We now exploit this fact by means of an averaging device and use Hölder’s inequality to obtain

$$\begin{aligned}
 \text{(I)} &= CE \left| \sum_{j=1}^T \sum_{i=1}^T f(X_i, \tilde{X}_j) \right| = CE \left\{ T^{\frac{1}{p}} \frac{\left| \sum_{j=1}^T \sum_{i=1}^T f(X_i, \tilde{X}_j) \right|}{T^{\frac{1}{p}}} \right\} \\
 &\leq C(ET^{\frac{1}{p-1}})^{\frac{p-1}{p}} \left( E \frac{\left| \sum_{j=1}^T \sum_{i=1}^T f(X_i, \tilde{X}_j) \right|^p}{T} \right)^{1/p} \\
 &\leq C(ET^{\frac{1}{p-1}})^{\frac{p-1}{p}} \left\{ E \frac{E[(\sum_{j=1}^T (\sum_{i=1}^T f(X_i, \tilde{X}_j))^2)^{\frac{p}{2}} | \sigma(\{X_r\}, T)]}{T} \right\}^{1/p}, \text{ by B-G,} \\
 &\leq C(ET^{\frac{1}{p-1}})^{\frac{p-1}{p}} \left\{ E \frac{\sum_{j=1}^T E[|\sum_{i=1}^T f(X_i, \tilde{X}_j)|^p | \sigma(\{X_r\}, T)]}{T} \right\}^{1/p} \\
 &= C(ET^{\frac{1}{p-1}})^{\frac{p-1}{p}} (E \sum_{i=1}^T |f(X_i, \tilde{X}_1)|^p)^{\frac{1}{p}} \\
 &\leq C(ET^{\frac{1}{p-1}})^{\frac{p-1}{p}} \{E(\sum_{i=1}^T f^2(X_i, \tilde{X}_1))^{p/2}\}^{1/p}, \text{ by B-G,} \\
 &\leq C(ET^{\frac{1}{p-1}})^{\frac{p-1}{p}} (E \sum_{i=1}^T |f(X_i, \tilde{X}_1)|^p)^{\frac{1}{p}} = C(ET^{\frac{1}{p-1}})^{\frac{p-1}{p}} (ETE|f(X_1, \tilde{X}_1)|^p)^{\frac{1}{p}},
 \end{aligned}$$

by concavity of the function  $f(x) = x^{p/2}$ , and by Theorem 1 applied conditionally.

To bound **(II)**, recall that  $\{\tilde{X}_i\}$  and  $\{\tilde{X}_i\}$  are independent copies of  $\{X_i\}$ , and set

$$\tilde{d}_j = \left( \sum_{i=1}^{j-1} f(X_j, \tilde{X}_i) \right)^2 1(T \geq j), \quad \tilde{e}_j = \left( \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i) \right)^2 1(T \geq j).$$

Then  $\{\tilde{e}_j\}$  and  $\{\tilde{d}_j\}$  are tangent with respect to  $\{\tilde{\mathcal{F}}_n\}$ , where

$$\tilde{\mathcal{F}}_n = \sigma(X_1, \dots, X_n; \tilde{X}_1, \dots, \tilde{X}_n; \tilde{X}_1, \dots, \tilde{X}_n).$$

Next, Davis' inequality and (3) yield

$$\begin{aligned}
(\mathbf{II}) &= CE \left| \sum_{1 \leq i < j \leq T} f(X_j, \tilde{X}_i) \right| = CE \left| \sum_{j=2}^T \sum_{i=1}^{j-1} f(X_j, \tilde{X}_i) \right| \\
&\leq CE \sqrt{\sum_{j=2}^{\infty} \left( \sum_{i=1}^{j-1} f(X_j, \tilde{X}_i) \right)^2 \mathbf{1}(T \geq j)} \\
&\leq CE \sqrt{\sum_{j=2}^{\infty} \left( \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i) \right)^2 \mathbf{1}(T \geq j)} \leq CE \left| \sum_{j=2}^T \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i) \right|,
\end{aligned}$$

where the last inequality follows by applying first Davis' inequality and then Lemma 1. Observe that we have now completely decoupled the stopping time. Moreover, by the B-G inequality and the concavity of  $f(x) = x^{p/2}$ ,

$$\begin{aligned}
E \left| \sum_{j=2}^n \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i) \right| &\leq \left\{ E \left| \sum_{j=2}^n \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i) \right|^p \right\}^{\frac{1}{p}} \\
&\leq C \left\{ E \left( \sum_{j=2}^n \left( \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i) \right)^2 \right)^{\frac{p}{2}} \right\}^{\frac{1}{p}} \leq C \left\{ \sum_{j=2}^n E \left| \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i) \right|^p \right\}^{\frac{1}{p}} \\
&\leq C \left\{ \sum_{j=2}^n (j-1) E |f(\tilde{X}_j, \tilde{X}_1)|^p \right\}^{\frac{1}{p}} \leq C n^{\frac{2}{p}} \left\{ E |f(X_1, X_2)|^p \right\}^{\frac{1}{p}}.
\end{aligned}$$

Putting these two observations together, we get

$$\begin{aligned}
(\mathbf{II}) &= CE \left| \sum_{1 \leq i < j \leq T} f(X_j, \tilde{X}_i) \right| \leq CE \left| \sum_{j=2}^T \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i) \right| \\
&= C \sum_{n=1}^{\infty} E \left| \sum_{j=2}^n \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i) \right| P(T = n) \\
&\leq C (E |f(X_1, X_2)|^p)^{\frac{1}{p}} \sum_{n=1}^{\infty} n^{\frac{2}{p}} P(T = n) \\
&= C (E |f(X_1, X_2)|^p)^{\frac{1}{p}} E T^{\frac{2}{p}} \leq C (E |f(X_1, X_2)|^p)^{\frac{1}{p}} E T^{\frac{1}{p-1}}.
\end{aligned}$$

(**III**) is the simplest quantity to bound. By Hölder's inequality and the B-G inequality,

$$\begin{aligned}
E \left| \sum_{i=1}^T f(X_i, \tilde{X}_i) \right| &\leq \left( E \left| \sum_{i=1}^T f(X_i, \tilde{X}_i) \right|^p \right)^{\frac{1}{p}} \leq C \left\{ E \left( \sum_{i=1}^T f^2(X_i, \tilde{X}_i) \right)^{\frac{p}{2}} \right\}^{\frac{1}{p}} \\
&\leq C \left( E \sum_{i=1}^T |f(X_i, \tilde{X}_i)|^p \right)^{\frac{1}{p}} = C (E T E |f(X_1, \tilde{X}_1)|^p)^{\frac{1}{p}},
\end{aligned}$$

where we have applied Theorem 1 in the last equality. This completes the proof of Theorem 2.  $\square$

The moment conditions on  $T$  and on  $f(X_1, X_2)$  for the case  $p = 2$  in Theorem 2 are

$$(4) \quad ET < \infty \quad \text{and} \quad Ef^2(X_1, X_2) < \infty.$$

Comparison of (4) with the moment conditions,  $ET < \infty$  and  $E|f(X_1)| < \infty$ , of Theorem 1 dealing with  $U$ -statistics of degree 1 suggests that extension of Wald's equation to  $U$ -statistics of degree  $k$  may involve higher moment conditions on  $f(X_1, \dots, X_k)$  and/or  $T$ . For the special case of a multilinear kernel  $f(x_1, \dots, x_k) = x_1 \cdots x_k$ , Chow, de la Peña and Teicher [CdlPT] established that  $E(\sum_{1 \leq i_1 < \dots < i_k \leq T} X_{i_1} \cdots X_{i_k}) = 0$  under the moment conditions

$$(5) \quad ET^{k-1} < \infty \quad \text{and} \quad EX_1^2 < \infty, \quad EX_1 = 0,$$

or more generally, under  $ET^{(k-1)/(p-1)} < \infty$  and  $E|X_1|^p < \infty$  for some  $1 < p \leq 2, EX_1 = 0$ . Concerning the sharpness of these results, observe that the case  $k = 2, p = 2$  along with Wald's first equation give Wald's second equation (involving the second moment of the randomly stopped sum). We are able to extend Theorem 2 to  $U$ -statistics of degree  $k > 2$ , with symmetric kernels and under the moment conditions  $E|f(X_1, \dots, X_k)|^p < \infty$  for some  $1 < p \leq 2$ ,

$$E\{f(X_1, \dots, X_k) | X_1, \dots, X_{h-1}, X_{h+1}, \dots, X_k\} = 0 \text{ a.s.}$$

for all  $1 \leq h \leq k$ , and  $ET^{\rho(k,p)} < \infty$ , where

$$\rho(k, p) = (p^{1/(k-1)} - 1)^{-1} < (k - 1)/\ln p.$$

The proof, which involves deeper decoupling ideas and is considerably more complicated than that of Theorem 2, will be presented elsewhere together with the decoupling results needed.

### 3. MEANS OF NORMALIZED $U$ -STATISTICS IN SEQUENTIAL EXPERIMENTS

For notational simplicity we shall again focus on  $U$ -statistics of degree 2 in this section, since the same arguments can be applied to the general case. Let  $g : S \times S \rightarrow \mathbb{R}$  be a symmetric Borel measurable function. As pointed out in Section 1, although the normalized  $U$ -statistic

$$(6) \quad U_n = \sum_{1 \leq i < j \leq n} g(X_i, X_j) / \binom{n}{2}$$

based on i.i.d. observations  $X_1, \dots, X_n$  is an unbiased estimate of  $\theta = Eg(X_1, X_2)$ , the corresponding  $U$ -statistic  $U_T$  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. We shall analyze the bias  $EU_T - \theta$  in sequential experiments whose stopping rules are of the form

$$(7) \quad T = \inf\{n \geq 2 : \sum_{i=1}^n Y_i + \xi_n \geq a\},$$

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

$$(8) \quad EY_1 > 0, \quad EY_1^2 < \infty, \quad \sum_1^\infty nP\{\xi_n < (d - EY_1)n\} < \infty,$$

(9)

$\{[(\sum_1^n Y_i + \xi_n - n/c)^+]^p, n \geq 1\}$  is uniformly integrable for some  $p \geq 3$ ,

$$(10) \quad \limsup_{\delta \rightarrow 0} P\{\max_{n \geq 1} \max_{k \leq n\delta} |\xi_{n+k} - \xi_n| > \epsilon\} = 0 \quad \text{for all } \epsilon > 0.$$

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

(11)

$$\sum_{n=1}^{\infty} nP\left(\bigcup_{k=n}^{\infty} A'_k\right) < \infty \text{ and } \{\max_{k \leq n} |\xi_{n+k} 1(A_{n+k})|^\alpha, n \geq 1\} \text{ is uniformly integrable,}$$

where  $A'_k$  denotes the complement of  $A_k$ .

Stopping rules of the form (7) are commonly used in sequential testing problems. Here the primary objective of the sequential experiment is to test a certain statistical hypothesis, and the experiment terminates as soon as the test statistic  $Z_n$  exceeds some threshold  $a$ . Typically  $Z_n$  can be represented in the form of a random walk  $\sum_{i=1}^n Y_i$  plus a remainder  $\xi_n$ . Regularity conditions of the type (10) and (11) on  $\xi_n$  were first introduced by Lai and Siegmund [LS] to develop a renewal theory for the perturbed random walk  $\sum_1^n Y_i + \xi_n$  and have been shown to hold for many commonly used test statistics in sequential analysis, cf. [S2], [W2]. In addition to the primary objective of testing the hypothesis, the experiment also has secondary objectives of estimating certain parameters of interest, and therefore the problem of bias in estimation with  $U$ -statistics or other commonly used estimators following a sequential test is of fundamental interest in sequential analysis. Stopping rules of the form (7) also arise naturally in sequential estimation problems. Here the primary objective of the sequential experiment is to estimate certain parameters with fixed accuracy or with an optimal balance between sampling cost and mean squared error, and stopping rules of the form (7) provide asymptotically efficient solutions to these problems, cf. [AW], [W1], [W2], [S2].

Aras and Woodroffe [AW, Theorem 2] recently proved the following asymptotic formula for the bias of the stopped sample mean  $\bar{W}_T = T^{-1} \sum_1^n W_i$  as  $a \rightarrow \infty$ :

$$(12) \quad E(\bar{W}_T) = E(W_1) + a^{-1}\{\text{Cov}(W_1, Y_1) + o(1)\},$$

where  $T (= T_a)$  is the stopping time defined in (7) with the  $Y_i$  and  $\xi_i$  satisfying (8)–(11), and  $(W_1, Y_1), (W_2, Y_2), \dots$  are i.i.d. bivariate random vectors such that  $E|W_1|^q < \infty$  for some  $q \geq \max\{4, 2\alpha/(\alpha - 1), 2p/(p - 2)\}$ . The following theorem extends (12) to the case where the sample mean  $\bar{W}_n$  is replaced by the normalized  $U$ -statistic (6).

**Theorem 3.** *Assume (8)–(11) and define  $T (= T_a)$  by (7). Let  $g : S \times S \rightarrow \mathbb{R}$  be a symmetric Borel measurable function such that*

$$(13) \quad E|g(X_1, X_2)|^q < \infty \text{ for some } q \geq \max\{6, 2\alpha/(\alpha - 1)\},$$

and define  $U_n$  by (6). Let  $\psi(x) = Eg(X_1, x), \theta = Eg(X_1, X_2)$ . Then as  $a \rightarrow \infty$ ,

$$(14) \quad E(U_T) = \theta + 2a^{-1}\{\text{Cov}(\psi(X_1), Y_1) + o(1)\}.$$

*Proof.* Let  $\mu = EY_1 (> 0)$  and assume without loss of generality that  $c < 1/\mu$  in (9). By Proposition 2 of [AW],  $T/a \rightarrow 1/\mu$  a.s. By (1),

$$(15) \quad U_T - \theta = 2T^{-1} \sum_{i=1}^T \psi(X_i) + 2\{T(T-1)\}^{-1} \sum_{1 \leq i < j \leq T} f_2(X_i, X_j),$$

noting that  $T \geq 2$  by (7). The function  $f_2$  in (15) is symmetric, and

$$E\{f_2(X_1, X_2) | X_1\} = 0 \quad \text{a.s.}$$

By (13) and (12), as  $a \rightarrow \infty$ ,

$$(16) \quad E\{T^{-1} \sum_{i=1}^T \psi(X_i)\} = a^{-1}\{\text{Cov}(\psi(X_1), Y_1) + o(1)\}.$$

In view of (15) and (16), (14) follows if it can be shown that

$$(17) \quad \lim_{a \rightarrow \infty} aE\tilde{U}_T = 0, \quad \text{where } \tilde{U}_n = \{n(n-1)\}^{-1} \sum_{1 \leq i < j \leq n} f_2(X_i, X_j).$$

By Lemma 1 of [LW],  $E|\sum_{1 \leq i < j \leq m} f_2(X_i, X_j)|^q = O(m^q)$ . Therefore  $E|\tilde{U}_m|^q = O(m^{-q})$ . Since  $\{\tilde{U}_n, n \geq 1\}$  is a reverse martingale (cf. [GS]), it then follows from Doob's inequality that

$$(18) \quad E(\sup_{n \geq m} |\tilde{U}_n|^q) = O(E|\tilde{U}_m|^q) = O(m^{-q}).$$

By Hölder's inequality and (18), since  $P\{T \geq 2a/\mu\} \rightarrow 0$ ,

$$(19) \quad E(|\tilde{U}_T|1\{T \geq 2a/\mu\}) \leq [E(\sup_{n \geq 2a/\mu} |\tilde{U}_n|^q)]^{1/q} [P\{T \geq 2a/\mu\}]^{1-q^{-1}} = o(a^{-1}).$$

Let  $\tau = \min\{T, [2a/\mu]\}$  and note that  $\tilde{U}_T - \tilde{U}_\tau = (\tilde{U}_T - \tilde{U}_{[2a/\mu]})1\{T > 2a/\mu\}$ . In view of (18) and (19), (17) would follow if it can be shown that  $\lim_{a \rightarrow \infty} aE\tilde{U}_\tau = 0$ . Since  $E\{\tau(\tau-1)\tilde{U}_\tau\} = 0$  by Theorem 2, it suffices to show that

$$(20) \quad \lim_{a \rightarrow \infty} E\{(a - \frac{\mu^2\tau(\tau-1)}{a})\tilde{U}_\tau\} = 0.$$

Using (18) and an argument similar to the proof of Proposition 6 of [AW], it can be shown that as  $a \rightarrow \infty$ ,

$$(21) \quad E(|a - \mu^2\tau(\tau-1)/a| |\tilde{U}_\tau| 1\{\tau \leq ca/4\}) \rightarrow 0.$$

Moreover, noting that  $\tau \leq 2a/\mu$ ,

$$(22) \quad \begin{aligned} E(|(a - \frac{\mu^2\tau(\tau-1)}{a})\tilde{U}_\tau|^q 1\{ca/4 \leq \tau \leq 2a/\mu\}) \\ \leq (3a)^q E(\sup_{n \geq ca/4} |\tilde{U}_n|^q) = O(1), \end{aligned}$$

by (18). Since  $(\tau - a/\mu)/\sqrt{a} = O_P(1)$  (cf. [AW]) and since  $\sup_{n \geq ca/4} |\tilde{U}_n| = O_P(a^{-1})$  by (18), it follows that  $\{a - \mu^2\tau(\tau-1)/a\}\tilde{U}_\tau 1\{\tau \geq ca/4\} \xrightarrow{P} 0$ . Combining this with the uniform integrability result implied by (22) then yields

$$(23) \quad \lim_{a \rightarrow \infty} E(|a - \mu^2\tau(\tau-1)/a| |\tilde{U}_\tau| 1\{\tau \geq ca/4\}) = 0.$$

From (21) and (23), (20) follows. □



## REFERENCES

- [AW] Aras, G. and Woodroffe, M., *Asymptotic expansions for the moments of a randomly stopped average*, Ann. Statist. **21** (1993), 503–519. MR **94a**:62120
- [CdIPT] Chow, Y. S., de la Peña, V. H. and Teicher, H., *Wald's equation for a class of de-normalized  $U$ -statistics*, Ann. Probab. **21** (1993), 1151–1158. MR **94c**:60064
- [CT] Chow, Y. S., and Teicher, H., *Probability Theory: Independence, Interchangeability, Martingales*, Springer-Verlag, New York, 1978. MR **80a**:60004
- [Co] Cox, D. R., *A note on sequential estimation of means*, Proc. Cambridge Philos. Soc. **48** (1952), 447–450. MR **14**:190e
- [GS] Grams, W. F. and Serfling, R. J., *Convergence rates for  $U$ -statistics and related statistics*, Ann. Statist. **1** (1973), 153–160. MR **49**:1561
- [Ha] Halmos, P. R., *The theory of unbiased estimation*, Ann. Math. Statist. **17** (1946), 34–43. MR **7**:463g
- [H] Hoeffding, W., *A class of statistics with asymptotically normal distribution*, Ann. Math. Statist. **19** (1948), 293–325. MR **10**:134g
- [K] Klass, M. J., *A method of approximating expectations of functions of sums of independent random variables*, Ann. Probab. **9** (1981), 413–428. MR **82f**:60119
- [LS] Lai, T. L. and Siegmund, D., *A nonlinear renewal theory with applications to sequential analysis I*, Ann. Statist. **5** (1977), 946–954; **II**, Ann. Statist. **7** (1979), 60–76. MR **56**:3935; MR **80c**:62100
- [LW] Lai, T. L. and Wang, J. Q., *Edgeworth expansions for symmetric statistics with applications to bootstrap methods*, Statistica Sinica **3** (1993), 517–542. MR **95c**:62019
- [S1] Siegmund, D., *Estimation following sequential tests*, Biometrika **65** (1978), 341–349. MR **80f**:62079
- [S2] Siegmund, D., *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlag, New York, 1985. MR **87h**:62145
- [W1] Woodroffe, M., *Second order approximation for sequential point and interval estimation*, Ann. Statist. **5** (1977), 985–995. MR **58**:13548
- [W2] Woodroffe, M., *Nonlinear Renewal Theory in Sequential Analysis*, SIAM, Philadelphia, 1982. MR **83j**:62118
- [Z] Zinn, J., *Comparison of martingale difference sequences*, Probability in Banach spaces V, Lecture Notes in Math., vol. 1153, Springer-Verlag, Berlin, 1985, pp. 453–457. MR **87f**:60070

DEPARTMENT OF STATISTICS, COLUMBIA UNIVERSITY, 617 MATHEMATICS BLDG., NEW YORK, NEW YORK 10027

*E-mail address:* [vp@wald.stat.columbia.edu](mailto:vp@wald.stat.columbia.edu)

DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY, SEQUOIA HALL, STANFORD, CALIFORNIA 94305-4065

*E-mail address:* [karola@playfair.stanford.edu](mailto:karola@playfair.stanford.edu)