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

BY

VICTOR H. DE LA PEÑA and TZE LEUNG LAI

TECHNICAL REPORT L94-6

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT DMS-9403794

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
WALD’S EQUATION AND ASYMPTOTIC BIAS OF RANDOMLY STOPPED U-STATISTICS

VICTOR H. DE LA PEÑA AND TZE LEUNG LAI

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.


Key words and phrases. Hoeffding decomposition, decoupling, martingales, Wald’s equation, stopping times.
1. INTRODUCTION

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

$$U_n = \{ \sum_{1 \leq i_1 < \cdots < i_k \leq n} g(X_{i_1}, \ldots, X_{i_k}) / n(n-1) \cdots (n-k+1) \},$$

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 $\theta(= \theta(F)) = \int \cdots \int g(x_1, \ldots, x_k) dF(x_1) \cdots dF(x_k)$ of the population distribution $F$. Hoeffding [H] subsequently developed a comprehensive theory of $U$-statistics. In particular, he introduced the basic decomposition

$$U_n - \theta = n^{-1} \sum_{i=1}^{n} f_1(X_i) + \{ n(n-1)^{-1} \sum_{1 \leq i < k \leq n} f_2(X_i, X_j) + \cdots$$

$$+ \{ n \cdots (n-k+1)^{-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} f_k(X_{i_1}, \ldots, X_{i_k}) \},$$

in which $Ef_1(X_1) = 0$ and $f_j$ is symmetric with $Ef_j(X_1, \ldots, X_j | X_1, \ldots, X_{j-1}) = 0$ for $2 \leq j \leq k$. In fact, $f_1(x) = Eg(x, X_2, \ldots, X_k) - \theta, f_2(x, y) = Eg(x, y, X_3, \ldots, X_k) - f_1(x) - f_1(y) + \theta, \text{ etc.}$ 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 nonparametric 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", 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 $E W_i = \mu$, and let $S_n = \sum_{i=1}^{n} (W_i - \mu)$. Let $T$ be a stopping time adapted to $\{W_i\}$ such that $E T < \infty$. Then $E S_T = 0$. 

1
In Section 2 we generalize Theorem 1 to the (de-normalized) \( U \)-statistics \( S_n = \sum_{1 \leq i_1 < \cdots < i_j \leq n} f_j(X_{i_1}, \cdots, X_{i_j}) \). The generalization involves much deeper probabilistic ideas, including recent results from decoupling theory. By using a different method, Chow, de la Peña and Teicher [CPdT] recently generalized Theorem 1 to the multilinear \( U \)-statistics \( S_n = \sum_{1 \leq i_1 < \cdots < i_k \leq n} X_{i_1} \cdots X_{i_k} \). 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 on 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

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

Let \( f : S \times S \to \mathbb{R} \) be a Borel measurable function such that

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

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

**Proof.** By the optimal sampling theorem of martingales, the desired conclusion holds if we replace \( T \) by the bounded stopping time \( \min(T, n) \). Hence by the dominated convergence theorem, the desired conclusion follows if it can be shown that

\[
E \max_{n \leq T | \sum_{1 \leq i < j \leq n} f(X_i, X_j)| < \infty.
\]

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.

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

\[
d_j = (\sum_{i=1}^{j-1} f(X_i, X_j))^2 1(T \geq j), \quad e_j = (\sum_{i=1}^{j-1} f(X_i, \tilde{X}_j))^2 1(T \geq j).
\]
obtain
\[(I) = CE\left\{ \sum_{j=1}^{T} \sum_{i=1}^{T} f(X_i, \bar{X}_j) \right\} = CE\left\{ T^{1/p} \frac{\sum_{j=1}^{T} \sum_{i=1}^{T} f(X_i, \bar{X}_j)}{T^{1/p}} \right\}\]
\[\leq C(ET^{-1/p})^{\frac{p-1}{p}} \left\{ E \frac{\sum_{j=1}^{T} \sum_{i=1}^{T} f(X_i, \bar{X}_j)^p}{T^{1/p}} \right\}^{1/p}, \text{ by B-D-G,}\]
\[\leq C(ET^{-1/p})^{\frac{p-1}{p}} \left\{ E \frac{\sum_{j=1}^{T} \sum_{i=1}^{T} f(X_i, \bar{X}_j)^p}{T^{1/p}} |\sigma(\{X_r\}, T)\right\}^{1/p}, \text{ by B-D-G,}\]
\[\leq C(ET^{-1/p})^{\frac{p-1}{p}} \left\{ E \sum_{i=1}^{T} f(X_i, \bar{X}_1)^p \sigma(\{X_r\}, T) \right\}^{1/p} = C(ET^{-1/p})^{\frac{p-1}{p}} (ET E f(X_1, \bar{X}_1)^p)^{1/p},\]
by concavity of the function \(f(x) = x^{p/2}\), and by Theorem 1 applied conditionally.

To bound (II), recall that \(\{\bar{X}_i\}\) and \(\{\tilde{X}_i\}\) are independent copies of \(\{X_i\}\), and set
\[\tilde{d}_j = (\sum_{i=1}^{j-1} f(X_j, \bar{X}_i))^2 1(T \geq j), \quad \tilde{\varepsilon}_j = (\sum_{i=1}^{j-1} f(\tilde{X}_j, \bar{X}_i))^2 1(T \geq j).\]

Then \(\{\tilde{\varepsilon}_j\}\) and \(\{\tilde{d}_j\}\) are tangent with respect to \(\{\tilde{\mathcal{F}}_n\}\), where
\[\tilde{\mathcal{F}}_n = \sigma(X_1, ..., X_n; \bar{X}_1, ..., \bar{X}_n; \tilde{X}_1, ..., \tilde{X}_n).\]

Next, the B-D-G inequality and (3) yield
\[(II) = CE|\sum_{1 \leq i < j \leq T} f(X_j, \bar{X}_i)| = CE|\sum_{j=2}^{T} \sum_{i=1}^{j-1} f(X_j, \bar{X}_i)|\]
\[\leq CE \sqrt{\sum_{j=2}^{\infty} \sum_{i=1}^{j-1} f(X_j, \bar{X}_i)^2 1(T \geq j)}\]
\[\leq CE \sqrt{\sum_{j=2}^{\infty} \sum_{i=1}^{j-1} f(\tilde{X}_j, \bar{X}_i)^2 1(T \geq j)} \leq CE|\sum_{j=2}^{T} \sum_{i=1}^{j-1} f(\tilde{X}_j, \bar{X}_i)|,\]
where the last inequality follows by the (lower) Burkholder-Davis-Gundy inequality and Lévy's inequality (cf. [CT]). Observe that we have now completely decoupled the stopping time. Moreover,
by the B-D-G inequality and the concavity of \( f(x) = x^{p/2} \),

\[
E[\sum_{j=2}^{n} \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i)] \leq \{E[\sum_{j=2}^{n} \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i)|T]|^{p/2} \}^{1/2} \\
\leq C\{E(\sum_{j=2}^{n} \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i)^p)^{1/2} \}^{1/2} \leq C\{\sum_{j=2}^{n} E[\sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i)|T]|^{p/2} \}^{1/2} \\
\leq C\sum_{j=2}^{n} \{j-1\} \{E[\sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i)|T]|^{p/2} \}^{1/2} \leq Cn^{1/2} \{E[|f(X_1, X_2)|T]|^{p/2} \}^{1/2}.
\]

Putting these two observations together, we get

\[
(II) = CE[\sum_{1 \leq i < j \leq T} f(X_j, X_i)] \leq CE[\sum_{j=2}^{T} \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i)] \\
= C\sum_{n=1}^{\infty} E[\sum_{j=2}^{n} \sum_{i=1}^{j-1} f(\tilde{X}_j, \tilde{X}_i)|T = \infty] P(T = n) \\
\leq C\{E[|f(X_1, X_2)|T]|^{p/2} \}^{1/2} \sum_{n=1}^{\infty} n^{1/2} P(T = n) \\
= C\{E[|f(X_1, X_2)|T]|^{p/2} \}^{1/2} ET^{1/2} \leq C\{E[|f(X_1, X_2)|T]|^{p/2} \}^{1/2} ET^{1/2}.
\]

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

\[
E|\sum_{i=1}^{T} f(X_i, \tilde{X}_i)| \leq \{E[\sum_{i=1}^{T} |f(X_i, \tilde{X}_i)|T]|^{p/2} \}^{1/2} \leq C\{E(\sum_{i=1}^{T} f^2(X_i, \tilde{X}_i))^{\frac{p}{2}} \}^{1/2} \\
\leq C\{E\sum_{i=1}^{T} |f(X_i, \tilde{X}_i)|T]|^{p/2} \}^{1/2} = C(ET|f(X_1, \tilde{X}_i)|T]|^{p/2} \}^{1/2},
\]

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

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

\[
(4) \quad ET < \infty \quad \text{and} \quad E f^2(X_1, X_2) < \infty.
\]

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 \( f(X_1, \cdots, X_k) \) and/or \( T \). For the special case of a multilinear kernel \( f(x_1, \cdots, x_k) = x_1 \cdots x_k \), Chow, de la Peña and Teicher [CPdT] established that \( E(\sum_{1 \leq i_1 < \cdots < i_k \leq T} X_{i_1} \cdots X_{i_k}) = 0 \) under the moment conditions

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

5
or more generally, under $ET^{(k-1)/(p-1)} < \infty$ and $E|X_1|^p < \infty$ for some $1 < p \leq 2$, $EX_1 = 0$. We are able to extend Theorem 2 to $U$-statistics of degree $k > 2$ under the moment conditions $Ef^2(X_1, \cdots, X_k) < \infty$, $E\{f(X_1, \cdots, X_k)X_1, \cdots, X_{h-1}, X_{h+1}, \cdots, X_k\} = 0$ a.s. for all $1 \leq h \leq k$ and $ETr^{(k)} < \infty$. 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 $g : S \times S \to \mathbb{R}$ be a symmetric Borel measurable function. As pointed out in Section 1, although the normalized $U$-statistic

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

based on i.i.d. observations $X_1, \cdots, X_n$ is an unbiased estimate of $\theta = Eg(X_1, X_2)$, the corresponding $U$-statistic $U_T$ based on a sample $\{X_1, \cdots, X_T\}$ 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

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

where $(X_1, Y_1), (X_2, Y_2), \cdots$ are i.i.d. random vectors and

$$EY_1 > 0, \ EY_1^2 < \infty, \ \sum_{i=1}^{\infty} nP\{\xi_n < -\delta n\} < \infty \text{ for some } 0 < \delta < EY_1,$$

$$\{(\sum_{i=1}^{n} Y_i + \xi_n - cn)^+\}, \ n \geq 1 \text{ is uniformly integrable for some } c > EY_1,$$

$$\lim_{\delta \to 0} \sup_{n \geq 1} P\{\max_{0 \leq k \leq n} |\xi_{n+k} - \xi_n| > \epsilon\} = 0 \text{ for all } \epsilon > 0.$$

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

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

where $A'_n$ denotes the complement of $A_n$.

Stopping rules of the form (7) are commonly used in sequential testing problems. Here the primary objective of the sequential experiment is to test 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 walks $\sum_{i=1}^{n} Y_i + \xi_n$ 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 Woodroofe [AW, Theorem 2] recently proved the following asymptotic formula for the bias of the stopped sample mean $\bar{W}_T = T^{-1} \sum_{i=1}^{n} W_i$ as $a \to \infty$:

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

where $T(= T_a)$ is the stopping time defined in (7) with the $Y_i$ and $\xi_i$ satisfying (8)–(11) and $(W_1, Y_1), (W_2, Y_2), \cdots$ are i.i.d. bivariate random vectors such that $E|W_1|^q < \infty$ for some $q \geq \max\{6, 2\alpha/(\alpha - 1)\}$. 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 $g : S \times S \to \mathbb{R}$ be a symmetric Borel measurable function such that

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

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

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

**Proof.** Let $\mu = EY_1(> 0)$. By Proposition 2 of [AW], $T/a \to 1/\mu$ a.s. By (1),

$$U_T - \theta = T^{-1} \sum_{i=1}^{T} \psi(X_i) + \{T(T - 1)\}^{-1} \sum_{1 \leq i < j \leq T} f_2(X_i, X_j),$$
noting that \( T \geq 2 \) by (7). The function \( f_2 \) in (15) is symmetric and \( E\{f_2(X_1, X_2)|X_1\} = 0 \) a.s. By (13) and (12), as \( a \to \infty \),

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

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

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

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

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

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

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

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 \( \lim_{a \to \infty} aE\tilde{U}_\tau = 0 \). Since \( E\{\tau(\tau - 1)\tilde{U}_\tau\} = 0 \) by Theorem 2, it suffices to show that

\[
\lim_{a \to \infty} E\{(a - \frac{\tau(\tau - 1)}{a\mu^2})\tilde{U}_\tau\} = 0.
\]

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

\[
E(|a - \tau(\tau - 1)/(a\mu^2)||\tilde{U}_\tau|1\{\tau \leq c^{-1}a/4\}) \to 0,
\]

where \( c > \mu \) is given by (9). Moreover,

\[
E\{(a - \frac{\tau(\tau - 1)}{a\mu^2})\tilde{U}_\tau|1\{c^{-1}a/4 \leq \tau \leq 2a/\mu\}\}) \leq (a + 4a/\mu^4)^q E(\sup_{n \geq c^{-1}a/4} |\tilde{U}_n|^q) = O(1),
\]

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

\[
\lim_{a \to \infty} E\{(a - \tau(\tau - 1)/(a\mu^2)||\tilde{U}_\tau|1\{\tau \geq c^{-1}a/4\}) = 0.
\]
From (21) and (23), (20) follows.

REFERENCES


