The Annals of Statistics
1985, Vol. 13, No. 1, 228–235

SEQUENTIAL NONPARAMETRIC FIXED-WIDTH CONFIDENCE INTERVALS FOR U-STATISTICS[1]

BY RAYMOND N. SPROULE

Malaspina College

A sequential fixed-width confidence interval for the mean of a U-statistic, having coverage probability approximately equal to preassigned α, is presented. The main result, Theorem 2, shows that the sequential procedure is asymptotically efficient in the sense of Chow and Robbins (1965) and assumes only finiteness of the second moment of the kernel, the weakest possible condition. The paper follows naturally from Sproule (1974) and Sproule (1969), the primary reference.

1. Introduction.

Let $X_1, \cdots, X_n$ be independent and identically distributed (i.i.d.) random variables having d.f. F. Let $f(x_1, \cdots, x_r)$ be a symmetric (measurable) function of r arguments. Then for $n \geq r$, Hoeffding (1948) defines a U-statistic by

$$U_n = \binom{n}{r}^{-1} \sum{}^{(n,r)} f(X_{\alpha_1}, \cdots, X_{\alpha_r}) \tag{1}$$

where $\sum^{(n,r)}$, here and in the sequel, represents the summation over all combinations $(\alpha_1, \cdots, \alpha_r)$ formed from the integers $\{1, 2, \cdots, n\}$. Thus U_n is an "average" of the function $f(x_1, \cdots, x_r)$ over the random sample $X_1, \cdots, X_n$. Particular examples are the sample mean (where $r = 1$ and $f(x_1) = x_1$) and the sample variance (where $r = 2$ and $f(x_1, x_2) = (x_1 - x_2)^2/2$). Let $\theta = E\{f(X_1, \cdots, X_r)\}$ so that $E\{U_n\} = \theta$. We develop a sequential confidence interval for θ of fixed-width $2d$, where $d > 0$, and such that the coverage probability approaches (as $d \to 0$) a specified α, where $0 < \alpha < 1$. Chow and Robbins (1965) solve the problem for the sample mean using $n^{-1}s_n^2$ where s_n^2 is the sample variance to estimate the unknown variance of the sample mean. We introduce an estimate for the unknown variance of the U-statistic and then consider a sequential procedure.

2. Estimation of the variance of U_n.

Define

$$f_c(x_1, \cdots, x_c) = E\{f(x_1, \cdots, x_c, X_{c+1}, \cdots, X_r)\}$$

for $c = 1, 2, \cdots, r$. Note that $f_r(x_1, \cdots, x_r) = f(x_1, \cdots, x_r)$. We interpret $E\{f(x_1, \cdots, x_c, X_{c+1}, \cdots, X_r)\}$ as the conditional expectation of $f(X_1, \cdots, X_r)$

Received April 1983; revised August 1984.

[1] Supported, in part, by the U.S. Public Health Service under grant GM-10397 at the University of North Carolina, Chapel Hill.

AMS 1980 *subject classifications*. Primary 62G15; secondary 62E20.

Key words and phrases. U-Statistics, asymptotic, large sample, fixed width confidence interval, generalized mean, sequential estimation, efficient, consistent, stopping variable, martingale, central limit theorem, nonparametric, distribution free, sample average.

given that $X_1, \cdots, X_c$ are fixed at the values $x_1, \cdots, x_c$, respectively. Clearly $E\{f_c(X_1, \cdots, X_c)\} = \theta$ for $c = 1, 2, \cdots, r$. Let $\rho_c = \text{Var}\{f_c(X_1, \cdots, X_c)\}$ for $c = 1, 2, \cdots, r$. In particular $f_1(x_1) = E\{f(x_1, X_2, \cdots, X_r)\}$ and $\rho_1 = \text{Var}\{f_1(X_1)\}$. If $E\{f(X_1, \cdots, X_r)\}^2 < \infty$ then it follows from Hoeffding (1948) that the variance of U_n can be represented by

$$\text{Var}\{U_n\} = n^{-1}r^2\rho_1 + O(n^{-2}). \tag{2}$$

If the terms of order n^{-2} in (2) can be considered negligible, the problem of estimating $\text{Var}\{U_n\}$ reduces to that of estimating the usually unknown functional ρ_1.

For each $i = 1, 2, \cdots, n$ define a U-statistic based on $X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n$ by

$$U_{(i)n} = \binom{n-1}{r}^{-1} \textstyle\sum_i^{(n-1,r)} f(X_{\alpha_1}, \cdots, X_{\alpha_r})$$

where the summation $\sum_i^{(n-1,r)}$ is over all combinations $(\alpha_1, \cdots, \alpha_r)$ formed from $\{1, 2, \cdots, i-1, i+1, \cdots, n\}$. Define the W-statistics by setting $W_{in} = nU_n - (n-r)U_{(i)n}$ for $i = 1, 2, \cdots, n$ and notice that they are identically distributed. Furthermore, $\bar{W}_n = n^{-1}\sum_{i=1}^n W_{in} = rU_n$. Let

$$s_{wn}^2 = (n-1)^{-1} \textstyle\sum_{i=1}^n (W_{in} - \bar{W}_n)^2.$$

It is well known that if $E\{f(X_1, \cdots, X_r)\}^2 < \infty$, then, letting $\sigma^2 = r^2\rho_1$,

$$\lim_{n\to\infty} s_{wn}^2 = \sigma^2 \quad \text{(a.s.)} \tag{3}$$

For more details refer to Sproule (1969) and Sen (1977). A first-order expression for the mean and variance of s_{wn}^2 is given in Sproule (1969). For $c = 0, 1, \cdots, r$ define

$$q^{(c)}(x_1, \cdots, x_{2r-c}) = \binom{2r-c}{r}^{-1}\binom{r}{c}^{-1} \textstyle\sum^{(c)} f(x_{\alpha_1}, \cdots, x_{\alpha_r}) f(x_{\beta_1}, \cdots, x_{\beta_r})$$

where the summation $\sum^{(c)}$ is over all combinations $(\alpha_1, \cdots, \alpha_r)$ and $(\beta_1, \cdots, \beta_r)$ each formed from $\{1, 2, \cdots, 2r-c\}$ and such that there are exactly c integers in common.

Then, for each $c = 0, 1, \cdots, r$ define a U-statistic by

$$U_n^{(c)} = \binom{n}{2r-c}^{-1} \textstyle\sum^{(n,2r-c)} q^{(c)}(X_{\alpha_1}, \cdots, X_{\alpha_{2r-c}}).$$

Let $\rho_0 = 0$. Then $E\{U_n^{(c)}\} = E\{q^{(c)}(X_1, \cdots, X_{2r-c})\} = \rho_c + \theta^2$ for $c = 0, 1, \cdots, r$. Next, by a direct combinatorial argument,

$$s_{wn}^2 = (n-1)^{-1} n \binom{n}{r}^{-1} \textstyle\sum_{c=0}^r \binom{r}{c}\binom{n-r}{r-c} [cn - r^2] U_n^{(c)}. \tag{4}$$

A rearrangement of (4) then yields

$$s_{wn}^2 = r^2(U_n^{(1)} - U_n^{(0)}) + \textstyle\sum_{c=0}^r \alpha_n(c) U_n^{(c)} \tag{5}$$

where $\alpha_n(c) = O(n^{-1})$ for $c = 0, 1, \cdots, r$. The representation for s_{wn}^2 in expression

(5) will be required in the proof of Theorem 2. Note that we will not need an explicit expression for $\alpha_n(c)$ although one is readily available.

3. A confidence interval for θ. Let $\Phi(x)$ denote the standard normal d.f. and let $0 < \alpha < 1$. Define a constant $a > 0$ by setting $\Phi(a) = (\alpha + 1)/2$. Let $\{a_n\}$ be a sequence of positive real numbers such that $\lim_{n\to\infty} a_n = a$. For $d > 0$ define the stopping variable

$$N(d) = \text{smallest integer } k \geq r \text{ such that } s_{wk}^{2*} \leq kd^2 a_k^{-2} \tag{6}$$

where $s_{wk}^{2*} = s_{wk}^2 + k^{-\gamma}$ for suitably chosen $\gamma > 0$. The $k^{-\gamma}$ term makes $N(d)$ a "delayed" stopping variable and prevents very early stopping in situations where the d.f. F is discrete and there is a very high probability that s_{wk}^2 is very small. Chow and Robbins (1965) chose $\gamma = 1$. Define a closed confidence interval $I_N = [U_N - d, U_N + d]$ of width $2d$. Then the following theorem is generally useful:

THEOREM 1. *Assume* $E\{f(X_1, \cdots, X_r)\}^2 < \infty$ *and* $\rho_1 > 0$. *Then*

(i) $N(d)$ *is well-defined and is a nonincreasing function of* d,

(ii) $\lim_{d\to 0} N(d) = \infty$ *(a.s.)*,

(iii) $\lim_{d\to 0} E\{N(d)\} = \infty$, *and*

(iv) $\lim_{d\to 0} a^{-2}\sigma^{-2}d^2 N(d) = 1$ *(a.s.)*.

(v) $\lim_{d\to 0} P\{\theta \in I_{N(d)}\} = \alpha$.

PROOF. From (3) we obtain $\lim_{n\to\infty} s_{wn}^{2*} = \sigma^2$ (a.s.). Let $y_n = \sigma^{-2}s_{wn}^{2*}$, $f(n) = a_n^{-2}na^2$ and $t = d^{-2}a^2\sigma^2$. Then parts (i)–(iv) of the theorem follow from Lemma 1 of Chow and Robbins (1965).

Let N_t be defined by (6) with d^2 replaced by $t^{-1}a^2\sigma^2$. (Note that $N_t = N(d)$.) Part (v) then follows from Theorem 6 of Sproule (1974) by identifying t with n_s and N_t with N_s.

The main theorem is

THEOREM 2. *Assume* $E\{f(X_1, \cdots, X_r)\}^2 < \infty$ *and* $\rho_1 > 0$. *Then*

$$\lim_{d\to 0} d^2 a^{-2}\sigma^{-2} E\{N(d)\} = 1. \tag{7}$$

The sequential procedure may be simply described as follows: at each stage of sampling, the U-statistic U_n and an estimate of its variance are calculated, and sampling is terminated as soon as the approximate coverage probability for the interval $[U_n - d, U_n + d]$, based on a normal approximation, is at least α. The coverage probability is, in a certain sense, asymptotically α; that is, the sequential procedure is consistent (Theorem 1(v)). Also, the expected sample size of the procedure is asymptotically equal to the sample size of the corresponding non-sequential scheme used when the variance of the U-statistic is known (Theorem 2); that is, the sequential procedure is efficient.

Sequential fixed-width confidence intervals (of the Chow and Robbins, 1965, type) for the mean of a U-statistic first appeared in Sproule (1969). Results similar to Theorems 1 and 2 appeared in Sen and Ghosh (1981) but the stronger assumption of finiteness of the $(2 + \delta)$th moment of the kernel for some $\delta > 0$ was needed. Here, the strongest possible result is achieved by utilizing the reverse martingale property of U-statistics.

We now introduce two lemmas required in the proof of Theorem 2.

LEMMA 1. *If* $E\{|f(X_1, \cdots, X_r)|\} < \infty$, *then for any* $\varepsilon > 0$

$$E\{\sup_n n^{-\varepsilon} |U_n|\} < \infty.$$

PROOF. Truncate $f(\cdots)$ by letting

$$f'(X_{\alpha_1}, \cdots, X_{\alpha_r}) = \begin{cases} f(X_{\alpha_1}, \cdots, X_{\alpha_r}) & \text{if } |f(\cdots)| \le (\max_j \alpha_j)^{\varepsilon/2} \\ 0 & \text{otherwise} \end{cases}$$

and set

$$f''(X_{\alpha_1}, \cdots, X_{\alpha_r}) = f(X_{\alpha_1}, \cdots, X_{\alpha_r}) - f'(X_{\alpha_1}, \cdots, X_{\alpha_r}).$$

Then let

$$S_n = \textstyle\sum^{(n,r)} f(X_{\alpha_1}, \cdots, X_{\alpha_r}), \quad S'_n = \sum^{(n,r)} f'(X_{\alpha_1}, \cdots, X_{\alpha_r})$$

and

$$S''_n = \textstyle\sum^{(n,r)} f''(X_{\alpha_1}, \cdots, X_{\alpha_r}).$$

(a) To prove $E\{\sup_n n^{-(r+\varepsilon)} |S'_n|\} < \infty$, note that

$$\begin{aligned} \sup_n n^{-(r+\varepsilon)} |S'_n| &\le \sup_n n^{-(r+\varepsilon)} \textstyle\sum^{(n,r)} |f'(X_{\alpha_1}, \cdots, X_{\alpha_r})| \\ &\le \sup_n n^{-(r+\varepsilon)} \textstyle\sum^{(n,r)} (\max_j \alpha_j)^{\varepsilon/2} < \sup_n \sum^{(n,r)} (\max_j \alpha_j)^{-r-\varepsilon/2} \\ &\le \sup_n \textstyle\sum_{j=r}^{n} \binom{j-1}{r-1} j^{-r-\varepsilon/2} < \sum_{j=r}^{\infty} j^{-1-\varepsilon/2} < \infty. \end{aligned}$$

(b) Next,

$$\begin{aligned} &E\{\sup_n n^{-(r+\varepsilon)} |S''_n|\} \\ &\quad \le E\{\sup_n n^{-(r+\varepsilon)} \textstyle\sum^{(n,r)} |f''(X_{\alpha_1}, \cdots, X_{\alpha_r})|\} \\ &\quad = E\{\sup_n n^{-(r+\varepsilon)} \textstyle\sum_{j=r}^{n} \sum_{\alpha's}^{(j-1,r-1)} |f''(X_j, X_{\alpha_2}, \cdots, X_{\alpha_r})|\} \\ &\quad \le E\{\textstyle\sum_{j=r}^{\infty} j^{-(r+\varepsilon)} \sum_{\alpha's}^{(j-1,r-1)} |f''(X_j, X_{\alpha_2}, \cdots, X_{\alpha_r})|\} \\ &\quad \le \textstyle\sum_{j=r}^{\infty} j^{-(r+\varepsilon)} \sum_{\alpha's}^{(j-1,r-1)} E\{|f''(X_j, X_{\alpha_2}, \cdots, X_{\alpha_r})|\} \\ &\quad = \textstyle\sum_{j=r}^{\infty} j^{-(r+\varepsilon)} \binom{j-1}{r-1} \int_{[|f(\cdots)| > j^{\varepsilon/2}]} |f(x_1, \cdots, x_r)| \prod_{i=1}^{r} dF(x_i) \\ &\quad \le \textstyle\sum_{j=r}^{\infty} j^{-1-\varepsilon} b_j \le b_r \sum_{j=r}^{\infty} j^{-1-\varepsilon} < \infty \end{aligned}$$

where we have set

$$b_j = \int_{[|f(x_1,\cdots,x_r)|>j^{e/2}]} |f(x_1, \cdots, x_r)| \prod_{i=1}^r dF(x_i)$$

for $j = r, r+1, \cdots$, so that $b_j \geq 0$ and $b_j \geq b_{j+1}$.

(c) Finally, the lemma follows from (a) and (b) and the fact that $S_n = S_n' + S_n''$.

A positive integer-valued random variable M depending on $(X_1, X_2, \cdots)$ such that, for $n = 1, 2, \cdots$, the event $\{M = n\}$ is in $\mathscr{B}_n'$, the σ-field generated by $\{X_n, X_{n+1}, \cdots\}$, is called a "reverse stopping variable". The following lemma appears in Simons (1968) and is a consequence of Theorem 2.2 on page 302 of Doob (1953).

LEMMA 2. *Let $Z_{-m_2}, \cdots, Z_{-m_1}$ be a martingale where $-\infty < m_1 < m_2 \leq \infty$ and let M be a reverse stopping variable with $P\{m_1 \leq M \leq m_2\} = 1$. Then $E\{Z_{-M}\} = E\{Z_{-m_1}\}$.*

PROOF OF THEOREM 2. (a) As in Simons (1968), define a reverse stopping variable for $d > 0$

$$M = \begin{cases} \text{last integer } n \geq n_0 \\ \quad \text{such that } s_{wn}^{2*} > nd^2a_n^{-2} & \text{if there is such an } n \\ n_0 - 1 & \text{if } s_{wn}^{2*} \leq nd^2a_n^{-2} \text{ for all } n \geq n_0 \\ \infty & \text{if } s_{wn}^{2*} > nd^2a_n^{-2} \text{ infinitely often} \end{cases} \tag{8}$$

where $n_0 \geq r + 1$. Let I represent the indicator function and define t and N_t as in the proof of Theorem 1. Then for every $t > 0$

$$\begin{aligned} N_t &\leq n_0 I_{[M=n_0-1]} + (M+1) I_{[M \geq n_0]} \\ &= M I_{[M\geq n_0]} + n_0 I_{[M=n_0-1]} + I_{[M\geq n_0]} \\ &\leq d^{-2} a_M^2 s_{wM}^{2*} + n_0 I_{[M\geq n_0-1]} \leq t a^{-2} \sigma^{-2} a_M^2 s_{wM}^{2*} + n_0. \end{aligned}$$

Thus, for every $t > 0$,

$$t^{-1}E\{N_t\} \leq a^{-2}\sigma^{-2}E\{a_M^2 s_{wM}^{2*}\} + t^{-1}n_0. \tag{9}$$

(b) We prove $\lim_{t\to\infty} E\{s_{wM}^{2*}\} = \sigma^2$. From expression (5) we have

$$E\{s_{wM}^2\} = r^2 E\{U_M^{(1)} - U_M^{(0)}\} + \textstyle\sum_{c=0}^r E\{\alpha_M^{(c)} U_M^{(c)}\}. \tag{10}$$

Define $Z_{-n}^{(c)} = U_n^{(c)}$ and $Z_{-\infty}^{(c)} = \lim_{n\to\infty} Z_{-n}^{(c)}$ for $c = 0, 1, \cdots, r$. Then $Z_{-\infty}^{(c)} = \lim_{n\to\infty} U_n^{(c)} = \rho_c + \theta^2$ (a.s.) for $c = 0, 1, \cdots, r$. (Recall that $\rho_0 = 0$.) Then, $\{Z_{-\infty}^{(c)}, \cdots, Z_{-r}^{(c)}\}$ is a martingale. Therefore, from Lemma 2 with $m_1 = n_0 - 1$ and

$m_2 = \infty$,

$$E\{U_M^{(c)}\} = E\{U_{n_0-1}^{(c)}\} = \rho_c + \theta^2 \tag{11}$$

for $c = 0, 1, \cdots, r$. In particular, $E\{U_M^{(1)}\} = \rho_1 + \theta^2$ and $E\{U_M^{(0)}\} = \theta^2$. From (6) and (8), for every $t > 0$, $N_t \leq M + 1$, so that, as a consequence of Theorem 1(ii), $\lim_{t\to\infty} M = \infty$ (a.s.). Also, $\lim_{t\to\infty} U_M^{(c)} = \rho_c + \theta^2$ (a.s.) for $c = 0, 1, \cdots, r$. Now $\alpha_n(c) = O(n^{-1})$, so that $\lim_{t\to\infty} \alpha_M(c) U_M^{(c)} = 0$ (a.s.) for $c = 0, 1, \cdots, r$. Furthermore, by Lemma 1, $E\{\sup_n \alpha_n(c) \mid U_n^{(c)} \mid\} < \infty$ for $c = 0, 1, \cdots, r$. We then use the Lebesgue dominated convergence theorem to obtain

$$\lim_{t\to\infty} \textstyle\sum_{c=0}^{r} E\{\alpha_M(c) U_M^{(c)}\} = 0. \tag{12}$$

Then, from (10), (11) and (12) we conclude that $\lim_{t\to\infty} E\{s_{wM}^2\} = \sigma^2$. Finally, since $\lim_{t\to\infty} M^{-\gamma} = 0$ (a.s.), the Lebesgue dominated convergence theorem again implies that $\lim_{t\to\infty} E\{M^{-\gamma}\} = 0$. Thus $\lim_{t\to\infty} E\{s_{wM}^{2*}\} = \sigma^2$.

(c) We prove that

$$\lim_{t\to\infty} E\{a_M^2 s_{wM}^{2*}\} = a^2\sigma^2. \tag{13}$$

First, note that $\lim_{t\to\infty} a_M^2 s_{wM}^{2*} = a^2\sigma^2$ (a.s.). Now, let $A = \inf_n a_n^2$ and $B = \sup_n a_n^2$. Then, for every $t > 0$, $A s_{wM}^{2*} \leq a_M^2 s_{wM}^{2*} \leq B s_{wM}^{2*}$. Thus

$$0 \leq a^2\sigma^2 - A\sigma^2 = E\{\lim_{t\to\infty}(a_M^2 s_{wM}^{2*} - A s_{wM}^{2*})\}$$

and, by Fatou's lemma,

$$\begin{aligned} 0 \leq a^2\sigma^2 - A\sigma^2 &\leq \liminf_{t\to\infty} E\{a_M^2 s_{wM}^{2*} - A s_{wM}^{2*}\} \\ &= \liminf_{t\to\infty} E\{a_M^2 s_{wM}^{2*}\} - A\sigma^2. \end{aligned} \tag{14}$$

Also,

$$0 \leq B\sigma^2 - a^2\sigma^2 = E\{\lim_{t\to\infty}(B s_{wM}^{2*} - a_M^2 s_{wM}^{2*})\}$$

and, by invoking Fatou's lemma once more,

$$\begin{aligned} 0 \leq B\sigma^2 - a^2\sigma^2 &\leq \liminf_{t\to\infty} E\{B s_{wM}^{2*} - a_M^2 s_{wM}^{2*}\} \\ &= B\sigma^2 - \limsup_{t\to\infty} E\{a_M^2 s_{wM}^{2*}\}. \end{aligned} \tag{15}$$

Then (13) follows from (14) and (15).

(d) We conclude from (9) and (13) that $\limsup_{t\to\infty} t^{-1} E\{N_t\} \leq 1$. However, Fatou's lemma implies that $\liminf_{t\to\infty} t^{-1} E\{N_t\} \geq 1$. This completes the proof of Theorem 2.

4. Examples.

EXAMPLE 1. *The population variance.* Let $\mu = E\{X_1\}$ and $\mu_j = E\{(X_1 - \mu)^j\}$ for $j = 2, 3, \cdots$ (when existent). Assume $0 < \mu_2^2 < \mu_4 < \infty$. Let $f(x_1, x_2) = (x_1 - x_2)^2/2$ so that $\theta = E\{(X_1 - X_2)^2/2\} = \mu_2$. The corresponding U-statistic is the sample variance $U_n = (n-1)^{-1}s_2$ where $s_j = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^j$ for $j = 2, 3, \cdots$ and $\bar{X}_n$ is the sample mean. Also $f_1(x_1) = E\{(x_1 - X_2)^2/2\} = (x_1 - \mu)^2/2 + \mu_2/2$ and $\rho_1 = \delta_1 = (\mu_4 - \mu_2^2)/4$ so that $\sigma^2 = r^2\rho_1 = \mu_4 - \mu_2^2$.

From the definition of s_{wn}^2, after some manipulation, we obtain

$$s_{wn}^{2*} = n^3(n-1)^{-3}[s_4 - s_2^2] + n^{-\gamma}. \tag{16}$$

The factor $n^3(n-1)^3$ in (16) may be omitted without affecting the properties of s_{wn}^{2*} to any appreciable extent.

For the sake of simplicity let $a_k = a$ for $k = 2, 3, \cdots$ although any positive sequence $\{a_k\}$ such that $\lim_{k\to\infty} a_k = a$ would do since we are investigating asymptotic behavior. (There is some justification for taking the a_k to be percentage points of the t-distribution.) Define

$$N(d) = \text{smallest integer } k \geq 2 \text{ such that } s_{wk}^{2*} \leq kd^2a^{-2}$$

where s_{wk}^{2*} is given by (16). Then $I_N = [U_N - d, U_N + d]$ is a sequential confidence interval for $\theta = \mu_2$ having width equal to $2d$ and coverage probability approximately equal to α, for small values of d. Note, also, that in addition to being efficient in the sense of Theorem 2, the sequential procedure is invariant under a location shift.

EXAMPLE 2. *The probability of concordance.* Suppose that $X_1 = (X_1^{(1)}, X_1^{(2)}), \cdots, X_n = (X_n^{(1)}, X_n^{(2)})$ is a bivariate random sample of a random variable $X = (X^{(1)}, X^{(2)})$ having continuous marginal distribution functions. Let $s(u) = -1, 0$ and $+1$ when $u < 0$, $u = 0$ and $u > 0$, respectively, and let $f(x_1, x_2) = s(x_1^{(1)} - x_2^{(1)})\cdot s(x_1^{(2)} - x_2^{(2)})$. The corresponding U-statistic is

$$U_n = n^{-1}(n-1)^{-1}\sum_{\alpha_1\neq\alpha_2} s(X_{\alpha_1}^{(1)} - X_{\alpha_2}^{(1)})\cdot s(X_{\alpha_1}^{(2)} - X_{\alpha_2}^{(2)}) \tag{17}$$

and is referred to as the difference sign covariance of the sample. See Hoeffding (1948). Two points x_1 and x_2 are "concordant" if $s(x_1^{(1)} - x_2^{(1)})s(x_1^{(2)} - x_2^{(2)}) = +1$ and are "discordant" if $s(x_1^{(1)} - x_2^{(1)})s(x_1^{(2)} - x_2^{(2)}) = -1$. Let

$$\pi_1 = P\{X_1 \text{ and } X_2 \text{ are concordant}\} = P\{(X_1^{(1)} - X_2^{(1)})\cdot(X_1^{(2)} - X_2^{(2)}) > 0\}$$

and π_2 equal the probability that X_1 and X_2 are either both concordant, or, both discordant, with X_3. Then the expectation of the U-statistic is $\theta = 2\pi_1 - 1$ and, after some calculation, $\rho_1 = 2\pi_2 - 1 - \theta^2$. Assume that $\rho_1 > 0$. Now, let C_n equal the number of concordant pairs among $\{X_1, \cdots, X_n\}$. Then (17) becomes

$$U_n = 4n^{-1}(n-1)^{-1}C_n - 1 = 2\bar{C}_n - 1$$

where $\bar{C}_n = C_n/\binom{n}{2}$.

To determine s_{wn}^2, for $i = 1, 2, \cdots, n$, let T_{in} equal the number of points among $\{X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n\}$ concordant with X_i. Then $W_{in} = 4(n-1)^{-1}T_{in} - 2$ and, after some manipulation,

$$s_{wn}^2 = 16(n-1)^{-3}[\textstyle\sum_{i=1}^{n} T_{in}^2 - 4n^{-1}C_n^2]. \tag{18}$$

Then (18) may be used to generate a sequential fixed-width confidence interval for π_1, the probability of concordance.

Acknowledgments. I am indebted to Professor W. J. Hall for originating (circa 1964) the problem of adapting the Chow and Robbins (1965) type sequential procedure to U-statistics. Moreover, the referee made several significant recommendations including the idea of "delaying" the sequential procedure.

REFERENCES

CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

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

SEN, P. K. (1977). Some invariance principles relating to jackknifing and their role in sequential analysis. *Ann. Statist.* **5** 316–329.

SEN, P. K. and GHOSH, M. (1981). Sequential point estimation of estimable parameters based on U-statistics. *Sankhya Ser. A* **43** 331–344.

SIMONS, G. (1968). On the cost of not knowing the variance when making a fixed-width confidence interval for the mean. *Ann. Math. Statist.* **39** 1946–1952.

SPROULE, R. N. (1969). A sequential fixed-width confidence interval for the mean of a U-statistic. Institute of Statistics Mimeo Series No. 636, University of North Carolina, Chapel Hill.

SPROULE, R. N. (1974). Asymptotic properties of U-statistics. *Trans. Amer. Math. Soc.* **199** 55–64.

MALASPINA COLLEGE
NANAIMO, BC V9R 5S5
CANADA