ROBUST SEQUENTIAL CONFIDENCE INTERVALS FOR THE BEHRENS-FISHER PROBLEM

by

Malay Ghosh

Department of Biostatistics
University of North Carolina, Chapel Hill

Institute of Statistics Mimeo Series No. 685

May, 1970

ROBUST SEQUENTIAL CONFIDENCE INTERVALS FOR THE BEHRENS-FISHER PROBLEM*

by Malay Ghosh

University of North Carolina at Chapel Hill

0. Summary. The problem of providing a bounded length (sequential) confidence interval for the median of a symmetric (but otherwise unknown) distribution based on a general class of one-sample rank-order statistics was investigated in [6]. The purpose of the present note is to indicate how the techniques developed there can be extended to the two-sample problem. It has been shown that in particular for the Behrens-Fisher situation (see e.g., [3] or [4]), when the proposed procedure is based on the "normal-scores" statistic, under very general conditions on the unknown distribution function (d.f.), it is asymptotically at least as efficient as an analogous procedure suggested in [5].

1. Introduction. Consider two independent sequences of random variables (rv's) $\{X_1, X_2, \ldots\}$ and $\{Y_1, Y_2, \ldots\}$, the X's independent and identically distributed (iid) with d.f. $F(x-\rho)$ and Y's iid with d.f. $G(x-\rho-\Delta)$, ρ, Δ unknown. Our goal is to find a confidence interval I of width 2d for Δ such that $P\{\Delta \varepsilon I\} \geq 1-\alpha$ (the desired confidence coefficient), where, $0<d<\infty$, $0<\alpha<1$ are preassigned constants. F. and G. being unknown, no fixed sample size procedure for the above problem seems feasible.

A sequential procedure for the above problem based on the mean difference $\bar{X}-\bar{Y}$ was proposed in [5]. This, obviously, is vulnerable to gross errors or outlying observations, and may be quite inefficient for heavy-tailed distributions. In the

* Work supported by National Institutes of Health GM-12868.

present note, an alternative procedure is proposed based on rank-order statistics, which is not, or at least, much less vulnerable to the above criticism.

To motivate the procedure, we use the same logic as in [5]. For a fixed sample $(X_1, X_2, \ldots, X_m)$ of size $m \equiv m(p)$ and $(Y_1, Y_2, \ldots, Y_n)$ of size $n \equiv n(p)$ $(p=m+n)$, from the two populations with d.f.'s $F(x-\rho)$ and $G(x-\eta)$ respectively ($\Delta = \eta - \rho$; F and G assumed to be symmetric about zero), we construct estimators of η and ρ based on one-sample rank-order statistics and take their difference as estimator of Δ. With this end in view, define the following one-sample rank-order statistics:

$$h_1(\underset{\sim}{X}_m) = \sum_{\alpha=1}^{m} u(X_\alpha) J_m(R_{m\alpha}/(m+1)); \tag{1.1}$$

$$h_2(\underset{\sim}{Y}_n) = \sum_{\beta=1}^{n} u(Y_\beta) J_n(R'_{n\beta}/(n+1)), \tag{1.2}$$

where, $u(t) = 0$ or 1 according as $t<0$ or not; $R_{m\alpha} = \sum_{i=1}^{m} u(|X_\alpha| - |X_i|)$, $R'_{n\beta} = \sum_{j=1}^{n} u(|Y_\beta| - |Y_j|)$, and $J_m(u)$, $J_n(v)$ $(0<u, v<1)$ are generated by a score function $J(u)$ $(0 \leq u < 1)$ in either of the following two ways:

(a) $J_m(u) = i/(m+1)$, $J_n(v) = j/(n+1)$, $(i-1)/m < u \leq i/m$, $(j-1)/n < v \leq j/n$, $1 \leq i \leq m$, $1 \leq j \leq n$.

(b) $J_m(u) = EJ(U_{mi})$, $J_n(v) = EJ(U_{nj})$, $(i-1)/m < u \leq i/m$, $(j-1)/n < v \leq j/n$, $1 \leq i \leq m$, $1 \leq j \leq n$,

where, $U_{m1} \leq \cdots \leq U_{mm}$ $(U_{n1} \leq \cdots \leq U_{nn})$ are m(n) ordered random variables in a sample of size m(n) from a rectangular (0,1) distribution. Further, assume that $J(u) = \Psi^{-1}((1+u)/2)$, $0 \leq u < 1$, $\Psi(x)$ being a d.f. defined on $(-\infty, \infty)$ satisfying

(1.3) (a) $\Psi(x) + \Psi(-x) = 1$, for all real x;

(1.4) (b) $-\log[1-\Psi(x)]$ is convex for all $x \geq x_0$ (x_0 real, ≥ 0). Also, let

(1.5) $F, G \in \mathcal{F}_0(J)$, the class of all absolutely continuous d.f.'s $\Pi(x)$ symmetric about zero for which both $\Pi'(x)$ and $\Pi''(x)$ are bounded for almost all $x(a{\cdot}a{\cdot}x)$ and

(1.6) $$\lim_{x\to\infty} \Pi'(x)J(\Pi(x)-\Pi(-x)) \text{ is bounded.}$$

Introduce the following notations:

(1.7) $\bar{E}_m = m^{-1}\sum_{i=1}^{m} J_m(i/(m+1))$; $A_m^2 = m^{-1}\sum_{i=1}^{m} J_m^2(i/(m+1))$, (similarly $\bar{E}_n, A_n^2$);

(1.8) $$\mu = \int_0^1 J(u)du,\ A^2 = \int_0^1 J^2(u)du.$$

It follows from (1.1) that $h_1(\underset{\sim}{X}_m - a\,\underset{\sim}{1}_m)$ ($\underset{\sim}{1}_m$ - an m-component row vector with all elements 1) is $\downarrow$ in a. Also, under $\rho=0$, it has a distribution independent of F, symmetric about $\frac{1}{2}\bar{E}_m$. Defining now,

(1.9) $$\hat{\rho}_{m1}^{(1)} = \sup\{a:\ h_1(\underset{\sim}{X}_m - a\underset{\sim}{1}_m) > \tfrac{1}{2}\bar{E}_m\},$$

(1.10) $$\hat{\rho}_{m2}^{(1)} = \inf\{a:\ h_1(\underset{\sim}{X}_m - a\underset{\sim}{1}_m) < \tfrac{1}{2}\bar{E}_m\},$$

we propose $\hat{\rho}_m^{(1)} = \frac{1}{2}(\hat{\rho}_{m1}^{(1)} + \hat{\rho}_{m2}^{(1)})$ as a point estimator of ρ. Similarly let $\hat{\eta}_n^{(2)}$ be a point estimator of η based on h_2. Propose $\hat{\Delta}_p = \hat{\eta}_n^{(2)} - \hat{\rho}_m^{(1)}$ as a point estimator of Δ. Under (1.3)-(1.6), and the assumption that

(1.11) $$0<\lambda<1, \text{ where } \lambda = \lim_{p\to\infty} m(p)/p,$$

it is well-known (see e.g., [4]) that $\sqrt{p}\,(\hat{\Delta}_p - \Delta)$ is asymptotically (as $p\to\infty$) normal $(0,\sigma_0^2)$, $\sigma_0^2 = \sigma_1^2/\lambda + \sigma_2^2/(1-\lambda)$, $\sigma_1^2 = A^2/B^2(F)$, $\sigma_2^2 = A^2/B^2(G)$, where,

(1.12) $$B(\Pi) = \int_0^\infty \frac{d}{dx} J(2\Pi(x)-1)d\Pi(x).$$

Thus, if $I = [\hat{\Delta}_p - d, \hat{\Delta}_p + d]$, $\lim_{p\to\infty} P\{\Delta \varepsilon I\} = 2\Phi(a)-1$, $a = \lim_{p\to\infty} \sqrt{p}d/\sigma_0$, Φ being the d.f. of a normal (0,1) distribution. Equating $2\Phi(a)-1$ to $1-\alpha$, we find $a = \tau_{\alpha/2}$, $\Phi(\tau_{\alpha/2}) = 1-\tfrac{1}{2}\alpha$. Next we find strongly consistent estimators of σ_1^2 and σ_2^2.

It follows from the remarks after (1.8) that there exist (known) constants $h_{11\alpha}^{(m)}$ and $h_{12\alpha}^{(m)} = \bar{E}_m - h_{11\alpha}^{(m)}$, such that

$$P\{h_{11\alpha}^{(m)} \le h_1(\underset{\sim}{X}_m) \le h_{12\alpha}^{(m)}\} = 1-\alpha_m \to 1-\alpha \text{ as } m\to\infty. \tag{1.13}$$

Define

$$\hat{\rho}_{L,m}^{(1)} = \sup\{a:\ h_1(\underset{\sim}{X}_m - a\underset{\sim}{1}_m) > h_{11\alpha}^{(m)}\}; \tag{1.14}$$

$$\hat{\rho}_{U,m}^{(1)} = \inf\{a:\ h_1(\underset{\sim}{X}_m - a\underset{\sim}{1}_m) < h_{12\alpha}^{(m)}\}. \tag{1.15}$$

Then, $P\{\hat{\rho}_{L,m}^{(1)} < \rho < \hat{\rho}_{U,m}^{(1)}\}\ 1-\alpha_m$, and it follows from lemma 5.2 of [6] that for every $\delta > 0$, there exists an integer m_δ such that for $m \ge m_\delta$ with probability $\ge 1-O(m^{-1-\delta})$,

$$\left|\{\sqrt{m}\ (\hat{\rho}_{U,m}^{(1)} - \hat{\rho}_{L,m}^{(1)})/\tau_{\alpha/2}\} - \sigma_1\right| = O(m^{-1/4}(\log m)^3). \tag{1.16}$$

Thus, if $\hat{\sigma}_{1m}^2 = m(\hat{\rho}_{U,m} - \hat{\rho}_{L,m})^2/\tau_{\alpha/2}^2$, $|\hat{\sigma}_{1m}^2 - \sigma_1^2| = O(m^{-1/4}(\log m)^3)$ with a probability statement as above. Define, now $\hat{\eta}_{L,n}^{(2)}$ and $\hat{\eta}_{U,n}^{(2)}$ similarly as in (1.14) and (1.15), and let $\hat{\sigma}_{2n}^2 = n(\hat{\eta}_{U,n}^{(2)} - \hat{\eta}_{L,n}^{(2)})^2/\tau_{\alpha/2}^2$. Then, for every $\delta > 0$, there exists an integer n_δ such that for $n \ge n_\delta$, $|\hat{\sigma}_{2n}^2 - \sigma_2^2| = O(n^{-1/4}(\log n)^3)$ with probability $\ge 1-O(n^{-1-\delta})$. Following Robbins, Simons and Starr [5] we propose a sampling scheme which says that if at any stage, we have taken m observations on X and n observations on Y with $p = m+n \ge 2p_0$, the next observation is taken on X or Y according as $m/n \le \hat{\sigma}_{1m}/\hat{\sigma}_{2n}$ or $m/n > \hat{\sigma}_{1m}/\hat{\sigma}_{2n}$. The procedure generates an infinite sequence of observations, and, from the definition of $\hat{\sigma}_{1m}$ and $\hat{\sigma}_{2n}$, it is seen to depend on α, but not on d. Further as in [5], one can easily show that $m/n \to \sigma_1/\sigma_2$ as $p \to \infty$.

Also, in a similar fashion as [5], we define the following three stopping rules:

Stop with the first $N \geq 2p_0$ such that if r observations on X and s observations on Y have been taken with $r+s \geq 2p_0$, then

(1.17)
$$R_1: \quad N \geq b(\hat{\sigma}_{1r}+\hat{\sigma}_{2s})^2, \; b = (\tau_{\alpha/2}/d)^2, \text{ i.e.,}$$
$$N \geq \{r(\hat{\rho}^{(1)}_{U,r}-\hat{\rho}_{L,r})^2 + s(\hat{\eta}^{(2)}_{U,s}-\hat{\eta}^{(2)}_{L,s})^2\}/d^2$$

(1.18)
$$R_2: \quad \hat{\sigma}^2_{1r}/r+\hat{\sigma}^2_{2s}/s \leq b^{-1}, \text{ i.e.,}$$
$$(\hat{\rho}^{(1)}_{U,r}-\hat{\rho}^{(1)}_{L,r})^2 + (\hat{\eta}_{U,s}-\hat{\eta}^{(2)}_{L,s})^2 \leq d^2$$

(1.19)
$$R_3: \quad r \geq b\hat{\sigma}_{1r}(\hat{\sigma}_{1r}+\hat{\sigma}_{2s}) \text{ and } s \geq b\hat{\sigma}_{2s}(\hat{\sigma}_{1r}+\hat{\sigma}_{2s}), \text{ i.e.,}$$
$$r \geq (\hat{\rho}^{(1)}_{U,r}-\hat{\rho}^{(1)}_{L,r})(\hat{\rho}^{(1)}_{U,r}-\hat{\rho}^{(1)}_{L,r}+\hat{\eta}_{U,s}-\hat{\eta}^{(2)}_{L,s})$$
$$\text{and } s \geq (\hat{\eta}^{(2)}_{U,s}-\hat{\eta}^{(2)}_{L,s})(\hat{\rho}^{(1)}_{U,r}-\hat{\rho}_{L,r}+\hat{\eta}^{(2)}_{U,s}-\hat{\eta}^{(2)}_{L,s})$$

It follows that if N_k are the stopping variables corresponding to R_k(R=1,2,3), then, $N_1 \leq N_2 \leq N_3$. The confidence interval proposed is $I_{N(d)} = [\hat{\Delta}_{N(d)}-d, \hat{\Delta}_{N(d)}+d]$.

2. The Main Results. The basic theorem of the paper is as follows:

THEOREM 2.1. *Under* (1.3)-(1.6), *and following the sampling scheme and any of the stopping rules in the earlier section, we have,* (2.1) $N(=N(d))$ <u>is a non-increasing function of</u> d, $N(d)$ <u>is finite a.s. and</u> $EN(d)<\infty$ <u>for all</u> $d>0$; $\lim_{d\to 0} N(d) = \infty$ <u>a.s. and</u> $\lim_{d\to 0} EN(d) = \infty$;

(2.2) $$\lim_{d\to 0} N(d)/(b(\sigma_1+\sigma_2)^2) = 1 \text{ a.s.};$$

(2.3) $$\lim_{d\to 0} P\{\Delta\varepsilon I_{N(d)}\} = 1-\alpha;$$

(2.4) $$\lim_{d\to 0} EN(d)/(b(\sigma_1+\sigma_2)^2) = 1.$$

Proof: Since (1.11) holds, from the remarks following (1.16), one gets that for every $\delta>0$, there exists an integer p_δ such that for $p\geq p_\delta$, $|\hat{\sigma}^2_{1m(p)}-\sigma_1^2| = O(p^{-1/4}(\log p)^3)$ with probability $\geq 1-O(p^{-1-\delta})$. A similar probability statement holds for $|\hat{\sigma}^2_{2n(p)}-\sigma_2^2|$. Because of the above, using the same technique as in lemma 5.5 of [6], under rule R_3, we get $EN(d) <\infty$ for all $d>0$. Since $N_i(d)\leq N_3(d)$ $(i=1,2,3)$, the same is true for R_1 and R_2 also. (2.1) now follows from the definition of the stopping rules and the monotone convergence theorem.

The proof of (2.2) follows in the same manner as the corresponding result in [5]. (2.4) follows along the lines of [6]. The details are omitted for brevity.

Since, $\sqrt{p}\ (\hat{\Delta}_p-\Delta)$ is asymptotically normal $(0, \sigma_0^2)$, and (2.2) holds, to prove (2.3), it is sufficient to show that the sequence $\{\hat{\Delta}_p\}$ is uniformly continuous with respect to $p^{-1/2}$ (see e.g., [1]). Use the inequality

$$P\{\sup_{|p'-p|<\delta p}|\sqrt{p}\,(\hat{\Delta}_{p'}-\hat{\Delta}_{p})|>\xi\},\ \xi \text{ some positive constant}$$

$$(2.5)\qquad \leq P\{\sup_{|p'-p|<\delta p}|\sqrt{m(p)}\,(\hat{\rho}_{m(p')}-\hat{\rho}_{m(p)})| > \frac{\xi\sqrt{m(p)}}{2\sqrt{p}}\}$$

$$+ P\{\sup_{|p'-p|<\delta p}|\sqrt{n(p)}\,(\hat{\eta}_{n(p')}-\hat{\eta}_{n(p)})| > \frac{\xi\sqrt{n(p)}}{2\sqrt{p}}\}.$$

The proof is now completed by using (1.11) and the same technique as in lemma 5.3 of [6].

The theorem shows that our procedure is "asymptotically consistent" and "asymptotically efficient" in the sense of [2].

We compare the asymptotic relative efficiency (ARE) of the proposed procedure to the one proposed in [5]. For a definition of ARE, see [6]. If R and P stand respectively for the procedures in the present paper and in [5], let $e_{R,P}$ = ARE of R with respect to P. Then (see (2.4) and (9) of [5]) under the assumption that the variances $\sigma_1'^2$ and $\sigma_2'^2$ of the two populations with d.f. $F(x-\rho)$ and $G(x-\eta)$ respectively, are non-zero and finite, we have,

$$(2.6)\qquad e_{R,p} = \lim_{d\to 0} b(\sigma_1'+\sigma_2')^2/b(\sigma_1+\sigma_2)^2 = (\sigma_1'+\sigma_2')^2/(\sigma_1+\sigma_2)^2,$$

which is independent of d. In particular, for the Behrens-Fisher situation $G(x) = F(cx)$, $c>0$, i.e. The two d.f. F and G differ only in scale, $\sigma_2' = \sigma_1'/c$, $B(G) = cB(F)$; so $\sigma_2^2 = \sigma_1^2/c^2$. Hence, from (2.6), $e_{R,P} = \sigma_1'^2 B^2(F)/A^2$ which is the ARE of a general rank-order test with respect to student's t-test. The conclusions made in [6] can be repeated in this case, and in particular for $J(u) = \Phi^{-1}((1+u)/2)$, $e_{R,P} \geq 1$, uniformly in F with a finite second moment, equality being attained if and only if F is normal $(0,\sigma_1'^2)$ d.f.