October, 1965

AN EXTENSION OF A RESULT OF LEHMANN ON THE ASYMPTOTIC EFFICIENCY OF SELECTION PROCEDURES BASED ON RANKS

G. G. Woodworth*

Technical Report No. 66

University of Minnesota
Minneapolis, Minnesota

*This research was supported in part by the National Science Foundation under Grant No. GP-3813.

There are k populations $\pi_1, \pi_2, \ldots, \pi_k$ differing only in location. In fact the distribution of an observation from π_j is $F(x-\theta_j)$, where F is assumed to have density f.

Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ be the ordered values of $\theta_1, \ldots, \theta_k$ and let $\pi_{[j]}$ correspond to $\theta_{[j]}$. For fixed $\frac{1}{\binom{k}{t}} < P^* < 1$ and $\delta^* > 0$, the goal is to select the t best populations $\pi_{[k-t+1]}, \ldots, \pi_{[k]}$ subject to the following restriction on the P{CS} (probability of correct selection):

$$(1) \qquad P\{CS\} \geq P^* \quad \text{when} \quad \theta_{[k-t+1]} - \theta_{[k-t]} \geq \delta^*.$$

Let $\{X_{ij};\ i = 1,\ldots,k;\ j = 1,\ldots,n\}$ be k samples of size n, one sample from each population, where X_{ij} is the $j^{\underline{th}}$ observation from π_i. Let R_{ij} be the rank of X_{ij} in the combined sample.

Let $\{J_{n,j};\ j = 1,\ldots,n,\ n = 1,\ldots\} = \mathcal{J}$ be a sequence of real numbers with $J_{n,i} \leq J_{n,j}$ $i \leq j$, $n = 1,2,\ldots,$ and define functions $\{J_n;\ n = 1,2,\ldots\}$ as follows:

$$J_n(u) = J_{n,i} \qquad \frac{i-1}{n} < u \leq \frac{i}{n}, \quad i = 1,\ldots,n.$$

We assume that there exists a function J, square integrable on (0,1) such that

$$(2) \qquad \int_0^1 (J_n(u) - J(u))^2\, du \to 0 \qquad \text{as} \quad n \to \infty.$$

We define the statistics T_{ni} $i = 1,\ldots,k$ as follows:

$$T_{ni} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} J_{nR_{ij}} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} J_n\left(\frac{R_{ij}}{nk+1}\right) .$$

Lehmann [4] suggests the following procedure, which we shall call $R_{\mathcal{J}}$: take a sample of fixed size n from each population and select the populations having the t largest values of T_{ni} $i = 1,\ldots,k$, where n is the smallest integer such that (1) is satisfied.

Let $T_{n[i]}$ correspond to $\pi_{[i]}$ $i = 1,\ldots,k$, then under procedure $R_{\mathcal{J}}$,

(3) $$P\{CS\} = P\{\max_{1\le i\le(k-t)} T_{n[i]} \le \min_{(k-t)<j\le k} T_{n[j]}$$

Let $\gamma(x_{11},\ldots,x_{kn})$ be the indicator of the set on the right side of (3). γ is a non-increasing function of each $x_{[i]j}$ for $i \le k-t$ and a non-decreasing function of each $x_{[i]j}$ for $i > k-t$ because $T_{n[i]}$ is non-decreasing* in $x_{[i]j}\, j = 1,\ldots,n$ while $T_{n[i']}$ is non-increasing* in $x_{[i]j}$ $j = 1,\ldots,n$, $i \neq i'$. Since $F(x-\theta)$ is stochastically increasing, it follows from a slight generalization of Rizvi [5] Theorem 1, or Lehmann [3] Chapter 3 Lemma 2(i), that the probability of a correct selection ($P\{CS\}$) attains its minimum subject to $\theta_{[k-t+1]} - \theta_{[k-t]} \ge \delta^*$ when

(4) $$\theta_{[1]} = \cdots = \theta_{[k-t]} = \theta_{[k-t+1]} - \delta^* = \cdots = \theta_{[k]} - \delta^* .$$

It is clear that this minimum depends only on δ^*, n and $\mathcal{J}$, hence, we shall denote it by $P\{CS|\delta^*,n,\mathcal{J}\}$.

For fixed P^* and δ^* let $n(P^*,\delta^*,\mathcal{J})$ be the smallest sample size for which $P\{CS|\delta^*,n,\mathcal{J}\} \ge P^*$. Now consider the class C of all procedures $R_{\mathcal{J}}$ such that $\mathcal{J}$ satisfies (2); a procedure $R_{\mathcal{J}_0}$ is <u>asymptotically most efficient in C</u> if $\liminf_{\delta^*\to 0} \frac{n(P^*,\delta^*,\mathcal{J})}{n(P^*,\delta^*,\mathcal{J}_0)} \ge 1$ for all $\frac{1}{\binom{k}{t}} < P^* < 1$ and all $R_{\mathcal{J}}$ in C.

The following theorem extends Lemma 2 of Lehmann [4] by weakening the restrictions on J.

<u>Theorem:</u> If $J_0(u) = -\frac{f'}{f}(F^{-1}(u))$ is square integrable on (0,1), then for fixed P^* $$n(P^*,\delta^*,\mathcal{J}) \sim \left(\frac{v^*}{\delta^*}\right)^2 \frac{\int_0^1 J^2(u)du - (\int_0^1 J(u)du)^2}{(\int_0^1 J(u)J_0(u)du)^2} ,$$

where v^* is the solution of

*Observe that $J_n(u)$ is non-decreasing in u for each n.

(5) $$P^* = t \int \Phi^{k-t}(x+v^*)\ (1-\Phi(x))\ d\Phi(x),$$

and Φ is the normal c.d.f.

<u>Proof</u>: Assume, without loss of generality, that

(6) $$\theta_{[1]} = \dots = \theta_{[k-t]} = -\delta^* \frac{t}{k}, \quad \theta_{[k-t+1]} = \dots = \theta_{[k]} = \delta^* \left(1-\frac{t}{k}\right)$$

Let $L_n = \sum_{i=1}^{k} \sum_{j=1}^{n} \log\left(\frac{f(x_{[i]j} - \theta_{[i]})}{f(x_{[i]j})}\right)$, where $x_{[i]j}$ is the j^{th} observation from the population with parameter $\theta_{[i]}$, let Q_n denote the distribution of $\{x_{ij};\ i = 1,\dots,k;\ j = 1,\dots,n\}$ when (6) holds and let P_n denote the distribution when $\theta_{[1]} = \theta_{[k]} = 0$, and let

$$W_n^* = \sum_{i=1}^{k} \sum_{j=1}^{n} \theta_{[i]} \left(-\frac{f'(x_{[i]j})}{f(x_{[i]j})}\right).$$

If we set $\delta^* = a/\sqrt{n}$ for some fixed $a > 0$, then by Hajek [2] (5.21)

(7) $$P_n \lim_{n\to\infty} (L_n - E_{P_n} L_n - W_n^*) = 0,$$

where $P_n \lim_{n\to\infty} (Z_n)$ is the limit of Z_n in P_n-probability. Consequently, if $\mathcal{L}(Y_n, W_n^*)$ is asymptotically normal with correlation ρ under P_n then so is $\mathcal{L}(Y_n, L_n)$.

Let $\alpha_1,\dots,\alpha_k$ be arbitrary real numbers and let $T_n = \sum_{i=1}^{k} \alpha_i T_{n[i]}$. Let $T_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{k} (\alpha_i - \bar{\alpha}) \sum_{j=1}^{n} J_n(F(X_{[i]j}))$, where $\bar{\alpha} = \frac{1}{k} \sum_{i=1}^{k} \alpha_i$. It follows from Hajek [1] Theorem 3.1 that

(8) $$P_n \lim_{n\to\infty} \frac{1}{\sigma_{P_n}(T_n^*)} (T_n - E_{P_n} T_n - T_n^*) = 0$$

and from (8) and Hajek [2] Lemma 4.2 (1) that

(9) $$Q_n \lim_{n\to\infty} \frac{1}{\sigma_{P_n}(T_n^*)} (T_n - E_{P_n} T_n - T_n^*) = 0.$$

From Hajek [1] Theorem 4.1 and (8) we conclude that, under P_n, T_n is asymptotically normal with mean $\sqrt{n} \sum_{i=1}^{k} \alpha_i \int_0^1 J(u)du$ and variance

$$k \sum_{i=1}^{k} (\alpha_i - \bar{\alpha}) \left(\int_0^1 J^2(u)du - \left(\int_0^1 J(u)du\right)^2\right)$$

and from Hajek [1] Theorem 4.1, (9) and the remark following (7) that, under Q_n, T_n is asymptotically normal with the same variance as above but with mean $\sqrt{n} \sum_{i=1}^{k} \alpha_i \int J(u)du$

$$+ a \sum_{i=1}^{k} (\alpha_i - \bar{\alpha}) c_i \int_0^1 J(u)\left(- \frac{f'}{f} (F^{-1}(u))\right) du \ , \quad \text{where} \quad c_i = \begin{cases} -\frac{t}{k} & i = 1,\ldots,k-t \\ 1-\frac{t}{k} & i = k-t,\ldots,k. \end{cases}$$

It follows that the vector $(T_{n[i]}, T_{n[2]},\ldots,T_{n[k]})$ is asymptotically normal with mean vector given by:

$$T_{n[i]} = \sqrt{n} \int_0^1 J(u)du + \begin{cases} 0 & \text{under } P_n \\ a\, C_i \int_0^1 J(u)J_0(u)du, & \text{under } Q_n, \end{cases}$$

where $J_0(u) = - \frac{f'}{f} (F^{-1}(u))$, and covariance matrix given by:

$$\sigma(T_{n[i]},T_{n[j]}) = \left(\delta_{ij} - \frac{1}{k}\right)\left(\int_0^1 J^2(u)du - \left(\int_0^1 J(u)du\right)^2\right)$$

under P_n and under Q_n.

Let $\xi_i = ((T_{n[i]} - T_{n[k]}) - d_i)/A$, $i = 1,\ldots,n-1$, where

$$d_i = \begin{cases} -a \int JJ_0 & i = 1,2,\ldots,k-t \\ 0 & i = k-t+1,\ldots,k \end{cases} \quad \text{and} \quad A^2 = \int_0^1 J^2(u)du - \left(\int_0^1 J(u)du\right)^2 \ ,$$

then $E_{Q_n}\xi_i \to 0$, $\sigma^2_{Q_n}(\xi_i) \to 2$, $\sigma_{Q_n}(\xi_i,\xi_j) \to \frac{1}{2}$ and

$$P\{CS\} = P\{\max_{i \le k-t} \xi_i - \frac{a}{A} \int_0^1 J(u)J_0(u)du \le \min_{k-t<k\le k-1} (\xi_j,0)\}. \tag{10}$$

If $\xi_1,\ldots,\xi_n$ are independent $N(0,1)$ random variables, then the right side of (10) converges to

$$P\{\max_{i\le k-t}(\xi_i-\xi_n) - \frac{a}{A}\int_0^1 J(u)J_0(u)du \le \min_{k-t<j\le k-1}((\xi_j-\xi_n),0)\}$$

$$= P\{\max_{i\le k-t}\xi_i - \frac{a}{A}\int_0^1 J(u)J_0(u)du \le \min_{k-t<j\le k}\xi_j\}$$

$$= t\int \Phi^{k-t}(x+\frac{a}{A}\int JJ_0)\,(1-\Phi(x))^{t-1}\,d\Phi(x).$$

Thus if $P\{CS\} = P^*$, then $\frac{a}{A}\int JJ_0 = v^*$, where v^* is the solution of

(11) $$P^* = t\int \Phi^{k-t}(x+v)\,(1-\Phi(x))^{t-1}\,d\Phi(x).$$

Recall that $\delta^* = \frac{a}{\sqrt{n}}$; consequently, the sample size required to satisfy (1) is given by:

(12) $$n(P^*,\delta^*,\mathcal{J}) \sim (\frac{v^*}{\delta^*})^2 \frac{\int J^2(u)du - (\int J(u)du)^2}{[\int J(u)\,J_0(u)du]^2}$$

for $\delta^* \sim 0$. QED

If $\mathcal{J}_0$ is any sequence with limit $J_0(u) = -\frac{f'}{f}(F^{-1}(u))$, then $n(P^*,\delta^*,\mathcal{J}_0) \sim (\frac{v^*}{\delta^*})^2 \frac{1}{\int J_0^2(u)du}$. Thus $\lim_{\delta^*\to 0}\frac{n(P^*,\delta^*,\mathcal{J}_0)}{n(P^*,\delta^*,\mathcal{J})} = \frac{(\int J^2-(\int J)^2)(\int J_0^2)}{(\int J\cdot J_0)^2}$

and we have the following

Corollary: Under the conditions of the theorem above, the procedure $R_{\mathcal{J}_0}$ is asymptotically most efficient among all procedures of the $R_{\mathcal{J}}$, with $\mathcal{J}$ satisfying (2).

Bibliography

[1] Hájek, Jaroslav, "Some extensions of the Wald-Wolfowitz-Noether theorem," Ann. Math. Statist. 32 (1961), 283-288.

[2] Hájek, Jaroslav, "Asymptotically most powerful rank-order tests," Ann. Math. Statist. 33 (1962), 1124-1147.

[3] Lehmann, E. L., Testing Statistical Hypotheses, John Wiley and Sons, New York (1959).

[4] Lehmann, E. L., "A class of selection procedures based on ranks," Math. Ann. 150 (1963), 268-275.

[5] Rizvi, M. H., "Ranking and selection problems of normal populations using the absolute values of their means: fixed sample size case," University of Minnesota, Department of Statistics Technical Report No. 31 (1963).