Open access • Journal Article • DOI:10.1007/BF02479757

On the consistency of single-stage ranking procedures — Source link

Yung Liang Tong

Institutions: University of Nebraska–Lincoln

Published on: 01 Dec 1972 - Annals of the Institute of Statistical Mathematics (Kluwer Academic Publishers)

Topics: Consistency (statistics) and Ranking (information retrieval)

ON THE CONSISTENCY OF SINGLE-STAGE RANKING PROCEDURES*

by

Yung Liang Tong

Technical Report No. 131

December 1969

University of Nebraska
and
University of Minnesota

*Research supported by National Science Foundation under Grant No. NSF-GP 11021.

1. Introduction. Let $\mathfrak{F} = \{F(x, \theta): \theta \in \Theta\}$ be a stochastically increasing family of distributions such that the parameter space Θ is a subset of the real line, for every $\theta \in \Theta$ $F(x, \theta)$ is absolutely continuous w.r.t. a fixed (Lebesgue or counting) measure, and $F(x, \theta)$ depends on θ only through its functional form. Let $\delta: R_2 \to R_1$ be the distance function considered by Bechhofer, Kiefer and Sobel [3: p.37] satisfying the following conditions: (i) $\delta(a, b) \geq 0$, (ii) $\delta(a, b) = 0$ iff $a = b$, (iii) $\delta(a, b) = \delta(b, a)$ and $\delta(a, b)$ is strictly increasing (decreasing) in a for fixed b when $a \geq b$ $(a \leq b)$. Then for any two members $F(x, \theta_1)$ and $F(x, \theta_2)$ in $\mathfrak{F}$, the distance between them can be reasonably measured by $\delta(\theta_1, \theta_2)$. In particular $\delta(a, b) = |a - b|$ can be used for the location parameter family and $\delta(a, b) = |\log \frac{a}{b}|$ can be used for the scale parameter family. We note that in general δ is not a metric because the triangle inequality is not assumed. However, the triangle inequality is satisfied in most applications, including the location and scale parameter families and the exponential family.

We first formulate the following ranking and selection problem in the usual way: Let $\pi_1, \pi_2, \ldots, \pi_k$ $(k \geq 2)$ be k populations with distributions $F(x, \theta_i) \in \mathfrak{F}$ $(i = 1, 2, \ldots, k)$, and let $\theta_{[1]} \leq \theta_{[2]} \leq \cdots \leq \theta_{[k]}$ denote the ordered parameter values. For arbitrary but preassigned $\delta^* > 0$ let Ω be a subset of the product parameter space such that

$$(1.1) \qquad \Omega = \Omega(\delta^*) = \{\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k): \delta(\theta_{[k]}, \theta_{[k-1]}) \geq \delta^*\}.$$

Then for every $\underline{\theta} \in \Omega$ the small parameters $\theta_{[1]}, \ldots, \theta_{[k-1]}$ are sufficiently distinct from the large parameter $\theta_{[k]}$. The statistical problem concerned is to find a procedure R for the selection of the "greatest" population associated with parameter $\theta_{[k]}$. Since there is no knowledge about $\underset{\sim}{\psi}$, it is desired to have the probability of correct selection (CS) under R uniformly controlled in Ω such that

$$(1.2) \qquad \inf_{\underset{\sim}{\psi} \varepsilon \Omega} P_{\underset{\sim}{\psi}}[CS|R] \geq \gamma$$

where $\gamma \varepsilon (\frac{1}{k}, 1)$ is arbitrary but preassigned.

Throughout this paper we shall consider only single-stage ranking procedures with an equal number of observations from each one of the k populations. For fixed n let $\{X_j\}$ and $\{X_{ij}\}$ $(j = 1, 2, \ldots, n)$ be independent random variables with distributions $F(x, \theta)$ and $F(x, \theta_i)$ $(i = 1, 2, \ldots, k)$, respectively. Let $T = T^{(n)}: R_n \to R_1$ be a real-valued statistic and

$$t = t^{(n)} = T(X_1, X_2, \ldots, X_n),$$

$$t_i = t_i^{(n)} = T(X_{i1}, X_{i2}, \ldots, X_{in}) \quad \text{for} \quad i = 1, 2, \ldots, k;$$

and denote $G_n(y, \theta)$, $g_n(y, \theta)$ to be the corresponding c.d.f. and density function of t. It is well-known ([1] and [14]) that under reasonable assumptions about $g_n(y, \theta)$ the natural decision rule "always select the population corresponding to the maximum of $(t_1, t_2, \ldots, t_k)$" uniformly minimizes the risk among a class of invariant decision rules based on $(t_1, t_2, \ldots, t_k)$. Hence if the ranking procedure R depends on $\{X_{ij}\}$ only through $(t_1, t_2, \ldots, t_k)$, then R is completely specified by T. We say that

Definition 1. T is consistent w.r.t. $(\mathfrak{Z}, \delta)$ if for every $\delta^* > 0$ and every $\gamma \varepsilon (\frac{1}{k}, 1)$ there exists an $N = N(\mathfrak{Z}, \delta^*, \gamma)$ such that (1.2) holds for every $n > N$.

An equivalent statement to Definition 1 is that the probability on the l.h.s. of (1.2) converges to 1 uniformly in $\underset{\sim}{\psi}$ for $\underset{\sim}{\psi} \in \Omega$; i.e.,

$$\text{(1.3)} \qquad \lim_{n \to \infty} \inf_{\underset{\sim}{\psi} \in \Omega} P_{\underset{\sim}{\psi}}[CS|R] = 1.$$

Under most circumstances the ranking statistic T is chosen to be a consistent estimator of θ. In particular the means procedure (under which T is the sample mean) has been widely accepted for a large number of families of distributions. But in general a consistent estimator of θ is not always consistent for the ranking and selection problem. Moreover, the consistency of a ranking statistic for a certain family of distributions also depends on the distance function δ. The following are some typical examples:

(A) If $\mathfrak{F}$ is the Poisson family with parameter θ, then the means procedure is not consistent w.r.t. $\delta(\theta_{[k-1]}, \theta_{[k]}) = \theta_{[k]} - \theta_{[k-1]}$ or $\delta(\theta_{[k-1]}, \theta_{[k]}) = \log \frac{\theta_{[k]}}{\theta_{[k-1]}}$ (see [17]).

(B) If $\mathfrak{F}$ is the Cauchy family with location parameter θ, then the means procedure is not consistent but the procedure based on the sample medians is consistent.

(C) If $\mathfrak{F}$ is the family of uniform distributions on $[0, \theta]$ for $\theta \in (0, \infty)$, then the means procedure is consistent w.r.t. $\delta(\theta_{[k-1]}, \theta_{[k]}) = \log \frac{\theta_{[k]}}{\theta_{[k-1]}}$, but it is not consistent w.r.t. $\delta(\theta_{[k-1]}, \theta_{[k]}) = \theta_{[k]} - \theta_{[k-1]}$.

In Section 2 we consider the conditions for the consistency of single-stage ranking procedures in general. It is first shown that the consistency of a ranking procedure does not depend on the number of populations involved; and that the consistency of a ranking procedure is related to the uniform consistency of a testing hypothesis problem (Theorem 1). Sufficient conditions for the consistency of ranking procedures are also given in terms of convergence in distributions uniformly in θ to a degenerate distribution (Theorem 2) or to the standard normal distribution (Theorem 3). The conditions on uniform convergence in θ can be found in a paper of Parzen [16].

In Section 3 the general conditions are applied to investigate the consistency of individual ranking procedures which include the means procedure, the procedure based on the maximum likelihood estimator and the procedure based on linear combinations of order statistics for location parameter family. Consistency of ranking procedures for the exponential family is investigated in Section 4.

2. <u>Some General Results</u>. We shall follow the notations developed in Section 1 and we shall assume that $G_n(y, \theta)$ is continuous in y for fixed $\theta \in \Theta$. If $G_n(y, \theta)$ is discrete, randomized decision rules should be considered and most of the following results can be modified easily. We shall also assume that $\mathcal{G} = \{G_n(y, \theta): \theta \in \Theta\}$ is a stochastically increasing (SI) family of distributions. However, the relationship between the SI property of $\mathfrak{F}$ and the SI property of $\mathcal{G}$ is studied in the Appendix of this paper. It is shown that if $t^{(n)}$ satisfies the condition given in (A.1), then $\mathfrak{F}$ is SI will imply $\mathcal{G}$ is SI and the above assumption is not required.

We first show that the consistency of a ranking statistic T does not depend on the number of populations involved.

<u>Lemma 1</u>. T is consistent w.r.t. $(\mathfrak{F}, \delta)$ for any k iff T is consistent w.r.t. $(\mathfrak{F}, \delta)$ for $k = 2$.

Proof: To consider the probability of correct selection under the procedure R based on T we can, without any loss, assume that $\theta_k = \theta_{[k]}$. For $i = 1, 2, \ldots, (k-1)$ let $\Omega_i = \{(\theta_i, \theta_k): \delta(\theta_i, \theta_k) \geq \delta^*\}$ and $A_i = [t_i \leq t_k]$. Then for every k and every ψ in the product parameter space the probability of correct selection is

$$P_\psi[CS|R] = P_\psi[\bigcap_{i=1}^{k-1} A_i].$$

It suffices to show that if $\inf_{(\theta_i,\theta_k)\epsilon\Omega_i} P_{(\theta_i,\theta_k)}(A_i) \to 1$ as $n \to \infty$ for every $i = 1, 2, \ldots, (k-1)$, then $\inf_{\psi\epsilon\Omega} P_\psi[\bigcap_{i=1}^{k-1} A_i] \to 1$ as $n \to \infty$.

Let $\varepsilon > 0$ be arbitrarily small but fixed. Then for every i there exists an N_i such that for every $n > N_i$ we have

$$\inf_{(\theta_i, \theta_k)\epsilon\Omega_i} P_{(\theta_i,\theta_k)}(A_i) \geq 1 - \varepsilon$$

hence

$$\sup_{(\theta_i, \theta_k)\epsilon\Omega_i} P_{(\theta_i,\theta_k)}(\bar{A}_i) \leq \varepsilon.$$

Since $t_1, t_2, \ldots, t_k$ are independent it is easy to see that for every ψ in the product parameter space $P_\psi(A_i)$ depends on ψ only through (θ_i, θ_k). Therefore for every $n > N = \max_{1\leq i\leq(k-1)} N_i$ we have

$$\begin{aligned}\inf_{\psi\epsilon\Omega} P_\psi [\bigcap_{i=1}^{k-1} A_i] = 1 - \sup_{\psi\epsilon\Omega} P_\psi [\bigcup_{i=1}^{k-1} \bar{A}_i] &\geq 1 - \sup_{\psi\epsilon\Omega} \sum_{i=1}^{k-1} P_\psi(\bar{A}_i) \\ &\geq 1 - \sum_{i=1}^{k-1} \sup_{\psi\epsilon\Omega_i} P_\psi(\bar{A}_i) \\ &\geq 1 - (k-1)\varepsilon\end{aligned}$$

which completes the proof.

Following from this lemma we shall, without loss of generality, restrict our attention to the case $k = 2$ for the remainder of this paper. We first observe a relationship between the consistency of ranking procedures and the uniform consistency of hypothesis-testing procedures. Consider a two-sample testing hypothesis problem H_1: $\theta_1 \le \theta_2$ v.s. H_2: $\theta_1 > \theta_2$ where the test ϕ depends on $\{X_{ij}\}$, $j = 1, 2, \ldots, n$; $i = 1, 2$ only through (t_1, t_2) and

$$(2.1) \qquad \phi = \phi(t_1, t_2) = \begin{cases} 1 & \text{if } t_1 \le t_2 \\ 0 & \text{otherwise,} \end{cases}$$

where H_1 is accepted iff $\phi = 1$. Then for every ψ in the product parameter space the expected value of ϕ is $E_{\psi}\phi = P_{\psi}[t_1 \le t_2] = \int_{-\infty}^{\infty} G_n(t, \theta_1) dG_n(t, \theta_2)$. Denote

$$(2.2) \qquad \begin{aligned} \Omega^{(1)} &= \{(\theta_1, \theta_2): \theta_1 < \theta_2, \delta(\theta_1, \theta_2) \ge \delta^*\}, \\ \Omega^{(2)} &= \{(\theta_1, \theta_2): \theta_1 > \theta_2, \delta(\theta_1, \theta_2) \ge \delta^*\}. \end{aligned}$$

The test ϕ is said to be uniformly consistent on $\Omega^{(1)} \cup \Omega^{(2)}$ if $\inf_{\psi \in \Omega^{(1)}} E_{\psi}\phi$ converges to one and $\sup_{\psi \in \Omega^{(2)}} E_{\psi}\phi$ converges to zero as $n \to \infty$.

Theorem 1. T is consistent w.r.t. $(\mathfrak{Z}, \delta)$ iff the test ϕ defined in (2.1) is uniformly consistent on $\Omega^{(1)} \cup \Omega^{(2)}$ for every $\delta^* > 0$.

Proof: We first note that π_2 is selected iff $\phi = 1$. For arbitrary but fixed $\delta^* > 0$, since $\Omega = \Omega^{(1)} \cup \Omega^{(2)}$ the probability of correct selection of the ranking procedure based on T is at least $(1 - \varepsilon)$ iff $\inf_{\psi \in \Omega^{(1)}} E_{\psi}\phi \ge 1 - \varepsilon$ and $\sup_{\psi \in \Omega^{(2)}} E_{\psi}\phi \le \varepsilon$. This completes the proof.

The necessary and sufficient conditions for the existence of uniformly consistent tests have been studied by Berger [4], Kraft [10] and others. Some of those conditions require the compactness of the parameter space Θ. In many applications it is not easy in general to justify whether a certain test is uniformly consistent even if it is known that a uniformly consistent test exists. However, if the power of the test depends on (θ_1, θ_2) only through $\delta(\theta_1, \theta_2)$ and it has the desired monotone property, then a consistent test will be uniformly consistent. In particular, if the test is consistent and has certain invariant property, then it is uniformly consistent.

Corollary. Let there exist a group $\mathcal{G}$ of transformations on $\mathfrak{X} = \{(t_1, t_2)\}$ and let $\bar{\mathcal{G}}$ be the induced group of transformations on the product parameter space. If the test ϕ defined in (2.1) is invariant and if the distance function δ is a maximal invariant w.r.t. $\bar{\mathcal{G}}$, then T is consistent w.r.t. $(\mathfrak{Z}, \delta)$ iff (1) the test ϕ is consistent, (2) the power of the test ϕ is monotonically increasing (decreasing) in $\delta(\theta_1, \theta_2)$ for $(\theta_1, \theta_2) \in \Omega^{(1)}(\Omega^{(2)})$.

Proof: Let U be a maximal invariant w.r.t. $\mathcal{G}$. If ϕ is invariant and δ is a maximal invariant w.r.t. $\bar{\mathcal{G}}$, then ϕ depends on (t_1, t_2) only through U and the distribution of U depends on (θ_1, θ_2) only through $\delta(\theta_1, \theta_2)$. Hence the power of ϕ depends on (θ_1, θ_2) only through $\delta(\theta_1, \theta_2)$ and ϕ is uniformly consistent on $\Omega^{(1)} \cup \Omega^{(2)}$ iff ϕ is consistent and the power function of ϕ has the desired monotone property.

We note that in particular the above corollary applies to location and scale parameter families.

We now observe a nature of the ranking and selection problems. If a ranking statistic T is consistent, then any linear transformation $T' = aT + b$ $(a > 0)$ of T is also consistent. We say

Definition 2. Two ranking statistics T and T' are equivalent if

(2.3) $$P_{\underset{\sim}{\psi}}[t_1 \leq t_2] = P_{\underset{\sim}{\psi}}[t_1' \leq t_2']$$

holds for every n and every $\underset{\sim}{\psi}$ in the product parameter space. Clearly if T' is a strictly monotonically increasing function of T, then T and T' are equivalent and T' is consistent w.r.t. $(\mathfrak{Z}, \delta)$ iff T is. Hence to consider the consistency of different ranking procedures we need to consider only those statistics which are not equivalent. In most cases T is such that $t^{(n)}$ converges to $\tau(\theta)$ in probability as $n \to \infty$ where τ is a continuous, strictly increasing function of θ. Let $T' = \tau^{-1}(T)$. Since convergence in probability is preserved by continuous mappings, it follows that $t'^{(n)}$ converges to θ in probability as $n \to \infty$. Hence without any loss we can consider T to be a consistent estimator of θ in the following theorem.

Theorem 2. Assume that δ is a metric, i.e., in addition to conditions (i) - (iv) on δ the triangle inequality is also satisfied. If, as $n \to \infty$, $\delta(t^{(n)}, \theta)$ converges to 0 in probability uniformly in θ, then T is consistent w.r.t. $(\mathfrak{Z}, \delta)$.

Proof: Let (θ_1, θ_2) be a point in the product parameter space. For arbitrary but fixed $\delta^* > 0$ denote

$$A_1 = \{x: \delta(x, \theta_1) < \frac{\delta^*}{2}\} \text{ and}$$
$$A_2 = \{x: \delta(x, \theta_2) < \frac{\delta^*}{2}\}.$$

Then by conditions (i) - (iv) on δ both A_1 and A_2 are intervals on the real line and θ_1 (θ_2) is an interior point of $A_1(A_2)$. Following from the triangle inequality if there is an $s \in A_1 \cap A_2$, then we must have

$$\delta(\theta_1, \theta_2) \leq \delta(s, \theta_1) + \delta(s, \theta_2) < \delta^*.$$

Therefore $\underset{\sim}{\psi} = (\theta_1, \theta_2) \in \Omega$ implies $A_1 \cap A_2 = \phi$. For arbitrary but fixed $\varepsilon > 0$ since there exists an N such that

$$(2.4) \qquad P_\theta[\delta(t^{(n)}, \theta) < \frac{\delta^*}{2}] \geq 1 - \varepsilon$$

for every $\theta \in \Theta$ whenever $n > N$, it follows that for every $\underset{\sim}{\psi} \in \Omega^{(1)}$ we have

$$P_{\underset{\sim}{\psi}}[CS|R] = P_{\underset{\sim}{\psi}}[t_1 \leq t_2] \geqq P_{\underset{\sim}{\psi}}[t_1 \in A_1, t_2 \in A_2] = (1-\varepsilon)^2$$

whenever $n > N$. Similar argument holds for $\underset{\sim}{\psi} \in \Omega^{(2)}$.

An equivalent statement of (2.4) is that the sequences of distributions of $\delta(t^{(n)}, \theta)$ converges to the distribution function

$$(2.5) \qquad K(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & x \geq 0 \end{cases}$$

uniformly in θ. Some general conditions on the convergence of a family of sequence of distributions to a limiting distribution $H(x)$ uniformly in a parameter θ have been studied by Parzen [16] and others. In particular if $H(x) = K(x)$, then the conditions for the convergence of $\delta(t^{(n)}, \theta)$ to 0 in probability uniformly in θ can be found in Parzen's paper.

Another interesting case is to take $H(x) = \Phi(x)$ where Φ is the standard normal distribution, because in many ranking problems the distributions of the ranking statistics are asymptotically normal. It is well-known that for fixed θ if the sequence of distributions $H_n(x, \theta)$, $n = 1, 2, \ldots$ converges to $\Phi(x)$, then the convergence is uniform in x. However, for fixed x the convergence may not be uniform in θ for θ in Θ. We first observe in the following lemma that if for fixed x the convergence is uniform in θ, then the convergence is uniform both in x and in θ (in fact, we need only the continuity of $\Phi(x)$ in the proof of the lemma).

Lemma 2. Let $\{H_n(x, \theta): \theta \in \Theta\}$, $n = 1, 2, \ldots$ be a family of sequences of distributions and let $\Phi(x)$ denote the standard normal distribution. If for every x there exists an $N' = N'(x, \varepsilon)$ such that

$$(2.6) \qquad |H_n(x, \theta) - \Phi(x)| < \frac{\varepsilon}{5} \text{ for every } \theta$$

whenever $n > N'$, then there exists an $N = N(\varepsilon)$ such that

$$(2.7) \qquad |H_n(x, \theta) - \Phi(x)| < \varepsilon \text{ for every } x \text{ and every } \theta$$

whenever $n > N$.

Proof: For arbitrary but fixed $\varepsilon > 0$ let C be large enough so that $\Phi(-C) < \frac{\varepsilon}{2}$ and $\Phi(C) > 1 - \frac{\varepsilon}{2}$. Let M be large enough so that for every x', x'' in $[-C, C]$ $|x' - x''| \leq \frac{2C}{M}$ implies $|\Phi(x') - \Phi(x'')| < \frac{\varepsilon}{5}$ (this is obviously possible because $\Phi(x)$ is uniformly continuous in $[-C, C]$). Consider the partition $-C = x_0 < x_1 < \ldots < x_M = C$ where $x_{i+1} - x_i = \frac{2C}{M}$, and denote $N_i' = N_i'(x_i, \varepsilon)$ such that (2.6) is satisfied whenever $n > N_i'$ for $i = 0, 1, \ldots, M$. Let $N = \max_{0 \leq i \leq M} N_i'$. Then for every θ and every $x \in [-C, C]$ there is an i such that $x_i \leq x \leq x_{i+1}$ and

$$\begin{aligned} |H_n(x, \theta) - \Phi(x)| &\leq [H_n(x_{i+1}, \theta) - H_n(x_i, \theta)] \\ &\quad + |H_n(x_i, \theta) - \Phi(x_i)| + |\Phi(x_i) - \Phi(x)| \\ &< [\{\Phi(x_{i+1}) + \tfrac{\varepsilon}{5}\} - \{\Phi(x_i) - \tfrac{\varepsilon}{5}\}] + \tfrac{2\varepsilon}{5} < \varepsilon \end{aligned}$$

whenever $n > N$. The cases that $x < -C$ and $x > C$ can easily be taken care of. Hence the lemma is proved.

We now proceed to investigate the consistency of a class of ranking statistics which have asymptotically normal distributions. Let T be the ranking statistic under consideration and denote $\tau_n(\theta) = E_\theta t^{(n)}$, $\sigma_n^2(\theta) = E_\theta[t^{(n)} - \tau_n(\theta)]^2$,

$$H_n(x, \theta) = P_\theta[\frac{t^{(n)} - \tau_n(\theta)}{\sigma_n(\theta)} \leq x].$$

Since by assumption the family of distributions $\{G_n(y, \theta): \theta \in \Theta\}$ is a stochastically increasing family, we have $\tau_n(\theta_1) < (>) \tau_n(\theta_2)$ if $\theta_1 < (>) \theta_2$.

Theorem 3. Assume $H_n(x, \theta) \to \Phi(x)$ uniformly in θ. Then T is consistent w.r.t. $(\mathfrak{J}, \delta)$ iff the absolute value of

$$C_n(\underline{\psi}) = \frac{[\tau_n(\theta_2) - \tau_n(\theta_1)]}{[\sigma_n^2(\theta_1) + \sigma_n^2(\theta_2)]^{\frac{1}{2}}} \tag{2.8}$$

approaches to ∞ uniformly in $\underline{\psi}$ for $\underline{\psi} \in \Omega$. In particular, if $t^{(n)}$ is an unbiased estimator of θ and $\delta(\theta_1, \theta_2) = |\theta_1 - \theta_2|$, then the above condition reduces to

$$\sigma_n^2(\theta) \to 0 \text{ uniformly in } \theta. \tag{2.9}$$

Proof: Let $\underline{\psi} = (\theta_1, \theta_2)$ be any point in the product parameter space. Without any loss assume $\theta_1 < \theta_2$. Then the probability of correct selection at $\underline{\psi}$ is

$$P_{\underline{\psi}}[CS] = P_{\underline{\psi}}[t_1 \leq t_2] = P_{\underline{\psi}}[Z_1 \leq r_n Z_2 + S_n]$$

$$= \int_{-\infty}^{\infty} H_n(r_n x + S_n, \theta_1) dH_n(x, \theta_2)$$

where $Z_i = \frac{t_i - \tau_n(\theta_i)}{\sigma_n(\theta_i)}$ for $i = 1, 2$; $r_n = r_n(\underline{\psi}) = \frac{\sigma_n(\theta_2)}{\sigma_n(\theta_1)}$

and $S_n = S_n(\underset{\sim}{\psi}) = \dfrac{\tau_n(\theta_2) - \tau_n(\theta_1)}{\sigma_n(\theta_1)} > 0$. By Lemma 2 there exists an N_1 (which does not depend on x and θ) so that (2.7) holds whenever $n > N_1$. Thus for every $n > N_1$ we have

$$(2.10) \qquad \left|\int_{-\infty}^{\infty} H_n(r_n x + S_n, \theta_1) dH_n(x, \theta_2) - \int_{-\infty}^{\infty} \Phi(r_n x + S_n) dH_n(x, \theta_2)\right| < 2\epsilon.$$

Now we claim that there exists an N_2 (which does not depend on $\underset{\sim}{\psi}$) such that for every $n > N_2$ we have

$$(2.11) \qquad \left|\int_{-\infty}^{\infty} \Phi(r_n x + S_n) dH_n(x, \theta_2) - \int_{-\infty}^{\infty} \Phi(r_n x + S_n) d\Phi(x)\right| < \epsilon.$$

This is not an immediate consequence of the well-known Helly-Bray Theorem or the uniform version of the Helly-Bray Theorem given by Parzen [16: p.30], because $H_n(x, \theta_2)$ depends on $\underset{\sim}{\psi}$ through θ_2 and the integrand $\Phi(r_n x + S_n)$ depends on both $\underset{\sim}{\psi}$ and n. However, since $\Phi(x)$ is uniformly continuous on the real line and $H_n(x, \theta)$ converges to $\Phi(x)$ uniformly in x and in θ, the proof of (2.11) is similar to the proof of the Helly-Bray Theorem; so the detail is omitted here.

Since for every fixed $\underset{\sim}{\psi}$ and n

$$\int_{-\infty}^{\infty} \Phi(r_n x + S_n) d\Phi(x) = \Phi(C_n(\underset{\sim}{\psi})),$$

combining (2.10) and (2.11) for every $n > N = \max\{N_1, N_2\}$ we have

$$(2.12) \qquad \Phi(C_n(\underset{\sim}{\psi})) - 3\epsilon < P_{\underset{\sim}{\psi}}[CS] < \Phi(C_n(\underset{\sim}{\psi})) + 3\epsilon$$

where N does not depend on $\underset{\sim}{\psi}$. Hence $\lim\limits_{n\to\infty} \inf\limits_{\underset{\sim}{\psi}\in\Omega} P_{\underset{\sim}{\psi}}[CS] \to 1$ iff $\lim\limits_{n\to\infty} \inf\limits_{\underset{\sim}{\psi}\in\Omega} C_n(\underset{\sim}{\psi}) = \infty$ and the proof of the theorem is completed.

Remark. We make the following remark which can easily be justified from the way we prove the above theorem: Let there be k populations involved in a ranking problem for any $k \geq 2$. If the ranking statistic T is such that $H_n(x, \theta) \to \Phi(x)$ uniformly in θ, then for every $\underset{\sim}{\psi} = (\theta_1, \theta_2, \ldots, \theta_k)$ in the product parameter space the probability of correct selection $P_{\underset{\sim}{\psi}}[CS]$ converges to

$$\int_{-\infty}^{\infty} [\prod_{i=1}^{k-1} \Phi(r_{n,i}x + S_{n,i})]d\Phi(x)$$

uniformly in $\underset{\sim}{\psi}$, where

$$r_{n,i} = r_{n,i}(\underset{\sim}{\psi}) = \frac{\sigma_n(\theta_{[k]})}{\sigma_n(\theta_{[i]})}, \; s_{n,i} = s_{n,i}(\underset{\sim}{\psi}) = \frac{\tau_n(\theta_{[k]}) - \tau_n(\theta_{[i]})}{\sigma_n(\theta_{[i]})}$$

and $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ is the ordered θ values; this limiting probability can be computed from multivariate normal probabilities.

3. Consistency of Some Commonly-Used Ranking Procedures. In this section we apply the results developed in Section 2 to investigate the consistency of some individual ranking statistics which have been used or can be used in most cases.

A. The Means Procedure. It appears that the means procedure (under which the ranking statistic is the sample mean) has been the most important ranking procedure considered among the literature of ranking and selection problems. Individual applications of this procedure have been made to normal [2], Binomial [18], Poisson [17], Gamma [7] and location parameter family [13]. Applications to other families of distributions have also been considered.

Perhaps the importance of the means procedure can be partially justified by the Weak Law of Large Numbers. If the family of distributions $\mathfrak{F} = \{F(x, \theta): \theta \in \Theta\}$ is a stochastically increasing family and if $E_\theta X = \tau(\theta)$ exists for $\theta \in \Theta$, then $\overline{X}^{(n)}$ converges to $\tau(\theta)$ in probability as $n \to \infty$ and $\tau(\theta)$ is a monotonically increasing function of θ. Following from the argument in Section 2 again we can assume, without loss of generality, that $\tau(\theta) = \theta$. If $\mathfrak{J}$ is a location (or scale) parameter family, then $\delta(\overline{X}^{(n)}, \theta) = |\overline{X}^{(n)} - \theta|$ (or $\delta(\overline{X}^{(n)}, \theta) = |\log \frac{\overline{X}^{(n)}}{\theta}|$) converges to 0 in probability uniformly in θ. It follows from Theorem 2 and the Weak Law of Large Numbers that if the first moment exists, then the means procedure is consistent w.r.t. location and scale parameter families.

Another important case of the means procedure is a consequence of the Uniform Central Limit Theorem. If the second moment also exists, then $P_\theta[\frac{\sqrt{n}(\overline{X}^{(n)} - \theta)}{\sigma(\theta)} \leq x] \to \Phi(x)$ as $n \to \infty$ for every $\theta \in \Theta$ when $E_\theta X = \theta$ and $E_\theta(X - \theta)^2 = \sigma^2(\theta)$. It follows from Theorem 3 that if the convergence is uniform in θ, then the means procedure is consistent iff $\sigma^2(\theta)$ is bounded in Θ. If $\sigma^2(\theta)$ is continuous in θ, then the means procedure is consistent if the parameter space Θ is bounded. We note that the conditions for the Uniform Central Limit Theorem given in [16] can be easily verified in many applications.

B. The Procedure Based on the Maximum Likelihood Estimator of θ. The maximum likelihood principle has played an important role in statistical estimation theory and the asymptotic behavior of the maximum likelihood estimator has been fully studied. But the role of the procedure based on the maximum likelihood estimator in ranking and selection problems has not been clarified yet because among most of the ranking problems considered this procedure is identical to the means procedure. It was shown in [16] that under certain conditions the maximum likelihood estimator converges to θ in probability uniformly in θ, and the c.d.f. of the standardized maximum likelihood estimator (with mean 0 and variance 1) converges to $\Phi(x)$ uniformly in θ. Hence if those conditions in [16] are satisfied, then the ranking procedure based on the maximum likelihood estimator is consistent when the distance function $\delta(\theta_1, \theta_2) = |\theta_1 - \theta_2|$ is used.

C. The Procedure Based on Linear Combination of Order Statistics for Location Parameter Family. We first look at the estimators for θ for the location parameter family $\mathfrak{F} = \{F(x, \theta) = F(x - \theta): \theta \in \Theta\}$. If the first moment exists and $E_\theta X = \theta$, then certainly the sample mean $\bar{X}^{(n)}$ can serve as an estimator of θ. However, in several occasions either the first moment does not exist or the estimator $\bar{X}^{(n)}$ is inefficient, other estimators have been considered. The estimator of θ based on linear combinations of order statistics and its asymptotic behavior has been studied recently by Chernoff, Gastwirth and Johns [5] and others. Following their notations let $Y_1 \leq Y_2 \leq \cdots \leq Y_n$ be the order statistics of n random samples from a population with c.d.f. $F(x - \theta)$ and let $t^{(n)} = \frac{1}{n} \sum_{j=1}^{n} c_j Y_j$ where $\underset{\sim}{c} = (c_1, c_2, \ldots, c_n)$ can be found in [5]. Then it is easy to check that (1) $\sigma^2 = E_\theta[\frac{\partial}{\partial\theta} \log f(x - \theta)]^2$ does not depend on θ and (2) for every $\underset{\sim}{c}$ $H_n(x, \theta) = P_\theta[\frac{t^{(n)} - \tau(\theta)}{\sigma} \leq x]$ does not depend on θ where $\tau(\theta) = E_\theta t^{(n)}$. Results in [5] assert that if the regularity conditions are satisfied, then $H_n(x, \theta) \to \Phi(x)$ for every $\theta \in \Theta$. This implies that $H_n(x, \theta) \to \Phi(x)$ uniformly in θ and (by Theorem 3) the ranking procedure based on this linear combination of order statistics is consistent for location parameter family.

4. Remarks on the Exponential Family. Consider a family of distributions $\mathfrak{F}$ with density functions $\{f(x, \theta): \theta \in \Theta\}$. $\mathfrak{F}$ is said to be in the exponential family if $f(x, \theta)$ has the form

$$f(x, \theta) = A(x)B(\theta)e^{Q(\theta)R(x)}$$

where A, B, Q, R are real-valued functions. Let $\{X_{ij}\}$, $j = 1, 2, \ldots, n$; $i = 1, 2$ be random samples taken from two populations with densities $f(x, \theta_1)$ and $f(x, \theta_2)$, respectively and let (θ_1^o, θ_2^o) be any constant vector such that $\theta_1^o < \theta_2^o$. Then in the testing hypothesis problem

$$H_1': \ (\theta_1, \theta_2) = (\theta_1^o, \theta_2^o) \ , \text{ v.s.}$$

$$H_2': \ (\theta_1, \theta_2) = (\theta_2^o, \theta_1^o)$$

H_1' is accepted iff

$$(4.1) \qquad \lambda_n = \log \frac{\prod_{j=1}^{n} [f(x_{1j}, \theta_1^o) f(x_{2j}, \theta_2^o)]}{\prod_{j=1}^{n} [f(x_{1j}, \theta_2^o) f(x_{2j}, \theta_1^o)]}$$

$$= [Q(\theta_1^o) - Q(\theta_2^o)][\sum_{j=1}^{n} R(X_{1j}) - \sum_{j=1}^{n} R(X_{2j})] > C$$

where C is a real number. If $Q(\theta)$ is monotonically increasing in θ the test

$$(4.2) \qquad \phi = \begin{cases} 1 & \text{if } \sum_{j=1}^{n} R(X_{1j}) \leq \sum_{j=1}^{n} R(X_{2j}) \\ 0 & \text{otherwise} \end{cases}$$

is uniformly most powerful for the hypothesis

$$H_1: \ \theta_1 \leq \theta_2 \ \text{ v.s.}$$

$$H_2: \ \theta_1 > \theta_2$$

where H_1 is accepted iff $\phi = 1$; and ϕ is uniformly consistent on Ω iff

$$\inf_{\psi\epsilon\Omega_1} P_{\psi}[\frac{1}{n}\sum_{j=1}^{n} R(X_{1j}) \le \frac{1}{n}\sum_{j=1}^{n} R(X_{2j})] \to 1,$$

(4.3)

$$\sup_{\psi\epsilon\Omega_2} P_{\psi}[\frac{1}{n}\sum_{j=1}^{n} R(X_{ij}) \le \frac{1}{n}\sum_{j=1}^{n} R(X_{2j})] \to 0$$

as $n \to \infty$, where Ω_1 and Ω_2 are defined in (2.2).

In many ranking and selection problems for families of distributions in the exponential family the distance function

$$\delta(\theta_1, \theta_2) = |Q(\theta_1) - Q(\theta_2)| \tag{4.4}$$

has been used when Q is monotonically increasing in θ. It follows from Theorem 1 that the ranking procedure for the exponential family based on $\sum_j R(X_j)$ is consistent iff (4.3) holds when $|Q(\theta_1) - Q(\theta_2)| > \delta^*$ for arbitrary but fixed $\delta^* > 0$.

We note that Bechhofer, Kiefer and Sobel have defined a "Rankability Condition" in their book [3: p.41] for sequential ranking procedures. It can be seen that under their rankability condition the O.C. curve of the Sequential Probability Ratio Test for our hypothesis H_1' v.s. H_2' defined by λ_n depends on (θ_1^o, θ_2^o) only through $\delta(\theta_1^o, \theta_2^o)$; hence a solution for the identification problem will lead to a solution for the ranking problem. In particular they have proved that the rankability condition is satisfied for all the families of distributions in the exponential family. But we do not have this advantage when single-stage ranking procedures are used. Because under the single-stage sampling rule the power of the test ϕ defined in (4.2) does not always depend on (θ_1, θ_2) only through $\delta(\theta_1, \theta_2)$. In fact, for some families of distributions (4.3) can not be satisfied and the single-stage ranking procedure is not consistent (the Poisson family is one of the examples).

5. Acknowledgement. I wish to thank Professors Milton Sobel and Michael D. Perlman for some comments.

APPENDIX

On a Property of Stochastically Increasing Families

Let $\mathfrak{F} = \{F(x, \theta) \colon \theta \in \Theta\}$ be a family of distribution functions such that for every $\theta \in \Theta$ $F(x, \theta)$ is absolutely continuous with respect to a fixed (Lebesgue or counting) measure μ, and $F(x, \theta)$ depends on θ only through its functional form (Θ is referred as the parameter space and is usually an interval on the real line). $\mathfrak{F}$ is said to be a stochastically increasing (SI) family of distributions if $\theta_1, \theta_2 \in \Theta$ and $\theta_1 < \theta_2$ implies $F(x, \theta_2) \leq F(x, \theta_1)$ for every x. It is well-known that the class of SI families contains most of the familiar distributions; also, in most cases the distribution of a statistic with random samples from a SI family also belongs to a SI family. Hence it is a natural thing to ask: under what condition(s) this SI property will be preserved?

In this appendix we apply some results of Lehmann in [11], [12] to give a solution to this problem. It is shown that a certain monotone property of the statistic serves as a sufficient condition. For $n \geq 1$ let $t = t^{(n)} \colon R_n \to R_1$ be a Borel measurable function such that for every $i = 1, 2, \ldots, n$,

$$t^{(n)}(\underset{\sim}{x}) = t(x_1, x_2, \ldots, x_n) \uparrow x_i \tag{A.1}$$

for every fixed $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ (where $\uparrow$ means non-decreasing). Let $X_1, X_2, \ldots, X_n$ be independently, identically distributed random variables with distribution function $F(x, \theta) \in \mathfrak{F}$, and let $G_n(y, \theta)$ denote the distribution function of $t^{(n)}(\underset{\sim}{X}) = t(X_1, X_2, \ldots, X_n)$.

Theorem. If $\mathfrak{F}$ is a SI family and t_n satisfies (A.1), then $\mathcal{G} = \{G_n(y, \theta): \theta \in \Theta\}$ is a SI family.

Proof: We need to show that if $\mathfrak{F}$ is SI, then for every $\theta_1, \theta_2 \in \Theta$ such that $\theta_1 < \theta_2$ and every real number c the inequality

$$\text{(A.2)} \qquad P_{\theta_2}[t(X_1, X_2, \ldots, X_n) \leq c] \leq P_{\theta_1}[t(X_1, X_2, \ldots, X_n) \leq c]$$

holds.

By a lemma of [12: p.73], there exist two non-decreasing, real-valued functions h_1 and h_2 and independently, identically distributed random variables $Z_1, Z_2, \ldots, Z_n$ such that

$$h_1(z) \leq h_2(z) \qquad \text{for every } z,$$

and for $i = 1, 2, \ldots, n$

$$F(x, \theta_1) = P_{\theta_1}[X_i \leq x] = P_{\theta_1}[h_1(Z_i) \leq x],$$

$$F(x, \theta_2) = P_{\theta_2}[X_i \leq x] = P_{\theta_2}[h_2(Z_i) \leq x]$$

for every x. Hence by taking $g_1 = g_2 = \ldots = g_n = h_1^{-1}$ and $g_1' = g_2' = \ldots = g_n' = h_2^{-1}$ the Condition (A) in [11] is satisfied. For arbitrary but fixed real number c let the Borel measurable set S in R_n be

$$\text{(A.3)} \qquad S = \{\underset{\sim}{x}: t(x_1, x_2, \ldots, x_n) > c\}.$$

Then by the condition imposed on $t^{(n)}$ in (A.1) the set S is an increasing set. Hence by Theorem 1 of [11] we have

$$P_{\theta_1}(S) \leq P_{\theta_2}(S),$$

or equivalently,

$$P_{\theta_2}[t(X_1, X_2, \ldots, X_n) \leq c] \leq P_{\theta_1}[t(X_1, X_2, \ldots, X_n) \leq c]$$

which completes the proof of the theorem.

We observe that [11] defines a large class of statistics including the mean (or any linear combination of the observations with non-negative coefficients), the median, the maximum or minimum or any other order statistics. We also observe the property of the distributions of the maximum likelihood estimators $\hat{\theta}$, which play an important role in estimation theory. Assume the regular conditions are satisfied so that $\hat{\theta}$ is the solution of the equation $\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(x_i, \theta) = 0$; where $f(x, \theta)$ denotes the corresponding density function. If $f(x, \theta)$ depends on x and θ only through $u = u(x, \theta)$ and if $(\frac{\partial}{\partial\theta} u(x, \theta))(\frac{\partial}{\partial x} u(x, \theta)) < 0$, then $\hat{\theta}$ satisfies (A.1) and the distribution of $\hat{\theta}$ belongs to a SI family. In particular, this applies to location and scale (of non-negative random variables) parameter families.

REFERENCES

[1] Bahadur, R. R. and Goodman, L. A. (1952). Impartial decision rules and sufficient statistics. Ann. Math. Statist., 23, 553-562.

[2] Bechhofer, R. E. (1954). A single-stage multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist., 25, 16-39.

[3] Bechhofer, R. E., Kiefer, J. and Sobel, M. (1968). Sequential Identification and Ranking Procedures. I.M.S. Monograph Series, Chicago University Press.

[4] Berger, A. (1951). On uniformly consistent tests. Ann. Math. Statist., 22, 289-293.

[5] Chernoff, H., Gastwirth, J. L. and Johns, M. V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. Ann. Math. Statist., 38, 52-72.

[6] Ferguson, T. S. (1967). Mathematical Statistics--A Decision Theoretic Approach. Academic Press.

[7] Gupta, S. S. (1963). On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math., Tokyo. 14, 199-216.

[8] Hall, W. J. (1959). The most economical character of some Bechhofer and Sobel decision rules. Ann. Math. Statist., 30, 964-969.

[9] Hoeffding, W. and Wolfowitz, J. (1958). Distinguishability of sets of distributions (the case of independent and identically distributed chance variables). Ann. Math. Statist., 29, 700-718.

[10] Kraft, C. H. (1955). Some conditions for consistency and uniform consistency of statistical procedures. Univ. of Calif. Publ. in Statist., 2 (No. 6), 125-142.

[11] Lehmann, E. L. (1955). Ordered families of distributions. Ann. Math. Statist., 26, 399-419.

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

[13] Lehmann, E. L. (1963). A class of selection procedures based on ranks. Math. Ann. (Part III), 268-275.

[14] Lehmann, E. L. (1966). On a theorem of Bahadur and Goodman. Ann. Math. Statist., 37, 1-6.

[15] Mahamunulu, D. M. (1967). Some fixed-sample ranking and selection problems. Ann. Math. Statist., 38, 1079-1091.

[16] Parzen, E. (1954). On uniform convergence of families of sequences of random variables. Univ. of Calif. Publ. in Statist., 2 (No. 2), 23-54.

[17] Sobel, M. (1963). Single-sample ranking problems with Poisson populations. Tech. Report No. 19, Dept. of Stat., Univ. of Minn., Minneapolis, Minnesota.

[18] Sobel, M. and Huyett, M. (1957). Select the best one of several binomial populations. Bell System Tech. J. 36, 537-576.

[19] Tong, Y. L. (1969). On partitioning a set of normal populations by their locations with respect to a control. Ann. Math. Statist., 40, 1300-1324.

[20] Tong, Y. L. (1970). Asymptotic efficiency of ranking procedures for location parameter family. In preparation.