A LARGE DEVIATION APPROACH TO ASYMPTOTIC EFFICIENCY OF SELECTION PROCEDURES

by

John A. Gaynor
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #474

*This research is based on the author's Ph.D. dissertation submitted at the University of Illinois at Urbana-Champaign, August, 1976.

AMS 1970 subject classification. 62F07

Summary

A Large Deviation Approach to Asymptotic Efficiency of Selection Procedures

For $i = 1,\ldots,k$, consider a sample of size n from $f(x - \theta_i)$, where f is known. The problem is to select the single population with the largest location parameter. Assume that the difference between the two largest parameters is $\Delta > 0$. For unbiased selection procedures, an optimal value is given for the exponential rate of convergence to 0 of the error probability as $n \to \infty$. Selection procedures based on Bahadur-efficient estimators attain this optimal value as $\Delta \to 0$. The scale parameter case is similar. Examples of the computation of the rate of convergence are given for procedures based on sample means or on sample medians, and also for a nonparametric procedure introduced by Bechhofer and Sobel.

A Large Deviation Approach to Asymptotic Efficiency of Selection Procedures[1]

(Efficiency of Selection Procedures)

by

John A. Gaynor
Purdue University

1. Introduction and summary. Consider k univariate populations $\pi_1,\ldots,\pi_k$, where for $i = 1,\ldots,k$, π_i has density $g(\cdot|\theta_i)$ with respect to Lebesque measure. Let $\vec{\theta} = (\theta_1,\ldots,\theta_k)$ be the unknown vector of real-valued parameters. Let $X_{i1},\ldots,X_{in}$ be a sample from π_i, $i = 1,\ldots,k$. Consider selection procedures (of sample size n) which select one of the k populations. The selection is based on the samples, and the goal is to choose the population corresponding to the largest parameter.

Two cases will be considered:

(1) Location parameter case: $g(x|\theta_i) = f(x - \theta_i)$, $i = 1,\ldots,k$, and f is known. Assume that $\theta_k = \max_i \{\theta_i\}$ and that $\min_i \{\theta_k - \theta_i\} = \Delta > 0$. The latter assumption is motivated by the "indifference zone" approach of Bechhofer (1954). Without loss of generality, assume that the true value of $\vec{\theta}$ is

(1.1) $$\vec{\theta}_0 = (0, \theta_2,\ldots,\theta_{k-1},\Delta),$$

where

(1.2) $$\theta_i \leq 0 \qquad \text{for } i = 2,\ldots,k-1.$$

---

[1]This research is based on the author's Ph.D. dissertation submitted at the University of Illinois at Urbana-Champaign, August, 1976.

AMS 1970 subject classification. 62F07

Key words and phrases. Selection procedure, large deviation.

(2) Scale parameter case: $g(x|\theta_i) = (1/\theta_i)f(x/\theta_i)$, $\theta_i > 0$, $i = 1,\ldots,k$, and f is known. Assume that $\theta_k = \max_i\{\theta_i\}$ and that $\min_i \{\theta_k/\theta_i\} = \Delta > 1$. The latter assumption is also motivated by the indifference zone approach. Without loss of generality, assume that the true value of $\vec{\theta}$ is

$$\vec{\theta}_0 = (1, \theta_2,\ldots,\theta_{k-1}, \Delta), \tag{1.3}$$

where

$$0 < \theta_i \le 1 \qquad \text{for } i = 2,\ldots,k-1. \tag{1.4}$$

Throughout this paper, $\vec{\theta}_0$ will satisfy either (1.1) and (1.2), or (1.3) and (1.4), depending on which case is being considered.

Let R be a selection procedure and define $P(i|\vec{\theta}, R, n)$ as the probability that procedure R selects π_i when $\vec{\theta}$ holds and the sample size is n. Every selection procedure R considered in this paper is assumed to be eventually unbiased in the sense that it satisfies the following property:

(1.5) If $\theta_i \ge \theta_j$, then $P(i|\vec{\theta}, R, n) \ge P(j|\vec{\theta}, R, n)$ for n sufficiently large, $i, j = 1,\ldots,k$.

For any selection procedure R, define

$$\bar{c}(\vec{\theta}_0, R) = \limsup_{n\to\infty} n^{-1}(-\log P(E|\vec{\theta}_0, R, n)), \tag{1.6}$$

where E denotes "error", i.e., selection of some π_i, $i = 1,\ldots,k-1$. Similarly, define $\underline{c}$ and c, if the limit exists.

Let R_1 and R_2 be selection procedures, and assume $c(\vec{\theta}_0, R_i)$ exists, $i = 1,2$. The ratio $c(\vec{\theta}_0, R_1)/c(\vec{\theta}_0,R_2) = r$, say, can be interpreted as the limiting ratio of the sample size for R_2 to that for R_1 as the common probability of error for the two procedures approaches zero. So r is a measure of the asymptotic relative efficiency (ARE) at $\vec{\theta}_0$ of R_1 with respect to R_2. This is essentially the definition of ARE introduced by Dudewicz (1971) in the location parameter case. It is analagous to the ARE for tests defined by Hodges and Lehmann (1956, p. 330). Another measure of ARE for selection procedures is lim(r), if it exists, where the limit is computed as $\Delta \to 0$ in the location parameter case, and as $\Delta \to 1$ in the scale parameter case. This limit may still depend on $\theta_2,\ldots,\theta_{k-1}$.

For any procedure R, the quantity $c(\vec{\theta}_0, R)$, if it exists, can be used to get an approximation to the sample size n required for procedure R to achieve a specified error probability P': $n \sim -\log P'/c(\vec{\theta}_0,R)$. The ratio tends to 1 as $n \to \infty$. The accuracy of this approximation has been studied by Dudewicz and Zaino (1971) in the location parameter case for normal populations.

Section 2 of this paper contains some preliminary results and definitions. In Section 3, an upper bound is obtained for $\bar{c}(\vec{\theta}_0, R)$ in the location parameter case. A definition of asymptotic efficiency of a selection procedure is given, and it is shown that if an estimator is efficient in the Bahadur (large deviation) sense, then the selection procedure based on that estimator is also efficient. In Section 4, the results of Section 3 are shown to hold for the scale parameter case also. Section 5 contains examples of the computation of c for selection procedures based on sample means or on sample medians, and also for a nonparametric procedure introduced by Bechhofer and Sobel (1958).

2. Preliminary results and definitions. The following well-known lemma will be needed.

LEMMA 2.1. For $n = 1,2,\ldots$, let P_n denote a probability measure on (R^m, B^m), where R^m is m-dimensional Euclidean space, and B^m is the σ-field of Borel subsets. If $A_i^{(n)} \in B^m$ for $i = 1,\ldots,s$ and $n = 1,2,\ldots$, then

$$\limsup_{n\to\infty} n^{-1} \log P_n(\bigcup_{i=1}^{s} A_i^{(n)}) = \max_i\{\limsup_{n\to\infty} n^{-1} \log P_n(A_i^{(n)})\}.$$

A similar identity holds for "lim inf" under the additional assumption that for some fixed j, $\max_i\{P_n(A_i^{(n)})\} = P_n(A_j^{(n)})$ for n sufficiently large.

The proof is based on the fact that

$$\max_i P_n(A_i^{(n)}) \leq P_n(\bigcup_{i=1}^{s} A_i^{(n)}) \leq k \max_i\{P_n(A_i^{(n)})\} \qquad \text{for } n = 1,2,\ldots.$$

Since E is the union of the events $\{\pi_i$ is selected$\}$, $i = 1,\ldots,k-1$, it follows from (1.5), (1.6), and Lemma 2.1 that

$$\bar{c}(\vec{\theta}_0, R) = \limsup_{n\to\infty} n^{-1}(-\log P(1|\vec{\theta}_0, R, n)), \tag{2.1}$$

with similar results for $\underline{c}$ and c.

Now consider a selection procedure R based on k independent statistics $T_{in}(X_{i1},\ldots,X_{in})$, $i = 1,\ldots,k$, which selects π_i if $T_{in} = \max_j\{T_{jn}\}$. If there is a tie for the maximum, one of the tied populations is selected at random. To compute $\bar{c}(\vec{\theta}_0, R)$ for such procedures, only π_1 and π_k need to be considered. To see this, let $q_n = P(T_{1n} > T_{kn}) + (1/2)P(T_{1n} = T_{kn})$. So q_n is the error probability for procedure R if π_1 and π_k are the only populations. Then $P(1|\vec{\theta}_0, R, n) \leq q_n \leq P(E|\vec{\theta}_0, R, n)$. It follows from (1.6) and (2.1) that $\bar{c}(\vec{\theta}_0, R) = \limsup_{n\to\infty} n^{-1}(-\log q_n)$. In particular, $\bar{c}(\vec{\theta}_0, R)$ is independent of k. Similar conclusions hold for $\underline{c}$ and c.

Define $I(\alpha,\beta) = \int g(x|\alpha) \log[g(x|\alpha)/g(x|\beta)]dx$ and $J(\alpha,\beta) = I(\alpha,\beta) + I(\beta,\alpha)$. I and J are measures of information defined by Kullback and Leibler (1951). Note that in the location parameter case, $I(\alpha,\beta) = I(0,\beta-\alpha)$, and in the scale parameter case, $I(\alpha,\beta) = I(1,\beta/\alpha)$.

3. Location parameter case. Assume $g(x|\theta_i) = f(x - \theta_i)$, $i = 1,\ldots,k$, and f is known. Let $\vec{\theta}_0$ be fixed and satisfy (1.1) and (1.2), with $\Delta > 0$.

THEOREM 3.1. For any eventually unbiased selection procedure R, $\bar{c}(\vec{\theta}_0, R) \leq J(0, \Delta/2)$.

PROOF. Consider a test of the hypothesis H_0: $\vec{\theta} = \vec{\theta}_0$ against the alternative H_1: $\vec{\theta} = (\Delta/2, \theta_2,\ldots,\theta_{k-1}, \Delta/2) = \vec{\theta}_1$, say. The test is based on the n observations $(X_{1j},\ldots,X_{kj})$, $j = 1,\ldots,n$.

Under H_0 the density function of the observations is $g_0(\vec{x}) = f(x_1)f(x_k - \Delta) \prod_{i=2}^{k-1} f(x_i - \theta_i)$, where $\vec{x} = (x_1,\ldots,x_k)$. Under H_1 the density function is $g_1(\vec{x}) = f(x_1 - \Delta/2)f(x_k - \Delta/2) \prod_{i=2}^{k-1} f(x_i - \theta_i)$. Lemma 6.1 of Bahadur (1971) implies that for any β, $0 < \beta < 1$, if for each n, $\alpha_n(\beta)$ is the infimum of all sizes of tests with power $1-\beta$ against H_1, then $\lim_{n\to\infty} n^{-1} \log \alpha_n(\beta) = -\int g_1(\vec{x})\log[g_1(\vec{x})/g_0(\vec{x})]d\vec{x} = J(0,\Delta/2)$.

Now consider any selection procedure R satisfying (1.5). R can be viewed as a test of H_0 versus H_1 by rejecting H_0 if π_1 is selected, and accepting H_0 otherwise. The power of this test against H_1 is $P(1|\vec{\theta}_1,R,n) \geq 1/k$, for n sufficiently large, by (1.5). Therefore, taking $\beta = 1 - 1/k$, it follows that $P(1|\vec{\theta}_0,R,n) \geq \alpha_n(\beta)$, for n sufficiently large, since $P(1|\vec{\theta}_0,R,n)$ is the size of the test based on R. Therefore, $\liminf_{n\to\infty} n^{-1} \log P(1|\vec{\theta}_0,R,n) \geq -J(0,\Delta/2)$, which implies $\bar{c}(\vec{\theta}_0,R) \leq J(0,\Delta/2)$. This completes the proof.

It should be noted that the upper bound in Theorem 3.1 is independent of $\theta_2,\ldots,\theta_{k-1}$, and so the bound in uniform for all values of $\vec{\theta}_0$ satisfying (1.1) and (1.2).

In view of Theorem 3.1, if $J(0, \Delta/2)$ is finite, define the (asymptotic) efficiency at $\vec{\theta}_0$ of a location parameter selection procedure R as

(3.1) $$e(\vec{\theta}_0, R) = \liminf_{\Delta\to 0} \underline{c}(\vec{\theta}_0, R)/J(0, \Delta/2).$$

Then $0 \le e(\vec{\theta}_0, R) \le 1$ for all R and $\vec{\theta}_0$. Note that $e(\vec{\theta}_0, R)$ may depend on the values of $\theta_2,\ldots,\theta_{k-1}$. When this is not the case, then $e(\vec{\theta}_0, R)$ will be written more simply as e(R).

Now consider selection procedures based on estimators $T_i = T_{in}$ of θ_i, where $T_i = T(X_{i1},\ldots,X_{in})$, $i = 1,\ldots,k$, for some estimator $T = T_n$. As shown in Section 2, for such procedures it is sufficient to consider the case of two populations. The efficiency as defined in (3.1) is independent of $\theta_2,\ldots,\theta_{k-1}$. So let $\vec{\theta}_0 = (0,\Delta)$. Then the estimators are T_1 and T_2.

In the following theorem, T is assumed to be translation equivariant, i.e., if b is a constant, then $T(X_{11}+b,\ldots,X_{1n} + b) = T(X_{11},\ldots,X_{1n}) + b$.

THEOREM 3.2. Let $P = P(\cdot|\vec{\theta}_0)$. Assume that for $\alpha > 0$,

$$\limsup_{\alpha\to 0}\ \limsup_{n\to\infty} (\alpha^2 n)^{-1}\log P\{T_1 \ge \alpha\} = K_1,$$

$$\limsup_{\alpha\to 0}\ \limsup_{n\to\infty} (\alpha^2 n)^{-1}\log P\{T_1 \le -\alpha\} = K_2,$$

$-\infty < K_1 + K_2 < 0$, and T is translation equivariant. Then

$$\limsup_{\Delta\to 0}\ \limsup_{n\to\infty} (\Delta^2 n)^{-1}\log P\{T_1 > T_2\}$$
$$= \limsup_{\Delta\to 0}\ \limsup_{n\to\infty} (\Delta^2 n)^{-1}\log P\{T_1 \ge T_2\} = K_1K_2/(K_1 + K_2).$$

PROOF. Define $H\{A\} = \limsup_{\Delta\to 0}\ \limsup_{n\to\infty} (\Delta^2 n)^{-1}\log P\{A\}$, for any event A. It is sufficient to show

(3.2) $$H\{T_1 > T_2\} \ge K_1K_2/(K_1 + K_2)$$

and

(3.3) $$H\{T_1 \ge T_2\} \le K_1K_2/(K_1 + K_2).$$

Let $S = \{(t_1,t_2): t_1 \geq t_2\}$, and let S^0 denote the interior of S. Let $\beta < 0$ be arbitrary, and define

$$W_\beta = \{(t_1,t_2): t_1 \geq (K_2+\beta)\Delta/(K_1+K_2),\ t_2 \leq (K_2-\beta)\Delta/(K_1+K_2)\} \subset S^0.$$

Then
$$\begin{aligned} H\{T_1 > T_2\} &\geq H\{(T_1,T_2) \in W_\beta\} \\ &= H\{T_1 \geq (K_2+\beta)\Delta/(K_1+K_2)\} + H\{T_2 \leq (K_2-\beta)\Delta/(K_1+K_2)\} \\ &= H\{T_1 \geq (K_2+\beta)\Delta/(K_1+K_2)\} + H\{T_2-\Delta \leq -(K_1+\beta)\Delta/(K_1+K_2)\} \\ &= [(K_2+\beta)/(K_1+K_2)]^2\cdot K_1 + [(K_1+\beta)/(K_1+K_2)]^2\cdot K_2 \\ &= K_1K_2/(K_1+K_2) + o(1) \text{ as } \beta \to 0. \end{aligned}$$

Since β is arbitrary, (3.2) is proved.

To prove (3.3), define $S_1 = S \cap \{(t_1,t_2): t_1 < 0\}$, $S_2 = S \cap \{(t_1,t_2): t_2 > \Delta\}$, and $S_3 = S - S_1 - S_2$. Since $S = S_1 \cup S_2 \cup S_3$, it follows from Lemma 2.1 that

$$\limsup_{n\to\infty} n^{-1} \log P\{T_1 \geq T_2\} = \max_i \limsup_{n\to\infty} n^{-1}\log P\{(T_1,T_2) \in S_i\}.$$

Therefore, it is sufficient to show

$$(3.4) \qquad H\{(T_1,T_2) \in S_i\} \leq K_1K_2/(K_1+K_2) \text{ for } i = 1,2,3.$$

For S_1, $H\{(T_1,T_2) \in S_1\} \leq H\{T_2 \leq 0\} = H\{T_1 \leq -\Delta\} = K_2 \leq K_1K_2/(K_1+K_2)$. Similarly, $H\{(T_1,T_2) \in S_2\} \leq H\{T_1 \geq \Delta\} = K_1 \leq K_1K_2/(K_1+K_2)$.

It remains to prove (3.4) for $i = 3$. Let M be an arbitrary positive integer greater than 1. Consider the M quadrants $V_1,\ldots,V_M$, where $V_j = \{(t_1,t_2): t_1 \geq (j-1)\Delta/M,\ t_2 \leq j\Delta/M\}$. Then $S_3 = \bigcup_{j=1}^{M} (V_j \cap S_3) \subset \bigcup_{j=1}^{M} V_j$. So $H\{(T_1,T_2) \in S_3\} \leq \max_j H\{(T_1,T_2) \in V_j\}$. Therefore, to prove (3.4) for $i = 3$, it is enough to show that for every $\varepsilon > 0$, there exists a positive integer $M = M(\varepsilon)$, such that

(3.5) $$H\{(T_1,T_2) \in V_j\} \le K_1K_2/(K_1+K_2) + \varepsilon \quad \text{for } j = 1,\ldots,M.$$

Let $\varepsilon > 0$ be arbitrary. For M sufficiently large,

$$H\{(T_1,T_2) \in V_1\} \le H\{T_2 \le \Delta/M\} = H\{T_1 \le (1-M)\Delta/M\}$$
$$= [(1-M)/M]^2 \cdot K_2 \le K_2 + \varepsilon \le K_1K_2/(K_1+K_2) + \varepsilon.$$

A similar result holds for V_M. For $j = 2,\ldots,M-1$,

$$H\{(T_1,T_2) \in V_j\} = H\{T_1 \ge (j-1)\Delta/M\} + H\{T_2 - \Delta \le (j-M)\Delta/M\} =$$

(3.6) $$[(j-1)/M]^2\, K_1 + [(j-M)/M]^2\, K_2.$$

Since $\max\limits_{0<x<1} \{(x-1/M)^2\, K_1 + (x-1)^2\, K_2\}$ occurs for $x = (K_1/M + K_2)/(K_1 + K_2)$, (3.6) is less than or equal to $[K_2(1-1/M)/(K_1+K_2)]^2 \cdot K_1 + [K_1(1/M-1)/(K_1+K_2)]^2 \cdot K_2$ which goes to $K_1K_2/(K_1+K_2)$ as $M \to \infty$. Therefore, for M sufficiently large, $H\{(T_1,T_2) \in V_j\} \le K_1K_2/(K_1+K_2) + \varepsilon$ for $j = 2,\ldots,M-1$. This completes the proof of (3.5) and of the theorem.

Assume now that

(3.7) $$I(0,\alpha) = I_0\alpha^2/2 + o(\alpha^2) \text{ as } \alpha \to 0 \text{ for some constant } I_0 > 0.$$

This implies that $I(\alpha,0) = I(0, -\alpha) = I_0\, \alpha^2/2 + o(\alpha^2)$ and so

(3.8) $$J(0,\alpha) = I_0\alpha^2 + o(\alpha^2) \text{ as } \alpha \to 0.$$

It follows from the proof of Theorem 6.1 of Bahadur (1971) that if T is a consistent estimator, then for $\alpha > 0$ and $P = P(\cdot|\vec{\theta}_0)$,

(3.9) $$\liminf_{\alpha\to 0}\ \liminf_{n\to\infty}\ (\alpha^2 n)^{-1}\log P\{T_1 \ge \alpha\} \ge -I_0/2$$

and

(3.10) $$\liminf_{\alpha\to 0}\ \liminf_{n\to\infty}\ (\alpha^2 n)^{-1}\log P\{T_1 \le -\alpha\} \ge -I_0/2.$$

This implies that

(3.11) $$\min\{K_1, K_2\} \geq -I_0/2,$$

where K_1 and K_2 are defined in Theorem 3.2.

In view of (3.9) and (3.10), define the Bahadur (asymptotic) efficiency of a consistent estimator T as

(3.12) $$e_0(T) = \liminf_{\alpha\to 0} \liminf_{n\to\infty} [-2/(I_0\alpha^2 n)]\log P\{|T_1| \geq \alpha\}.$$

We then have the following theorem.

THEOREM 3.3. Assume that T is a consistent estimator and that (3.7) holds. Let R be the selection procedure based on T. Then under the assumptions of Theorem 3.2, the following five statements hold:

(i) $e_0(T) = -2\max\{K_1, K_2\}/I_0$.

(ii) $e(R) = -4/I_0[K_1K_2/(K_1+K_2)]$.

(iii) $0 \leq e_0(T) \leq e(R) \leq 1$.

(iv) $e_0(T) = e(R)$ iff $K_1 = K_2$ or $K_1K_2 = 0$.

(v) $e_0(T) = 1$ iff $e(R) = 1$ iff $K_1 = K_2 = -I_0/2$.

PROOF. (i) follows from (3.12) and Lemma 2.1. Since $P\{T_1 > T_2\} \leq P(E|\vec{\theta}_0, R, n) \leq P\{T_1 \geq T_2\}$, (ii) follows from (3.1) and (3.8). (iii) and (iv) are proved easily from (i) and (ii). It remains to prove (v). Clearly, $K_1 = K_2 = -I_0/2$ implies $e_0(T) = 1$, which implies $e(R) = 1$. So it is enough to show that if $e(R) = 1$, then $K_1 = K_2 = -I_0/2$. Let $K_1 = -aI_0$ and $K_2 = -bI_0$. By (ii) and (3.11), $1/2 \geq a, b > 0$. Then $1 = e(R) = 4ab/(a+b)$, which implies that $a = b(4a-1) \leq b$ and $b = a(4b-1) \leq a$. Therefore, $a = b = 1/2$. This completes the proof.

4. Scale parameter case. Assume $g(x|\theta_i) = (1/\theta_i)f(x/\theta_i)$, $\theta_i > 0$, $i = 1,\ldots,k$, and f is known. Let $\vec{\theta}_0$ be fixed and satisfy (1.3) and (1.4), with $\Delta > 1$.

THEOREM 4.1. For any eventually unbiased selection procedure R, $\bar{c}(\vec{\theta}_0,R) \le J(1,\Delta^{\frac{1}{2}})$.

The proof is analagous to that of Theorem 3.1 with $\vec{\theta}_1 = (\Delta^{\frac{1}{2}},\theta_2,\ldots,\theta_{k-1},\Delta^{\frac{1}{2}})$.

If $J(1,\Delta^{\frac{1}{2}})$ is finite, define the (asymptotic) efficiency at $\vec{\theta}_0$ of a scale parameter selection procedure R as

$$e(\vec{\theta}_0,R) = \liminf_{\Delta\to 1} \underline{c}(\vec{\theta}_0,R)/J(1,\Delta^{\frac{1}{2}}). \tag{4.1}$$

Then $0 \le e(\vec{\theta}_0,R) \le 1$ for all R and $\vec{\theta}_0$. If $e(\vec{\theta}_0,R)$ does not depend on the values of $\theta_2,\ldots,\theta_{k-1}$, then the efficiency will be written as $e(R)$.

Now consider selection procedures based on an estimator $T = T_n$. As in the location parameter case, it is sufficient to consider the case of two populations. So let $\vec{\theta}_0 = (1,\Delta)$, and let $T_1 = T_{1n}$ and $T_2 = T_{2n}$ be defined by $T_i = T(X_{i1},\ldots,X_{in})$, $i = 1,2$.

In the following theorem, T is assumed to be scale equivariant, i.e., if b is a positive constant, then $T(bX_{11},\ldots,bX_{1n}) = b\cdot T(X_{11},\ldots,X_{1n})$.

THEOREM 4.2. Let $P = P(\cdot|\vec{\theta}_0)$. Assume that for $\alpha > 0$,

$$\limsup_{\alpha\to 0}\ \limsup_{n\to\infty}\ (\alpha^2 n)^{-1}\log P\{T_1 - 1 \ge \alpha\} = K_1,$$

$$\limsup_{\alpha\to 0}\ \limsup_{n\to\infty}\ (\alpha^2 n)^{-1}\log P\{T_1 - 1 \le -\alpha\} = K_2,$$

$-\infty < K_1 + K_2 < 0$, and T is scale equivariant. Then

$$\limsup_{\Delta\to 1}\ \limsup_{n\to\infty}\ [(\Delta-1)^2 n]^{-1}\log P\{T_1 > T_2\}$$
$$= \limsup_{\Delta\to 1}\ \limsup_{n\to\infty}\ [(\Delta-1)^2 n]^{-1}\log P\{T_1 \ge T_2\} = K_1K_2/(K_1+K_2).$$

The proof is analagous to that of Theorem 3.2.

Assume now that

(4.2) $\qquad I(1,\alpha) = I_0(\alpha-1)^2/2 + o(\alpha-1)^2$ as $\alpha \to 1$, for some constant $I_0 > 0$. Then $I(\alpha,1) = I(1,1/\alpha) = I_0(\alpha-1)^2/2 + o(\alpha-1)^2$ as $\alpha \to 1$. So

(4.3) $$J(1,\alpha) = I_0(\alpha-1)^2 + o(\alpha-1)^2 \text{ as } \alpha \to 1.$$

It follows from the results of Bahadur (1971) that if T is a consistent estimator, then for $\alpha > 0$ and $P = P(\cdot|\vec{\theta}_0)$,

(4.4) $$\liminf_{\alpha\to 0}\ \liminf_{n\to\infty}\ (\alpha^2 n)^{-1} \log P\{T_1-1 \geq \alpha\} \geq -I_0/2$$

and

(4.5) $$\liminf_{\alpha\to 0}\ \liminf_{n\to\infty}\ (\alpha^2 n)^{-1} \log P\{T_1-1 \leq -\alpha\} \geq -I_0/2.$$

This implies that

(4.6) $$\min\{K_1,K_2\} \geq -I_0/2,$$

where K_1 and K_2 are defined in Theorem 4.2.

Define the Bahadur (asymptotic) efficiency of a consistent estimator T as

$$e_0(T) = \liminf_{\alpha\to 0}\ \liminf_{n\to\infty}\ [-2/(I_0\alpha^2 n)]\log P\{|T_1-1| \geq \alpha\}.$$

We then have the following theorem.

THEOREM 4.3. Assume that T is a consistent estimator and that (4.2) holds. Let R be the selection procedure based on T. Then under the assumptions of Theorem 4.2, statements (i)-(v) of Theorem 3.3 hold.

The proof is analagous to that of Theorem 3.3.

In the special case where the support of f is the set of positive real numbers, all of the theorems of this section follow as corollaries to the theorems of Section 3, since, in this case, the scale parameter case reduces to the location parameter case by taking logarithms.

5. Examples. In this section we will consider three selection procedures R_1, R_2, and R_3, where R_1 is the procedure based on the sample means, R_2 is the procedure based on the sample medians, and R_3 is a nonparametric procedure proposed by Bechhofer and Sobel (1958). R_3 is defined as follows: For $i = 1,\ldots,k$, let $X_{i1},\ldots,X_{in}$ be a sample from π_i. For $i = 1,\ldots,k$; $j = 1,\ldots,n$, define

$$(5.1) \qquad T_{ij} = \begin{cases} 1 & \text{if } X_{ij} = \max\{X_{1j},\ldots,X_{kj}\}, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$(5.2) \qquad T_i = n^{-1} \sum_{j=1}^{n} T_{ij}.$$

Then R_3 selects π_i if $T_i = \max_j \{T_j\}$. As usual, in case of ties, one of the tied populations is selected at random. R_3 was considered by Dudewicz (1971) in this context.

To compute $c(\vec{\theta}_0,R_i)$ for $i = 1, 2$, it is sufficient to consider the case of two populations. However, $c(\vec{\theta}_0, R_3)$ depends on k.

Procedure R_1; Location Parameter Case

Let $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_n$ be samples from $f(x)$ and $f(x-\Delta)$, where $\Delta > 0$. So $\vec{\theta}_0 = (0,\Delta)$. Let $\bar{X}$ and $\bar{Y}$ be the sample means. Then

$$P(\bar{X} \geq \bar{Y}) = P\{n^{-1} \sum_{i=1}^{n} [X_i - (Y_i - \Delta)] \geq \Delta\} = P\{\bar{Z} \geq \Delta\},$$

where $\bar{Z}$ is the mean of n independent random variables, each with the same distribution as $X_1 - X_2$. Let $m(t) = E(e^{tX_1})$. Then by Chernoff's Theorem (1952), if τ and ρ satisfy $m'(\tau)/m(\tau) - m'(-\tau)/m(-\tau) = \Delta$ and $\rho = \exp(-\Delta\tau)m(\tau)m(-\tau)$, then $c(\vec{\theta}_0,R_1) = -\lim_{n\to\infty} n^{-1}\log P\{\bar{X} \geq \bar{Y}\} = -\log \rho$.

In the special case where f is symmetric about 0, $m(t) = m(-t)$ for all t. So if τ and ρ satisfy $m'(\tau)/m(\tau) = \Delta/2$ and $\rho = \exp(-\Delta\tau)m^2(\tau)$, then $c(\vec{\theta}_0, R_1) = -\log \rho$.

For normal populations $(f(x) = (2\pi)^{-\frac{1}{2}}\exp(-x^2/2))$, it can be shown that $c(\vec{\theta}_0, R_1) = J(0,\Delta/2) = \Delta^2/4$. (This result for c was obtained by Dudewicz (1969)). So equality holds in Theorem 3.1 for all Δ, and, by (3.1), $e(R_1) = 1$. The case where the normal populations have common variance $\sigma^2 \neq 1$ reduces to the standard normal case if $\vec{\theta}_0$ is taken to be $(0,\Delta\sigma)$.

For double exponential populations $(f(x) = \exp(-|x|)/2)$, it can be shown that

$$c(\vec{\theta}_0,R_1) = (4+\Delta^2)^{\frac{1}{2}} - 2 - 2\log[((4+\Delta^2)^{\frac{1}{2}} + 2)/4]$$

and $e(R_1) = 1/2$.

Procedure R_1; Scale Parameter Case

Let $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_n$ be samples from $f(x)$ and $(1/\Delta)f(x/\Delta)$, where $\Delta > 1$. So $\vec{\theta}_0 = (1,\Delta)$. Then $P(\bar{X} \geq \bar{Y}) = P(\bar{Z} \geq 0)$, where $\bar{Z}$ is the mean of $Z_1,\ldots,Z_n$, with $Z_i = X_i - Y_i$ for $i = 1,\ldots,n$. Let $m_1(t) = E(e^{tX_1})$ and $m_2(t) = E(e^{tY_1})$. So by Chernoff's Theorem (1952), if τ and ρ satisfy $m_1'(\tau)/m_1(\tau) - m_2'(-\tau)/m_2(-\tau) = 0$ and $\rho = m_1(\tau)m_2(-\tau)$, then $c(\vec{\theta}_0,R_1) = -\log \rho$.

For exponential populations $(f(x) = e^{-x}, x > 0)$, it can be shown that $c(\vec{\theta}_0,R_1) = \log[(\Delta+1)^2/(4\Delta)]$ and $e(R_1) = 1$.

Procedure R_2; Location or Scale Parameter Case

Let $X_1,\ldots,X_n$ be a sample from a continuous distribution with cdf F_0, density f_0, and median u_0. Let $Y_1,\ldots,Y_n$ be a sample from a continuous distribution with cdf G_0, density g_0, and median v_0. Assume that G_0 is stochastically larger than F_0. So $u_0 < v_0$.

This situation includes the location parameter case with $G_0(x) = F_0(x-\Delta)$, $\Delta > 0$, and $\vec{\theta}_0 = (0,\Delta)$. It also includes the scale parameter case with $G_0(x) = F_0(x/\Delta)$, $\Delta > 1$, and $\vec{\theta}_0 = (1,\Delta)$, provided G_0 is stochastically larger than F_0.

Let F_n and G_n be the empirical cdf's of the X-sample and the Y-sample, respectively, and let T_1 and T_2 be the respective sample medians.

For any cdf's F and G, define $T(F,G) = \int \operatorname{sgn}(F+G-1)dF$. Then

$$P(T_1 > T_2) = P(T_1 \geq T_2) = P\{T(F_n,G_n) \geq 0\}. \tag{5.3}$$

By Theorem 1 of Hoadley (1967),

$$\begin{aligned} c(\vec{\theta}_0,R_2) &= -\lim_{n\to\infty} n^{-1}\log P\{T(F_n,G_n) \geq 0\} \\ &= \inf_{(F,G)\in\Omega} \left[\int f \log(f/f_0)dx + \int g \log(g/g_0)dx\right] \end{aligned} \tag{5.4}$$

where $\Omega = \{(F,G): \int \operatorname{sgn}(F+G-1)dF \geq 0;\ F,\ G \text{ absolutely continuous}\}$

$= \{(F,G): \text{med } F \geq \text{med } G;\ F,\ G \text{ absolutely continuous}\}$,

and where f and g are the densities of F and G, respectively.

By Theorem 6 of Sanov (1957) (or Lemma 3.2 of Hoadley (1967)),

$$\inf_{f:\ \text{med } F=u} \int f \log(f/f_0)dx = -(1/2)\log[4F_0(u)(1-F_0(u))] = H(u),$$

say. Similarly,

$$\inf_{g:\ \text{med } G=v} \int g \log(g/g_0)dx = -(1/2)\log[4G_0(v)(1-G_0(v))] = K(v),$$

say. Therefore, by (5.4), $c(\vec{\theta}_0,R_2) = \inf_{u\geq v} [H(u) + K(v)]$. It is easy to show that this infimum occurs for $u = v$ and $u_0 \leq u \leq v_0$. So

$c(\vec{\theta}_0,R_2) = \inf_{u_0\leq u\leq v_0} [H(u) + K(u)]$, or

$$c(\vec{\theta}_0,R_2) = -(1/2) \sup_{u_0\leq u\leq v_0} \log[16\, F_0(u)(1-F_0(u))G_0(u)(1-G_0(u))]. \tag{5.5}$$

Assume now that the following four conditions hold:

(1) $G_0(x) = F_0(x-\Delta)$; F_0 has density f_0.

(2) f_0 is symmetric about 0.

(3) f_0 is unimodal.

(4) f_0/F_0 is decreasing (i.e. nonincreasing) over $\{x: x < 0, F_0(x) > 0\}$.

Then $u_0 = 0$ and $v_0 = \Delta$. It is shown in Gaynor (1976) that conditions (1)-(4) imply that the supremum in (5.5) is attained when $u = \Delta/2$. Conditions (2)-(4) are satisfied for standard normal, double exponential, logistic ($f(x) = e^{-x}(1+e^{-x})^{-2}$), and uniform (over [-1/2, 1/2]) distributions. So in the location parameter case, we have from (5.5) that

$$c(\vec{\theta}_0,R_2) = \begin{cases} -\log[4\Phi(\Delta/2)(1-\Phi(\Delta/2))] & \text{for normal populations,} \\ -\log[2\exp(-\Delta/2)-\exp(-\Delta)] & \text{for double exponential populations,} \\ \log[(1+\exp(-\Delta/2))^2\exp(\Delta/2)/4] & \text{for logistic populations,} \\ -\log(1-\Delta^2),\ \Delta < 1, & \text{for uniform populations.} \end{cases}$$

It can be shown that in the normal population case, $e(R_2) = 2/\pi$, and that in the double exponential case, $e(R_2) = 1$.

Note that for normal populations with unit variance, the efficiency of the medians procedure relative to the means procedure is

$$c(\vec{\theta}_0,R_2)/c(\vec{\theta}_0,R_1) = (-4/\Delta^2)\log[4\Phi(\Delta/2)(1-\Phi(\Delta/2))].$$

This is the same as the ARE (as defined by Hodges and Lehmann (1956, equation (3.3))) of the sign test relative to the t test for testing the hypothesis that the mean of a normal population is zero against the alternative that the mean is $\Delta/2$.

It can be shown that a sufficient condition for conditions (3) and (4) is that f_0'/f_0 is decreasing, i.e., that f_0 is strongly unimodal.

Procedure R_3; Location or Scale Parameter Case

For $1 \le i \le k$, $1 \le j \le n$, let T_{ij} and T_i be defined by (5.1) and (5.2). Let $p = P\{T_{k1} = 1\}$, and let $q_i = P\{T_{i1} = 1\}$, $i = 1,\ldots,k-1$. Assume that $q_i = q$, $i = 1,\ldots,k-1$, and that $p > q$. Then $q = (1-p)/(k-1)$. This situation applies to the location parameter case for $\vec{\theta}_0 = (0,\ldots,0,\Delta)$, $\Delta > 0$. It also applies to the scale parameter case for $\vec{\theta}_0 = (1,\ldots,1,\Delta)$, $\Delta > 1$, provided $p > q$. For the remainder of this section, $\vec{\theta}_0$ will be restricted to these values.

If we take into account the possibility of ties, it follows from (2.1) that

$$-\limsup_{n\to\infty} n^{-1}\log P\{T_1 \ge \max_{j\ne1} T_j\} \le \underline{c}(\vec{\theta}_0,R_3) \le \bar{c}(\vec{\theta}_0,R_3) \tag{5.6}$$
$$\le -\liminf_{n\to\infty} n^{-1} \log P\{T_1 > \max_{j\ne1} T_j\}.$$

Define $M = \{\vec{x} = (x_1,\ldots,x_k): \sum_{i=1}^{k} x_i = 1,\ x_i \ge 0, \text{ all } i\}$. For $\vec{x}$, $\vec{p} \in M$, define $I(\vec{x},\vec{p}) = \sum_{i=1}^{k} x_i \log(x_i/p_i)$, where $x_i \log(x_i/p_i) = 0$ if $x_i = 0$. For $A \subset M$, define $I(A,\vec{p}) = \inf\{I(\vec{x},\vec{p}): \vec{x} \in A\}$. Let $A^{(n)} = \{\vec{x} \in A: nx_i \text{ is an integer, for } i = 1,\ldots,k\}$. Finally, define $A_1 = \{\vec{x} \in M: x_1 > x_j,\ j = 2,\ldots,k\}$ and $A_2 = \{\vec{x} \in M: x_1 \ge x_j,\ j = 2,\ldots,k\}$.

Since $(T_1,\ldots,T_k)$ has a multinomial distribution with parameter $\vec{p}_0 = (q,\ldots,q,p)$, it follows from Theorem 2.1 of Hoeffding (1965) that

$$\lim_{n\to\infty} n^{-1}\log P\{T_1 > \max_{j\ne1} T_j\} = -\lim_{n\to\infty} I(A_1^{(n)},\vec{p}_0)$$

and

$$\lim_{n\to\infty} n^{-1}\log P\{T_1 \ge \max_{j\ne1} T_j\} = -\lim_{n\to\infty} I(A_2^{(n)},\vec{p}_0).$$

Since $\lim_{n\to\infty} I(A_1^{(n)},\vec{p}_0) = \lim_{n\to\infty} I(A_2^{(n)},\vec{p}_0) = I(A_2,\vec{p}_0)$, then by (5.6),

$$c(\vec{\theta}_0, R_3) = I(A_2, \vec{p}_0). \tag{5.7}$$

It is shown in Gaynor (1976) that $I(A_2,\vec{p}_0) = I(\vec{x},\vec{p}_0)$, where for $k = 2$, $x_1 = x_2 = 1/2$, and for $k > 2$,

$$x_1 = x_k = [2 + (k-2)(q/p)^{\frac{1}{2}}]^{-1} = a, \tag{5.8}$$

say, and $x_i = (1-2a)/(k-2)$ for $i = 2,\ldots,k-1$. So by (5.7),

$$c(\vec{\theta}_0,R_3) = \begin{cases} -\log[2(p(1-p))^{\frac{1}{2}}] & \text{if } k = 2, \\ a \log[a^2/(pq)]+(1-2a) \log[(1-2a)/((k-2)q)] & \text{if } k > 2, \end{cases} \tag{5.9}$$

where a is defined by (5.8). This result for $k = 2$ was obtained by Dudewicz (1971) by a different method.

Dudewicz (1971) conjectured that $c(\vec{\theta}_0,R_3) \to 0$ as $k \to \infty$. This can be verified with (5.9).

Acknowledgement. The author would like to thank Professor Kumar Jogdeo and Professor Robert A. Wijsman for their advice concerning this paper.

References

[1] Bahadur, R. R. (1971). Some Limit Theorems in Statistics. SIAM, Philadelphia, PA.

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

[3] Bechhofer, R. E. and Sobel, M. (1958). Nonparametric multiple-decision procedures for selecting that one of k populations which has the highest probability of yielding the largest observation (preliminary report). Abstract, Ann. Math. Statist., 29, 325.

[4] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Statist., 23, 493-507.

[5] Dudewicz, Edward J. (1969). An approximation to the sample size in selection problems. Ann. Math. Statist., 40, 492-497.

[6] Dudewicz, Edward J. (1971). A nonparametric selection procedure's efficiency: largest location parameter case. J. Amer. Statist. Assoc., 66, 152-161.

[7] Dudewicz, Edward J. and Zaino, Nicholas A., Jr., Sample size for selection, in Statistical Decision Theory and Related Topics, edited by Shanti S. Gupta and James Yackel, Academic Press, 1971.

[8] Gaynor, John A. (1976). Asymptotic efficiencies of selection procedures. Ph.D. Thesis, Univ. of Illinois.

[9] Hoadley, A. B. (1967). On the probability of large deviations of functions of several empirical cdf's. Ann. Math. Statist., 38, 360-381.

[10] Hodges, J. L., Jr. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t test. Ann. Math. Statist., 27, 324-335.

[11] Hoeffding, Wassily. (1965). Asymptotically optimal tests for multi nomial distributions. Ann. Math. Statist., 36, 369-401.

[12] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. Ann. Math. Statist., 22, 79-86.

[13] Sanov, I. N. (1957). On the probability of large deviations of random variables. Sel. Transl. Math. Statist. Prob., 1, 213-244.