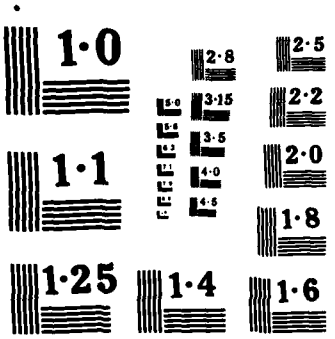
AD-A162 824 OPTIMUM TWO-STAGE SELECTION PROCEDURES FOR WEIBULL POPULATIONS(U) PURDUE UNIV LAFAYETTE IN DEPT OF STATISTICS S S GUPTA ET AL. OCT 85 TR-85-30 1/1

UNCLASSIFIED N00014-84-C-0167 F/G 12/1 NL

END

FILMED

DTIC



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

Optimum Two-Stage Selection Procedures for Weibull Populations

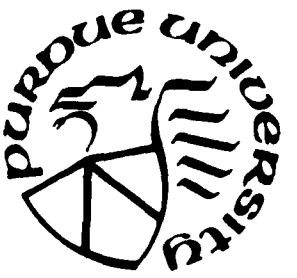
by

Shanti S. Gupta[1]
Purdue University

Klaus J. Miescke[2]
University of Illinois at Chicago

Technical Report#85-30

PURDUE UNIVERSITY



DEPARTMENT OF STATISTICS

Optimum Two-Stage Selection Procedures for Weibull Populations

by

Shanti S. Gupta[1]
Purdue University

Klaus J. Miescke[2]
University of Illinois at Chicago

Technical Report#85-30

Department of Statistics
Purdue University

October 1985

AMS (1980) Subject Classification: Primary 62F07; Secondary 62N05, 62L99

Key Words and Phrases: Selection in two stages; Finding the most reliable component; Optimum screening and selection; the best Weibull population.

[1] Research supported by the Office of Naval Research Contract N00014-84-C-0167 at Purdue University.

[2] Research supported by the Air Force Office of Scientific Research Contract AFOSR-85-0347 at the University of Illinois at Chicago.

(-A-)

Optimum Two-Stage Selection Procedures for Weibull Populations

by

Shanti S. Gupta[1]
Purdue University

Klaus J. Miescke[2]
University of Illinois at Chicago

Abstract

Let $\pi_1, \ldots, \pi_k$ be Weibull populations with a common known shape parameter, and with unknown scale parameters. The goal is to find the population with the largest scale parameter. From each population, Type II-censored observations are available at two stages, where censoring at stage 1 (2) occurs at the q-th (r-th) failure. Two-stage procedures with screening at the first stage are considered which are optimum permutation invariant in terms of the risk with respect to a large class of loss functions. For the procedure with a fixed subset size at stage 1, the least favorable parameter configuration under the indifference zone approach is of the slippage type, which makes it feasible to control the infimum of the probability of a correct selection. Some extensions of the results are discussed at the end.

[1] Research supported by the Office of Naval Research Contract N00014-84-C-0167 at Purdue University.

[2] Research supported by the Air Force Office of Scientific Research Contract AFOSR-85-0347 at the University of Illinois at Chicago.

1. Introduction.

Suppose we wish to find the most reliable of k types of components $\pi_1, \ldots, \pi_k$, say. Let the failure times follow a Weibull model with density w.r.t. the Lebesgue measure, confined to the positive real line,

$$f(x|\theta_i) = \alpha\theta_i x^{\alpha-1} exp(-\theta_i x^\alpha), \qquad x > 0, \tag{1}$$

where $\alpha > 0$ is fixed known, and $\theta_i^{-1/\alpha} > 0$ is the unknown scale parameter associated with π_i, $i = 1, \ldots, k$. Thus our goal becomes to find that one π_i which is associated with the smallest θ_i, $i = 1, \ldots, k$. We may assume that it is unique, to keep the presentation of our material simple. This is not a serious restriction, since in case of ties, we may be content with the selection of any of the most reliable components, and our results hold in this case analogously.

We shall consider 2-stage selection procedures, with screening at the first stage, as discussed in Miescke (1984), which are based on Type II-censored failure times observed at both stages. In the following, all failure times considered are assumed to be independent. To describe the sampling process from the k populations $\pi_1, \ldots, \pi_k$, it is sufficient to do so for one particular population π_i, say.

At stage 1, $n_i > 0$ components of type π_i are tested simultaneously until the q-th failure occurs. If this type of component is not screened out at stage 1, $m_i \geq 0$ components are added, or $-m_i \geq 0$ components are withdrawn, and then the $n_i - q + m_i$ components are further tested simultaneously until the r-th failure at this stage occurs. Hereby we assume, by obvious reasons, that $q \geq 1$, $r \geq 1$, $n_i \geq q$, and $n_i + m_i \geq q + r$. It should be pointed out that both q and r do not depend on i. We shall see later that this is very crucial to gain permutation invariance in our decision problem, which in turn provides the basis for finding optimum permutation invariant decision procedures.

The next step is a reduction by sufficiency. Because of the assumed independence of all failure times, this can be done as well for one particular population, π_i, say. Let $U_{i,1}, U_{i,2}, \ldots, U_{i,q}$ denote the ordered failure times to failure q of π_i at stage 1. Likewise, let $V_{i,1}, V_{i,2}, \ldots, V_{i,r}$ denote the ordered failure times to failure r of π_i at stage 2, measured from $U_{i,q}$ onwards. Let

$$\begin{aligned} U_i &= \sum_{j=1}^{q} U_{i,j}^\alpha + (n_i - q)U_{i,q}^\alpha \\ V_i &= \sum_{j=1}^{r} V_{i,j}^\alpha + (n_i + m_i - q - r)V_{i,r}^\alpha \ , \text{ and} \\ T_i &= U_i + V_i \ . \end{aligned} \tag{2}$$

Then the distributional properties of U_i, V_i, and T_i can be summarized as follows.

Lemma 1. *For every population π_i, the following holds.*

(a) $U_i(V_i)$ is sufficient for θ_i at stage 1 (2), and T_i is sufficient for θ_i at both stages combined.

(b) U_i and V_i are independent.

(c) $2\theta_i U_i(2\theta_i V_i, 2\theta_i T_i)$ is chi-squared distributed with $2q(2r, 2q+2r)$ degrees of freedom

Proof: Most of these facts are well known. Thus we outline the proof only briefly. The statement concerning U_i in (a) follows from looking at the likelihood function at stage 1, as it is done in Tsokos and Rao (1979). The statement concerning U_i in (c) is proved in Gnedenko et al. (1969), sec. 3.3, for the case of $\alpha = 1$, and it can be extended to the case of any other value of α immediately. Finally, by considering, instead of the original failure times, the α-th powers of the same, which are exponentially distributed with scale parameter θ_i^{-1}, the proof can be completed by using the lack of memory of the exponential law or, more precisely, the strong Markov property inherent in the sampling process described in terms of these exponential random variables, as it is discussed in Feller (1971), section I.6.

Returning to the joint consideration of our k populations, all facts stated so far can be carried over in a natural way, since (U_i, V_i, T_i), $i = 1, \ldots, k$, are independent random vectors. For notational convenience, let in the following $\underline{U} = (U_1, \ldots, U_k)$, $\underline{V} = (V_1, \ldots, V_k)$, and $\underline{T} = (T_1, \ldots, T_k)$. Next we introduce a class of 2-stage selection procedures for the given decision problem.

Definition. A 2-stage selection procedure acts as follows. After all observations at stage 1 have been made, a non-empty subset $s \subseteq \{1, \ldots, k\}$ is selected. All population π_i with $i \notin s$ are discarded. If $s = \{j\}$, say, the final decision "π_j is the most reliable type" is made. If s contains more than one element, stage 2 is entered. The sampling process is continued for all π_i with $i \in s$ as described before, and then for some $j \in s$ the final decision "π_j is the most reliable type" is made.

Let $L(\underline{\theta}, (s, i))$ be a real-valued loss which occurs at $\underline{\theta} = (\theta_1, \ldots, \theta_k)$, if s is selected at stage 1, and the final decision is in favor of π_i, $i \in s \subseteq \{1, \ldots, k\}$. We assume that it is integrable such that the associated risk function exists. Moreover, let it be permutation invariant as defined in Gupta and Miescke (1984), and let it favor the selection of more reliable components in the following way:

$$L(\underline{\theta}, (s, i)) \leq L(\underline{\theta}, (\tilde{s}, j)) \text{ if} \tag{3}$$

(a) $i, j \in s = \tilde{s} \quad \theta_i \leq \theta_j$, or

(b) $s \backslash \{i\} = \tilde{s} \backslash \{j\}, \quad \theta_i \leq \theta_j$, or

(c) $i = j, \quad s \backslash \{u\} = \tilde{s} \backslash \{v\}, \quad \theta_u \leq \theta_v$.

It can be seen that under such a loss function, the decision problem is invariant under the group of permutations. This justifies restricting our further considerations to 2-stage selection procedures which are permutation invariant. A rigorous definition of this class can be found in Gupta and Miescke (1984). The optimum rules within this class are, as we shall see later, of the following form.

Definition: Let $R:(0,\infty)^k \to \{1,2,\ldots,k\}$ be a symmetric, Borel-measurable function. Let us consider $R(\underline{U})$ as a decision rule to determine the size of the subset to be selected at stage 1. Then $P(R)$ is the 2-stage selection procedure which selects at stage 1 in terms of the $R(\underline{U})$-largest U_i's, and makes the final decision at stage 2 in terms of the largest observed T_i.

Optimality of $P(R)$ among all procedures which employ the same subset-size rule R will be shown in Section 2. Further properties of $P(R)$ will also be discussed there. Of special interest is the case of a constant R, $R = t$, say. This will be considered in Section 3 under the 0–1 loss function, which is zero if and only if the most reliable type of component is finally selected. The risk, or probability of an incorrect selection, respectively, of $P(t)$ will be shown to have a natural least favorable parameter configuration in the indifference zone approach of Bechhofer (1954), which makes it feasible to control the infimum of the probability of a correct selection, denoted by $P_{\underline{\theta}}(CS|P(t))$, on a certain subset of the parameter space. Finally, in Section 4, extensions of our results to p-stage selection procedures will be described, and some open questions for further research will be presented.

2. Optimality of the Procedures $P(R)$.

The first of our results establishes optimality of procedure $P(R)$ within the class of all permutation invariant 2-stage selection procedures which employ the same subset size rule R.

Theorem 1. *Let R be a subset-size rule, and let L be a loss function, with properties as described in Section 1. Then for every permutation invariant 2-stage selection procedure P which employs R at stage 1, we have*

$$\mathcal{R}(\underline{\theta}, P(R)) \leq \mathcal{R}(\underline{\theta}, P), \quad \underline{\theta} \in (0,\infty)^k, \tag{4}$$

where $\mathcal{R}(\underline{\theta}, P)$ denotes the risk, i.e. expected loss, of P at $\underline{\theta}$.

Proof: $\underline{U}$ and $\underline{V}$, respectively, can be considered as two random vectors which are generated through

$$U_i = \sum_{j=1}^{q} C_{i,j}, \quad V_i = \sum_{j=q+1}^{q+r} C_{i,j}, \quad i = 1,\ldots,k, \tag{5}$$

where the $C_{i,j}$'s are generic random variables which are mutually independent. For $i \in \{1,\ldots,k\}$ and $j \in \{1,\ldots,q+r\}$, $2\theta_i C_{i,j}$ follows a chi-squared distribution with 2 degrees of freedom, i.e. $C_{i,j}$ has the following density on the positive real line:

$$g(x|\theta_i) = |\tilde{\theta}_i| exp(\tilde{\theta}_i x), \quad x > 0, \quad \tilde{\theta}_i = -\theta_i. \tag{6}$$

It can be seen now that almost all assumptions which are made in Gupta and Miescke (1984) are fulfilled, where the underlying exponential family of densities is of the form

$$h(x|\theta) = c(\theta) exp(\theta x) d(x), \quad x \in \quad , \quad \theta \in \Omega \subseteq \quad . \tag{7}$$

If d were log-concave, i.e. if the exponential family were strongly unimodal, our proof would be completed by applying Corollary 2 of Gupta and Miescke (1984). However, the function d in (6) is the indicator function of the positive real line, which is obviously not log-concave.

A careful examination of the proof of the key Lemma 2 in Gupta and Miescke (1984) fortunately shows that if the family (7) has the positive real line as a common support, then all results in that paper remain valid if every n-fold convolution of d is log-concave on the positive real line. And since the n-fold convolution of the indicator function of the positive real line at $x > 0$ is equal to $x^{n-1}/(n-1)!$, which is indeed log-concave for $n > 1$, the proof of this theorem is completed.

The next result establishes uniqueness of the optimum procedure $P(R)$, and some consequences. Let $\mathcal{D}_I(R)$ denote the class of all permutation invariant 2-stage selection procedures which employ R at stage 1, and let $\mathcal{D}_I = \{P | P \in \mathcal{D}_I(R), R \text{ subset-size rule}\}$. For a moment, let us also consider the larger classes $\mathcal{D}(R)$ and $\mathcal{D}$, say, where the procedures, including their subset-size rules at stage 1, are not necessarily permutation invariant, i.e. symmetric for brevity.

Theorem 2. *Let L be any loss function with properties as described in Section 1. Then the following holds.*

(a) For every symmetric subset-size rule R, $P(R)$ is the unique optimum procedure in $\mathcal{D}_I(R)$ in the sense (4). Moreover, $P(R)$ is admissible in $\mathcal{D}(R)$.

(b) If there exists a minimax procedure in $\mathcal{D}_I$, which employs R_0, say, at stage 1, then $P(R_0)$ is minimax in $\mathcal{D}_I$, and both procedures are minimax in $\mathcal{D}$.

(c) The class $\{P(R) | R \text{ subset-size rule}\}$ is essentially complete in $\mathcal{D}_I$.

Proof. Let L be a loss function which has the assumed properties. Let R be a symmetric subset-size rule, and let $\underline{\theta} \in (0,\infty)^k$, where not all of the θ_i's are equal, be fixed. Then it can be shown, as in Gupta and Miescke (1984), that $P(R)$ is the unique Bayes rule in $\mathcal{D}(R)$ with respect to the symmetric prior which gives probability mass $1/k!$ to each of the $k!$ permutations of $\underline{\theta} = (\theta_1, \ldots, \theta_k)$. Since the Bayes risk is equal to $\mathcal{R}(\underline{\theta}, P(R))$, the first parts of (a) and (b), as well as (c), follow from Theorem 1. The second part of (a) holds since a unique Bayes rule is always admissible. The second part of (b) follows from Blackwell and Girshick (1954), sec. 8.6 and the fact that the group of permutations is finite. This completes the proof of the theorem.

The last result in this section confirms the intuitive conjecture that sampling of more information improves the optimum procedure. Let $P(R; \underline{n}, \underline{m}, q, r)$ be the procedure $P(R)$, which employes R at stage 1, where $q \geq 1$, $r \geq 1$, $\underline{n} = (n_1, \ldots, n_k)$, $n_i \geq q$, $\underline{m} = (m_1, \ldots, m_k)$, $m_i + n_i \geq q + r$, $i = 1, \ldots, k$, may now be variable. Then we can state the following.

Theorem 3. *Let L be any loss function, which does not depend on $\underline{n}$ and $\underline{m}$, with properties as described in Section 1. Then for every symmetric subset-size rule R, the risk of $P(R)$ at $\underline{\theta}$ does not depend on $\underline{n}$ and $\underline{m}$. Let it be denoted by $\rho(\underline{\theta}; R, q, r)$, say. At every $\underline{\theta} \in (0,\infty)^k$, where not all of the θ_i's are equal, it has the following properties.*

(8) $\rho(\underline{\theta}; R, q, r) > \rho(\underline{\theta}; R, q+1, r)$, if $n_i > q$, $n_i + m_i > q + r$, $i = 1, \ldots, k$,

(9) $\rho(\underline{\theta}; R, q, r) > \rho(\underline{\theta}; R, q, r+1)$, if $n_i + m_i > q + r$, $i = 1, \ldots, k$,

(10) $\rho(\underline{\theta}; R, q, r) > \rho(\underline{\theta}; R, q+1, r-1)$, if $n_i > q$, $i = 1, \ldots, k$.

Proof. Let L and R be given as stated in the theorem. $P(R)$ utilizes all of the relevant information contained in the observations through $\underline{U}$ and $\underline{T}$ at stage 1 and stage 2, respectively. In view of Lemma 1, it is seen that the joint distribution of $\underline{U}$ and $\underline{T}$ does not depend on $\underline{n}$ and $\underline{m}$. Hence the risk function of $P(R)$ has the same property.

Because of the similarity of arguments, we give only a proof for (10). First it is important to note that $P(R)$ remains to be the unique Bayes procedure with respect to any symmetric prior within the class $\mathcal{D}(R)$, if all procedures were included in $\mathcal{D}(R)$ which make use of the available observations of all of the π_i's at stage 2, but which still restrict final selections to those π_i's which have been selected at stage 1.

If now $P(R; \underline{n}, \underline{m}, q+1, r-1)$ is considered to be based on all available failure times up to the $(q+1)$-th and the $(r-1)$-th failures at stage 1 and stage 2, respectively, then $P(R; \underline{n}, \underline{m}, q, r)$ can be considered to be based on the same observations. The latter would just ignore the k $(q+1)$-th failure times at the subset selection at stage 1. The former is the unique Bayes procedure with respect to the symmetric prior on all permutations of any fixed $\underline{\theta} \in (0, \infty)^k$, as long as not all of the θ_i's are equal. For every prior of this type, risk and Bayes risk coincide for each of the two procedures, and thus (10) is seen to be true. This completes the proof of the theorem.

Remark. If L would depend on $\underline{n}$ and $\underline{m}$, it would naturally be non-decreasing in n_i and m_i, $i = 1, \ldots, k$. In this case, of course, one would take $n_1 = \cdots = n_k = q$ and $m_1 = \cdots = m_k = r$ as the best allocation of components to be tested. In a more complicated approach, L could also be non-decreasing in the time until a final decision is made. This would lead to an opposite requirement of sufficiently large n_i's and m_i's. We shall not discuss further such more difficult problems.

The final topic to be considered here is the choice of a suitable subset-size rule R for the decision at stage 1. This is a very challenging problem, indeed. Clearly, there does not exist any R_0, say, such that $P(R_0)$ is optimum in terms of the risk, uniformly in $\underline{\theta} \in (0, \infty)^k$, within the class $\mathcal{D}_I$. On the other hand, in a Bayes-approach, the optimum choice of R would depend heavily on L and on the chosen prior. Therefore, it seems to be justified to consider in more detail the natural rule $P(t)$, say, where $R = t$ is constant, $t \in \{2, \ldots, k-1\}$. This will be done in the next section.

3. Properties of the Procedure $P(t)$.

For a fixed $t \in \{2, \ldots, k-1\}$, let $P(t)$ be the following 2-stage selection procedure. At stage 1, the t π_i's are selected which are associated with the t largest U_i's. Then, at stage 2, the final decision is made in favor of that one π_j which is associated with the largest of the t T_i's from the π_i's which have been selected at stage 1.

A natural way of implementing the procedure $P(t)$ is to employ the so-called "indifference zone approach", which is due to Bechhofer (1954). It allows to control the probability of a correct selection, i.e. of finding the best π_i, over a range of parameter configurations, where the best π_i is sufficiently better than the other $k-1$ π_j's. Thus let us adopt in the following the 0-1 loss function, which is 0 (1) if the best π_i is (is not) finally selected. One

minus the risk, i.e. the expected loss, is then the probability of a correct selection, which will be denoted by $P_{\underline{\theta}}(CS|\mathcal{P}(t))$, $\underline{\theta} \in (0,\infty)^k$, in the sequel.

Let $\Delta > 1$ be fixed. For $\underline{\theta} \in (0,\infty)^k$, let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ denote the ordered values of $\theta_1, \ldots, \theta_k$. Then let

$$\Omega(\Delta) = \{\underline{\theta} | \underline{\theta} \in (0,\infty)^k, \quad \Delta\theta_{[1]} \leq \theta_{[i]}, \quad i = 2, \ldots, k\}. \tag{11}$$

The next result states that the probability of a correct selection with procedure $\mathcal{P}(t)$ can be controlled on $\Omega(\Delta)$. More precisely, if a value $P^* \in (1/k, 1)$ is predetermined, then values for q and r can be found such that the infimum probability of a correct selection with procedure $\mathcal{P}(t)$ is at least P^* if $\underline{\theta}$ is restricted to $\Omega(\Delta)$. It will also be shown that the parameter configuration, at which the infimum occurs, i.e. the least favorable configuration (LFC), is of the "slippage"-type.

Theorem 4. *For every $t \in \{2, \ldots, k-1\}$, and $\Delta > 1$,*

$$Inf\{P_{\underline{\theta}}(CS|\mathcal{P}(t)) | \underline{\theta} \in \Omega(\Delta)\} = P_{\underline{e}}(CS|\mathcal{P}(t)), \tag{12}$$

where $\underline{e} = (1, \Delta, \Delta, \ldots, \Delta)$ with k coordinates.

Proof. The probability of a correct selection for the procedure $\mathcal{P}(t)$ at $\underline{\theta} \in \Omega(\Delta)$ with, say, $\theta_1 = \theta_{[1]}$, has the form

$$P_{\underline{\theta}}(CS|\mathcal{P}(t)) = \sum_{\substack{\tilde{s} \subseteq \{2,\ldots,k\} \\ |\tilde{s}| = t-1}} P_{\underline{\theta}}\{U_\ell < U_i,\ \ell \notin s,\ i \in s;\ U_j + V_j < U_1 + V_1,\ j \in \tilde{s}\}, \tag{13}$$

where here in the sequel, $s = \tilde{s} \cup \{1\}$, if both s and $\tilde{s}$ appear simultaneously in an expression.

Let $B_i = \theta_i V_i$, $i = 1, \ldots, k$, be auxiliary random variables to be used in the following. It is easy to see that a lower bound to (13) is attained if for $j = 2, \ldots, k$, V_j is replaced by $B_j/\Delta\theta_1$ in the events appearing in (13). Since the distribution of the random vector $\underline{W} = (B_1/\theta_1 - B_2/\Delta\theta_1, B_1/\theta_1 - B_3/\Delta\theta_1, \ldots, B_1/\theta_1 - B_k/\Delta\theta_1)$ is seen to be permutation symmetric, this lower bound can be represented by an integral over $\{\underline{a} = (a_1, \ldots, a_{t-1}) \mid a_1 < a_2 < \ldots a_{t-1}\}$, where the integrand is a product of the joint density of the first $t-1$ coordinates of $\underline{W}$ at $\underline{a}$ and the following function of $\underline{a}$.

$$\sum_{\substack{2 \leq i_1 < i_2 < \cdots < i_{t-1} \leq k \\ \tilde{s} = \{i_1, i_2, \ldots, i_{t-1}\}}} \sum_{\sigma} P_{\underline{\theta}}\{U_\ell < U_i, \quad \ell \notin s, \quad i \in s;$$

$$U_{i_j} < U_1 + a_{\sigma(j)}, \quad j = 1, \ldots, t-1\}, \tag{14}$$

where in the second summation, σ runs over all $(t-1)!$ permutations of $(1, 2, \ldots, t-1)$. To show now that (14) is nondecreasing in $\theta_2, \ldots, \theta_k$, it suffices to prove it for θ_k. To do so, we first replace in (14) all probabilities by the corresponding conditional probabilities, given $U_1 = y_1, \ldots, U_{k-1} = y_{k-1}$, where we may assume without loss of generality that $y_2 < y_3 < \ldots < y_{k-1}$ holds. Let $b_j = y_1 + a_j$, $j = 1, \ldots, t-1$. Then we get

$$\sum_{\sigma} P_{\underline{\theta}}\{U_k, U_{k-t} < U_{k-t+1}, U_1;\ U_{k-t+1} \leq b_{\sigma(1)}, \ldots,$$

$$U_{k-1} \leq b_{\sigma(t-1)} | U_1 = y_1, \ldots, U_{k-1} = y_{k-1}\} \tag{15}$$

$$+\sum_{\sigma} P_{\underline{\theta}}\{U_{k-t+1} < U_k, U_1;\ \ U_k \le b_{\sigma(1)}, U_{k-t+2} \le b_{\sigma(2)}, \ldots,$$

$$U_{k-1} \le b_{\sigma(t-1)} | U_1 = y_1, \ldots, U_{k-1} = y_{k-1}\}.$$

In case of $y_{k-t+1} < y_1$, this reduces to

$$\sum_{\sigma} P_{\underline{\theta}}\{U_k, U_{k-t+1} \le b_{\sigma(1)}, U_{k-t+2} \le b_{\sigma(2)}, \ldots,$$

$$U_{k-1} \le b_{\sigma(t-1)} | U_1 = y_1, \ldots, U_{k-1} = y_{k-1}\}, \tag{16}$$

whereas in case of $y_1 \le y_{k-t+1}$, it reduces to

$$\sum_{\sigma} P_{\underline{\theta}}\{U_k, U_{k-t} < U_1;\ \ U_{k-t+1} \le b_{\sigma(1)}, \ldots,$$

$$U_{k-1} \le b_{\sigma(t-1)} | U_1 = y_1, \ldots, U_{k-1} = y_{k-1}\}. \tag{17}$$

Since now both, (16) and (17), are seen to be nondecreasing in θ_k, the proof of the theorem is completed by noting that $P_{\tau\underline{\theta}}(CS|P(t))$, $\tau > 0$, does not depend on τ.

It should be noted that Theorem 4 holds also for $t = 1$ and for $t = k$. But $P(1)$ and $P(k)$ are actually 1-stage selection procedures. Procedures of this type have been studied extensively in the past, and an overview of the literature in this respect can be found in Gupta and Panchapakesan (1979).

A very natural and interesting question is now to find sufficient conditions under which a 2-stage procedure of the type $P(t)$ performs better than a 1-stage procedure. This could be done, for example, on the basis of a common total number of failures. To be more specific, let us assume that there exists an integer $d > 2$, say, such that $k = dt$. Thus, if $P(t)$ is based on q failures at stage 1 and on $r = dq$ failures at stage 2, respectively, then the total number of failures becomes $k(q) + t(dq) = k(2q)$, and $P(t)$ can be compared with the optimum 1-stage procedure, P_1, say, which is based on $2q$ failures from $\pi_1, \ldots, \pi_k$.

If the $k - t$ largest θ_i's tend to large values compared to the t smallest θ_i's, then it is not difficult to see that the $P(CS)$ of $P(t)$ will be larger than the $P(CS)$ of P_1. However, at other parameter configurations, P_1 may be the better procedure. It appears thus to be more promising to compare the infima of probabilities of correct selection on $\Omega(\Delta)$, which are both attained at $\underline{\theta} = (1, \Delta, \ldots, \Delta)$. The answer to the stated question would then of course depend on Δ. No results in this respect are known.

4. Some Extensions.

A natural extension of the topics of the previous sections is to consider p-stage selection procedures for $p \ge 2$, which are based on q_i failures at stage i, $i = 1, \ldots, p$, where the selected subsets at consecutive stages are nested. For permutation invariant subset size rules $R_1 \ge R_2 \ge \cdots \ge R_{p-1}$, let $P(R_1, \ldots, R_{p-1})$ denote the procedure which selects at stage i in terms of the R_i-largest sufficient statistics for stages 1 through i combined, $i = 1, \ldots, p$, where $R_p = 1$. As in Gupta and Miescke (1984), if the loss function is generalized accordingly, it can be shown that all final decision rules, including those which are made whenever an R_i turns out to be 1, as well as the subset selection at stage $p - 1$,

are optimum in an analogous way to (4). And it can be shown that $P(R_1,\ldots,R_{p-1})$ is the unique Bayes procedure with respect to every i.i.d. prior among all procedures which employ $R_1,\ldots,R_{p-1}$, if the loss at stage i depends only on the parameters of the actually selected populations at this stage. In this case, the procedure turns out to be admissible in $\mathcal{D}(R_1,\ldots,R_{p-1})$, the natural generalization of $\mathcal{D}(R)$.

The situation becomes more favorable if $R_1 = t_1,\ldots,R_{p-1} = t_{p-1}$ are fixed, where of course $t_1 \geq t_2 \geq \cdots \geq t_{p-1}$ holds. The procedure $P(t_1,\ldots,t_{p-1})$ can be shown to be the unique Bayes procedure, with respect to every symmetric prior, within the class $\mathcal{D}(t_1,\ldots,t_{p-1})$. Thus it is also, uniformly in $\underline{\theta}$, optimum p-stage procedure in an analogous way to (4) within the class $\mathcal{D}_I(t_1,\ldots,t_{p-1})$. The proof of these facts is essentially the same as in Gupta and Miescke (1984), where one has only to take care of the slight technical modification concerning the function d in (7), which has been discussed at the end of the proof of Theorem 1. One problem, however, remains open: The least favorable parameter configuration in $\Omega(\Delta)$ under a 0–1 loss, or more specifically, for the probability of finally selecting the best population, is not known for $p > 2$.

Finally, some comments about α, the shape parameter of the underlying Weibull family, have to be made. We could have allowed from the very beginning that $\pi_1,\ldots,\pi_k$ have known shape parameters $\alpha_1,\ldots,\alpha_k$, which are not necessarily identical. If U_i, V_i, and T_i were defined as in (2), but now with α_i instead of α, $i = 1,\ldots,k$, then all subsequent results would still be true because of the facts stated in Lemma 1. However, we did not follow this idea since the statistical relevance of selecting the population with the smallest θ_i would become rather questionable.

A more interesting problem would be the following. Suppose that $\pi_1,\ldots,\pi_k$ have a common shape parameter α, which is unknown. Great difficulties arise in this situation, mainly because a reduction by sufficiency, as it was done before quite successfully in (2), is no longer possible here. The maximum likelihood approach, as it was utilized in the two papers of Kingston and Patel (1980), would lead in $\mathcal{D}(t)$ to a procedure which is almost identical with $P(t)$. The only difference is that at stage 1, α is replaced by the maximum likelihood estimator $\hat{\alpha}_1$, and at stage 2, α in U_i and V_i, $i = 1,\ldots,k$, is replaced by the overall maximum likelihood estimator $\hat{\alpha}_2$, which is based on all observed failure times. It is not known how good this procedure actually is.

If a bound $\alpha_* \leq \alpha$ (or $\alpha^* \geq \alpha$) were known, this would not be of any help. Since if we used $P(t)$ with α_* in (2) as a substitute for the unknown α, the resulting procedure would perform at the actual α worse than if α_* were the right shape parameter. This is a direct consequence of Theorem 1. Therefore, a bound of α is of no use if we wish to control the probability of a correct selection from below.

The most promising problem, however, appears to be the search for conditions under which a 2-stage procedure $P(t)$ is preferable to a 1-stage procedure P_1 which employs the same total number of failures. It has been formulated at the end of Section 3.

References

[1] 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.

[2] Blackwell, D., and Girshick, M.A. (1954). *Theory of Games and Statistical Decisions.* J. Wiley, New York.

[3] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II.* J. Wiley, New York.

[4] Gnedenko, B.V., et al. (1969). *Mathematical Methods of Reliability.* Academic Press, New York.

[5] Gupta, S.S., and Miescke, K.J. (1984). Sequential selection procedures—a decision theoretic approach. *Ann. Statist.* **12**, 336–350.

[6] Gupta, S.S., and Panchapakesan, S. (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations.* J. Wiley, New York.

[7] Kingston, J.V., and Patel, J.K. (1980). Selecting the best one of several Weibull populations. *Commun. Statist. Theor. Meth.* **A9**, 383–398.

[8] Kingston, J.V., and Patel, J.K. (1980). A restricted subset selection procedure for Weibull populations. *Commun. Statist. Theor. Meth.* **A9** 1371-1383.

[9] Miescke, K.J. (1984). Recent results on multi-stage selection procedures. *Proceedings of the 7th Conference on Probability Theory.* Iosifescu M. et al. eds. Ed. Acad. Rep. Soc. Rumania, Bucharest.

[10] Tsokos, C.P., and Rao, A.N.V. (1979). Robustness studies for Bayesian developments in Reliability. *Decision Information.* Tsokos, C.P. and Thrall, R.M. eds. Academic Press, New York.

END

FILMED

2-86

DTIC