AN ASYMPTOTIC MINIMAX TYPE PROCEDURE FOR NONPARAMETRIC SELECTION OF THE "BEST" POPULATION WITH PARTIALLY CLASSIFIED DATA

by

Tommaso Gastaldi*
Università degli Studi
di Roma "La Sapienza"
and Purdue University

and

Shanti S. Gupta**
Purdue University

Technical Report #92-28C

Department of Statistics
Purdue University

July 1992

---

* The research of this author was supported in part by the C.N.R. (Consiglio Nazionale delle Ricerche), Italy.

** The research of this author was supported in part by the NSF Grant DMS-8923071.

*

AN ASYMPTOTIC MINIMAX TYPE PROCEDURE FOR NONPARAMETRIC SELECTION OF THE "BEST" POPULATION WITH PARTIALLY CLASSIFIED DATA

by

Tommaso Gastaldi*
Dipartimento di Statistica, Probabilità
e Statistiche Applicate
Università "La Sapienza"
00185 – Roma, ITALIA
and Purdue University

and

Shanti S. Gupta**
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

Abstract

Consider a sample $\underset{\sim}{x} \equiv (x_1, \ldots, x_n)$ of n independent observations, in which each observation x_i is a realization from either one of k_i given populations, chosen among k populations $\pi_1, \ldots, \pi_k$. Our main objective is to derive a procedure for the selection of the most reliable population π_j at any arbitrarily fixed time ξ, when no assumptions about the k populations are made. A numerical example is presented.

Key Words and Phrases: Selection procedures; Survival function; Reliability; Confidence interval; Multicomponent systems; Competing risk model; Masked data.

* The research of this author was supported in part by the C.N.R. (Consiglio Nazionale delle Ricerche), Italy.

** The research of this author was supported in part by the NSF Grant DMS-8923071.

1. INTRODUCTION

This paper is concerned with a problem in multiple decision theory, where one is interested in selecting the "best" population, using partially classified observations. For a general approach to ranking and selection methodology, a reference can be made to Gupta and Panchapakesan (1979). For a broader decision theoretic formulation and Bayesian analysis, Berger (1985) is a useful reference.

Let $\underset{\sim}{x} \equiv (x_1, \ldots, x_n)$ be a sample of n independent observations, each one of which is known to be a realization of either one (and only one) of k real-valued random variables (r.v.'s) associated with k populations $\pi_1, \ldots, \pi_k$, having cumulative distribution functions (c.d.f.'s) $F_j(u)$, $j = 1, \ldots, k$, respectively.

We denote by $\bar{F}_j(u) \equiv 1 - F_j(u)$ the survival function or reliability at time u.

When there is no confusion, we will use the symbol π_j both for the j-th population and the associated r.v.

We assume that the exact parent population of each x_i, $i = 1, \ldots, n$, may *not* be precisely known, that is, in general, it may not be exactly known which one of the k populations $\pi_1, \ldots, \pi_k$ generated a given x_i; however, for each $x_i \in \underset{\sim}{x}$, a set $PP_i \equiv \{j_{i1}, \ldots, j_{ik_i}\}$, $PP_i \subset \{1, \ldots, k\}$, is *known* which lists the subscripts of k_i populations $\pi_{j_{i1}}, \ldots, \pi_{j_{ik_i}}$ such that $P[\underset{j \in PP_i}{\cup} (x_i \text{ is a realization of } \pi_j)\ ] = 1$; such populations are referred to as the "possible" parent populations of x_i.

We denote our data by $\underset{\sim}{d} \equiv \{(x_i, PP_i);\ i = 1, \ldots, n\}$. Let us denote by $|PP_i|$ the cardinality of the set PP_i. We refer to the observations with $|PP_i| = 1$ as "(fully) classified" observations. The remaining observations will be referred to as "partially classified"

observations.

It should be pointed out that, in the literature, the kind of data used here is sometimes called "masked" data. Furthermore, the model discussed in this paper can be looked upon as a slight generalization of the so-called "competing risk model." A few recent references on the selection procedures with fully classified data are: Gupta, Liang and Rau (1992), Gupta and Liang (1989, 1991), Gupta (1990), Fong and Berger (1992), and Barlow and Gupta (1969) for the problem of selecting the large quantiles. It should be noted that, while there exists a great deal of literature on selection procedures for the "best" of "good" populations, these procedures are based on data whose origin is completely known.

Examples. The above set-up formalizes the situation in which we have observations whose exact origin may not be precisely known, and *generalizes the usual model in which* k *samples, of sizes* n_j, $\sum_{j=1}^{k} n_j = n$, *drawn from* k *populations* π_j, $j = 1, \ldots, k$, *are given.* Such a situation is commonly encountered, for instance, in life-testing on multimodule, or multicomponent, serial systems with time-censored search for the cause of failure. As an example, consider a life-test conducted on n equal independent multimodule systems, each consisting of k modules *in series* (that means that the failure of any one of the modules in a system causes the failure of the entire system itself), and the object of the experiment be an analysis of the reliabilities of the modules. When any system fails, the failure time x_i is recorded and an investigation about the possible cause of failure is carried out which leads to the identification of a subset of modules, one of which certainly is the cause of failure (the cause of failure is here assumed to be unique). Such a subset of modules, which in particular could have cardinality 1, is, in general, thought of as constituted by an arbitrary

number $k_i \in \{1, \ldots, k\}$ of elements, in order to take into consideration even the case in which the search for the failed module is curtailed (censored), due to constraints of time or convenience.

Another example of an applied situation in which partially classified observations may arise is a clinical experiment k diseases are being studied on n patients, some of which may present a quantitative symptom that can be imputable to two or more of the diseases, and we are interested in processing even those the observations x_i whose origin is not precisely established but narrowed to a certain subset $\{\pi_{j_{i_1}}, \ldots, \pi_{j_{ik_i}}\}$ of possibilities.

2. FORMULATION

Let ξ be a number chosen in the intersection of the supports of the r.v.'s associated with the populations $\pi_1, \ldots, \pi_k$. Given ξ and a (high) threshold $1 - \gamma_\xi$, referred to as "survival" or "reliability level," $\gamma_\xi \in (0, 1)$, we define as "good" those populations π_j, $j = 1, \ldots, k$, such that $\bar{F}_j(\xi) > 1 - \gamma_\xi$; the remaining populations will be called "non-good."

The population(s) such that its (their) reliability at ξ is higher than or equal to that of any other population is (are) referred to as the "best."

The following propositions may serve to clarify the preceding concepts.

The number of good populations depends on ξ and γ_ξ, and can be any number of the set $\{0, \ldots, k\}$.

If the set of the good populations is not empty, a best population is necessarily an element of such a set. Consequently, if there exists a unique good population, it is also the best population.

A best population exists, and is *not* necessarily good.

For (conceptual) consistency, we assume that the threshold $1-\gamma_\xi$, which is to be fixed by the user of the procedure, whenever viewed as a function of ξ, be monotone decreasing; the virtual population π with c.d.f. $F_\pi(u)$ such that $\bar{F}_\pi(\xi) = 1-\gamma_\xi$ may be called "transition" or "control population." Intuitively, for any ξ, the good populations allocate more probability mass to the right of ξ, than the non-good populations do.

For example, if π_j, $j = 1,\ldots,k$, are lifelengths of comparable products of different brands, we may be interested in selecting the *best* one, whose survival (or reliability) at $\xi = 1$ year is the greatest and, anyway, we may consider the selected product as *good* only if its reliability at 1 year is greater than a certain level $1-\gamma_\xi$.

The selection problem. A selection rule will be denoted by $\underset{\sim}{s}(\underset{\sim}{d}) \equiv (s_1(\underset{\sim}{d}),\ldots,s_k(\underset{\sim}{d}))$, where $s_j(\underset{\sim}{d}) : \mathbf{R}^k \mapsto [0,1]$ (in general, the domain of the application could be thought of as an arbitrary space) is the probability that π_j is selected as the best population after that the data $\underset{\sim}{d}$ is observed. A selection rule $\underset{\sim}{s}(\underset{\sim}{d})$ is called nonrandomized if all its components $s_j(\underset{\sim}{d})$ can be only 0 or 1, otherwise it is a randomized rule. In this work, we restrict our attention to nonrandomized selections.

For fixed ξ and γ_ξ, we define "λ-good" a selection which, given $\underset{\sim}{d}$, ensures a confidence level at least equal to $1-\lambda$, $\lambda \in (0,1)$, that the chosen population π_j is good, i.e., $P[\bar{F}_{j^*}(u) > 1-\gamma_\xi] \geq 1-\lambda$. In general, the existence of λ-good selections depends on the chosen levels γ_ξ and λ; in fact, it is clear that (especially if the values λ and γ_ξ are chosen too small) there may be no populations which are good with probability not less than $1-\lambda$. When no λ-good selections exist at level $1-\lambda$, we may decide that all the considered populations are not good, or we may want to "redefine" our concept of goodness

by lowering the value $1-\lambda$ or $1-\gamma_\xi$, or both.

The goal of our selection problem is to derive a minimax type selection rule for choosing a best population, given the data $\underset{\sim}{d}$. We shall also provide a sufficient condition for our selection to be λ-good.

For each given value $x_i \in \underset{\sim}{x}$, let Π_i be the random population which generated the observed value x_i and let $p_i(j) \equiv P[\Pi_i = \pi_j]$ be a *given* prior positive evaluation of the probability of the given value x_i being a realization of π_j, $j \in PP_i$.

Given the data $\underset{\sim}{d}$, we denote by $\mathcal{P}$ the space of all possible choices of the distributions $\underset{\sim}{p} \equiv [p_i(j)]$, $i = 1, \ldots, n$, $j = j_{i1}, \ldots, j_{ik_i}$:

$$\mathcal{P} \equiv \{\underset{\sim}{p} \equiv [p_i(j)] : \; p_i(j) > 0, \; \sum_{j \in PP_i} p_i(j) = 1; \; i = 1, \ldots, n; \; j = j_{i1}, \ldots, j_{ik_i}\}. \qquad (2.1)$$

For each choice of a distribution $\underset{\sim}{p}$ in $\mathcal{P}$, a classification of our set of observations can be generated by randomly choosing, *for each* x_i, one parent population from the set $\{\pi_j; \; j = j_{i1}, \ldots, j_{ik_i}\}$ from a multinomial distribution with probabilities $p_i(j_{i1}), \ldots, p_i(j_{ik_i})$.

For each $i \in \{1, \ldots, n\}$, we denote by $PP_i^* \equiv \{j_i\}$ the set of cardinality 1 containing the subscript of that particular population π_{j_i}, $j_i \in PP_i$, which our random classification attributes to the observation x_i, and call a "classification" any set $(PP_1^*, \ldots, PP_n^*)$.

The space of all possible classifications which can be generated in the manner described above, is denoted by $\mathcal{Z}$:

$$\mathcal{Z} \equiv \{(PP_1^*, \ldots, PP_n^*) : \; PP_i^* \equiv \{j_i\}, \; j_i \in PP_i \text{ for any } i \in \{1, \ldots, n\}\}. \qquad (2.2)$$

The reliability at ξ of each population π_j can be estimated through the respective empirical c.d.f.. However, in our problem, together with the sampling error component,

which makes, for each $j \in \{1, \ldots, k\}$, the empirical c.d.f., say $\hat{F}_j(u)$, differ from the underlying c.d.f. $F_j(u)$, there is another source of uncertainty, due to the fact that the true parent population of some of the observed values x_i is not precisely known, and the empirical c.d.f. computed after the classification $(PP_1^*, \ldots, PP_n^*)$, say $\hat{F}_j^*(u)$, may randomly differ from the empirical c.d.f. $\hat{F}_j(u)$, which is not known (since some observations are only partially classified).

We adopt a conservative point of view, according to which we want to protect ourselves against the selection of a population with high unreliability at time ξ. This is reflected in the following minimax type criterion, referred to as α-minimax.

In determining a selection rule $\underset{\sim}{s}(\underset{\sim}{d})$ which is α-minimax with respect to a loss function $\mathcal{L}$, the quantity to be minimized for selecting the best population is $\hat{U}_\xi(j)$ defined as $\hat{U}_\xi(j) \equiv \sup_y \{y : \ y \in I_{\underset{\sim}{d},\alpha}\}$, and $I_{\underset{\sim}{d},\alpha}$ is the shortest interval which contains the loss function $\mathcal{L}$ with probability $(1-\alpha), \alpha \in (0,1)$. Note that the loss function is chosen in an appropriate way in order to penalize the choice of "unreliable" populations. In particular, we define the loss function as:

$$\mathcal{L} \equiv \mathcal{L}(\underset{\sim}{s}(\underset{\sim}{d}), \bar{F}_1(\xi), \ldots, \bar{F}_k(\xi)) \equiv 1 - \bar{F}_{j^*}(\xi) \tag{2.3}$$

where j^* is such that $s_{j^*}(\underset{\sim}{d}) = 1$, i.e., our loss is given by the *unreliability* at time ξ of the selected population. Briefly, we can state our α-minimax criterion as follows:

$$\min_{\pi_1, \ldots, \pi_k} \sup_y \{y : \ y \in I_{\underset{\sim}{d},\alpha}\}. \tag{2.4}$$

where the set $I_{\underset{\sim}{d},\alpha}$ is such that

$$P[\mathcal{L} \in I_{\underset{\sim}{d},\alpha}] = 1 - \alpha. \tag{2.5}$$

3. MAIN RESULTS : LARGE SAMPLE APPROXIMATION

We are not able to construct an interval which contains the loss function with probability exactly equal to $(1-\alpha)$, however we can provide an asymptotic approximation. For each $j \in \{1,\ldots,k\}$, let ν_j represent the number of observations which certainly have been drawn from π_j

$$\nu_j \equiv \sum_{i=1}^{n} 1_{(x_i \in \underset{\sim}{x}:\ PP_i=\{j\})} \tag{3.1}$$

where $1_{(\mathcal{A})}$ is the indicator function of the event $\mathcal{A}$, and let $\hat{\hat{F}}_j(u)$ be the corresponding empirical c.d.f.:

$$\hat{\hat{F}}_j(u) \equiv \hat{\hat{F}}_j(u;\ x_i \in \underset{\sim}{x}:\ PP_i = \{j\}) \equiv \frac{1}{\nu_j} \sum_{i=1}^{n} 1_{(x_i \le u\ \wedge\ PP_i=\{j\})}. \tag{3.2}$$

Denote

$$l_j(u) \equiv \sum_{i=1}^{n} 1_{(x_i \in \underset{\sim}{x}:\ x_i \le u\ \wedge\ j \in PP_i\ \wedge\ |PP_i|>1)} \tag{3.3}$$

and

$$r_j(u) \equiv \sum_{i=1}^{n} 1_{(x_i \in \underset{\sim}{x}:\ x_i > u\ \wedge\ j \in PP_i\ \wedge\ |PP_i|>1)}. \tag{3.4}$$

Let $\hat{F}_j(u)$ be the (unknown) empirical c.d.f. of the r.v. π_j.

Finally, denote by $\hat{F}^*_j(u)$ the empirical c.d.f. of the r.v. π_j computed using the observations $x_i,\ i = 1,\ldots n$, such that $PP^*_i = \{j\}$, which, according to our classification are considered realizations of π_j:

$$\hat{F}^*_j(u) \equiv \hat{F}^*_j(u;\ x_i \in \underset{\sim}{x}:\ PP^*_i = \{j\}) \equiv \frac{1}{\sum_{i=1}^{n} 1_{(PP^*_i=\{j\})}} \sum_{i=1}^{n} 1_{(x_i \le u\ \wedge\ PP^*_i=\{j\})}. \tag{3.5}$$

The following lemma holds.

Lemma 1. For any $j \in \{1, \ldots, k\}$ and any u, we have

$$\hat{I}_j(u) \equiv \inf_{\mathcal{Z}} \hat{F}_j^*(u) = \frac{\nu_j}{\nu_j + r_j(u)} \hat{\hat{F}}_j(u) \tag{3.6}$$

$$\hat{S}_j(u) \equiv \sup_{\mathcal{Z}} \hat{F}_j^*(u) = \frac{\nu_j}{\nu_j + l_j(u)} \hat{\hat{F}}_j(u) + \frac{l_j(u)}{\nu_j + l_j(u)} \tag{3.7}$$

and, hence, as to the unknown empirical c.d.f. $\hat{F}_j(u)$, which, for some classification $(PP_1^*, \ldots, PP_n^*)$, must necessarily be equal to $\hat{F}_j^*(u)$, we have

$$0 \leq \hat{I}_j(u) \leq \hat{F}_j(u) \leq \hat{S}_j(u) \leq 1 \tag{3.8}$$

We also have, for any u

$$\lim_{\nu_j \to \infty} \hat{I}_j(u) = \hat{F}_j(u) = \lim_{\nu_j \to \infty} \hat{S}_j(u) \tag{3.9}$$

if, as $\nu_j \to \infty$, $\frac{\nu_j(u) + l_j(u)}{\nu_j} \to 0$.

Thus, for ν_j large and a relatively small number of partially classified observations, any point in the interval $[I_j(u),\ S_j(u)]$ can be considered an approximate value of the empirical c.d.f. $\hat{F}_j(u)$. We will, hence, consider this approximate value

$$\hat{\mathcal{F}}_j(u) \equiv \frac{1}{2} [I_j(u) + S_j(u)] \approx \hat{F}_j(u). \tag{3.10}$$

Now, observing that $\nu_j \leq n_j \leq \nu_j + l_j(u) + r_j(u)$, and hence asymptotically, as $\nu_j \to \infty$ and $\frac{r_j(u) + l_j(u)}{\nu_j} \to 0$, $n_j \approx \nu_j$, a $100(1-\alpha)$ percent approximate confidence interval for our loss function $\mathcal{L} \equiv F_{j^*}(\xi)$ is obtainable from the following normal approximation:

$$P\left[\hat{L}_\xi(j^*) \leq \mathcal{L} \leq \hat{U}_\xi(j^*)\right] \approx 1 - \alpha \tag{3.11}$$

where

$$\hat{L}_j(j^*) = \max\{0, L_\xi(j^*)\}, \hat{U}_\xi(j^*) = \min\{1, U_\xi(j^*)\} \tag{3.12}$$

and

$$L_\xi(j^*) = \hat{\mathcal{F}}_{j^*}(u) - Z_{(\alpha/2)}\left[\frac{\hat{\mathcal{F}}_{j^*}(u)(1-\hat{\mathcal{F}}_{j^*}(u))}{\nu_{j^*}}\right]^{\frac{1}{2}} \tag{3.13}$$

$$U_\xi(j^*) = \hat{\mathcal{F}}_{j^*}(u) + Z_{(\alpha/2)}\left[\frac{\hat{\mathcal{F}}_{j^*}(u)(1-\hat{\mathcal{F}}_{j^*}(u))}{\nu_{j^*}}\right]^{\frac{1}{2}} \tag{3.14}$$

and $Z_{(\alpha/2)}$ is the value cutting off the area $\alpha/2$ in the upper tail of the standard normal distribution.

Alternate limits of the interval, taking into account a correction for continuity, obtained by replacing (3.13) and (3.14) with the two following expressions:

$$L'_\xi(j^*) = \frac{(2\nu_{j^*}\hat{\mathcal{F}}_{j^*}(u) + Z^2_{(\alpha/2)} - 1) + Z_{(\alpha/2)}\left[Z^2_{(\alpha/2)} - (2 + \frac{1}{\nu_{j^*}}) + 4\hat{\mathcal{F}}_{j^*}(u)(\nu_{j^*}(1-\hat{\mathcal{F}}_{j^*}(u)) + 1)\right]^{\frac{1}{2}}}{2(\nu_{j^*} + Z^2_{(\alpha/2)})} \tag{3.15}$$

$$U'_\xi(j^*) = \frac{(2\nu_{j^*}\hat{\mathcal{F}}_{j^*}(u) + Z^2_{(\alpha/2)} + 1) - Z_{(\alpha/2)}\left[Z^2_{(\alpha/2)} + (2 - \frac{1}{\nu_{j^*}}) + 4\hat{\mathcal{F}}_{j^*}(u)(\nu_{j^*}(1-\hat{\mathcal{F}}_{j^*}(u)) - 1)\right]^{\frac{1}{2}}}{2(\nu_{j^*} + Z^2_{(\alpha/2)})} \tag{3.16}$$

As a consequence of (3.11), an α-minimax selection is $\underset{\sim}{s}^*(\underset{\sim}{d}) \equiv (s_1^*(\underset{\sim}{d}), \ldots, s_k^*(\underset{\sim}{d}))$ such that $s^*_{j^*}(\underset{\sim}{d}) = 1$, $s^*_j(\underset{\sim}{d}) = 0$ for any other $j \neq j^*$, and j^* is such that $\hat{U}_\xi(j^*) = \min_{j\in\{1,\ldots,k\}} \hat{U}_\xi(j)$.

In the event there are more than one population with the same value $\hat{U}_\xi(j^*)$, we might choose, among these populations, the one (or one of those) for which $\hat{L}_\xi(j^*)$ is minimum. In

fact, once we know, for instance, that with two given selections we have the same (probable) maximum loss, in absence of any type of information about the underlying distributions, we may want to further subselect a population for which the (probable) minimum loss is smaller.

Finally, it may be noticed that a sufficient condition SC for an α-minimax selection $\underset{\sim}{s}^*(\underset{\sim}{d})$ to be λ-good is $SC \equiv \{(\hat{U}_\xi(j^*) < \gamma_\xi) \ \wedge \ (1 - \lambda \leq 1 - \alpha)\}$.

Proof of Lemma 1. Denote, for any classification $(PP_1^*, \ldots, PP_n^*)$, by the symbol $l_j^*(u)$ (the symbol $r_j^*(u)$) the number of the *partially classified* observations, less than or equal to (greater than) u, which the classification $(PP_1^*, \ldots, PP_n^*)$ attributes to π_j, $j = 1, \ldots, k$:

$$l_j^*(u) \equiv \#(x_i \in \underset{\sim}{x} : \ x_i \leq u \ \wedge \ PP_i^* = \{j\} \ \wedge \ |PP_i| > 1), \tag{3.17}$$

$$r_j^*(u) \equiv \#(x_i \in \underset{\sim}{x} : \ x_i > u \ \wedge \ PP_i^* = \{j\} \ \wedge \ |PP_i| > 1). \tag{3.18}$$

By (3.3), (3.4), (3.17), and (3.18), we have

$$l_j(u) = \max_{\underset{\sim}{Z}} l_j^*(u), \text{ and } r_j(u) = \max_{\underset{\sim}{Z}} r_j^*(u) \tag{3.19}$$

(i.e., the value $l_j(u)$ (the value $r_j(u)$) represents the maximum number *of partially classified observations* less or equal (greater) than u which could be classified as realizations of π_j, varying in all the possible ways the prior distribution $\underset{\sim}{p}$, and, hence, the consequent classification).

Thus, for any $j \in \{1, \ldots, k\}$, $\tau_j \in \{1, \ldots, \nu_j\}$ and u, we have

$$(l_j(u) - l_j^*(u))(\nu_j - \tau_j) + (\tau_j + l_j(u))r_j^*(u) \geq 0 \tag{3.20}$$

$$\Leftrightarrow \qquad \tau_j l_j(u) + l_j^*(u)\nu_j \leq \tau_j l_j^*(u) + \tau_j r_j^*(u) + l_j(u)\nu_j + l_j(u)r_j^*(u)$$

$$\Leftrightarrow \qquad \tau_j\nu_j + \tau_j l_j(u) + l_j^*(u)\nu_j + l_j^*(u)l_j(u)$$

$$\leq \tau_j\nu_j + \tau_j l_j^*(u) + \tau_j r_j^*(u) + l_j(u)\nu_j + l_j(u)l_j^*(u) + l_j(u)r_j^*(u)$$

$$\Leftrightarrow \qquad (\tau_j + l_j^*(u))(\nu_j + l_j(u)) \leq (\tau_j + l_j(u))(\nu_j + l_j^*(u) + r_j^*(u))$$

$$\Leftrightarrow \qquad \frac{\tau_j + l_j^*(u)}{\nu_j + l_j^*(u) + r_j^*(u)} \leq \frac{\tau_j + l_j(u)}{\nu_j + l_j(u)}. \tag{3.21}$$

Since for some τ_j, dependent on u, we must have $\hat{\hat{F}}_j(u) = \frac{\tau_j}{\nu_j}$, we can write

$$\frac{\tau_j + l_j^*(u)}{\nu_j + l_j^*(u) + r_j^*(u)} \leq \frac{\nu_j}{\nu_j + l_j(u)}\hat{\hat{F}}_j(u) + \frac{l_j(u)}{\nu_j + l_j(u)} \tag{3.22}$$

from which, statement (3.7) follows, for, by definition (cf., formulae (3.17), (3.18), and (3.5)), the left hand side of the inequality is the empirical c.d.f. $\hat{F}_j^*(u)$ of π_j at u under the classification $(PP_1^*, \ldots, PP_n^*)$. (In particular, the value τ_j such that $\hat{\hat{F}}_j(u) = \frac{\tau_j}{\nu_j}$ is $\tau_j = \sum_{i=1}^n 1_{(x_i \leq u \ \wedge \ PP_i = \{j\})}$.

The proof of statement (3.6) is analogous. For any $j \in \{1, \ldots, k\}$, $\tau_j \in \{1, \ldots, \nu_j\}$ and u, we have

$$(\nu_j - \tau_j)l_j^*(u) + (r_j(u) - r_j^*(u))\tau_j + l_j^*(u)r_j(u) \geq 0 \tag{3.23}$$

$$\Leftrightarrow \qquad \tau_j r_j(u) + l_j^*(u)\nu_j + l_j^*(u)r_j(u) - \tau_j l_j^*(u) - \tau_j r_j^*(u) \geq 0$$

$$\Leftrightarrow \qquad \tau_j\nu_j + \tau_j r_j(u) + l_j^*(u)\nu_j + l_j^*(u)r_j(u) \geq \tau_j\nu_j + \tau_j l_j^*(u) + \tau_j r_j^*(u)$$

$$\Leftrightarrow \qquad (\tau_j + l_j^*(u))(\nu_j + r_j(u)) \geq \tau_j(\nu_j + l_j^*(u) + r_j^*(u))$$

$$\Leftrightarrow \qquad \frac{\tau_j + l_j^*(u)}{\nu_j + l_j^*(u) + r_j^*(u)} \geq \frac{\tau_j}{\nu_j + r_j(u)} \tag{3.24}$$

and, hence, we must have

$$\frac{\tau_j + l_j^*(u)}{\nu_j + l_j^*(u) + r_j^*(u)} \geq \frac{\nu_j}{\nu_j + r_j(u)} \hat{\bar{F}}_j(u). \tag{3.25}$$

4. A NUMERICAL EXAMPLE

In this section, we present an example to show how the above procedure works.

A simulation was carried out in order to reproduce the conditions of a life-test involving $(n =)1000$ systems s_i, of the same type, made of $(k =)3$ components C_j, $j = 1, 2, 3$, in series. In particular, in order to generate the data, for each system s_i, three lifelengths t_{i1}, t_{i2}, t_{i3} have been drawn from three different independent c.d.f. $F_1(u)$, $F_2(u)$, $F_3(u)$, (which does not matter to specify here) and, then, the failure time x_i of the system has been computed as $x_i = \min\{t_{i1}, t_{i2}, t_{i3}\}$. Finally, a masking has been appropriately carried out on the causes of failure, in order to simulate a time-censored search for the failed component on a small proportion of systems.

In Table 1, a portion of the data is listed (the complete data is available from the authors).

To the purposes of our example, the data generated as described above may be thought of as the outcome of a life-testing experiment conducted on 1000 3-component devices s_i in which, when a device s_i fails (because one of its component fails), the failure time x_i is recorded together with a set $PP_i = \{j_{i1}, \ldots, j_{ik_i}\}$ such that $P[$ the subscript of the failed component is in $PP_i] = 1$. The set PP_i might be, for instance, the result of a time-censored

search for the cause of failure.

The distributions of the lifelengths of the components are assumed to be totally unknown and no hypothesis is made about them, except that one of independence. (In presence of some distributional information, it is obvious that selection rules "better" than the one developed here can be found.)

We want to select the "best" component in the α-minimax sense.

We put $1 - \alpha = 0.95$ and show in Table 2 the outcome of the procedure when applied for several different values of ξ.

Table 1

Simulated Data

x_i	PP_i	x_i	PP_i	x_i	PP_i	x_i	PP_i	x_i	PP_i
0.4918	(2)	3.6590	(2)	5.0039	(2)	6.1849	(3 2)	7.4050	(3)
0.6219	(2)	3.6639	(3 2 1)	5.0059	(2)	6.1870	(2)	7.4124	(2 1)
0.6466	(2)	3.6893	(2)	5.0084	(2)	6.1889	(3)	7.4160	(3)
0.6651	(3)	3.6931	(3)	5.0096	(3)	6.1912	(1)	7.4161	(1 2)
0.9001	(1)	3.6949	(2 3)	5.0127	(1)	6.1953	(2 3 1)	7.4199	(2)
0.9128	(2)	3.6951	(1)	5.0172	(1)	6.2049	(1 3)	7.4388	(3)
here 192 rows have been omitted									
3.6424	(2)	4.9835	(3)	6.1613	(1)	7.4010	(2)	9.6431	(3)
3.6567	(3)	4.9852	(2)	6.1688	(2)	7.4029	(1)	9.7265	(2)

Table 2 shows the selected populations, in correspondence to some values of ξ. In particular, with reference to our notation, the structure of Table 2 is as follows:

ξ	$\hat{U}_\xi(1)$	$\hat{U}_\xi(2)$	$\hat{U}_\xi(3)$	(subscript of selected π_j)
	$\hat{L}_\xi(1)$	$\hat{L}_\xi(2)$	$\hat{L}_\xi(3)$	

repeated for various ξ, where the $\hat{U}_\xi$'s and $\hat{L}_\xi$'s are computed according to formulae (3.12), (3.13), and (3.14).

Table 2

ξ	$\hat{U}_\xi(1)$ & $\hat{L}_\xi(1)$	$\hat{U}_\xi(2)$ & $\hat{L}_\xi(2)$	$\hat{U}_\xi(3)$ & $\hat{L}_\xi(3)$	Selected population
2.0	0.03775	0.09040	0.04420	(1)
	0.00161	0.04451	0.00330	
3.0	0.10543	0.18571	0.14962	(1)
	0.03825	0.11988	0.06627	
4.0	0.22858	0.33866	0.31965	(1)
	0.12889	0.25507	0.20173	
5.0	0.41320	0.47143	0.46945	(1)
	0.28902	0.38095	0.33767	
6.0	0.56181	0.65182	0.63783	(1)
	0.43172	0.56246	0.50491	
7.0	0.77018	0.78926	0.79663	(1)
	0.65227	0.70999	0.67845	
7.9	0.90118	0.89822	0.91976	(2)
	0.80968	0.83613	0.83106	
8.0	0.91102	0.90542	0.93037	(2)
	0.82261	0.84494	0.84567	
8.8	0.97785	0.97834	0.98901	(1)
	0.92079	0.94271	0.93897	
8.9	0.97785	0.97990	0.99176	(1)
	0.92079	0.94515	0.94463	
9.0	0.98063	0.97990	0.99441	(2)
	0.92565	0.94515	0.95041	
9.5	1.00000	1.00000	1.00000	(3)
	1.00000	0.99020	0.98709	
9.6	1.00000	1.00000	1.00000	(3)
	1.00000	0.99020	0.98709	
9.7	1.00000	1.00000	1.00000	(2)
	1.00000	0.99390	1.00000	
9.8	1.00000	1.00000	1.00000	
	1.00000	1.00000	1.00000	

In Table 3, the values of $\hat{I}_j(u)$ and $\hat{S}_j(u)$, in correspondence to different values of u, are reported. According to formula (3.8), for any u, the set $[\hat{I}_j(u),\ \hat{S}_j(u)]$ certainly contains the unknown empirical c.d.f. $\hat{F}_j(u)$. The structure of Table 3 is the following:

u	$\hat{S}_1(u)$	$\hat{S}_2(u)$	$\hat{S}_3(u)$
	$\hat{I}_1(u)$	$\hat{I}_2(u)$	$\hat{I}_3(u)$

repeated for several u.

Table 3

u	$\hat{S}_1(u)$ & $\hat{I}_1(u)$	$\hat{S}_2(u)$ & $\hat{I}_2(u)$	$\hat{S}_3(u)$ & $\hat{I}_3(u)$
0.4	0.00000	0.00000	0.00000
	0.00000	0.00000	0.00000
0.5	0.00000	0.00218	0.00000
	0.00000	0.00184	0.00000
1.0	0.00441	0.01089	0.00469
	0.00327	0.00919	0.00373
3.5	0.14103	0.25214	0.21818
	0.08696	0.20374	0.15709
4.0	0.21849	0.32770	0.30045
	0.13898	0.26604	0.22093
5.0	0.41600	0.46584	0.45652
	0.28622	0.38654	0.35060
6.0	0.57090	0.64940	0.62810
	0.42264	0.56487	0.51464
7.0	0.77305	0.78378	0.78039
	0.64940	0.71546	0.69469
8.0	0.90169	0.89474	0.90566
	0.83193	0.85563	0.87037
9.0	0.96689	0.97032	0.97753
	0.93939	0.95474	0.96729
9.7	1.00000	0.99816	1.00000
	1.00000	0.99782	1.00000
9.8	1.00000	1.00000	1.00000
	1.00000	1.00000	1.00000

Aknowledgment

The authors thank Professor TaChen Liang and Mr. Yuning Liao for a critical reading of an earlier draft of this paper.

References

Barlow, R. E. and Gupta, S. S. (1969). Selection procedures for restricted families of probability distributions, *Ann. Math. Stat.*, 40, 905-917.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*, Springer-Verlag, New York, Inc.

Fong, D. K. H. and Berger, J. O. (1992). Ranking, estimation and hypothesis testing in unbalanced two-way additive models: a Bayesian approach. To appear in *Statistics and Decisions.*

Gupta, S. S. and Liang, T.-C. (1989). Selecting the best binomial population: parametric empirical Bayes approach. *J. Statist. Plann. Inference*, 23, 21-31.

Gupta, S. S. (1990). Recent advances in statistical ranking and selection: theory and methods. In the *Proceedings of the 1990 Taipei Symposium in Statistics* (Eds: M. T. Chao and P. T. Cheng), Academia Sinica, Taiwan, 133-166.

Gupta, S. S., Liang, T.-C., and Rau, R.-B. (1992). Empirical Bayes two-stage procedures for selecting the best Bernoulli population compared with a control. To appear in *Statistical Decision Theory and Related Topics V*, Springer-Verlag, New York.

Gupta, S. S. and Panchapakesan, S. (1979). *Multiple decision procedures — Theory and methodology of selecting and ranking populations*, John Wiley & Sons, Inc.