MINIMAX TYPE PROCEDURES 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 #94-9C

(An expanded and revised
version of Technical Report #92-28C)

Department of Statistics
Purdue University

April 1994

△

MINIMAX TYPE PROCEDURES FOR NONPARAMETRIC SELECTION OF THE "BEST" POPULATION WITH PARTIALLY CLASSIFIED DATA

Tommaso Gastaldi

Dipartimento di Statistica, Probabilità e Statistiche Applicate
Università degli Studi di Roma "La Sapienza", 00185 - Roma, ITALIA
and
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

Shanti S. Gupta

Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

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

ABSTRACT

Consider a sequence $\underset{\sim}{x} \equiv (x_1, \cdots, x_n)$ of n independent observations, in which each observation x_i is known to be a realization from either one of k_i given populations, chosen among k ($\geq k_i$) populations $\pi_1, \cdots, \pi_k$. Our main objective is to study the problem of the selection of the most reliable population π_j at a fixed time ξ, when no assumptions about the k populations are made. Some numerical examples are presented.

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.

Prior to defining what we mean by "best" population within the context of our selection problem, we shall introduce a general set-up with "partially classified" data.

While there exists a great deal of literature on selection procedures for the "best" or "good" populations, these procedures are based on data whose origin is completely known. A few recent references on selection procedures with fully classified data are: Gupta, Liang and Rau (1992), Gupta and Liang (1989, 1991), Gupta (1990), Fong and Berger (1993), and Barlow and Gupta (1969) for the problem of selecting the largest quantiles.

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 $\underset{\sim}{x} \equiv (x_1, \cdots, x_n)$ be a sequence of n independent observations, each one of which is a realization of either one (and only one) of k real-valued random variables (r.v.'s) associated with k populations $\pi_1, \cdots, \pi_k$, having cumulative distribution functions (c.d.f.'s) $F_j(u)$, j=1,$\cdots$, k, and survival or *reliability* functions $\bar{F}_j(u) \equiv 1 - F_j(u)$, j=1,$\cdots$, k, respectively. We assume that the exact parent population of each x_i, i=1,$\cdots$, n, may *not* be precisely known, that is, in general, it may not be exactly known which one of the k populations $\pi_1, \cdots, \pi_k$ generated a given x_i. However, for each $x_i \in \underset{\sim}{x}$, a set $PP_i \equiv \{j_{i1}, \cdots, j_{ik_i}\}$, $PP_i \subset \{1, \cdots, k\}$, is *known* which lists the subscripts of k_i populations $\pi_{j_{i1}}, \cdots, \pi_{j_{ik_i}}$ such that $\mathbb{P}[\bigcup_{j \in PP_i}(x_i \text{ is a realization of } \pi_j)] = 1$. Such populations will be referred to as the "possible parent populations" of x_i.

When there is no confusion, we will use the symbol π_j both for the j-th population and the associated r.v. with c.d.f. $F_j(u)$.

We denote our data by $\mathfrak{D} \equiv \{(x_i, PP_i);\ i=1, \cdots, n\}$ and 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.

Examples. The above framework 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,$\cdots$, k, *are given.* Such a situation is commonly encountered in several practical instances. Some examples of applied situations where partially classified observations may arise are the following:

– **Loss of data.** We put n units of k different types of electronic component on test simultaneously to study their reliabilities. Due to faults of the testing equipment, for some lifelengths x_i's, the type of failed component is not recorded or this information is lost or confused.

– **Curtailed investigation on the origin of some x_i's.** Lifetime data obtained from n parallel multicomponent systems Σ_i is being used to establish the reliabilities of the components. To save time, when a failure is detected at time x_i on a system Σ_i, an entire module (group of components) could be replaced, while the system is still working, without attempting to find out the specific failed component.

– **Competing risk model with incomplete knowledge on the cause of failure.** A life-test is conducted on n equal independent multimodule systems, each consisting of k modules in series (this means that the failure of any one of the modules in a system causes the failure of the entire system), and the object of the experiment is an analysis of the reliabilities of the modules. When a system fails, the failure time x_i is recorded and an investigation about the possible cause of failure is carried out leading to the identification of a subset of modules, one of which certainly is the cause of failure. Such a subset of modules, which in particular could have cardinality 1, is, in general, thought of as constituted by an arbitrary number $k_i \in \{1,\cdots, k\}$ of elements, in order to take into consideration even the case when the search for the failed module is curtailed, due to constraints of time or convenience (this example is dealt with in detail in Section 4).

– **Uncertainty about the origin of some x_i's.** In a clinical experiment, k diseases are being studied on n patients, some of which may present a quantitative symptom x_i which can be imputed to two or more of the

diseases. We are interested in processing even the observations x_i whose origin, although not precisely established, is narrowed to a certain subset $\{\pi_{j_{i1}}, \cdots, \pi_{j_{ik_i}}\}$ 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 $\pi_1, \cdots, \pi_k$.

The population(s) such that its (their) reliability $\bar{F}_j(\xi)$ at ξ is higher than or equal to that of any other population is (are) referred to as the "best."

For example, if π_j, j=1,$\cdots$, k, are lifelengths of comparable products of different brands, we may be interested in selecting the *best* one, whose survival (or reliability) at ξ=1 year is the highest one.

The selection problem. A selection rule will be denoted by $\underset{\sim}{s}(\mathfrak{D}) \equiv (s_1(\mathfrak{D}), \cdots, s_k(\mathfrak{D}))$, $\sum_{j=1}^{k} s_j(\mathfrak{D})=1$, where $s_j(\mathfrak{D})$: $\mathbb{R}^k \mapsto [0,1]$ (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 $\mathfrak{D}$ is observed. A selection rule $\underset{\sim}{s}(\mathfrak{D})$ is called nonrandomized if all its components $s_j(\mathfrak{D})$ can be only 0 or 1, otherwise it is a randomized rule. In this work, we restrict our attention to nonrandomized selections.

The goal of our selection problem is to derive a minimax type selection rule for choosing a best population, given the data $\mathfrak{D}$.

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 $p_i(j) \equiv \mathbb{P}[\ \Pi_i = \pi_j\ ]$ be a *given* prior evaluation of the probability of the given value x_i being a realization of π_j, $j \in PP_i$. This probability could, for instance, be based on some information, possibly inferred from previous experiments, that we have on the $F_j(u)$'s.

For each $\underset{\sim}{p}$, a classification of our set of observations can be generated by randomly choosing for each x_i one parent population from the set $\{\pi_1, \cdots, \pi_k\}$ with respective probabilities $p_i(1), \cdots, p_i(k)$, where $p_i(j) \geq 0$ if $j \in PP_i$, or $p_i(j) = 0$ otherwise, $\sum_j p_i(j) = 1$.

For each $i \in \{1, \cdots, n\}$, we denote by $PP_i^* \equiv \{j_i\}$ the set of cardinality 1

containing the subscript of that specific 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 $(PP_1^*, \cdots, PP_n^*)$.

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

$$\mathcal{Z} \equiv \{ (PP_1^*, \cdots, PP_n^*): PP_i^* \equiv \{j_i\}, j_i \in PP_i \text{ for each } i \in \{1,\cdots, n\} \}. \tag{2.1}$$

The reliability at ξ of each population π_j can be estimated through the corresponding empirical c.d.f.. However, in our problem, together with the sampling error component, which makes, for each $j \in \{1,\cdots, k\}$, 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 $(PP_1^*, \cdots, PP_n^*)$, 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 $\underline{s}(\mathfrak{D})$ which is α-minimax with respect to a loss $\mathcal{L}$, the quantity to be minimized for selecting the best population is $\hat{U}_{\xi,\alpha}(j)$ defined as $\hat{U}_{\xi,\alpha}(j) = \sup_y \{y: y \in I_{\mathfrak{D},\alpha}(j)\}$, and $I_{\mathfrak{D},\alpha}(j)$ is the shortest interval which contains the loss $\mathcal{L}$ with probability $(1-\alpha)$, $\alpha \in (0,1)$. Note that the loss is chosen in an appropriate way, in order to penalize the choice of "unreliable" populations. In particular, we define the loss as:

$$\mathcal{L}(\underline{s}(\mathfrak{D}), \overline{F}_1(\xi), \cdots, \overline{F}_k(\xi)) \equiv 1-\overline{F}_{j^*}(\xi) \tag{2.2}$$

where j^* is such that $s_{j^*}(\mathfrak{D})=1$, 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:

$$\min_{j \in \{1,\cdots, k\}} \sup_y \{y: y \in I_{\mathfrak{D},\alpha}(j)\} \tag{2.3}$$

where the interval $I_{\mathfrak{D},\alpha}(j)$ is such that

$$\mathbb{P}[\, F_j(\xi) \in I_{\mathfrak{D},\alpha}(j) \,] = 1-\alpha, \qquad j = 1,\cdots, k. \tag{2.4}$$

3. MAIN RESULTS: LARGE SAMPLE APPROXIMATION

We are not able to construct an interval which contains the loss function with probability equal to $(1-\alpha)$, however, we can provide an asymptotic approximation.

For each $j \in \{1,\cdots, k\}$, let $\hat{F}_j(u)$ be the (unknown) empirical c.d.f. of the r.v. π_j:

$$\hat{F}_j(u) \equiv \frac{1}{n_j} \,|\, \{\, i \mid x_i \le u,\ x_i \text{ comes from } \pi_j \,\} \,|$$

where n_j is the (unknown) number of observations which are realizations from $F_j(u)$.

Denoting by ν_j the number of observations which certainly have been drawn from π_j

$$\nu_j \equiv |\, \{\, i \mid PP_i=\{j\} \,\} \,| \tag{3.1}$$

let $\hat{\hat{F}}_j(u)$ be the corresponding empirical c.d.f.:

$$\hat{\hat{F}}_j(u) \equiv \frac{1}{\nu_j} \,|\, \{\, i \mid x_i \le u,\ PP_i=\{j\} \,\} \,|. \tag{3.2}$$

Finally, denote by $\hat{F}^*_j(u)$ the empirical c.d.f. of the r.v. π_j computed using the observations x_i, $i=1,\cdots, n$, such that $PP^*_i=\{j\}$, which, *according to a specific classification* $(PP^*_1, \cdots, PP^*_n)$, are considered realizations of π_j:

$$\hat{F}^*_j(u) \equiv \frac{|\, \{\, i \mid x_i \le u,\ PP^*_i=\{j\} \,\} \,|}{|\, \{\, i \mid PP^*_i=\{j\} \,\} \,|}. \tag{3.3}$$

Denoting

$$\lambda_j(u) \equiv |\, \{\, i \mid x_i \le u,\ j \in PP_i,\ |PP_i| > 1 \,\} \,| \tag{3.4}$$

and

$$\rho_j(u) \equiv |\{ i \mid x_i > u, j \in PP_i, |PP_i| > 1 \}| \tag{3.5}$$

the following lemma holds.

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

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

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

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

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

We also have, for any fixed u and $\hat{\hat{F}}_j(u)$,

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

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

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

$$\hat{\mathfrak{F}}_j(u) \equiv \tfrac{1}{2} [\, \hat{I}_j(u) + \hat{S}_j(u) \,] \sim \hat{F}_j(u). \tag{3.10}$$

Now, observing that, for any u,

$$\nu_j \le n_j \le \nu_j + \lambda_j(u) + \rho_j(u)$$

and hence asymptotically, as $\nu_j \to \infty$ and $\frac{\rho_j(u) + \lambda_j(u)}{\nu_j} \to 0$, we have $n_j \sim \nu_j$, a $(1-\alpha)$ approximate confidence interval for our loss function is obtainable from the following normal approximation:

$$\mathbb{P}[\, \hat{L}_{\xi,\alpha}(j) \leq F_j(\xi) \leq \hat{U}_{\xi,\alpha}(j) \,] \approx 1-\alpha \tag{3.11}$$

where

$$\hat{L}_{\xi,\alpha}(j) \equiv \max\{0, L_{\xi,\alpha}(j)\}, \text{ and } \hat{U}_{\xi,\alpha}(j) \equiv \min\{1, U_{\xi,\alpha}(j)\} \tag{3.12}$$

and

$$L_{\xi,\alpha}(j) = \hat{\mathfrak{F}}_j(\xi) - z_{(\alpha/2)} \left[\frac{\hat{\mathfrak{F}}_j(\xi)\,(1-\hat{\mathfrak{F}}_j(\xi))}{\nu_j}\right]^{\frac{1}{2}} \tag{3.13}$$

$$U_{\xi,\alpha}(j) = \hat{\mathfrak{F}}_j(\xi) + z_{(\alpha/2)} \left[\frac{\hat{\mathfrak{F}}_j(\xi)\,(1-\hat{\mathfrak{F}}_j(\xi))}{\nu_j}\right]^{\frac{1}{2}} \tag{3.14}$$

where $z_{(\alpha/2)}$ 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 bringing normal curve probabilities into closer agreement with binomial probabilities, are obtained by replacing (3.13) and (3.14) with the following two expressions (cf., Fleiss (1981), p. 14):

$$L'_{\xi,\alpha}(j) \equiv \frac{(2\nu_j\,\hat{\mathfrak{F}}_j(\xi) + z^2_{(\alpha/2)} - 1)}{2(\nu_j + z^2_{(\alpha/2)})}$$

$$- \frac{z_{(\alpha/2)}\,[\, z^2_{(\alpha/2)} - (2+\frac{1}{\nu_j}) + 4\hat{\mathfrak{F}}_j(\xi)\,(\nu_j(1-\hat{\mathfrak{F}}_j(\xi)) + 1\,]^{\frac{1}{2}}}{2(\nu_j + z^2_{(\alpha/2)})} \tag{3.15}$$

$$U'_{\xi,\alpha}(j) \equiv \frac{(2\nu_j\,\hat{\mathfrak{F}}_j(\xi) + z^2_{(\alpha/2)} + 1)}{2(\nu_j + z^2_{(\alpha/2)})}$$

$$+ \frac{z_{(\alpha/2)} \left[z^2_{(\alpha/2)} + (2-\frac{1}{\nu_j}) + 4\hat{\mathfrak{F}}_j(\xi)\,(\nu_j(1-\hat{\mathfrak{F}}_j(\xi)) - 1 \right]^{\frac{1}{2}}}{2(\nu_j + z^2_{(\alpha/2)})}. \qquad (3.16)$$

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

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

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

$$\lambda_j^*(u) \equiv | \{ i \mid x_i \leq u, PP_i^*=\{j\}, |PP_i| > 1 \} | \qquad (3.17)$$

$$\rho_j^*(u) \equiv | \{ i \mid x_i > u, PP_i^*=\{j\}, |PP_i| > 1 \} |. \qquad (3.18)$$

By (3.4), (3.5), (3.17), and (3.18), we have

$$\lambda_j(u) = \max_{\mathcal{Z}} \lambda_j^*(u), \text{ and } \rho_j(u) = \max_{\mathcal{Z}} \rho_j^*(u) \qquad (3.19)$$

(i.e., the value $\lambda_j(u)$ (the value $\rho_j(u)$) 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 each $j \in \{1,\cdots, k\}$, $\tau_j \in \{1,\cdots, \nu_j\}$, and any fixed u, we have

$$(\lambda_j(u) - \lambda_j^*(u)) (\nu_j - \tau_j) + (\tau_j + \lambda_j(u)) \rho_j^*(u) \geq 0 \qquad (3.20)$$

$$\Leftrightarrow \quad \tau_j \lambda_j(u) + \lambda_j^*(u) \nu_j \leq \tau_j \lambda_j^*(u) + \tau_j \rho_j^*(u) + \lambda_j(u) \nu_j + \lambda_j(u) \rho_j^*(u)$$

$$\Leftrightarrow \quad \tau_j \nu_j + \tau_j \lambda_j(u) + \lambda_j^*(u) \nu_j + \lambda_j^*(u) \lambda_j(u)$$

$$\leq \tau_j \nu_j + \tau_j \lambda_j^*(u) + \tau_j \rho_j^*(u) + \lambda_j(u) \nu_j + \lambda_j(u) \lambda_j^*(u) + \lambda_j(u) \rho_j^*(u)$$

$$\Leftrightarrow \quad (\tau_j + \lambda_j^*(u)) (\nu_j + \lambda_j(u)) \leq (\tau_j + \lambda_j(u)) (\nu_j + \lambda_j^*(u) + \rho_j^*(u))$$

$$\Leftrightarrow \quad \frac{\tau_j + \lambda_j^*(u)}{\nu_j + \lambda_j^*(u) + \rho_j^*(u)} \leq \frac{\tau_j + \lambda_j(u)}{\nu_j + \lambda_j(u)}. \qquad (3.21)$$

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

$$\frac{\tau_j(u) + \lambda_j^*(u)}{\nu_j + \lambda_j^*(u) + \rho_j^*(u)} \leq \frac{\nu_j}{\nu_j + \lambda_j(u)} \hat{\hat{F}}_j(u) + \frac{\lambda_j(u)}{\nu_j + \lambda_j(u)} \qquad (3.22)$$

from which statement (3.7) follows, for, by definition (cf., formulae (3.3), (3.4), 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 $(PP_1^*, \cdots, PP_n^*)$. Clearly, the value $\tau_j(u)$ such that $\hat{\hat{F}}_j(u) = \frac{\tau_j(u)}{\nu_j}$ is $\tau_j(u) = | \{ i \mid x_i \leq u, PP_i = \{j\} \} |$.

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

$$(\nu_j - \tau_j) \lambda_j^*(u) + (\rho_j(u) - \rho_j^*(u)) \tau_j + \lambda_j^*(u) \rho_j(u) \geq 0 \qquad (3.23)$$

$$\Leftrightarrow \quad \tau_j \rho_j(u) + \lambda_j^*(u) \nu_j + \lambda_j^*(u) \rho_j(u) - \tau_j \lambda_j^*(u) - \tau_j \rho_j^*(u) \geq 0$$

$$\Leftrightarrow \quad \tau_j\nu_j + \tau_j\,\rho_j(u) + \lambda_j^*(u)\,\nu_j + \lambda_j^*(u)\,\rho_j(u) \;\geq\; \tau_j\nu_j + \tau_j\,\lambda_j^*(u) + \tau_j\,\rho_j^*(u)$$

$$\Leftrightarrow \qquad (\,\tau_j + \lambda_j^*(u)\,)\,(\nu_j + \rho_j(u)\,) \;\geq\; \tau_j\,(\,\nu_j + \lambda_j^*(u) + \rho_j^*(u)\,)$$

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

and, hence, similar to (3.21), we must have

$$\Leftrightarrow \qquad \frac{\tau_j(u) + \lambda_j^*(u)}{\nu_j + \lambda_j^*(u) + \rho_j^*(u)} \geq \frac{\nu_j}{\nu_j + \rho_j(u)}\,\hat{\bar{F}}_j(u). \tag{3.25}$$

□

Note. In line with the α-minimax criterion, we might define as "λ-good" a selection which, given the data $\mathfrak{D}$, ensures a confidence level at least equal to $1-\lambda$, $\lambda \in (0,1)$, that the chosen population has a reliability at ξ higher than a certain level, $1-\gamma_\xi$ say, i.e., $\mathbb{P}[\bar{F}_{j^*}(\xi) > 1-\gamma_\xi] \geq 1-\lambda$. 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 λ-good populations. When this happens, we may decide that all the considered populations are not good, or we may want to "redefine" our concept of goodness by reducing the value $1-\lambda$ or $1-\gamma_\xi$, or both.

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

4. A NUMERICAL EXAMPLE

In this section, we present an example of application of the proposed procedure within a generalized version of the competing risk model, where there may be more than one possible cause of failure in correspondence to each failure time. A reference on parametric estimation within this model is

(Gastaldi, 1993), where further references may be found.

A simulation was carried out in order to reproduce the conditions of a life-test involving (n=)1000 systems Σ_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 Σ_i, $i = 1,\cdots, n$, three lifelengths t_{i1}, t_{i2}, t_{i3} have been drawn from three different independent absolutely continuous c.d.f.'s $G_1(u)$, $G_2(u)$, $G_3(u)$, (with increasing failure rates functions and limited lifelengths) and, then, the failure time x_i of the system has been computed as $x_i=\min\{t_{i1}, t_{i2}, t_{i3}\}$. The lifelengths t_{ij}, $i = 1,\cdots, n$, $j = 1,\cdots, k$, are assumed to be independent and, for each $j = 1,\cdots, k$, the r.v.'s $t_{1j}, \cdots, t_{nj}$ are identically distributed (they are lifelengths of the same component). Finally, a masking has been 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. The percentage of fully classified observation is 89.9%, while there is 8.4% observations with $PP_i = \{j_1, j_2\}$, $\{j_1, j_2\} \subset \{1, 2, 3\}$, and 1.7% observations with $PP_i = \{1, 2, 3\}$. Within the 899 fully classified observations, the 25.25% are from π_1, the 51.06% are from π_2, and the 23.69% are from π_3; in particular, we have $\nu_1 = 227$, $\nu_2 = 459$, $\nu_3 = 213$. Also, denoting by $\nu_{j_1 j_2}$ the number of occurrences of the event $PP_i = \{j_1, j_2\}$, we have $\nu_{12} = 46$, $\nu_{13} = 16$, $\nu_{23} = 22$.

In Table I, a portion of the data is listed (the complete data is available from the authors). The structure of Table I is as follows:

$$
\begin{array}{llllll}
x_{(1)} & PP_{(1)} & x_{(201)} & PP_{(201)} & \cdots & x_{(801)} \quad PP_{(801)} \\
\cdots & & & & & \\
x_{(200)} & PP_{(200)} & x_{(400)} & PP_{(400)} & \cdots & x_{(1000)} \quad PP_{(1000)}
\end{array}
$$

where $x_{(r)}$ denotes the r-th order statistic of the sample $\underset{\sim}{x}$ and $PP_{(i)}$ the associated set of subscripts of possible parent populations.

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 Σ_i in which, when a device Σ_i fails (because one of its component fails), the failure time x_i is recorded together with a set $PP_i=\{j_{i1},\cdots, j_{ik_i}\}$ such that $\mathbb{P}$[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; cf., discussion in Section 5.)

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

We put $1-\alpha = 0.95$ and show in Table II the outcome of the procedure when applied for several different values of ξ, where the $\hat{L}_{\xi,\alpha}(j)$'s and the $\hat{U}_{\xi,\alpha}(j)$'s are computed according to formulae (3.12), (3.13), and (3.14).

Table I

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 }	7.4160	{ 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 }
0.9912	{ 2 }	3.7006	{ 2 }	5.0259	{ 2 }	6.2052	{ 1 }	7.4443	{ 3 }
1.0793	{ 2 }	3.7013	{ 2 }	5.0284	{ 2 }	6.2171	{ 2 }	7.4486	{ 3 }
1.1322	{ 2 }	3.7015	{ 1 }	5.0376	{ 2 }	6.2342	{ 3 }	7.4548	{ 2 }
1.1552	{ 2 }	3.7099	{ 2 }	5.0385	{ 2 }	6.2352	{ 2 3 1 }	7.4561	{ 2 }
1.2384	{ 2 3 }	3.7174	{ 3 }	5.0425	{ 2 1 }	6.2355	{ 2 }	7.4606	{ 2 }
1.2571	{ 2 }	3.7265	{ 1 }	5.0443	{ 3 }	6.2418	{ 1 }	7.4613	{ 3 }
1.2649	{ 2 }	3.7446	{ 2 }	5.0508	{ 1 }	6.2482	{ 2 }	7.4726	{ 3 }
1.2800	{ 2 }	3.7446	{ 3 }	5.0543	{ 2 }	6.2762	{ 1 }	7.4791	{ 1 }
1.3024	{ 2 }	3.7489	{ 1 }	5.0614	{ 3 }	6.2774	{ 3 }	7.4845	{ 3 }
1.3039	{ 2 }	3.7548	{ 3 }	5.0628	{ 2 }	6.2804	{ 3 }	7.4873	{ 1 }
1.3110	{ 2 }	3.7600	{ 2 }	5.0727	{ 2 }	6.2848	{ 2 }	7.4890	{ 1 }
1.3182	{ 2 }	3.7644	{ 2 }	5.0752	{ 1 }	6.3056	{ 1 }	7.4938	{ 2 }
1.3245	{ 3 1 }	3.7701	{ 2 }	5.0809	{ 1 2 }	6.3105	{ 1 }	7.5068	{ 2 }
1.3305	{ 1 }	3.7725	{ 3 1 2 }	5.0831	{ 3 }	6.3144	{ 3 }	7.5223	{ 2 }
1.3305	{ 2 }	3.7744	{ 2 }	5.0905	{ 1 }	6.3162	{ 1 }	7.5335	{ 2 }
1.4326	{ 1 }	3.7796	{ 1 }	5.0938	{ 2 }	6.3207	{ 1 3 }	7.5345	{ 1 }
---									
here 165 rows have been omitted									

3.5585	{ 2 1 }	4.9279	{ 2 }	6.1117	{ 2 }	7.2990	{ 1 }	9.3427	{ 3 2 }
3.5597	{ 2 }	4.9330	{ 3 2 }	6.1136	{ 1 }	7.3071	{ 2 3 1 }	9.3627	{ 2 }

3.5630 { 3 }	4.9491 { 3 }	6.1166 { 2 }	7.3160 { 3 1 }	9.3895 { 1 }
3.5672 { 2 }	4.9497 { 3 }	6.1197 { 1 }	7.3166 { 1 }	9.3914 { 1 }
3.5838 { 3 }	4.9549 { 2 }	6.1272 { 3 }	7.3190 { 2 }	9.3992 { 2 }
3.5919 { 3 }	4.9593 { 2 }	6.1294 { 2 }	7.3341 { 2 1 }	9.4099 { 2 }
3.5924 { 2 }	4.9628 { 3 }	6.1318 { 3 }	7.3371 { 1 }	9.4311 { 1 }
3.5937 { 3 }	4.9669 { 2 }	6.1333 { 2 1 }	7.3378 { 2 }	9.4464 { 2 }
3.6014 { 2 }	4.9733 { 3 }	6.1430 { 3 }	7.3662 { 1 }	9.4929 { 1 }
3.6050 { 1 }	4.9763 { 2 }	6.1578 { 3 }	7.3725 { 2 }	9.4929 { 3 }
3.6211 { 1 }	4.9814 { 3 }	6.1612 { 2 }	7.3727 { 2 }	9.6042 { 2 }
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 II shows the selected populations, in correspondence to several values of ξ and with α=0.05. In particular, the structure of Table II is as follows:

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

repeated for various ξ's.

Table II

Confidence intervals

ξ	$\hat{U}_{\xi,\alpha}(1)\&\hat{L}_{\xi,\alpha}(1)$	$\hat{U}_{\xi,\alpha}(2)\&\hat{L}_{\xi,\alpha}(2)$	$\hat{U}_{\xi,\alpha}(3)\&\hat{L}_{\xi,\alpha}(3)$	π_{j^*}
0.4	0.00000	0.00000	0.00000	{ 1 or 2 or 3 }
	0.00000	0.00000	0.00000	
0.5	0.00000	0.00610	0.00000	{ 1 or 3 }
	0.00000	0.00000	0.00000	
0.6	0.00000	0.00610	0.00000	{ 1 or 3 }
	0.00000	0.00000	0.00000	
0.7	0.00000	0.01311	0.01291	{ 1 }
	0.00000	0.00000	0.00000	
0.8	0.00000	0.01311	0.01291	{ 1 }
	0.00000	0.00000	0.00000	
0.9	0.00000	0.01311	0.01291	{ 1 }
	0.00000	0.00000	0.00000	
1.0	0.01188	0.01916	0.01291	{ 1 }
	0.00000	0.00092	0.00000	
1.1	0.01188	0.02203	0.01291	{ 1 }
	0.00000	0.00207	0.00000	
1.2	0.01188	0.02757	0.01291	{ 1 }
	0.00000	0.00456	0.00000	
1.3	0.01188	0.03694	0.01738	{ 1 }
	0.00000	0.00941	0.00000	

1.4	0.02271	0.04960	0.02144	{ 3 }
	0.00000	0.01682	0.00000	
1.5	0.02881	0.05943	0.02144	{ 3 }
	0.00000	0.02305	0.00000	
1.6	0.02881	0.05943	0.02144	{ 3 }
	0.00000	0.02305	0.00000	
1.7	0.02881	0.06907	0.02831	{ 3 }
	0.00000	0.02947	0.00000	
1.8	0.02881	0.07982	0.03814	{ 1 }
	0.00000	0.03692	0.00096	
...	...	...	...	...
7.8	0.88625	0.89371	0.90715	{ 1 }
	0.79047	0.83063	0.81413	
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	
8.1	0.92239	0.91525	0.93037	{ 2 }
	0.83788	0.85715	0.84567	
8.2	0.93044	0.92324	0.93569	{ 2 }
	0.84893	0.86721	0.85315	
8.3	0.93841	0.92767	0.94944	{ 2 }
	0.86010	0.87284	0.87305	
8.4	0.94630	0.94077	0.95616	{ 2 }
	0.87142	0.88983	0.88315	
8.5	0.95875	0.95453	0.95948	{ 2 }
	0.88995	0.90828	0.88824	
8.6	0.96768	0.95958	0.97553	{ 2 }
	0.90391	0.91526	0.91424	
8.7	0.97060	0.96949	0.98163	{ 2 }
	0.90862	0.92940	0.92496	
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	
9.1	0.98748	0.98380	0.99691	{ 2 }
	0.93829	0.95140	0.95632	
9.2	0.99745	0.99401	1.00000	{ 2 }
	0.95984	0.96954	0.96873	
9.3	0.99955	0.99534	1.00000	{ 2 }
	0.96541	0.97222	0.97539	
9.4	1.00000	1.00000	1.00000	{ 3 }
	0.98098	0.98380	0.97930	
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 or 2 or 3 }
	1.00000	1.00000	1.00000.	

Interpretation of the results. The reader may have noticed that the set-up we have considered in our example presents some specific additional feature compared with the general framework described in Section 3. In particular, the fact that the observations x_i are defined as

$$x_i = \min\{t_{i1}, \cdots, t_{ik}\} \tag{4.1}$$

has important consequences on the interpretation of the results for, in this case, each one of the k samples

$$\{ x_i \mid i \in \{1, \cdots, n\}, x_i \text{ is generated by } G_j(u) \}, \tag{4.2}$$

by (4.1), is *not* a random samples drawn from $G_j(u)$, but a random sample from a different c.d.f., say $F_j(u)$, clearly with lower reliability for each u, namely,

$$F_j(\xi) \equiv K_j \int_{(0,\, \xi]} \frac{\lambda_j(u)}{\lambda_\Sigma(u)}\, dF_\Sigma(u) \tag{4.3}$$

for any $\xi > 0$, where K_j is a normalizing constant, $F_\Sigma(u)$ is the system lifelength distribution, and $\lambda_j(u)$, $\lambda_\Sigma(u)$ denote the failure rate of C_j and the system failure rate, respectively:

$$\lambda_j(u) \equiv \frac{dG_j(u)}{du} / [1 - G_j(u)], \quad \lambda_\Sigma(u) \equiv \sum_{j=1}^{k} \lambda_j(u). \tag{4.4}$$

The value $\int_{(0,\, \xi]} [\lambda_j(u)/\lambda_\Sigma(u)]\, dF_\Sigma(u)$ may be called *mean relative failure rate of* C_j up to time ξ and denoted by $MRFR_j(\xi)$. The values $K_j \equiv MRFR_j(\infty)$, $j = 1, \cdots, k$, may be considered mere constants which reduce the $MRFR_j(\xi)$'s to vary within the unit interval.

Recalling that, the α-minimax selected component C_{j^*} at time ξ is to be interpreted as the one for which $\hat{U}_{\xi,0.05}(j)$ is minimum w.r.t. j, where $\hat{U}_{\xi,0.05}(j)$ is the probable maximum unreliability $\overline{F}_j(\xi)$, we can rephrase our

interpretation in terms of G_j (lifelengths of the components), to which we are mostly interested, as follows: C_{j^*} is the component for which the probable maximum $MRFR_{j^*}(\xi)$ is the lowest.

In our specific example, C_1 appears to be, almost uniformly w.r.t. ξ, the "best" component in the sense specified above.

In Figures I and II the upper and lower limits of the confidence intervals, viewed as functions of ξ, are represented in order to get an idea about the shape of the c.d.f.'s $F_j(\xi)$.

In Table III, the values of the bounds $\hat{I}_j(u)$ and $\hat{S}_j(u)$, in correspondence to different values of u, are reported. According to inequality (3.8), for any fixed u, the set $[\hat{I}_j(u), \hat{S}_j(u)]$ certainly contains the unknown empirical c.d.f. $\hat{F}_j(u)$ relative to the c.d.f. $F_j(u)$.

The structure of Table III 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 some values of u.

Place of Fig. I

Place of Fig. II

Fig. I. *Plots of* $\hat{U}_{\xi,0.05}(1)$, $\hat{U}_{\xi,0.05}(2)$, $\hat{U}_{\xi,0.05}(3)$

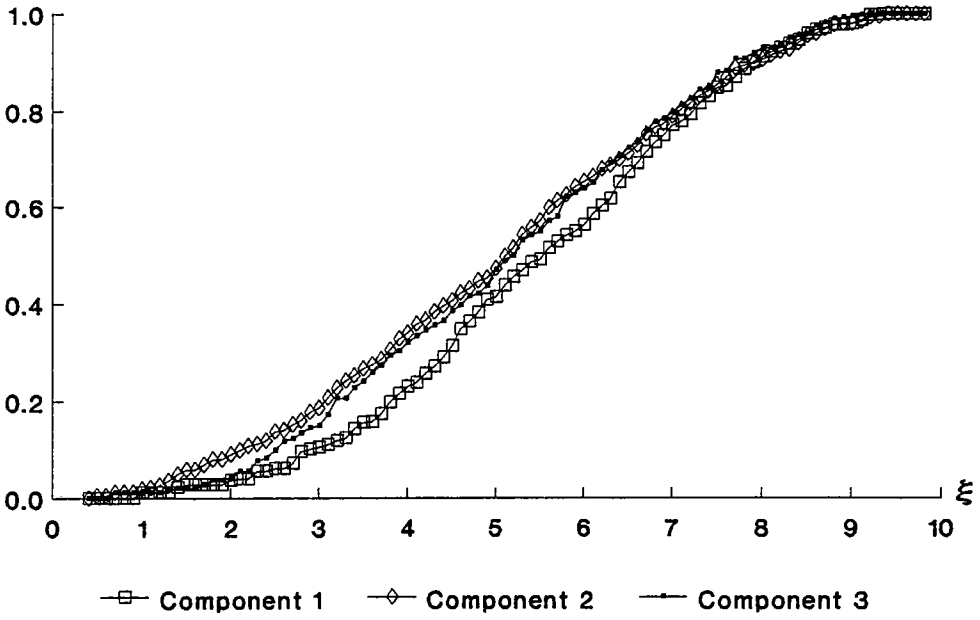

Fig. II. *Plots of* $\hat{L}_{\xi,0.05}(1)$, $\hat{L}_{\xi,0.05}(2)$, $\hat{L}_{\xi,0.05}(3)$

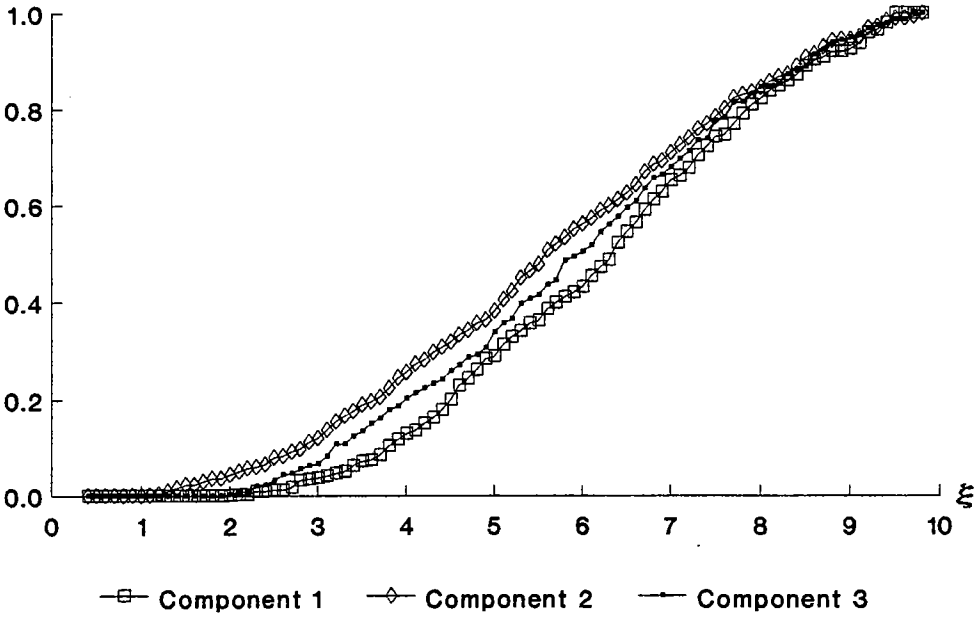

Table III

Bounds for the empirical c.d.f.'s

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
1.5	0.01754	0.04565	0.01395
	0.00984	0.03683	0.00376
2.0	0.02620	0.07576	0.03241
	0.01316	0.05915	0.01509
2.5	0.04783	0.12095	0.07870
	0.02640	0.09630	0.05283
3.0	0.09052	0.16989	0.12844
	0.05316	0.13569	0.08745
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
4.5	0.31020	0.39749	0.36726
	0.20139	0.32571	0.27451
5.0	0.41600	0.46584	0.45652
	0.28622	0.38654	0.35060
5.5	0.49805	0.56707	0.53846
	0.35507	0.48141	0.42510
6.0	0.57090	0.64940	0.62810
	0.42264	0.56487	0.51464
6.5	0.68000	0.70866	0.71084
	0.53876	0.62828	0.60776
7.0	0.77305	0.78378	0.78039
	0.64940	0.71546	0.69469
7.5	0.84429	0.84601	0.85496
	0.74590	0.79245	0.79909
8.0	0.90169	0.89474	0.90566
	0.83193	0.85563	0.87037
8.5	0.94352	0.94238	0.93609
	0.90517	0.92043	0.91163
9.0	0.96689	0.97032	0.97753
	0.93939	0.95474	0.96729
9.5	1.00000	0.99632	0.99627
	1.00000	0.99564	0.99531
9.8	1.00000	1.00000	1.00000
	1.00000	1.00000	1.00000.

5. DISCUSSION: THE SMALL SAMPLE CASE

While the advantages of the model with partially classified data should be readily apparent, it is also evident that it will, in general, require a larger amount of data than a similar situation with fully classified data. This need increases with increased unclassification. Thus, while small sample situations can certainly be considered, in practice, statistical analyses based on partially classified data will require relatively large amount of data to be meaningful, especially if the populations are many and the unclassification is extensive.

The reader may have noticed that, when we are dealing with large amount of data and *relatively* few unclassified or partially classified observations, there is no practical difference between using the α-minimax criterion or making a decision on the basis of the (unknown) empirical reliability function $1-\hat{F}_j(\xi)$ ($\sim 1-\hat{\hat{F}}_j(\xi)$). In fact, it is immediate to notice that, asymptotically, the α-minimax criterion (2.3) is equivalent to the following:

$$\min_{j \in \{1,\cdots, k\}} \hat{F}_j(\xi). \tag{5.1}$$

In other words, under the asymptotic conditions of our approximation, our procedure tends to the most intuitive selection rule which we may think of when dealing with fully classified data, consisting of choosing (one of) the population(s) with highest empirical reliability at the fixed time ξ, and, hence, by the Glivenko-Cantelli theorem:

$$\mathbb{P}[\ \lim_{\nu_j \to \infty} \sup_{u \in \mathbb{R}} |\ \hat{F}_j(u) - F_j(u)\ | = 0\ ] = 1, \tag{5.2}$$

it is asymptotically equivalent to

$$\min_{j \in \{1,\cdots, k\}} F_j(\xi). \tag{5.3}$$

We will refer to this property of our selection procedures as *consistency*, in consideration of the fact that the probability of a correct selection (PCS)

$$PCS \equiv \mathbb{P}[\, \hat{F}_i(\xi) \leq \min_{j \in \{1,\cdots,k\}} \hat{F}_j(\xi) \mid F_i(\xi) \leq \min_{j \in \{1,\cdots,k\}} F_j(\xi) \,] \tag{5.4}$$

increases to 1 as $\nu_j \to \infty$.

The agreement with the intuitive rule (5.1) for choosing the most reliable π_j at time ξ with fully classified data assures us that, at least asymptotically, the α-minimax criterion is justifiable and acceptable.

The fact that, asymptotically and with relatively small unclassification, using a procedure substantially based on the $\hat{\hat{F}}_j(\xi)$'s to make a decision is equivalent, to practical purposes, to using the (unknown) $\hat{F}_j(\xi)$'s should be intuitive and expected. However, our study yields a formal elicitation of the analytical conditions under which this fact occurs. Several problems, instead, arise in the small sample case, especially if the ν_j's are very different and there is extensive unclassification. (By small or moderate sample case, clearly, we mean a situation where *the ratio* n/k is small or moderate). In fact, even if we assume that confidence intervals for the loss function were available, their lengths would be affected by the values ν_j's, so that some confidence intervals could be relatively larger than others as a result of small values of ν_j. Thus, the comparisons of the underlying reliabilities would be disturbed by the circumstance that we possess interval estimates for them which have different degrees of precision.

In the small or moderate sample case, we turn, hence, to a different selection criterion which also generalizes, as the α-minimax criterion, the intuitive rule (5.1) we would apply if the data were fully classified and that, furthermore, takes into account the "prior" distributions $\underset{\sim}{p}_i$. In fact, as in the small sample case the interval $[\hat{I}_j(\xi), \hat{S}_j(\xi)]$ which contains $\hat{F}_j(\xi)$ is much larger, the necessity of distinguishing, within such an interval, the value(s) which, according to our prior beliefs, are most likely equal to the (unknown) $\hat{F}_j(\xi)$ becomes more impelling, while, when the amount of data is large, prior beliefs do not count very much.

Bearing in mind the above considerations, in the small sample case, we could consider the following empirical minimax type criterion:

$$\min_{j \in \{1,\cdots,k\}} y_j^*(\xi) \tag{5.5}$$

where $y_j^*(\xi)$ is such that

$$\mathbb{P}[\ \hat{F}_j(\xi) = y_j^*(\xi)\ ] = \max_{y \in \mathfrak{Y}_j(\mathfrak{Z})} \mathbb{P}[\ \hat{F}_j(\xi) = y\ ] \qquad (5.6)$$

and $\mathfrak{Y}_j(\mathfrak{Z}) \equiv \{y \mid \exists\ \underset{\sim}{p}$ such that the consequent classification $(PP_1^*, \cdots, PP_n^*)$ yields $\hat{F}_j^*(\xi) = y\}$, $\mathfrak{Y}_j(\mathfrak{Z}) \subset [\hat{I}_j(\xi), \hat{S}_j(\xi)]$, which consists in choosing (one of) the population(s) π_j such that the most probable (a priori) value assumed by the unknown c.d.f. $\hat{F}_j(\xi)$ is minimum. In such a way, we include the prior information in our analysis by using it to weight the various possible values in the interval $[\hat{I}_j(\xi), \hat{S}_j(\xi)]$.

Virtually, problem (5.2) can always be solved since, clearly, the set $\mathfrak{Y}_j(\mathfrak{Z})$ is finite, although, in practice, computational problems may arise when n is large (especially if also k is large).

Recalling that the α-minimax criterion is asymptotically equivalent to the intuitive rule (5.1) and that, by (3.9), the length of $[\hat{I}_j(\xi), \hat{S}_j(\xi)]$ tends to zero, clearly, the above criterion (5.2) is also asymptotically equivalent to the α-minimax and, hence, consistent.

Example 5.1. Below, we show an example where we analyze the data contained in Table IV, which can be thought of as the outcome of a life-test conducted on 25 two-component series systems.

Table IV

Simulated Data

x_i	PP_i	x_i	PP_i	x_i	PP_i	x_i	PP_i	x_i	PP_i
0.3028	{ 1 }	4.3360	{ 1 }	7.7756	{ 1 }	9.2408	{ 2 }	10.0576	{ 2 }
1.9814	{ 1 2 }	4.8281	{ 1 2 }	8.1553	{ 2 1 }	9.3783	{ 2 1 }	10.1221	{ 2 }
3.5253	{ 1 }	5.1428	{ 1 2 }	9.1514	{ 2 }	9.5131	{ 2 1 }	10.2950	{ 2 1 }
3.7418	{ 1 }	7.3832	{ 1 2 }	9.2050	{ 1 2 }	9.9456	{ 2 1 }	11.1203	{ 2 }
4.2188	{ 1 2 }	7.5144	{ 1 2 }	9.2304	{ 2 1 }	9.9936	{ 2 1 }	11.7295	{ 2 }.

We suppose that we are interested in comparing the two populations (lifelengths of the two components) at time $\xi = 7$. For simplicity, we assume that the prior distributions $\underset{\sim}{p}_i \equiv (p_i(1),\ p_i(2)) \equiv \underset{\sim}{p}$, *for the unclassified observations* $\{x_i \mid i \in \{1, \cdots, n\},\ |PP_i| = 2\}$, are constant w.r.t. i, and show in

Tables V – VIII, the distribution of $\hat{F}_1(\xi)$ and $\hat{F}_2(\xi)$ for various choices of $\underset{\sim}{p} \equiv (p(1),\ p(2))$. Clearly, the probability values in Tables V – VIII have been computed as

$$\mathbb{P}[\ \hat{F}_j(\xi) = y\ ] = \sum_{\mathcal{Z}} \mathbb{P}[\ (PP_1^*, \cdots, PP_n^*)\ ]\ I_{\{y\}}(\hat{F}_j^*(\xi)) \qquad (5.7)$$

where

$$\mathbb{P}[\ (PP_1^*, \cdots, PP_n^*)\ ] = \prod_{i=1}^{n} \sum_{j=1}^{k} p_i(j)\ I_{PP_i^*}(j). \qquad (5.8)$$

We have processed the data also for $\underset{\sim}{p} \equiv (1,\ 0)$ and for $\underset{\sim}{p} \equiv (0,\ 1)$ and obtained:

$$\mathbb{P}[\ \hat{F}_1(\xi) =\ 0.4211\ ] = 1,\ \mathbb{P}[\ \hat{F}_2(\xi) = 0\ ] = 1$$

in the first case and

$$\mathbb{P}[\ \hat{F}_1(\xi) =\ 0.8\ ] = 1,\ \mathbb{P}[\ \hat{F}_2(\xi) = 0.2\ ] = 1$$

in the second case.

Analyzing the data for different choices of the prior distributions, as we have done below, allows the user to control in part the effect of his prior beliefs and to find out at what extent the decision that the outcome of the procedure suggests is influenced by his own prior information. This is important in presence of small samples, when it is clear that, especially in situations in which the differences of the reliabilities at the fixed time are not marked, wrong prior beliefs could easily result in incorrect decisions.

The following are the values of $y_j^*(\xi)$ obtained in correspondence of the various prior distributions considered with $\xi = 7$, cf., Tables V – VIII:

$y_1^*(\xi) = 0.4211,$	$y_2^*(\xi) = 0$	with $\underset{\sim}{p} \equiv (1,\ 0)$
$y_1^*(\xi) = 0.5,$	$y_2^*(\xi) = 0$	with $\underset{\sim}{p} \equiv (0.6,\ 0.4)$
$y_1^*(\xi) = 0.5455,$	$y_2^*(\xi) = 0.2$	with $\underset{\sim}{p} \equiv (0.4,\ 0.6)$
$y_1^*(\xi) = 0.8,$	$y_2^*(\xi) = 0.2$	with $\underset{\sim}{p} \equiv (0,\ 1).$

For all the prior distributions considered, we have obtained

$$y_1^*(\xi) > y_2^*(\xi).$$

Thus, if our prior beliefs are well represented by either one of the above four distributions, we should decide, according to (5.2), that the second component has a (probable) maximum MRFR(7) lower than the first because the most probable value of $\hat{F}_2(7)$ is lower than the most probable value of $\hat{F}_1(7)$. This is, we think, also the decision that any user would intuitively make just on the basis of the data in Table IV, without any further statistical analysis beyond the simple observation that all the fully classified x_i's which are less than $\xi = 7$ are due to failure of the first component.

We should, also, be aware that our decision is certainly influenced by the assumption of constancy of the $\underset{\sim}{p}_i$'s of the unclassified observations and that certainly there exist some distributions which would make us change our decision, even though we may retain that such distribution are not reasonable. In fact, by Lemma 1, there must exist two distribution such that

$$y_1^*(\xi) = \hat{I}_1(\xi) = 0.2667 \text{ and } y_2^*(\xi) = \hat{S}_2(\xi) = 0.4$$

or, conversely,

$$y_1^*(\xi) = \hat{S}_1(\xi) = 0.8889 \text{ and } y_2^*(\xi) = \hat{I}_2(\xi) = 0.$$

In particular, by (3.6) and (3.7), such distributions clearly are:

$\underset{\sim}{p}_i = (0, 1)$ for all unclassified $x_i \le u$, $\underset{\sim}{p}_i = (1, 0)$ for all unclassified $x_i > u$

$\underset{\sim}{p}_i = (1, 0)$ for all unclassified $x_i \le u$, $\underset{\sim}{p}_i = (0, 1)$ for all unclassified $x_i > u$,

respectively. In fact, it can easily be verified from Table IV that, if we assign the $\lambda_j(\xi) = 4$ unclassified observations less than $\xi = 7$ to π_2 and the remaining $\rho_j(\xi) = 10$ unclassified observations greater than $\xi = 7$ to π_1, as dictated by the first of the two above distributions, we have

$$\hat{F}_1(7) = \frac{4}{4+11} = 0.2667, \text{ and } \hat{F}_2(7) = \frac{4}{4+6} = 0.4,$$

while, if we assign the $\lambda_j(\xi) = 4$ unclassified observations less than $\xi = 7$ to π_1

and the remaining $\rho_j(\xi) = 10$ unclassified observations greater than $\xi = 7$ to π_2, as dictated by the second prior distribution, we obtain

$$\hat{F}_1(7) = \frac{8}{8+1} = 0.8889, \text{ and } \hat{F}_2(7) = \frac{0}{0+16} = 0.$$

Note that the assumption of constancy of the $\underset{\sim}{p}_i$'s of the unclassified observations w.r.t. i could be justified, for instance, in a situation where our prior beliefs (or information drawn from previous experiments) make us think that the failure rates functions $\lambda_j(u)$, $j = 1, 2$, of the $F_j(u)$'s remain constant (at least approximately) w.r.t. the time. In this case, our prior beliefs about the relative magnitude of the failure rate λ_j could be represented by $p_i(j) \equiv \lambda_j/(\lambda_1 + \lambda_2) = \lambda_j/\lambda_\Sigma$ for every unclassified x_i. Clearly, different beliefs about the failure rate functions (increasing, decreasing, bathtub shaped) can be reflected through appropriate choices of the $\underset{\sim}{p}_i$'s.

Table V

Possible values for $\hat{F}_1(7)$ *and corresponding probabilities when* $\underset{\sim}{p} \equiv (0.6, 0.4)$

.2667	0.00015479	.5000	0.20210456
.2857	0.00103196	.5333	0.03250662
.3077	0.00309587	.5385	0.06934745
.3125	0.00092876	.5455	0.03852636
.3333	0.01169550	.5556	0.00652298
.3529	0.00208971	.5714	0.02627708
.3571	0.01857521	.5833	0.03852636
.3636	0.00642106	.6000	0.01467671
.3750	0.01393141	.6154	0.01444738
.3846	0.03302259	.6250	0.00163075
.3889	0.00208971	.6364	0.01467671
.4000	0.04693107	.6667	0.00921321
.4118	0.01393141	.7000	0.00366918
.4167	0.03852636	.7143	0.00024159
.4211	0.00078364	.7273	0.00137594
.4286	0.07430084	.7500	0.00054358
.4375	0.04179422	.7778	0.00054358
.4444	0.00807808	.8000	0.00020653
.4545	0.03082109	.8333	0.00001611
.4615	0.08668431	.8571	0.00003624
.4667	0.07430084	.8750	0.00003624
.4706	0.01567283	.8889	0.00001359.

Table VI

Possible values for $\hat{F}_2(7)$ *and corresponding probabilities when* $\underset{\sim}{p} \equiv (0.6, 0.4)$

.0000	0.12960000	.1818	0.07430084
.0588	0.00003624	.1875	0.00652298
.0625	0.00054358	.2000	0.05891973
.0667	0.00366918	.2105	0.00004027
.0714	0.01467671	.2143	0.03082109
.0769	0.03852636	.2222	0.01420320
.0833	0.06934745	.2308	0.03852636
.0909	0.08668431	.2353	0.00108716
.1000	0.07430084	.2500	0.03796611
.1111	0.04183046	.2667	0.00513685
.1176	0.00054358	.2727	0.01857521
.1250	0.01760058	.2857	0.00642106
.1333	0.01467671	.3000	0.00619174
.1429	0.04061607	.3077	0.00550377
.1538	0.06934745	.3333	0.00402463
.1579	0.00001611	.3636	0.00103196
.1667	0.08692590	.4000	0.00015479.
.1765	0.00163075		

Table VII

Possible values for $\hat{F}_1(7)$ *and corresponding probabilities when* $\underset{\sim}{p} \equiv (0.4, 0.6)$

.2667	0.00001359	.5000	0.20210456
.2857	0.00020384	.5333	0.00285380
.3077	0.00137594	.5385	0.03082109
.3125	0.00003624	.5455	0.08668431
.3333	0.00604735	.5556	0.07430084
.3529	0.00003624	.5714	0.02080968
.3571	0.00366918	.5833	0.03852636
.3636	0.01444738	.6000	0.07430084
.3750	0.00054358	.6154	0.00642106
.3846	0.01467671	.6250	0.04179422
.3889	0.00001611	.6364	0.03302259
.4000	0.02967447	.6667	0.05252226
.4118	0.00024159	.7000	0.01857521
.4167	0.03852636	.7143	0.01393141
.4211	0.00000268	.7273	0.00309587
.4286	0.01467671	.7500	0.01393141
.4375	0.00163075	.7778	0.00619174
.4444	0.03254688	.8000	0.00181560
.4545	0.06934745	.8333	0.00208971
.4615	0.03852636	.8571	0.00208971
.4667	0.00652298	.8750	0.00092876
.4706	0.00027179	.8889	0.00015479.

Table VIII

Possible values for $\hat{F}_2(7)$ *and corresponding probabilities when* $\underset{\sim}{p} \equiv (0.4, 0.6)$

.0000	0.02560000	.1818	0.01467671
.0588	0.00092876	.1875	0.07430084
.0625	0.00619174	.2000	0.09113713
.0667	0.01857521	.2105	0.00522428
.0714	0.03302259	.2143	0.06934745
.0769	0.03852636	.2222	0.01621641
.0833	0.03082109	.2308	0.03852636
.0909	0.01712283	.2353	0.02786281
.1000	0.00652298	.2500	0.04721956
.1111	0.00372046	.2667	0.02600529
.1176	0.01393141	.2727	0.00366918
.1250	0.04203581	.2857	0.01444738
.1333	0.07430084	.3000	0.00054358
.1429	0.08670042	.3077	0.00550377
.1538	0.06934745	.3333	0.00141218
.1579	0.00208971	.3636	0.00020384
.1667	0.05245777	.4000	0.00001359.
.1765	0.04179422		

Finally, it may be interesting to compare the previous results with the decision made using the α-minimax criterion, although, in this case, all the reservations and criticisms made previously for this method in presence of small samples apply fully. To this purpose, in Table IX we have reported the values of the $\hat{L}_{\xi,\alpha}(j)$'s and $\hat{U}_{\xi,\alpha}(j)$'s, with $\alpha = 0.05$, computed according to formulae (3.12), (3.13), and (3.14). In table X, confidence intervals are computed by (3.15) and (3.16), which take into account a correction for continuity.

Table IX

Confidence intervals

ξ	$\hat{U}_{\xi,\alpha}(1)\&\hat{L}_{\xi,\alpha}(1)$	$\hat{U}_{\xi,\alpha}(2)\&\hat{L}_{\xi,\alpha}(2)$	selected population
0.3	0.00000	0.00000	{ 1 or 2 }
	0.00000	0.00000	
1.0	0.41751	0.00000	{ 2 }
	0.00000	0.00000	
2.0	0.54135	0.27750	{ 2 }
	0.00000	0.00000	
3.0	0.54135	0.27750	{ 2 }

	0.00000	0.00000	
4.0	0.84881	0.27750	{ 2 }
	0.00000	0.00000	
5.0	0.99733	0.46487	{ 2 }
	0.12767	0.00000	
6.0	1.00000	0.52007	{ 2 }
	0.14484	0.00000	
7.0	1.00000	0.52007	{ 2 }
	0.14484	0.00000	
8.0	1.00000	0.59648	{ 2 }
	0.28775	0.00000	
9.0	1.00000	0.62415	{ 2 }
	0.30992	0.00000	
10.0	1.00000	0.93654	{ 2 }
	0.67440	0.13864	
11.0	1.00000	1.00000	{ 1 }
	1.00000	0.45369	
12.0	1.00000	1.00000	{ 1 or 2 }
	1.00000	1.00000.	

Table X

Corrected confidence intervals

ξ	$\hat{U}_{\xi,\alpha}(1)\&\hat{L}_{\xi,\alpha}(1)$	$\hat{U}_{\xi,\alpha}(2)\&\hat{L}_{\xi,\alpha}(2)$	selected population
0.4	0.64559	0.48319	{ 2 }
	0.00000	0.00325	
1.0	0.64559	0.48319	{ 2 }
	0.00000	0.00325	
2.0	0.69719	0.55262	{ 2 }
	0.00024	0.00000	
3.0	0.69719	0.55262	{ 2 }
	0.00024	0.00000	
4.0	0.83890	0.55262	{ 2 }
	0.07186	0.00000	
5.0	0.91148	0.63518	{ 2 }
	0.14235	0.00189	
6.0	0.91812	0.66187	{ 2 }
	0.15066	0.00955	
7.0	0.91812	0.66187	{ 2 }
	0.15066	0.00955	
8.0	0.96132	0.70001	{ 2 }
	0.21837	0.02393	
9.0	0.96636	0.71411	{ 2 }
	0.22862	0.03027	
10.0	0.99962	0.88049	{ 2 }
	0.38092	0.15527	
11.0	0.98132	0.97965	{ 2 }
	0.45376	0.31965.	

Table XI, gives the upper and lower bound of the unknown empirical c.d.f.'s $\hat{F}_j(u)$ for some values of u. In this case, the influence of the prior information is much more important, due to the great lengths of the intervals $[\hat{I}_j(u), \hat{S}_j(u)]$.

Table XI

Exact bounds for the empirical c.d.f.'s

u	$\hat{S}_1(u)\&\hat{I}_1(u)$	$\hat{S}_2(u)\&\hat{I}_2(u)$
0.3	0.00000	0.00000
	0.00000	0.00000
1.0	0.20000	0.00000
	0.05263	0.00000
2.0	0.33333	0.14286
	0.05556	0.00000
3.0	0.33333	0.14286
	0.05556	0.00000
4.0	0.66667	0.14286
	0.16667	0.00000
5.0	0.87500	0.33333
	0.25000	0.00000
6.0	0.88889	0.40000
	0.26667	0.00000
7.0	0.88889	0.40000
	0.26667	0.00000
8.0	1.00000	0.50000
	0.38462	0.00000
9.0	1.00000	0.53846
	0.41667	0.00000
10.0	1.00000	0.78947
	0.83333	0.28571
11.0	1.00000	0.90000
	1.00000	0.66667
12.0	1.00000	1.00000
	1.00000	1.00000.

The main difficulty we have encountered in our attempts of making inferences about the underlying distributions $F_j(u)$'s was the absence of any kind of assumptions about them, which has as a consequence the fact that we cannot say much more than giving an interval $[\hat{I}_j(u), \hat{S}_j(u)]$ which certainly contains $\hat{F}_j(u)$. In order to be able to say something more about the probability of the points in $[\hat{I}_j(u), \hat{S}_j(u)]$ to be equal to the (unknown)

empirical c.d.f., we had to introduce some prior knowledge about the probability of each value x_i being generated by each one of the populations π_j. The introduction of additional information was done in this way in order to remain within a nonparametric framework. Another way could have been that one of considering a parametric set-up, by making some distributional hypotheses on the $\hat{F}_j(u)$'s. The approach with prior probabilities has the advantage to express explicitly the probability of each x_i being from π_j, while, in a parametric approach, the additional information we need is somehow incorporated in the parametric model. On the other hand, it has the disadvantage to put the user in the condition to have to express explicitly his degree of belief of each value being from each π_j, which can be more embarrassing and may seem more arbitrary than making some distributional assumption. We have, however, seen in Example 5.1 that this arbitrariness may be partially controlled by considering various alternative prior distributions and comparing the results.

ACKNOWLEDGMENTS

The authors thank Professor TaChen Liang and Dr. Yuning Liao, for a critical reading of an earlier draft of this paper, and Professor Klaus Miescke and Dr. Re-Bin Rau, for help during the first revision.

The research of S. S. Gupta was supported in part by the NSF Grant DMS-8923071.

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

BIBLIOGRAPHY

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.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*, John Wiley & Sons, Inc., 13-14.

Fong, D. K. H. and Berger, J. O. (1993). Ranking, estimation and hypothesis testing in unbalanced two-way additive models: a Bayesian approach. *Statistics and Decisions*, **11**, 1-24.

Gastaldi, T. (1993). Improved maximum likelihood estimation for component reliabilities with Miyakawa-Usher-Hodgson-Guess' estimators under censored search for the cause of failure. To appear in *Statistics and Probability Letters*, 18, No. 5.

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 an 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.