EVALUATING THE CHOSEN POPULATION: A BAYES AND MINIMAX APPROACH*

Harold Sackrowitz

Rutgers University

and

Ester Samuel-Cahn

The Hebrew University

One of K populations is chosen according to some given selection procedure. Population i has the parameter θ_i associated with it. The θ value for the chosen population is to be estimated using only the data on which the selection was made. Some general results relating to Bayes and minimax rules as well as the minimax value are obtained. Applications of these results to particular problems are given.

1. Introduction.

One is often faced with the problem of having to choose one of a number of options. Once data are collected and criteria established the selection rule is usually straightforward. More often than not the selection procedure will be a simple function of the order statistics. For example, a manufacturer would simply select the machine from among K machines that was most productive during some trial period. It is often important to be able to say something about the likely result of the chosen option. The above manufacturer would want an estimate of the expected mean output of the selected machine.

From a practical point of view the main concern in problems such as these is the possibility of overestimation if only data from the chosen

*Research supported, in part, by NSF-DMS-84-18416.
AMS 1980 subject classifications. Primary 62C10, 62C20; Secondary 62F99.
Key words and phrases. Bayes, minimax, selected population.

population are used. It seems clear that a product's performance in a contest in which it has won is likely to be above its mean performance. The problem of inference after selection has received relatively little attention in the literature. Our bibliography and references therein contain the bulk of the work on this problem. When it has been studied it is usually with respect to some particular problem.

In this paper we develop some basic theory regarding minimax decision making for such problems. Minimaxity seems to be a well suited criterion in this setting. Unbiasedness on the other hand does not seem to be a good criterion in such problems. Putter and Rubinstein (1968) have shown that no unbiased estimator exists in the normal case and Sackrowitz and Samuel-Cahn (1984) have shown that the U.M.V.U.E. is inadmissible in the negative exponential case. The results given here yield methods for finding and identifying minimax estimators in some cases as well as answering questions left open in previous works in this area. We only consider continuous random variables so that ties can be avoided. We remark that different methods for breaking ties can lead to different minimax values.

2. Notation and definitions.

Our model is as follows. Let $\pi_1,\ldots,\pi_K$ be K populations, π_j characterized by some unknown parameter $\theta_j \in \Theta$, $j=1,\ldots,K$, for some set Θ. Let X_i, $i=1,\ldots,K$, be independent continuous random variables with density function $f(x|\theta_i)$. Let $\underline{X} = (X_1,\ldots,X_K)$ and $\underline{\theta} = (\theta_1,\ldots,\theta_K)$. $I(\underline{X})$ is called a selection rule if

$$I(\underline{X}) = j \quad \text{if } \underline{X} \in A_j,\ j=1,\ldots,K. \tag{2.1}$$

where A_j is a given partition of the space of $\underline{X}$ values. This terminology implies that population π_j is selected if $I(\underline{X}) = j$. We wish to make inference about $\theta_{I(\underline{X})}$ relative to the loss $L(a,\theta_{I(\underline{X})})$, which notedly is a random variable. A decision rule will be $\psi(\underline{X})$ with the risk function $R_I(\psi,\underline{\theta}) = E_{\underline{\theta}} L(\psi(\underline{X}),\theta_{I(\underline{X})})$

which is not a random variable. For a joint prior distribution, $\underline{G}$, over the space of vectors $\underline{\theta}$ denote by $G_I(.|\underline{x})$ the posterior distribution of $\theta_{I(\underline{X})}$ given $\underline{X} = \underline{x}$. The Bayes risk of ψ relative to $\underline{G}$ is denoted by $r_I(\psi,\underline{G})$ while the Bayes rule with respect to $\underline{G}$ will be denoted $\psi^I_{\underline{G}}(\underline{X})$. For short write $r^*_I(\underline{G})$ instead of $r_I(\psi^I_{\underline{G}},\underline{G})$.

We will find it very useful to relate this problem to the more standard inference problem which we shall call the "component problem." In the component problem the random variable X has density $f(x|\theta)$. The loss function is $L(a,\theta)$, a decision rule will be denoted by $\phi(x)$, $R(\phi,\theta) = E_\theta L(\phi(X),\theta)$ is the risk function. If G is a prior distribution on θ then the posterior distribution of θ given $X = x$ is denoted by $G(\cdot|x)$. The Bayes risk of ϕ with respect to G is $r(\phi,G)$, ϕ_G is a Bayes rule with respect to G and $r(\phi_G,G)$ is abbreviated to $r^*(G)$. Lastly the posterior risk of ϕ is $r_G(\phi,x) = E\{L(\phi(X),\theta)|X = x\}$ and $r_G(\phi_G,x)$ is abbreviated to $r_G(x)$.

3. Theoretical results.

We begin by exploring the relationship between Bayes rules for the selected parameter and component problems for independent priors. Lemmas 3.1 and 3.2 are stated without proof as they essentially follow from definitions.

LEMMA 3.1. Let the joint prior distribution $\underline{G}(\underline{\theta}) = \prod_{j=1}^{K} G_j(\theta_j)$ so that the θ_j are independent. Then the posterior distribution of $\theta_{I(X)}$ given $\underline{X} = \underline{x}$ satisfies

$$G_I(\cdot|\underline{x}) = G_j(\cdot|x_j) \text{ for } I(\underline{x}) = j. \tag{3.1}$$

LEMMA 3.2. Let $I(\cdot)$ be a selection rule and assume that $\underline{G}(\underline{\theta}) = \prod_{j=1}^{K} G_j(\theta_j)$. Then for each $j=1,\ldots,K$

$$\psi^I_{\underline{G}}(\underline{x}) = \phi_{G_j}(x_{I(\underline{x})}) \text{ for } I(\underline{x}) = j. \tag{3.2}$$

An interpretation of (3.2) is that when the θ_j are independent the

Bayes rule ignores the selection procedure in the sense that the component Bayes rule is always used (together with the observation of the selected population). A case of particular interest is $G_j = G$ all $j=1,\ldots,K$ so that the θ_j are i.i.d. each with c.d.f. G. Then (3.2) becomes $\psi_{\underline{G}}^{I}(\underline{x}) = \phi_G(x_{I(\underline{x})})$. It should be noted that even in this case it does not follow that $r_I^*(\underline{G}) = r^*(G)$. The following theorem given necessary and sufficient conditions for this to be true for all selection rules.

THEOREM 3.1. Let θ_j be i.i.d. with distribution G. A necessary and sufficient condition for the Bayes rule for $\theta_{I(\underline{X})}$, given in (3.2), to have Bayes risk equal to $r^*(G)$ for every selection rule $I(\cdot)$, is that the posterior risk for the Bayes component problem, $r_G(x)$, not depend on x.

Proof. Sufficiency follows immediately, since by Lemma 3.1 the posterior risk of (3.2) the Bayes rule of $\theta_{I(\underline{X})}$, is just $r_G(x_{I(\underline{x})})$. To prove necessity, suppose $r_G(x)$ is not constant (on a set with positive probability under the marginal distribution of X.) Without loss of generality take K = 2 and let $I(\underline{x}) = 1$ if $r_G(x_1) > r_G(x_2)$ and $I(\underline{x}) = 2$ otherwise. Also let the set $A^* = \{\underline{x} : I(\underline{x}) = 1\}$ and let f( ) denote the marginal density of X. Then simple manipulations of the integral yields

$$r_I^*(\underline{G}) = r^*(G) + \int_{A^*}[r_G(x_1) - r_G(x_2)]f(x_1)f(x_2)dx_1dx_2 > r^*(G)$$

by our assumption.

Remark. An interesting phenomenon occurs in the related problem of estimating all parameters in the following way. Let $I_1(\underline{x}),\ldots,I_K(\underline{x})$ be a set of selection rules such that for each $\underline{x}$, $I_i(\underline{x}) \neq I_j(\underline{x})$ for $i \neq j$ (i.e., no population is chosen by more than one rule). Let $\phi(\underline{x})$ be any decision rule for the component problem, and suppose $\psi^i(\underline{x}) = \phi(x_{I_i})$ is used to decide on θ_{I_i}. One may then be misled to believe that $\sum_{i=1}^{K} R_{I_i}(\psi^i,\underline{\theta}) = \sum_{i=1}^{K} R(\phi,\theta_i)$, arguing, for example, that "the order of the presentation of the problems can make no difference." This is not correct. The somewhat surprising fact is that the above statement is correct, for any ϕ, if the risk function is replaced by the Bayes risks. That is, if $\underline{G}$ is a prior distribution which

renders θ_j independent, θ_j with distribution G_j, then

$$\sum_{i=1}^{K} r_{I_i}(\psi^i,\underline{G}) = \sum_{i=1}^{K} r(\phi,G_i).$$

We can now turn our attention towards minimaxity.

THEOREM 3.2. A sufficient condition for $\psi(\underline{x})$ to be minimax for $\theta_{I(\underline{X})}$ is that there exists a sequence of Bayes rules (with respect to G_n) for the component problem, whose posterior risks are independent of x, and such that

$$R_I(\psi,\underline{\theta}) \leq \lim_{n\to\infty} r^*(G_n) \text{ for all } \underline{\theta}.$$

Proof. This follows immediately from Theorem 3.1 and Theorem 2 of Ferguson (1967), p. 90.

Up to now we have considered a fixed number, K, of populations. In the remainder of this section we will be interested in the minimax value as a function of K. We will be using a fixed order statistic as selection rule although it will be seen that the results obtained can be extended to other types of selection rules. To avoid confusion we set $\underline{X}_K = (X_1,\ldots,X_K)$, $\underline{\theta}_K = (\theta_1,\ldots,\theta_K)$, etc. To be specific we shall consider decisions about the parameter of the maximal observation only, and let $J_K = J_K(\underline{X}_K)$ denote the index i such that $X_i > X_j$ for all $j \neq i$, $j=1,\ldots,K$. For the remainder of this section we consider a fixed loss function, $L(a,\theta)$, bounded from below, and hence without loss of generality assumed to satisfy $L(a,\theta) \geq 0$. Let v_K denote the minimax value for deciding on θ_{J_K} and assume $v_K < \infty$, all K. Sometimes it is useful to think of $X_1,\ldots,X_K$ as the observations at hand at time K.

Theorem 3.3 states that under a simple condition $v_K \leq v_{K+1}$. The content and proof of this theorem is a formalization of the following intuitive reasoning, from Nature's point of view: If the distributions are such that by proper choice of θ_{K+1} Nature can make the probability that $J_{K+1}(\underline{X}_{K+1}) = J_K(\underline{X}_K)$ as close to one as desired, then in the K+1 dimensional problem Nature would be no worse off than in the K dimensional problem.

THEOREM 3.3. Let $X_1, X_2, \ldots$ be independent continuous r.v.s. where X_i has density $f(\cdot|\theta_i)$. Suppose the family of distributions $F(\cdot|\theta)$, $\theta \in \Theta$, satisfies

(I) For every $\varepsilon > 0$ if x is such that $P_\theta(X < x) > 0$ for some θ then there exists a $\theta_0 = \theta_0(x,\varepsilon)$ such that $P_{\theta_0}(X < x) \geq 1-\varepsilon$.

If $L(a,\theta) \geq 0$, then $v_K \leq v_{K+1}$, K=1,2,... .

Proof. We first give a proof which assumes that least favorable distributions exist (as proper priors) and that Bayes rules exist for all proper priors. In particular let $\underline{G}_K^*$ denote a least favorable distribution for the K dimensional problem. Also let ϕ_i^*, i=K,K+1 denote a minimax rule for the i dimensional problem. We note that ϕ_K^* is Bayes with respect to $\underline{G}_K^*$. Fix θ and define the K+1 dimensional distribution $H_\theta(\underline{\theta}_{K+1})$ as follows: $\underline{\theta}_K = (\theta_1,\ldots,\theta_K)$ have joint distribution $\underline{G}_K^*$ and are independent of θ_{K+1} which is equal to θ with probability 1. Let δ_{K+1}^θ denote the Bayes rule with respect to H_θ (for all K+1 dimensional problem). We define the set $A = \{\underline{x}_{K+1} : J_K(\underline{x}_K) = J_{K+1}(\underline{x}_{K+1})\}$ and examine properties of δ_{K+1}^θ. Since Bayes rules can be obtained by minimizing the posterior risk it follows from the definition of H_θ that

$$(3.3) \qquad \delta_{K+1}^\theta(\underline{x}_{K+1}) = \begin{cases} \phi_K^*(\underline{x}_K) & \text{for } \underline{x}_{K+1} \in A \\ \text{irrelevant} & \text{otherwise} \end{cases}$$

and

$$E_{H_\theta}\{L(\delta_{K+1}^\theta(\underline{X}_{K+1}),\theta_{J_{K+1}(\underline{X}_{K+1})})|\underline{X}_{K+1}\}$$

$$(3.4)\quad \begin{cases} = E_{\underline{G}^*_K}\{L(\phi_K(\underline{X}_K),\theta_{J_K(\underline{X}_K)})|X_K\} & \text{if } \underline{X}_{K+1} \varepsilon A \\ \geq 0 & \text{otherwise} \end{cases}$$

The Bayes risk is the expectation of the l.h.s. of (3.4) with respect to the marginal density. The marginal densities (for the K and K+1 dimensional problems, resp.) are

$$f_{\underline{G}^*_K}(\underline{x}_K) = \int_{\Theta^K} \prod_{i=1}^{K} f(x_i|\theta_i)dG^*_K(\underline{\theta}) \text{ and } f_{H_\theta}(\underline{x}_{K+1}) = f_{\underline{G}^*_K}(\underline{x}_K)f(x_{K+1}|\theta).$$

By the definition of the minimax value

$$(3.5)\qquad v_{K+1} = \sup_{\underline{G}_{K+1}} \inf_{\delta} r(\delta,\underline{G}_{K+1}) \geq r(\delta^{\theta}_{K+1},H_\theta) \quad \text{for all } \theta \varepsilon \Theta.$$

Also let $\varepsilon > 0$ be given. Since the r.v.s. are continuous

$$(3.6)\qquad \begin{aligned} v_K &= \int_{\chi^K} E_{G^*_K}\{L(\phi^*_K(\underline{x}_K),\theta_{J_K(\underline{x}_K)})|\underline{x}_K\}\ f_{\underline{G}^*_K}(\underline{x}_K) \prod_{i=1}^{K} dx_i \\ &= \lim_{y\to-\infty} \int_y^\infty\!\!\cdots\!\int E_{G^*_K}\{L(\phi^*_K(\underline{x}_K),\theta_{J_K(\underline{x}_K)})|\underline{x}_K\}\ f_{\underline{G}^*_K}(\underline{x}_K) \prod_{i=1}^{K} dx_i. \end{aligned}$$

Thus for given $\varepsilon > 0$ there exist a $y_0 = y_0(\varepsilon)$ such that the value of the integral on the r.h.s. of (3.6) is $\geq v_K-\varepsilon$ and $P_\theta(X < y_0) > 0$ for some θ. Now let $\theta_0 = \theta_0(y_0,\varepsilon)$ be defined through condition (I). Then

$$\begin{aligned} r(\delta^{\theta_0}_{K+1},H_{\theta_0}) &\geq \int_A\!\int E_{\underline{G}^*_K}\{L(\phi^*_K(\underline{x}_K),\theta_{J_K(\underline{x}_K)})|\underline{x}_K\}\ f_{\underline{G}^*_K}(\underline{x}_K)\ f(x_{K+1}|\theta_0)\prod_{i=1}^{K+1} dx_i \\ (3.7)\qquad &\geq \int_{-\infty}^{y_0}\Big[\int_{y_0}^{\infty}\!\!\cdots\!\int E_{\underline{G}^*_K}\{L(\phi^*_K(\underline{x}_K),\theta_{J_K(\underline{x}_K)})|\underline{x}_K\}\ f_{\underline{G}^*_K}(\underline{x}_K)\prod_{i=1}^{K} dx_i\Big]\ f(x_{K+1}|\theta_0)\ dx_{K+1} \\ &\geq \int_{-\infty}^{y_0}(v_K-\varepsilon)\ f(x_{K+1}|\theta_0)dx_{K+1} \geq (1-\varepsilon)(v_K-\varepsilon). \end{aligned}$$

Combining (3.5) and (3.7) yields $v_{K+1} \geq v_K$, since ε was arbitrary. If no least favorable distribution or Bayes or minimax rules exist one can obtain a proof by essentially repeating the above argument but using instead a sequence of

ε-minimax rules, ε-Bayes rules, etc... The only point that requires some care is the relationship between δ^{θ}_{K+1} and ϕ^*_K as expressed in (3.3) and (3.4). This is resolved by noting that one method of constructing ε-Bayes rules is to use a rule which comes within ε of minimizing the posterior risk for almost all $\underline{X}_{K+1}$ (i.e. the rule is conditional ε-Bayes as well as ε-Bayes).

In the following example we show that if Condition I of Theorem 3.3 is not fulfilled (yet all other assumptions are valid) then the result of Theorem 3.3 need not hold.

Example 3.1. Let X_1, X_2 be independent uniform distributions on $[0,\theta_1]$ and $[0,\theta_2]$ respectively. We wish to estimate $\theta_{J_K(\underline{X}_K)}$ with respect to squared error loss, $L(a,\theta) = (a-\theta)^2$. However the parameter space consists of only two points; $\Theta = \{1,2\}$. Easy computations show that in the 1 dimensional problem the minimax rule is ϕ^*_1 where $\phi^*_1(x_1) = \sqrt{2}$ if $0 \leq x_1 \leq 1$ and $\phi^*_1(x_1) = 2$ if $1 < x_1 \leq 2$ and the minimax value is $v_1 = (\sqrt{2}-1)^2 > .17$. In 2 dimensions the minimax rule is ϕ^*_2 where $\phi^*_2(x_1,x_2) = (1+\sqrt{3})/2$ if $0 \leq x_i \leq 1$, $i=1,2$ and $\phi^*_2(x_1,x_2) = 2$ otherwise and the minimax value is $v_2 = (\sqrt{3}-1)^2 < .14$. Thus $v_2 < v_1$.

The following corollary, which follows immediately from Theorem 3.3 indicates how the result of that theorem may be used to find minimax rules (see also example 4.2) for problems in which K is fixed.

COROLLARY 3.1. Under the assumptions of Theorem 3.3 if there exists an estimator $\psi(\underline{X}_K)$ for the K dimensional problem such that

$$\sup_{\underline{\theta}_K} R(\psi,\underline{\theta}_K) \leq v_i \qquad \text{some } i=1,\ldots,K$$

then ψ is minimax. (Usually v_i would be determined most easily for $i=1$.)

Another immediate consequence of Theorem 3.3 is the following rather general result.

COROLLARY 3.2. Let $X_1, X_2, \ldots$ be independent continuous random variables. If X_i has (location parameter) c.d.F. $F(x-\theta_i)$, $\Theta = (-\infty,\infty)$ and $L(a,\theta) > 0$ then $v_K < v_{K+1}$ all $K=1,2,\ldots$.

On the other hand v_K cannot increase with K too rapidly. We have

THEOREM 3.4. If $L(a,\theta) > 0$, then $v_{K+1} < (K+1)v_K/K$.

Proof. Let ϕ_K be a minimax rule for the K dimensional problem, and let $\tilde{\phi}$ be the rule for the K+1 dimensional problem defined as follows: Drop, at random, one of the observations different from $X_{J_{K+1}(\underline{X}_{K+1})}$, each with probability K^{-1}. Then decide on $\theta_{J_{K+1}}$ by using ϕ_K with the remaining observations.
Let $\underline{\theta}_{K+1}$ be fixed, and let $\underline{\theta}_{K+1(i)} = (\theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots\theta_{K+1})$, $i=1,\ldots,K+1$, and let $\underline{X}_{K+1(i)}$ be defined correspondingly. Let $A_j = \{\underline{x}_{K+1}:J_{K+1}(\underline{x}_{K+1})=j\}$, $j=1,\ldots,K+1$, and $B_j^{(i)} = \{\underline{x}_{K+1(i)}:J_K(\underline{x}_{K+1(i)})=j\}$, $j=1,\ldots,K+1$, $j\neq i$. Then

$$R_{J_{K+1}}(\tilde{\phi},\underline{\theta}_{K+1}) = \sum_{j=1}^{K+1} \int_{A_j} L(\tilde{\phi}(\underline{x}_{K+1}),\theta_j) \prod_{t=1}^{K+1} (f_{\theta_t}(x_t)dx_t)$$

$$= K^{-1} \sum_{j=1}^{K+1} \sum_{\substack{i=1\\ i\neq j}}^{K+1} \int_{A_j} L(\phi_K(\underline{x}_{K+1(i)}),\theta_j) \prod_{t=1}^{K+1} (f_{\theta_t}(x_t)dx_t)$$

$$(3.8) \qquad < K^{-1} \sum_{j=1}^{K+1} \sum_{\substack{i=1\\ i\neq j}}^{K+1} \int_{B_j^{(i)}} L(\phi_K(\underline{x}_{K+1(i)},\theta_j) \prod_{\substack{t=1\\ t\neq i}}^{K+1} (f_{\theta_t}(x_t)dx_t)$$

$$= K^{-1} \sum_{i=1}^{K+1} \sum_{\substack{j=1\\ j\neq i}}^{K+1} \int_{B_j^{(i)}} L(\phi_K(\underline{x}_{K+1(i)},\theta_j) \prod_{\substack{t=1\\ t\neq i}}^{K+1} (f_{\theta_t}(x_t)dx_t)$$

$$= K^{-1} \sum_{i=1}^{K+1} R_{J_K}(\phi_K,\underline{\theta}_{K+1(i)}) < K^{-1}(K+1)v_K,$$

where the first inequality in (3.8) uses $L > 0$, and the last follows since ϕ_K is a minimax rule for the K-dimensional problem. Taking supremum over $\underline{\theta}_{K+1}$ yields the theorem. Clearly the proof carries over also when no minimax rule exists for the K dimensional problem, and ϕ_K is chosen to be any ε-minimax rule.

We chose to concentrate on the maximal observation, J_K. Clearly, however, the arguments carry over quite easily to any other fixed order statistic. Similar results carry over also to other selection rules, such as the median. Note that if $\phi_K(\underline{x}_K)$ is any (fixed) function of X_{J_K} only, and one uses ψ_K to decide on θ_{J_K}, then an argument similar to that of Theorem 3.4 yields $b_{K+1} \leq b_K(K+1)/K$, where $b_K = \sup_{\underline{\theta}_K} R(\psi_K, \underline{\theta}_K)$.

In example 4.1,(b), we show that for the Normal distribution $v_K \to \infty$.

4. Examples.

In this section we investigate minimaxity for some particular distributions and try to relate the results to those of the previous section. Except for the normal case (where some theoretical results are also established) we will simply report the outcomes without exhibiting the calculations.

Example 4.1. The Normal distribution: Let X be $N(\theta,\sigma^2)$ and let θ be $N(\mu,\nu^2)$. The posterior distribution of θ given x is then $N(\mu_1,\nu_1^2)$ where $\mu_1 = (\sigma^2\mu + \nu^2 x)/(\sigma^2 + \nu^2)$ and $\nu_1^2 = \sigma^2\nu^2/(\sigma^2 + \nu^2)$. For the component problem μ_1 is therefore the Bayes estimator and ν_1^2 is the posterior risk for the Bayesian problem with squared error loss $L(a,\theta) = (a-\theta)^2$. Since ν_1^2 is independent of x it is the Bayes risk for the Bayes rule for any $\theta_{I(\underline{X})}$.

Now let $\nu^2 \to \infty$. The Bayes risk then tends to σ^2, and hence the minimax value for $\theta_{I(\underline{X})}$ and any selection rule I(.) and any K is at least σ^2. For K=2, $X_{(1)}$ (the largest order statistic) as an estimator of θ_{J_K} has risk function equal to the constant σ^2, and hence is minimax. This result is well known. See e.g., Cohen and Sackrowitz (1981), where the problem of finding "reasonable" estimators for θ_{J_K} for the normal distribution, is treated in detail. (By symmetry of the normal distribution, the problem of estimating the θ corresponding to the minimum X_i is analogous to that of estimating θ_{J_K}.)

The result above does not generalize to K > 2, and indeed, for K $\geq$ 3 both the minimax rule and value are unknown. Below we prove the following facts (using the antirank notation of $q_1 = J_K(\underline{X})$):

(a) $X_{(1)}$ is not minimax for $K \geq 3$; and (b) $v_K \to \infty$ as $K \to \infty$.

Proof of (a): For $C > 0$ let $\phi_c(\underline{X}) = X_{(1)}-C$. Then

$$(4.1) \qquad R_{q_1}(\phi_C,\underline{\theta}) = R_{q_1}(\phi_0,\underline{\theta}) + C^2 - 2C\, E_{\underline{\theta}}(X_{(1)} - \theta_{q_1})$$

Now $E_{\underline{\theta}}(X_{(1)}-\theta_{q_1}) > \max_i EX_i - E\theta_{q_1} \geq \max_i \theta_i - \max_i \theta_i = 0$ and $E_{\underline{\theta}}(X_{(1)} - \theta_{q_1})$ tends to 0 only if the difference between $\max_i \theta_i$ and all other θ_j tends to ∞. By Theorem 3.1 of Cohen and Sackrowitz (1981) $R_{q_1}(\phi_0,\underline{\theta})$ is maximal if $\theta_1=\ldots=\theta_K$. For this case $R_{q_1}(\phi_0,\underline{\theta}) = \sigma^2(a_K^2 + b_K^2)$ where $a_K^2 = \mathrm{Var}(\max_{i=1,\ldots,K} Z_i)$ and $b_K = E(\max_{i=1,\ldots,K} Z_i)$ and Z_i are i.i.d. $N(0,1)$. a_K^2 and b_K are extensively tabulated. For $K=2$, $a_K^2 + b_K^2 = 1$, and for $K \geq 3$, $a_K^2 + b_K^2 > 1$. For $K \geq 3$, $R(\phi_0,\underline{\theta})$ tends to σ^2 only when the difference between the largest, or two largest θ_i and all other θ_i tends to ∞. Otherwise it is larger than σ^2. Let $K \geq 3$, $0 < \varepsilon < (a_K^2 + b_K^2 - 1)\sigma^2$, and $\Omega = \{\underline{\theta}: \max_{\underline{\theta}} R_{q_1}(\phi_0,\underline{\theta}) - R_{q_1}(\phi_0,\underline{\theta}) \leq \varepsilon\}$. Also, let $\inf_{\underline{\theta}\varepsilon\Omega} E_\theta(X_{(1)} - \theta_{q_1}) = \delta$. By our previous argument $\delta > 0$. Now let $0 < C^* < \min\{\varepsilon^{1/2}, 2\delta\}$. By (4.1), for any $\underline{\theta}_0 \,\varepsilon\, \Omega$, $\max_{\underline{\theta}\varepsilon\Omega} R_{q_1}(\phi_0,\underline{\theta}) - R_{q_1}(\phi_{C^*},\underline{\theta}_0) \geq R_{q_1}(\phi_0,\underline{\theta}_0) - R_{q_1}(\phi_{C^*},\underline{\theta}_0) = C^*[2E_{\underline{\theta}_0}(X_{(1)} - \theta_{q_1}) - C^*] \geq C^*(2\delta - C^*)$, and for any

$\underline{\theta}_0 \,\varepsilon\, \Omega$, $\max_{\underline{\theta}\varepsilon\Omega} R_{q_1}(\phi_0,\underline{\theta}) - R_{q_1}(\phi_{C^*},\underline{\theta}_0) > \max_{\underline{\theta}\varepsilon\Omega} R_{q_1}(\phi_0,\underline{\theta}) - [R_{q_1}(\phi_0,\underline{\theta}_0) + C^{*2}] \geq \varepsilon - C^{*2}$. Thus $\sup_{\underline{\theta}\varepsilon\mathbf{R}} R_{q_1}(\phi_0,\underline{\theta}) - \sup_{\underline{\theta}\varepsilon\mathbf{R}}(\phi_{C^*},\underline{\theta}) = \inf_{\underline{\theta}\varepsilon\mathbf{R}} \{\max_{\underline{\theta}_0\varepsilon\mathbf{R}} R_{q_1}(\phi_0,\underline{\theta}_0) - R_{q_1}(\phi_{C^*},\underline{\theta})\} \geq \min\{C^*(2\delta - C^*), \varepsilon - C^*\} > 0$ and hence ϕ_0 is not minimax. Whether ϕ_0 is admissible or not, is still and open question.

Proof of (b). For simplicity set $\sigma^2 = 1$ and consider the Bayesian problem of estimating θ_{J_K}, where the θ_i are i.i.d. with prior $P\{\theta_i = 0\} = 1-P\{\theta_i=\xi\} = 1-K^{-1}$, $I=1,\ldots,K$. By Lemma 3.2 the Bayes estimator is $\xi_K[1+(K-1)\exp(-\xi_K X_{J_K} + \xi_K^2/2)]^{-1}$. Let B_K be the Bayes risk of this estimator. We shall show that for proper choice of ξ_K, $B_K \to \infty$ as $K \to \infty$. Since $v_K \geq B_K$ the result then follows.

$$B_K \geq P\{\theta_1=\ldots=\theta_K = 0\}\xi_K^2 E([1+(K-1)\exp(-\xi_K Y_K+\xi_K^2/2)]^{-2}) \tag{4.2}$$

where $Y_K = \max_{i=1,\ldots,K} Z_i$ and $Z_i \sim N(0,1)$. Let $\xi_K = (2 \log K)^{1/2}$. Then $E([1+(K-1)\exp(-\xi_K Y_K+\xi_K^2/2)]^{-2}) > \frac{1}{4}P\{\exp(-\xi_K Y_K+\xi_K^2/2) \leq K^{-1}\}$ $= \frac{1}{4}P\{Y_K \geq (2 \log K)^{1/2}\}$. Since $P\{\theta_1=\ldots=\theta_K= 0\} = (1-K^{-1})^K \to e^{-1}$, it follows by (4.2) that is suffices to show that

$$\lim_{K\to\infty} 2 \log K \, P(Y_K \geq (2 \log K)^{1/2}) = \lim_{K\to\infty} 2 \log K[1-\Phi^K((2 \log K)^{1/2})] = \infty. \tag{4.3}$$

For $x > 0$ (See e.g. Feller (1968), 2nd ed. p. 166)

$$\phi(x)x^{-1}\{1-x^{-2}\} < 1-\Phi(x) < \phi(x)x^{-1}. \tag{4.4}$$

From (4.4) we have

$$1 \geq \lim_{K\to\infty} \Phi^K(\sqrt{2\log K}) \geq \lim_{K\to\infty}[1 - \frac{1}{2K\sqrt{\pi \log K}}]^K = 1 \tag{4.5}$$

It follows from a well known inequality (see e.g. Hardy, Littlewood and Polya (1934) 2nd ed. p.39, paragraph 41) that for all x

$1-\Phi^K(x) > K \,\Phi^{K-1}(x)(1-\Phi(x))$, and thus by (4.4)

$$1-\Phi^K(x) > K \,\Phi^{K-1}(x)\phi(x)x^{-1}[1-x^{-2}] \tag{4.6}$$

Setting $x = \sqrt{2\log K}$ in (4.6), multiplying by 2 log K, we have

$$\lim_{K\to\infty} 2 \log K[1-\Phi^K(\sqrt{2\log K})] \geq \lim_{K\to\infty} \{(2 \log K)K \,\Phi^{K-1}(\sqrt{2\log K})(\sqrt{2\pi})K^{-1}$$

$$\times\ (2 \log K)^{-1/2}[1-(2 \log K)^{-1}]\} = \infty \text{ by } (4.5)$$

which establishes (4.3) and completes the proof.

A good example of the use of Corollary 3.1 is

Example 4.2: The Uniform Distribution. Let X_i be independent, $X_i \sim U(0,\theta_i)$, $i=1,\ldots,K$. The loss function is $L(a,\theta) = (a-\theta)^2/\theta^2$. It can be shown that for one dimension the minimax rule is $3X_1/2$ and has constant risk of 1/4, i.e. $v_1 = 1/4$. Consider the estimator $\psi(\underline{X}) = 3X_{J_K}/2$. Suppose, without loss of generality, that $\theta_1 \geq \ldots \geq \theta_K > 0$. After somewhat tedious calculations it can be shown that the risk function of ψ in the K dimensional problem is

$$(4.7)\quad \frac{1}{4} + 3\sum_{i=1}^{K-1} \frac{\theta_{i+1}^{i}}{\prod_{t=1}^{i}\theta_t}\left\{\frac{3}{4(i+3)}\left[1-\frac{1}{i+2}\sum_{s=1}^{i}\left(\frac{\theta_{i+1}}{\theta_s}\right)^2\right] - \frac{1}{i+2}\left[1-\frac{1}{i+1}\sum_{s=1}^{i}\left(\frac{\theta_{i+1}}{\theta_s}\right)\right]\right\}.$$

It is easily seen that the curly bracket portion of (4.7) is nonpositive for all $i \leq 3$ and the risk function is $\leq 1/4$ for $K \leq 4$. Hence $v_1=\ldots=v_4=1/4$ and $\psi(\underline{X})$ is minimax by Corollary 3.1.

Example 4.3. The Negative Exponential Distribution: Let $X_1,\ldots,X_K$ be i.i.d. each with density $f_\theta(x) = \theta \exp(-\theta x)$, $x > 0$, $\theta > 0$. Let the prior on θ be gamma with density $g(\theta) = \Gamma(r)^{-1}\xi^r\theta^{r-1}e^{-\theta\xi}$ for $\theta > 0$ where $r > 0$, $\xi > 0$. Using loss $L(a,\lambda) = (a-\lambda)^2/\lambda^2$ to estimate $\lambda = \theta^{-1}$, straight forward calculations yield $\phi(x) = (x+\xi)/(r+2)$ as Bayes estimator with posterior risk $(r+2)^{-1}$ for the component problem. By Theorem 3.1 this will be the Bayes risk for $\lambda_{I(\underline{X})}$ for any selection rule $I(.)$. This example was considered in detail by Sackrowitz and Samuel-Cahn (1984). Letting $r \to 0$ yields a minimax value of at least 1/2 for any selection rule $I(.)$. The corresponding Bayes rule for the component problem tends to $(x+\xi)/2$ and letting also $\xi \to 0$ we have $x/2$ as a candidate for a minimax rule for the component problem. It is easily seen to have constant risk 1/2, and hence is minimax. The "natural" candidate for minimax rule for λ_{J_K} is therefore $\frac{1}{2}X_{(1)}$. For small values of K it is minimax (with nonconstant risk), but for $K \geq 8$ it is not. It was however shown (see Sackrowitz and Samuel-Cahn (1984)) that the estimator of λ_{J_K} given by

$\frac{1}{2}(X_{(1)}-X_{(2)})$ has constant risk 1/2, for any K, and hence, by Theorem 3.2 is minimax. Note that this is an example of a rather unusual situation, where the limit of the Bayes risks yields the minimax value, but the corresponding limit of the Bayes rules does not yield a minimax rule.

REFERENCES

Cohen, A. and Sackrowitz, H. (1981). Estimating the mean of the selected population. Proceedings of Third Symp. Stat. Decision Theory and Related Topics. Purdue University.

Dahiya, R.C. (1974). Estimation of the mean of the selected population. J. Amer. Statist. Assoc. **69** 226-230.

Feller, W. (1968). An Introduction to Probability Theory and Its Applications. Vol. 1, 3rd Ed. Wiley, NY.

Ferguson, T.S. (1967). Mathematical Statistics Academic Press, New York and London.

Guttman, I. and Tiao, G.C. (1964). A Bayesian approach to some best population problems. Ann. Math. Statist., **35** 825-835.

Hardy, G.H., Littlewood, J.E. and Polya, G. (1934). Inequalities Cambridge University Press, London.

Putter, J. and Rubinstein, D. (1968). On estimating the mean of a selected population. Technical Report NO. 165, University of Wisconsin.

Sackrowitz, H. and Samuel-Cahn, E. (1984). Estimation of the mean of a selected negative exponential population. J. Roy. Statist. Soc. Ser. B, **46** 242-249.

Sarkadi, K. (1967). Estimation after selection. Studia Scientiarum Mathematicarum Hungarica, 2 341-350.