D-A140179
BAYES-P* SUBSET SELECTION PROCEDURES FOR THE BEST
POPULATION (U) PURDUE UNIVERSITY LAFAYETTE IN
DEPT OF STATISTICS
S. S. Gupta ET AL.
FEB 84 TR-84-2
UNCLASSIFIED N00014-75-C-0435
F/G 12/1
BAYES-P* SUBSET SELECTION PROCEDURES
FOR THE BEST POPULATION*

by

Shanti S. Gupta
Purdue University

Hwa-Ming Yang
University of Toledo

Technical Report #84-2

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited
BAYES-P* SUBSET SELECTION PROCEDURES
FOR THE BEST POPULATION*

by

Shanti S. Gupta                        Hwa-Ming Yang
Purdue University                     University of Toledo

Technical Report #84-2

Department of Statistics
Purdue University

February 1984

*This research was supported by the Office of Naval Research
Contract N00014-75-C-0455 at Purdue University. Reproduction
in whole or in part is permitted for any purpose of the United
States Government.

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited
BAYES-P* SUBSET SELECTION PROCEDURES
FOR THE BEST POPULATION*

by
Shanti S. Gupta
Purdue University

Hwa-Ming Yang
University of Toledo

ABSTRACT
Two new selection procedures, called nonrandomized and randomized Bayes-P* procedures are defined for selecting a small nonempty subset of k populations which contains the best population. It is shown that these procedures have some optimal properties. If we restrict attention to the class D(D*) of all nonrandomized (randomized) selection procedures, which satisfy the PP*-condition, that is the posterior probability of a correct selection, for any given observation \( x = x \), is not less than \( P* \), a predetermined number between \( 1/k \) and \( 1 \), then these two new selection procedures are shown also to be Bayes decision procedures in the class D and D* respectively, provided that some regularity conditions are satisfied. Robustness of these procedures and comparisons with some other selection procedures are studied by using Monte Carlo simulations.

1. INTRODUCTION
Suppose we have k populations \( \pi_1, \ldots, \pi_k \) whose distributions are determined by unknown real parameters \( \theta_1, \ldots, \theta_k \), respectively. In a subset selection problem, the goal is to select a subset of

*This research was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.
the populations which includes the population associated with the largest parameter with high probability and, possibly, includes the others with low probabilities. A population \( \pi_i \) will be called the best population if \( \pi_i \geq \pi_j \) for all \( j \neq i \). The remaining \( k-1 \) populations will be called non-best. If there are more than one population satisfying this condition we arbitrarily tag one of them and call it the best one.

A large body of literature exists in the area of subset selection procedures (see Gupta and Panchapakesan (1979)). As pointed out by many authors (see, for example, Bahadur (1950), Bechhofer (1954)) the testing of homogeneity of population means or variances is not a satisfactory solution to a comparison among several populations. Gupta (1956, 1965) gave a maximum-type subset selection procedure. The maximum-type procedures present a direct and relatively efficient way to meet our goal. Gupta and Hsu (1978) studied the performance of maximum-type procedure, average-type procedure (Seal (1955, 1957)) and Bayes procedures. Berger (1979) and Berger and Gupta (1980) proved that the maximum-type procedure is minimax under certain loss functions. The LFC (least favorable configurations) of the maximum-type procedure usually occurs when the distributions are identical, i.e., under the hypothesis of homogeneity. As usual, in many cases, the hypothesis of homogeneity is rejected at some small significance level. It seems then that the maximum-type procedure is still conservative. Therefore we may wish to relax (modify) the so-called \( P^* \)-condition. On the other hand, in the decision-theoretic approach, Bayes procedure always gives us a most economic decision under a certain loss; however, this
does not mean that its quality is good enough to pass a certain level. Suppose the loss function is a linear combination of \( L_i(e) \), \( i = 1, \ldots, k \), where \( L_i(e) \) is the loss if the \( i \)th population is selected in the subset, as was assumed by Bahadur and Goodman (1952), Dunnett (1960), Lehmann (1966), Eaton (1967) and Alam (1973). As pointed out by Goel and Rubin (1977), the decision-theoretic procedures mentioned above do not seem to be appropriate, mainly because they ignore a reasonable component of loss which depends on whether or not the selected subset contains the best population and secondly because they specify the subset size in advance, whereas it should depend on the information available from the sample. We may use some loss functions that involve an additional component, such as the loss function given by Gupta and Hsu (1978), which is associated with the probability of incorrect selection, or the one given by Goel and Rubin (1977), which is associated with the distance between the selected subset and the best population, or some others which are proposed by Chernoff and Yahav (1977), Bickel and Yahav (1977) and Kim (1979), to improve the quality of decision. However, the results are quite sensitive to the weights of each components, or equivalently, the ratio of the coefficient of the two components. In practice we always have some difficulties in figuring out the ratio whenever the two components of loss are not comparable, or they are comparable but the ratio is not a constant function or it is not completely known. In these situations we may wish to try some other methods of attack.

For guaranteeing the quality of selection procedures, we would like to have a 'quality control' about the class of all possible selection procedures, that is, any selection procedure with lower
quality will be removed, even though it might be the cheapest one under some loss. By using the PP*-condition (defined in Section 2) as a filter or control condition we get a class of selection procedures D and its randomized version D*. The PP*-condition represents the minimum quality (accuracy) level of a selection procedure under certain prior information. We try to derive one of these procedures, which gives the minimum risk under a large family of loss functions and has properties which we think an optimal selection procedure should have.

In Section 2 we define some notations, the PP*-condition (posterior-P*-condition), class D and D* and propose two selection procedures \( \psi^B \) and \( \psi^{B*} \). Some selection procedures close to \( \psi^B \) but restricted to normal populations were studied by Roth (1978) and Naik (1978). The optimal properties of \( \psi^B \) and \( \psi^{B*} \) in D and D*, respectively, such as ordered, justness, most efficient and Bayes with respect to a large family of loss functions are shown in Section 3. Their application to normal distributions is discussed in Section 4. In Section 5 procedure \( \psi^B \) is compared with the maximum-type procedure. In Section 6 we discuss their applications to the selection problem for Poisson distributions and Poisson processes and their relation to the selection of gamma distributions. Section 7 deals with comparisons of the performance of selection procedures \( \psi^{B*}, \psi^B, \psi^M \) and \( \psi^{MED} \). Here \( \psi^M \) and \( \psi^{MED} \) are the maximum-type selection procedures based on sample means and sample medians, respectively (see Gupta (1956, 1965) and Gupta and Singh (1980)). The comparisons are based on Monte Carlo studies. Robustness of these four procedures is studied in terms of the expected size of the selected subset and the efficiency (defined in Section 3) where the robustness is in the sense of the effect on the performance of the procedure when the k true distributions are not
normal but, say, logistic, double exponential distribution or the contaminated distribution (gross error model) (Tukey (1960)). Further discussion about the Bayes-P* selection procedures is given in Section 8.

2. NOTATION AND FORMULATION

Assume that we have $n_i$ independent observations $X_{ij}$, $j = 1,...,n_i$ for population $\pi_i$, $i = 1,...,k$. Let $X_i = T_i(X_{i1},...,X_{in_i})$ be a suitable estimator of $\theta_i$, $i = 1,...,k$; assume that $X_i$'s are independently distributed. Usually $X_i$ is a sufficient statistic for $\theta_i$. Let $\theta = (\theta_1,...,\theta_k) \in \Theta \subset R^k$ and let $X = (X_1,...,X_k)$ with cumulative distribution function (cdf) $F(x|\theta)$ and density (frequency) $f(x|\theta)$. A selection procedure will be denoted by $\psi(x) = (\psi_1(x),...,\psi_k(x))$ where $\psi_i(x) : R^k \rightarrow [0,1]$ is the probability that $\pi_i$ is included in the selected subset when $X = x$ is observed. A selection procedure $\psi$ is called nonrandomized if all $\psi_i$'s are 0 or 1, otherwise, it is a randomized procedure. A correct selection (CS) is defined to be the selection of any subset that includes the best population. Suppose we have a prior distribution $\tau$ for $\theta = (\theta_1,...,\theta_k)$ and our control condition is that for any given observation the posterior probability of CS must be greater than or equal to $P*$, a preassigned value between $1/k$ and 1. That is

$$P(CS|\psi, X = x) = \sum_{i=1}^{k} \psi_i(x) p_i(x) \geq P*, \text{ for all } x,$$

where

$$P_i(x) = P(\pi_i \text{ is the best}|X = x) = P(\pi_i \text{ is the largest}|X = x).$$

For convenience, we now assume that posterior cdf of $\theta$ is absolutely continuous. Then it is clear that
\[ \sum_{i=1}^{k} p_i(x) = 1, \]
hence this kind of selection procedures always exist. Let \( p[1](x) \leq \cdots \leq p[k](x) \) be the ordered \( p_i(x) \)'s and let \( \pi(i) \) be the population associated with \( p[i](x) \), \( i = 1, \ldots, k \), then a subset selection procedure is completely specified by \( \{\psi(1), \ldots, \psi(k)\} \)
where \( \psi(i) \) is defined by
\[
(2.3) \quad \psi_i(x) = P(\pi(i) \text{ is selected} | x), \quad i = 1, \ldots, k.
\]

**Definition 2.1.** Given a number \( P* \), \( 1/k < P* < 1 \), and a prior \( \tau \), we say a selection procedure \( \psi \) satisfies the PP*-condition (posterior-P*-condition) if
\[
(2.4) \quad \psi_k(x) = 1 \text{ and } P(\text{CS} | \psi, X = x) > P* \text{ for all } x.
\]

**Remark 1.** The PP*-condition implies the expected probability of CS with respect to a given prior is not less than \( P* \). Since the prior information is used in PP*-condition, it is different from the usual so-called P*-condition.

Given a prior \( \tau \), let \( D = D(\tau, P*) (D^* = D^*(\tau, P*)) \) be the class of all nonrandomized (randomized) selection procedures in which all procedures satisfy the PP*-condition for any given observation \( X = x \).

Now, we propose two selection procedures \( \psi^B \) and \( \psi^{B*} \) as follows:

**Definition 2.2.** Given a number \( P*(1/k < P* < 1) \), an observation \( X = x \) and a prior \( \tau \), the selection procedure \( \psi^B \) is defined by \( \{\psi^B(1), \ldots, \psi^B(k)\} \) where
\[
\psi^B(i)(x) = \begin{cases} 
1, & \text{if } i \geq j \\
0, & \text{otherwise}
\end{cases}
\]
and \( j \) is the largest integer between 1 and \( k \), such that
Definition 2.3. Given a number $P^*(1/k < P^* < 1)$, an observation $X = x$ and a prior $\tau$, the randomized selection procedure $\psi^{B^*}$ is defined by $(\psi_1, \ldots, \psi_k)$ where

$$\psi^B(k)(x) = 1,$$

and

$$\psi^B(j)(x) = \begin{cases} 1, & \text{if } \sum_{i=j+1}^{k} p[i](x) \leq P^*, j \neq k, \\ v, & \text{if } \sum_{i=j+1}^{k} p[i](x) < P^* \text{ and } \left(\sum_{i=j+1}^{k} p[i](x) > P^* \right), \\ 0, & \text{otherwise}, \end{cases}$$

where the constant $v$ is determined so that

$$vp[j](x) + \sum_{i=j+1}^{k} p[i](x) = P^*, 0 < v < 1.$$

Example. If $k = 3$, $P^* = 0.90$ and the posterior probabilities are: $p_1(x) = 0.05$, $p_2(x) = 0.80$, $p_3(x) = 0.15$, then selection procedure $\psi^{B^*}$ will select the population $\pi_2$ (corresponding to $p_3(x)$) with probability 1, and select $\pi_3$ with probability $v$ where $v$ is given by

$$0.15v + 0.80 = 0.90,$$

$$v = \frac{0.10}{0.15} = \frac{2}{3}.$$
and Bayes in its class, respectively.

3. **OPTIMAL PROPERTIES**

In this section some properties of selection procedures $\psi^B$ and $\psi^{B*}$ are studied.

**Definition 3.1.** A selection procedure $\psi$ is called ordered if for every $x \in \mathbb{R}^k$, $x_i \leq x_j$ implies $\psi_i(x) \leq \psi_j(x)$. It is called just if for every $i = 1, \ldots, k$, and $x, x' \in \mathbb{R}^k$, $\psi_i(x) \leq \psi_i(x')$ whenever $x_i \leq x'_i$, $x_j \geq x'_j$ for any $j \neq i$.

Just procedures were defined and investigated in more generality by Nagel (1970) and Gupta and Nagel (1971).

**Definition 3.2.** A selection procedure is translation invariant if for every $x \in \mathbb{R}^k$, for every $c \in \mathbb{R}$, $\psi_i(x+c) = \psi_i(x)$ for every $i = 1, \ldots, k$, where $\mathbb{1} = (1, \ldots, 1)$.

**Lemma 3.1.** (Berger and Gupta (1980)) A selection procedure $\psi(x) = (\psi_1(x), \ldots, \psi_k(x))$ is just and translation invariant if and only if the following two conditions hold:

1. for every $i = 1, \ldots, k$, $\psi_i$ is a function only of the set of differences $(x_j - x_i | j = 1, \ldots, k, j \neq i)$, and
2. if $x$ and $y$ satisfy $x_j - x_i \leq y_j - y_i$ for every $j \neq i$, then $\psi_i(x) \geq \psi_i(y)$.

Let $\rho = (p_1, \ldots, p_k)$, where $p_i$'s are defined by (2.2). If we treat $\rho$ as a selection procedure, then $\psi^B(\psi^{B*})$ is ordered, just and translation invariant if and only if $\rho$ is ordered, just and translation invariant, respectively. Therefore, we have Theorem 3.1.
Theorem 3.1. Selection procedure $\psi^B$ and $\psi^{B*}$ are just and translation invariant if

1. for every $i = 1, \ldots, k$, $p_i = p_i(x)$ is a function only of the set of differences $x_j - x_i | j = 1, \ldots, k$, $j \neq i$, and
2. if $x$ and $y$ satisfy $x_j - x_i \leq y_j - y_i$ for every $j \neq i$, then $p_i(x) \geq p_i(y)$.

Some sufficient conditions for $\rho$ to be ordered, just or translation invariant are given below:

Theorem 3.2. Let $H(\theta_i | x)$ be the posterior cdf of $\theta_i$, given $X = x$.

If $H(\theta_i | x)$ is absolutely continuous and has the generalized stochastic increasing property (GSIP), that is:

1. $H(\theta_i | x) = \prod_{i=1}^{k} H_i(\theta_i | x)$, $H_i(\cdot | x) = \text{posterior cdf of } \theta_i$.
2. $H_i(t|x) \geq H_j(t|x)$ for any $t$, whenever $x_i \leq x_j$.

Then both $\psi^B$ and $\psi^{B*}$ are ordered and just. If, in addition, $H_i$ has location parameter $x_i$, that is, $H_i(\theta_i | x) = H_i(\theta_i - x_i)$ for every $i = 1, \ldots, k$, then both $\psi^B$ and $\psi^{B*}$ are also translation-invariant.

Proof: The first part can be proved by using integration by parts. Since $x_i \leq x_j$, implies $H_i(t|x) \geq H_j(t|x)$ for all $t$, hence

$$p_i(x) = P(\theta_i = \theta[i] | x)$$

$$= \int_{m \neq i} \prod_{m \neq i} H_m(t|x) dH_i(t|x)$$

$$\leq \int_{m \neq j} \prod_{m \neq j} H_m(t|x) dH_i(t|x)$$

$$= 1 - \int H_i(t|x) d[\prod_{m \neq i} H_m(t|x)]$$

$$\leq 1 - \int H_j(t|x) d[\prod_{m \neq j} H_m(t|x)]$$
\[
H_m(t|x) dH_j(t|x)
\]

\[
= p_j(x) \text{ if } x_i \leq x_j.
\]

Therefore, both $\psi^B$ and $\psi^{B*}$ are ordered. The proof of justness is similar to the above and hence omitted. The proof of the second part is obvious.

Let $G$ denote the group of all permutations of the components of a $k$-component vector. A set $S \subset \mathbb{R}^k$ is called symmetric if $gS = S$ for all $g \in G$. A distribution $H$ is called symmetric if $H(S) = H(gS)$ for all measurable set $S$ and $g \in G$. A family of distributions $P_\theta$ is called invariant with respect to $G$ if $P_\theta(S) = P_{g\theta}(gS)$ for all measurable set $S$ and $g \in G$.

Given an observation $X = x$, suppose set $s$ is the selected subset under a selection procedure $\psi$ in $D$, then the loss can be described by a non-negative real-valued function $L(\theta, s)$ which has the properties below:

**Definition 3.3.** For all $g \in G$, $\theta \in \Theta$, a loss function $L$ has property $T$ if and only if

1. $L(\theta, s) = L(g\theta, gs)$,
2. $L(\theta, s)$ is non-increasing in $\theta_i$ for $i \in s$, and
3. $L(\theta, s) \leq L(\theta, s')$, if $s \subset s'$.

Let property $T'$ indicate that the loss function satisfies the first two conditions of property $T$, namely, invariance and monotonicity properties. The third condition of ordering of property $T$ is reasonable, because the indirect loss of an incorrect selection is controlled by the $PP^*$-condition and the direct loss of a selected subset is naturally more than its subset.
Example 3.1. The following loss functions satisfy property T:

a. \( L(\hat{\theta}, s) = s \)

b. \( L(\hat{\theta}, s) = \sum_{i \in S} L_i(\hat{\theta}) \) where \( L_i(\hat{\theta}) \), the loss if the ith population is selected in the subset, is invariant and monotonic, i.e.

\[ L_g_i(\hat{\theta}_j) = L_i(\hat{\theta}) \] for all \( g \in G \), and \( L_i(\hat{\theta}) \leq L_{i+1}(s) \) whenever \( \theta_i < \theta_{i+1} \), \( i = 1, \ldots, k-1 \). A useful form of this loss is that \( L_i(\hat{\theta}) = q(\hat{\theta}, \theta_i) \) which is non-increasing in \( \theta_i \), where \( q(\hat{\theta}) \) is a real-valued symmetric function of \( \hat{\theta} \). For example, \( q(\hat{\theta}, \theta_i) = C(\theta_i, \theta_i, \theta_j) \) for all \( \theta_i \) and \( \theta_j \) such that \( \theta_i < \theta_j < \theta_k \).

The next theorem (Theorem 3.3.) shows that under some regularity conditions selection procedure \( \psi^B(\psi^{B*}) \) is Bayes in \( D(D^*) \) if the loss function has property T.

Theorem 3.3. Suppose the prior distribution \( \pi \) is symmetric on \( \Theta \). Given \( \hat{\theta} \in \Theta \), \( X_1, \ldots, X_k \) are independently distributed and the pdf \( f(x|\theta) \) has monotone likelihood ratio (MLR) property. Then \( \psi^B(\psi^{B*}) \) is ordered and is a Bayes procedure in \( D(D^*) \) provided that the loss function has property T.

Proof: First we need to show that the selection procedure \( \psi^B(\psi^{B*}) \) or \( \psi \) is ordered. For any \( i \neq j \), if \( x_i \leq x_j \), let

\[ \Theta_1 = \{ \theta \in \Theta | \theta_i < \theta_j \} \]

\[ p_j(x) - p_i(x) = b/ \int_\Theta \int_{\Theta_{i,j}} f(x|\theta) \] d\( \theta_i \)

\[ = b/ \int_\Theta \int_{\Theta_{i,j}} f(x|\theta) \] d\( \theta_i \)

\[ \geq 0, \]

where \( b \) is a normalizing factor and \( \theta_i \) is obtained from \( \theta_j \) by interchange the components \( \theta_i \) and \( \theta_j \). The third equation above is an application of
the assumption that \( \tau \) is symmetric on \( \theta \). The last inequality is based on the fact that \( f(x|\theta) - f(x|\theta') \) is non-negative by the MLR property of \( f(x|\theta) \) and \( I_{\{\theta=j}\} \{\theta\} - I_{\{\theta'=j\}} \{\theta\} \) is nonnegative. Because for any given observation \( x = x^* \), \( \psi^B \) always has minimum size of the selected subset, say \( m \), in \( D \). Therefore, under the assumptions and the property \( \tau \), the selection problem turns into "the mth decision problem" as mentioned in Lemma 1 of Goel and Rubin (1977), then by Theorem 4.1. of Eaton (1967), the result holds. The proof for procedure \( \psi^B \) is analogous, and hence is omitted.

**Theorem 3.4.** Under the assumptions of Theorem 3.2., \( \psi^B (\psi^B) \) is Bayes procedure in \( D(D^*) \) provided that the loss function has property \( \tau \).

**Proof:** By Theorem 3.2., \( \psi^B (\psi^B) \) is ordered. By an argument similar to the argument in Theorem 3.3., the theorem is proved.

For any selection procedure \( \psi \in D \), the posterior efficiency of \( \psi \), given observation \( x = x \), is defined by

\[
eff(\psi|x) = P(CS|\psi,x)/E(S|\psi,x)
\]

where \( E(S|\psi,x) \) is the posterior expected size of the selected subset. The expectation of \( \eff(\psi|x) \) is the efficiency of procedure and is denoted by \( \eff(\psi) \). A selection procedure \( \psi \in D \) is called most efficient (ME) in \( D(D^*) \) if \( \eff(\psi) \geq \eff(\psi') \) for all \( \psi' \in D(D^*) \).
Theorem 3.5. The selection procedures $\psi^B$ and $\psi^{B*}$ are ME in $D$ and $D^*$, respectively.

Proof: In $D$, given any observation $X = x$, since $\psi^B$ has minimum size of selected subset, say $m$, any selection procedure $\psi'$ in $D$ should have its size of selected subset equal to $m+c$ for some $0 \leq c \leq k-1$. Now,

$$\text{eff}(\psi'|x) = \frac{\{p[k-m-c+1](x) + \ldots + p[k](x)\}}{m+c}$$

$$\leq \frac{\{cp[k-m+1](x) + p[k-m+1](x) + \ldots + p[k](x)\}}{m+c}$$

$$\leq \frac{\{p[k-m+1](x) + \ldots + p[k](x)\}}{m}$$

$$= \text{eff}(\psi|x),$$

the first part is proved. For $\psi^{B*}$ in $D^*$, the proof is similar hence is omitted.

4. EXTENSION AND APPLICATIONS

In this section, the formulas for the posterior probabilities which are necessary to carry out the selection procedures $\psi^B$ and $\psi^{B*}$ are given under various assumptions.

Suppose we have $k$ populations $\pi_1, \ldots, \pi_k$; $\pi_i$ has normal distribution $N(\mu_i, \sigma_i^2)$ where $\mu_i$'s are unknown. Assume that we have sample $X_{i1}, \ldots, X_{i n_i}$ for each population $\pi_i$. Let $\bar{X}_i$ be the sample mean and let $\bar{X} = (\bar{X}_1, \ldots, \bar{X}_k)$. Suppose we are interested in selecting a subset containing the population having the largest population mean under the PP*-condition, with respect to some prior distribution $\tau$ of $\mu = (\mu_1, \ldots, \mu_k)$.

A. No Prior Information

Under the situation where very little is known a priori, we may use a 'non-informative' prior (see Box and Tiao (1973)) provided that
the unknown parameters are locally independent a priori (see Guttman
and Tiao (1964)).

A.1. Common Variance \( \sigma^2 \) (Known) and Common Sample Size \( n \)

By using the non-informative prior \( \tau(y) = c \), we have

\[
P_i(x) = \int_{-\infty}^{\infty} \prod_{j \neq i} \phi(t + \sqrt{n} \sigma^{-1}(x_{[i]} - x_{[j]})) \phi(t), \quad i = 1, \ldots, k.
\]

A.2. Unequal Variance \( \sigma_i^2 \)'s (Known) and Unequal Sample Size \( n_i \)'s

By using the same non-informative prior \( \tau(y) = c \), we have

\[
p_i(x) = \int_{-\infty}^{\infty} \prod_{j \neq i} \phi(tv_i/v_j + (x_i - x_j)/v_j) \phi(t)
\]

where \( v_i = \sigma_i/\sqrt{n_i}, \quad i = 1, \ldots, k. \)

A.3. Unequal Variance \( \sigma_i^2 \)'s and Unequal Sample Size \( n_i \)'s

By using non-informative prior \( \tau(u_i, \sigma_i) = \sigma_i - 1 \) for each
population, we have

\[
p_i(x) = \int_{-\infty}^{\infty} \prod_{j \neq i} T_v \left( t \frac{s_i/\sqrt{n_i}}{s_j/\sqrt{n_j}} + \frac{x_i - x_j}{s_j/\sqrt{n_j}} \right) dt_v (t)
\]

where \( v_i = n_i - 1, \)

\[
v_is_i^2 = \sum_{j=1}^{n_i} (x_{ij} - x_i)^2, \quad i = 1, \ldots, k.
\]

and \( T_v \) is the cdf of \( t \)-distribution with \( v \) degrees of freedom.

For large \( v \), it can be approximated by the normal distribution.

B. Independent Normal Prior

B.1. Identical Prior

Assume \( u_i \)'s have common distribution \( N(0, \sigma_0^2) \) and given \( u_i \),

\( x_i \) has distribution \( N(u_i, \sigma^2/\sqrt{n}) \), then
(4.5) \[ p_i(x) = \int_{-\infty}^{\infty} \prod_{j \neq i} \varphi(t+b\sigma_n^{-2}(x_i-x_j))d\phi(t), \]

where \( b^2 = (\sigma_0^{-2} + n\sigma_i^{-2})^{-1} \), \( i = 1,\ldots,k \).

B.2. Non-Identical Prior

Assume \( \nu_i \)'s have independent normal prior distribution \( N(\nu_i, \nu_0^2) \).

Since given \( \nu_i \), \( X_i \) has normal distribution \( N(\nu_i, \nu_i^2/n_i) \). Let

\[ z(x_i) = b_i^2(\sigma_0^{-2} \nu_i + n_i \sigma_i^{-2} x_i) \]

\[ b_i^2 = (\sigma_0^{-2} + \nu_i^2/n_i)^{-1} \]

we have, for \( i = 1,\ldots,k \),

(4.8) \[ p_i(x) = \int_{-\infty}^{\infty} \prod_{j \neq i} \varphi((tb_i + (z(x_i)-z(x_j))/b_j)d\phi(t). \]

In Case A.1. \( \psi^B \) and \( \psi^{B*} \) are just, translation invariant and ordered, hence these will be a Bayes procedure in \( D \) and \( D^* \), respectively, provided loss function has property T.

C. General Normal Model

Suppose we have \( k \) normal populations with common known variance \( \sigma^2 > 0 \) and common sample size \( n \). The observation can be reduced to \( X = (X_1,\ldots,X_k) \) where \( X_i \) is the sample mean for population \( \nu_i \). Assume that, given \( \nu \), \( X \) has normal distribution \( N(\nu, \nu I) \), where \( \nu = \sigma^2/n \) and \( \nu \) itself has normal distribution \( N(\theta_0 I, rI + wU) \) with \( \theta_0 \in \mathbb{R}, r = \sigma_0^2 > 0, w > -r/k, I = (1,\ldots,1) \) and \( U = I' \). Note that here \( r > 0 \) and \( w > -r/k \) are sufficient and necessary for \( rI + wU \) to be positive definite. This model was chosen by Chernoff and Yahav (1977) \( (t > 0) \), Gupta and Hsu (1978) and Miescke (1979). In this model, the \( p_i(x) \)'s are exactly the same as that of the independent prior case B.1.
5. RELATION BETWEEN PROCEDURES \( \psi^B \) AND \( \psi^M \) IN THE NORMAL LOCATION PARAMETER CASE

Suppose there are \( k \) independent normal populations with a common known variance \( \sigma^2 \) and common sample size \( n \). For this case, Gupta (1956) proposed and studied the maximum-type procedure

\[ \psi^M: \text{Select } i \text{ iff } X_i > X_{[k]} - d\sigma/\sqrt{n}, \quad i = 1, \ldots, k; \]

where \( d = d(k, P^*) > 0 \) is determined by the \( P^* \)-condition, that is,

\[ \int_{-\infty}^{\infty} \phi^{k-1}(t+d)d\phi(t) = P^*. \]  

(5.1)

When the prior \( \tau \) is the non-informative prior, the following theorem shows that \( \psi^M \in D(\tau, P^*) \) and gives for \( \psi^M \) a lower bound for the posterior probability of CS. Let

\[ (5.2) \quad \mathcal{Z} = \{ \text{all possible observed values} \} = R^k, \]

\[ (5.3) \quad \mathcal{Z}_1 = \{ x \in \mathcal{Z} | x[k] - \sigma/\sqrt{n} \leq x[1] \}, \]

\[ (5.4) \quad \mathcal{Z}_i = \{ x \in \mathcal{Z} | x[i-1] < x[k] - \sigma/\sqrt{n} \leq x[i] \}, \quad 2 \leq i \leq k, \]

\[ (5.5) \quad \mathcal{Z}_{i1}^{(1)} = \{ x \in \mathcal{Z} | x[1] = x[i-1] < x[k] - \sigma/\sqrt{n} \leq x[i] \} \subset \mathcal{Z}_i, \]

\[ (5.6) \quad \mathcal{Z}_{i1}^{(2)} = \{ x \in \mathcal{Z} | x[1] = x[i-1] < x[k] - \sigma/\sqrt{n} = x[i] = x[k-1] \} \subset \mathcal{Z}_i^{(1)}. \]

It is clear that \( \{ \mathcal{Z}_1, \ldots, \mathcal{Z}_k \} \) is a partition of the sample space \( \mathcal{Z} \).

**Theorem 5.1.** Given \( P^*(1/k < P^* < 1) \) and non-informative prior \( \tau \). If the observation \( \bar{x} = x \in \mathcal{Z}_i \), then

\[ (5.7) \quad P(\text{CS}|\psi^M, \bar{x} = x) > Q^*(i) \]

where

\[ (5.8) \quad Q^*(i) = P^* + (1-P^*)((k-i)/(k-1)). \]

Therefore, \( \psi^M \in D(\tau, P^*) \).
Proof: Without loss of generality we can assume $\sigma/\sqrt{n} = 1$. Since $p[j](x)$ is nonincreasing for all $x[m]$, $m \leq j-1$, given $x \in \mathcal{X}_i$, we have

$$P(CS|\psi^M, x) \geq \inf_{x \in \mathcal{X}_i} P(CS|\psi^M, x)$$

$$= \inf_{x \in \mathcal{X}_i} \sum_{j=1}^{k} p[j](x)$$

$$= \inf_{x \in \mathcal{X}_i} \sum_{j=1}^{k} p[j](x)$$

$$= 1 - \sup_{x \in \mathcal{X}_i} \sum_{m=1}^{i-1} p[m](x)$$

$$= 1 - \sup_{x \in \mathcal{X}_i} \sum_{m=1}^{i-1} \int_{-\infty}^{\infty} \phi(t+x[m]-x[j]) d\phi(t)$$

$$= 1 - \sup_{x \in \mathcal{X}_i} \sum_{m=1}^{i-1} \int_{-\infty}^{\infty} \phi(t) \phi^{k-2}(t) d\phi(t)$$

$$= 1 - \sum_{m=1}^{i-1} \int_{-\infty}^{\infty} \phi(t) \phi^{k-2}(t) d\phi(t)$$

$$= Q^*(i).$$

The supremum occurs when $x \in \mathcal{X}_i^{(2)}$. The last equality follows from the identity

$$(5.9) \quad \int_{-\infty}^{\infty} \phi^{k-2}(t) \phi(t) d\phi(t)$$

$$(k-1) \int_{-\infty}^{\infty} \phi^{k-1}(t+d) d\phi(t)$$

which can be proved by integration by parts.
Remark 5.1. If the procedure $\psi^M$ selects one population only, i.e. $X = x \in X_k$, then by Theorem 2.5. We have $p_k(x) \geq P^*$, so both procedures $\psi^B$ and $\psi^{B^*}$ will select the same population. In other words, when the maximum-type selection procedure does an excellent job, so does the Bayes-$P^*$ selection procedures. However, the converse is not true. In general, under PP*-condition, the subset selected by $\psi^B$ or $\psi^{B^*}$ is always smaller than the one selected by $\psi^M$, see Roth (1978) for some discussion.

Remark 5.2. For the case $k = 2$, $\psi^B = \psi^M$ a.e. for any given $X = x$; if $x \in X_2$, then $p_2(x) > P^*$, hence both $\psi^M$ and $\psi^B$ select the population $\pi_2$ which is associated with $x_2$. If $x \in X_1$ and $x_2 - \alpha/\sqrt{n} < x_1$ then $\psi^M$ and $\psi^B$ select both populations $\pi_1$ and $\pi_2$. Since $P(X_2 - \alpha/\sqrt{n} = x_1) = 0$, we have $\psi^B = \psi^M$ a.e.

Remark 5.3. Under the assumption of Theorem 5.1., and from the proof of it we have a lower bound on the value of $p_1(x) + \ldots + p_k(x)$ for any observation $x \in X_i$.

6. Application to Poisson Distributions and Poisson Processes

6.1. Poisson Distributions Case

Suppose that $\pi_1, \ldots, \pi_k$ are $k$ independent Poisson populations, where the independent observations $X_{i1}, \ldots, X_{in_i}$ from $\pi_i$ have the Poisson distribution with parameter $\lambda_i$ denoted by $P(\cdot | \lambda_i)$, $i = 1, \ldots, k$. Under non-informative prior $\tau(\lambda) = \lambda^{-2}$ for each population, if the best population is associated with the maximum parameter, we have

\[ p_i(x) = P(\lambda_i = \lambda_{[k]} | x) \]

(6.1)
\[ = \int_0^{\infty} \prod_{j \neq i} x_{m_j}^2 (y) (yn_j/n_j) dy_{m_i}^2 (y) \]
\( (6.2) \quad = \int \prod_{j \neq i} \chi^2_{m_j}(y) dx^2_{m_i}(y), \text{ if } n_1 = \ldots = n_k \)

where

\( (6.3) \quad m_i = 2n_i x_i + 1, \quad x_i = \sum_{j=1}^{n_i} x_{ij}/n_i. \)

where \( \chi^2_m \) is the cdf of chi-squared distribution with \( m \) degrees of freedom.

On the other hand, if we are interested in selecting the population with the smallest parameter \( \lambda \), then

\( (6.4) \quad P_i(x) = \int \prod_{j \neq i} [1 - \chi^2_{m_j}(yn_j/n_i)] dx^2_{m_i}(y) \)

\( (6.5) \quad = \int \prod_{j \neq i} [1 - \chi^2_{m_j}(y)] dx^2_{m_i}(y) \text{ if } n_1 = \ldots = n_k. \)

For this case, the simulation results for procedures \( \psi^B \) and \( \psi^B^* \) are tabulated in Table 5.

6.2. Poisson Processes Case

Suppose we have \( k \) independent Poisson processes \( \{X^{(1)}(t), \ldots, X^{(k)}(t)\} \) with expected arrival times equal to \( 1/\lambda_1, \ldots, 1/\lambda_k \), respectively. Hence for the processes \( \{X^{(i)}(t)\} \), the probability that there are \( m_i \) arrivals until time \( t_i \) is

\( (6.6) \quad P(X^{(i)}(t_i) = m_i | \lambda_i, t_i) = (t_i \lambda_i)^{m_i} \exp(-t_i \lambda_i)/m_i! . \)

If there exists no prior information, then we use the non-informative prior \( \pi(\lambda_i) = \lambda_i^{-1} \) for all processes. Let \( \bar{m} = (m_1, \ldots, m_k) \) and \( \bar{t} = (t_1, \ldots, t_k) \), it can be shown that the \( i \)th Poisson processes has the maximum parameter, i.e. the minimum expected waiting time, given \( (\bar{m}, \bar{t}) \) is
Here we list two special cases which are of interest.

(a) Observations of all processes are obtained in common length time intervals \([s_i, t_0+s_i]\). Since Poisson process is stationary, we can assume that \(s_i = 0\). In this case \(p_i(m, t)\) is independent of \(t\).

(b) All \(m_i\)'s are equal to \(m_0\), i.e. we fix \(m_0\) first, then get observations, the waiting time, \(t\). Let \(T_i\) be the waiting time of the \(m_0\)th arrival in the \(i\)th process, then \(T_i\) has a gamma distribution with pdf given by

\[
(6.8) \quad f(t_i) = \frac{\lambda_i^{m_i-1}}{(m_i)^{m_i}} e^{-\lambda_i t_i}, \quad t_i > 0.
\]

By using the same non-informative prior for \(\lambda\) as before, we get the same formulas for \(p_i(m, t)\). That is under case (b) the selection problem is identical with the selection problem on populations with gamma or exponential distribution.

Remark 6.1. Under non-informative prior, in comparing the subset selection problem for \(k\) Poisson distributions with the problem for \(k\) Poisson processes, it is easily seen that Poisson distribution model is a special case of Poisson processes model, with, \(t_i = n_i\) and \(m_i = n_i x_i\).

7. **Comparison of the Performance of \(B^*\), \(B\), \(M\) and \(MED\)**

Let \(\pi_i, i = 1, \ldots, k\) be \(k\) independent populations, where \(\pi_i\) has the associated cdf \(F_i(x, \theta_i) = F(x-\theta_i)\) with unknown location parameter \(\theta_i\). Let \(f_i(x, \theta_i) = f(x-\theta_i)\) be the pdf. Suppose the goal is to find a small (nontrivial) subset which contains the best.
The following subset selection procedure $\psi^{\text{MED}}$ based on sample medians is due to Gupta and Singh (1980).

$\psi^{\text{MED}}$: Select $i$ if and only if $Y_i > Y[k] - d'$

where $Y_i$ is the median of the $2m+1$ random observations from population $i$ and $Y[k] = \max Y_i$. The value of $d'$ is determined by the following equation so that the P*-condition is met.

\[(7.1) \int W^{k-1}(u+d')A(u)du = P^* \]

where

\[(7.2) A(u) = \frac{((2m+1)!/(m!)^2)[F(u)]^m[1-F(u)]^m f(u)} \]
\[(7.3) W(u) = I_y(u)(m+1, m+1) \]

where $I_y(a,b)$ is the incomplete beta function.

In this section we use Monte Carlo simulation techniques to compare the performance of selection procedures $\psi^B, \psi^{B*}, \psi^M$ and $\psi^{\text{MED}}$ in the normal means problem. Because selection procedures $\psi^M$ and $\psi^{\text{MED}}$ are not based on any prior information about the unknown parameters, we assume that the prior distribution $\tau$ for both procedures $\psi^B$ and $\psi^{B*}$ is non-informative. Since procedure $\psi^M$ satisfies both the P*-condition and the PP*-condition with respect to the non-informative prior, it makes sense to compare the Bayes-P* procedures $\psi^B$ and $\psi^{B*}$ with $\psi^M$ and compare $\psi^M$ with $\psi^{\text{MED}}$ in terms of efficiency. Furthermore, we assume the true distributions to be non-normal distributions, namely, the logistic, Laplace (the double exponential) and the gross error model (the contaminated) distribution, but keep the selection procedure unchanged (i.e. still based on the normal assumption) and compute the efficiency. The Monte Carlo simulation results for both equal distances of the parameters and slippage cases are tabulated. In the
simulation study all generated random variables are adjusted to have variance one. Each time five random numbers with indicated distribution were generated for each population. All four procedures are applied to the same data. The simulation process was repeated one hundred times. The relative frequency of selecting population $\pi_i$ is used as an approximation to the probability of selecting population $\pi_i$. The sum of relative frequency of selecting each population is treated as an approximation of the expected selected size. The efficiency (EFF) of each selection procedure is approximated by the ratio of relative frequency of selecting the best one to the expected selection size.

The simulation results indicate that in all cases we have the performance

$$B^* > B > M,$$

where the symbol "\(\succ\)" stands for better than.

For small sample size, the efficiency of rule $M$ tends to be larger than $\psi_{MED}$ under the $P*$-condition.

Remark 7.1. The gross error model we used has the density function

$$f(x-\theta) = (1-c)\varphi(x-\theta) + \frac{c}{4}\varphi\left(x-\frac{\theta}{\sqrt{4}}\right), \quad c = .15$$

for which $\varphi$ is the pdf of $N(0,1)$ and the variance for this distribution is $(1-c) + 16c = 3.25$.

In the tables, the efficiency (EFF) of a procedure $\psi$, given parameter $\theta$, is defined by

$$\text{EFF}_{\theta}(\psi) = \frac{P_{\theta}(\text{CS}|\psi)}{E_{\theta}(S|\psi)}$$

where $E_{\theta}(S|\psi)$ is the expected selected size.
Discussion of the Tables:

For Table 1 and Table 2 (equal distances case) the value of $P^*$ is .99 and .90 respectively, the common sample size $n = 5$, and $k = 5$. If the $k$ populations have normal distributions with the unknown parameter configuration $(\theta, \theta + \Delta, \ldots, \theta + (k-1)\Delta)$ and common variance one, then from both tables the performance based on either the efficiency or the expected selected size is

$$\psi^{B^*} > \psi^B > \psi^M,$$

if the PP*-condition is considered; and

$$\psi^M > \psi^\text{MED}$$
under the P*-condition.

When the true distributions are not normal, but the logistic, Laplace or the gross error model, the simulation results are very close to the normal case. This suggests the four procedures are reasonably robust. From Table 2 all efficiencies are larger than the corresponding ones in Table 1. This is to be expected because the value of $P^*$ is smaller in the second table.

For Table 3 and Table 4 (slippage case) the value of $P^*$ is .99 and .90 respectively, the common sample size $n = 5$, and $k = 5$. If the $k$ populations have normal distributions with unknown parameter configurations $(\theta, \ldots, \theta, \theta + \Delta)$ and common variance one, then from both tables the performance is the same as the equal distances case.

Note that in both equal distances and slippage cases when $\Delta \sqrt{n} > 1$, that means the largest population mean and the second largest population mean are not very close, the Bayes-P* selection procedures $\psi^B$ and $\psi^{B^*}$, with respect to the locally uniform priors, always satisfy not only the PP*-condition but also the estimated $P(\text{CS}) \cdot P^*$, and the expected
selected size of the Bayes-P* procedures is much less than the selection procedures $\psi^M$ and $\psi^{\text{MED}}$. For example, in the normal equal distance case, $P^* = .99$, $k = 5$ and $\sqrt{n} = 4$, we have

$$E(S|\psi^{\text{MED}}) - E(S|\psi^B) = .370;$$

In the normal slippage case, $P^* = .99$, $k = 5$ and $\sqrt{n} = 4$, we have

$$E(S|\psi^{\text{MED}}) - E(S|\psi^B) = 1.560.

8. DISCUSSION:

The Bayes-P* selection procedures $\psi^B$ and $\psi^{B*}$ are highly-efficient and have the following advantages.

a. These procedures can apply to any family of distributions, even their mixtures, and do not need equal sample size for each population.

b. Good prior information will not be ignored. Even under non-informative situation, they still perform well.

c. They are robust in terms of the loss function. We do not even need to specify or to know the exact form of the loss function before we make a decision. There will automatically be a Bayes decision procedure under the control condition and the assumptions given by Theorem 3.3 and Theorem 3.4.

d. The weight or contribution of each population in the selected subset is known.

e. Compared with the classical maximum-type or average-type selection procedures, the Bayes-P* selection procedure is less sensitive to the total number of populations. For example, in the normal case, if there are some newly added populations with very small sample means, from Section 4 we can see that the selected subset for
Bayes-$P^*$ selection procedure is nearly unchanged; however, the selected subset for the classical procedures may increase rapidly.

f. Based on the simulation results of Section 7, Bayes-$P^*$ selection procedure is robust if the true family of distributions for each population is symmetric.

The only disadvantage in using the proposed selection procedures is that the computation of the posterior probabilities needs more work than the classical selection procedures which can use some precalculated tables; however, this disadvantage can be offset by the use of computers. In fact, we need not evaluate all $p_i$'s, but only a few of the large ones.
TABLE 1

Efficiency (EFF) and expected selected size (ES) (based on simulation) of \( \psi^*, \psi, \psi, \psi \) and \( \psi_{MED} \), under normal assumption, when the unknown means of the \( k \) populations are \( \theta, \theta + \Delta, \ldots, \theta + (k-1)\Delta \); the common variance = 1, common sample size \( n = 5 \) and the prior for \( \psi^* \) is the non-informative prior.

\[
k = 5, \quad P^* = .99
\]

<table>
<thead>
<tr>
<th>( \sqrt{n} )</th>
<th>( \psi^* )</th>
<th>( \psi )</th>
<th>( \psi )</th>
<th>( \psi_{MED} )</th>
<th>( \psi_{MED} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.5</td>
<td>.254</td>
<td>.238</td>
<td>.208</td>
<td>.203</td>
<td>.203</td>
</tr>
<tr>
<td></td>
<td>3.809</td>
<td>4.110</td>
<td>4.810</td>
<td>4.810</td>
<td>4.810</td>
</tr>
<tr>
<td></td>
<td>.250</td>
<td>.233</td>
<td>.207</td>
<td>.202</td>
<td>.202</td>
</tr>
<tr>
<td></td>
<td>3.963</td>
<td>4.290</td>
<td>4.840</td>
<td>4.940</td>
<td>4.940</td>
</tr>
<tr>
<td></td>
<td>3.774</td>
<td>4.120</td>
<td>4.720</td>
<td>4.940</td>
<td>4.940</td>
</tr>
<tr>
<td></td>
<td>.248</td>
<td>.250</td>
<td>.207</td>
<td>.201</td>
<td>.201</td>
</tr>
<tr>
<td>1</td>
<td>.333</td>
<td>.305</td>
<td>.250</td>
<td>.232</td>
<td>.232</td>
</tr>
<tr>
<td></td>
<td>2.977</td>
<td>3.200</td>
<td>4.000</td>
<td>4.310</td>
<td>4.310</td>
</tr>
<tr>
<td></td>
<td>.332</td>
<td>.304</td>
<td>.234</td>
<td>.224</td>
<td>.224</td>
</tr>
<tr>
<td></td>
<td>3.005</td>
<td>3.290</td>
<td>4.260</td>
<td>4.460</td>
<td>4.460</td>
</tr>
<tr>
<td></td>
<td>.336</td>
<td>.302</td>
<td>.246</td>
<td>.217</td>
<td>.217</td>
</tr>
<tr>
<td></td>
<td>2.941</td>
<td>3.280</td>
<td>4.030</td>
<td>4.600</td>
<td>4.600</td>
</tr>
<tr>
<td></td>
<td>.329</td>
<td>.304</td>
<td>.233</td>
<td>.214</td>
<td>.214</td>
</tr>
<tr>
<td></td>
<td>3.032</td>
<td>3.290</td>
<td>4.290</td>
<td>4.670</td>
<td>4.670</td>
</tr>
<tr>
<td>2</td>
<td>.541</td>
<td>.486</td>
<td>.417</td>
<td>.368</td>
<td>.368</td>
</tr>
<tr>
<td></td>
<td>1.847</td>
<td>2.050</td>
<td>2.400</td>
<td>2.720</td>
<td>2.720</td>
</tr>
<tr>
<td></td>
<td>.541</td>
<td>.481</td>
<td>.417</td>
<td>.351</td>
<td>.351</td>
</tr>
<tr>
<td></td>
<td>1.839</td>
<td>2.000</td>
<td>2.400</td>
<td>2.850</td>
<td>2.850</td>
</tr>
<tr>
<td></td>
<td>.541</td>
<td>.493</td>
<td>.515</td>
<td>.362</td>
<td>.362</td>
</tr>
<tr>
<td></td>
<td>1.884</td>
<td>2.010</td>
<td>2.410</td>
<td>2.760</td>
<td>2.760</td>
</tr>
<tr>
<td></td>
<td>.559</td>
<td>.510</td>
<td>.437</td>
<td>.368</td>
<td>.368</td>
</tr>
<tr>
<td></td>
<td>1.778</td>
<td>1.360</td>
<td>2.290</td>
<td>2.720</td>
<td>2.720</td>
</tr>
<tr>
<td>4</td>
<td>.825</td>
<td>.730</td>
<td>.676</td>
<td>.575</td>
<td>.575</td>
</tr>
<tr>
<td></td>
<td>1.212</td>
<td>1.370</td>
<td>1.480</td>
<td>1.740</td>
<td>1.740</td>
</tr>
<tr>
<td></td>
<td>.855</td>
<td>.806</td>
<td>.694</td>
<td>.573</td>
<td>.573</td>
</tr>
<tr>
<td></td>
<td>1.169</td>
<td>1.240</td>
<td>1.440</td>
<td>1.730</td>
<td>1.730</td>
</tr>
<tr>
<td></td>
<td>.821</td>
<td>.746</td>
<td>.660</td>
<td>.566</td>
<td>.566</td>
</tr>
<tr>
<td></td>
<td>1.217</td>
<td>1.340</td>
<td>1.470</td>
<td>1.700</td>
<td>1.700</td>
</tr>
<tr>
<td></td>
<td>.865</td>
<td>.800</td>
<td>.671</td>
<td>.577</td>
<td>.577</td>
</tr>
<tr>
<td></td>
<td>1.156</td>
<td>1.250</td>
<td>1.490</td>
<td>1.750</td>
<td>1.750</td>
</tr>
</tbody>
</table>
TABLE 2

Efficiency (EFF) and expected selected size (ES) (based on simulation) of $\psi^*, \psi, \psi$ and MED, under normal assumption, when the unknown means of the $k$ populations are $(0, \gamma A, \ldots, (k-1)A)$; the common variance = 1, common sample size $n = 5$ and the prior for $\psi^*$ and $\psi$ is the non-informative prior.

$$k = 5, \quad p^* = .90$$

<table>
<thead>
<tr>
<th>$\sqrt{\Delta n}$</th>
<th>B* $\psi$</th>
<th>B $\psi$</th>
<th>M $\psi$</th>
<th>MED $\psi$</th>
<th>normal EFF</th>
<th>normal ES</th>
<th>logistic EFF</th>
<th>logistic ES</th>
<th>Laplace EFF</th>
<th>Laplace ES</th>
<th>gross error EFF</th>
<th>gross error ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>.334</td>
<td>.303</td>
<td>.239</td>
<td>.226</td>
<td>2.449</td>
<td>2.870</td>
<td>4.020</td>
<td>4.200</td>
<td>2.447</td>
<td>2.359</td>
<td>.332</td>
<td>2.359</td>
</tr>
<tr>
<td></td>
<td>.353</td>
<td>.313</td>
<td>.252</td>
<td>.231</td>
<td>.334</td>
<td>.285</td>
<td>.333</td>
<td>.252</td>
<td>2.447</td>
<td>2.359</td>
<td>.332</td>
<td>2.359</td>
</tr>
<tr>
<td></td>
<td>2.611</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.333</td>
<td>2.611</td>
</tr>
<tr>
<td>1</td>
<td>.493</td>
<td>.444</td>
<td>.327</td>
<td>.297</td>
<td>1.807</td>
<td>2.160</td>
<td>3.030</td>
<td>3.270</td>
<td>.516</td>
<td>1.817</td>
<td>.502</td>
<td>1.839</td>
</tr>
<tr>
<td></td>
<td>1.839</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.502</td>
<td>1.839</td>
<td>.502</td>
<td>1.839</td>
</tr>
<tr>
<td>2</td>
<td>.773</td>
<td>.671</td>
<td>.535</td>
<td>.490</td>
<td>1.275</td>
<td>1.490</td>
<td>1.870</td>
<td>2.040</td>
<td>.756</td>
<td>1.257</td>
<td>.768</td>
<td>1.257</td>
</tr>
<tr>
<td></td>
<td>.770</td>
<td>.678</td>
<td>.541</td>
<td>.490</td>
<td>1.275</td>
<td>1.490</td>
<td>1.870</td>
<td>2.040</td>
<td>1.291</td>
<td>1.257</td>
<td>.768</td>
<td>1.257</td>
</tr>
<tr>
<td></td>
<td>1.227</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.786</td>
<td>1.227</td>
<td>.786</td>
<td>1.227</td>
</tr>
<tr>
<td>4</td>
<td>.992</td>
<td>.980</td>
<td>.877</td>
<td>.719</td>
<td>1.008</td>
<td>1.020</td>
<td>1.140</td>
<td>1.390</td>
<td>.956</td>
<td>1.020</td>
<td>.981</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td>.981</td>
<td>.935</td>
<td>.862</td>
<td>.800</td>
<td>1.008</td>
<td>1.020</td>
<td>1.140</td>
<td>1.390</td>
<td>1.042</td>
<td>1.020</td>
<td>.981</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td>1.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.990</td>
<td>1.010</td>
<td>.990</td>
<td>1.010</td>
</tr>
</tbody>
</table>


TABLE 3

Efficiency (EFF) and expected selected size (ES) (based on simulation of $B^*$, $B$, $M$ and $MED$, under normal assumption, when the unknown means of the $k$ populations are $\theta, \ldots, \theta + \Delta$; the common variance = 1, common sample size $n = 5$ and the prior for $B^*$ and $B^*$ is the non-informative prior.

$k = 5, \quad p^* = .99$

<table>
<thead>
<tr>
<th>$\sqrt{n}$</th>
<th>normal</th>
<th>logistic</th>
<th>Laplace</th>
<th>gross error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EFF</td>
<td>ES</td>
<td>EFF</td>
<td>ES</td>
</tr>
<tr>
<td>0.5</td>
<td>$B^*$</td>
<td>.207</td>
<td>4.192</td>
<td>.218</td>
</tr>
<tr>
<td></td>
<td>$B$</td>
<td>.204</td>
<td>4.520</td>
<td>.212</td>
</tr>
<tr>
<td>1</td>
<td>$B^*$</td>
<td>.228</td>
<td>4.101</td>
<td>.240</td>
</tr>
<tr>
<td></td>
<td>$B$</td>
<td>.220</td>
<td>4.400</td>
<td>.224</td>
</tr>
<tr>
<td></td>
<td>MED</td>
<td>.204</td>
<td>4.890</td>
<td>.201</td>
</tr>
<tr>
<td>2</td>
<td>$B^*$</td>
<td>.278</td>
<td>3.598</td>
<td>.274</td>
</tr>
<tr>
<td></td>
<td>$B$</td>
<td>.252</td>
<td>3.970</td>
<td>.253</td>
</tr>
<tr>
<td></td>
<td>MED</td>
<td>.208</td>
<td>4.800</td>
<td>.207</td>
</tr>
<tr>
<td>4</td>
<td>$B^*$</td>
<td>.513</td>
<td>1.950</td>
<td>.528</td>
</tr>
<tr>
<td></td>
<td>$B$</td>
<td>.457</td>
<td>2.190</td>
<td>.474</td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>.352</td>
<td>2.840</td>
<td>.356</td>
</tr>
<tr>
<td></td>
<td>MED</td>
<td>.267</td>
<td>3.750</td>
<td>.267</td>
</tr>
</tbody>
</table>
TABLE 4

Efficiency (EFF) and Expected selected Size (ES) (based on simulation of $\psi^{B*}$, $\psi^B$, $\psi^M$ and $\psi^{MED}$, under normal assumption when the unknown means of the k populations are $\theta_1, \ldots, \theta_k$; the common variance $= 1$, common sample size $n = 5$ and the prior for $\psi^B$ and $\psi^{B*}$ is the non-informative prior.

$k = 5, \quad P^* = .90$

<table>
<thead>
<tr>
<th>$\Delta \sqrt{n}$</th>
<th>$\psi^{B*}$</th>
<th>$\psi^B$</th>
<th>$\psi^M$</th>
<th>$\psi^{MED}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>normal EFF</td>
<td>ES</td>
<td>logistics EFF</td>
<td>ES</td>
</tr>
<tr>
<td>.5</td>
<td>.245</td>
<td>2.813</td>
<td>.267</td>
<td>2.632</td>
</tr>
<tr>
<td></td>
<td>.243</td>
<td>3.250</td>
<td>.262</td>
<td>3.020</td>
</tr>
<tr>
<td></td>
<td>.208</td>
<td>4.480</td>
<td>.219</td>
<td>4.430</td>
</tr>
<tr>
<td></td>
<td>.201</td>
<td>4.470</td>
<td>.204</td>
<td>4.600</td>
</tr>
<tr>
<td>1</td>
<td>.306</td>
<td>2.645</td>
<td>.349</td>
<td>2.508</td>
</tr>
<tr>
<td></td>
<td>.288</td>
<td>3.060</td>
<td>.315</td>
<td>2.890</td>
</tr>
<tr>
<td></td>
<td>.219</td>
<td>4.420</td>
<td>.231</td>
<td>4.160</td>
</tr>
<tr>
<td></td>
<td>.228</td>
<td>4.250</td>
<td>.238</td>
<td>4.420</td>
</tr>
<tr>
<td>2</td>
<td>.450</td>
<td>2.149</td>
<td>.463</td>
<td>2.167</td>
</tr>
<tr>
<td></td>
<td>.369</td>
<td>2.500</td>
<td>.385</td>
<td>2.520</td>
</tr>
<tr>
<td></td>
<td>.269</td>
<td>3.720</td>
<td>.272</td>
<td>3.670</td>
</tr>
<tr>
<td></td>
<td>.253</td>
<td>3.950</td>
<td>.247</td>
<td>4.050</td>
</tr>
<tr>
<td>4</td>
<td>.875</td>
<td>1.143</td>
<td>.862</td>
<td>1.155</td>
</tr>
<tr>
<td></td>
<td>.794</td>
<td>1.260</td>
<td>.775</td>
<td>1.290</td>
</tr>
<tr>
<td></td>
<td>.645</td>
<td>1.550</td>
<td>.610</td>
<td>1.640</td>
</tr>
</tbody>
</table>
TABLE 5

For procedure $\psi^B*$ and $\psi^B$ and the parameter configurations (.5,1,...,5k) of k Poisson populations, this table gives the values (based on simulation) of the probability of selecting the population with parameter $.5i$, $i = 1,...,k$ and the expected selected size $ES$. The prior distribution for each population is $\pi(\lambda) = \lambda^{-\frac{1}{2}}$.

$n = 10$

<table>
<thead>
<tr>
<th>k</th>
<th>p*</th>
<th>.99</th>
<th>.95</th>
<th>.90</th>
<th>.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi^B*$</td>
<td>$\psi^B$</td>
<td>$\psi^B*$</td>
<td>$\psi^B$</td>
<td>$\psi^B*$</td>
<td>$\psi^B$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.933</td>
<td>1.000</td>
<td>.966</td>
<td>.990</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.670</td>
<td>.820</td>
<td>.450</td>
<td>.670</td>
</tr>
<tr>
<td></td>
<td>ES</td>
<td>1.663</td>
<td>1.820</td>
<td>1.415</td>
<td>1.660</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.998</td>
<td>1.000</td>
<td>.990</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.741</td>
<td>.850</td>
<td>.338</td>
<td>.630</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.205</td>
<td>.330</td>
<td>.120</td>
<td>.130</td>
</tr>
<tr>
<td></td>
<td>ES</td>
<td>1.944</td>
<td>2.180</td>
<td>1.498</td>
<td>1.760</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>.988</td>
<td>1.000</td>
<td>.981</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.754</td>
<td>.820</td>
<td>.560</td>
<td>.730</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.369</td>
<td>.490</td>
<td>.070</td>
<td>.140</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.075</td>
<td>.110</td>
<td>.041</td>
<td>.060</td>
</tr>
<tr>
<td></td>
<td>ES</td>
<td>2.187</td>
<td>2.420</td>
<td>1.653</td>
<td>1.930</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>.996</td>
<td>1.000</td>
<td>.981</td>
<td>.990</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.737</td>
<td>.850</td>
<td>.503</td>
<td>.650</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.067</td>
<td>.090</td>
<td>.015</td>
<td>.030</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ES</td>
<td>2.154</td>
<td>2.410</td>
<td>1.607</td>
<td>1.840</td>
<td>1.498</td>
</tr>
</tbody>
</table>
REFERENCES


Two new selection procedures, called nonrandomized and randomized Bayes-P* procedures are defined for selecting a small nonempty subset of k populations which contains the best population. It is shown that these procedures have some optimal properties. If we restrict attention to the class D(D*) of all nonrandomized (randomized) selection procedures, which satisfy the PP*-condition, that is the posterior probability of a correct selection, for any given observation \( X = x \), is not less than \( P^* \), a predetermined number between \( 1/k \) and 1, then these two new selection procedures are shown also to be (over)