BAYESIAN PROCEDURES FOR RANKING
AND SELECTION PROBLEMS

Z. GOVINDARAJULU AND CHARLES HARVEY*

(Received Nov. 15, 1971)

1. Introduction and summary

It was noticed by several research workers, e.g. Bahadur [1], Bechhofer [3], that after carrying out a test of homogeneity, one is often interested in either ranking the treatments or selecting one or more of the 'best' treatments. Bahadur [1] and Bahadur and Goodman [2] proposed certain impartial decision rules for selecting one or more populations from among k populations. Bechhofer [3] proposed a procedure for selecting the 'best' of k populations with at least a preassigned probability P^* of 'correct selection' using an 'indifference zone' approach. Seal [12], Gupta [8], and Gupta and Sobel [9] proposed procedures for selecting a subset from k populations which would contain the best population with probability at least P^*. Somerville [14] and Fairweather [7] proposed minimax two-stage procedures for selecting the best population.

Procedures for selecting a single population using a Bayesian decision theoretic approach and linear loss functions were developed by Raiffa and Schlaiffer [11]. More recently, Bland and Bratcher [4] and Bratcher [5] have studied Bayesian procedures for ranking binomial populations by a succession of pairwise comparisons; and Deeley and Gupta [6] have studied Bayesian procedures for selecting a subset from k populations which involve a general class of linear loss functions.

In this paper we present a unified treatment of ranking and selection problems from a Bayesian decision theoretic point of view. Consider k populations $\Pi_1, \cdots, \Pi_k$ whose distributions belong to the same parametrized family. Let $f(x|\theta, \boldsymbol{\omega})$ denote the likelihood function for these distributions, where θ is the single parameter of interest and $\boldsymbol{\omega}$ denotes possible nuisance parameters. Let θ_i denote the true value of the parameter θ_i associated with the ith population, Π_i, $i=1, \cdots, k$. Assume that prior distributions have been assigned to the parameters $(\theta_i, \boldsymbol{\omega}_i)$, and random samples have been obtained from the populations Π_i, $i=1, \cdots, k$.

---

* Presently at Dickinson College.

Let $h_i(\theta)$ denote the resulting marginal posterior density concerning the parameter θ_i, $i=1,\cdots,k$. Also let P'' and E'' denote probabilities and expectations taken with respect to posterior distributions.

In Section 2, a ranking procedure is developed under certain terminal losses so that, roughly speaking, the decision-maker is led to assign higher rank to Π_j than to Π_i if he believes that θ_j is probably greater than θ_i in the sense that $P''(\theta_i<\theta_j)>P''(\theta_j<\theta_i)$. Ranking procedures are derived in Section 3 for several types of situations in which the terminal losses are specified by functions $w_i(\theta_i)$ of the true parameter values θ_i for the population or populations selected from the ranking according to a given procedure. In Section 4, the ranking procedures of Sections 2 and 3 are shown to be equivalent under certain circumstances. This equivalence is shown in part by obtaining sufficient conditions for the monotonicity of the ratio of two posterior densities, $h_j(\theta)/h_i(\theta)$. These conditions are similar to the sufficient conditions for the monotone likelihood ratio property. Section 5 consists of a discussion of certain problems of selecting a subset from the k populations.

2. Ranking procedures

Suppose that one wishes to report a ranking $R: \theta_{i_1}<\theta_{i_2}<\cdots<\theta_{i_k}$ of k populations having distributions as described in the introduction. Following a Bayesian decision theoretic approach, the decision-maker may assign a loss function $l(R,\boldsymbol{\theta})$ which depends on the ranking R selected, and on the true values $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)'$ of the parameters of interest. Having obtained a posterior density $h(\boldsymbol{\theta})$ concerning the parameters $\boldsymbol{\theta}$, he will then decide upon that ranking R which minimizes the posterior expected loss,

$$(2.1) \qquad L(R)=\int_\Omega\cdots\int_\Omega l(R,\boldsymbol{\theta})h(\boldsymbol{\theta})d\boldsymbol{\theta} \ .$$

A ranking $R: \theta_{i_1}<\theta_{i_2}<\cdots<\theta_{i_k}$ will be denoted by an ordered sequence $R=(i_1,i_2,\cdots,i_k)$ where $1\leq i_j\leq k$, $1\leq j\leq k$. Here i_j denotes the index of that population which is assigned rank j, $j=1,\cdots,k$. The indices i_j are sometimes called anti-ranks. Loss functions will be presented so that, intuitively speaking, if the decision-maker tends to believe that $\theta_i<\theta_j$ for two indices i, j, he will be led to assign higher rank to Π_j than to Π_i, i.e. he will choose $R=(i_1,\cdots,i_k)$ where the integer i precedes the integer j.

Perhaps the simplest loss function is $l_1(R,\boldsymbol{\theta})$, which assigns loss 1 to any ranking R which is not correct given the true parameter values $\boldsymbol{\theta}$, and loss 0 to the correct ranking. Assuming this loss function, the posterior expected loss for a ranking R is

$$(2.2) \qquad L_1(R) = 1 - P''(R) = 1 - \int_R h(\boldsymbol{\theta}) d\boldsymbol{\theta} \, ,$$

where by an abuse of terminology, R denotes the event that R is the correct ranking.

The loss function $l_1(R, \boldsymbol{\theta})$ is thus attractive because of its simplicity ; however it seems to possess the following two disadvantages.

(i) $l_1(R, \boldsymbol{\theta})$ may be an inappropriate formulation of the decision-maker's losses in that it assigns equal loss to any incorrect ranking, whether it be only mildly incorrect or drastically incorrect.

(ii) As the number k of populations increases, it rapidly becomes more difficult to evaluate by numerical quadrature the $k!$ distinct k-dimensional integrals (2.2) that are needed in order to determine the optimal ranking.

Thus it would be useful to modify l_1 so that differing degrees of error are reflected in the losses. It would also be desirable for such a modified loss function to lead to a more tractable computational procedure. This section is concerned with such a modification of l_1. We begin by introducing the following terminology.

A transposition of two adjacent indices in a ranking will be called an *interchange*. Thus, for $j = 2, \cdots, k$ the rankings $R = (i_1, \cdots, i_{j-1}, i_j, \cdots, i_k)$ and $R' = (i_1, \cdots, i_j, i_{j-1}, \cdots, i_k)$ differ by an interchange. A finite sequence of rankings $R_0, R_1, \cdots, R_m$ where any two adjacent rankings differ by an interchange will be referred to as a *chain*. The number m of interchanges will be called its length.

DEFINITION. For any two rankings R and R', the *distance* $d(R, R')$ from R to R' will be defined as the minimum number of interchanges in any chain from R to R'.

The function $d(R, R')$ satisfies the axioms of a metric distance, that is
(i) $d(R, R') = 0$ if and only if $R = R'$.
(ii) $d(R, R') = d(R', R)$.
(iii) $d(R, R'') \leq d(R, R') + d(R', R'')$.

As an example of this notion, consider $R : \theta_3 < \theta_1 < \theta_2$ and $R^* : \theta_1 < \theta_2 < \theta_3$. Then $d(R, R') = 2$ since a chain from R to R' of minimum length is given by ; $(\theta_3 < \theta_1 < \theta_2)$, $(\theta_1 < \theta_3 < \theta_2)$, $(\theta_1 < \theta_2 < \theta_3)$.

Let $R^0(\boldsymbol{\theta})$ denote the true or correct ranking for given parameter values $\boldsymbol{\theta}$. The distance $d(R, R^0(\boldsymbol{\theta}))$ between an assigned ranking R and the correct ranking $R^0(\boldsymbol{\theta})$ may be interpreted as the number of corrections (interchanges) needed to correct R.

A loss function $l_2(R, \boldsymbol{\theta})$ may be proposed in terms of the distance $d(R, R^0(\boldsymbol{\theta}))$, namely

$$(2.3) \qquad l_2(R, \boldsymbol{\theta}) = d(R, R^0(\boldsymbol{\theta})) \, .$$

It follows that the posterior expected loss under an assigned ranking R is given by

$$(2.4) \qquad\qquad L_2(R) = \Sigma_Q d(R, Q)\, P''(Q)$$

where the summation is over all $k!$ possible rankings Q, and $P''(Q)$ denotes the posterior probability that Q is the true ranking. The decision-maker can determine the optimal ranking by evaluating each of the $k!$ distinct k-dimensional integrals $P''(Q)$. Thus the procedure for determining the optimal ranking under l_2 appears to involve essentially the same computational effort as that for determining the optimal ranking under l_1. The following somewhat technical material will lead to a more effective computational procedure for determining the optimal ranking under the loss l_2.

Consider first those transformations γ of the class of all rankings which are defined as follows. For a permutation g of the integers $1, \cdots, k$, let γ denote the induced transformation of rankings $R = (i_1, \cdots, i_k)$,

$$(2.5) \qquad\qquad \gamma R = (g(i_1), \cdots, g(i_k)) .$$

Then for any rankings R, R', and any such transformation γ,

$$(2.6) \qquad\qquad d(\gamma R, \gamma R') = d(R, R') .$$

To verify (2.6), note that if two rankings Q and Q' differ by an interchange, e.g. of i_{p-1} and i_p, then γQ and $\gamma Q'$ also differ by an interchange, namely of $g(i_{p-1})$ and $g(i_p)$. Applying γ to each ranking in a minimal chain from R to R', i.e. a chain of length $d(R, R')$, one obtains a chain of length $d(R, R')$ from γR to $\gamma R'$. Hence $d(\gamma R, \gamma R') \leq d(R, R')$. Likewise, applying γ^{-1} to each ranking in a minimal chain from γR to $\gamma R'$, one concludes that $d(R, R') \leq d(\gamma R, \gamma R')$. Hence (2.6) follows.

The following algorithm constructs a chain from an arbitrary ranking $R = (i_1, \cdots, i_k)$ to the ranking $I = (1, 2, \cdots, k)$ which will be shown to be of minimal length. One proceeds inductively on the index $m = 1, 2, \cdots, k$ so that for each value of m a portion of the chain is constructed ending in a ranking R_m whose first m entries are in increasing order. Hence R_k must equal I. Let $R_1 = R$. For $m = 2$, interchange i_1 and i_2 in R if $i_1 > i_2$, otherwise let $R_2 = R$. Inductively, given the ranking $R_m = (i_1', \cdots, i_m', i_{m+1}, \cdots, i_k)$, compare i_{m+1} with $i_m', i_{m-1}', \cdots, i_1'$ successively, and interchange i_{m+1} with each i_l' for which $i_l' > i_{m+1}$. Letting $m = 1, 2, \cdots, k$, the chain is constructed. It should be noted that the number of interchanges leading from R_m to R_{m+1} equals the number of indices i_l', $1 \leq l \leq m$, such that $i_l' > i_{m+1}$.

As an example of this algorithm, the following chain leads from $R = (2, 1, 4, 3)$ to I; $R_1 = (2, 1, 4, 3)$, $R_2 = (1, 2, 4, 3)$, $R_3 = R_2$, $R_4 = (1, 2, 3, 4)$. Here $d(R, I) = 2$.

LEMMA 2.1. *For any ranking* $R=(i_1,\cdots,i_k)$,

(2.7) $$d(R,I)=\sum_{m=1}^{k}\sum_{l<m}H(i_l-i_m)$$

where $H(n)=1$ *if* $n\geq 0$, $H(n)=0$ *if* $n<0$ *is the Heaviside function.*

PROOF. The double sum in (2.7) equals the number of pairs of entries i_l, i_m in R, $l<m$, for which $i_l>i_m$. This equals the number of interchanges in the chain from R to I constructed according to the previous algorithm. On the other hand, consider any pair i_l, i_m, $l<m$ where $i_l>i_m$. Any chain leading from R to I must contain an interchange of i_l and i_m. Thus any chain from R to I must have length at least the quantity specified in (2.7), so the proof is complete.

LEMMA 2.2. *For any integer* h, $2\leq h\leq k$, *let* $R=(i_1,\cdots,i_{h-1},i_h,\cdots,i_k)$ *and* $R'=(i_1,\cdots,i_h,i_{h-1},\cdots,i_k)$ *denote two rankings which differ by an interchange of the hth and (h−1)st entries.*
  Then, for any ranking $Q=(j_1,\cdots,j_k)$, *the difference* $d(R,Q)-d(R',Q)$ *equals* -1 *if* i_{h-1} *preceeds* i_h *in* Q, *and* $+1$ *if* i_h *preceeds* i_{h-1} *in* Q.

PROOF. Consider first the case $Q=I$. Comparing the representations (2.7) of $d(R,I)$ and $d(R',I)$, the terms where $m\neq h-1$ or h cancel leaving,

(2.8) $$d(R,I)-d(R',I)=\sum_{l<h-1}H(i_l-i_{h-1})+\sum_{l<h}H(i_l-i_h)$$
$$-\sum_{l<h-1}H(i_l'-i_{h-1}')-\sum_{l<h}H(i_l'-i_h')\ .$$

Here i_j' refers to entries in R'. Note that $i_j'=i_j$ for $j\neq h-1$ or h, $i_{h-1}'=i_h$, and $i_h'=i_{h-1}$. Thus by combining the first and fourth terms, and the second and third terms, (2.8) reduces to

(2.9) $$d(R,I)-d(R',I)=-H(i_h-i_{h-1})+H(i_{h-1}-i_h)\ .$$

This proves the lemma when $Q=I$.

Consider next a general ranking Q. Let γ denote the transformation defined by (2.5) which carries Q into I. Thus γ is defined by the permutation $g(j_m)=m$, $m=1,\cdots,k$. Assume that i_{h-1} preceeds i_h in Q. Then $g(i_{h-1})<g(i_h)$, i.e. $g(i_{h-1})$ preceeds $g(i_h)$ in I. The rankings γR and $\gamma R'$ differ only by an interchange of $g(i_{h-1})$ and $g(i_h)$, so the above proof implies that $d(\gamma R,I)-d(\gamma R',I)=-1$. Thus by (2.6), $d(R,Q)-d(R',Q)=d(\gamma R,\gamma Q)-d(\gamma R',\gamma Q)=-1$. By a similar argument, if i_h preceeds i_{h-1}, then $d(R,Q)-d(R',Q)=+1$.

THEOREM 2.1. *If two rankings* R *and* R' *differ only by an interchange of* $\theta_i<\theta_j$ *in* R *to* $\theta_j<\theta_i$ *in* R', *then*

(2.10) $L_2(R) - L_2(R') = -\mathrm{P}''(\theta_i < \theta_j) + \mathrm{P}''(\theta_j < \theta_i)$.

PROOF. The difference $L_2(R) - L_2(R')$ equals $\Sigma_Q(d(R, Q) - d(R', Q)) \cdot \mathrm{P}''(Q)$ by (2.4) where the summation is over all $k!$ rankings. The event $\theta_i < \theta_j$, as a subset of the space of values of $\boldsymbol{\theta}$, is the disjoint union of those rankings $Q: \theta_{j_1} < \cdots < \theta_{j_k}$ for which θ_i preceeds θ_j, and $\theta_j < \theta_i$ is the disjoint union of those rankings for which θ_j preceeds θ_i. However, by the previous lemma, $d(R, Q) - d(R', Q)$ equals -1 for rankings of the first type and equals $+1$ for rankings of the second type. Thus (2.10) follows.

Theorem 2.1 provides a simple criterion for comparing any two rankings which differ only by an interchange. The decision-maker prefers that ranking such that Π_j is assigned higher rank than Π_i provided that he believes θ_j to be probably greater than θ_i in the sense that $\mathrm{P}''(\theta_i < \theta_j) > \mathrm{P}''(\theta_j > \theta_i)$. The result also justifies the following computational procedure for determining the ranking which minimizes the posterior expected loss $L_2(R)$.

Procedure for implementing $l_2(R, \boldsymbol{\theta})$

(i) Compute each of the $k(k-1)/2$ double integrals, $\mathrm{P}''(\theta_i < \theta_j)$, $1 \leq i < j \leq k$.

(ii) For each ranking R, determine the minimal chain from R to I according to the algorithm described above, and sum the differences $L_2(Q) - L_2(Q') = -\mathrm{P}''(\theta_i < \theta_j) + \mathrm{P}''(\theta_j < \theta_i)$ corresponding to each interchange from Q to Q' in the chain, thus obtaining $L_2(R) - L_2(I)$.

(iii) Choose a ranking which minimizes $L_2(R) - L_2(I)$.

Example 2.1. The following counter-expample seems to indicate that this procedure cannot be further simplified in general. Three independent distributions of θ_1, θ_2, and θ_3 are defined such that $\mathrm{P}(\theta_1 < \theta_2) > \mathrm{P}(\theta_2 < \theta_1)$, and $\mathrm{P}(\theta_2 < \theta_3) > \mathrm{P}(\theta_3 < \theta_2)$, but $\mathrm{P}(\theta_1 < \theta_3) < \mathrm{P}(\theta_3 < \theta_1)$.

Let $\mathrm{P}(\theta_1 = -0.5) = 2/3$, $\mathrm{P}(\theta_1 = 1) = 1/3$; $\theta_2 = 0$ with probability 1; and $\mathrm{P}(\theta_3 = -1) = 1/3$, $\mathrm{P}(\theta_3 = 0.5) = 2/3$. Then $\mathrm{P}(\theta_1 < \theta_2) = 2/3$, $\mathrm{P}(\theta_2 < \theta_3) = 2/3$, and yet $\mathrm{P}(\theta_1 < \theta_3) = 4/9$. Thus the relation $\mathrm{P}(\theta_i < \theta_j) > 1/2$ is not always transitive. In this counter-example, roughly speaking, θ_1 is probably less than θ_2, θ_2 is probably less than θ_3, but θ_1 is probably greater than θ_3. Similar counter-examples can be constructed where θ_1, θ_2, and θ_3 have continuous distributions.

3. Loss functions depending on θ_i

The loss function $l_2(R, \boldsymbol{\theta})$ has been proposed as an appropriate loss function for many situations in which the loss depends on the degree of error in the chosen ranking. This section considers four types of

situations in which the losses are direct functions of the true parameter values θ_i, $i=1,\cdots,k$. For these situations, it will be somewhat more convenient to consider utility rather than loss. Let $u(R,\boldsymbol{\theta})$ denote a utility function, and $U(R)$ a posterior expected utility. In each situation, the ranking is for the specific purpose of later selecting one or more populations according to a given procedure. After the selection, utilities will be realized which depend only on the true values θ_i for the populations selected. Assume that for preassigned functions w_i, a utility $w_i(\theta_i)$ is realized whenever population Π_i is selected, $i=1,\cdots,k$.

Consider first a situation in which a single population is to be selected from the k ranked populations according to a probability distribution depending only on the ranks of the populations. Let p_i denote the imputed probability that the population assigned rank i will be selected, $i=1,\cdots,k$. Also assume that

$$(3.1) \qquad\qquad 0<p_1<\cdots<p_k<1 .$$

Then, the utility function and the posterior expected utility for the ranking $R=(i_1,\cdots,i_k)$ are,

$$(3.2) \qquad u_1(R,\boldsymbol{\theta})=w_{i_1}(\theta_{i_1})p_1+\cdots+w_{i_k}(\theta_{i_k})p_k ,$$

and

$$(3.3) \qquad U_1(R)=\mathrm{E}''\,(w_{i_1}(\theta_{i_1}))p_1+\cdots+\mathrm{E}''\,(w_{i_k}(\theta_{i_k}))p_k .$$

The ranking $R=(i_1,\cdots,i_k)$ which maximizes U_1 is specified by

$$(3.4) \qquad \mathrm{E}''\,(w_{i_1}(\theta_{i_1}))<\mathrm{E}''\,(w_{i_2}(\theta_{i_2}))<\cdots<\mathrm{E}''\,(w_{i_k}(\theta_{i_k})) ,$$

since the representation (3.3) may be considered as a weighted sum of the quantities $\mathrm{E}''\,(w_{i_j}(\theta_{i_j}))$ with weights p_j, $j=1,\cdots,k$.

To implement this procedure, one needs only to determine the k single integrals, $\mathrm{E}''\,(w_i(\theta_i))$, $i=1,\cdots,k$.

It should be emphasized that whereas the expected utility $U_1(R)$ depends on the imputed probabilities, p_i, the choice of an optimal ranking does not, so long as Assumption (3.1) is valid. Thus for the purpose of choosing a ranking, the decision-maker does not need to assign specific values to the selection probabilities, p_i.

Consider next a situation in which a specified number r, $1\leq r<k$, of the ranked populations are to be selected, again according to a probability distribution depending only on the ranks of the populations. Let $t=\{t(1)\cdots t(r)\}$ denote a subset containing r of the indices $1,\cdots,k$. Let q_t denote the probability that the r populations with ranks $t(1),\cdots,t(r)$ will be selected, and let p_i denote the marginal probability that the population ranked ith will be among those selected. Assume that these marginal probabilities satisfy (3.1).

Then the utility function and the posterior expected utility may be written as,

$$(3.5) \qquad u_2(R, \boldsymbol{\theta}) = \Sigma_t (w_{i_{t(1)}}(\theta_{i_{t(1)}}) + \cdots + w_{i_{t(r)}}(\theta_{i_{t(r)}})) q_t \,,$$

and

$$(3.6) \qquad U_2(R) = \Sigma_t \{ \mathrm{E}'' \, (w_{i_{t(1)}}(\theta_{i_{t(1)}})) + \cdots + \mathrm{E}'' \, (w_{i_{t(r)}}(\theta_{i_{t(r)}})) \} q_t$$

where $R = (i_1, \cdots, i_k)$, and the summation is over all subsets t of size r.

In order to simplify these expressions consider the functions $I_j(t) = 1$ if $j \in t$, $I_j(t) = 0$ if $j \notin t$. Then

$$(3.7) \qquad u_2(R, \boldsymbol{\theta}) = \Sigma_t \left(\sum_{j=1}^{k} I_j(t) w_{i_j}(\theta_{i_j}) \right) q_t$$

$$= w_{i_1}(\theta_{i_1}) \Sigma_t I_1(t) q_t + \cdots + w_{i_k}(\theta_{i_k}) \Sigma_t I_k(t) q_t$$

$$= w_{i_1}(\theta_{i_1}) p_1 + \cdots + w_{i_k}(\theta_{i_k}) p_k \,.$$

Consequently,

$$(3.8) \qquad U_2(R) = \mathrm{E}'' \, (w_{i_1}(\theta_{i_1})) p_1 + \cdots + \mathrm{E}'' \, (w_{i_k}(\theta_{i_k})) p_k \,.$$

Thus the problem of ranking in order to select a subset of r populations is reduced to the corresponding problem for selecting a single population, the optimal ranking being specified by (3.4).

For the third type of situation, assume that the highest ranked populations are to be selected for certain, but the number r of populations which will be selected is now indeterminate, with an imputed distribution q_i, $i = 1, \cdots, k$, where q_i will denote the probability that the i populations ranking highest are selected. If r is determinate, then the full information conveyed by the ranking is not utilized, and one has a selection problem rather than a ranking problem. In this sense a ranking can be considered as being more informative than a selection.

The utility function for this situation is

$$(3.9) \qquad u_2(R, \boldsymbol{\theta}) = w_{i_k}(\theta_{i_k}) q_1 + \{ w_{i_k}(\theta_{i_k}) + w_{i_{k-1}}(\theta_{i_{k-1}}) \} q_2 + \cdots$$

$$+ \{ w_{i_k}(\theta_{i_k}) + \cdots w_{i_1}(\theta_{i_1}) \} q_k$$

$$= w_{i_1}(\theta_{i_1}) q_k + \cdots + w_{i_k}(\theta_{i_k}) \{ q_k + \cdots + q_1 \} \,,$$

where $R = (i_1, \cdots, i_k)$. Thus the expected utility is,

$$(3.10) \qquad U_3(R) = \mathrm{E}'' \, (w_{i_1}(\theta_{i_1})) q_1 + \cdots + \mathrm{E}'' \, (w_{i_k}(\theta_{i_k})) \{ q_k + \cdots + q_1 \}$$

for the ranking $R = (i_1, \cdots, i_k)$. Once again the posterior expected utility is a weighted sum of the quantities $\mathrm{E}'' \, (w_{i_j}(\theta_{i_j}))$, $j = 1, \cdots, k$, with weights increasing as j increases. Hence the optimal ranking is again specified by (3.4).

Whereas in the three previous situations, the selection probabilities were based entirely on the ranks of the populations, and the higher ranked populations were more likely to be selected, neither of these properties will be true in general for the following situation.

Suppose that a single population (or none) will be selected from the reported ranking, $R=(i_1,\cdots,i_k)$. The population Π_{i_k} with highest rank will be considered first. It will 'accept' and hence be selected with a probability $p(i_k)$ which depends on the population Π_{i_k} itself rather than on its ranking. If Π_{i_k} is not selected, the population $\Pi_{i_{k-1}}$ having next highest rank will be considered. It will accept, and hence be selected, with probability $p(i_{k-1})$. In general, let $p(j)$, $j=1,\cdots,k$, denote the probability that population Π_j will accept if considered, $j=1,\cdots,k$. Assume that $0<p(j)<1$, and that the events of different populations accepting or not accepting are mutually independent. Let u_0 denote the utility if none of the k populations accept.

Under these assumptions, the utility function and the posterior expected utility for $R=(i_1,\cdots,i_k)$ may be expressed as follows:

$$(3.11) \quad u_4(R,\boldsymbol{\theta})=w_{i_k}(\theta_{i_k})p(i_k)+w_{i_{k-1}}(\theta_{i_{k-1}})p(i_{k-1})(1-p(i_k))+\cdots$$
$$+w_{i_1}(\theta_{i_1})p(i_1)\prod_{l=2}^{k}(1-p(i_l))+u_0\prod_{l=1}^{k}(1-p(i_l)),$$

and

$$(3.12) \quad U_4(R)=\mathrm{E}''\,(w_{i_k}(\theta_{i_k}))p(i_k)+\mathrm{E}''\,(w_{i_{k-1}}(\theta_{i_{k-1}}))p(i_{k-1})(1-p(i_k))+\cdots$$
$$+\mathrm{E}''\,(w_{i_1}(\theta_{i_1}))p(i_1)\prod_{l=2}^{k}(1-p(i_l))+u_0\prod_{l=1}^{k}(1-p(i_l)).$$

THEOREM 3.1. *For a ranking problem with utility function $u_4(R,\boldsymbol{\theta})$, the optimal ranking is specified by (3.4).*

PROOF. Let $R=(i_1,\cdots,i_k)$ denote a ranking which maximizes U_4. As in the discussion of $l_2(R,\boldsymbol{\theta})$, we consider rankings $R'=(i_1,\cdots,i_h,i_{h-1},\cdots,i_k)$ which differ from R by an interchange of i_{h-1} and i_h. The difference $U_4(R)-U_4(R')$ may be determined as follows, where the first equality is obtained from (3.12) by cancellation.

$(3.13) \quad U_4(R)-U_4(R')$
$$=\mathrm{E}''\,(w_{i_h}(\theta_{i_h}))p(i_h)\prod_{l>h}(1-p(i_l))+\mathrm{E}''\,(w_{i_{h-1}}(\theta_{i_{h-1}}))p(i_{h-1})$$
$$\cdot(1-p(i_h))\prod_{l>h}(1-p(i_l))-\mathrm{E}''\,(w_{i_{h-1}}(\theta_{i_{h-1}}))p(i_{h-1})$$
$$\cdot\prod_{l>h}(1-p(i_l))-\mathrm{E}''\,(w_{i_h}(\theta_{i_h}))p(i_h)(1-p(i_{h-1}))\prod_{l>h}(1-p(i_l))$$
$$=p(i_{h-1})p(i_h)\prod_{l>h}(1-p(i_l))\{\mathrm{E}''\,(w_{i_h}(\theta_{i_h}))-\mathrm{E}''\,(w_{i_{h-1}}(\theta_{i_{h-1}}))\}.$$

Since R maximizes U_4, in particular $U_4(R)-U_4(R')\geqq 0$ for each $h=2$,

$\cdots, k$. Thus $\mathrm{E}''(w_{i_{h-1}}(\theta_{i_{h-1}})) \leqq \mathrm{E}''(w_{i_h}(\theta_{i_h}))$, $h=2, \cdots, k$, i.e. R must equal the ranking specified by (3.4).

4. Comparison of ranking procedures

This section is concerned with those circumstances in which the decision-maker is assured that the ranking procedures of Sections 2 and 3 will lead to the same optimal ranking. This will be the case, broadly speaking, whenever either (i) the marginal posterior densities $h_i(\theta)$, $i=1, \cdots, k$, have the same 'content' in the sense that for each population the sample size plus the corresponding prior parameter are equal, or (ii) the densities $h_i(\theta)$, $i=1, \cdots, k$, are symmetric.

Consider first the case in which the distributions of the populations $\Pi_1, \cdots, \Pi_k$ belong to a family of distributions which depend on a single parameter θ, and admit a sufficient statistic, $T(\boldsymbol{x})$, $\boldsymbol{x}=(x_1, \cdots, x_n)$. Such a family is said to have the monotone likelihood ratio property, MLR, in $T(\boldsymbol{x})$, if for any $\theta_1 < \theta_2$, the ratio of likelihoods $f(\boldsymbol{x}|\theta_2)/f(\boldsymbol{x}|\theta_1)$ is a strictly increasing function of $T(\boldsymbol{x})$, (see, for instance, Lehmann [10], p. 68).

DEFINITION. Consider a family $\mathscr{H}$ of posterior densities $h(\theta)$ whose members are indexed by two parameters, η'' and τ''. Let $\mathscr{H}(\eta_0'')$ denote the subfamily of those densities such that $\eta''=\eta_0''$. The family $\mathscr{H}$ will be said to have the property of monotone posterior ratio (MPR) if for any two densities $h_1(\theta)$, $h_2(\theta)$ in the same subfamily $\mathscr{H}(\eta_0'')$ with parameter values τ_1'', τ_2'' respectively; $\tau_1'' < \tau_2''$ implies that the ratio $h_2(\theta)/h_1(\theta)$ is an increasing function of θ.

This notion of MPR can be used to effect orderings of the posterior densities $h_i(\theta)$, much as the notion of MLR is used to effect orderings among likelihood functions.

A_1: Assume that the likelihoods $f(x|\theta_i)$, $i=1, \cdots, k$, associated with $\Pi_1, \cdots, \Pi_k$ belong to the exponential family,

$$(4.1) \qquad f(x|\theta_i) = c(\theta_i) \exp[u(x)q(\theta_i)]k(x)$$

where $q(\theta_i)$ is a strictly increasing function.

A_2: Assume that for each population Π_i, $i=1, \cdots, k$, an independent prior density $g_i(\theta)$ has been assigned from the natural conjugate family, i.e. for some constants η_i and τ_i,

$$(4.2) \qquad g_i(\theta) = [c(\theta)]^{\eta_i} \exp[\tau_i q(\theta)]k$$

and that an independent sample $\boldsymbol{x}_i=(x_{i1}, \cdots, x_{in_i})'$ of size n_i has then

been taken. (This leads to posterior densities of the form

(4.3) $$h_i(\theta) = [c(\theta)]^{\eta_i + n_i} \exp [(\tau_i + T(\boldsymbol{x}_i))q(\theta)]k(\boldsymbol{x}_i)$$

where n_i and $T(\boldsymbol{x}_i) = \sum_{j=1}^{n_i} u(x_{ij})$ are sufficient statistics.) Let $\eta_i'' = \eta_i + n_i$, and $\tau_i'' = \tau_i + T(\boldsymbol{x}_i)$, $i = 1, \cdots, k$.

A_3: Assume that the posterior densities $h_i(\theta)$ belong to the same sub-family $\mathscr{H}(\eta_0'')$, $i = 1, \cdots, k$.

THEOREM 4.1. *If assumptions* A_1–A_3 *are satisfied, then*
(i) *The densities* $h_i(\theta)$ *have the MPR property with respect to* $\tau_i'' = \tau_i + T(\boldsymbol{x}_i)$.
(ii) *For any indices* $1 \leq i$, $j \leq k$, $\tau_i + T(\boldsymbol{x}_i) < \tau_j + T(\boldsymbol{x}_j)$ *impleis that*
   (a) $P''(\theta_i < \theta_j) > P''(\theta_j < \theta_i)$.
   (b) *For any monotone increasing function* $w(\theta)$, $E''(w(\theta_i)) < E''(w(\theta_j))$.

PROOF. Assume that for given indices i, j, $\tau_i'' < \tau_j''$. Since by A_3, $\eta_i'' = \eta_j''$, the ratio of the posterior densities may be written as

(4.4) $$\frac{h_j(\theta)}{h_i(\theta)} = \exp [(\tau_j + T(\boldsymbol{x}_j) - \tau_i - T(\boldsymbol{x}_i))q(\theta)] \frac{k(\boldsymbol{x}_j)}{k(\boldsymbol{x}_i)}.$$

The exponential factor is a strictly increasing function of θ since $q(\theta)$ is strictly increasing, and the coefficient of $q(\theta)$ is assumed to be positive. Thus the posterior densities have the MPR property as stated in (i). It follows that for increasing values of θ, $\theta_i < \theta_j$,

(4.5) $$h_j(\theta_i)h_i(\theta_j) < h_i(\theta_i)h_j(\theta_j).$$

Part (ii) will be shown to be a consequence of (i). Consider

(4.6) $$P''(\theta_i < \theta_j) - P''(\theta_j < \theta_i)$$
$$= \int \cdots \int_{\theta_i < \theta_j} h_i(\theta_i)h_j(\theta_j) \prod_{l \neq i,j} h_l(\theta_l)d\theta_1 \cdots d\theta_k$$
$$- \int \cdots \int_{\theta_j < \theta_i} h_i(\theta_i)h_j(\theta_j) \prod_{l \neq i,j} h_l(\theta_l)d\theta_1 \cdots d\theta_k$$

where $\theta_1, \cdots, \theta_k$ denote dummy variables of integration.
By interchanging θ_i and θ_j in the second integral, we have,

(4.7) $$P''(\theta_i < \theta_j) - P''(\theta_j < \theta_i)$$
$$= \int \cdots \int_{\theta_i < \theta_j} (h_i(\theta_i)h_j(\theta_j) - h_i(\theta_j)h_j(\theta_i)) \prod_{l \neq i,j} h_l(\theta_l)d\theta_1 \cdots d\theta_k.$$

By (4.5), the difference in parentheses is positive, so $\mathrm{P}''(\theta_i < \theta_j) > \mathrm{P}''(\theta_j < \theta_i)$.

Consider next the difference,

$$(4.8) \qquad \mathrm{E}''(w(\theta_j)) - \mathrm{E}''(w(\theta_i)) = \int_{\Omega} w(\theta)(h_j(\theta) - h_i(\theta)) d\theta .$$

The integral $\int_{\Omega}(h_j(\theta) - h_i(\theta)) d\theta$ equals zero, so unless $h_j \equiv h_i$, there exists points θ at which $h_j(\theta) < h_i(\theta)$ and other points θ at which $h_j(\theta) > h_i(\theta)$.

The ratio $h_j(\theta)/h_i(\theta)$ is strictly increasing, so there must exists a unique point $\theta_0 \in \Omega$ at which this ratio equals unity. Moreover $h_j(\theta) < h_i(\theta)$ for all $\theta < \theta_0$ and $h_j(\theta) > h_i(\theta)$ for all $\theta > \theta_0$. This implies that

$$(4.9) \quad \mathrm{E}''(w(\theta_j)) - \mathrm{E}''(w(\theta_i))$$
$$= \int_{\theta < \theta_0} w(\theta)(h_j(\theta) - h_i(\theta)) d\theta + \int_{\theta > \theta_0} w(\theta)(h_j(\theta) - h_i(\theta)) d\theta$$
$$> w(\theta_0) \int_{\theta < \theta_0} (h_j(\theta) - h_i(\theta)) d\theta + w(\theta_0) \int_{\theta > \theta_0} (h_j(\theta) - h_i(\theta)) d\theta = 0 .$$

Theorem 4.1 implies that whenever the posterior densities $h_i(\theta_i)$ satisfy the specified assumptions, and the utilities $w(\theta_i)$ are increasing in θ_i, $i = 1, \cdots, k$, then for the procedures of Sections 2–3, the optimal ranking $R = (i_1, \cdots, i_k)$ is specified by

$$(4.10) \qquad \tau_{i_1} + T(x_{i_1}) < \cdots < \tau_{i_k} + T(x_{i_k}) .$$

The assumption that all $\eta_i + n_i$ are equal, i.e. that for each population the sample size plus the analogous prior parameter is the same, will be satisfied, for example, when the sample sizes are equal and the prior densities are improper and vague.

Some results similar to Theorem 4.1, but concerning the distribution of the sampling random variables, have been obtained by Seal [13] for the problem of ranking type III populations

$$f(x | a_i, k) = a_i^k \Gamma(k)^{-1} x_i^{k-1} \exp(-a_i x_i) , \qquad 0 < x < \infty ,$$

where the equality of the known parameters k plays the role of our assumption that $h_i(\theta) \in \mathcal{H}(\eta_0'')$, $i = 1, \cdots, k$.

Example 4.1. Suppose the Π_i represent Bernoulli processes with parameters p_i, $0 < p_i < 1$, and data x_{ij}, $j = 1, \cdots, n_i$, having values 0 or 1. If the prior densities $g_i(p)$ are chosen from the beta family,

$$(4.11) \qquad g(p) = [B(\tau, \eta - \tau)]^{-1} p^{\tau-1}(1-p)^{\eta-\tau-1} , \qquad 0 < \tau < \eta ,$$

where $B(a, b) = \Gamma(a+b)/\Gamma(a)\Gamma(b)$, then the posterior densities will be

$$(4.12) \quad h_i(p) = [B(\tau_i'', \eta_i'' - \tau_i'')]^{-1}[p(1-p)]^{-1}(1-p)^{\eta_i''} \exp[\tau_i'' \log\{p/(1-p)\}]$$

where $\eta_i'' = \eta_i + n_i$, $\tau_i'' = \tau_i + T(\boldsymbol{x}_i)$, and $T(\boldsymbol{x}_i) = \sum_{j=1}^{n_i} x_{ij}$, $i = 1, \cdots, k$. The function $q(p) = \log \{p(1-p)\}$ is strictly increasing in p, $0 < p < 1$. Thus if the densities $h_i(p)$ are in the same subfamily $\mathcal{H}(\eta_0'')$, the Bernoulli processes should be ranked according to $\tau_i'' = \tau_i + \sum_{j=1}^{n_i} x_{ij}$, $i = 1, \cdots, k$.

We consider next several situations in which the distributions associated with the populations Π_i do not satisfy the assumptions of Theorem 4.1; however it is still possible to provide sufficient conditions under which the ranking procedures become identical.

Example 4.2. The likelihoods $f(x|\theta, \boldsymbol{\omega})$ may depend on more than a single parameter θ. Suppose, for example, that the populations Π_i represent normal processes with both μ and σ^2 unknown, which are to be ranked according to σ^2. The decision-maker wishes to assign higher rank to those populations with smaller variances, and he believes that the precision $\rho_i = 1/\sigma_i^2$ is a reasonable measure of his utility if the population Π_i should be selected. He assigns independent normal-gamma prior distributions (Raiffa and Schlaifer [11], pp. 298–303) to the parameters μ_i, ρ_i, $i = 1, \cdots, k$, i.e. in a notation suitable for the present purpose,

$$(4.13) \qquad g(\mu, \rho \,|\, m, l, \tau, \eta) = K\rho^{1/2} e^{-l\rho(\mu-m)^2/2} \rho^{(n/2)-1} e^{-\tau\rho/2}$$

where $l, \tau, \eta > 0$. Independent data x_{ij}, $j = 1, \cdots, n_i$ are taken from Π_i leading to the statistics $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i$ and $S_i^2 = \sum_{i=1}^{n_i} (x_{ij} - \bar{x}_i)^2$. Then the marginal posterior density on ρ_i belongs to the gamma family,

$$(4.14) \qquad h_i(\rho) = K_i'' \rho^{(\eta_i''/2)-1} e^{-\tau_i''\rho/2}$$

where $\eta_i'' = \eta_i + n_i$, and $\tau_i'' = \tau_i + S_i^2 + (n_i^{-1} + l_i^{-1})^{-1}(\bar{x}_i - m_i)^2$.

Therefore, it may be directly verified that whenever the posterior densities $h_i(\rho)$, $i = 1, \cdots, k$, are in the same subfamily $\mathcal{H}(\eta_0'')$, they have the MPR property with respect to τ_i'', $i = 1, \cdots, k$, and hence consequence (ii) of Theorem 4.1 is valid.

Example 4.3. Suppose that the likelihoods associated with the k populations are uniform distributions $f_i(x|\theta_i) = \theta_i^{-1}$ for $0 < x < \theta_i$. Let the assigned independent prior densities be of the form

$$(4.15) \qquad g(\theta) = \eta \tau^\eta \cdot (1/\theta)^{\eta+1} H(\theta - \tau)$$

for some parameters τ, $\eta > 0$, where H denotes the Heaviside function as in (2.7). Then the posterior densities are given by

$$(4.16) \qquad h_i(\theta) = \eta_i'' \cdot (\tau_i'')^{\eta_i''} (1/\theta)^{\eta_i''} H(\theta - \tau_i'')$$

where $\tau_i'' = \max_{1 \le j \le n_i} \{\tau_i, x_{ij}\}$, $\eta_i'' = \eta_i + n_i$. Thus, if the desities $h_i(\theta)$ are in

the same class $\mathcal{H}(\eta_0'')$, and $\tau_i'' < \tau_j''$, then

$$h_j(\theta)/h_i(\theta) = (\tau_j''/\tau_i'')^{\eta_0''} H(\theta - \tau_j'') \ ,$$

which is an increasing function of θ. Therefore the conclusions of Theorem 4.1 remain valid for uniform distributions.

Example 4.4. The decision-maker may wish to rank according to the predicted performances of the k populations under specified conditions, rather than according to a parameter θ. For example, if the populations represent Bernoulli processes as in Example 4.1, the decision-maker may wish to rank according to the predicted distribution of successes in future Bernoulli processes of specified length N for each population. Assuming that the independent posterior densities $h_i(p)$ are given by (4.12), the probability distribution $h_i(r)$, $r = 0, 1, \cdots, N$ for the predicted number r of successes is a beta-binomial (or hyper-binomial) mass function (Raiffa and Schlaifer [11], p. 237).

$$(4.17) \qquad h_i(r) = \binom{N}{r} B(\tau_i'', \eta_0'' - \tau_i'')^{-1} B(r + \tau_i'', N + \eta_0'' - r - \tau_i'')$$

where the parameters τ_i'', η_0 have the meanings given in Example 4.1.

Consider two indices i, j such that $\tau_i'' < \tau_j''$. The ratio $W_{ji}(r) = h_j(r)/h_i(r)$ may be reduced in general to the expression,

$$(4.18) \qquad W_{ji}(r) = \frac{B(\tau_i'', \eta_0'' - \tau_i'')}{B(\tau_j'', \eta_0'' - \tau_j'')} \cdot \frac{\Gamma(r + \tau_j'')\Gamma(N + \eta_0'' - r - \tau_j'')}{\Gamma(r + \tau_i'')\Gamma(N + \eta_0'' - r - \tau_i'')} \ .$$

Thus for $r = 0, 1, \cdots, N-1$, one may compare,

$$(4.19) \qquad W_{ji}(r+1)/W_{ji}(r) = \frac{(r + \tau_j'')}{(r + \tau_i'')} \cdot \frac{(N + \eta_0'' - r - \tau_i'' - 1)}{(N + \eta_0'' - r - \tau_j'' - 1)} \ .$$

Each quotient in (4.19) is greater than 1 since $\tau_j'' > \tau_i''$. Thus $W_{ji}(r) = h_j(r)/h_i(r)$ is an increasing function of r, i.e. $h_i(r)$ has the MPR property with respect to τ_i''. Thus consequence (ii) of Theorem 4.1 may be deduced by arguments which differ from those given in that summations replace integrations.

The marginal posterior densities $h_i(\mu_i)$ for the means μ_i in Example 4.2 are Student distributions. It may be verified that, assuming η_i'' equal, $i = 1, \cdots, k$, the densities $h_i(\mu_i)$ do not in general have the MPR property. Nevertheless, as a consequence of the following general result, Conclusion (ii) of Theorem 4.1 remains valid for such densities, at least in the important special case, $w(\theta_i) = \alpha\theta_i + \beta$, $\alpha > 0$.

THEOREM 4.2. *Assume that each posterior density* $h_i(\theta)$, $-\infty < \theta < +\infty$, *is symmetric about some point,* $i = 1, \cdots, k$. *Let* τ_i'' *denote the point*

of symmetry of $h_i(\theta)$, $i=1,\cdots,k$.   *Then for any indices* $1 \le i$, $j \le n$, τ_i''
$< \tau_j''$ *implies that*,
   (a) $\mathrm{P}''\,(\theta_i < \theta_j) > \mathrm{P}''\,(\theta_j < \theta_i)$.
   (b) $\mathrm{E}''\,(w(\theta_i)) < \mathrm{E}''\,(w(\theta_j))$ *for any linear function*, $w(\theta) = \alpha\theta + \beta$, $\alpha > 0$.

PROOF.   By the assumption of symmetry, $\mathrm{E}''\,(\theta_i) = \tau_i''$, $i=1,\cdots,k$.
Thus (b) is immediate.   By the same assumption the median m_i of $h(\theta_i)$
also equals τ_i'', $i=1,\cdots,k$.   Now

$$\mathrm{P}''\,(\theta_i < \theta_j) = \mathrm{P}''\,((\theta_i - m_i) - (\theta_j - m_j) < -(m_i - m_j)) \ .$$

But the distribution of $(\theta_i - m_i) - (\theta_j - m_j)$ is symmetric about the origin,
hence has median zero.   Since $-(m_i - m_j) = \tau_j'' - \tau_i'' > 0$, it follows that
$\mathrm{P}''\,(\theta_i < \theta_j) > 1/2$ which implies (a).

Note that Theorem 4.2 does not assume that the densities $h_i(\theta)$ be-
long to the same class $\mathscr{H}(\eta_0'')$.   Indeed, in sampling from normal pop-
ulations, the conclusions (a), (b) remain valid even if the variances are
known for some of the populations and unknown for others.

5.   Subset selection procedures

The subset selection problem consists of choosing a subset S from
the set of k populations so as to include in S those populations which
are believed to have large true parameter values θ_i.   Assume that the
purpose of selecting the subset S is to restrict further sampling to those
populations which are most promising.   The populations Π_i in S are to
receive further exhaustive examination at a cost c_i for the population
Π_i, $i=1,\cdots,k$.   Then a single population will be selected from S, and
a utility realized depending on this decision.

Consider first those situations in which the utility gained after a
single population Π_i has been chosen depends on whether an error has
been committed, more specifically, whether or not the selected popula-
tion Π_i is the 'best' population in the sense that it has the largest
value θ_i among all k populations.   Suppose that the utility gained is d
if S contains the best population, zero otherwise.   As a notational aid,
let $\delta(S, \boldsymbol{\theta})$ denote that function which is one in the first case, zero in
the second.   Since the decision-maker will obtain very precise informa-
tion concerning the values θ_j for $\Pi_j \in S$, it may be assumed as an ideal-
ization that the population chosen will be optimal provided S contains
the best population.   The utility function is then

(5.1) $u_1(S, \boldsymbol{\theta}) = -\sum_{i \in s} c_i + \delta(S, \boldsymbol{\theta}) \cdot d$

where s denotes the subset of indices i such that $\Pi_i \in S$.

Let A_i denote the event that populatin Π_i has the largest true value θ_i among all k populations, $i = 1, \cdots, k$. The event, $\cup_{i \in s} A_i$, that some Π_i in S has the largest θ_i will be represented by the standard notation CS, i.e. 'correct selection', introduced by Bechhofer [3]. Then the posterior expected utility may be written as

$$(5.2) \qquad U_1(S) = -\sum_{i \in s} c_i + d \cdot \mathrm{P}''(CS) = \sum_{i \in s} [-c_i + d \cdot \mathrm{P}''(A_i)] \ .$$

It is clear from the second expression for $U_1(S)$ that the procedure for determining the optimal subset S is to include in S those Π_i for which $-c_i + d \cdot \mathrm{P}''(A_i) > 0$, i.e.

$$(5.3) \qquad\qquad \mathrm{P}''(A_i)/c_i > (1/d) \ .$$

Intuitively, the decision-maker selects those populations with the greatest value in terms of the probabilities $\mathrm{P}''(A_i)$, scaled according to the costs c_i.

This procedure is implemented by calculating, for each $i = 1, \cdots, k$, the k-dimensional integral $\mathrm{P}''(A_i)$ with sufficient accuracy so that its value can be compared with c_i/d.

If values have been imputed for the losses due to costs, how can the decision-maker assign an appropriate value to the utility d of selecting the best population? One method is to choose a value for d by comparing, for a fixed index i, the following three lotteries, $\Lambda_1, \Lambda_2, \Lambda_3$. Let Λ_1 consist of winning 0 for certain. Let Λ_2 consist of paying a proportion $a_i c_i$ of the cost c_i as an entrance fee to win a utility d for certain, i.e. to be given perfect information concerning θ. Let Λ_3 consist of paying c_i as an entrance fee to win utility d with probability p_i, or utility 0 with probability $1 - p_i$. Suppose that for the index i under consideration, one can choose a constant a_i so that one is indifferent between lotteries Λ_1 and Λ_2. Then d should be assigned the value $a_i c_i$. If one can choose a probability p_i so that one is indifferent between lotteries Λ_1 and Λ_3, then d should be assigned value c_i/p_i.

It may be necessary to reassign values to a_i and p_i in order to satisfy the consistency requirements of utility theory. If the costs c_i are different, and values are assigned to d on the basis of more than one c_i, then even more retrospective comparisons may be needed to achieve consistency.

Consider next a situation in which constraints have been imposed on either the total sampling costs or on the probability of error. Suppose, for example, that one wishes to maximize the expected utility due to selecting the best population, given that the total sampling costs cannot exceed C. Then (5.2) leads to the optimization problem,

$$(5.4) \qquad\qquad \text{maximize } \mathrm{P}''(CS) \text{ subject to } \sum_{i \in s} c_i \leq C \ .$$

Suppose next that an upper bound α has been placed on the probability of error, and the decision-maker wishes to minimize the total sampling cost. This leads to the optimization problem,

$$(5.5) \qquad \text{minimize} \sum_{i \in s} c_i \text{ subject to } P''(CS) \geqq 1 - \alpha .$$

As a second utility function to be considered, suppose that the utility depends only on the single population Π_i finally chosen, rather than on its relation to the other populations as with $u_1(S, \boldsymbol{\theta})$. Suppose that this utility is a function $w_i(\theta_i)$ of the true value θ_i of the parameter associated with Π_i. Then the utility function and the posterior expected utility are,

$$(5.6) \qquad u_2(S, \boldsymbol{\theta}) = -\sum_{i \in s} c_i + \max_{i \in s} w_i(\theta_i)$$

$$(5.7) \qquad U_2(S) = -\sum_{i \in s} c_i + E'' (\max_{i \in s} w_i(\theta_i)) .$$

In general, the subset S maximizing (5.7) can be determined by calculating the $2^k - 1$ integrals $E'' (\max_{i \in s} w_i(\theta_i))$.

If the cost of further sampling of S is constrained not to exceed C, then the decision-maker selects that subset S so as to,

$$(5.8) \qquad \text{maximize } E'' (\max_{i \in s} w_i(\theta_i)) \text{ subject to } \sum_{i \in s} c_i \leqq C .$$

If the expected performance of the single population chosen must be at least a predetermined value M, then the decision-maker selects S so as to,

$$(5.9) \qquad \text{minimize} \sum_{i \in s} c_i \text{ subject to } E'' (\max_{i \in s} w_i(\theta_i)) \geqq M .$$

Consider next those situations in which the decision-maker needs to select a subset S from a set of k populations which will be reported as the more desirable in some sense, but where the horizon of the formal analysis cannot be effectively placed beyond this subset selection. Perhaps it is not clear whether the selection decision will lead to further sampling, immediate selection of one of the populations in S, utilization of all the selected populations, rejection of all the selected populations, or possibly some other goal. The first question, of course, is whether some loss function or utility function can be proposed which appears to be fairly appropriate, at least over a wide range of subset selection problems. As the reader may anticipate, we are unable to bring forth such a procedure. The primary difficulty, which can assume various forms, seems to be that it is not clear how to compare subsets of different sizes. If the purpose of the subset selection is non-

specific, and one wishes to compare a subset S with a subset $S'=S \cup \{\Pi_i\}$ containing an additional population, how should one scale the 'costs' of including Π_i against the 'advantages' of this inclusion?

Suppose that the problem is restricted to selecting a subset S, given that its size has been predetermined. Is it possible to define the 'quality' of a subset of populations so that subsets of equal size can be compared?

Assume that a single population Π_i will be chosen from S and a utility realized which is a function $w_i(\theta_i)$ of the parameter θ_i only, but that the decision-maker has no idea which population in S will be chosen. This situation leads one to the familiar choice between minimax procedures, Bayes' procedures, or Laplace's procedure (i.e. assigning equal probability to each state of nature). The reader who has a definite opinion as to which procedure is appropriate for the problem should observe that the original restricted problem of comparing subsets of equal size is a more general problem.

In Section 2, where some measure of the 'quality' of a subset of size r was required, we used Laplace's procedure, i.e. a utility function equal to the average of $w_i(\theta_i)$, $\Pi_i \in S$, or equivalently,

$$(5.10) \qquad\qquad u(S, \boldsymbol{\theta})= \sum_{i \in s} w_i(\theta_i) \ .$$

This choice was motivated in large part for reasons of mathematical tractability. The above considerations seem to indicate that, whenever possible, a decision theoretic approach to a subset selection problem should be based on a precise understanding of the purpose of that particular selection.

University of Kentucky

References

[1] Bahadur, R. R. (1950). On a problem in the theory of k populations, *Ann. Math. Statist.*, 21, 362-375.
[2] Bahadur, R. R. and Goodman, L. A. (1952). Impartial decision rules and sufficient statistics, *Ann. Math. Statist.*, 23, 553-562.
[3] Bechhofer, R. E. (1954). A single-sample multiple-decision procedure for ranking means of normal populations with known variances, *Ann. Math. Statist.*, 25, 16-39.
[4] Bland, R. P. and Bratcher, T. L. (1968). A Bayesian approach to the problem of ranking binomial probabilities, *SIAM J. Appl. Math.*, 16, 843-850.
[5] Bratcher, T. L. (1970). A Bayesian treatment of a multiple comparison problem for binomial probabilities, An unpublished manuscript.
[6] Deeley, J. J. and Gupta, S. S. (1968). On the properties of subset selection procedures, *Sankhyā*, Ser. A, 30, 37-50.
[7] Fairweather, W. R. (1968). Some extensions of Somerville's procedure for ranking means of normal populations, *Biometrika*, 55, 411-418.
[8] Gupta, S. S. (1956). *On a Decision Rule for a Problem in Ranking Means*, Mimeograph

Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, N.C.

[9]   Gupta, S. S. and Sobel, M. (1960).   Selecting a subset containing the best of several binomial populations, *Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling*, Stanford: Stanford University Press, 224-248.

[10] Lehmann, E. L. (1959).   *Testing Statistical Hypotheses*, New York: Wiley.

[11] Raiffa, H. and Schlaifer, R. (1961).   *Applied Statistical Decision Theory*, Boston: Harvard University Graduate School of Business Administration.

[12] Seal, K. C. (1955).   On a class of decision procedures for ranking means of normal populations, *Ann. Math. Statist.*, **26**, 387-398.

[13] Seal, K. C. (1958).   On ranking parameters of scale in Type III populations, *J. Amer. Statist. Ass.*, **53**, 164-175.

[14] Somerville, P. N. (1954).   Some problems of optimal sampling, *Biometrika*, **41**, 420-429.