The Annals of Statistics
1973, Vol. 1, No. 4, 750–755

ON A MULTIPLE DECISION RULE[1]

By Khursheed Alam

Clemson University

Let $X = (X_1, \cdots, X_k)$ be a random vector whose distribution depends on a parameter vector $\theta = (\theta_1, \cdots, \theta_k)$. A standard procedure ϕ^* is considered for selecting a set of $m < k$ coordinate values corresponding to the m largest components of θ. ϕ^* is given as follows: Select the m coordinates corresponding to the m largest components of x, the observed value of X. Break ties, if any, with randomization. Some optimal properties of ϕ^* are known, given that the loss function and the distribution of X have certain invariance and monotonicity properties. It is shown in this paper that ϕ^* is a Bayes decision rule if X is "stochastically increasing" in θ.

1. Introduction. Let $X = (X_1, \cdots, X_k)$ be a random vector whose distribution depends on a vector parameter θ. We consider a standard procedure ϕ^* for selecting a set of $m < k$ coordinate values corresponding to the m largest components of the unknown parameter θ. Let $\gamma = (\gamma_1, \gamma_2)$ denote a partition of the set $\{1, \cdots, k\}$ into two disjoint subsets γ_1 and γ_2, consisting of m and $k - m$ elements, respectively. Let Γ denote the set of all such partitions, and let x denote the observed value of X. A general decision rule for the given problem is a function $\phi = \{\phi_\gamma(x) : \gamma \in \Gamma\}$ where $0 \leqq \phi_\gamma(x) \leqq 1$ and $\sum_{\gamma \in \Gamma} (x) = 1$. $\phi_\gamma(x)$ is the probability, given x, of selecting the set of coordinate values which are the elements of γ_1.

$\phi^* = \{\phi_\pi^*(x) : \gamma \in \Gamma\}$ is given as follows: Let

$$A_\gamma = \{x : x_i \geqq x_j \quad \text{for all} \quad i \in \gamma_1 \quad \text{and} \quad j \in \gamma_2\},$$
$$C(x) = \{\gamma : \gamma \in \Gamma, x \in A_\gamma\}$$

and let $n(x)$ denote the number of elements in the set $C(x)$. Then

$$\phi_\gamma^*(x) = 1/n(x) \quad \text{if} \quad \gamma \in C(x), \qquad (1.1)$$
$$= 0 \quad \text{otherwise.}$$

For the problem of selecting the best one of several populations, Bahadur [1] and Bahadur and Goodman [2] have shown that for certain families of distributions, the natural selection procedure ϕ^* (for $m = 1$) uniformly minimizes the risk among all symmetric procedures for a general class of loss functions. Lehmann [6] has given another proof of this result, and has indicated other properties of ϕ^*. Eaton [3] has extended the results for a more general problem of ranking and a family of distributions of X.

In this paper we consider a family of distributions with the property SIP, defined below. Let a partial ordering $\prec$ be defined as follows: $x \prec x'$ iff $x_i \leqq x_i'$,

Received April 1972; revised August 1972.

[1] The author's work was supported in part by the office of Naval Research under contract N00014-71-0339-0002 Task NR942-271.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*. www.jstor.org

$i = 1, \cdots, k$. Similarly, $\theta \prec \theta'$ iff $\theta_i \leqq \theta_i'$, $i = 1, \cdots, k$. A measurable subset of the sample space is called monotone non-decreasing (with respect to $\prec$) if $x \in S$ and $x \prec x'$ implies $x' \in S$. Let $P_\theta(S)$ denote the probability measure of S under the conditional distribution of X, given θ. The distribution is said to have stochastically increasing property (SIP) in θ if $P_\theta(S) \leqq P_{\theta'}(S)$ for every monotone non-decreasing set S and $\theta \prec \theta'$. SIP was first introduced by Lehmann [5].

A characterization of SIP is given by the following lemma (for proof see Lehmann [5], page 400). A function $\psi(x)$ is said to be non-decreasing (with respect to $\prec$) if $\psi(x) \leqq \psi(x')$ for $x \prec x'$. Let E_θ denote expectation with respect to the distribution P_θ.

LEMMA 1.1. *A family of distributions P_θ has* SIP *in θ if $E_\theta\psi(x) \leqq E_{\theta'}\psi(x)$ for all non-decreasing integrable function $\psi(x)$, and $\theta \prec \theta'$.*

From Lemma 1.1 it follows that if P_θ has SIP in θ, and if $\psi(x)$ is non-decreasing in x_j then $E_\theta\psi(X)$ is non-decreasing in θ_i and non-increasing in θ_j.

The results of Eaton [3] are applicable to a family of distributions with densities which have a certain property called M. If θ is a location parameter of the distribution of X then the distribution has SIP but not Property M, in general. On the other hand, the multinomial distribution has both SIP and Property M.

In Section 2 we show that ϕ^* is a Bayes decision rule with respect to a given prior distribution, when the posterior distribution has SIP in x, and the loss function has certain invariance and monotonicity properties. Some applications of this result are given in Section 3.

2. Properties of ϕ^*. The Bayes character of ϕ^* is derived basically from certain properties of the distribution of X and the loss function. First we give same preliminary results and describe the basic assumptions.

Let g denote a permutation of the components of a k-component vector, and let G denote the group of all such permutations. A set $A \subset R_k$ (k-dimensional Euclidean space) is called symmetric if $gA = A$ for all $g \in G$ where gA denotes the image of A under g. We assume that the sample space χ and the parameter space Ω are symmetric Borel subsets of R_k. A distribution Π is called symmetric if $\Pi(A) = \Pi(gA)$ for all measurable set A and $g \in G$, where $\Pi(A)$ denotes the probability measure of A under Π. A family of distributions P_θ is called invariant with respect to G if $P_\theta(A) = P_{g\theta}(gA)$ for all measurable set A and $g \in G$.

Let Π be a given symmetric prior distribution on Ω, and let P_x denote the family of posterior distributions of θ, corresponding to P_θ. Theorem 2.1, below, gives the dual relation between the invariance properties of P_θ and P_x.

THEOREM 2.1. *If P_θ is invariant with respect to G then P_x is invariant with respect to G.*

PROOF. Let $P(A)$ denote the probability measure of a measurable set $A \subset \chi$

under the marginal distribution of X. For any $g \in G$, we have

$$P(A) = \int P_\theta(A)\, d\Pi(\theta) = \int P_{g\theta}(gA)\, d\Pi(\theta) = \int P_{g\theta}(gA)\, d\Pi(g\theta) = \int P_\theta(gA)\, d\Pi(\theta) = P(gA)\,. \tag{2.1}$$

On the right-hand side of (2.1), the second line follows from the invariance property of P_θ, and the third line follows from the symmetry of the a priori distribution Π. From (2.1) we have for any measurable set $B \subset \Omega$

$$P\{\theta \in B \mid X \in A\} = P\{X \in A \mid \theta \in B\}\Pi(B)/P(A) = P\{X \in gA \mid \theta \in gB\}\Pi(gB)/P(gA) = P\{\theta \in gB \mid X \in gA\}\,. \tag{2.2}$$

The theorem follows from (2.2). □

For any partition $\gamma \in \Gamma$, let $L_\gamma(\theta) \geqq 0$ denote the loss in the presence of θ, due to selecting the coordinates which are the elements of γ_1. We assume that the loss function has the property defined below.

Property L. For all $\gamma \in \Gamma$, $g \in G$ and $\theta \in \Omega$

(i) $L_\gamma(\theta) = L_{g\gamma}(g\theta)$ and
(ii) $L_\gamma(\theta)$ is non-increasing (non-decreasing) in θ_i for $i \in \gamma_1(\gamma_2)$.

It follows from Property L that

$$L_\gamma(\theta) \leqq L_{g\gamma}(\theta) \tag{2.3}$$

for $\theta_i \geqq \theta_j$, $i \in \gamma_1$, $j \in \gamma_2$ and $g = (i, j)$, denoting the element of G which interchanges the ith and jth components but leaves the other components unchanged.

The optimal properties of ϕ^* are given by Theorem 2.2 and Corollary 2.1, below.

THEOREM 2.2. *If the loss function satisfies Property L and if the family of posterior distributions P_x with respect to a prior distribution Π, is invariant with respect to G, and stochastically increasing in x then ϕ^* is a Bayes procedure with respect to Π, for the problem of selecting a set of m coordinate values corresponding to the m largest components of θ.*

PROOF. Let $\phi = \{\phi_\gamma(x) : \gamma \in \Gamma\}$ be a decision rule for the given problem, and let

$$t_\gamma(x) = \int L_\gamma(\theta)\, dP_x(\theta)\,. \tag{2.4}$$

From the invariance of P_x and the loss function (Property L, (i)) we have

$$t_\gamma(x) = t_{g\gamma}(gx) \tag{2.5}$$

for all $g \in G$.

As the family of distributions P_x is stochastically increasing in x, we have from SIP (using Lemma 1.1) and the monotonicity property of the loss function (Property L, (ii)) that $t_\gamma(x)$ is non-increasing (non-decreasing) in x_i for $i \in \gamma_1(\gamma_2)$.

From the monotonicity property of $t_\gamma(x)$, shown above, and the invariance property given by (2.5), we have that for $i \in \gamma_1$, $j \in \gamma_2$, $g = (i, j)$ and $x_i \geqq x_j$

$$t_\gamma(x) = t_{g\gamma}(gx) \leqq t_{g\gamma}(x) . \tag{2.6}$$

The Bayes risk of ϕ with respect to Π is given by

$$r_\phi = E_X(\sum_{\gamma \in \Gamma} \phi_\gamma(x) t_\gamma(x)) \tag{2.7}$$

where E_X denotes expectation with respect to the marginal distribution of X. That ϕ^* is a Bayes procedure follows from (2.6) and (2.7). □

Let (A) denote the condition that the integral $\int \xi(x)\, dP_\theta(x)$ is a continuous function of θ for all bounded measurable functions ξ. We have the following corollary.

COROLLARY 2.1. *If the parameter space $\Omega = R_k$ and the support of Π is Ω, if (A) holds and if $L_\gamma(\theta)$ is bounded and continuous in θ for each $\gamma \in \Gamma$ then ϕ^* is admissible, under the conditions of Theorem* 2.2.

The conditions of the corollary imply that the risk is continuous. The admissibility of ϕ^* follows from its Bayes character. The proof is standard (see, for example, Ferguson [4], Theorem 2.3.3 and Theorem 3.7.1).

We make the following remarks on the results given above.

REMARK 1. The optimal property of ϕ^*, given by Theorem 2.2 is not as strong as the one obtained by Eaton [3] in the presence of Property M. When Property M holds, ϕ^* is seen to be a Bayes rule for any symmetric prior. Therefore, ϕ^* minimizes the risk uniformly among all symmetric procedures, and is admissible and minimax.

REMARK 2. If the conditional distribution of X has SIP in θ, under what conditions will there exist a proper prior for which the posterior distribution has SIP in x? We do not have a satisfactory answer to this question. However, we conjecture that if θ is a location parameter then there exists a proper prior for which the posterior distribution is stochastically increasing in x.

3. Application. For illustration we consider below several distributions which satisfy the conditions of Theorem 2.2. These distributions do not have Property M.

EXAMPLE 1. Let the distribution of X have density function

$$f(x, \theta) = U(x - \theta)V(x)/W(\theta)$$

where $W(\theta) = \int U(x - \theta)V(x)\, dx$ and $dx = dx_1 \cdots dx_k$. If $\int W(\theta)\, d\theta = 1$ where

$d\theta = d\theta_1 \cdots d\theta_k$, then $W(\theta)$ represents the density function of a prior distribution Π, say, and $U(x-\theta)$ represents the posterior density function. As x is a location parameter of the posterior distribution, the distribution has SIP in x. If $U(x)$ is a symmetric function of x then the conditions of Theorem 2.2 are satisfied.

For example, let

$$U(x-\theta) = \left(\frac{\lambda}{2}\right)^k \exp(-\lambda \textstyle\sum_{k=1}^k |x_i - \theta_i|) ,$$

$$V(x) = (2\Pi)^{-k/2} \exp(-\tfrac{1}{2} \textstyle\sum_{i=1}^k x_i^2)$$

and

$$W(\theta) = \left(\frac{\lambda}{2}\right)^k e^{k\lambda^2/2} \textstyle\prod_{i=1}^k (e^{-\lambda\theta_i}\phi(\theta_i - \lambda) + e^{\lambda\theta_i}\phi(-\theta_i - \lambda))$$

where λ is a positive number, and $\Phi(y)$ denotes the standard normal distribution function.

EXAMPLE 2. Let the components of X be independently distributed, and let X_i have the double exponential distribution with density

$$f(y; \theta_i) = \tfrac{1}{2} e^{-|y-\theta_i|} , \qquad -\infty < y < \infty , \tag{3.1}$$

$i = 1, \cdots, k$. Let the prior Π be multivariate normal with the null vector as mean and covariance I, so that the components of θ are independently distributed. As the components of X are independently distributed under the conditional distribution, the components of θ are independently distributed under the posterior distribution. From (3.1), the posterior distribution function of θ_i, given x_i, is given by

$$\begin{aligned} H(\theta_i; x_i) &= e^{-x_i}\Phi(\theta_i - 1)/\phi(x_i) \qquad &&\text{for} \quad \theta_i \leqq x_i \\ &= 1 - e^{x_i}\Phi(-\theta_i - 1)/\phi(x_i) &&\text{for} \quad \theta_i > x_i \end{aligned} \tag{3.2}$$

where

$$\phi(y) = e^{-y}\Phi(y-1) + e^{y}\Phi(-y-1) .$$

It is easy to show that $H(\theta_i; x_i)$ is non-increasing in x_i. Thus the posterior distribution of θ has SIP in x, and Theorem 2.2 is applicable.

EXAMPLE 3. Let X be normally distributed with mean θ and arbitrary but known covariance Σ. Suppose that a sample of n observations is taken from the given distribution. Let $\bar{X}$ denote the sample mean. As $\bar{X}$ is a sufficient statistic we consider decision rules based on $\bar{X}$. Therefore, in what follows we substitute $\bar{X}$ for X and $n^{-1}\Sigma$ for Σ.

Let the prior Π be multivariate normal with mean η and covariance $\lambda = (W^{-1} - \Sigma^{-1})^{-1}$ where $W = (W_{ij})$ is given by

$$\begin{aligned} W_{ij} &= a^2 \qquad &&\text{for} \quad i = j \\ &= \rho a^2 &&\text{for} \quad i \neq j , \end{aligned}$$

$0 < \rho < 1$ and $a^2 > 0$. Let U denote the smallest characteristic root of Σ, and let $a^2 < U/(1 + (k - 1)\rho)$. As the largest characteristic root of W is equal to $a^2(1 + (k - 1)\rho)$, we see that λ is positive definite.

The posterior distribution of θ is normal with mean $y = \lambda\Sigma^{-1}(x - \eta) + \eta$ and covariance W. The distribution has SIP in y and is invariant with respect to G. The decision rule $\phi^* = \{\phi_\gamma^*(y) : \gamma \in \Gamma\}$, based on y, is Bayes with respect to Π.

The distribution of X has the property M (see Eaton [3], Proposition 2.2) if and only if $\Sigma^{-1} = c_1 I - c_2 e'e$ where $c_1 > 0$, $-\infty < c_2/c_1 < 1/k$, $e = (1, \cdots, 1)$ and I denotes the $k \times k$ identity matrix.

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] EATON, L. M. (1967). Some optimal properties of ranking procedures. *Ann. Math. Statist.* **38** 124-137.

[4] FERGUSON, T. S. (1967). *Mathematical Statistics, A Decision Theoretic Approach.* Academic Press, New York.

[5] LEHMANN, E. L. (1955). Ordered families of distributions. *Ann. Math. Statist.* **26** 399-419.

[6] LEHMANN, E. L. (1966). On a theorem of Bahadur and Goodman. *Ann. Math. Statist.* **37** 1-6.

DEPARTMENT OF MATHEMATICAL SCIENCES
CLEMSON UNIVERSITY
CLEMSON, SOUTH CAROLINA 29631