The Annals of Statistics
1980, Vol. 8, No. 5, 1164–1167

A NOTE ON OPTIMAL SUBSET SELECTION PROCEDURES[1]

BY SHANTI S. GUPTA AND DENG-YUAN HUANG

Purdue University and Academia Sinica, Taipei

A result for constructing an "optimal" selection rule for selecting a subset of k ($\geqslant 2$) populations is given. Attention is restricted to the class of rules for which the infimum of the probability of a correct selection, over a subset of the parameter space, is guaranteed to be a specified number. In this class a rule is derived which minimizes the supremum of the expected size of the selected subset.

Let $\pi_1, \pi_2, \cdots, \pi_k$ represent k ($\geqslant 2$) independent populations (treatments) and let $X_{i1}, \cdots, X_{in_i}$ be n_i independent random observations from π_i. The quality of the ith population π_i is characterized by a real-valued parameter θ_i, usually unknown. Let $\Omega = \{\underline{\theta} | \underline{\theta}' = (\theta_1, \cdots, \theta_k)\}$ denote the parameter space. Let $\tau_{ij} = \tau_{ij}(\underline{\theta})$ be a measure of separation between π_i and π_j. We assume that there exists a monotonically nonincreasing function h such that $\tau_{ji} = h(\tau_{ij})$. Let $\Omega_i = \{\underline{\theta} | \tau_{ij} \geqslant \tau_0, \forall j \neq i\}$, $1 \leqslant i \leqslant k$, and $\Omega_0 = \Omega - \bar{\Omega}$, where $\bar{\Omega} = \cup_{i=1}^{k} \Omega_i$. For this problem, we assume τ_0 and τ_{ii} as known with $\tau_0 > \tau_{ii}$ for all i. Let $\tau_i = \min_{j \neq i} \tau_{ij}$, $1 \leqslant i \leqslant k$. We define $\tau^* = \max_{1 \leqslant l \leqslant k} \tau_l$. The population associated with τ^* will be called the best population. It should be pointed out that if $\underline{\theta} \in \Omega_i$, then $\tau_i \geqslant \tau_j$ for all j, since for some $j, j \neq i, \tau_{ji} = h(\tau_{ij}) \leqslant h(\tau_0) \leqslant h(\tau_{ii}) = \tau_{ii} < \tau_0$. Thus if $\underline{\theta} \in \Omega_i$, then π_i is the best population. A selection of a subset containing the best population is called a correct selection (CS).

To illustrate the above notation, we assume that independent observations are drawn from π_i which has a normal distribution with unknown mean $\theta_i (i = 1, \cdots, k)$ and known variance σ^2. We define $\tau_{ij} = \theta_i - \theta_j$; then $\tau_i = \theta_i - \theta_{[k]}$ if $\theta_i < \theta_{[k]}$ and $\tau_i = \theta_i - \theta_{[k-1]}$ if $\theta_i = \theta_{[k]}$, where $\theta_{[1]} \leqslant \cdots \leqslant \theta_{[k]}$. In this case, $\tau_{ii} = 0$ for all i and the population with the largest mean, $\theta_{[k]}$, is the best. If, instead, $\tau_{ij} = \theta_j - \theta_i$ then the population with the smallest mean, $\theta_{[1]}$, would be the best. In the above example, $h(t) = -t$, which is a decreasing function.

Let the observed sample vector be denoted by $\underline{X}' = (\underline{X}_1', \cdots, \underline{X}_k')$ where $\underline{X}_i$ has components $X_{i1}, \cdots, X_{in_i}$, $i = 1, \cdots, k$. Let $\delta = (\delta_1, \cdots, \delta_k)$ be a selection procedure where $\delta_i(\underline{x})$ is the probability of selecting $\pi_i (1 \leqslant i \leqslant k)$ based on the observed vector $\underline{X} = \underline{x}$. As measures of goodness of a selection rule, consider two quantities (cf. Lehmann [5]) $R(\underline{\theta}, \delta)$ and $S(\underline{\theta}, \delta)$. We define $S(\underline{\theta}, \delta) = P_{\underline{\theta}}(\mathrm{CS} | \delta)$ and $R(\underline{\theta}, \delta) = \sum_{i=1}^{k} R^{(i)}(\underline{\theta}, \delta_i)$, where $R^{(i)}(\underline{\theta}, \delta_i) = P\{\text{Selecting } \pi_i | \delta\}$. Thus $R(\underline{\theta}, \delta)$ is the expected size of the selected subset. For a specified γ, $(0 < \gamma < 1)$, we restrict attention to

Received September 1976; revised June 1978.

[1]This research was supported by Office of Naval Research Contract N00014-75-C-0455 at Purdue University.

AMS 1970 *subject classifications.* Primary 62F07; secondary 62G30.

Key words and phrases. Subset selection, restricted minimax.

the class $\mathcal{C}$ of all δ such that

$$S(\underline{\theta}, \delta) \geqslant \gamma \quad \text{for} \quad \underline{\theta} \in \bar{\Omega}. \tag{1}$$

We are interested in constructing an optimal procedure δ^0 in $\mathcal{C}$ which minimizes the supremum of $R(\underline{\theta}, \delta)$ over Ω for all $\delta \in \mathcal{C}$, i.e.,

$$\sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta^0) = \min_{\delta \in \mathcal{C}} \sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta). \tag{2}$$

REMARK. For some basic results and the motivation of the subset selection approach, reference can be made to Gupta [4]. Some (different) optimality results assuming a slippage configuration are given by Studden [7] for the exponential family. Recently Bjørnstad [2] has obtained some results on the minimaxity aspects of the procedures of Gupta [4], Seal [6] and Studden [7].

We restrict attention to those selection procedures which depend on the observations only through a sufficient statistic for $\underline{\theta}$.

Let the statistic Z_{ij} be based on the n_i and n_j independent observations from π_i and $\pi_j (i, j = 1, 2, \cdots, k)$, respectively, and suppose that for any i, the statistic $\underline{Z}'_i = (Z_{i1}, \cdots, Z_{ik})$ is invariant sufficient under a transformation group G and let $\underline{\tau}'_i = (\tau_{i1}, \cdots, \tau_{ik})$ be a maximal invariant under the induced group $\bar{G}$. It is well known (see Ferguson [3]) that the distribution of $\underline{Z}_i$ depends only on $\underline{\tau}_i$. For any i, let the joint density of $Z_{ij}, \forall j \neq i$, be $p_{\underline{\theta}}(\underline{z}_i)$ with SIP. Let $p_{\underline{\theta}}(\underline{z}_i)$ be denoted by $p_0(\underline{z}_i)$ when $\tau_{i1} = \cdots = \tau_{ik} = \tau_{ii} =$ constant and by $p_i(\underline{z}_i)$ when $\tau_{i1} = \cdots = \tau_{ik} = \tau_0$, $1 \leqslant i \leqslant k$. In the normal means example, a choice of Z_{ij} might be $\bar{X}_i - \bar{X}_j$, where $\bar{X}_i = 1/n_i \Sigma_{l=1}^{n_i} X_{il}$ and $\bar{X}_j = 1/n_j \Sigma_{l=1}^{n_j} {}_j X_{jl}$. Let ν be a σ-finite measure on $\mathbb{R}^{k-1}$.

Now we state and prove a theorem which provides a solution to the restricted minimax problem as stated in (1) and (2) (cf. Lehmann [5]).

THEOREM. *Suppose that for any* $i, p_i(\underline{z}_i)/p_0(\underline{z}_i)$ *is nondecreasing in* $\underline{z}_i$ *and that* $p_\theta(\underline{z})$ *has the stochastically increasing property. If* $R(\underline{\theta}, \delta^0)$ *is maximized at* $\tau_{ij} = \tau_{ii} =$ *constant, for all* i, j, *where* δ^0 *is given by*

$$\begin{aligned} \delta_i^0(\underline{z}_i) &= 1 \quad \text{if} \ \ p_i(\underline{z}_i) > c p_0(\underline{z}_i), \\ &= \lambda_i \quad \text{if} \ \ p_i(\underline{z}_i) = c p_0(\underline{z}_i), \\ &= 0 \quad \text{if} \ \ p_i(\underline{z}_i) < c p_0(\underline{z}_i), \end{aligned}$$

such that $c(>0)$ *and* λ_i *are determined by* $\int \delta_i^0 p_i = \gamma$, $1 \leqslant i \leqslant k$. *Then* $\delta^0 = (\delta_1^0, \cdots, \delta_k^0)$ *minimizes* $\sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta)$ *subject to* $\inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta) \geqslant \gamma$.

PROOF. For any $\delta \in \mathcal{C}$, $\underline{\theta} \in \bar{\Omega}$ implies $\underline{\theta} \in \Omega_i$ for some i, thus

$$S(\underline{\theta}, \delta) = \int \delta_i(\underline{z}_i) p_{\underline{\theta}}(\underline{z}_i)\, d\nu(\underline{z}_i) \geqslant \min_{1 < i < k} \inf_{\underline{\theta} \in \Omega_i} \int \delta_i(\underline{z}_i) p_{\underline{\theta}}(\underline{z}_i)\, d\nu(\underline{z}_i).$$

We have

$$\inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta) = \min_{1 < i < k} \inf_{\underline{\theta} \in \bar{\Omega}_i} \int \delta_i(\underline{z}_i) p_{\underline{\theta}}(\underline{z}_i)\, d\nu(\underline{z}_i).$$

Hence for any $\delta \in \mathcal{C}$, $\inf_{\underline{\theta} \in \Omega_i} \int \delta_i(\underline{z}_i) p_{\underline{\theta}}(\underline{z}_i)\, d\nu(\underline{z}_i) \geqslant \gamma$, $1 \leqslant i \leqslant k$, and by the assumption that $\int \delta_i^0 p_i = \gamma$, it follows that

$$\int (\delta_i - \delta_i^0)(p_i - cp_0) \leqslant 0$$

which implies

$$\int \delta_i^0 p_0 \leqslant \int \delta_i p_0.$$

By our assumption, $\delta_i^0(\underline{z}_i)$ is nondecreasing in $\underline{z}_i$, hence

$$\inf_{\underline{\theta} \in \bar{\Omega}} S(\underline{\theta}, \delta^0) = \min_{1 \leqslant i \leqslant k} \int \delta_i^0 p_i = \gamma.$$

If $R(\underline{\theta}, \delta^0)$ is maximized at $\tau_{ij} = \tau_{ii} =$ constant, for all i, j, then

$$\sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta) \geqslant \Sigma_{i=1}^k \int \delta_i p_0 \geqslant \Sigma_{i=1}^k \int \delta_i^0 p_0 = \sup_{\underline{\theta} \in \Omega} R(\underline{\theta}, \delta^0),$$

which completes the proof.

As an application of the preceding result, consider the following example:

EXAMPLE. Let $\pi_1, \pi_2, \cdots, \pi_k$ be k independent normal populations with means $\theta_1, \cdots, \theta_k$ and common known variance $\sigma^2 = 1$. The ordered θ_i's are denoted by $\theta_{[1]} \leqslant \cdots \leqslant \theta_{[k]}$. It is assumed that there is no prior knowledge of the correct pairing of the ordered and the unordered θ_i's. Our goal is to select a nonempty subset of the k populations so as to include the population associated with $\theta_{[k]}$.

Let $\bar{X}_i$, $1 \leqslant i \leqslant k$, denote the sample means of independent samples of size n from these populations. The likelihood function of $\underline{\theta}$ is then

$$p_{\underline{\theta}}(\underline{x}) = \Pi_{i=1}^k p_{\theta_i}(\bar{x}_i),$$

where

$$p_{\theta_i}(\bar{x}_i) = \frac{n^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} e^{-(n/2)(\bar{x}_i - \theta_i)^2}, \qquad 1 \leqslant i \leqslant k.$$

Let $\tau_{ij} = \tau_{ij}(\underline{\theta}) = \theta_i - \theta_j$, $1 \leqslant i, j \leqslant k$, $\tau_0 = \Delta > 0$, $\bar{\Omega} = \{\underline{\theta} | \theta_{[k]} - \theta_{[k-1]} \geqslant \Delta\}$ and $Z_{ij} = \bar{X}_i - \bar{X}_j$, $1 \leqslant i, j \leqslant k$. Let $\underline{z}_i' = (z_{i1}, \cdots, z_{ik})$, $\underline{\tau}_i' = (\tau_{i1}, \cdots, \tau_{ik})$, then since $Z_{ii} = 0$ and $\tau_{ii} = 0$, $\forall i$, the joint density of $Z_{ij}, j \neq i$, is given by

$$p_{\underline{\theta}}(\underline{z}_i) = (2\pi)^{(k-1)/2} |\Sigma|^{-\frac{1}{2}} \exp\{-(\underline{z}_i - \underline{\tau}_i)' \Sigma^{-1} (\underline{z}_i - \underline{\tau}_i)\},$$

where

$$\Sigma_{(k-1)\times(k-1)} = \frac{1}{n} \begin{bmatrix} 2 & & 1 \\ & \ddots & \\ 1 & & 2 \end{bmatrix}$$

is the covariance matrix of Z_{ij}'s. Since

$$\frac{p_i(\underline{z}_i)}{p_0(\underline{z}_i)} = \exp\{\underline{z}_i' \Sigma^{-1} \underline{\Delta} + \underline{\Delta}' \Sigma^{-1} \underline{z}_i - \underline{\Delta}' \Sigma^{-1} \underline{\Delta}\} = \exp\left\{\frac{n\Delta}{k}(z_{i1} + \cdots + z_{ik})\right\}$$

is nondecreasing in $z_{ij}, j \neq i$, where $\underline{\Delta}' = (\Delta, \cdots, \Delta)$. Hence

$$\frac{p_i(\underline{z}_i)}{p_0(\underline{z}_i)} > c$$

is equivalent to

$$\bar{x}_i > \frac{1}{k-1}\Sigma_{j\neq i}\bar{x}_j + d.$$

Since $R(\underline{\theta}, \delta^0) = \Sigma_{i=1}^{k} P\{\bar{X}_i > 1/(k-1)\Sigma_{j\neq i}\bar{X}_j + d\}$ is the expected size of the selected subset for Seal's average-type procedure δ^0 [6], the following result of Berger [1] and Bjørnstad [2] applies

$$\sup_{\underline{\theta}\in\Omega} R(\underline{\theta}, \delta^0) = R(\underline{\theta}, \delta^0) \text{ iff } \inf_{\underline{\theta}\in\Omega} S(\underline{\theta}, \delta^0) \geqslant \frac{k-1}{k}.$$

Since the right-hand side is equivalent to $\Phi(((k-1)/k)^{\frac{1}{2}} n^{\frac{1}{2}} d) \leqslant 1/k$, the left-hand side for every fixed $\Delta > 0$ holds if and only if

$$\gamma = 1 - \Phi\left(\left(\frac{k-1}{k}\right)^{\frac{1}{2}} n^{\frac{1}{2}}(d-\Delta)\right) \geqslant 1 - \Phi\left(\Phi^{-1}\left(\frac{1}{k}\right) - \left(\frac{k-1}{k}\right)^{\frac{1}{2}} n^{\frac{1}{2}}\Delta\right),$$

where $\Phi(\cdot)$ is the cdf of the standard normal. Therefore, if for $\Delta > 0$, γ is the chosen in such a way that the preceding inequality holds, then the result of the theorem can be applied.

Acknowledgment. The authors wish to thank the referees and the associate editor for their comments and suggestions which have improved and simplified the presentation.

REFERENCES

[1] Berger, R. L. (1977). Minimax subset selection for loss measured by subset size. Mimeo Ser. #189, Depart. Statist., Purdue Univ.

[2] Bjørnstad, J. (1978). The subset selection problem, II. On the optimality of some subset selection procedures. Mimeo. Series #78-27, Depart. Statist., Purdue Univ.

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

[4] Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics* **7** 225–245.

[5] Lehmann, E. L. (1961). Some model I problems of selection. *Ann. Math. Statist.* **32** 990–1012.

[6] Seal, K. C. (1955). On a class of decision procedures for ranking means of normal populations. *Ann. Math. Statist.* **36** 387–397.

[7] Studden, W. J. (1967). On selecting a subset of k populations containing the best. *Ann. Math. Statist.* **38** 1072–1078.

Department of Statistics
Purdue University
Mathematical Sciences Building
West Lafayette, Indiana 47907

Academia Sinica
Taipei, Taiwan