PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

ON SOME METHODS FOR CONSTRUCTING
OPTIMAL SUBSET SELECTION PROCEDURES

by

Shanti S. Gupta, Purdue University
and
Deng-Yuan Huang, Academia Sinica, Taipei, Taiwan

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #470

September 1976

*This research was supported by 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.

On Some Methods for Constructing
Optimal Subset Selection Procedures

by

Shanti S. Gupta, Purdue University

and

Deng-Yuan Huang, Academia Sinica, Taipei, Taiwan

A

During the past decade, selection and ranking theory has developed rapidly. Many reasonable rules have been proposed. Some good properties have been studied. However, very little work has been done to consider the optimality of a selection procedure, especially in the subset selection approach. In this paper, we are interested in working on some results for optimality. Some classical selection rules are constructed as special cases.

Let $\prec$ denote a partial ordering on the k-dimensional Euclidean space. Let $\underline{x} = (x_1,\ldots,x_k)$ and $\underline{\theta} = (\theta_1,\ldots,\theta_k)$, and $\underline{x} \prec \underline{x}'$ be defined by, $x_i \le x_i'$, $1 \le i \le k$; similarly, $\underline{\theta} \prec \underline{\theta}' \Leftrightarrow \theta_i \le \theta_i'$ $1 \le i \le k$. A measurable subset S of the sample space is called monotone non-decreasing (with respect to $\prec$) if $\underline{x} \in S$ and $\underline{x} \prec \underline{x}'$ implies $\underline{x}' \in S$. Let $P_{\underline{\theta}}(S)$ denote the probability measure of S under the conditional distribution of $\underline{X}$ given $\underline{\theta}$. The distribution is said to have stochastically increasing property (SIP) in $\underline{\theta}$ if $P_{\underline{\theta}}(S) \le P_{\underline{\theta}'}(S)$ for every monotone non-decreasing set S and all $\underline{\theta} \prec \underline{\theta}'$. A function $\varphi(\underline{x})$ is said to be non-decreasing (with respect to $\prec$) if $\varphi(\underline{x}) \le \varphi(\underline{x}')$ for all $\underline{x} \prec \underline{x}'$. Let $E_{\underline{\theta}}$ denote the expectation with respect to the distribution $P_{\underline{\theta}}$.

A characterization of SIP is given by the following lemma (for proof see Lehmann [5], p.400; or see Alam [2]).

*This research was supported by 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.

Lemma 1. A family of distributions $P_{\underline{\theta}}$ has SIP in $\underline{\theta}$ iff $E_{\underline{\theta}}\psi(\underline{X}) \le E_{\underline{\theta}'}\psi(\underline{X})$ for all non-decreasing integrable function $\psi(\underline{x})$ and all $\underline{\theta} < \underline{\theta}'$.

It is easy to generalize Alam's result in [1] as follows.

Theorem 1. Let the distribution of $\underline{X}$ have stochastically increasing property in $\underline{\theta}$, and let $\varphi(\underline{x},\underline{\theta})$ be non-decreasing in $\underline{x}$ and $\underline{\theta}$. Then $E_{\underline{\theta}}\varphi(\underline{X},\underline{\theta})$ is non-decreasing in $\underline{\theta}$.

Proof. Let $\underline{\theta} < \underline{\theta}'$. From the fact that $E_{\underline{\theta}}\varphi(\underline{X},\underline{\theta}) \le E_{\underline{\theta}}\varphi(\underline{X},\underline{\theta}')$ and Lemma 1, we know that $E_{\underline{\theta}}\varphi(\underline{X},\underline{\theta}') \le E_{\underline{\theta}'}\varphi(\underline{X},\underline{\theta}')$. This completes the proof.

Remark 1. It should be pointed out that the dimensionality of the vectors $\underline{x}$ and $\underline{\theta}$ need not be the same. The above results are true, in general.

There are given k populations $\pi_1,\pi_2,\ldots,\pi_k$ of which we wish to select a subset. The quality of the ith population is characterized by a real-valued parameter θ_i. Let Ω denote the whole parameter space.

Let δ be a selection procedure and let $R(\underline{\theta},\delta)$ and $S(\underline{\theta},\delta)$ be some quantities such that $S(\underline{\theta},\delta)$ is large and $R(\underline{\theta},\delta)$ is small.

Lehmann [7] has proved the following result.

Lemma 2. Let $\mathcal{B}$ be a σ-field of subsets of the parameter space Ω and let λ and μ be probability distributions over $(\Omega,\mathcal{B})$. Let A,B be two positive constants and let δ_0 maximize the integral

(1) $$B\int S(\underline{\theta},\delta)d\mu(\underline{\theta})-A\int R(\underline{\theta},\delta)d\lambda(\underline{\theta}).$$

Then δ_0 minimizes $\sup R(\underline{\theta},\delta)$ subject to

(2) $$\inf S(\underline{\theta},\delta) \ge \gamma$$

provided

(3) $$\int R(\underline{\theta},\delta_0)d\lambda(\underline{\theta}) = \sup R(\underline{\theta},\delta_0)$$

and

(4) $$\int S(\underline{\theta},\delta_0)d\mu(\underline{\theta}) = \inf S(\underline{\theta},\delta_0) = \gamma.$$

Let $\Omega_i = \{\underline{\tau}_i | \tau_{ij} \geq \tau_0, j \neq i\}$, $1 \leq i \leq k$, be a partition of Ω such that $\Omega = \bigcup_{i=0}^{k} \Omega_i$, let τ_{ij} be a distance between π_i and π_j, $\underline{\tau}_i = (\tau_{i1}, \ldots \tau_{ik})$ with k-1 elements (since τ_{ii} is not considered), τ_{ii} and τ_0 are known constants. Let $\overline{\Omega} = \bigcup_{1}^{k} \Omega_i$ and $\tau_i = \min_{j \neq i} \tau_{ij}$ and let $\tau_j = \max_{1 \leq \ell \leq k} \tau_\ell$ be associated with π_j which we call the best.

Let $\delta = (\delta_1, \ldots, \delta_k)$ and $S(\underline{\tau}, \delta) = S(\underline{\tau}, \delta_i) = P_{\underline{\tau}}\{\text{Selecting } \pi_i | \delta_i\} = \int \delta_i p_{\underline{\tau}}$ if $\underline{\tau} \in \Omega_i$, $1 \leq i \leq k$. Let $R^{(i)}(\underline{\tau}, \delta_i) = P_{\underline{\tau}}\{\text{Selecting } \pi_i | \delta_i\}$, if $\tau \in \Omega$, and $R(\underline{\tau}, \delta) = \sum_{i=1}^{k} R^{(i)}(\underline{\tau}, \delta_i)$. Let Z_{ij} be a sufficient and maximal invariant statistic for τ_{ij}, $1 \leq i, j \leq k$, $j \neq i$. We know that the distribution of Z_{ij}, depends only on τ_{ij} (see [6]).

Oosterhoff [8] defines a monotone likelihood ratio density for a random vector $\underline{X}$ with m components as follows: Let $\underline{\theta}$ be an m-dimensional vector of parameters. A partial ordering of points in R^k is defined by $\underline{x}' \underset{\neq}{<} \underline{x}$, meaning $x_i' \leq x_i$ for $i = 1,2,\ldots,k$, and the inequality is strict for at least one component. The density $f(\underline{x},\underline{\theta})$ has monotone likelihood ratio if for all $\underline{\theta}'' \underset{\neq}{<} \underline{\theta}'$, $[f(\underline{x},\theta')/f(\underline{x},\underline{\theta}'')]$ is non-decreasing in $\underline{x}$.

Let the joint density of Z_{ij}, $j \neq i$, be $p_{\underline{\tau}_i}(\underline{z}_i)$. Let $p_{\underline{\tau}_i}$ be denoted by p_0 when $\tau_{i1} = \ldots = \tau_{ik} = \tau_{ii}$ and by p_i when $\tau_{i1} = \ldots = \tau_{ik} = \tau_0$, $\tau_0 > \tau_{ii}$.

Restricted Minimax Selection Procedures

Theorem 2. Suppose that $p_{\underline{\tau}}(\underline{z})$ has monotone likelihood ratio. Assume that the supremum of $R(\underline{\tau},\delta^0)$ over Ω occurs at $\tau_{ij} = \tau_{ii}$ for all i,j, where for any i,

$$\delta_i^0 = \begin{cases} 1 & \text{if } \dfrac{p_i(\underline{z}_i)}{p_0(\underline{z}_i)} > c, \\ a & \qquad\qquad = \ , \\ 0 & \qquad\qquad < \ , \end{cases}$$

where c and a are determined by

$$\int \delta_i^0(\underline{z}_i)p_i(\underline{z}_i)d\underline{z}_i = \gamma .$$

Then $\delta^0 = (\delta_1^0,\ldots,\delta_k^0)$ minimizes $\sup_{\underline{\tau}\in\Omega} R(\underline{\tau},\delta)$ subject to $\inf_{\underline{\tau}\in\Omega} S(\underline{\tau},\delta) \geq \gamma$.

Proof. Let λ be the distribution which assigns probability one to the point $\underline{\tau}_i = (\tau_{ii},\ldots,\tau_{ii})$ and μ the distribution which assigns probability $\frac{1}{k}$ to the points $\underline{\tau}_0 = (\tau_0,\ldots,\tau_0)$. Then by Theorem 1, we have

$$\inf_{\underline{\tau}\in\Omega} S(\underline{\tau},\delta^0) = \min_{1\leq i\leq k} \inf_{\underline{\tau}\in\Omega_i} S(\underline{\tau},\delta^0) = \min_{1\leq i\leq k} S(\underline{\tau}_0,\delta_i^0)$$

$$= \frac{1}{k}\sum_{i=1}^{k} \int \delta_i^0 p_i = \int_{\Omega} S(\underline{\tau},\delta^0)d\mu(\underline{\tau}),$$

(since $\int \delta_i^0 p_i = \gamma$, $1 \leq i \leq k$).

$$\sup_{\underline{\tau}\in\Omega} R(\underline{\tau},\delta^0) = \int(\delta_1^0 + \ldots + \delta_k^0)\, p_0, \text{ and}$$

hence (1) reduces to

$$B\int_{\Omega} S(\underline{\tau},\delta^0)d\mu(\underline{\theta}) - A\int R(\underline{\tau},\delta^0)d\lambda(\underline{\theta}) \tag{5}$$

$$= \frac{B}{k}\sum_{i=1}^{k} \int \delta_i^0 p_i - A\sum_{i=1}^{k} \int \delta_i^0 p_0$$

$$= \int \sum_{i=1}^{k} \delta_i^0\left(\frac{B}{k} p_i - Ap_0\right).$$

Since $0 \leq \delta_i \leq 1$, (5) is maximized by putting $\delta_i^0 = 0$ or 1 as $\frac{B}{k} p_i <$ or $> Ap_0$.

Example. Let $g_{\underline{\theta}}(\underline{x}) = \prod_{j=1}^{k} g_{\theta_i}(\overline{x}_i)$, where $g_{\theta_i}(\overline{x}_i) = \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n}{2}(\overline{x}_i-\theta_i)^2}$. Let $\tau_{ij} = \theta_i-\theta_j$, $1 \leq j \leq k$, $j\neq i$, $\tau_{ii} = 0$, $\tau_0 = \Delta>0$ and $z_{ij} = \overline{X}_i-\overline{X}_j$, $j\neq i$. Let $\underline{\tau}_i' = (\tau_{i1},\ldots,\tau_{ik})$ and $\underline{z}_i' = (z_{i1},\ldots,z_{ik})$.

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

where $\Sigma_{(k-1)\times(k-1)} = \frac{1}{n}\begin{pmatrix} 2 & & 1 \\ & \ddots & \\ 1 & & 2 \end{pmatrix}$ is the positive definite covariance matrix of z_{ij}'s. We know that

$$\frac{p_i(\underline{z}_i)}{p_0(\underline{z}_i)} = \exp\{\underline{z}_i' \Sigma^{-1}\underline{\Delta} + \underline{\Delta}' \Sigma^{-1}\underline{z}_i - \underline{\Delta}' \Sigma^{-1}\underline{\Delta}\}$$

is non-decreasing in z_{ij}, $j \neq i$, and is equivalent to

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

We can show that the supremum of $R(\underline{\tau},\delta^0)$ over Ω occurs at $\theta_1 = \ldots = \theta_k$. Note that the above is the Seal's procedure [9] to select a subset containing the population associated with the largest θ_i's, so called the "best" population.

<u>An essential complete class of multiple decision procedures</u>.

A point x_0 is called a change point for a function h if in some neighborhood of x_0,

$$h(x)h(x^*) \leq 0,$$

whenever $x \leq x_0 \leq x^*$, and for some $x_1 \leq x_0 \leq x_1^*$, $h(x_1) \neq 0$ and $h(x_1^*) \neq 0$ with $x_1 \neq x_1^*$.

Karlin and Rubin [4] have proved the following result.

<u>Lemma 3</u>. If φ changes sign at most once in one-dimensional Euclidean space R^1, then

$$\psi(w) = \int p(x|w)\varphi(x)d\mu(x)$$

changes sign at most once, where μ is a σ-finite measure on R' and $p(x|w)$ is the density of X with monotone likelihood ratio (MLR).

<u>Remark</u>: It is useful to note that ψ changes sign in the same direction as φ if it changes sign at all.

A decision rule $d(\underline{x})$ is called monotone if

$$d(\underline{x}) = \begin{cases} 1 & \text{if } \underline{x} \succ \underline{x}_0, \\ 0 & \text{otherwise.} \end{cases}$$

Let $F_{\underline{\tau}}$ be the continuous distribution function of the pdf $p_{\underline{\tau}}(\underline{z})$ for any vector $\underline{\tau}$, and let $F_{\underline{\tau}}^{-1}(y)$ be the supremum of the set $\{\underline{x}: F_{\underline{\tau}}(\underline{x}) = y\}$ with respect to the partial order "<" for any real y and $\underline{\tau}$.

Let δ_i be any non-monotone rule and for any given real vector $\underline{b}_i = (b_{i1},\ldots,b_{ik})$ with k-1 elements, such that

$$R^i(\underline{\tau}_i,\delta_i) = \int(1-\delta_i(\underline{z}_i))p_{\underline{\tau}_i}(\underline{z}_i)d\nu(\underline{z}_i),$$

where ν is a σ-finite measure on k-1 dim. Euclidean space R^{k-1}. Let

$$\delta_i^0(\underline{z}_i) = \begin{cases} 1 & \text{if } \underline{z}_i > F_{\underline{b}_i}^{-1}(R^i(\underline{b}_i,\delta_i)), \\ 0 & \text{otherwise.} \end{cases}$$

Then $R^i(\underline{b}_i,\delta_i) = \int(1-\delta_i^0)p_{\underline{b}_i}(\underline{z}_i)d\nu(\underline{z}_i)$

$$= \int_{-\infty}^{F_{\underline{b}_i}^{-1}(R^i(\underline{b}_i,\delta_i))} p_{\underline{b}_i}(\underline{z}_i)d\nu(\underline{z}_i) = R^i(\underline{b}_i,\delta_i).$$

Suppose that δ_i is not monotone in z_{ij}, for fixed $z_{i\ell}$, $\ell \neq j$. For each fixed $z_{i\ell}$, $(\ell \neq j)$ define $\delta_i^0(\underline{z}_i)$ as above is monotone in z_{ij}, and to satisfy $R^i(\underline{b}_i,\delta_i^0) = R^i(\underline{b}_i,\delta_i)$. And

$$R^i(\underline{\tau}_i,\delta_i^0)-R^i(\underline{\tau}_i,\delta_i)$$

$$= \int[\delta_i-\delta_i^0]p_{\underline{\tau}_i}(\underline{z}_i)d\mu(z_{ij}).$$

Since δ_i^0 is monotone, $\delta_i-\delta_i^0$ as a function of z_{ij} has at most one sign change in the order of plus to minus. Using this fact, the MLR of $p_{\underline{\tau}_i}(\underline{z}_i)$ and Lemma 3, we have

$$R^i(\underline{\tau}_i,\delta_i^0) \leq R^i(\underline{\tau}_i,\delta_i) \text{ for } \tau_{ij} \geq b_{ij}.$$

Hence $$R^i(\underline{\tau},\delta_i^0) \leq R^i(\underline{\tau}_i\delta_i) \text{ for } \underline{\tau}_i > \underline{b}_i.$$

As the normal means example, a monotone procedure is the following form:

$$\delta_i^0(\underline{z}_i) = \begin{cases} 1 & \text{if } \underline{z}_i > \underline{C}, \\ 0 & \text{otherwise,} \quad \underline{C} = (C_j;y\neq i), \end{cases}$$

this is equivalent to

$$\delta_i^0 = \begin{cases} 1 & \text{if } \bar{x}_i \geq \max_{j \neq i}(\bar{x}_j + C_j), \\ 0 & \text{otherwise.} \end{cases}$$

It should be pointed out that the monotone procedure $\delta^0 = (\delta_1^0, \ldots, \delta_k^0)$ is the usual Gupta type procedure (see Gupta [3])to select a subset containing the best population.

[1] Alam, K. (1972). Unimodality of the distribution of an order statistic. Ann. Math. Statist. 43, 2041-2044.

[2] Alam, K. (1973). On a multiple decision rule. Ann of Statist. 1, 750-755.

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

[4] Karlin, S. and Rubin, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. Ann. Math. Statist. 27, 272-299.

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

[6] Lehmann, E. L. (1959). Testing Statistical Hypotheses. Wiley.

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

[8] Oosterhoff, J. (1969). Combination of One-Sided Statistical Tests. Mathematical Centre Tracts, Amsterdam.

[9] Seal, K. C. (1955). On a class of decision procedures for ranking means of normal populations. Ann. Math. Statist. 36, 387-897.

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #470	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On Some Methods For Constructing Optimal Subset Selection Procedures	9	5. TYPE OF REPORT & PERIOD COVERED Technical rept. 6. PERFORMING ORG. REPORT NUMBER Mimeo. Series #470
7. AUTHOR(s) Shanti S. Gupta and D. Y. Huang	15	8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Dept. of Statistics, W. Lafayette, IN 47907	16	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-243
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC	11	12. REPORT DATE Sept. 1976 13. NUMBER OF PAGES 8
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	12 10p.	15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited

14 Mimeograph ser-470

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

restricted minimax, optimal selection, stochastically increasing property, monotone likelihood ratio, monotone procedure, essential complete

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper, we are concerned with the construction of "optimal" subset selection procedures. Some classical selection procedures are considered as special cases.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601 291730 unclassified