ON BAYES AND EMPIRICAL BAYES RULES FOR SELECTING GOOD POPULATIONS*

by

Shanti S. Gupta and Lii-Yuh Leu
Purdue University

Technical Report #83-37

Department of Statistics
Purdue University

September 1983

(Revised August 1985)

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

ON BAYES AND EMPIRICAL BAYES RULES
FOR SELECTING GOOD POPULATIONS

by

Shanti S. Gupta and Lii-Yuh Leu
Purdue University and National Central University

Abstract

This paper deals with the problem of selecting all populations which are close to a control or standard. A general Bayes rule for the above problem is derived. Empirical Bayes rules are derived when the populations are assumed to be uniformly distributed. Under some conditions on the marginal and prior distributions, the rate of convergence of the empirical Bayes risk to the minimum Bayes risk is investigated. The rate of convergence is shown to be $n^{-\delta/3}$ for some δ, $0 < \delta < 2$.

Key words: Bayes rules, empirical Bayes rules, selection procedures, asymptotically optimal, rate of convergence.

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

ON BAYES AND EMPIRICAL BAYES RULES FOR SELECTING GOOD POPULATIONS

by

Shanti S. Gupta and Lii-Yuh Leu
Purdue University and National Central University

1. Introduction

Empirical Bayes rules have been considered for multiple decision problems by Deely (1965), Van Ryzin (1970), Van Ryzin and Susarla (1977), Singh (1977), and Gupta and Hsiao (1983). Most of the papers are concerned with the selection of the best population where best is usually defined in terms of the largest or smallest unknown parameter. Gupta and Hsiao (1983) considered the problem which is concerned with the selection of populations better than a control. In some practical applications, one may be interested in selecting populations which are close to a control. We will consider such a problem in this paper.

In Section 2, we propose a general Bayes rule for selecting good populations. In Section 3, assuming that the populations are uniformly distributed, empirical Bayes rules are derived for both the known control parameter and the unknown control parameter cases. Under some conditions on the marginal and prior distributions, the rate of convergence of the empirical Bayes risk to the minimum Bayes risk is investigated. The rate of convergence is shown to be $n^{-\delta/3}$ for some δ, $0 < \delta < 2$.

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

2. A General Bayes Rule for Selecting Good Populations

Let $\pi_0,\pi_1,\ldots,\pi_k$ be (k+1) independent populations which are characterized by parameters $\theta_0,\theta_1,\ldots,\theta_k$, respectively. Assume that π_0 is the control population with parameter θ_0 which may be known or unknown. When θ_0 is unknown, let $\underline{\theta} = (\theta_0,\theta_1,\ldots,\theta_k)$ and $\underline{X} = (X_0,X_1,\ldots,X_k)$ where X_i is an observation from π_i, $i = 0,1,\ldots,k$. When θ_0 is known, no observation from population π_0 is taken, and θ_0, X_0 are deleted from $\underline{\theta}$ and $\underline{X}$, respectively. When there is no confusion, $\underline{\theta}$ and $\underline{X}$ are used to represent either case. We define population π_i to be a good population if $|\theta_i-\theta_0| < \Delta$ and a bad population if $|\theta_i-\theta_0| \geq \Delta$, where $\Delta > 0$ is a pre-assigned constant. Our goal is to find a Bayes rule which selects all good populations and rejects bad ones. We assume that given θ_i, X_i has a probability density function $f(x_i|\theta_i)$ with respect to a σ-finite measure μ, for $i = 0,1,\ldots,k$, and $\underline{\theta}$ has a prior distribution $G(\underline{\theta}) = \prod_{i=0}^{k} G_i(\theta_i)$ on the parameter space Ω. Let $\mathcal{G} = \{s|s \subseteq \{1,2,\ldots,k\}\}$ be the action space and let

$$(2.1)\qquad L(\underline{\theta},s) = \sum_{i\in s} \{c_1(\theta_0-\Delta-\theta_i)I_{\{\theta_i\leq\theta_0-\Delta\}}(\theta_i) + c_2(\theta_i-\theta_0-\Delta)I_{\{\theta_0+\Delta\leq\theta_i\}}(\theta_i)\} + \sum_{i\notin s} \{c_3(\theta_i-\theta_0+\Delta)I_{\{\theta_0-\Delta<\theta_i\leq\theta_0\}}(\theta_i) + c_4(\theta_0+\Delta-\theta_i)I_{\{\theta_0<\theta_i<\theta_0+\Delta\}}(\theta_i)\}$$

be the loss function defined on $\Omega \times \mathcal{G}$, where c_i, $i = 1,2,3,4$ are positive constants and I is the indicator function.

Since the action space is finite, attention can be restricted to the non-randomized rules for deriving the Bayes rules. For a non-randomized decision function δ: $\mathcal{X} \to \mathcal{G}$, the corresponding Bayes risk with respect to G is given by

$$(2.2)\qquad r(G,\delta) = \int_{\mathcal{X}} \int_{\Omega} L(\underline{\theta},\delta(\underline{x}))f(\underline{x}|\underline{\theta})dG(\underline{\theta})d\mu(\underline{x}),$$

where $\mathcal{X}$ is the sample space and $f(\underline{x}|\underline{\theta}) = \prod_i f(x_i|\theta_i)$.

In the sequel we consider the special case where $c_1 = c_2 = c_3 = c_4$ is a constant which can be taken to be unity without loss of generality. If ϕ is the empty set, (2.1) can be expressed as

$$(2.3) \qquad L(\underline{\theta},s) = L(\underline{\theta},\phi) + \sum_{i\in s}\{(\theta_0-\Delta-\theta_i)I_{\{\theta_i\le\theta_0\}}(\theta_i) + (\theta_i-\theta_0-\Delta)I_{\{\theta_0<\theta_i\}}(\theta_i)\}.$$

Hence, for any δ, we have

$$(2.4) \qquad r(G,\delta) - r(G,\phi) = \int_{\mathcal{X}} \sum_{i\in\delta(\underline{x})} \{\int_{\Omega}(\theta_0-\Delta-\theta_i)f(\underline{x}|\underline{\theta})dG(\underline{\theta}) + 2\int_{\{\theta_0<\theta_i\}}(\theta_i-\theta_0)f(\underline{x}|\underline{\theta})dG(\underline{\theta})\}d\mu(\underline{x}).$$

From (2.4), it follows that if

$$(2.5) \qquad \int_{\Omega}(\theta_0-\Delta-\theta_i)f(\underline{x}|\underline{\theta})dG(\underline{\theta})+2\int_{\{\theta_0<\theta_i\}}(\theta_i-\theta_0)f(\underline{x}|\underline{\theta})dG(\underline{\theta}) < 0,$$

then the Bayes rule is $\delta_B(\underline{x})$, $i \in \delta_B(\underline{x})$.

Let $m_i(x_i) = \int_{\Omega} f(x_i|\theta_i)dG_i(\theta_i)$ be the marginal distribution of X_i, $\pi(\theta_i|x_i)$ be the posterior distribution of θ_i given $X_i = x_i$, and $E(\theta_i|x_i)$ be the expected value of θ_i given $X_i = x_i$. If $m_i(x_i) > 0$ for all x_i, then (2.5) is equivalent to

$$(2.6) \qquad (\theta_0-\Delta)-E(\theta_i|x_i) + 2\int_{\{\theta_0<\theta_i\}}(\theta_i-\theta_0)\pi(\theta_i|x_i)d\theta_i < 0$$

if θ_0 is known, or

$$(2.7) \qquad E(\theta_0|x_0)-E(\theta_i|x_i)-\Delta+2\int_{\{\theta_0<\theta_i\}}(\theta_i-\theta_0)\pi(\theta_i|x_i)\pi(\theta_0|x_0)d\theta_i d\theta_0 < 0$$

if θ_0 is unknown.

From the above discussion, we have the following main result:

Theorem 2.1. Under the loss function (2.3), the Bayes rule $\delta_B(\underline{x})$ with respect to G is as follows:

(a) If θ_0 is known, then $i \in \delta_B(\underline{x})$ if the inequality (2.6) holds.

(b) If θ_0 is unknown, then $i \in \delta_B(\underline{x})$ if the inequality (2.7) holds

Example:

Suppose that

$$(2.8) \qquad f(x_i|\theta_i) = e^{-\theta_i}\theta_i^{x_i}/(x_i!), \; x_i = 0,1,\ldots; \quad \theta_i > 0$$

and θ_i has a prior distribution $g_i(\theta_i) = G_i'(\theta_i)$ which is given by

$$(2.9) \qquad g_i(\theta_i) = \beta_i^{\alpha_i}\theta_i^{\alpha_i-1} e^{-\beta_i\theta_i} I_{(0,\infty)}(\theta_i)/\Gamma(\alpha_i),$$

where $\alpha_i > 0$ and $\beta_i > 0$ are known. Then the Bayes rule $\delta_B(\underline{x})$ is given by

(a) If θ_0 is known, then $i \in \delta_B(\underline{x})$ if

$$\frac{x_i+\alpha_i}{1+\beta_i}\{1-2\Gamma(\theta_0(1+\beta_i);\, x_i+\alpha_i+1)\} - \theta_0\{1-2\Gamma(\theta_0(1+\beta_i);\, x_i+\alpha_i)\} < \Delta,$$

where

$$\Gamma(a;\, \alpha) = \int_0^a \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-x}dx, \quad a > 0, \quad \alpha > 0.$$

(b) If θ_0 is unknown and $\beta_i = \beta$, $i = 0,1,\ldots,k$, then $i \in \delta_B(\underline{x})$ if

$$\frac{x_i+\alpha_i}{1+\beta}\{2I(\tfrac{1}{2};\, x_0+\alpha_0, x_i+\alpha_i)-1\} +$$

$$\frac{x_0+\alpha_0}{1+\beta}\{1 + 2I(\tfrac{1}{2};\, x_0+\alpha_0, x_i+\alpha_i)-4I(\tfrac{1}{2};\, x_0+\alpha_0+1, x_i+\alpha_i)\} < \Delta,$$

where

$$I(z;\, \alpha,\beta) = \int_0^z \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx, \quad \alpha > 0, \quad \beta > 0,$$

and

$$B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta).$$

3. Empirical Bayes Rules for Uniform Populations

In this section we will assume that X_i has the probability density function $f(x_i|\theta_i) = \frac{1}{\theta_i} I_{(0,\theta_i)}(x_i)$, where $\theta_i > 0$ is unknown. Suppose that $\underset{\sim}{\theta}$ has a prior distribution $G(\underset{\sim}{\theta}) = \prod_i G_i(\theta_i)$ on Ω and G_i has a continuous positive probability density function g_i. Let $m_i(x_i)$ and $M(x_i)$ be the marginal pdf and cdf of X_i, respectively.

3.1. Known Control Population

When θ_0 is known, we assume $\theta_0 > \Delta$ and let

$$\Delta_{G_i}(x_i) = (\theta_0-\Delta)m_i(x_i)-\int_{x_i}^{\infty} dG_i(\theta_i)+2\int_{(x_i,\infty)\cap(\theta_0,\infty)}(\theta_i-\theta_0)f(x_i|\theta_i)dG_i(\theta_i). \tag{3.1}$$

From (2.5), we have $i \in \delta_B(\underset{\sim}{x})$ if $\Delta_{G_i}(x_i) < 0$.

We can show that

$$\Delta_{G_i}(x_i) = \begin{cases} \Delta_{1,G_i}(x_i) = (\theta_0-\Delta-x_i)m_i(x_i)+1-2M_i(\theta_0)+M_i(x_i), & \text{if } x_i \le \theta_0. \quad (3.2) \\ \Delta_{2,G_i}(x_i) = (x_i-\theta_0-\Delta)m_i(x_i)+1-M_i(x_i), & \text{if } x_i > \theta_0. \quad (3.3) \end{cases}$$

Therefore

$$\delta_B(\underset{\sim}{x}) = \{i|x_i \le \theta_0,\Delta_{1,G_i}(x_i) < 0\}\cup\{i|x_i > \theta_0,\Delta_{2,G_i}(x_i) < 0\}. \tag{3.4}$$

Remarks:

(1) $\Delta_{G_i}(x_i)$ is strictly decreasing for $0 < x_i < \theta_0-\Delta$, strictly increasing for $\theta_0-\Delta < x_i < \theta_0+\Delta$, and strictly decreasing for $\theta_0+\Delta < x_i$.

(2) If $x_i \ge \theta_0+\Delta$, then $\Delta_{G_i}(x_i) \ge 0$. Hence $i \notin \delta_B(\underset{\sim}{x})$ if $x_i \ge \theta_0+\Delta$.

(3) If G_i is such that $1-2M_i(\theta_0) + M_i(\theta_0-\Delta) \ge 0$, then $\delta_B(\underset{\sim}{x}) = \phi$. Otherwise, $i \in \delta_B(\underset{\sim}{x})$ if $(\theta_0-\Delta)-d_1 < x_i < (\theta_0+\Delta)-d_2$, for some positive real numbers d_1 and d_2. Hence a selection rule of this type is a Bayes rule relative to some prior distribution.

If G is unknown, the Bayes rules are not obtainable. In this case, we consider a sequence $(\underline{X}_1,\underline{\Delta}_1)$, $(\underline{X}_2,\underline{\Delta}_2),\ldots,$ of independent pairs of random vectors where each $\underline{\Delta}_i$ is distributed as G on Ω and $\underline{X}_i = (X_{i1},\ldots,X_{ik})$ has conditional density function $f(\underline{x}|\underline{\theta})$ given $\underline{\Delta}_i = \underline{\theta}$. The empirical Bayes approach, which was introduced by Robbins (1956), attempts to construct a decision rule concerning $\underline{\Delta}_{n+1}$ at stage n+1 based on $\underline{X}_1,\ldots,\underline{X}_{n+1}$. The risk at stage n+1 by taking action $\delta_n(\underline{x};\ \underline{X}_1,\ldots,\underline{X}_n) = \delta_n(\underline{x})$ is given by

$$(3.5)\qquad r_n(G,\delta_n) = \int_{\mathcal{X}} E_n\{\sum_{i\in\delta_n(\underline{x})}[\int_{\Omega}(\theta_0-\Delta-\theta_i)f(\underline{x}|\underline{\theta})dG(\underline{\theta}) + 2\int_{(\theta_0,\infty)}(\theta_i-\theta_0)f(\underline{x}|\underline{\theta})dG(\underline{\theta})]\}d\underline{x} + r(G,\phi),$$

where E_n denotes the expectation with respect to the n independent random vectors $\underline{X}_1,\ldots,\underline{X}_n$ each with a common density function

$$m(\underline{x}) = \int_{\Omega} f(\underline{x}|\underline{\theta})dG(\underline{\theta}) = \prod_{i=1}^{k} m_i(x_i).$$

<u>Definition 3.1</u>. The sequence of procedures $\{\delta_n\}$ is said to be asymptotically optimal (a.o.) relative to G if $r_n(G,\delta_n) - r(G) = o(1)$ as $n \to \infty$, where $r(G) = \inf_{\delta} r(G,\delta)$.

In order to find an a.o. sequence of rules, let $\delta_{1,B}(\underline{x}) = \{i|x_i \le \theta_0,\Delta_{1,G_i}(x_i) < 0\}$ and $\delta_{2,B}(\underline{x}) = \{i|\theta_0 < x_i < \theta_0+\Delta,$ $\Delta_{2,G_i}(x_i) < 0\}$. From (3.4) and Remark (2), we have $\delta_B(\underline{x}) = \delta_{1,B}(\underline{x}) \cup \delta_{2,B}(\underline{x})$. For any $i = 1,2,\ldots,k$ and $\ell = 1,2$, let $\Delta_{\ell,i,n}(x_i) = \Delta_{\ell,i}(x_i;\ X_{1i},\ldots,X_{ni})$, $n = 1,2,\ldots$ be two sequences of real-valued measurable functions, and define δ_n by

$$(3.6)\qquad \delta_n(\underline{x}) = \delta_{1,n}(\underline{x}) \cup \delta_{2,n}(\underline{x}),$$

where

$$\delta_{1,n}(\underline{x}) = \{i|x_i \le \theta_0,\ \Delta_{1,i,n}(x_i) < 0\}$$

and

$$\delta_{2,n}(\underline{x}) = \{i|\theta_0 < x_i < \theta_0+\Delta,\Delta_{2,i,n}(x_i) < 0\}.$$

Then we have the following theorem:

<u>Theorem 3.1</u>. If $\int_0^\infty \theta dG_i(\theta) < \infty$, $i = 1,2,\ldots,k$ and $\Delta_{1,i,n}(x_i) \xrightarrow{P} \Delta_{1,G_i}(x_i)$, for almost all $x_i \leq \theta_0$ and $\Delta_{2,i,n}(x_i) \xrightarrow{P} \Delta_{2,G_i}(x_i)$, for almost all $\theta_0 < x_i < \theta_0+\Delta$, where $\xrightarrow{P}$ means convergence in probability, ~~then $\{\delta_n(\underline{x})\}$~~ defined by (3.6) is a.o. relative to G.

Proof. Analogous to the proof of Theorem 2.1 of Gupta and Hsiao (1983), it can be shown that

$$0 \leq \int_\Omega L(\underline{\theta},\delta_n(\underline{x}))f(\underline{x}|\underline{\theta})dG(\underline{\theta})-\int_\Omega L(\underline{\theta},\delta_B(\underline{x}))f(\underline{x}|\underline{\theta})dG(\underline{\theta}) \leq 4\varepsilon \sum_{i=1}^{k} \Big(\prod_{\substack{j=1\\ j\neq i}}^{k} m_j(x_j)\Big)$$

with probability near 1, for large n. Hence

$$\int_\Omega L(\underline{\theta},\delta_n(\underline{x}))f(\underline{x}|\underline{\theta})dG(\underline{\theta}) \xrightarrow{P} \int_\Omega L(\underline{\theta},\delta_B(\underline{x}))f(\underline{x}|\underline{\theta})dG(\underline{\theta})$$

for almost $\underline{x}$. By Corollary 1 of Robbins (1964), $\{\delta_n(\underline{x})\}$ is a.o. relative to G.

From Theorem 3.1, our problem is reduced to finding consistent estimators of $\Delta_{1,G_i}(x_i)$ and $\Delta_{2,G_i}(x_i)$. Let

$$M_{in}(x_i) = \frac{1}{n}\sum_{j=1}^{n} I_{(-\infty,x_i]}(X_{ji}), \tag{3.7}$$

then $M_{in}(x_i) \xrightarrow{P} M_i(x_i)$ for all $x_i > 0$. Next, let $\varphi(x) \geq 0$ be a Borel function satisfying the following conditions:

(3.8) (i) $\sup_{-\infty<x<\infty} \varphi(x) < \infty$, (ii) $\int_{-\infty}^{\infty} \varphi(x)dx = 1$, and (iii) $\lim_{x\to\infty} x\varphi(x) = 0$.

Let $\{h(n)\}$ be a sequence of positive constants satisfying the following conditions:

(3.9) (i) $h(n) \to 0$ as $n \to \infty$ and (ii) $nh(n) \to \infty$ as $n \to \infty$.

If we define

$$m_{in}(x) = \frac{1}{nh(n)}\sum_{j=1}^{n} \varphi\Big(\frac{x-X_{ji}}{h(n)}\Big), \tag{3.10}$$

then $m_{in}(x) \xrightarrow{P} m_i(x)$ for all x (see Parzen (1962)). For $i = 1,2,\ldots,k$, let

(3.11) $\Delta_{1,i,n}(x_i) = (\theta_0-\Delta-x_i)m_{in}(x_i)+1-2M_{in}(\theta_0)+M_{in}(x_i)$

and

(3.12) $\Delta_{2,i,n}(x_i) = (x_i-\theta_0-\Delta)m_{in}(x_i)+1-M_{in}(x_i)$.

Then

$$\Delta_{1,i,n}(x_i) \xrightarrow{P} \Delta_{1,G_i}(x_i) \quad \text{for all } x_i \leq \theta_0$$

and

$$\Delta_{2,i,n}(x_i) \xrightarrow{P} \Delta_{2,G_i}(x_i) \text{ for all } \theta_0 < x_i < \theta_0+\Delta.$$

Thus the sequence of procedures $\{\delta_n\}$ defined by

$$\delta_n(\underline{x}) = \{i | x_i \leq \theta_0, \Delta_{1,i,n}(x_i) < 0\} \cup \{i | \theta_0 < x_i < \theta_0+\Delta,\ \Delta_{2,i,n}(x_i) < 0\},$$

is a.o. relative to G.

3.2. Unknown Control Population

In this subsection we consider the case of the unknown parameter θ_0 of the control population π_0. As indicated in Section 2, the notations $\underline{\theta}$, Ω, $\underline{X}$, $\mathcal{X}$, $G(\underline{\theta})$ and $f(\underline{x}|\underline{\theta})$ should be interpreted accordingly. For example, the observation at stage n is denoted by $\underline{x}_n = (x_{n0}, x_{n1}, \ldots, x_{nk})$. Under the loss function (2.3), the Bayes rule $\delta_B(\underline{x})$ is given as follows:

It can be shown that

$$\Delta_{G_0,G_i}(x_0,x_i) = m_i(x_i)(1-M_0(x_0)) + (1+M_i(x_i))m_0(x_0) + \tag{3.13}$$

$$(x_0-x_i-\Delta)m_i(x_i)m_0(x_0)-2\int_{x_0}^{\infty} \frac{M_i(\theta_0)}{\theta_0}\, dG_0(\theta_0)$$

$$= \Delta_{1,G_0,G_i}(x_0,x_i) \text{ (say)}, \quad \text{if } 0 < x_i \leq x_0$$

and

(3.14) $$\Delta_{G_0,G_i}(x_0,x_i) = (1-M_i(x_i))m_0(x_0)+(1+M_0(x_0)-2M_0(x_i))m_i(x_i) + (x_i-x_0-\Delta)m_i(x_i)m_0(x_0)+2M_i(x_i)m_0(x_i)-2\int_{x_i}^{\infty} \frac{M_i(\theta_0)}{\theta_0}\, dG_0(\theta_0)$$

$$= \Delta_{2,G_0,G_i}(x_0,x_i) \quad \text{(say), if } 0 < x_0 < x_i.$$

Thus

(3.15) $$\delta_B(\underline{x}) = \delta_{1,B}(\underline{x}) \cup \delta_{2,B}(\underline{x})$$

where

$$\delta_{1,B}(\underline{x}) = \{i \mid 0 < x_i \le x_0,\ \Delta_{1,G_0,G_i}(x_0,x_i) < 0\}$$

and

$$\delta_{2,B}(\underline{x}) = \{i \mid 0 < x_0 < x_i,\ \Delta_{2,G_0,G_i}(x_0,x_i) < 0\}.$$

Similar to Theorem 3.1, we have the following result.

Theorem 3.2. If $\int_0^{\infty} \theta dG_i(\theta) < \infty$, $i = 0,1,\ldots,k$ and for all $1 \le i \le k$, $\Delta_{1,i,n}(x_0,x_i) \xrightarrow{P} \Delta_{1,G_0,G_i}(x_0,x_i)$ for $x_i \le x_0$ and $\Delta_{2,i,n}(x_0,x_i) \xrightarrow{P} \Delta_{2,G_0,G_i}(x_0,x_i)$ for $x_0 < x_i$, the sequence $\{\delta_n\}$ defined by $\delta_n(\underline{x}) = \{i \mid x_i \le x_0,\ \Delta_{1,i,n}(x_0,x_i) < 0\} \cup \{i \mid x_0 < x_i, \Delta_{2,i,n}(x_0,x_i) < 0\}$, is a.o. relative to G.

Now our problem is to find a consistent estimator of $\int_a^{\infty} \frac{M_i(\theta_0)}{\theta_0}\, dG_0(\theta_0)$ for $x_0 \le a$.

Theorem 3.3. Let $M_{in}(x)$ and $m_{in}(x)$ be defined by (3.7) and (3.10), respectively. Then

$$\int_a^{\infty} \frac{M_{in}(\theta_0)}{\theta_0}\, dG_{0n}(\theta_0) \xrightarrow{P} \int_a^{\infty} \frac{M_i(\theta_0)}{\theta_0}\, dG_0(\theta_0) \quad \text{for } x_0 \le a,$$

where $G_{0n}(\theta_0) = M_{0n}(\theta_0)-\theta_0 m_{0n}(\theta_0)$.

Proof. $\left|\int_a^\infty \frac{M_{in}(\theta_0)}{\theta_0}\, dG_{0n}(\theta_0) - \int_a^\infty \frac{M_i(\theta_0)}{\theta_0}\, dG_{0n}(\theta_0)\right|$

$$\leq \int_a^\infty \frac{|M_{in}(\theta_0)-M_i(\theta_0)|}{\theta_0}\, dG_{0n}(\theta_0)$$

$$\leq \frac{1}{a} \sup_{-\infty<x<\infty} |M_{in}(x)-M_i(x)| \leq \varepsilon$$

with probability near 1, for large n, by Glivenko-Cantelli Theorem. Since $\frac{M_i(\theta_0)}{\theta_0}$ is bounded continuous and $G_{0n}(\theta_0) \xrightarrow{P} G_0(\theta_0)$, we have $\int_a^\infty \frac{M_i(\theta_0)}{\theta_0}\, dG_{0n}(\theta_0) \xrightarrow{P} \int_a^\infty \frac{M_i(\theta_0)}{\theta_0}\, dG_0(\theta_0)$.

Thus

$$\left|\int_a^\infty \frac{M_{in}(\theta_0)}{\theta_0}\, dG_{0n}(\theta_0) - \int_a^\infty \frac{M_i(\theta_0)}{\theta_0}\, dG_0(\theta_0)\right|$$

$$\leq \left|\int_a^\infty \frac{M_{in}(\theta_0)}{\theta_0}\, dG_{0n}(\theta_0) - \int_a^\infty \frac{M_i(\theta_0)}{\theta_0}\, dG_{0n}(\theta_0)\right| +$$

$$\left|\int_a^\infty \frac{M_i(\theta_0)}{\theta_0}\, dG_{0n}(\theta_0) - \int_a^\infty \frac{M_i(\theta_0)}{\theta_0}\, dG_0(\theta_0)\right|$$

$\leq \varepsilon$ with probability near 1, for large n.

From Theorem 3.3, if we define

$$\Delta_{1,i,n}(x_0,x_i) = m_{in}(x_i)(1-M_{0n}(x_0)) + m_{0n}(x_0)(1+M_{in}(x_i))$$
$$+ (x_0-x_i-\Delta)m_{in}(x_i)m_{0n}(x_0) - 2\int_{x_0}^\infty \frac{M_{in}(\theta_0)}{\theta_0}\, dG_{0n}(\theta_0),$$

and

$$\Delta_{2,i,n}(x_0,x_i) = m_{0n}(x_0)(1-M_{in}(x_i)) + m_{in}(x_i)(1+M_{0n}(x_0)-2M_{0n}(x_i))$$
$$+ (x_i-x_0-\Delta)m_{0n}(x_0)m_{in}(x_i) + 2M_{in}(x_i)m_{0n}(x_i) - 2\int_{x_i}^\infty \frac{M_{in}(\theta_0)}{\theta_0}\, dG_{0n}(\theta_0),$$

then we have

$$\Delta_{\ell,i,n}(x_0,x_i) \xrightarrow{P} \Delta_{\ell,G_0,G_i}(x_0,x_i),\ \ell = 1,2.$$

Therefore, the sequence of rules $\{\delta_n\}$ defined by

$$\delta_n(\underline{x}) = \{i \mid x_i \le x_0, \Delta_{1,i,n}(x_0,x_i) < 0\} \cup \{i \mid x_0 < x_i, \Delta_{2,i,n}(x_0,x_i) < 0\},$$

is a.o. relative to G by Theorem 3.2.

3.3. Rate of Convergence of the Empirical Bayes Rules

In this section we will consider the rate of convergence of the empirical Bayes rule derived in Section 3.1.

Definition 3.2. The sequence of procedures $\{\delta_n\}$ is said to be asymptotically optimal of order α_n relative to G if $r_n(G,\delta_n)-r(G) = O(\alpha_n)$ as $n \to \infty$, where $\lim_{n\to\infty} \alpha_n = 0$.

The main result (Theorem 3.8) of this section is based on a series of lemmas.

Lemma 3.4. Let $\Delta_{1,G_i}(x_i)$, $\Delta_{2,G_i}(x_i)$, $\Delta_{1,i,n}(x_i)$ and $\Delta_{2,i,n}(x_i)$ be defined by (3.2), (3.3), (3.11) and (3.12), respectively. Then we have

$$\begin{aligned}0 &\le r_n(G,\delta_n)-r(G)\\ &\le \sum_{i=1}^{k} \int_0^{\theta_0} |\Delta_{1,G_i}(x_i)|^{1-\delta} E|\Delta_{1,i,n}(x_i)-\Delta_{1,G_i}(x_i)|^{\delta} dx_i +\\ &\quad \sum_{i=1}^{k} \int_{\theta_0}^{\theta_0+\Delta} |\Delta_{2,G_i}(x_i)|^{1-\delta} E|\Delta_{2,i,n}(x_i)-\Delta_{2,G_i}(x_i)|^{\delta} dx_i, \ \delta > 0.\end{aligned}$$

Proof. The proof is similar to that of Lemma 3 of Van Ryzin and Susarla (1977) and hence omitted.

Lemma 3.5. Assume that $h(n)$ satisfies the condition (3.9) and that $\varphi(x)$ satisfies the following;

(i) $\varphi(x) = 0$ if $x \notin (0,a)$ for some finite $a > 0$

(ii) $\int_0^a \varphi(x)dx = 1$

(iii) $\sup_x |\varphi(x)| < \infty$.

Then, for $m_{in}(x_i)$ defined by (3.10), we have

$$|E\, m_{in}(x_i) - m_i(x_i)| \leq h(n) f_\epsilon(x_i) \int_0^a |u\, \varphi(u)| du$$

for large n where $f_\epsilon(x_i) = \sup_{0 \leq y \leq \epsilon} |m_i'(x_i+y)|$, $\epsilon > 0$.

Proof. $Em_{in}(x_i) - m_i(x_i)$

$$= \frac{1}{h(n)} \int \varphi\left(\frac{y-x_i}{h(n)}\right) m_i(y)dy - m_i(x_i)$$

$$= \int_0^a \varphi(u)[m_i(x_i+uh(n)) - m_i(x_i)]du$$

$$= \int_0^a \varphi(u)[uh(n)m_i'(x_i+\eta_n(x_i,u))]du$$

where $0 < \eta_n(x_i,u) < uh(n)$.

For $\varepsilon > 0$, let n be large enough so that $ah(n) \leq \varepsilon$, then

$$|Em_{in}(x_i) - m_i(x_i)| \leq h(n) f_\varepsilon(x_i) \int_0^a |u\, \varphi(u)| du.$$

Lemma 3.6. Under the assumptions of Lemma 3.5, we have

$$\mathrm{Var}\, m_{in}(x_i) \leq \frac{1}{nh(n)} m_i(x_i) \int_0^a \varphi^2(u)du.$$

Proof. $\mathrm{Var}\, m_{in}(x_i) = \mathrm{Var}\left\{\frac{1}{nh(n)} \sum_{j=1}^n \varphi\left(\frac{X_{ji}-x_i}{h(n)}\right)\right\}$

$$\leq \frac{1}{nh(n)} \int_0^a \varphi^2(u) m_i(x_i+uh(n))du$$

$$\leq \frac{1}{nh(n)} m_i(x_i)\int_0^a \varphi^2(u)du, \text{ since } m_i(x_i) \text{ ~~is non-increasing.~~}$$

Remark: From Lemma 3.5 and Lemma 3.6, we have

$$m_{in}(x_i) \xrightarrow{P} m_i(x_i) \quad \text{if} \quad f_\varepsilon(x_i) < \infty.$$

Lemma 3.7. Under the assumptions of Lemma 3.5, we have

(a) $\operatorname{Var} \Delta_{1,i,n}(x_i) \leq M(\theta_0-\Delta-x_i)^2 m_i(x_i)(nh(n))^{-1} + \frac{5}{n}$,

(b) $\operatorname{Var} \Delta_{2,i,n}(x_i) \leq M(x_i-\theta_0-\Delta)^2 m_i(x_i)(nh(n))^{-1} + \frac{1}{2n}$,

where $M = 2\int_0^a \varphi^2(u)du$.

Proof. Since $\operatorname{Var}(X+Y) \leq 2\{\operatorname{Var}(X) + \operatorname{Var}(Y)\}$ for any random variables X and Y, it follows from Lemma 3.6 that

$$\operatorname{Var} \Delta_{1,i,n}(x_i) \leq 2\{(\theta_0-\Delta-x_i)^2 \operatorname{Var} m_{in}(x_i) + \operatorname{Var}(M_{in}(x_i)-2M_{in}(\theta_0))\}$$

$$\leq 2\{(\theta_0-\Delta-x_i)^2 m_i(x_i)(nh(n))^{-1}\int_0^a \varphi^2(u)du + \frac{5}{2n}\},$$

which proves (a). Similarly we have the result (b).

Theorem 3.8. Assume the conditions of Lemma 3.5 and the following for $0 < \delta < 2$;

(i) $\int_0^{\theta_0} |\Delta_{1,G_i}(x_i)|^{1-\delta}dx_i < \infty$ and $\int_0^{\theta_0} |\Delta_{1,G_i}(x_i)|^{1-\delta}|\theta_0-\Delta-x_i|^\delta m_i^{\delta/2}(x_i)dx_i < \infty$,

(ii) $\int_{\theta_0}^{\theta_0+\Delta} |\Delta_{2,G_i}(x_i)|^{1-\delta}dx_i < \infty$ and $\int_{\theta_0}^{\theta_0+\Delta} |\Delta_{2,G_i}(x_i)|^{1-\delta}|x_i-\theta_0-\Delta|^\delta m_i^{\delta/2}(x_i)dx_i < \infty$,

(iii) $\int_0^{\theta_0} |\Delta_{1,G_i}(x_i)|^{1-\delta}|\theta_0-\Delta-x_i|^\delta f_\varepsilon^\delta(x_i)dx_i < \infty$

(iv) $\int_{\theta_0}^{\theta_0+\Delta} |\Delta_{2,G_i}(x_i)|^{1-\delta} |x_i-\theta_0-\Delta|^{\delta} f_{\varepsilon}^{\delta}(x_i)dx_i < \infty.$

Then we have

$$r_n(G,\delta_n)-r(G) = O(\max\{(\frac{1}{nh(n)})^{\delta/2}, (h(n))^{\delta}\}) \text{ as } n \to \infty.$$

Proof. For $0 < \delta < 2$, by Hölder inequality and Lemma 3.4, we have

$$0 \le r_n(G,\delta_n)-r(G)$$

$$\le \sum_{i=1}^{k} \{\max(1,2^{\delta-1})[\int_0^{\theta_0} |\Delta_{1,G_i}(x_i)|^{1-\delta}(\text{Var } \Delta_{1,i,n}(x_i))^{\delta/2}dx_i +$$

$$\int_0^{\theta_0} |\Delta_{1,G_i}(x_i)|^{1-\delta} |(\theta_0-\Delta-x_i)(Em_{in}(x_i)-m_i(x_i))|^{\delta}dx_i]\} +$$

$$\sum_{i=1}^{k} \{\max(1,2^{\delta-1})[\int_{\theta_0}^{\theta_0+\Delta} |\Delta_{2,G_i}(x_i)|^{1-\delta}(\text{Var } \Delta_{2,i,n}(x_i))^{\delta/2}dx_i +$$

$$\int_{\theta_0}^{\theta_0+\Delta} |\Delta_{2,G_i}(x_i)|^{1-\delta} |(x_i-\theta_0-\Delta)(Em_{in}(x_i)-m_i(x_i))|^{\delta}dx_i]\}.$$

Since $(a+b)^{\delta/2} \le a^{\delta/2} + b^{\delta/2}$ for $a > 0$, $b > 0$ and $0 < \delta < 2$, it follows from Lemma 3.7, (i) and (ii) that

$$\int_0^{\theta_0} |\Delta_{1,G_i}(x_i)|^{1-\delta}(\text{Var } \Delta_{1,i,n}(x_i))^{\delta/2}dx_i = O((nh(n))^{-\delta/2})$$

and

$$\int_{\theta_0}^{\theta_0+\Delta} |\Delta_{2,G_i}(x_i)|^{1-\delta}(\text{Var } \Delta_{2,i,n}(x_i))^{\delta/2}dx_i = O((nh(n))^{-\delta/2}).$$

By Lemma 3.5 (iii) and (iv),

$$\int_0^{\theta_0} |\Delta_{1,G_i}(x_i)|^{1-\delta} |\theta_0-\Delta-x_i|^{\delta} |E\, m_{in}(x_i)-m_i(x_i)|^{\delta}dx_i = O((h(n))^{\delta})$$

and

$$\int_{\theta_0}^{\theta_0+\Delta} |\Delta_{2,G_i}(x_i)|^{1-\delta}|(x_i-\theta_0-\Delta)(E\, m_{in}(x_i)-m_i(x_i))|^{\delta} dx_i = O((h(n))^{\delta}).$$

Hence

$$r_n(G,\delta_n)-r(G) = O(\max\{(nh(n))^{-\delta/2}, (h(n))^{\delta}\}) \text{ as } n \to \infty.$$

Corollary 3.9. Assume the conditions of Theorem 3.8. If we take $h(n) = n^{-\alpha}$, $0 < \alpha < 1$, then the optimal choice of α is 1/3 and $r_n(G,\delta_n)-r(G) = O(n^{-\delta/3})$ as $n \to \infty$.

Remark: If the prior distribution G_i is such that both $g_i(x)/x$ and $m_i(x)$ are bounded on $(0, \theta_0+\Delta+\varepsilon)$, it is easy to check that the conditions of Theorem 3.8 are satisfied for $0 < \delta \leq 1$.

Acknowledgment: The authors wish to thank Dr. Woo-Chul Kim of Seoul National University for a critical reading of this paper and for several helpful suggestions.

REFERENCES

1. Deely, J. J. (1965). Multiple decision procedures from an empirical Bayes approach. Ph.D. Thesis (Mimeo. Ser. No. 45), Dept. of Statist., Purdue University.

2. Gupta, S. S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. Journal of Statistical Planning and Inference 8, 87-101.

3. Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065-1076.

4. Robbins, H. (1956). An empirical Bayes approach to statistics. Proc. 3rd Berkeley Symp. Vol. I (Ed. Jerzy Neyman) Univ. of California Press, Berkeley, 157-163.

5. Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist. 35, 1-20.

6. Singh, A. K. (1977). On slippage tests and multiple decision (selection and ranking) procedures. Ph.D. Thesis (Mimeo. Ser. No. 494), Dept. of Statist., Purdue University.

7. Van Ryzin, J. (1970). On some nonparametric empirical Bayes multiple decision problems. Proc. First Internat. Symp. Nonparametric Techniques in Statistic Inference (Ed. M. L. Puri), 585-603.

8. Van Ryzin, J. and Susarla, V. (1977). On the empirical Bayes approach to multiple decision problems. Ann. Statist. 5, 172-181.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report #83-37	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ON BAYES AND EMPIRICAL BAYES RULES FOR SELECTING GOOD POPULATIONS		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Technical Report #83-37
7. AUTHOR(s) Shanti S. Gupta and Lii-Yuh Leu		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. REPORT DATE September 1983 (Revised 6-85)
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

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)

Bayes rules, empirical Bayes rules, selection procedures, asymptotically optimal, rate of convergence.

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

This paper deals with the problem of selecting all populations which are close to a control or standard. A general Bayes rule for the above problem is derived. Empirical Bayes rules are derived when the populations are assumed to be uniformly distributed. Under some conditions on the marginal and prior distributions, the rate of convergence of the empirical Bayes risk to the minimum Bayes risk is investigated. The rate of convergence is shown to be $n^{-\delta/3}$ for some δ, $0 < \delta < 2$.