ON BAYES AND EMPIRICAL BAYES RULES FOR SELECTING GOOD POPULATIONS(?) PURDUE UNIV LAFAYETTE IN DEPT OF STATISTICS S S GUPTA ET AL. SEP 83 TR-83-37
UNCLASSIFIED N00014-75-C-0455
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

*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

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 \( \delta, 0 < \delta < 2 \).

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

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 $\delta$, $0 < \delta < 2$.

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

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 (1981). 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 (1981) 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 this kind of 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 $\delta$, $0 < \delta < 2$.

*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 $\pi_0$ is the control population with parameter $\theta_0$ which may be known or unknown. When $\theta_0$ is unknown, let $\bar{\theta} = (\theta_0, \theta_1, \ldots, \theta_k)$ and $\bar{X} = (X_0, X_1, \ldots, X_k)$ where $X_i$ is an observation from $\pi_i$, $i = 0, 1, \ldots, k$. When $\theta_0$ is known, no observation from population $\pi_0$ is taken, and $\theta_0, X_0$ are deleted from $\bar{\theta}$ and $\bar{X}$, respectively. When there is no confusion, $\bar{\theta}$ and $\bar{X}$ are used to represent either case. We define population $\pi_i$ 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 $\theta_i, X_i$ has probability density function $f(x_i | \theta_i)$ with respect to a $\sigma$-finite measure $\mu$, for $i = 0, 1, \ldots, k$, and $\bar{\theta}$ has a prior distribution

$G(\bar{\theta}) = \prod_{i=0}^{k} G_i(\theta_i)$ on the parameter space $\Omega$. Let $G = \{ s | s \subseteq \{1, 2, \ldots, k\} \}$ be the action space and let

$$(2.1) \quad L(\bar{\theta}, s) = \sum_{i \in s} \left[ c_1(\theta_0 - \Delta - \theta_i) I(\theta_i < \theta_0 - \Delta)(\theta_i) + c_2(\theta_i - \theta_0 - \Delta) I(\theta_0 + \Delta < \theta_i)(\theta_i) \right] + \sum_{i \notin s} \left[ c_3(\theta_i - \theta_0 + \Delta) I(\theta_0 - \Delta < \theta_i < \theta_0)(\theta_i) + c_4(\theta_0 + \Delta - \theta_i) I(\theta_0 < \theta_i < \theta_0 + \Delta)(\theta_i) \right]$$

be the loss function defined on $\Omega \times G$, where $c_i, i = 1, 2, 3, 4$ are positive constants and $I$ is the indicator function. The Bayes risk with respect to $G$ can be expressed as

$$(2.2) \quad r(G, s) = \int_{\Omega} L(\bar{\theta}, s) f(x | \bar{\theta}) dG(\bar{\theta}) du(x),$$

where $\Omega$ is the sample space and $f(x | \bar{\theta}) = \pi f(x_i | \theta_i).$
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 \( \delta: \mathcal{X} \rightarrow \mathcal{C} \), the corresponding Bayes risk with respect to \( G \) is given by

\[
(2.3) \quad r(G, \delta) = \int_{\mathcal{X}} \int_{\mathcal{E}} L(\theta, \delta(x)) f(x|\theta) dG(\theta) d\mu(x).
\]

In the sequel we consider the special case where \( c_1 = c_2 = c_3 = c_4 = \text{constant} \) which can be taken to be unity without loss of generality. If \( \emptyset \) is the empty set, we have

\[
L(\theta, \emptyset) = \sum_{i=1}^{k} \left( (\theta_i - \theta_0 + \Delta) I_{\{\theta_0 - \Delta < \theta_i < \theta_0\}}(\theta_i) + (\theta_0 + \Delta - \theta_i) I_{\{\theta_0 < \theta_i < \theta_0 + \Delta\}}(\theta_i) \right),
\]

and (2.1) can be expressed as

\[
(2.4) \quad L(\theta, s) = L(\theta, \emptyset) + \sum_{i \in s} \left( (\theta_0 - \Delta - \theta_i) I_{\{\theta_i < \theta_0\}}(\theta_i) + (\theta_i - \theta_0 - \Delta) I_{\{\theta_0 < \theta_i \}}(\theta_i) \right).
\]

Hence, for any \( \delta \), we have

\[
(2.5) \quad r(G, \delta) - r(G, \emptyset) = \int \sum_{\mathcal{X}} \left( \int_{\mathcal{E}} L(\theta, \delta(x)) f(x|\theta) dG(\theta) + 2 \int \int_{\{\theta_0 < \theta_i \}}(\theta_i - \theta_0) f(x|\theta) dG(\theta) d\mu(x) \right).
\]

From (2.5), \( \delta_B(x) \) is given by \( i \in \delta_B(x) \) if

\[
(2.6) \quad \int_{\{\theta_0 - \Delta - \theta_i \}} f(x|\theta) dG(\theta) + 2 \int \int_{\{\theta_0 < \theta_i \}}(\theta_i - \theta_0) f(x|\theta) dG(\theta) < 0,
\]

then \( \delta_B(x) \) is a Bayes rule with respect to \( G \).
Let \( m_i(x_i) = \int f(x_i | \theta_i) dG_i(\theta_i) \) be the marginal distribution of \( X_i \), 
\( \pi(\theta_i | x_i) \) be the posterior distribution of \( \theta_i \) given \( X_i = x_i \), and \( E(\theta_i | x_i) \) 
be the expected value of \( \theta_i \) given \( X_i = x_i \). If \( m_i(x_i) > 0 \) for all \( x_i \), then 
(2.6) is equivalent to 

(2.7) \( (\theta_i - \theta) - E(\theta_i | x_i) + 2/\pi(\theta_i | x_i) \pi(\theta_i | x_i) d\theta_i < 0 \)

if \( \theta_i \) is known, or 

(2.8) \( E(\theta_i | x_i) - E(\theta_i | x_i) - (\theta_i - \theta) + 2/\pi(\theta_i | x_i) \pi(\theta_i | x_i) d\theta_i < 0 \)

if \( \theta_i \) is unknown.

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

**Theorem 2.1.** Under the loss function (2.4), the Bayes rule \( \delta_B(x) \) with 
respect to \( G \) is given by 

(a) If \( \theta_0 \) is known, then \( i \in \delta_B(x) \) if the inequality (2.7) holds. 

(b) If \( \theta_0 \) is unknown, then \( i \in \delta_B(x) \) if the inequality (2.8) holds.

**An Application:**

Suppose that 

(2.9) \( f(x_i | \theta_i) = e^{\theta_i x_i} / (x_i!) \), \( x_i = 0, 1, \ldots, \theta_i > 0 \) 

and \( \theta_i \) has a prior distribution \( g_i(\theta_i) = G_i^i(\theta_i) \) which is given by 

(2.10) \( g_i(\theta_i) = \beta_i \alpha_i^{\theta_i - 1} e^{\theta_i / \Gamma(\alpha_i)} I_0(\alpha_i, \theta_i) \), 

where \( \alpha_i > 0 \) and \( \beta_i > 0 \) are known. Then 

(2.11) \( \pi(\theta_i | x_i) = (1+\beta_i)^{-1} \theta_i^{\alpha_i + x_i} e^{-\theta_i (1+\beta_i)} / \Gamma(\alpha_i + x_i) \) 

and 

(2.12) \( E(\theta_i | x_i) = (x_i + \alpha_i) / (1+\beta_i) \).
Lemma 2.2. If \( \pi(x_i | x_i) \) is defined by (2.11), then

\[
\int_{\theta_i \leq \theta_i} (\theta_i - \theta_0) \pi(x_i | x_i) d\theta_i
\]

\[
= \frac{x_i + \alpha_i}{1+\beta_i} (1 - \Gamma(\theta_0(1+\beta_i) + x_i + \alpha_i + 1) - \theta_i (1 - \Gamma(\theta_0(1+\beta_i) + x_i + \alpha_i)),
\]

where \( \theta_0 > 0 \) is known and

\[
\Gamma(a; \beta) = \int_0^\infty \frac{x^{a-1} e^{-x} dx}{\Gamma(a)}, \quad a > 0, \ \beta > 0.
\]

Proof. Proof is simple and hence omitted.

Lemma 2.3. If \( \pi(x_i | x_i) \) is defined by (2.11) and \( \theta_i = \beta, \ i = 0, 1, \ldots, k \) and \( \theta_0 \) is unknown, then

\[
\int_{\theta_0 \leq \theta_i} (\theta_i - \theta_0) \pi(x_i | x_i) \pi(\theta_0 | x_0) d\theta_i d\theta_0
\]

\[
= \frac{(x_i + \alpha_i + \theta_0)}{1+\beta_i} I(\frac{1}{2}; x_0 + \alpha_0, x_i + \alpha_i) - \frac{2(x_0 + \alpha_0)}{1+\beta_i} I(\frac{1}{2}; x_0 + \alpha_0 + 1, x_i + \alpha_i),
\]

where

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

and

\[
B(a, \beta) = \Gamma(a) \Gamma(\beta)/\Gamma(a+\beta).
\]

Proof. \( \int_{\theta_0 \leq \theta_i} (\theta_i - \theta_0) \pi(x_i | x_i) \pi(\theta_0 | x_0) d\theta_i d\theta_0 \)

\[
= \int_0^\infty \frac{(1+\beta)}{\Gamma(x_i + \alpha_i) \Gamma(x_0 + \alpha_0)} \theta_0 \theta_0 e^{-(\theta_0 + \theta_i) \Gamma(1+\beta)} d\theta_i d\theta_0
\]

\[
= \int_0^\infty \frac{(1+\beta)}{\Gamma(x_i + \alpha_i) \Gamma(x_0 + \alpha_0)} \theta_0 \theta_0 e^{-(\theta_0 + \theta_i) \Gamma(1+\beta)} d\theta_i d\theta_0
\]

\[
= \int_0^\infty \frac{(1+\beta)}{\Gamma(x_i + \alpha_i) \Gamma(x_0 + \alpha_0)} \theta_0 \theta_0 e^{-(\theta_0 + \theta_i) \Gamma(1+\beta)} d\theta_i d\theta_0
\]
\[ \int (x_{i}^{\alpha_{i}} + x_{0}^{\alpha_{0}}) f_{\alpha_{i}}(x_{i}) \text{d}x_{i} = \int (1 + \beta) B(x_{0}^{\alpha_{0}}, x_{i}^{\alpha_{i}}) \text{d}x_{i} \]

\[ = \frac{(x_{i}^{\alpha_{i}} + x_{0}^{\alpha_{0}})}{(1 + \beta) B(x_{0}^{\alpha_{0}}, x_{i}^{\alpha_{i}})} \int (1-w)v (1+w) \text{d}v \]

\[ = \frac{(x_{i}^{\alpha_{i}} + x_{0}^{\alpha_{0}})}{(1 + \beta) B(x_{0}^{\alpha_{0}}, x_{i}^{\alpha_{i}})} \int (1-2s)(1-s) s^{\alpha_{i}-1} x_{i}^{\alpha_{i}-1} x_{0}^{\alpha_{0}-1} \text{d}s \]

\[ = \frac{(x_{i}^{\alpha_{i}} + x_{0}^{\alpha_{0}})}{1 + \beta} I\left(\frac{1}{2}; x_{0}^{\alpha_{0}}, x_{i}^{\alpha_{i}}\right) - \frac{2(x_{0}^{\alpha_{0}})}{1 + \beta} I\left(\frac{1}{2}; x_{0}^{\alpha_{0}+1}, x_{i}^{\alpha_{i}}\right). \]

From Theorem 2.1, Lemma 2.2 and Lemma 2.3, we have the following theorem:

**Theorem 2.4.** If \( f(x_{i}|\theta_{i}) \) is defined by (2.9) and \( g_{i}(\theta_{i}) \) is defined by (2.10). Under the loss function (2.4), the Bayes rule \( \delta_{B}(x) \) is given by

(a) If \( \theta_{0} \) is known, then \( i \in \delta_{B}(x) \) if

\[ \frac{x_{i}^{\alpha_{i}}}{1 + \beta_{i}} (1-2\Gamma(\theta_{0}(1+\beta_{i}); x_{i}^{\alpha_{i}+1})) - \theta_{0}(1-2\Gamma(\theta_{0}(1+\beta_{i}); x_{i}^{\alpha_{i}})) < \Delta. \]

(b) If \( \theta_{0} \) is unknown and \( \beta_{i} = \beta, i = 0,1,...,k \), then \( i \in \delta_{B}(x) \) if

\[ \frac{x_{i}^{\alpha_{i}}}{1 + \beta_{i}} (2\Gamma\left(\frac{1}{2}; x_{0}^{\alpha_{0}}, x_{i}^{\alpha_{i}}\right) - 1) + \frac{x_{0}^{\alpha_{0}}}{1 + \beta_{i}} (1 + 2\Gamma\left(\frac{1}{2}; x_{0}^{\alpha_{0}+1}, x_{i}^{\alpha_{i}}\right)) < \Delta. \]

3. **Empirical Bayes Rules for Uniform Populations**

In this section we will assume that \( X_{i} \) has probability density function

\[ f(x_{i}|\theta_{i}) = \frac{1}{\theta_{i}} I(0,\theta_{i})(x_{i}), \text{ where } \theta_{i} > 0 \text{ is unknown. Suppose that } \theta \text{ has a prior distribution } G(\theta) = \pi G_{i}(\theta_{i}) \text{ on } \Omega \text{ and } G_{i} \text{ has a continuous probability distribution on } \theta_{i}. \]
density function \( g_i \) and \( g_i \) is positive. Let \( m_i(x_i) \), \( M_i(x_i) \) be the marginal pdf and cdf of \( X_i \), respectively. Then

\[
(3.1) \quad m_i(x_i) = \int_{x_i}^{\infty} \frac{1}{\theta_i} \, dG_i(\theta_i)
\]

and

\[
(3.2) \quad M_i(x_i) = x_i m_i(x_i) + G_i(x_i).
\]

From (3.2), we have

\[
(3.3) \quad G_i(x_i) = M_i(x_i) - x_i m_i(x_i).
\]

It follows that

\[
(3.4) \quad \int_a^b \frac{1}{\theta_i} \, dG_i(\theta_i) = m_i(a) - m_i(b)
\]

and

\[
\int_a^\infty \frac{1}{\theta_i} \, dG_i(\theta_i) = m_i(a)
\]

for any \( x_i < a < b < \infty \).

3.1. \( \theta_0 \) known

In the case where \( \theta_0 \) is known, let

\[
(3.5) \quad \Delta G_i(x_i) = (\theta_0 - \Delta) m_i(x_i) - \int_{x_i}^{\infty} \frac{1}{\theta_i} \, dG_i(\theta_i) + 2 \int_{x_i}^{\infty} (\theta_i - \theta_0) f(x_i | \theta_i) \, dG_i(\theta_i)
\]

From (2.6), we have \( i \in \delta_g(x) \) if \( \Delta G_i(x_i) < 0 \). If \( x_i \leq \theta_0 \),

\[
(3.6) \quad \Delta G_i(x_i) = (\theta_0 - \Delta) m_i(x_i) - \int_{x_i}^{\infty} \frac{1}{\theta_i} \, dG_i(\theta_i) + 2 \int_{x_i}^{\infty} (\theta_i - \theta_0) \frac{1}{\theta_i} \, dG_i(\theta_i)
\]

\[
= (\theta_0 - \Delta) m_i(x_i) + 1 - 2M_i(\theta_0) + M_i(x_i)
\]

\[
= \Delta_{1, G_i}(x_i) \quad \text{(say)}.
\]

If \( x_i > \theta_0 \),
\[ (3.7) \quad \Delta_{G_1}(x_i) = (\theta_0 - \Delta) m(x_i) + \int_{x_i}^{\infty} dG_i(\theta_i) + 2 \int_{x_i}^{\infty} (\theta_i - \theta_0) \frac{1}{\theta_i} dG_i(\theta_i) \]
\[ = (x_i - \theta_0 + \Delta) m_i(x_i) + 1 - M_i(x_i) \]
\[ = \Delta_{Z_i, G_1}(x_i) \text{ (say)}. \]

Therefore
\[ (3.8) \quad \delta_B(x) = \{ i | x_i \leq \theta_0, \Delta_1, G_i(x_i) < 0 \} \cup \{ i | x_i > \theta_0, \Delta_2, G_i(x_i) < 0 \}. \]

Remarks:
1. \( \Delta_{G_1}(x_i) \) 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 \( \theta_0 + \Delta < x_i \) (we assume that \( \theta_0 - \Delta > 0 \)).
2. If \( x_i \geq \theta_0 + \Delta \), then \( \Delta_{G_1}(x_i) \geq 0 \). Hence \( i \notin \delta_B(x) \) if \( x_i \geq \theta_0 + \Delta \).
3. If \( G_i \) is such that \( 1 - 2M_i(\theta_0) + M_i(\theta_0 - \Delta) \geq 0 \), then \( \delta_B(x) = \emptyset \).
   Otherwise, \( i \in \delta_B(x) \) if \( (\theta_0 - \Delta) - d_1 < x_i < (\theta_0 + \Delta) - d_2 \), for some positive real numbers \( d_1 \) and \( d_2 \). Hence this type of selection rules are Bayes rules relative to some prior distribution.

If \( G \) is unknown, the Bayes rules are not obtainable. In this case, we consider a sequence \((X_1, G_1), (X_2, G_2), \ldots \), which are independent pairs of random vectors, each \( G_i \) is distributed as \( G \) on \( \Theta \) and \( X_i = (X_{i1}, \ldots, X_{ik}) \) has conditional density function \( f(x_i, g) \) given \( G_i = g \). The empirical Bayes approach, which was introduced by Robbins (1955), attempts to construct a decision rule concerning \( \Delta_{n+1} \) at stage \( n+1 \) based on \( X_1, \ldots, X_n \). The risk at stage \( n+1 \) taking action \( \delta_n(x; X_1, \ldots, X_n) = \delta_n(x) \) is given by
\[ (3.9) \quad r_n(G, \delta_n) = \int_{\Theta} \left[ \sum_{i \in \delta_n(x)} \int_{\Theta} \left[ (\theta_0 - \theta_i - \theta_0) f(x_i, g) dG(g) \right] dx \right] + r(G, \delta) \]
where $E_n$ denotes the expectation with respect to the $n$ independent random variables $X_1,\ldots,X_n$ each with common density function

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

**Definition 3.1.** 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. sequences of rules, let

$$\delta_{1,B}(x) = \{i|\theta < \theta_0 + \Delta_1, G_i(x_i) < 0\}$$

and

$$\delta_{2,B}(x) = \{i|\theta_0 < x_i < \theta_0 + \Delta_2, G_i(x_i) < 0\}.$$  

From (3.8) and Remark (2), we have

$$\delta_{B}(x) = \delta_{1,B}(x) \cup \delta_{2,B}(x).$$

For any $i = 1,2,\ldots,k$ and $\epsilon = 1,2$, let

$$\Delta_{\epsilon,i,n}(x_i) = \Delta_{\epsilon,i,n}(x_i; x_1,\ldots,x_n),$$

$n = 1,2,\ldots$ be two sequences of real-valued measurable functions. We define

$$\delta_{n}(x) = \delta_{1,n}(x) \cup \delta_{2,n}(x),$$

where

$$\delta_{1,n}(x) = \{i|\theta_0 < \theta_0 + \Delta_1, G_i(x_i) < 0\}$$

and

$$\delta_{2,n}(x) = \{i|\theta_0 < x_i < \theta_0 + \Delta_2, G_i(x_i) < 0\}.$$  

We have the following theorem:

**Theorem 3.1.** If $\int_0^\infty dG_i(\theta) < \infty$, $i = 1,2,\ldots,k$ and $\Delta_{1,i,n}(x_i) \Rightarrow \Delta_{1,G_i}(x_i)$,

for almost all $x_i \leq \theta_0$ and $\Delta_{2,i,n}(x_i) \Rightarrow \Delta_{2,G_i}(x_i)$, for almost all

$\theta_0 < x_i < \theta_0 + \Delta$, where $\Rightarrow$ means convergence in probability. Then $\{\delta_n(x)\}$

defined by (3.10) is a.o. relative to $G$. 


Proof. \( 0 \leq \int_{\Omega} (g_0(x) - \delta(x)) \, dG(\cdot) - \int_{\Omega} (g_0(x) - \delta_B(x)) \, dG(\cdot) \)

\[
(3.11) = \left( \sum_{i \in \delta_1, n(x)} \Delta_{1,i,n}(x_i) \prod_{j=1}^{k} m_j(x_j) - \sum_{i \in \delta_1, B(x)} \Delta_{1,i,n}(x_i) \prod_{j=1}^{k} m_j(x_j) \right) + \left( \sum_{i \in \delta_2, n(x)} \Delta_{2,i,n}(x_i) \prod_{j=1}^{k} m_j(x_j) - \sum_{i \in \delta_2, B(x)} \Delta_{2,i,n}(x_i) \prod_{j=1}^{k} m_j(x_j) \right)
\]

The first term of (3.11) can be expressed as

\[
(3.12) = \left( \sum_{i \in \delta_1, n(x)} \Delta_{1,i,n}(x_i) \prod_{j=1}^{k} m_j(x_j) - \sum_{i \in \delta_1, B(x)} \Delta_{1,i,n}(x_i) \prod_{j=1}^{k} m_j(x_j) \right) + \left( \sum_{i \in \delta_1, B(x)} \Delta_{1,i,n}(x_i) \prod_{j=1}^{k} m_j(x_j) - \sum_{i \in \delta_1, B(x)} \Delta_{1,i,n}(x_i) \prod_{j=1}^{k} m_j(x_j) \right)
\]

Since by the definition of \( \delta_1, n(x) \), the second sum of (3.12) is less than or equal to zero.

The second term of (3.11) has a similar result.

Hence, if \( \Delta_{1,i,n}(x_i) \geq \Delta_{1,1,n}(x_i), i = 1, 2 \), then...
\[ 0 \leq \int_{\Omega} \left( L(\theta, \delta_n(x)) f(x|\theta) dG(\theta) - \int_{\Omega} L(\theta, \delta_B(x)) f(x|\theta) dG(\theta) \right) \]

\[ \leq 2 \sum_{i=1}^{k} |\Delta_{1,G_i}(x_i) - \Delta_{1,i,n(x_i)}| \prod_{j=1}^{k} m_j(x_j) + 2 \sum_{i=1}^{k} |\Delta_{2,G_i}(x_i) - \Delta_{2,i,n(x_i)}| \prod_{j=1}^{k} m_j(x_j) \]

\[ \leq 4 \varepsilon \sum_{i=1}^{k} \left( \prod_{j=1}^{k} m_j(x_j) \right) \]

with probability near 1, for large n. Hence

\[ \int_{\Omega} L(\theta, \delta_n(x)) f(x|\theta) dG(\theta) \leq \int_{\Omega} L(\theta, \delta_B(x)) f(x|\theta) dG(\theta) \]

for almost all \( x \).

By Corollary 1 of Robbins (1964), \( \{\delta_n(x)\} \) 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

\( (3.13) \quad M_{1n}(x_i) = \frac{1}{n} \sum_{j=1}^{n} I(-\infty, x_i)(x_j) \),

then \( M_{1n}(x_i) \overset{P}{\rightarrow} M_{1}(x_i) \) for all \( x_i > 0 \). Next, let \( \varphi(x) \geq 0 \) be a Borel function satisfying the following conditions:

\( (3.14) \quad (i) \sup_{-\infty < x < \infty} \varphi(x) < \infty, (ii) \int_{-\infty}^{\infty} \varphi(x) dx = 1, \) and \( (iii) \lim_{x \to \infty} x \varphi(x) = 0 \)

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

\( (3.15) \quad (i) \ h(n) \to 0 \) as \( n \to \infty \) and \( (ii) nh(n) \to \infty \) as \( n \to \infty \).

We define

\( (3.16) \quad m_{1n}(x) = \frac{1}{nh(n)} \sum_{j=1}^{n} \varphi(\frac{x-x_j}{h(n)}) \),
then \( m_{in}(x) \stackrel{P}{\rightarrow} m_i(x) \) for all \( x \) (see Parzen (1962)). For \( i = 1,2,\ldots,k \), let

\[
\Delta_{1,i,n}(x_i) = (e_0-\Delta-x_i)m_{in}(x_i)+1-2m_{in}(e_0)+M_{in}(x_i)
\]

and

\[
\Delta_{2,i,n}(x_i) = (x_i-e_0-\Delta)m_{in}(x_i)+1-M_{in}(x_i).
\]

Then

\[
\Delta_{1,i,n}(x_i) \stackrel{P}{\rightarrow} \Delta_{1,i,G_i}(x_i) \text{ for all } x_i < e_0
\]

and

\[
\Delta_{2,i,n}(x_i) \stackrel{P}{\rightarrow} \Delta_{2,i,G_i}(x_i) \text{ for all } e_0 < x_i < e_0+\Delta.
\]

Finally, we define

\[
\delta_n(x) = \{i| x_i \leq e_0, \Delta_{1,i,n}(x_i) < 0\} \cup \{i| e_0 < x_i < e_0+\Delta, \Delta_{2,i,n}(x_i) < 0\}.
\]

Then \( \delta_n(x) \) is a.o. relative to \( G \).

3.2. \( \theta_0 \) unknown

If \( \theta_0 \) is unknown, let \( \pi_0 \) be the control population and \( X_0 \) be the random variable from \( \pi_0 \). We assume that \( X_0 \) has conditional pdf

\[
f(x_0|\theta_0) = \frac{1}{\theta_0} I(0,\theta_0)(x_0), \theta_0 > 0.
\]

In this case

\[
\Omega = (\theta = (\theta_0,\theta_1,\ldots,\theta_k)| \theta_i > 0, 1 = 0,1,\ldots,k), Z = (x = (x_0,x_1,\ldots,x_k)
\]

\[
|x_i > 0, i = 0,1,\ldots,k), G(\theta) = \prod_{i=0}^{k} G_i(\theta_i), f(x|\theta) = \prod_{i=0}^{k} f(x_i|\theta_i), \text{ and}
\]

at stage \( n \) we observed \( z_n = (x_{n0},x_{n1},\ldots,x_{nk}) \). Under the loss function

\((2.4)\), the Bayes rule \( \delta_B(x) \) is given by

\[
i \in \delta_B(x) \text{ if } \Delta_{G_0,G_i}(x_0,x_i) < 0
\]

where

\[
\Delta_{G_0,G_i}(x_0,x_i) = \int_{\theta_0}^{e_0} f(x_0|\theta_0) dG_0(\theta_0)m_i(x_i)-\Delta m_0(x_0)m_i(x_i)-
\]

\[
\int_{e_0}^{\theta_0} f(x_i|\theta_i) dG_i(\theta_i)m_0(x_0)+2\int_{\theta_0<\theta_i} (\theta_i-\theta_0)f(x_i|\theta_i)f(x_0|\theta_0)dG_i(\theta_i)dG_0(\theta_0).
\]
Using formula (3.4) if \(0 < x_i \leq x_0\), we have
\[
\int_{\{\theta \in \Theta_i\}} (\theta_i - \theta_0) f(x_i | \theta_i) f(x_0 | \theta_0) dG_i(\theta_i) dG_0(\theta_0)
\]
\[
= M_i(\theta_0) - \int_{x_0}^{\infty} \frac{M_i(\theta_0)}{\theta_0} dG_0(\theta_0),
\]
and if \(0 < x_0 < x_i\), we have
\[
\int_{\{\theta \in \Theta_i\}} (\theta_i - \theta_0) f(x_i | \theta_i) f(x_0 | \theta_0) dG_i(\theta_i) dG_0(\theta_0)
\]
\[
= (1 - G_i(x_i))(m_0(x_0) - m_i(x_i)) - m_i(x_i)(G_0(x_i) - G_0(x_0))
\]
\[
- \int_{x_i}^{\infty} \frac{M_i(\theta_0)}{\theta_0} dG_0(\theta_0) + m_0(x_i).
\]

Hence
(3.19) \(\Delta_{G_0, G_i}(x_0, x_1) = m_i(x_i)(1 - M_0(x_0)) + (1 + M_i(x_i))m_0(x_0) +
\[
(x_i - x_1) m_i(x_i)m_0(x_0) - 2 \int_{x_0}^{\infty} \frac{M_i(\theta_0)}{\theta_0} dG_0(\theta_0)
\]
\[
= \Delta_{1, G_0, G_i}(x_0, x_1) \text{ (say), if } 0 < x_i \leq x_0
\]
and
(3.20) \(\Delta_{G_0, G_i}(x_0, x_1) = (1 - M_i(x_i))m_0(x_0) + (1 + M_0(x_0) - 2M_0(x_i))m_i(x_i) +
\[
(x_i - x_0 - \delta) m_i(x_i)m_0(x_0) + 2M_i(x_i)m_0(x_0) - 2 \int_{x_i}^{\infty} \frac{M_i(\theta_0)}{\theta_0} dG_0(\theta_0)
\]
\[
= \Delta_{2, G_0, G_i}(x_0, x_1) \text{ (say), if } 0 < x_0 < x_i.
\]

Thus
(3.21) \(\delta_B(x) = \delta_{1, B}(x) \cup \delta_{2, B}(x)
\)
where
\[ \delta_{1,B}(x) = \{ i | 0 < x_i \leq x_0, \Delta_1, G_0, G_i(x_0, x_i) < 0 \} \]

and

\[ \delta_{2,B}(x) = \{ i | 0 < x_0 < x_i, \Delta_2, G_0, G_i(x_0, x_i) < 0 \} \]

Similar to Theorem 3.1, we have the following result.

**Theorem 3.2.** If \( \int_0^\infty \delta_i dG_i(\theta) < \infty \), \( i = 0, 1, \ldots, k \) and for all \( 1 \leq i \leq k \),

\[ \Delta_{1,i,n}(x_0, x_i) \overset{p}{\to} \Delta_{1,0,0}, G_i(x_0, x_i) \text{ for } x_i \leq x_0 \]

and

\[ \Delta_{2,i,n}(x_0, x_i) \overset{p}{\to} \Delta_{2,0,0}, G_i(x_0, x_i) \text{ for } x_0 < x_i. \]

Let

\[ \delta_n(x) = \{ i | x_i \leq x_0, \Delta_{1,i,n}(x_0, x_i) < 0 \} \cup \{ i | x_0 < x_i, \Delta_{2,i,n}(x_0, x_i) < 0 \}, \]

then \( \{ \delta_n(x) \} \) is a.o. relative to \( G \).

Hence our problem is to find a consistent estimator of

\[ \int_a^x \frac{M_i(\theta)}{\theta} dG_0(\theta) \text{ for } x_0 \leq a. \]

**Theorem 3.3.** Let \( M_{i,n}(x) \) and \( m_{i,n}(x) \) be defined by (3.13) and (3.16), respectively. Then

\[ \int_a^x \frac{M_{i,n}(\theta)}{\theta} dG_{0n}(\theta) \overset{p}{\to} \int_a^x \frac{M_i(\theta)}{\theta} dG_0(\theta) \text{ for } x_0 \leq a, \]

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

**Proof.**

\[ \int_a^x \frac{M_{i,n}(\theta)}{\theta} dG_{0n}(\theta) - \int_a^x \frac{M_i(\theta)}{\theta} dG_0(\theta) \]

\[ \leq \int_a^x \frac{|M_{i,n}(\theta) - M_i(\theta)|}{\theta} dG_{0n}(\theta) \]

\[ \leq \frac{1}{a} \sup_{-a < x < a} |M_{i,n}(x) - M_i(x)| \leq \varepsilon \]

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
\[ \int_a^{\infty} \frac{M_i(\theta_0)}{\theta_0} \, dG_{0n}(\theta_0) \leq \int_a^{\infty} \frac{M_i(\theta_0)}{\theta_0} \, dG_0(\theta_0). \]

Thus
\[ \left| \int_a^{\infty} \frac{M_i(\theta_0)}{\theta_0} \, dG_{0n}(\theta_0) - \int_a^{\infty} \frac{M_i(\theta_0)}{\theta_0} \, dG_0(\theta_0) \right| \]
\[ \leq \left| \int_a^{\infty} \frac{M_i(\theta_0)}{\theta_0} \, dG_{0n}(\theta_0) - \int_a^{\infty} \frac{M_i(\theta_0)}{\theta_0} \, dG_0(\theta_0) \right| + \left| \int_a^{\infty} \frac{M_i(\theta_0)}{\theta_0} \, dG_{0n}(\theta_0) - \int_a^{\infty} \frac{M_i(\theta_0)}{\theta_0} \, dG_0(\theta_0) \right| \]
\[ \leq c \quad \text{with probability near 1, for large } n. \]

From Theorem 3.3, if we define
\[ \Delta_{1,1,n}(x_0,x_i) = m_{in}(x_i)(1-M_{0n}(x_0)) + m_{0n}(x_0)(1+M_{in}(x_i)) \]
\[ + (x_0-x_i-\Delta)m_{in}(x_i) \, m_{0n}(x_0) - 2\int \frac{M_{in}(\theta_0)}{\theta_0} \, dG_{0n}(\theta_0), \]
where \( G_{0n}(\theta_0) = M_{0n}(\theta_0) - \theta_0 m_{0n}(\theta_0) \) and \( M_{in}(x) \), \( m_{in}(x) \) are defined by (3.13) and (3.16), respectively, and
\[ \Delta_{2,1,n}(x_0,x_i) = m_{0n}(x_0)(1-M_{in}(x_i)) + m_{in}(x_i)(1+M_{0n}(x_0)-2m_{0n}(x_i)) \]
\[ + (x_i-x_0-\Delta)m_{0n}(x_0)m_{in}(x_i) + 2m_{in}(x_i)m_{0n}(x_0) - 2\int \frac{M_{in}(\theta_0)}{\theta_0} \, dG_{0n}(\theta_0). \]
Then
\[ \Delta_{\varepsilon,1,n}(x_0,x_i) \xrightarrow{P} \Delta_{\varepsilon,G_0,G_i}(x_0,x_i), \ \varepsilon = 1,2. \]

Now, let
\[ \delta_n(x) = \{ 1 | x_i \leq x_0, \Delta_{1,1,n}(x_0,x_i) < 0 \} \cup \{ 1 | x_0 < x_i, \Delta_{2,1,n}(x_0,x_i) < 0 \}. \]
From Theorem 3.2, we have \( \delta_n(x) \) is a.o. relative to \( G \).
3.3. Rate of Convergence of the Empirical Bayes Rules

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

Definition 3.2. The sequence of procedures $\{\delta_n\}$ is said to be asymptotically optimal of order $\alpha_n$ 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 $\Delta_1, G_1(x_i), \Delta_2, G_1(x_i), \Delta_1, i, n(x_i)$ and $\Delta_2, i, n(x_i)$ be defined by (3.6), (3.7), (3.17) and (3.18) respectively. Then $0 < r_n(G, \delta_n) - r(G) < \infty \sum_{i=1}^{k} \int_0^{\delta_0} |\Delta_1, G_1(x_i)|^{1-\delta} E|\Delta_1, i, n(x_i) - \Delta_1, G_1(x_i)|^{\delta} dx_i + \infty \sum_{i=1}^{k} \int_0^{\delta_0} |\Delta_2, G_1(x_i)|^{1-\delta} E|\Delta_2, i, n(x_i) - \Delta_2, G_1(x_i)|^{\delta} dx_i, \delta > 0$.

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

Lemma 3.5. Let $\varphi(x)$ satisfy the conditions (i) $\varphi(x) = 0$ if $x \notin (0, a)$ for some finite $a > 0$, (ii) $\int_0^a \varphi(x)dx = 1$, and (iii) $\sup_{x} |\varphi(x)| < \infty$ and define $m_{in}(x_i) = \frac{1}{nh(n)} \sum_{j=1}^{n} \varphi\left(\frac{x_j - x_i}{h(n)}\right)$, where $(h(n))$ satisfy the conditions (3.15) (see Johns and Van Ryzin (1972)). Then $|Em_{in}(x_i) - m_i(x_i)| \leq h(n)f(x_i)\int_0^{\delta} \sup_{0 \leq y \leq \epsilon} |m_i(x_i + y)| du$, for large $n$, where $f(x_i) = \sup_{0 \leq y \leq \epsilon} |m_i(x_i + y)|$, $\epsilon > 0$. 


Proof. \( E_m(x_i) - m_i(x_i) \)

\[
= \frac{1}{h(n)} \int \frac{y-x_i}{h(n)} m_i(y) dy - m_i(x_i)
\]

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

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

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

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

\[
|E_m(x_i) - m_i(x_i)| \leq h(n)f_\varepsilon(x_i) \int_0^a \phi(u) |du|
\]

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

\[
\operatorname{var} m_i(x_i) \leq \frac{1}{nh(n)} m_i(x_i) \int_0^a \phi^2(u) du.
\]

Proof. \( \operatorname{var} m_i(x_i) = \operatorname{var} \left( \frac{1}{nh(n)} \sum_{j=1}^n \frac{X_{ij} - x_i}{h(n)} \right) \)

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

\[
\leq \frac{1}{nh(n)} m_i(x_i) \int_0^a \phi^2(u) du, \text{ since } m_i(x_i) +.
\]

Remark: From Lemma 3.5 and Lemma 3.6, we have

\[ m_i(x_i) \text{ is } \phi \text{ if } f_\varepsilon(x_i) < \infty. \]

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

(a) \( \operatorname{Var} \Delta_{1,1,n}(x_i) = O((e_0 - \Delta - x_i)^2 m_i(x_i) \frac{1}{nh(n)}), \)

(b) \( \operatorname{Var} \Delta_{2,1,n}(x_i) = O((x_i - e_0 - \Delta)^2 m_i(x_i) \frac{1}{nh(n)}). \)
Proof. (a) \( \text{Var} \Delta_{1,i,n}(x_i) \)
\[
\leq 2((\theta_0-\Delta-x_i)^2 \text{Var} m_i(x_i) + \text{var}(M_{i,n}(x_i) - 2M_{i,n}(\theta_0)))
\]
\[
\leq 2((\theta_0-\Delta-x_i)^2 m_i(x_i) \frac{1}{\text{nh}(n)} \int_0^a \varphi^2(u)du + \frac{5}{2n}) \quad \text{(By Lemma 3.6)}
\]
\[
\leq M \frac{1}{\text{nh}(n)} (\theta_0-\Delta-x_i)^2 m_i(x_i), \quad \text{for some} \ M > 0.
\]

Similarly, we have the result (b).

Theorem 3.8. Under the conditions of Lemma 3.5. If

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

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

(iii) \( \int_0^{\theta_0} |\Delta_{1,G_i}(x_i)|^{1-\delta} |\theta_0-\Delta-x_i|^{\delta/2} f_\xi(x_i)dx_i < \infty, \)

and

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

where \( 0 < \delta < 2, \) then

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

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

\[
0 \leq r_n(G,\delta_n)-r(G)
\]
\[
\leq \sum_{i=1}^k (\max(1,2^{\delta-1})\int_0^{\theta_0} |\Delta_{1,G_i}(x_i)|^{1-\delta} (\text{Var} \Delta_{1,i,n}(x_i))^{\delta/2}dx_i +
\]

\[ \int_{0}^{\theta_0} |\Delta_1, G_i(x_i)|^{1-\delta} |(\theta_0 - \Delta - x_i)(E m_{1n}(x_i) - m_i(x_i))|^{\delta} dx_i + \]
\[ \sum_{i=1}^{k} \{ \max(1, 2^{\delta - 1}) \int_{\theta_0}^{\delta + \Delta} |\Delta_2, G_i(x_i)|^{1-\delta} (\text{var } \Delta_2, i, n(x_i))^{\delta/2} dx_i + \]
\[ \int_{\theta_0}^{\delta + \Delta} |\Delta_2, G_i(x_i)|^{1-\delta} (E m_{1n}(x_i) - m_i(x_i))|^{\delta} dx_i] \].

By Lemma 3.7, we have
\[ \int_{0}^{\theta_0} |\Delta_1, G_i(x_i)|^{1-\delta} (\text{var } \Delta_1, i, n(x_i))^{\delta/2} dx_i = O((n h(n))^{-\delta/2}) \]
and
\[ \int_{\theta_0}^{\delta + \Delta} |\Delta_2, G_i(x_i)|^{1-\delta} (\text{var } \Delta_2, i, n(x_i))^{\delta/2} dx_i = O((n h(n))^{-\delta/2}). \]

By Lemma 3.5, we have
\[ \int_{0}^{\theta_0} |\Delta_1, G_i(x_i)|^{1-\delta} |(\theta_0 - \Delta - x_i)|^{\delta} |E m_{1n}(x_i) - m_i(x_i)|^{\delta} dx_i = O((h(n))^{\delta}) \]
and
\[ \int_{\theta_0}^{\delta + \Delta} |\Delta_2, G_i(x_i)|^{1-\delta} |(x_i - \theta_0 - \Delta)(E m_{1n}(x_i) - m_i(x_i))|^{\delta} dx_i = O((h(n))^{\delta}). \]

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

**Corollary 3.9.** Under the conditions of Theorem 3.8. If we take \( h(n) = n^{-\alpha}, 0 < \alpha < 1 \), then the optimal choice of \( \alpha \) 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 \( g_i(x)/x \) and \( m_i(x) \) are both bounded on \( (0, \theta_0^+\delta+c) \), it is easy to check that the conditions of Theorem 3.8 are satisfied for \( 0 < \delta \leq 1 \).
REFERENCES


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 $\delta$, $0 < \delta < 2$. 

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