Empirical Bayes Rules for
Selecting Good Populations*

by
Shanti S. Gupta
Purdue University

and
Ping Hsiao
Wayne State University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #81-5

March, 1981
(Revised September 1982)

*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.
Empirical Bayes Rules for
Selecting Good Populations*

by
Shanti S. Gupta
Purdue University

and
Ping Hsiao
Wayne State University

Abstract

A problem of selecting populations better than a control is considered. When the populations are uniformly distributed, empirical Bayes rules are derived for a linear loss function for both the known control parameter and the unknown control parameter cases. When the priors are assumed to have bounded supports, empirical Bayes rules for selecting good populations are derived for distributions with truncation parameters (i.e., the form of the pdf is \( f(x|\theta) = p_i(x)c_i(\theta)I(0,\theta)(x) \)). Monte Carlo studies are carried out which determine the minimum sample sizes needed to make the relative errors less than \( \epsilon \) for given \( \epsilon \)-values.

AMS Subject Classification: Primary 62F07, Secondary 62C10.

Key Words: Empirical Bayes; Asymptotically Optimal; Selection and Ranking; Truncation Parameter; Better than a Control.

*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.
Empirical Bayes Rules for Selecting Good Populations*

by

Shanti S. Gupta
Purdue University

and

Ping Hsiao
Wayne State University

1. Introduction

We assume that G is an unknown prior distribution on \( \Theta \), and denote the minimum Bayes risk in a decision problem by \( r(G) \). Robbins, in his pioneering papers (1955), (1964), proposed sequences of decision rules, based on data from \( n \) independent repetitions of the same decision problem, whose \( (n+1) \)st stage Bayes risk converges to \( r(G) \) as \( n \to \infty \). Such sequences of rules are called empirical Bayes rules. Empirical Bayes rules have been derived for multiple decision problems by Deely (1965), Van Ryzin (1970), Huang (1975), Van Ryzin and Susarla (1977), and Singh (1977). However, the forms of densities of the populations that these authors considered are either \( c(\theta)h(x)e^{\theta x} \), for continuous case or \( c(\theta)h(x)e^{\theta x} \), for discrete case, and the loss functions are either squared error or merely \( \max_{1 \leq j \leq k} |\theta_j - \theta_1| \) type. Fox (1978) discussed some estimation problems under the squared error loss, in which empirical Bayes rules were derived for uniform distributions for the first time. Barr and Rizvi (1966), and McDonald (1974) also considered selection problems related to uniform distribution by subset selection approach.

*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.
The problem considered in this paper is related to truncation parameters and can be illustrated by the following example. Suppose that there are $k$ treatments for a certain disease, and the effect of the treatment $i$ follows an unknown distribution $G_i, 1 \leq i \leq k$. The effectiveness of the treatment $i$ has been tested on $n$ subjects (for different treatments, different groups of subjects are used. If the same subject has to be used for more than one test, let there be a wash-out period between tests, so the effects of different treatments are independent.). Let $\theta_{ij}$ be the parameter which represents the effectiveness of the treatment $i$ on the subject $j$. The treatment $i$ is good for the subject $j$ if $\theta_{ij} > \theta_0$ and hence is called a good treatment, otherwise it is called a bad treatment. $\theta_0$ is called the control parameter. Let $Y_{ij}$ be an observable result of the treatment $i$ on the subject $j$ and assume that $\theta_{ij}$ is underestimated by $Y_{ij}$. The pdf of $Y_{ij}$ is $f(x|\theta_{ij}) = P_i(x)c_i(\theta_{ij})I(0,\theta_{ij})(x)$. Our purpose is to find an empirical Bayes rule which decides on the quality (better or worse than the control) of the treatment $i$ based on $Y_{ij} (1 \leq i \leq k, 1 \leq j \leq n)$ and $X_i (1 \leq i \leq k)$, where $X_i$ is the endpoint result of the treatment $i$ on the present patient. In Section 2, a general formulation is given and empirical Bayes rules are derived for selecting populations better than a known control when the populations are uniformly distributed (i.e. $p_i(x) = 1$). In Section 3, the same problem is considered except that the control parameter is unknown. In Section 4, empirical Bayes rules are found for truncation parameters (that is the densities are of the form $p_i(x)c_i(\theta_i)I(0,\theta_i)(x)$). Rate of convergence is also discussed. Monte Carlo studies are carried out for the priors $G(\theta) = \frac{\sigma^2}{C^2}I(0,c)(\theta)$. The smallest sample size $N$ is determined to guarantee that the relative error is less than $\epsilon$.

2. $\theta_0$ known

Assume that $\pi_1, \pi_2, \ldots, \pi_k$ are $k$ populations and $X_i$ is a random observation for a certain characteristic of $\pi_i$. Assume that $X_i \sim U(0,\theta_i)$, where $\theta_i$ is unknown for $1 \leq i \leq k$. Let $\theta_0$ be a known control parameter, we define $\pi_i$
to be a good population according to the specified characteristic if \( \theta_i > \theta_0 \) and to be a bad population if \( \theta_i \leq \theta_0 \). Let \( \Theta = \{ \theta = (\theta_1, \ldots, \theta_k) | \theta_i > 0 \text{ for all } 1 \leq i \leq k \} \). For any \( \theta \in \Theta \), let \( A(\theta) = \{ i | \theta_i > \theta_0 \} \) and \( B(\theta) = \{ i | \theta_i \leq \theta_0 \} \), then \( A(\theta) \) (\( B(\theta) \)) is the set of indices of good (bad) populations. Our goal is to select all the good populations and reject the bad ones. We formulate the problem in the empirical Bayes framework as follows:

1) Let \( G = \{ S | S \subseteq \{1, 2, \ldots, k\} \} \) be the action space.

   When we take action \( S \), we say \( \pi_i \) is good if \( i \in S \) and \( \pi_i \) is bad if \( i \notin S \).

2) \( L(\theta, S) = L_1 \sum_{i \in A(\theta) \cap S} (\theta_i - \theta_0) + L_2 \sum_{i \in B \cap S} (\theta_0 - \theta_i) \)

   is the loss function.

3) Let \( dG(\theta) = \prod_{i=1}^{k} dG_i(\theta_i) \) be an unknown prior distribution on \( \Theta \), where \( G_i \) has a continuous pdf \( g_i \) with respect to the Lebesgue measure.

4) Let \( \theta_{ij}, \ldots, \theta_{in}, Y_{in} \) be pairs of random variables from \( \pi_i \) and \( Y_{ij} | \theta_{ij} \sim U(0, \theta_{ij}) \) for \( 1 \leq i \leq k, 1 \leq j \leq n \). Let \( Y_j = (Y_{1j}, \ldots, Y_{kj}) \), then \( Y_j \) denotes the previous j-th observations from \( \pi_1, \ldots, \pi_k \).

5) Let \( X = (X_1, \ldots, X_k) \) be the present observation and \( f(x | \theta) = \prod_{i=1}^{k} \frac{1}{\theta_i} I(0, \theta_i)(x_i) \). Note that \( X_i = Y_{i,n+1} \) and \( \theta_i = \theta_{i,n+1} \). Since we are interested in Bayes rules, we can restrict our attention to the non-randomized rules.

6) Let \( D = \{ \delta | \delta: \mathcal{X} \to G \text{ is measurable} \} \), then \( r(G) = \inf_{\delta \in D} r(G, \delta) \) is the minimum Bayes risk.

The decision rules \( \{ \delta_n(x; Y_1, \ldots, Y_n) \}_{n=1}^{\infty} \) are said to be asymptotically optimal (a.o.) or empirical Bayes (e.B.) relative to \( G \) if \( r_n(G, \delta_n) = \int_E L(\theta, \delta_n(x; Y_1, \ldots, Y_n)) f(x | \theta) dG(\theta) dx \to r(G) \) as \( n \to \infty \). For simplicity,
\( \delta_n(x; Y_1, \ldots, Y_n) \) will be denoted by \( \delta_n(x) \).

Let \( m_1(x) \) be the marginal pdf of \( X_1 \) and \( M_1(x) \) be the marginal distribution of \( X_1 \). Then we have

\[
m_1(x) = \int_0^\infty \frac{1}{x} \, dG_1(\theta) \quad \text{for all} \quad x > 0,
\]

and

\[
M_1(x) = \int_0^\infty \int_0^\infty \frac{1}{x} \, dG_1(\theta) \, dt = \int_0^\infty \int_0^x \frac{1}{t} \, dG_1(\theta) + \int_0^\infty \int_0^x \frac{1}{t} \, dG_1(\theta)
\]

\[
= x m_1(x) + G_1(x).
\]

Hence, \( G_1(x) = M_1(x) - x m_1(x) \). (2.3)

Now, the loss function defined in (2.1) can be expressed as

\[
L(\theta, S) = \sum_{i \in S} [L_2(\theta_i - \theta_0) I(0, \theta_i) (\theta_i) - L_1(\theta_i - \theta_0) I(\theta_i, \infty)](\theta_i)
\]

\[
+ \sum_{i \not\in S} L_1(\theta_i - \theta_0) I(\theta_i, \infty). \quad (2.4)
\]

Since the second sum in (2.4) does not depend on the action \( S \), we can omit it and need only to consider the first sum as our loss for finding an empirical Bayes rules from now on. Therefore,

\[
r(G, \delta) = \int \sum_{i \in \delta(x)} \left[ \int_{\theta_i \leq \theta_0} L_2(\theta_i - \theta_0) f(x|\theta) \, dG(\theta)
\right.
\]

\[
- \int_{\theta_i > \theta_0} L_1(\theta_i - \theta_0) f(x|\theta) \, dG(\theta)] \, dx.
\]

So, if \( \delta_B(x) = S^* \) is the Bayes rule wrt \( G \), one finds \( i \in S^* \) if

\[
\int_{(0, \theta_0) \cap (x_i, \infty)} L_2(\theta_0 - \theta_i) \frac{1}{\theta_i} \, dG_1(\theta_i)
\]

\[
\leq \int_{\theta_0 \cap (x_i, \infty)} L_1(\theta_i - \theta_0) \frac{1}{\theta_i} \, dG_1(\theta_i). \quad \text{Hence},
\]
\[ S^* = \{ i | x_i \geq \theta_0 \} \cup \{ i | x_i < \theta_0 \text{ and } H_i(x_i) \leq c_i(\theta_0) \}, \]

where

\[ H_i(x_i) = L_2 \theta_0 \int_{x_i}^{\theta_0} \frac{1}{\theta_i} \ dG_i(\theta_i) + L_2 G_i(x_i) \quad \text{and} \]

\[ c_i(\theta_0) = L_2 G_i(\theta_0) + L_1 (1 - G_i(\theta_0)) - L_1 \theta_0 \int_{\theta_0}^{\infty} \frac{1}{\theta_i} \ dG_i(\theta_i). \]

Since \( H_i(x_i) \) is decreasing in \( x_i \) for \( x_i < \theta_0 \) and \( H(\theta_0) \leq c_i(\theta_0) \), so

\[ S^* = \{ i | x_i \geq \theta_0 - b_i \}, \text{where } b_i \geq 0 \text{ satisfies } H(\theta_0 - b_i) = c_i(\theta_0). \]

This shows for any \( G \), the above type rules are Bayes rules [see Gupta and Sobel (1958) and Gupta (1963, 1965)].

Now, \( G \) is unknown; the Bayes rules are not obtainable. We wish to find a sequence of rules \( \{ \delta_n(x) \}_{n=1}^{\infty} \) to be a.o. Let

\[ \Delta_{G_i}(x_i) = H_i(x_i) - c_i(\theta_0) \]

and

\[ S_{G_i}(x) = \{ i | x_i < \theta_0, \Delta_{G_i}(x_i) \leq 0 \}. \]

Also, for any \( i \) \((1 \leq i \leq k)\), let \( \Delta_{i,n}(x_i) = \Delta_i(x_i, Y_i, \ldots, Y_i^n) \) for all \( n = 1, 2, \ldots \), be a sequence of real-valued measurable functions; we define

\[ S_i(x) = \{ i | x_i < \theta_0 \text{ and } \Delta_{i,n}(x_i) \leq 0 \} \quad (2.5) \]

and

\[ \delta_n(x) = \{ i | x_i \geq \theta_0 \} \cup S_i(x). \quad (2.6) \]

One can show that

Theorem 2.1. If \( \int_{0}^{\infty} dG_i(\theta) < \infty \), \( i = 1, 2, \ldots, k \), and \( \Delta_{i,n}(x_i) \to \Delta_{G_i}(x_i) \) in (p) for almost all \( x_i < \theta_0 \). Then \( \{ \delta_n(x) \}_{n=1}^{\infty} \) defined by (2.6) is empirical Bayes.

Proof: For all \( S \in G \), let
\[ \mathcal{Z}_S = \{ x \mid x_i \geq \theta_0 \text{ if } i \in S \text{ and } x_i < \theta_0 \text{ if } i \notin S \}. \]

Now, for any \( x \in \mathcal{Z}_S \), \( \delta_B(x) = S \cup S_B(x) \). Therefore

\[ \int_{\mathcal{Z}_S} L(\theta, \delta_B(x))f(x|\theta)dG(\theta) = \sum_{i \in \delta_B(x)} \left[ \int_{\{ \theta_i \leq \theta_0 \}} L_2(\theta_0 - \theta_i)f(x|\theta)dG(\theta) - \int_{\{ \theta_i > \theta_0 \}} L_1(\theta_i - \theta_0)f(x|\theta)dG(\theta) \right] \]

\[ = \sum_{i \in S} (-Q(x)) + \sum_{i \in S_B(x)} \Delta_{G_i}(x_i) \prod_{j \neq i} m_j(x_j), \]

where \( Q(x) = \int_{\{ \theta_i > \theta_0 \}} L_1(\theta_i - \theta_0)f(x|\theta)dG(\theta). \)

Similarly, for \( x \in \mathcal{Z}_S \), we have

\[ \int_{\mathcal{Z}_S} L(\theta, \delta_n(x))f(x|\theta)dG(\theta) \]

\[ = \sum_{i \in S} (-Q(x)) + \sum_{i \in S_n(x)} \Delta_{G_i}(x_i) \prod_{j \neq i} m_j(x_j). \]

Hence, if \( \Delta_{i,n}(x_i) - \Delta_{i,n}(x_i) \) in (p), then

\[ 0 \leq \int_{\mathcal{Z}} [L(\theta, \delta_n(x)) - L(\theta, \delta_B(x))]f(x|\theta)dG(\theta) \]

\[ \leq \sum_{i \in S_n(x)} |\Delta_{G_i}(x_i) - \Delta_{i,n}(x_i)| \prod_{j \neq i} m_j(x_j) \]

\[ + \left( \sum_{i \in S_n(x)} - \sum_{i \in S_B(x)} \right) \Delta_{i,n}(x_i) \prod_{j \neq i} m_j(x_j) \]

\[ + \sum_{i \in S_B(x)} |\Delta_{i,n}(x_i) - \Delta_{G_i}(x_i)| \prod_{j \neq i} m_j(x_j) \]

\[ \leq 2\varepsilon \sum_{i=1}^{k} \prod_{j \neq i} m_j(x_j) \]

with probability near 1 for large \( n \). Note that (2.7) is non-positive by the definition of \( S_n(x) \). Thus, we have proved
in (p) for all most all x. By Corollary 1 of Robbins (1964), \( \{\delta_n(x)\}_{n=1}^{\infty} \) is empirical Bayes. This completes the proof.

In view of (2.2) and (2.3), we have

\[
\Delta_{G_i}(x_i) = L_2 m_i(x_i) (\theta_0 - x_i) + L_2 [M_i(x_i) - M_i(\theta_0)] + L_1 [M_i(\theta_0) - 1].
\]

Hence, if we define

\[
\Delta_{i,n}^*(x_i) = L_2 m_i, n(x_i) (\theta_0 - x_i) + L_2 [M_i, n(x_i) - M_i, n(\theta_0)] + L_1 [M_i, n(\theta_0) - 1],
\]

where

\[
M_i, n(x) = \frac{1}{n} \sum_{j=1}^{n} I_{(-\infty, x]}(Y_{ij})
\]

and

\[
m_i, n(x) = \frac{1}{h} [M_i, n(x+h) - M_i, n(x)], h > 0,
\]

then \( \Delta_{i,n}^*(x_i) \rightarrow \Delta_{G_i}(x_i) \) in (p) a.e. in x, if \( h = h(n) \rightarrow 0 \) and \( nh \rightarrow \infty \) as \( n \rightarrow \infty \).

So, by Theorem 1, \( \delta_{n,x}^*(x) = \{i | x_i \geq \theta_0\} \cup \{i | x_i < \theta_0, \Delta_{i,n}^*(x_i) \leq 0\} \) is empirical Bayes.

Remark: In (2.8), \( M_i, n(x) \) and \( m_i, n(x) \) can be defined as any functions such that \( M_i, n(x) \rightarrow M_i(x) \) in (p) and \( m_i, n(x) \rightarrow m_i(x) \) in (p) for almost all x.

For example, let \( m_{i,n}^0(x) = \frac{1}{nh} \sum_{j=1}^{n} w\left(\frac{x - Y_{ij}}{h}\right) \) where \( w(\cdot) \geq 0 \) satisfies

(i) \( \sup_{-\infty < x < \infty} w(x) \leq K \) for some constant K,

(ii) \( \int_{-\infty}^{\infty} w(x)dx = 1 \)

(iii) \( \lim_{x \rightarrow -\infty} xw(x) = 0 \)
and $h = h(n)$ satisfies $h \to 0$, $nh \to \infty$ as $n \to \infty$ then $m_{i,n}^0(x)$ is a consistent estimator of $m_i(x)$ (see Parzen (1962)).

3. $\theta_0$ unknown

Let $\pi_0$ be the control population and assume that $X_0$, a certain observable characteristic of $\pi_0$, follows $U(0,\theta_0)$. Let $Y_{01}, \ldots, Y_{0n}$ be the past data collected from $\pi_0$. Based on this further information, we will search for empirical Bayes rules for selecting populations better than the control.

Note that now $\theta = (\theta_0, \theta_1, \ldots, \theta_k)$, $x = (x_0, x_1, \ldots, x_k)$ and

$$G(\theta) = \prod_{i=0}^{k} G_i(\theta_i).$$

Under the loss function in (2.4), the Bayes rule $\delta_B$ is:

$$i \in \delta_B(x) \text{ if}$$

$$L_2 \int_{x_0}^{\infty} \frac{1}{\theta_0} \int_{\theta_0}^{\theta_i} \frac{1}{\theta_i} (\theta_0 - \theta_i) dG_i(\theta_i) dG_0(\theta_0)$$

$$\leq L_1 \int_{x_0}^{\infty} \frac{1}{\theta_0} \int_{\theta_0}^{\theta_i} \frac{1}{\theta_i} (\theta_i - \theta_0) dG_i(\theta_i) dG_0(\theta_0).$$

Hence, $i \in \delta_B(x)$ if

(i) $x_i \geq x_0$ and $\Delta_{G_0,G_i}^1(x_0,x_i) \leq 0$, where

$$\Delta_{G_0,G_i}^1(x_0,x_i) = (L_1-L_2)[\int_{x_i}^{\infty} m_i(\theta_0) dG_0(\theta_0) + \int_{x_i}^{\infty} m_0(\theta_1) dG_1(\theta_1)]$$

$$- L_1[1-G_i(x_i)]m_0(x_0) + m_i(x_i)[L_2+(L_1-L_2)G_0(x_i)-L_1G_0(x_0)]$$

or

(ii) $x_i < x_0$ and $\Delta_{G_0,G_i}^2(x_0,x_i) \leq 0$, where

$$\Delta_{G_0,G_i}^2(x_0,x_i) = (L_1-L_2)[\int_{x_0}^{\infty} m_i(\theta_0) dG_0(\theta_0) + \int_{x_0}^{\infty} m_0(\theta_1) dG_1(\theta_1)]$$

$$- m_0(x_0)[L_1+(L_2-L_1)G_i(x_0)-L_2G_i(x_i)] + L_2m_i(x_i)(1-G_0(x_0)).$$

(3.1)
When $L_1 = L_2 = L$, the Bayes rule is greatly simplified. We find

\[ i \in \delta_B(x) \text{ if} \]

\[ \Delta_{0,G_i}(x_0,x_i) = m_0(x_0)[1-G_i(x_i)] - m_i(x_i)[1-G_0(x_0)] \geq 0. \]

Let \( \delta_n(x) = \{i | \Delta_{i,n}(x_i,x_0) \geq 0\} \), where

\[ \Delta_{i,n}(x_i,x_0) = m_{0,n}(x_0)[1-G_{i,n}(x_i)] - m_{i,n}(x_i)[1-G_{0,n}(x_0)], \]

\( M_{i,n}(x_i) \) and \( m_{i,n}(x_i) \) are defined in (2.9), and \( G_{i,n}(x_i) = M_{i,n}(x_i) - x_i m_{i,n}(x_i) \). Then, \( \{\delta_n(x)\}_{n=1}^\infty \) is e.b. by Theorem 3.2. When \( L_1 \neq L_2 \), one needs to find consistent estimators of \( \int_a^\infty m_i(\theta_0)dG_{0}(\theta_0) \) and \( \int_a^\infty m_0(\theta_1)dG_{1}(\theta_1) \).

**Theorem 3.1.** Let \( M_{i,n}(x) \) and \( m_{i,n}(x) \) be defined by (2.9) with $h = h(n)$ satisfying $h \to 0$, $nh^2 \to \infty$ as $n \to \infty$. If \( \int_0^\infty dG_i(\theta) < \infty \) for all $i = 0,1,\ldots,k$, then

\[ \int_a^\infty x m_{i,n}(x)dm_{0,n}(x) \to \int_a^\infty m_i(x)dG_0(x) \]

in (p) for any $a > 0$.

**Proof:** See Appendix A.

**Theorem 3.2.** Assume that \( \int_0^\infty dG_i(\theta) < \infty \) for all $0 \leq i \leq k$. If for all

\[ 1 \leq i \leq k, \Delta_{i,n}(x_0,x_i) + \Delta_{i,G_0}(x_0,x_i) \to \Delta_{i,n}(x_0,x_i) \text{ in (p)} \] for $x_i > x_0$, and \( \Delta_{i,n}(x_0,x_i) \to \Delta_{i,G_0}(x_0,x_i) \) in (p) for $x_i < x_0$. Then

\[ \delta_n(x) = \delta_{n,1}(x) \cup \delta_{n,2}(x) \]

\[ = \{i | x_i \geq x_0 \text{ and } \Delta_{i,n}(x_0,x_i) \leq 0\} \cup \]

\[ \{i | x_i < x_0 \text{ and } \Delta_{i,n}(x_0,x_i) \leq 0\} \]

(3.3)

defines an empirical Bayes rule.
Proof: \[ \int_{\Theta} \mathcal{L}(\theta, \delta_B(x)) f(x | \theta) dG(\theta) \]

\[ = \sum_{i \in S_1^*(x)} \Delta_{G_i, G_0}^1(x_0, x_i) \prod_{j \neq i} m_j(x_j) + \sum_{i \in S_2^*(x)} \Delta_{G_i, G_0}^2(x_0, x_i) \prod_{j \neq i} m_j(x_j), \]

where \( S_1^*(x) = \{i | x_i \geq x_0 \text{ and } \Delta_{G_i, G_0}^1(x_0, x_i) \leq 0\} \)

\( S_2^*(x) = \{i | x_i < x_0 \text{ and } \Delta_{G_i, G_0}^2(x_0, x_i) \leq 0\}, \)

and \[ \int_{\Theta} \mathcal{L}(\theta, \delta_n^*(x)) f(x | \theta) dG(\theta) \]

\[ = \sum_{i \in S_1_n^*(x)} \Delta_{G_i, G_0}^1(x_0, x_i) \prod_{j \neq i} m_j(x_j) + \sum_{i \in S_2_n^*(x)} \Delta_{G_i, G_0}^2(x_0, x_i) \prod_{j \neq i} m_j(x_j). \]

Now, following the same method as in the proof of Theorem 2.1, we can show

\[ \sum_{i \in S_2_n^*(x)} \Delta_{G_i, G_0}^2(x_0, x_i) \prod_{j \neq i} m_j(x_j) \rightarrow \sum_{i \in S_1^*(x)} \Delta_{G_i, G_0}^1(x_0, x_i) \prod_{j \neq i} m_j(x_j) \]

in (p) for \( \ell = 1, 2 \). Hence \( \delta_n^*(x) \rightarrow n=1 \text{ is empirical Bayes}. \) This completes the proof.

Now, let

\[ \Delta_{i,n}^1(x_0, x_i) = (L_2-L_1) \{ \int_{x_i}^\infty x m_{i,n}(x) dm_{0,n}(x) + \int_{x_0}^\infty x m_{0,n}(x) dm_{i,n}(x) \}

- L_1[1-G_i,n(x_i)]m_{0,n}(x_0) + m_{i,n}(x_i) [L_2+(L_1-L_2)] G_{0,n}(x_i)-L_1 G_{0,n}(x_0), \]

(3.4)

and

\[ \Delta_{i,n}^2(x_0, x_i) = (L_2-L_1) \{ \int_{x_0}^\infty x m_{i,n}(x) dm_{0,n}(x) + \int_{x_0}^\infty x m_{0,n}(x) dm_{i,n}(x) \}

+ L_2[1-G_{0,n}(x_0)]m_{i,n}(x_i)-m_{0,n}(x_0) [L_1+(L_2-L_1)] G_{i,n}(x_0),

- L_2 G_{i,n}(x_i)], \]

where \( G_{i,n}(x) = M_{i,n}(x) - x m_{i,n}(x) \).

(3.5)
Then, by Theorem 3.1 and Theorem 3.2, (3.3), (3.4), and (3.5) define an empirical Bayes rule.

4. Generalization and Simulation

Let \( p_i(x) \) be a positive continuously differentiable function which is defined over \((0, \infty)\) for \(1 \leq i \leq k\). Let \( c_i(\theta)^{-1} = \int_0^\theta p_i(x)dx \) for \( \theta > 0 \), then \( f_i(x|\theta) = p_i(x)c_i(\theta)I(0,\theta)(x) \) is a density function and \( \theta \) is a truncation parameter. In this section, we assume that \( \pi_i = f_i(x|\theta_i) \) for \(1 \leq i \leq k\). Under the formulation of Section 2, we wish to find empirical Bayes rules for these more general density functions. For simplicity, we assume that \( L_1 = L_2 = L \) and that \( \theta_0 \) is known. Also we assume \( G_i(\theta) \) has a continuous density \( g_i(\theta) \) with a bounded support \([0, \alpha_i]\) with a known \( \alpha_i \) for all \(1 \leq i \leq k\). We find

\[
m_i(x) = \int_0^{\alpha_i} f_i(x|\theta)dG_i(\theta) = p_i(x)\int_{X_i}^{\alpha_i} c_i(\theta)dG_i(\theta).
\]

If we follow the same discussion as in Section 2, we can show that the Bayes rule \( \delta_B \) is \( i \in \delta_B(x) \) iff

(i) \( x_i \geq \theta_0 \), or

(ii) \( x_i < \theta_0 \) and \( \theta_0 \int_{X_i}^{\alpha_i} c_i(x)dG_i(x) \leq \int_{X_i}^{\alpha_i} x c_i(x)dG_i(x) \).

Hence, \( \delta_B(x) = \{i|x_i \geq \theta_0-d_i\} \), where \( d_i \geq 0 \) satisfies \( \int_{d_i}^{\alpha_i} (\theta_0-x) c_i(x)dG_i(x) = 0 \).

Let \( d_{i,n} = d_{i,n}(Y_1, \ldots, Y_{in}) \) be a consistent estimation of \( d_i \), then \( \delta^0_n(x) = \{i|x_i \geq \theta_0-d_{i,n}\} \) is e.B. and they are (weak) admissible in the sense that \( \delta^0_n(x_Y, \ldots, Y_n) \) is an admissible rule for the non-empirical problem for all \( Y_1, \ldots, Y_n \) and \( n \) (see Houwelingen (1976). Meeden (1972)). However,
to find such a sequence \( \{d_{i,n}\}_{n=1}^{\infty} \) is very difficult. In view of Theorem 2.1, a more practical way to find empirical Bayes rules is to estimate

\[
\alpha_i \int_{x_i} x c_i(x) dG_i(x).
\]

**Theorem 4.1.** Let \( p_i(x) \) and \( G_i(x) \) be defined as above. If \( m_{i,n}(x) \) is defined by (2.9) with \( h \to 0, nh \to \infty \), then

\[
\alpha_i \int_{x_i} x p_i(x) dx - \int_{x_i} p_i(x) m_{i,n}(x) dx - \int_{x_i} p_i(x) m_{i,n}(x) dx + \int_{x_i} x c_i(x) dG_i(x) \text{ in (p)}.
\]

**Proof:** See Appendix B.

Now, let

\[
\Delta_{i,n}(x_i) = \frac{\theta_0 m_{i,n}(x_i)}{p_i(x_i)} + \int_{x_i} \frac{x}{p_i(x)} dm_{i,n}(x) - \int_{x_i} \frac{x}{p_i(x)} m_i(x) dx,
\]

(4.1) then \( \delta_n^*(x) = \{i|x_i \geq \theta_0\} \cup \{i|x_i < \theta_0 \text{ and } \Delta_{i,n}(x_i) \leq 0\} \)

(4.2) defines an empirical Bayes rule.

The following lemma is a direct result of Lemma 3 of Van Ryzin and Susarla (1977).

**Lemma 4.2.** Let \( \Delta_{G_i}(x) = \int \frac{x_{\alpha_i}}{x} (\theta_0 - t) c_i(t) dG_i(t) I(0, \alpha_i)(x) \),

then

\[
0 \leq r_n(G, \delta_n^*) - r(G) = \sum_{i=1}^{k} \left[ \int_{H_i} \Delta_{G_i}(x) |p_i(x)| P[\Delta_{i,n}(x) < 0] dx + \int_{H_i} \Delta_{G_i}(x) |p_i(x)| P[\Delta_{i,n}(x) \geq 0] dx \right],
\]
where $\Delta_{i,n}^* (x)$ and $\delta_n^*$ are defined by (4.1) and (4.2) respectively, and

\[ H_1 = \{ x | x \leq \theta_0 \text{ and } \Delta_{G_i} (x) > 0 \} \quad \text{and} \quad H_2 = \{ x | x > \theta_0 \text{ and } \Delta_{G_i} (x) < 0 \} \]

Now, let $O(\alpha_n)$ denote a quantity such that $0 \leq \lim_{n \to \infty} \frac{O(\alpha_n)}{\alpha_n} < \infty$. Then

since $|\Delta_{G_i} (x)| p_i(x) \leq M_i$ for all $x \leq \theta_0$ for some constant $M_i$, so

\[
\begin{align*}
\frac{r_n(G, \delta_n^*)}{r(G)} &\leq \sum_{i=1}^{k} M_i \left\{ \int_{H_1} \mathbb{P}[\Delta_{i,n}^* (x) < 0] dx \right. \\
&\quad \left. + \int_{H_2} \mathbb{P}[\Delta_{i,n}^* (x) > 0] dx \right\}.
\end{align*}
\]

Therefore, if for all $x \leq \theta_0$

\[
P\left[ |\Delta_{i,n}^* (x) - \Delta_{G_i} (x)| > |\Delta_{G_i} (x)| \right] = O(\alpha_n) \quad \text{as } n \to \infty
\]

then

\[
r_n(G, \delta_n^*) - r(G) = O(\alpha_n).
\]

Now, by the inequality

\[
P\left[ |\Delta_{i,n}^* (x) - \Delta_{G_i} (x)| > |\Delta_{G_i} (x)| \right] \leq \frac{\text{Var}[\Delta_{i,n}^* (x)\big]}{\left[ |\Delta_{G_i} (x)| - \mathbb{E}[\Delta_{i,n}^* (x)\big] \right]^2},
\]

we conclude that if $\text{Var}[\Delta_{i,n}^* (x)] = O(\alpha_n)$ for all $x \leq \theta_0$ then

\[
r_n(G, \delta_n^*) - r(G) = O(\alpha_n). \quad \text{Note that if } \alpha_n \to 0, \text{ then } \delta_n^* \text{ is empirical Bayes.}
\]

In the following, we have carried out some Monte Carlo study to see how fast the derived empirical Bayes rules converge. We let $X_i \sim U(0, \theta_i)$ for $i = 0, 1$. $\theta_0$ is treated as unknown. Assume that $g_i(\theta) = \frac{2\theta}{c} I(0, c)(\theta)$ for $i = 0, 1$ and $L_1 = L_2 = 1$. The smallest sample size $N$ such that

Relative error: $\frac{|r_m(G, \delta_m^*) - r(G)|}{r(G)} \leq \varepsilon$
for \( N-4 \leq m \leq N \) is determined. Here 
\[ r(S) = P_{\theta}(\theta_1 > \theta_0, X_1 < X_0) \cup \]
\[ (\theta_1 < \theta_0, X_1 > X_0) = \frac{c}{15}. \]
The Monte Carlo studies are repeated for 55 times and the values of \( N \) and the associated standard deviations corresponding to selected \( \varepsilon \) and \( c \) are shown in the next table for \( h = n^{-1/4} \), for \( h = n^{-1/5} \) and for \( h = n^{-1/6} \), where \( h \) is used to define (2.9).
Table 1

Upper entries are the lists of values of the smallest $N$ such that \( \frac{|r_m(G, \delta^*) - r(G)|}{r(G)} \leq \epsilon \) for $N-4 \leq m \leq N$, lower entries are the list of associated standard deviations for each corresponding case where the density of the priors is $g_i(\theta) = \frac{2\theta}{c^2} I(0, c)(\theta)$ for $i = 0, 1$.

<table>
<thead>
<tr>
<th>$h = n^{-1/4}$</th>
<th>$h = n^{-1/5}$</th>
<th>$h = n^{-1/6}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>$\epsilon$</td>
<td>.25</td>
</tr>
<tr>
<td>1/3</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>1/2</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>42</td>
<td>46</td>
</tr>
</tbody>
</table>

Note: "-" means that $N > 400$ (Monte Carlo study was curtailed because of limited resources)
Appendix A

Proof of Theorem 3.1.

For \( i \) fixed, \( \int_0^\infty x m_i, n(x) dm_0, n(x) \)

\[
= \frac{1}{n^2} \sum_{j=1}^n \sum_{\ell=1}^n \int_a^x I(x, x+h)(Y_{ij})dI[Y_{0x-h}, Y_{0x}](x)
\]

\[
= \frac{1}{n^2} \sum_{j=1}^n \sum_{\ell=1}^n (U_{j\ell} - V_{j\ell}), \text{ where}
\]

\[U_{j\ell} = (Y_{0x-h})I(a_{\ell})(Y_{0x-h})I(Y_{0x-h}, Y_{0x})(Y_{ij})\]

\[V_{j\ell} = Y_{0x}I(a_{\ell})(Y_{0x})I(Y_{0x}, Y_{0x}+h)(Y_{ij}).\]

Since \( Y_{0x} \sim M_0(x) \) and \( Y_{ij} \sim M_i(x) \) for \( 1 \leq j, \ell \leq n \), so

\[
\mathbb{E} \int_a^\infty x m_i, n(x) dm_0, n(x) = \frac{1}{h^2} \mathbb{E}[U_{11} - V_{11}]
\]

\[
= \int_a^\infty \frac{1}{h} \int_x^{x+h} dM_1(y) \frac{1}{h} [m_0(x+h) - m_0(x)] dx.
\]

Now, by (2.2) \( m_i(x) \) is decreasing in \( x \), hence

\[
\frac{1}{h} \int_x^{x+h} dM_i(y) \leq m_i(x) \leq \frac{1}{x} [1 - G_i(x)]. \tag{A.1}
\]

Then

\[
|x| \frac{1}{h} \int_x^{x+h} dM_1(y) \frac{1}{h} [m_0(x+h) - m_0(x)]
\]

\[
\leq \frac{1}{x} [1 - G_i(x)] \frac{1}{h} \int_x^{x+h} \frac{1}{\delta} dG_0(\delta) \leq \frac{1}{x} g_0(x+\delta h), \text{ for some } \delta \in [0,1].
\]

The last term is integrable over \((a, \infty)\), then by Lebesgue Dominated Convergence Theorem (LDCT),
\[
E \int_{a}^{\infty} x m_{i,n}(x) dm_{0,n}(x) + \int_{a}^{\infty} x m_{i}(x)m_{0}(x)dx
= - \int_{a}^{\infty} m_{i}(x)dG_{0}(x) \text{ in (p) if } h \to 0 \text{ as } n \to \infty. \quad \text{(A.2)}
\]

Now, \( \text{Var} \int_{a}^{\infty} x m_{i,n}(x) dm_{0,n}(x) = \text{Var} \frac{1}{n^{2}} \frac{1}{h^{2}} \sum_{j \neq k} (U_{j,k} - V_{j,k}) \)
\[
= \frac{1}{n^{2}h^{2}} \text{Var}(U_{11} - V_{11}) + \frac{2(n-1)}{n^{2}h^{2}} \text{Cov}(U_{11} - V_{11}, U_{12} - V_{12}). \quad \text{(A.3)}
\]

But \( \text{Var}(U_{11} - V_{11}) \leq E[(U_{11} - V_{11})^{2}] = E(U_{11}^{2}) + E(V_{11}^{2}) \) [because \( U_{11}V_{11} = 0 \)], and \( \frac{1}{h} E(U_{11}^{2}) \)
\[
\leq \int_{a}^{\infty} x^{2} \cdot \frac{1}{h} \int_{x}^{x+h} dM_{i}(y)dM_{0}(x+h)
\leq \int_{a}^{\infty} x^{2} \cdot \frac{1}{h} (1-G_{i}(x))dM_{0}(x+h) \leq \int_{a}^{\infty} x dM_{0}(x+h)
\leq E_{i}[X] = E_{0}[E[X|\theta_{0}]] = \frac{1}{2} E_{0}[\theta_{0}] < \infty,
\]

hence \( \frac{1}{h} \text{Var}(U_{11} - V_{11}) \leq E_{0}[\theta_{0}] \) for all \( h > 0. \) \( \quad \text{(A.4)} \)

Meanwhile, \( \text{Cov}(U_{11} - V_{11}, U_{12} - V_{12}) = \text{Cov}(U_{11}, U_{12}) + \text{Cov}(V_{11}, V_{12}) - \text{Cov}(U_{11}, V_{12}) - \text{Cov}(V_{11}, U_{12}) \), and \( \left| \frac{1}{h^{2}} \text{Cov}(U_{11}, U_{12}) \right| \leq \frac{1}{h^{2}} [E(U_{11}U_{12}) + E(U_{11}E(U_{12}))] \) because \( U_{j,k} > 0 \) for all \( 1 \leq j, k \leq n. \)

Now, \( \frac{1}{h^{2}} E(U_{11}U_{12}) = \frac{1}{h^{2}} \int_{h}^{\infty} \left[ \int_{0}^{\infty} ydM_{0}(y+h) \right]^{2} dM_{i}(x) \)
\[
= \frac{1}{h^{2}} \int_{h}^{\infty} \left[ \int_{a+h}^{x} ydM_{0}(y+h) \right]^{2} dM_{i}(x) + \frac{1}{h^{2}} \int_{a+h}^{x} \int_{a}^{x} ydM_{0}(y+h) dM_{i}(x).
\]

Because \( \int_{x-h}^{x} ydM_{0}(y+h) = \int_{x-h}^{x} \int_{y}^{\infty} \frac{1}{y} dG_{0}(\theta)dy \)
\[
\leq \int_{x-h}^{x} y \cdot \frac{1}{y+h} \cdot dy \leq h, \text{ and similarly},
\]
\[ \int_a^x ydM_0(y+h) \leq \int_a^{a+h} ydM_0(y+h) \leq h \text{ for } a < x < a+h, \]

we get \[ \frac{1}{h^2} E(U_{11}U_{12}) \leq 1-M_i(a+h) + M_i(a+h) - M_i(a) = 1-M_i(a). \]

The same argument shows that \[ \frac{1}{h} E(U_{11}) \leq 1-M_i(a) \]

\[ \frac{1}{h} E(V_{11}) \leq 1-M_i(a), \]

hence \[ \frac{1}{h^2} \text{ Cov}(U_{11}, U_{12}) \leq 2[1-M_i(a)]. \] This implies that

\[ \frac{1}{h^2} |\text{ Cov}(U_{11}-V_{11}, U_{12}-V_{12})| \leq 8[1-M_i(a)] \text{ for any } h > 0. \] (A.5)

By (A.3), (A.4) and (A.5)

\[ \text{Var} \int_a^x \mu_i(x)dm_0(x) \to 0 \text{ if nh}^2 \to 0 \text{ and } h \to 0. \] (A.6)

Now, (A.2) and (A.6) implies that

\[ \int_a^\infty \mu_i(x)dm_0(x) \to -\int_a^\infty \mu_i(x)dxG_0(x) \text{ in } (p). \]

This finishes the proof.
Appendix B

Proof of Theorem 4.1.

First, \[ E \int_{x_i}^{x_i+\delta} \frac{x}{p_1(x)} \, dm_i, n(x) = \int_{x_i}^{x_i+\delta} \frac{x}{p_1(x)} \, \frac{1}{h} \left[ m_i(x+h) - m_i(x) \right] \, dx \]

\[ \to \int_{x_i}^{x_i+\delta} \frac{x}{p_1(x)} \, dm_i(x) \quad \text{by LDCT.} \]

Now, \[ \text{Var} \int_{x_i}^{x_i+\delta} \frac{x}{p_1(x)} \, dm_i, n(x) = \text{Var} \left[ \frac{1}{nh} \sum_{j=1}^{n} (U_j - V_j) \right], \]

where \[ U_j = \frac{Y_i - \frac{x}{p_1(Y_{ij}) - h}}{p_1(Y_{ij} - h)} \quad \text{I} \left[ x_i, x_i + \delta \right] (Y_{ij} - h), \]

and \[ V_j = \frac{Y_{ij}}{p_1(Y_{ij} - h)} \quad \text{I} \left[ x_i, x_i + \delta \right] (Y_{ij}). \]

Hence, \[ \text{Var} \int_{x_i}^{x_i+\delta} \frac{x}{p_1(x)} \, dm_i, n(x) = \frac{1}{nh^2} \text{Var}(U_1 - V_1) \]

\[ \leq \frac{1}{nh^2} E \left[ (U_1 - V_1)^2 \right] \leq \frac{1}{n} \int_{x_i}^{x_i+\delta} \left[ \frac{x}{p_1(x)} - \frac{x-h}{p_1(x-h)} \right]^2 \, dM_1(x) \]

\[ + \frac{1}{nh} \int_{x_i}^{x_i+\delta} \left[ \frac{x-h}{p_1(x-h)} \right]^2 \, dM_1(x) \]

\[ + \frac{1}{nh} \int_{x_i}^{x_i+\delta} \frac{x-h}{p_1(x-h)} \, dM_1(x) \]

\[ \leq \frac{1}{n} \max_{x \in [x_i, x_i+\delta]} \left[ \frac{dx}{p_1(x)} \right]^2 + \frac{2}{nh} \max_{x \in [x_i, x_i+\delta]} \left[ \frac{x}{p_1(x)} \right]^2 \]

\[ \to 0 \quad \text{if} \quad nh \to \infty. \]

We see that

\[ \int_{x_i}^{x_i+\delta} \frac{x}{p_1(x)} \, dm_i, n(x) \to \int_{x_i}^{x_i+\delta} \frac{x}{p_1(x)} \, dm_i(x) \quad \text{in} \ (p). \]
Similarly \[ \int \frac{\alpha_i x p_i'(x)}{x_i p_i'(x)} m_i(x) dx \rightarrow \int \frac{\alpha_i x p_i'(x)}{x_i p_i'(x)} m_i(x) dx \text{ in } (p). \]

Since \[ \int x c_i(x)dG_i(x) = \int -x \frac{d}{dx} \left[ \frac{m_i(x)}{p_i(x)} \right] \]

\[ = \int \frac{\alpha_i x p_i'(x)}{x_i p_i'(x)} m_i(x) dx - \int \frac{x}{x_i p_i'(x)} dm_i(x), \]

the proof is completed.

Acknowledgement

We would like to thank the referee for some helpful suggestions.
Bibliography


Empirical Bayes Rules for Selecting Good Populations

Shanti S. Gupta and Ping Hsiao

March 1981

Approved for public release, distribution unlimited.

Asymptotically optimal, empirical Bayes, truncation parameter, rate of convergence, Monte Carlo study.

A problem of selecting populations better than a control is considered. When the populations are uniformly distributed, empirical Bayes rules are derived for a linear loss function for both the known control parameter and the unknown control parameter cases. When the priors are assumed to have bounded supports, empirical Bayes rules for selecting good populations are derived for distributions with truncation parameters (i.e. the form of the pdf is \( f(x|\theta) = p_{i}(x)c_{i}(\theta)I(0,\theta)(x) \)). Monte Carlo studies are carried out which determine the
minimum sample sizes needed to make the relative errors less than $\varepsilon$ for given $\varepsilon$-values.