Empirical Bayes Rules for Selecting the Best Binomial Population*

by

Shanti S. Gupta
Purdue University
and
TaChen Liang
Southern Illinois University
Technical Report#86-13

Department of Statistics
Purdue University

May 1986

* This research was partially supported by the Office of Naval Research Contract N00014-84-C-0167 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

EMPIRICAL BAYES RULES FOR SELECTING THE BEST BINOMIAL POPULATION*

Shanti S. Gupta and TaChen Liang
Purdue University and Southern Illinois University

Abstract

Consider k populations π_i, $i = 1,\ldots,k$, where an observation from π_i has binomial distribution with parameters N and p_i (unknown). Let $p_{[k]} = \max_{1\le j\le k} p_j$. A population π_i with $p_i = p_{[k]}$ is called a best population. We are interested in selecting the best population. Let $\underset{\sim}{p} = (p_1,\ldots,p_k)$ and let i denote the index of the selected population. Under the loss function $\ell(\underset{\sim}{p},i) = p_{[k]}-p_i$, this statistical selection problem is studied via empirical Bayes approach.

Some selection rules based on monotone empirical Bayes estimators of the binomial parameters are proposed. First, it is shown that, under the squared error loss, the Bayes risks of the proposed monotone empirical Bayes estimators converge to the related minimum Bayes risks with rates of convergence at least of order $0(n^{-1})$, where n is the number of accumulated past experiences at hand. Further, for the selection problem, the rates of convergence of the proposed selection rules are shown to be at least of order $0(\exp(-cn))$ for some $c > 0$.

Abbreviated Title: Empirical Bayes Selection Rules

AMS 1980 Subject Classification: 62F07, 62C12

Key Words and Phrases: Bayes rule, empirical Bayes rule, monotone estimation, monotone selection rule, Asymptotically optimal, rate of convergence

*This research was partially supported by the Office of Naval Research Contract N00014-84-C-0167 at Purdue University.

EMPIRICAL BAYES RULES FOR SELECTING THE BEST BINOMIAL POPULATION

1. Introduction

In many situations, an experimenter is often confronted with choosing a model which is the best in some sense among those under study. For example, consider k different competing drugs for a certain ailment. We would like to select the best among them in the sense that it has the highest probability of success (cure of the ailment). This kind of binomial model occurs in many fields, such as medicine, engineering, and sociology. The problem of selecting a binomial model associated with the largest probability of success was first considered by Sobel and Huyett (1957) and Gupta and Sobel (1960). The former used the indifference zone formulation and the latter studied the subset selection approach; see Gupta and Huang (1976) and Gupta, Huang and Huang (1976), and Gupta and McDonald (1986) for further variations in goals and procedures for this problem.

Now, consider a situation in which one will be repeatedly dealing with the same selection problem independently. This will be the case with an on-going testing with drugs, for example. In such instances, it is reasonable to formulate the component problem in the sequence as a Bayes decision problem with respect to an unknown prior distribution on the parameter space, and then, use the accumulated observations to improve the decision rule at each stage. This is the empirical Bayes approach of Robbins (see Robbins (1956, 1964 and 1983)). Many such empirical Bayes rules have been shown to be asymptotically optimal in the sense that the risk for the nth decision problem converges to the minimum Bayes risk which would have been obtained if the prior distribution was known and the Bayes rule with respect to this prior distribution was used.

Empirical Bayes rules have been derived for subset selection goals by Deely (1965). Recently, Gupta and Hsiao (1983) and Gupta and Leu (1983) have studied empirical Bayes rules for selecting good populations with respect to a standard or a control, with the underlying distributions being uniformly distributed. Gupta and Liang (1984) studied empirical Bayes rules for selecting binomial populations better than a standard or a control.

In this paper, we obtain empirical Bayes procedures for selecting the best among k different binomial populations. These rules are based on monotone empirical Bayes estimators of the binomial success probabilities. First, it is shown that, under the squared error loss, the Bayes risks of the proposed monotone empirical Bayes estimators converge to the related minimum Bayes risks with rates of convergence at least of order $O(n^{-1})$. Further, for the selection problem, the rates of convergence of the proposed selection rules are shown to be at least of order $O(\exp(-cn))$ for some $c > 0$.

2. Formulation of the Empirical Bayes Approach

Consider k binomial populations π_i, $i = 1,\ldots,k$, each consisting of N trials. For each i, $i = 1,\ldots,k$, let p_i be the probability of success for each trial in π_i, and let X_i denote the number of successes among the associated N trials. Then, $X_i|p_i$ is binomially distributed with probability function $f_i(x|p_i) = \binom{N}{x_i} p_i^{x_i}(1-p_i)^{N-x_i}$, $x_i = 0,1,\ldots,N$. Let $f(\underset{\sim}{x}|\underset{\sim}{p}) = \prod_{i=1}^{k} f_i(x_i|p_i)$ where $\underset{\sim}{x} = (x_1,\ldots,x_k)$ and $\underset{\sim}{p} = (p_1,\ldots,p_k)$. For each $\underset{\sim}{p}$, let $p_{[1]} \le \cdots \le p_{[k]}$ be the ordered parameters of $p_1,\ldots,p_k$. It is assumed that the exact matching between the ordered and the unordered parameters is unknown. Any population π_i with $p_i = p_{[k]}$ is considered as the best population. Our goal is to derive empirical Bayes rules to select the best population.

Let $\Omega = \{\underset{\sim}{p}|\underset{\sim}{p} = (p_1,\ldots,p_k),\ p_i \in (0,1),\ i = 1,\ldots,k\}$ be the parameter space and $G(\underset{\sim}{p}) = \prod_{i=1}^{k} G_i(p_i)$ be the prior distribution over Ω. Let $\mathcal{A} = \{i|i = 1,\ldots,k\}$ be the action space. When action i is taken, it means that population π_i is selected as the best population. For the parameter $\underset{\sim}{p}$ and action i, the loss function $\ell(\underset{\sim}{p},i)$ is defined as:

(2.1) $$\ell(\underset{\sim}{p},i) = p_{[k]} - p_i,$$

the difference between the best and the selected population.

Let $\mathfrak{X} = \prod_{i=1}^{k} \{0,1,\ldots,N\}$ be the sample space. A selection rule $d = (d_1,\ldots,d_k)$ is a mapping from $\mathfrak{X}$ to $[0,1]^k$ such that for each observation $\underset{\sim}{x} = (x_1,\ldots,x_k)$, the function $d(\underset{\sim}{x}) = (d_1(\underset{\sim}{x}),\ldots,d_k(\underset{\sim}{x}))$ satisfies that $0 \le d_i(\underset{\sim}{x}) \le 1$, $i = 1,\ldots,k$, and $\sum_{i=1}^{k} d_i(\underset{\sim}{x}) = 1$. Note that $d_i(\underset{\sim}{x})$, $i = 1,\ldots,k$, is the probability of selecting population π_i as the best population when $\underset{\sim}{x}$ is observed.

Let $\mathfrak{D} = \{d \mid d : \mathfrak{X} \to [0,1]^k$, being measurable$\}$ be the set of all selection rules. For each $d \in \mathfrak{D}$, let $r(G,d)$ denote the associated Bayes risk. Then, $r(G) = \inf_{d\in\mathfrak{D}} r(G,d)$ is the minimum Bayes risk.

From (2.1), the Bayes risk associated with selection rule d is:

$$\begin{aligned} r(G,d) &= \int_{\Omega} \sum_{\underset{\sim}{x}\in\mathfrak{X}} \ell(\underset{\sim}{p}, d(\underset{\sim}{x}))f(\underset{\sim}{x}|\underset{\sim}{p})dG(\underset{\sim}{p}) \\ &= C - \sum_{\underset{\sim}{x}\in\mathfrak{X}} \Big[\sum_{i=1}^{k} d_i(\underset{\sim}{x})\varphi_i(x_i)\Big] f(\underset{\sim}{x}), \end{aligned} \tag{2.2}$$

where $f(\underset{\sim}{x}) = \prod_{i=1}^{k} f_i(x_i)$, $\varphi_i(x) = \frac{W_i(x)}{f_i(x)}$,

$$f_i(x) = \int_0^1 f_i(x|p)dG_i(p), \qquad W_i(x) = \int_0^1 pf_i(x|p)dG_i(p)$$

$$C = \sum_{\underset{\sim}{x}\in\mathfrak{X}} \int_{\Omega} p_{[k]}dG(\underset{\sim}{p}|\underset{\sim}{x})f(\underset{\sim}{x}),\ \text{being a constant,}$$

and $G(\underset{\sim}{p}|\underset{\sim}{x})$ is the posterior distribution of $\underset{\sim}{p}$ given $\underset{\sim}{x}$.

For each $\underset{\sim}{x} \in \mathfrak{X}$, let

$$(2.3) \qquad A(\underset{\sim}{x}) = \{i \mid \varphi_i(x_i) = \max_{1 \le j \le k} \varphi_j(x_j)\}.$$

Thus, a randomized Bayes rule is

$d_G = (d_{1G}, \ldots, d_{kG})$, where

$$(2.4) \qquad d_{iG}(\underset{\sim}{x}) = \begin{cases} |A(\underset{\sim}{x})|^{-1}, & \text{if } i \in A(\underset{\sim}{x}); \\ 0 & \text{otherwise;} \end{cases}$$

and $|A|$ denotes the size of the set A.

When the prior distribution G is unknown, it is impossible to apply the Bayes rules. In this case, we use the empirical Bayes approach. Note that, for each i, $\varphi_i(x_i)$ is the posterior mean of the binomial probability p_i given that $X_i = x_i$ is observed. Due to the surprising quirk that $\varphi_i(x_i)$ can not be consistently estimated in the usual empirical Bayes sense (see Robbins (1964), Samuel (1963) and Vardeman (1978)), we use below an idea of Robbins in setting up the empirical Bayes framework for our selection problem.

For each i, $i = 1, \ldots, k$, at stage j, consider N+1 trials from π_i. Let X_{ij} and Y_{ij}, respectively, stand for the number of successes in the first N trials and the last trial. Let P_{ij} stand for the probability of success for each of the N+1 trials. P_{ij} has distribution G_i. Conditional on $P_{ij} = p_{ij}$, $X_{ij}|p_{ij} \sim B(N, p_{ij})$, $Y_{ij}|p_{ij} \sim B(1, p_{ij})$, and $X_{ij}|p_{ij}$ and $Y_{ij}|p_{ij}$ are independent. Let $\underset{\sim}{Z}_j = ((X_{1j}, Y_{1j}), \ldots, (X_{kj}, Y_{kj}))$ denote the observations at the jth stage, $j = 1, \ldots, n$. We also let $\underset{\sim}{X}_{n+1} = \underset{\sim}{X} = (X_1, \ldots, X_k)$ denote the present observations.

Consider an empirical Bayes selection rule $d_n(\underset{\sim}{x}; \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n) = (d_{1n}(\underset{\sim}{x}; \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n),\ldots,d_{kn}(\underset{\sim}{x}; \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n))$. Let $r(G,d_n)$ be the Bayes risk associated with the selection rule $d_n(\underset{\sim}{x}; \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n)$. Then,

$$(2.5)\quad r(G,d_n) = \sum_{\underset{\sim}{x}\in\mathfrak{X}} E\int_{\Omega} \ell(\underset{\sim}{p}, d_n(\underset{\sim}{x}; \underset{\sim}{Z}_1,\ldots,\underset{\sim}{z}_n))f(\underset{\sim}{x}|\underset{\sim}{p})dG(\underset{\sim}{p}),$$

where the expectation is taken with respect to $(\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n)$. For simplicity, $d_n(\underset{\sim}{x}; \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n)$ will be denoted by $d_n(\underset{\sim}{x})$.

Definition 2.1. A sequence of selection rules $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal relative to the prior distribution G if $r(G,d_n) \to r(G)$ as $n \to \infty$.

From (2.4), a natural empirical Bayes selection rule can be defined as follows:

For each $i = 1,\ldots,k$, and $n = 1,2,\ldots$, let $\varphi_{in}(x) \equiv \varphi_{in}(x; (X_{i1},Y_{i1}),\ldots,(X_{in},Y_{in}))$ be an estimator of $\varphi_i(x)$. Let $A_n(\underset{\sim}{x}) = \{i|\varphi_{in}(x_i) = \max_{1\le j\le k} \varphi_{jn}(x_j)\}$, and define $d_n(\underset{\sim}{x}) = (d_{1n}(\underset{\sim}{x}),\ldots, d_{kn}(\underset{\sim}{x}))$ where

$$(2.6)\qquad d_{in}(\underset{\sim}{x}) = \begin{cases} |A_n(\underset{\sim}{x})|^{-1} & \text{if } i \in A_n(\underset{\sim}{x}); \\ 0 & \text{otherwise.} \end{cases}$$

If $\varphi_{in}(x) \xrightarrow{P} \varphi_i(x)$ for all $x = 0,1,\ldots,N$ and $i = 1,\ldots,k$ (where "$\xrightarrow{P}$" means convergence in probability), then, by the boundedness of the loss function $\ell(\underset{\sim}{p},i)$ and Corollary 2 of Robbins (1964), it follows that $r(G,d_n) \to r(G)$ as $n \to \infty$. Thus, the sequence of selection rules $\{d_n\}_{n=1}^{\infty}$ defined in (2.6) is asymptotically optimal. Hence, our task is only to find the sequence of estimators $\{\varphi_{in}(x)\}$ possessing the above mentioned convergence property.

3. The Proposed Empirical Bayes Selection Rules

Before we go further to construct empirical Bayes estimators $\{\varphi_{in}(x)\}$, we first investigate some property related to the Bayes rule d_G defined in (2.4).

Definition 3.1. A selection rule $d = (d_1,\ldots,d_k)$ is said to be monotone if for each $i = 1,\ldots,k$, $d_i(\underset{\sim}{x})$ is increasing in x_i while all other variables x_j are fixed, and decreasing in x_j for each $j \neq i$ while all other variables are fixed.

Note that $\varphi_i(x)$ is the Bayes estimator of the binomial parameter p_i under the squared error loss given that $X_i = x$ is observed. It is also easy to see that $\varphi_i(x)$ is increasing in x for $x = 0,1,\ldots,N$.

Definition 3.2. An estimator $\varphi(\cdot)$ is called a monotone estimator if $\varphi(x)$ is an increasing function of x.

By the monotone property of the Bayes estimators $\varphi_i(x)$, $i = 1,\ldots,k$, one can see that the Bayes selection rule d_G is a monotone selection rule.

Under the squared error loss, the problem of estimating the binomial parameter p_i is a monotone estimation problem. By Theorem 8.7 of Berger (1980), for a monotone estimation problem, the class of monotone decision rules form an essentially complete class. With this consideration, it is reasonable to require that the concerned estimators $\{\varphi_{in}(x)\}$ possess the above monotone property.

In the literature, Robbins (1956) and Vardeman (1978), among others, proposed some estimators for $\varphi_i(x)$. Those estimators are consistent in that they converge to $\varphi_i(x)$ in probability. However, they do not possess the monotone property. We now propose some monotone estimators.

For each $i = 1,\ldots,k$, $n = 1,2,\ldots,$ and $x = 0,1,\ldots,N$, define

$$(3.1) \qquad f_{in}(x) = \frac{1}{n}\sum_{j=1}^{n} I_{\{x\}}(X_{ij}) + n^{-1};$$

$$(3.2) \qquad W_{in}(x) = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}I_{\{x\}}(X_{ij}) + n^{-1};$$

where $I_A(\cdot)$ denotes the indicator function of the set A. Also, let $V_{ij} = X_{ij} + Y_{ij}$ for each $i = 1,\ldots,k$ and $j = 1,2,\ldots$.
Define

$$(3.3) \qquad \tilde{W}_{in}(x) = \left\{\left[\frac{x+1}{n(N+1)}\sum_{j=1}^{n} I_{\{x+1\}}(V_{ij})\right]\wedge\left[\frac{1}{n}\sum_{j=1}^{n} I_{\{x\}}(X_{ij})\right]\right\} + n^{-1},$$

where $a \wedge b = \min\{a,b\}$. Let

$$(3.4) \qquad \varphi_{in}(x) = W_{in}(x)/f_{in}(x);$$

$$(3.5) \qquad \tilde{\varphi}_{in}(x) = \tilde{W}_{in}(x)/f_{in}(x);$$

and, for each $0 \le x \le N$, define

$$(3.6) \qquad \varphi^*_{in}(x) = \max_{0\le s\le x}\ \min_{s\le t\le N}\left\{\sum_{y=s}^{t}\varphi_{in}(y)/(t-s+1)\right\};$$

$$(3.7) \qquad \tilde{\varphi}^*_{in}(x) = \max_{0\le s\le x}\ \min_{s\le t\le N}\left\{\sum_{y=s}^{t}\tilde{\varphi}_{in}(y)/(t-s+1)\right\}.$$

Note that by (3.6) and (3.7), both $\varphi^*_{in}(x)$ and $\tilde{\varphi}^*_{in}(x)$ are increasing in x. We propose $\varphi^*_{in}(x)$ (or $\tilde{\varphi}^*_{in}(x)$) as an estimator of $\varphi_i(x)$. Let

$$(3.8) \qquad A^*_n(\underset{\sim}{x}) = \{i \mid \varphi^*_{in}(x_i) = \max_{1\le j\le k} \varphi^*_{jn}(x_j)\};$$

$$(3.9) \qquad \tilde{A}^*_n(\underset{\sim}{x}) = \{i \mid \tilde{\varphi}^*_{in}(x_i) = \max_{1\le j\le k} \tilde{\varphi}^*_{jn}(x_j)\}.$$

Two selection rules $d^*_n = (d^*_{1n},\ldots,d^*_{kn})$ and $\tilde{d}^*_n = (\tilde{d}^*_{1n},\ldots,\tilde{d}^*_{kn})$ analogous to the Bayes selection rule d_G are proposed as follows:

For each i = 1,...,k, let

$$(3.10) \qquad d^*_{in}(\underset{\sim}{x}) = \begin{cases} |A^*_n(\underset{\sim}{x})|^{-1} & \text{if } i \in A^*_n(\underset{\sim}{x}); \\ 0 & \text{otherwise;} \end{cases}$$

and

$$(3.11) \qquad \tilde{d}^*_{in}(\underset{\sim}{x}) = \begin{cases} |\tilde{A}^*_n(\underset{\sim}{x})|^{-1} & \text{if } i \in \tilde{A}^*_n(\underset{\sim}{x}); \\ 0 & \text{otherwise.} \end{cases}$$

Due to the monotone property of the estimators $\{\varphi^*_{in}(x_i); i = 1,\ldots,k\}$ and $\{\tilde{\varphi}^*_{in}(x_i); i = 1,\ldots,k\}$, one can see that d^*_n and $\tilde{d}^*_n$ are both monotone selection rules.

4. Asymptotic Optimality of the Monotone Estimators

In this section, we study the asymptotic optimality property of the estimators $\varphi^*_{in}(x)$ and $\tilde{\varphi}^*_{in}(x)$. Under the squared error loss, $\varphi_i(x)$ is the Bayes estimator of p_i. The associated Bayes risk is

$$(4.1) \qquad R_i(G_i) = E[(P_i - \varphi_i(X_i))^2].$$

Let $\psi_i(\cdot)$ be any estimator of p_i with the associated Bayes risk $R_i(G_i, \psi_i)$. Then,

(4.2) $$R_i(G_i,\psi_i) - R_i(G_i) = E[(\psi_i(X_i) - \varphi_i(X_i))^2].$$

Let $\{\psi_{in}(x; (X_{i1},Y_{i1}),\ldots,(X_{in},Y_{in})) \equiv \psi_{in}(x)\}$ be a sequence of empirical Bayes estimators based on $(x; (X_{i1},Y_{i1}),\ldots,(X_{in},Y_{in}))$.

Definition 4.1. A sequence of empirical Bayes estimators $\{\psi_{in}\}_{n=1}^{\infty}$ is said to be asymptotically optimal at least of order α_n relative to the prior G_i if $R_i(G_i,\psi_{in}) - R_i(G_i) \le O(\alpha_n)$ as $n \to \infty$ where $\{\alpha_n\}$ is a sequence of positive values satisfying $\lim_{n\to\infty} \alpha_n = 0$.

Theorem 4.1. Let $\{\varphi^*_{in}\}$ and $\{\tilde{\varphi}^*_{in}\}$ be the sequences of epirical Bayes estimators defined in (3.6) and (3.7), respectively. Then,

$$R_i(G_i,\varphi^*_{in}) - R_i(G_i) \le O(n^{-1})$$

and $$R_i(G_i,\tilde{\varphi}^*_{in}) - R_i(G_i) \le O(n^{-1}).$$

The following lemmas are useful in presenting a concise proof of Theorem 4.1.

Lemma 4.1. Let Z be a random variable and z be a real number such that $-\infty \le a \le Z,\ z \le b \le \infty$. Then, for any $s > 0$,

$$E[|Z-z|^s] = \int_0^{z-a} st^{s-1}P\{Z-z < -t\}dt + \int_0^{b-z} st^{s-1}P\{Z-z > t\}dt,$$

provided that the expectation exists.

Proof: Straightforward computation.

Lemma 4.2. For the estimators φ_{in} and φ^*_{in} defined in (3.4) and (3.6), respectively, we have

a) $\varphi^*_{in}(0) \le \varphi_{in}(0)$, $\varphi^*_{in}(N) \ge \varphi_{in}(N)$.

b) For $1 \le x \le N-1$,

$\varphi^*_{in}(x) > \varphi_{in}(x)$ iff there is some $y < x$ such that $\varphi_{in}(y) > \varphi_{in}(x)$;

$\varphi^*_{in}(x) < \varphi_{in}(x)$ iff there is some $y > x$ such that $\varphi_{in}(y) < \varphi_{in}(x)$.

c) For $0 \le x \le N$,

$$P\{\varphi^*_{in}(x) - \varphi_i(x) > t\} \le \sum_{y=0}^{x} P\{\varphi_{in}(y) - \varphi_i(y) > t\};$$

$$P\{\varphi^*_{in}(x) - \varphi_i(x) < -t\} \le \sum_{y=x}^{N} P\{\varphi_{in}(y) - \varphi_i(y) < -t\}.$$

Proof: Parts a) and b) are straightforward from (3.6). Part c) is a result of parts a) and b) and an application of Bonferroni's inequality.

Remark 4.1. Lemma 4.2 is also true if φ_{in} and φ^*_{in} are replaced by $\tilde{\varphi}_{in}$ and $\tilde{\varphi}^*_{in}$, respectively.

Lemma 4.3. For $0 < t < 1-\varphi_i(x)$ and $0 \le y \le x$,

a) $P\{\varphi_{in}(y) - \varphi_i(y) > t\} \le \exp\{-2na_1^2(t,y,n,i)\}$; and

b) $P\{\tilde{\varphi}_{in}(y) - \varphi_i(y) > t\} \le \exp\{-\frac{n}{2} a_1^2(t,y,n,i)\}$,

if $t > b(n,y,i)$, where $b(n,y,i) = (1-\varphi_i(y))n^{-1}/(f_i(y)+n^{-1})$ and $a_1(t,y,n,i) = t(f_i(y) + n^{-1}) - n^{-1}(1-\varphi_i(y))$.

For $0 < t < \varphi_i(x)$ and $x \le y \le N$,

c) $P\{\varphi_{in}(y) - \varphi_i(y) < -t\} \le \exp\{-2na_2^2(t,y,n,i)\}$; and

d) $P\{\tilde{\varphi}_{in}(y) - \varphi_i(y) < -t\} \le 2\exp\{-\frac{n}{2} a_2^2(t,y,n,i)\}$, where $a_2(t,y,n,i) = -t(f_i(y) + n^{-1}) - n^{-1}(1-\varphi_i(y))$.

Proof: Here we prove part a) only. Other parts follow by a similar reasoning.

For $0 < t < 1-\varphi_i(x)$ and $0 \le y \le x$, by (3.1), (3.2), (3.4) and the fact that $\varphi_i(y) = W_i(x)/f_i(x)$, following a straightforward computation, one can obtain

$$\begin{aligned} &P\{\varphi_{in}(y) - \varphi_i(y) > t\} \\ &= P\{W_{in}(y) - (\varphi_i(y) + t)f_{in}(y) > 0\} \\ &= P\Big\{\frac{1}{n}\sum_{j=1}^{n} I_{\{y\}}(X_{ij})[Y_{ij} - \varphi_i(y) - t] + tf_i(y) > a_1(t,y,n,i)\Big\}. \end{aligned} \tag{4.3}$$

Note that $I_{\{y\}}(X_{ij})[Y_{ij} - \varphi_i(y) - t]$, $j = 1,2,\ldots,n$ are i.i.d., $-\varphi_i(y) - t \le I_{\{y\}}(X_{ij})[Y_{ij} - \varphi_i(y) - t] \le 1 - \varphi_i(y) - t$ for all j, and $E[I_{\{y\}}(X_{ij})[Y_{ij} - \varphi_i(y) - t]] = -tf_i(y)$. Also, $a_1(t,y,n,i) > 0$ iff $t > b(n,y,i)$. Hence, by (4.3) and Theorem 2 of Hoeffding (1963), $P\{\varphi_{in}(y) - \varphi_i(y) > t\} \le \exp\{-2na_1^2(t,y,n,i)\}$ if $t > b(n,y,i)$.

Remark 4.2. Lemma 4.3 is still true if the strict inequality $<$ ($>$) is replaced by $\le$ ($\ge$).

Lemma 4.4. For $0 \le y \le x$,

a) $\int_0^{1-\varphi_i(x)} tP\{\varphi_{in}(y) - \varphi_i(y) > t\}dt \le O(n^{-1})$; and

b) $\int_0^{1-\varphi_i(x)} tP\{\tilde{\varphi}_{in}(y) - \varphi_i(y) > t\}dt \le O(n^{-1})$.

For $x \le y \le N$,

c) $\displaystyle\int_0^{\varphi_i(x)} tP\{\varphi_{in}(x) - \varphi_i(y) < -t\}dt \le O(n^{-1})$; and

d) $\displaystyle\int_0^{\varphi_i(x)} tP\{\varphi_{in}(y) - \varphi_i(y) < -t\}dt \le O(n^{-1})$.

Proof: We prove part a) only.

Case 1. As $b(n,y,i) \ge 1 - \varphi_i(x)$, then

$$\int_0^{1-\varphi_i(x)} tP\{\varphi_{in}(y) - \varphi_i(y) > t\}dt$$

$$\le \int_0^{b(n,y,i)} t\ dt$$

$$= b^2(n,y,i)/2$$

$$= O(n^{-2}).$$

Case 2. As $b(n,y,i) < 1 - \varphi_i(x)$, then, by Lemma 4.3.a) and a direct computation,

$$\int_0^{1-\varphi_i(x)} tP\{\varphi_{in}(y) - \varphi_i(y) > t\}dt$$

$$\le \int_0^{b(n,y,i)} t\ dt + \int_{b(n,y,i)}^{1-\varphi_i(x)} tP\{\varphi_{in}(y) - \varphi_i(y) > t\}dt$$

$$\le O(n^{-2}) + O(n^{-1})$$

$$= O(n^{-1}).$$

Proof of Theorem 4.1.

By (4.2),

$$\begin{aligned} 0 &\le R_i(G_i, \varphi^*_{in}) - R_i(G_i) \\ &= E[(\varphi^*_{in}(X) - \varphi_i(X))^2] \qquad (4.4) \\ &= \sum_{x=0}^{N} E[(\varphi^*_{in}(X) - \varphi_i(X))^2 | X = x] f_i(x). \end{aligned}$$

By Lemmas 4.1 ~ 4.3 and the fact that $0 \le \varphi^*_{in}(x)$, $\varphi_i(x) \le 1$, one can obtain that

$$\begin{aligned} &E[(\varphi^*_{in}(X) - \varphi_i(X))^2 | X = x] \\ &= \int_0^{\varphi_i(x)} 2tP\{\varphi^*_{in}(x) - \varphi_i(x) < -t\}dt \\ &\quad + \int_0^{1-\varphi_i(x)} 2tP\{\varphi^*_{in}(x) - \varphi_i(x) > t\}dt \qquad (4.5) \\ &\le \sum_{y=x}^{N} \int_0^{\varphi_i(x)} 2tP\{\varphi_{in}(y) - \varphi_i(y) < -t\}dt \\ &\quad + \sum_{y=0}^{x} \int_0^{1-\varphi_i(x)} 2tP\{\varphi_{in}(y) - \varphi_i(y) > t\}dt. \end{aligned}$$

Then, by Lemma 4.4, (4.4), (4.5) and the fact that N is a finite number, therefore, $R_i(G_i, \varphi^*_{in}) - R_i(G_i) \le O(n^{-1})$.

The similar claim for $\tilde{\varphi}^*_{in}$ is established on the same lines.

5. Asymptotic Optimality of the Selection Rules

Let $\{d_n\}_{n=1}^{\infty}$ be a sequence of empirical Bayes selection rules relative to the prior distribution G. Since the Bayes rule d_G achieves the minimum Bayes risk $r(G)$, $r(G,d_n) - r(G) \geq 0$ for all $n = 1,2,\ldots$. Thus, the nonnegative difference $r(G,d_n) - r(G)$ is used as a measure of the optimality of the sequence of empirical Bayes rules $\{d_n\}_{n=1}^{\infty}$.

Definition 5.1. The sequence of empirical Bayes rules $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal at least of order β_n relative to the prior G if $r(G,d_n) - r(G) \leq O(\beta_n)$ as $n \to \infty$ where $\{\beta_n\}$ is a sequence of positive numbers such that $\lim_{n\to\infty} \beta_n = 0$.

For each $\underset{\sim}{x} \in \mathfrak{X}$, let $A(\underset{\sim}{x})$ be that defined in (2.3) and let $B(\underset{\sim}{x}) = \{1,\ldots,k\} - A(\underset{\sim}{x})$. Thus, for each $\underset{\sim}{x} \in \mathfrak{X}$, $\varphi_i(x_i) > \varphi_j(x_j)$ for $i \in A(\underset{\sim}{x})$ and $j \in B(\underset{\sim}{x})$. Let $\epsilon = \min_{\underset{\sim}{x}\in\mathfrak{X}} \{\varphi_i(x_i) - \varphi_j(x_j) |$ $i \in A(\underset{\sim}{x}),\ j \in B(\underset{\sim}{x})\}$. Hence, $\epsilon > 0$ since $\mathfrak{X}$ is a finite space. Then,

$$\begin{aligned} 0 &\leq r\,(G,d_n^*) - r(G) \\ &\leq \sum_{\underset{\sim}{x}\in\mathfrak{X}} P\{\max_{i\in A(\underset{\sim}{x})} \varphi_{in}^*(x_i) \leq \max_{j\in B(\underset{\sim}{x})} \varphi_{jn}^*(x_j)\} \qquad (5.1) \\ &\leq \sum_{\underset{\sim}{x}\in\mathfrak{X}} \sum_{i\in A(\underset{\sim}{x})} \sum_{j\in B(\underset{\sim}{x})} P\{\varphi_{in}^*(x_i) \leq \varphi_{jn}^*(x_j)\}. \end{aligned}$$

Now, for each $\underset{\sim}{x} \in \mathfrak{X}$, $i \in A(\underset{\sim}{x})$, $j \in B(\underset{\sim}{x})$,

$$P\{\varphi^*_{in}(x_i) \le \varphi^*_{jn}(x_j)\}$$

$$= P\{[\varphi^*_{in}(x_i)-\varphi_i(x_i)]-[\varphi^*_{jn}(x_j)-\varphi_j(x_j)] \le \varphi_j(x_j)-\varphi_i(x_i)\}$$

(5.2)

$$\le P\{[\varphi^*_{in}(x_i)-\varphi_i(x_i)]-[\varphi^*_{jn}(x_j)-\varphi_j(x_j)] \le -\varepsilon\}$$

$$\le P\{\varphi^*_{in}(x_i)-\varphi_i(x_i) \le -\varepsilon/2\} + P\{\varphi^*_{jn}(x_j)-\varphi_j(x_j) \ge \varepsilon/2\}.$$

In (5.2), the first inequality is due to the definition of ε. From (2.3), it suffices to consider the asymptotic behavior of the probabilities $P\{\varphi^*_{jn}(x_j)-\varphi_j(x_j) \ge \varepsilon/2\}$ and $P\{\varphi^*_{in}(x_i)-\varphi_i(x_i) \le -\varepsilon/2\}$.

Let $c_1 = \min_{1\le i\le k} \min_{0\le y\le N} \{\varepsilon^2 f_i^2(y)/2\}$. Then $c_1 > 0$. From the definitions of ε and $b(n,y,i)$, we see that, for sufficiently large n, $\varepsilon > 2 \max_{1\le i\le k} \max_{0\le y\le N} \{b(n,y,i)\}$. Therefore, by Lemma 4.2 c) and remark 4.2, for n large enough,

$$P\{\varphi^*_{in}(x_i) - \varphi_i(x_i) \ge \varepsilon/2\}$$

$$\le \sum_{y=0}^{x_i} P\{\varphi_{in}(y) - \varphi_i(y) \ge \varepsilon/2\}$$

(5.3)

$$\le \sum_{y=0}^{x_i} \exp\{-2na_1^2(\varepsilon/2,y,n,i)\}$$

$$\le O(\exp(-c_1 n)).$$

The last step of (5.3) follows from the fact that $\exp\{-2na_1^2(t,y,n,i)\} \le O(\exp(-c_1 n))$ for all $0 \le y \le N$ and $1 \le i \le k$, which is established easily by a straightforward computation and definitions of $a_1(\epsilon/2,y,n,i)$ and c_1.

Similarly, one can prove that

$$P\{\varphi^*_{in}(x_i) - \varphi_i(x_i) \leq -\varepsilon/2\}$$

(5.4) $$\leq \sum_{y=x_i}^{N} \exp\{-2na_2^2(\varepsilon/2, y, n, i)\}$$

$$\leq O(\exp(-c_1 n)).$$

Therefore, from (5.1) to (5.4), and the finiteness of the space $\mathcal{X}$, we have

$$0 \leq r(G, d_n^*) - r(G) \leq O(\exp(-c_1 n)).$$

Similarly, for the sequence of empirical Bayes selection rules $\{\tilde{\varphi}_n^*\}_{n=1}^{\infty}$, we can prove that $0 \leq r(G, \tilde{d}_n^*) - r(G) \leq 0(\exp(-c_2 n))$ for some $c_2 > 0$.

We now state these results as a theorem.

<u>Theorem 5.1</u>. Let $\{d_n^*\}_{n=1}^{\infty}$ and $\{\tilde{d}_n^*\}_{n=1}^{\infty}$ be the sequences of empirical Bayes selection rules defined in (3.10) and (3.11), respectively. Then,

$$r(G, d_n^*) - r(G) \leq O(\exp(-c_1 n)),$$

and

$$r(G, \tilde{d}_n^*) - r(G) \leq O(\exp(-c_2 n))$$

for some $c_i > 0$, $i = 1, 2$.

REFERENCES

Berger, J. O. (1980). Statistical Decision Theory. Springer-Verlag, New York.

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

Gupta, S. S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. J. Statist. Plan. Infer. 8 87-101.

Gupta, S. S. and Huang, D. Y. (1976). Subset selection procedures for the entropy function associated with the binomial populations. Sankhyā Ser. A 38 153-173.

Gupta, S. S. and Huang, D. Y. and Huang, W. T. (1976). On ranking and selection procedures and tests of homogeneity for binomial populations. Essays in Probability and Statistics (Eds. S. Ikeda et. al.), Shinko Tsusho Co. Ltd. Tokyo, Japan, 501-533.

Gupta, S. S. and Leu, L. Y. (1983). On Bayes and empirical Bayes rules for selecting good populations. Technical Report 83-37, Dept. of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S. S. and Liang, T. (1984). Empirical Bayes rules for selecting good binomial populations, will appear in The Proceedings of the Symposium on Adaptive Statistical Procedures and Related Topics (ed. J. Van Ryzin).

Gupta, S. S. and Sobel, M. (1960). Selecting a subset containing the best of several binomial populations. Contributions to Probability and Statistics (Eds. I. Olkin et. al.), Stanford University Press, California, 224-248.

Gupta, S. S. and McDonald, G. C. (1986). A statistical selection approach to binomial models. Journal of Quality Technology, 18, 103-115.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. Ann. Math. Statist. 34 13-30.

Robbins, H. (1956). An empirical Bayes approach to statistics. Proc. Third Berkeley Symp. Math. Statist. Probab. 1 157-163, University of California Press.

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

Robbins, H. (1983). Some thoughts on empirical Bayes estimation. Ann. Statist. 11 713-723.

Samuel, E. (1963). An empirical Bayes approach to the testing of certain parametric hypotheses. Ann. Math. Statist. 34 1370-1385.

Sobel, M. and Huyett, M. J. (1957). Selecting the best one of several binomial populations. Bell System Tech. J. 36 537-576.

Vardeman, S. B. (1978). Bounds on the empirical Bayes and compound risks of truncated versions of Robbins's estimator of a binomial parameter. J. Statist. Plan. Infer. 2 245-252.

BEFORE COMPLETING FORM

1. REPORT NUMBER Technical Report #86-13	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Empirical Bayes Rules for Selecting the Best Binomial Population		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Technical Report #86-13
7. AUTHOR(s) Shanti S. Gupta and TaChen Liang		8. CONTRACT OR GRANT NUMBER(s) N00014-84-C-0167
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 May 1986
		13. NUMBER OF PAGES 20
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 rule; Empirical Bayes Rules, Monotone Estimation, Monotone Selection Rules, Asymptotically Optimal, Rate of Convergence.

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

Consider k populations π_i, $i = 1,\ldots,k$, where an observation from π_i has binomial distribution with parameters N and p_i (unknown). Let $p_{[k]} = \max_{1 \le j \le k} p_j$. A population π_i with $p_i = p_{[k]}$ is called a best population. We are interested in selecting the best population. Let $\underset{\sim}{p} = (p_1,\ldots,p_k)$ and let i denote the index of the selected population.

(over)

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Under the loss function $\ell(\underset{\sim}{p}, i) = p_{[k]} - p_i$, this statistical selection problem is studied via empirical Bayes approach.

Some selection rules based on monotone empirical Bayes estimators of the binomial parameters are proposed. First, it is shown that, under the squared error loss, the Bayes risks of the proposed monotone empirical Bayes estimators converge to the related minimum Bayes risks with rates of convergence at least of order $O(n^{-1})$, where n is the number of accumulated past experiences at hand. Further, for the selection problem, the rates of convergence of the proposed selection rules are shown to be at least of order $O(\exp(-cn))$ for some $c > 0$.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)