SELECTING THE BEST BINOMIAL POPULATION:
PARAMETRIC EMPIRICAL BAYES APPROACH*

by
Shanti S. Gupta
Purdue University

TaeChen Liang
Wayne State University

Technical Report#88-1

Department of Statistics
Purdue University

February 1988
Revised August 1988

* This research was supported in part by the Office of Naval Research Contracts N00014-84-C-0167, N00014-88-K-0170 and NSF Grants DMS-8606964, DMS-8702620 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.
SELECTING THE BEST BINOMIAL POPULATION:
PARAMETRIC EMPIRICAL BAYES APPROACH*

by

Shanti S. Gupta
Purdue University
West Lafayette, IN 47907

TaChen Liang
Wayne State University
Detroit, MI 48202

Abstract

Consider \( k \) populations \( \pi_1, \ldots, \pi_k \), where an observation from population \( \pi_i \) has a binomial distribution with parameters \( N \) and \( p_i \) (unknown). Let \( p[k] = \max_{1 \leq j \leq k} p_j \). A population \( \pi_i \) with \( p_i = p[k] \) is called a best population. We are interested in selecting the best population. Let \( \mathbf{p} = (p_1, \ldots, p_k) \) and let \( a \) denote the index of the selected population. Under the loss function \( L(p, a) = p[k] - p_a \), this statistical selection problem is studied via a parametric empirical Bayes approach. It is assumed that the binomial parameters \( p_i, i = 1, \ldots, k \), follow some conjugate beta prior distributions with unknown hyperparameters. Under the binomial-beta statistical framework, an empirical Bayes selection rule is proposed. It is shown that the Bayes risk of the proposed empirical Bayes selection rule converges to the corresponding minimum Bayes risk with rates of convergence at least of order \( O(\exp(-cn)) \) for some positive constant \( c \), where \( n \) is the number of accumulated past experience (observations) at hand.

* This research was supported in part by the Office of Naval Research Contracts N00014-84-C-0167, N00014-88-K-0170 and NSF Grants DMS–8606964, DMS–8702620 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.
AMS 1980 Subject Classification: Primary 62F07; Secondary 62C12.

*Key words and phrases:* Asymptotically optimal, Bayes rules, empirical Bayes rules, best population, binomial-beta model, rate of convergence.
1. Introduction

Consider $k$ populations $\pi_1, \ldots, \pi_k$, where an observation from $\pi_i$ has a binomial distribution with parameters $N$ and $p_i$ (unknown). Let $p_{[1]} \leq \ldots \leq p_{[k]}$ denote the ordered values of the parameters $p_1, \ldots, p_k$. It is assumed that the exact pairing between the ordered and the unordered parameters is unknown. Any population associated with $p_{[k]}$ is considered as the best population. A number of statistical procedures based on single sampling or sequential sampling rules have been studied in the literature for selecting the best binomial population. Sobel and Huyett (1957) have studied a fixed sample procedure through indifference zone approach. Gupta and Sobel (1960), Gupta and Huang (1976), and Gupta, Huang and Huang (1976) have studied this selection problem using a subset selection approach. Bechhofer and Kulkarni (1982) and Kulkarni and Jennison (1986) have studied a sequential selection procedure. Recently, Abghalous and Miescke (1987) have studied Bayes selection procedures under "0–1" loss and some linear loss for certain priors (also, see Gupta and McDonald (1986) for some new work and an application).

Now, consider a situation in which one will be repeatedly dealing with the same selection problem independently. 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 (1956, 1964 and 1983). Recently, Gupta and Liang (1987) have studied the problem of selecting the best binomial population by using the nonparametric empirical Bayes approach. They assume that the form of the prior distribution is completely unknown. However, in many
cases, an experimenter may have some prior information about the parameters of interest, and he would like to use this information to make appropriate decisions. Usually, it is suggested that the prior information is quantified through a class of subjectively plausible priors. In view of this situation, in this paper, it is assumed that the binomial parameters \( p_i, \ i = 1, \ldots, k \), follow some conjugate beta prior distributions with unknown hyperparameters. Under the binomial-beta statistical framework, an empirical Bayes selection rule is proposed. It is shown that the proposed empirical Bayes selection rules possess the following asymptotic optimality property: The Bayes risk of the proposed empirical Bayes selection rules converges to the minimum Bayes risk with rate of convergence at least of order \( O(\exp(-cn)) \) for some positive constant \( c \), where \( n \) is the number of accumulated past experience (observations) at hand.

2. A Bayesian Formulation of the Selection Problem

Let \( \pi_1, \ldots, \pi_k \) denote \( k \) populations, each consisting of \( N \) trials. For each \( i = 1, \ldots, k \), let \( p_i \) be the probability of success for each independent trial in \( \pi_i \), and let \( X_i \) denote the number of successes among the associated \( N \) trials. Then, conditional on \( p_i, \) \( X_i \) is binomially distributed with probability function \( f_i(x_i|p_i) = \binom{N}{x_i} p_i^{x_i} (1 - p_i)^{N-x_i}, \ x_i = 0, 1, \ldots, N \). Let \( f(\mathbf{x}|\mathbf{p}) = \prod_{i=1}^{k} f_i(x_i|p_i), \) where \( \mathbf{x} = (x_1, \ldots, x_k) \) and \( \mathbf{p} = (p_1, \ldots, p_k) \). For each \( \mathbf{p}, \) let \( p_{[i]} \leq \cdots \leq p_{[k]} \) be the ordered values of the parameters \( p_1, \ldots, p_k \). It is assumed that the exact pairing between the ordered and the unordered parameters is unknown. A population \( \pi_i \) with \( p_i = p_{[k]} \) is considered as a best population. Our goal is to derive an empirical Bayes rule to select the best binomial population.
Let $\Omega = \{p|p = (p_1, \ldots, p_k), p_i \in (0, 1), i = 1, \ldots, k\}$ be the parameter space. It is assumed that the parameter $p$ has a prior distribution $G$ with a joint probability density function $g(p) = \prod_{i=1}^{k} g_i(p_i)$, where for each $i = 1, \ldots, k$,

$$g_i(p_i) = \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i \mu_i) \Gamma(\alpha_i (1-\mu_i))} p_i^{\alpha_i \mu_i - 1} (1 - p_i)^{\alpha_i (1-\mu_i) - 1}, \quad (2.1)$$

and where $0 < \mu_i < 1, \alpha_i > 0$, both $\mu_i$ and $\alpha_i$ are unknown. Thus, we call the statistical model under study as a binomial-beta model.

Let $A = \{i|i = 1, \ldots, k\}$ be the action space. When action $i$ is taken, it means that population $\pi_i$ is selected as a best population. For the parameter $p$ and action $i$, the loss function $L(p, i)$ is defined as

$$L(p, i) = p_{[k]} - p_i, \quad (2.2)$$

the difference between the best and the selected population.

Let $\bar{X} = (X_1, \ldots, X_k)$ and let $\mathcal{X}$ be the sample space generated by $\bar{X}$. A selection rule $d = (d_1, \ldots, d_k)$ is a mapping from the sample space $\mathcal{X}$ to $[0,1]^k$ such that for each observation $\bar{x} = (x_1, \ldots, x_k)$ in $\mathcal{X}$, the function $d(\bar{x}) = (d_1(\bar{x}), \ldots, d_k(\bar{x}))$ satisfies that $0 \leq d_i(\bar{x}) \leq 1, \ i = 1, \ldots, k$, and $\sum_{i=1}^{k} d_i(\bar{x}) = 1$. Note that $d_i(\bar{x}), \ i = 1, \ldots, k$, is the probability of selecting the population $\pi_i$ as the best population when $\bar{x}$ is observed.

Let $D$ be the class of all selection rules defined above. For each $d \in D$, let $r(G, d)$ denote the associated Bayes risk. Then, $r(G) = \inf_{d \in D} r(G, d)$ is the minimum Bayes risk. From (2.1) and (2.2), the Bayes risk associated with the selection rule $d$ is:

$$r(G, d) = \int_{\Omega} \sum_{\bar{x} \in \mathcal{X}} \sum_{i=1}^{k} L(p, i) d_i(\bar{x}) f(\bar{x}|p) dG(p)$$

$$= C - \sum_{\bar{x} \in \mathcal{X}} \left[ \sum_{i=1}^{k} d_i(\bar{x}) \phi_i(x_i) \right] f(\bar{x}), \quad (2.3)$$

3
where \( f(x) = \prod_{i=1}^{k} f_i(x_i), \) \( f_i(x_i) = \int_{0}^{1} f_i(x_i|p)g_i(p)dp, \) \( \varphi_i(x_i) = E[p_i|x_i] = (x_i + \alpha_i\mu_i)/(N + \alpha_i), \) the posterior mean of \( p_i \) given \( X_i = x_i, \) \( C = \sum_{\bar{x} \in \mathcal{X}} \int_{\Omega} p[k]g(\bar{y}|\bar{x})d\bar{y}f(\bar{x}), \) being a constant, and \( g(\bar{y}|\bar{x}) = \prod_{i=1}^{k} g_i(p_i|x_i), \) where \( g_i(p_i|x_i) = \frac{\Gamma(N+\alpha_i)}{\Gamma(x_i+\alpha_i\mu_i)\Gamma(N-x_i+\alpha_i(1-\mu_i))}p_i^{x_i+\alpha_i\mu_i-1}(1-p_i)^{N-x_i+\alpha_i(1-\mu_i)-1} \) is the posterior density function of \( p_i \) given \( X_i = x_i. \)

For each \( \bar{x} \in \mathcal{X}, \) let

\[
A(\bar{x}) = \{i | x_i + \alpha_i\mu_i = \max_{1 \leq j \leq k} \frac{x_j + \alpha_j\mu_j}{N + \alpha_j}\}. \tag{2.4}
\]

Thus, a randomized Bayes selection rule, say \( \hat{d}_G = (d_{1G}, \ldots, d_{kG}) \), can be obtained as follows:

\[
d_iG(\bar{x}) = \begin{cases} |A(\bar{x})|^{-1} & \text{if } i \in A(\bar{x}), \\ 0 & \text{otherwise}. \end{cases} \tag{2.5}
\]

where \( |A(\bar{x})| \) denotes the cardinality of the set \( A(\bar{x}). \)

It should be noted that nonrandomized Bayes selection rules do also exist. For example, let \( I(\bar{x}) = \min\{i | i \in A(\bar{x})\}. \) Then, a nonrandomized Bayes selection rule, say \( \check{d}_G = (\check{d}_{1G}, \ldots, \check{d}_{kG}) \), can be obtained as follows:

\[
\check{d}_iG(\bar{x}) = \begin{cases} 1 & \text{if } i = I(\bar{x}), \\ 0 & \text{otherwise}. \end{cases}
\]

Note that the Bayes selection rule \( d_G \) is dependent of the values of the parameters \( (\alpha_i, \mu_i), i = 1, \ldots, k. \) However, since the values of these parameters are unknown, it is impossible to apply the Bayes selection rule \( d_G \) for the selection problem at hand. As we mentioned above, we study this selection problem via empirical Bayes approach.
3. The Proposed Empirical Bayes Selection Rule

For each $i = 1, \ldots, k$, at stage $j$, consider $N$ independent trials from population $\pi_i$. Let $X_{ij}$ stand for the number of successes among the $N$ trials. Let $p_{ij}$ stand for the probability of a success for each of the $N$ trials. Let $P_j = (P_{1j}, \ldots, P_{kj})$. We assume that $P_j, j = 1, 2, \ldots$ are iid with a prior density $g(p) = \prod_{i=1}^{k} g_i(p_i)$, where $g_i(p_i), i = 1, \ldots, k$, are given by (2.1). Conditional on $P_{ij} = p_{ij}$, $X_{ij} | p_{ij} \sim B(N, p_{ij})$. Let $X_j = (X_{1j}, \ldots, X_{kj})$ denote the random observations at the $j$th stage, $j = 1, \ldots, n$. We also let $X_{n+1} = X = (X_1, \ldots, X_k)$ denote the random observation at the present stage.

Under the binomial-beta statistical model, we have, for each $i = 1, \ldots, k$,

\[
\begin{align*}
E[X_i/N] &= \mu_i \\
E[(X_i/N)^2] &= \mu_i/N + (\alpha_i \mu_i + 1)\mu_i(N-1)/(N(\alpha_i+1)) \equiv \mu_{i2} \text{ (say).}
\end{align*}
\]

From (3.1), through direct computation, the parameter $\alpha_i$ can be written as $\alpha_i = B_i/A_i$, where

\[
\begin{align*}
B_i &= \mu_i - \mu_{i2} \\
A_i &= \mu_{i2} - \mu_i N^{-1} + \mu_i^2 N^{-1} - \mu_i^2.
\end{align*}
\]

Note that under the binomial-beta model, $0 \leq (X_i/N)^2 \leq X_i/N \leq 1$ and therefore $B_i > 0$ since $X_i$ is a non-degenerate random variable. Also, $A_i > 0$ since $\alpha_i > 0$. Thus, $\mu_i$ and $\mu_{i2}$ satisfy the following inequalities: $\mu_i N^{-1} - \mu_i^2 N^{-1} + \mu_i^2 < \mu_{i2} < \mu_i$. From (3.2), $\alpha_i$ can be viewed as a function of $\mu_i$ and $\mu_{i2}$ for $\mu_i \in (0, 1)$ and $\mu_{i2} \in (\mu_i N^{-1} - \mu_i^2 N^{-1} + \mu_i^2, \mu_i)$.

For each fixed $\mu_i$, $\alpha_i$ is decreasing in $\mu_{i2}$ and $\lim_{\mu_{i2} \to \mu_i} \alpha_i = 0$, $\lim_{\mu_{i2} \to \alpha_i} \alpha_i = \infty$, where $\alpha_i = \mu_i N^{-1} - \mu_i^2 N^{-1} + \mu_i^2$. 

5
Let \( \mu_{in} \) and \( \mu_{i2n} \) be the moment estimators of \( \mu_i \) and \( \mu_{i2} \), respectively, based on the \( n \) past observations at hand. That is,

\[
\begin{aligned}
\mu_{in} &= \frac{1}{n} \sum_{j=1}^{n} X_{ij}/N, \\
\mu_{i2n} &= \frac{1}{n} \sum_{j=1}^{n} (X_{ij}/N)^2.
\end{aligned}
\] (3.3)

Also, let

\[
\begin{aligned}
A_{in} &= \mu_{i2n} - \mu_{in}N^{-1} + \mu_{in}^2N^{-1} - \mu_{in}^2, \\
B_{in} &= \mu_{in} - \mu_{i2n}.
\end{aligned}
\] (3.4)

Then, we propose some empirical Bayes estimators for the unknown parameter \( \alpha_i \) and the posterior mean \( \varphi_i(x_i) = (x_i + \alpha_i\mu_i)/(N + \alpha_i) \) as follows:

\[
\alpha_{in} = \begin{cases} 
B_{in}/A_{in} & \text{if } A_{in} > 0, \\
\infty & \text{otherwise};
\end{cases}
\] (3.5)

\[
\varphi_{in}(x_i) = \begin{cases} 
(x_i + \alpha_{in}\mu_{in})/(N + \alpha_{in}) & \text{if } \alpha_{in} < \infty, \\
\mu_{in} & \text{if } \alpha_{in} = \infty.
\end{cases}
\] (3.6)

We then propose an empirical Bayes selection rule \( d^*_n = (d^*_{1n}, \ldots, d^*_{kn}) \) for the selection problem under study, as follows:

For each \( \bar{x} \in \mathcal{X} \), let

\[
A^*_n(\bar{x}) = \{ i | \varphi_{in}(x_i) = \max_{1 \leq j \leq k} \varphi_{jn}(x_j) \}
\] (3.7)

and for each \( i = 1, \ldots, k \), let

\[
d^*_{in}(\bar{x}) = \begin{cases} 
|A^*_n(\bar{x})|^{-1} & \text{if } i \in A^*_n(\bar{x}), \\
0 & \text{otherwise}.
\end{cases}
\] (3.8)

We denote the associated Bayes risk of the proposed empirical Bayes selection rule \( d^*_n \) by \( r(G, d^*_n) \). Then, from (2.3),

\[
r(G, d^*_n) = C - \sum_{\bar{x} \in \mathcal{X}} \left[ k \sum_{i=1}^{k} d^*_{in}(\bar{x})\varphi_i(x_i) \right] f(\bar{x}).
\] (3.9)
1. Note that $\alpha_i = \infty \implies \text{Var}(P_i) = 0$, which means that the prior density $g_i(p_i)$ is degenerate at the point $p_i = \mu_i$. In this situation, the posterior mean $\varphi_i(x_i) = \mu_i$. Hence, it is reasonable to estimate $\varphi_i(x_i)$ by $\mu_{in}$ when $\alpha_{in} = \infty$. We consider the case where $\alpha_i = \infty$ as an extreme case for the family of beta distributions.

2. Definition 3.1. A selection rule $d = (d_1, \ldots, d_k)$ is said to be monotone if for each $i = 1, \ldots, k$, $d_i(x)$ is nondecreasing in $x_i$ while all the other variables $x_j$ are kept fixed, and nonincreasing in $x_j$ for each $j \neq i$ while all the other variables are kept fixed.

For the fixed past observations $X_1, \ldots, X_n$, we see from (3.6) that for each $i = 1, \ldots, k$, $\varphi_{in}(x_i)$, the estimator of the posterior mean $\varphi_i(x_i)$, is increasing in $x_i$. Thus, from (3.7) and (3.8), one can see that the proposed empirical Bayes selection rule $d_n^*$ possesses the monotonicity property.

4. Asymptotic Optimality of the Selection Rules $\{d_n^*\}$

Consider an empirical Bayes selection rule $d_n = (d_{1n}, \ldots, d_{kn})$. Let $r(G, d_n)$ be the associated Bayes risk. Then, $r(G, d_n) - r(G) \geq 0$ since $r(G)$ is the minimum Bayes risk. Thus $E[r(G, d_n)] - r(G) \geq 0$, where

$$E[r(G, d_n)] = C - \sum_{x \in \mathcal{X}} \left[ \sum_{i=1}^{k} E[d_{in}(x)|\varphi_i(x_i)] \right] f(x)$$

and the expectation $E[d_{in}(x)]$ is taken with respect to $(X_1, \ldots, X_n)$. The nonnegative difference $E[r(G, d_n)] - r(G)$ is always used as a measure of performance of the selection rule $d_n$.
Definition 4.1. A sequence of empirical Bayes rules \( \{d_n\}_{n=1}^{\infty} \) is said to be asymptotically optimal at least of order \( \beta_n \) relative to the unknown prior distribution \( G \) if \( E[r(G, d_n)] - r(G) \leq O(\beta_n) \) as \( n \to \infty \), where \( \{\beta_n\} \) is a sequence of positive values such that \( \lim_{n \to \infty} \beta_n = 0 \).

In order to investigate the asymptotic optimality of the empirical Bayes selection rules \( \{d_n^*\} \), we need the following lemmas.

Lemma 4.1. If random variables \( Y_1, \ldots, Y_n \) are iid such that \( a \leq Y_i \leq b, \ i = 1, \ldots, k \), then for each \( t > 0 \),

\[
P\{\overline{Y} - \mu \geq t\} \leq \exp\left\{-2nt^2/(b-a)^2\right\},
\]

where \( \overline{Y} = \frac{1}{n} \sum_{j=1}^{n} Y_j \), and \( \mu = E[\overline{Y}] \).

Proof: This lemma is a special case of Theorem 1 of Hoeffding (1963).

Lemma 4.2. Let \( \mu_i, \mu_{i2}, \mu_{in} \) and \( \mu_{i2n} \) be as defined in (3.1) and (3.3), respectively. Then, for any \( c > 0 \),

a) \( P\{\mu_{in} - \mu_i \leq -c\} \leq O(\exp(-2nc^2)) \),
b) \( P\{\mu_{in} - \mu_i \geq c\} \leq O(\exp(-2nc^2)) \),
c) \( P\{\mu_{i2n} - \mu_{i2} \leq -c\} \leq O(\exp(-2nc^2)) \) and
d) \( P\{\mu_{i2n} - \mu_{i2} \geq c\} \leq O(\exp(-2nc^2)) \).

Proof: Note that under the framework of the statistical model under consideration, \( X_{ij}/N, j = 1, \ldots, n \), are iid and \( 0 \leq X_{ij}/N \leq 1 \). Then, \( 0 \leq \mu_{in} = \frac{1}{n} \sum_{j=1}^{n} X_{ij}/N \leq 1 \). Thus, \( P\{\mu_{in} - \mu_i \geq c\} = 0 \) if \( \mu_i + c > 1 \), and \( P\{\mu_{in} - \mu_i \geq c\} \leq \exp\{-2nc^2\} \) if \( \mu_i + c \leq 1 \), which follows from Lemma 4.1. This completes the proof of part b).
The proof for the other inequalities are analogous and hence omitted.

**Lemma 4.3.** Let $A_i$, $B_i$, $A_{in}$ and $B_{in}$ be as given in (3.2) and (3.4), respectively. Then, for any $c > 0$, we have

a) $P\{A_{in} - A_i \leq -c\} \leq O(\exp(-nc^2/8)),$

b) $P\{A_{in} - A_i \geq c\} \leq O(\exp(-nc^2/8)),$

c) $P\{B_{in} - B_i \leq -c\} \leq O(\exp(-nc^2/2)),$ and

d) $P\{B_{in} - B_i \geq c\} \leq O(\exp(-nc^2/2)).$

**Proof:** The techniques used to prove these four inequalities are similar. Here, we give the proof of part a) only.

\[
P\{A_{in} - A_i \leq -c\}
= P\{\left(\frac{1}{N} - 1\right)(\mu_{in} + \mu_i) - \frac{1}{N}\}(\mu_{in} - \mu_i) \leq -c\}
\leq P\left\{\left[\left(\frac{1}{N} - 1\right)(\mu_{in} + \mu_i) - \frac{1}{N}\right](\mu_{in} - \mu_i) \leq -\frac{c}{2}\right\} + P\{\mu_{i2n} - \mu_{i2} \leq -\frac{c}{2}\}.
\]

Since $0 \leq \mu_{in} \leq 1$, $0 < \mu_i < 1$, and $N$ is a positive integer, then $0 > (\frac{1}{N} - 1)(\mu_{in} + \mu_i) - \frac{1}{N} \geq (\frac{1}{N} - 1)2 - \frac{1}{N} = \frac{1-2N}{N}$. Therefore,

\[
P\{\left[\left(\frac{1}{N} - 1\right)(\mu_{in} + \mu_i) - \frac{1}{N}\right](\mu_{in} - \mu_i) \leq -\frac{c}{2}\}
\leq P\{\frac{1-2N}{N}(\mu_{in} - \mu_i) \leq -\frac{c}{2}\}
= P\{\mu_{in} - \mu_i \geq \frac{Nc}{2(2N-1)}\}
\leq P\{\mu_{in} - \mu_i \geq \frac{c}{4}\}.
\]
Thus,
\[ P\{A_i - A_i \leq c\} \]
\[ \leq P\{c/2 \leq \mu_i \leq c/4\} + P\{\mu_i \geq c/4\} \]
\[ \leq P\{c/2 \leq \mu_i \leq -c/4\} + P\{\mu_i \geq c/4\} \]
\[ \leq O(\exp(-nc^2/8)), \]
which follows from Lemma 4.2.

For each \( \mathcal{X} \in \mathcal{X} \), let \( A(x) \) be as defined in (2.4), and let \( B(x) = \{1, 2, \ldots, k\}\). That is, \( B(x) \) is the set consisting of the indices of nonbest populations given \( X = x \). Thus, for each \( x \in \mathcal{X} \), \( i \in A(x) \), \( j \in B(x) \), \( \varphi_i(x_i) > \varphi_j(x_j) \). From (2.4) and (4.1), following straightforward computation and using the fact that \( 0 < \varphi_i(x_i) < 1 \), \( 0 < f(x) < 1 \), we see that for the empirical Bayes selection rule \( d_n^* \),
\[
0 \leq E[r(G, d_n^*)] - r(G)
\]
\[ \leq \sum_{x \in \mathcal{X}} \sum_{i \in A(x)} \sum_{j \in B(x)} P\{\varphi_i(x_i) \leq \varphi_j(x_j)\}. \tag{4.2} \]

Since the sample space \( \mathcal{X} \) is finite and for each \( x \in \mathcal{X} \), \( |A(x)| + |B(x)| = k \), therefore, it suffices to evaluate the asymptotic behavior of the probability \( P\{\varphi_i(x_i) \leq \varphi_j(x_j)\} \) where \( i \in A(x), j \in B(x) \). Now, for each \( x \in \mathcal{X} \), \( i \in A(x) \), \( j \in B(x) \),
\[
P\{\varphi_i(x_i) \leq \varphi_j(x_j)\}
\]
\[ = P\{\varphi_i(x_i) \leq \varphi_j(x_j) \text{ and } (\alpha_{i\infty} < \alpha_{jn} < \alpha_{jn < \alpha_{jn < \alpha_{jn}}})\} \tag{4.3} \]
\[ + P\{\varphi_i(x_i) \leq \varphi_j(x_j) \text{ and } (\alpha_{i\infty} = \alpha_{jn} = \alpha_{jn = \infty})\}. \]

Let
\[ a = \min\{A_i|i = 1, \ldots, k\} \tag{4.4} \]
where $A_i, \ i = 1, \ldots, k$, are defined in (3.2). Then, $a > 0$ since $A_i > 0$ for all $i = 1, \ldots, k$, and $k$ is a finite number.

**Lemma 4.4.** For each $x \in X, \ i \in A(x), \ j \in B(x)$,

$$P\{\varphi_{in}(x_i) \leq \varphi_{jn}(x_j) \text{ and } (\alpha_{in} = \infty, \text{ or } \alpha_{jn} = \infty)\} \leq O(\exp(-na^2/8)).$$

**Proof:** Note that

$$P\{\varphi_{in}(x_i) \leq \varphi_{jn}(x_j) \text{ and } (\alpha_{in} = \infty \text{ or } \alpha_{jn} = \infty)\}$$

$$\leq P\{\alpha_{in} = \infty\} + P\{\alpha_{jn} = \infty\}$$

$$= P\{A_{in} \leq 0\} + P\{A_{jn} \leq 0\}$$

$$= P\{A_{in} - A_i \leq -A_i\} + P\{A_{jn} - A_j \leq -A_j\}$$

$$\leq P\{A_{in} - A_i \leq -a\} + P\{A_{jn} - A_j \leq -a\}$$

$$\leq O(\exp(-na^2/8)),$$

which is obtained from Lemma 4.3.

For each $i = 1, \ldots, k$, and $n = 1, 2, \ldots$, let $C_i(x_i) = x_iA_i + B_i\mu_i$, $D_i = NA_i + B_i$, $C_{in}(x_i) = x_iA_{in} + B_{in}\mu_{in}$ and $D_{in} = NA_{in} + B_{in}$. Also, let

$$b = \min_{x \in X}\{C_i(x_i)D_j - C_j(x_j)D_i | i \in A(x), \ j \in B(x)\}. \quad (4.5)$$

Then, $b > 0$ which is a consequence of the definitions of the sets $A(x)$ and $B(x)$ and the fact that the sample space $X$ is a finite space. Thus, for $i \in A(x), \ j \in B(x)$,

$$P\{\varphi_{in}(x_i) \leq \varphi_{jn}(x_j) \text{ and } (\alpha_{in} < \infty \text{ and } \alpha_{jn} < \infty)\}$$

$$= P\{C_{in}(x_i)D_{jn} - C_{jn}(x_j)D_{in} \leq 0\}$$

$$\leq P\{[C_{in}(x_i)D_{jn} - C_{jn}(x_j)D_{in}] - [C_i(x_i)D_j - C_j(x_j)D_i] \leq -b\} \quad (4.6)$$

$$\leq P\{C_{in}(x_i)D_{jn} - C_i(x_i)D_j \leq -\frac{b}{2}\} + P\{C_{jn}(x_j)D_{in} - C_j(x_j)D_i \geq \frac{b}{2}\}.$$
Now,

\[ P\{C_{in}(x_i)D_{jn} - C_i(x_i)D_j \leq -\frac{b}{2} \} \]

\[ = P\{[C_{in}(x_i)D_{jn} - C_i(x_i)D_j] + [C_i(x_i)D_j - C_i(x_i)D_j] \leq -\frac{b}{2} \} \]

\[ \leq P\{C_{in}(x_i)[D_{jn} - D_j] \leq -\frac{b}{4} \} + P\{[C_i(x_i) - C_i(x_i)]D_j \leq -\frac{b}{4} \}. \]  (4.7)

Similarly,

\[ P\{C_{jn}(x_j)D_{in} - C_j(x_j)D_i \geq \frac{b}{2} \} \]

\[ \leq P\{C_{jn}(x_j)(D_{in} - D_i) \geq \frac{b}{4} \} + P\{[C_{jn}(x_j) - C_j(x_j)]D_i \geq \frac{b}{4} \}. \]  (4.8)

**Lemma 4.5.**

a) \( P\{C_{in}(x_i)[D_{jn} - D_j] \leq -\frac{b}{4} \} \leq O(\exp(-b(n))), \)

b) \( P\{C_{in}(x_i)[D_{jn} - D_j] \geq \frac{b}{4} \} \leq O(\exp(-b(n))), \)

c) \( P\{[C_{in}(x_i) - C_i(x_i)]D_j \leq -\frac{b}{4} \} \leq O(\exp(-b(n))), \) and

d) \( P\{[C_{in}(x_i) - C_i(x_i)]D_j \geq \frac{b}{4} \} \leq O(\exp(-b(n))), \) where \( b(n) = nb^2/(512N^2(N + 1)^2) \).

**Proof:** We prove part a) and c) only. Proofs for b) and d) are similar.

a)

\[ P\{C_{in}(x_i)[D_{jn} - D_j] \leq -\frac{b}{4} \} \]

\[ = P\{(x_iA_{in} + B_{in}\mu_{in})(NA_{jn} + B_{jn} - NA_j - B_j) \leq -\frac{b}{4} \} \]

\[ \leq P\{(N + 1)[(N(A_{jn} - A_j) + (B_{jn} - B_j)] \leq -\frac{b}{4} \} \]

(since \( 0 \leq x_iA_{in} + B_{in}\mu_{in} \leq N + 1 \))

\[ \leq P\{A_{jn} - A_j \leq -\frac{b}{8N(N + 1)} \} + P\{B_{jn} - B_j < -\frac{b}{8(N + 1)} \} \]

\[ \leq O(\exp(-\frac{nb^2}{512N^2(N + 1)^2})), \]
which follows from Lemma 4.3.

c) 
\[
P\{[C_{in}(x_i) - C_i(x_i)]D_j \leq -\frac{b}{4}\} \\
\leq P\{C_{in}(x_i) - C_i(x_i) \leq -\frac{b}{4(N+1)}\} (\text{since } 0 < D_j < N + 1) \\
= P\{x_i(A_{in} - A_i) + \mu_i(B_{in} - B_i) \leq -\frac{b}{4(N+1)}\} \\
\leq P\{x_i(A_{in} - A_i) \leq -\frac{b}{8(N+1)}\} + P\{\mu_i(B_{in} - B_i) \leq -\frac{b}{8(N+1)}\}
\]

where 
\[
P\{\mu_i(B_{in} - B_i) \leq -\frac{b}{8(N+1)}\} \\
\leq P\{B_{in} - B_i \leq -\frac{b}{8(N+1)}\} (\text{since } 0 < \mu_i < 1) \\
\leq O(\exp(-\frac{nb^2}{128(N+1)^2})),
\]

and 
\[
P\{x_i(A_{in} - A_i) \leq -\frac{b}{8(N+1)}\} = 0 \text{ if } x_i = 0
\]

and for \( x_i > 0 \), 
\[
P\{x_i(A_{in} - A_i) \leq -\frac{b}{8(N+1)}\} \\
\leq P\{A_{in} - A_i \leq -\frac{b}{8(N+1)}\} \\
\leq O(\exp(-\frac{nb^2}{512N^2(N+1)^2})).
\]

Thus, 
\[
P\{[C_{in}(x_i) - C_i(x_i)]D_j \leq -\frac{b}{4}\} \leq O(\exp(-b(n))).
\]

Therefore, from (4.6) to (4.8) and Lemma 4.5, we conclude that: For \( i \in A(\bar{x}), j \in B(\bar{x}) \),

\[
P\{\varphi_{in}(x_i) \leq \varphi_{jn}(x_j) \text{ and } (\alpha_{in} < \infty \text{ and } \alpha_{jn} < \infty)\} \leq O(\exp(-b(n))), \quad (4.9)
\]
where the expression at the right-hand-side of (4.9) is independent of the present observation $x$.

Now, by the finiteness of the sample space $\mathcal{X}$ and from (4.2), (4.9) and Lemma 4.4, we conclude the following theorem:

**Theorem 4.1.** Let $\{d_n^x\}$ be the sequence of empirical Bayes selection rules defined in Section 3. Then,

$$E[r(G, d_n^x)] - r(G) \leq O(\exp(-cn)),$$

where $c = \min\left(\frac{b^2}{512N^2(N+1)^2}, \frac{a^2}{8}\right) > 0$ and $a$ and $b$ are defined in (4.4) and (4.5), respectively.

**References**


16. SUPPLEMENTARY NOTATION

17. COSATI CODES

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Asymptotically optimal, Bayes rules, empirical Bayes rules, best population, binomial-beta model, rate of convergence.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
Consider $k$ populations $\pi_1, \ldots, \pi_k$, where an observation from population $\pi_i$ has a binomial distribution with parameters $N$ and $p_i$ (unknown). Let $p_i = \max_{1 \leq j \leq k} p_j$. A population $\pi_i$ with $p_i = p[k]$ is called a best population. We are interested in selecting the best population. Let $\underline{p} = (p_1, \ldots, p_k)$ and let $a$ denote the index of the selected population. Under the loss function $L(p, a) = p[k] - p_a$, this statistical selection problem is studied via a parametric empirical Bayes approach. It is assumed that the binomial parameters $p_i$, $i = 1, \ldots, k$, follow some conjugate beta prior distributions with unknown hyperparameters. Under the binomial-beta statistical framework, an empirical selection rule is proposed. It is shown that the Bayes risk of the proposed empirical Bayes selection rule converges to the corresponding minimum Bayes risk with rates of convergence at least of order $O(\exp(-cn))$ for some positive constant $c$.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED  SAME AS RPT.  DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Shanti S. Gupta

22b. TELEPHONE (Include Area Code)
317-494-6031
19. $c$, where $n$ is the number of accumulated past experience (observations) at hand.