SELECTING THE BEST BINOMIAL POPULATION: PARAMETRIC EMPIRICAL BAYES APPROACH*

by

Shanti S. Gupta
Purdue University

TaChen 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 π_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 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 a denote the index of the selected population. Under the loss function $L(\underset{\sim}{p}, a) = p_{[k]} - p_a$, 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.

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 π_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, Abughalous 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 π_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 $f(\underset{\sim}{x}|\underset{\sim}{p}) = \prod_{i=1}^{k} f_i(x_i|p_i)$, where $\underset{\sim}{x} = (x_1, \ldots, x_k)$ and $\underset{\sim}{p} = (p_1, \ldots, p_k)$. For each $\underset{\sim}{p}$, let $p_{[i]} \leq \ldots \leq p_{[k]}$ 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 π_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 $\Omega = \{\underset{\sim}{p} | \underset{\sim}{p} = (p_1, \ldots, p_k),\ p_i \in (0,1),\ i = 1, \ldots, k\}$ be the parameter space. It is assumed that the parameter $\underset{\sim}{p}$ has a prior distribution G with a joint probability density function $g(\underset{\sim}{p}) = \prod_{i=1}^{k} g_i(p_i)$, where for each $i = 1, \ldots, k$,

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

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

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 a best population. For the parameter $\underset{\sim}{p}$ and action i, the loss function $L(\underset{\sim}{p}, i)$ is defined as

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

the difference between the best and the selected population.

Let $\underset{\sim}{X} = (X_1, \ldots, X_k)$ and let $\mathcal{X}$ be the sample space generated by $\underset{\sim}{X}$. 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 $\underset{\sim}{x} = (x_1, \ldots, x_k)$ in $\mathcal{X}$, 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 the population π_i as the best population when $\underset{\sim}{x}$ 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:

$$\begin{aligned} r(G, d) &= \int_{\Omega} \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i=1}^{k} L(\underset{\sim}{p}, i) d_i(\underset{\sim}{x}) f(\underset{\sim}{x} | \underset{\sim}{p}) dG(\underset{\sim}{p}) \\ &= C - \sum_{\underset{\sim}{x} \in \mathcal{X}} \left[\sum_{i=1}^{k} d_i(\underset{\sim}{x}) \varphi_i(x_i) \right] f(\underset{\sim}{x}), \end{aligned} \tag{2.3}$$

where $f(\underset{\sim}{x}) = \prod_{i=1}^{k} f_i(x_i)$, $f_i(x_i) = \int_0^1 f_i(x_i|p)g_i(p)dp$, $\varphi_i(x_i) = E[p_i|x_i] = (x_i + \alpha_i\mu_i)/(N + \alpha_i)$, the posterior mean of p_i given $X_i = x_i$, $C = \sum_{\underset{\sim}{x} \in \mathcal{X}} \int_\Omega p_{[k]} g(\underset{\sim}{p}|\underset{\sim}{x}) d\underset{\sim}{p} f(\underset{\sim}{x})$, being a constant, and $g(\underset{\sim}{p}|\underset{\sim}{x}) = \prod_{i=1}^{k} g_i(p_i|x_i)$, 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 $\underset{\sim}{x} \in \mathcal{X}$, let

$$A(\underset{\sim}{x}) = \{i | \frac{x_i + \alpha_i\mu_i}{N + \alpha_i} = \max_{1 \le j \le k} \frac{x_j + \alpha_j\mu_j}{N + \alpha_j}\}. \tag{2.4}$$

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

$$d_{iG}(\underset{\sim}{x}) = \begin{cases} |A(\underset{\sim}{x})|^{-1} & \text{if } i \in A(\underset{\sim}{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

where $|A(\underset{\sim}{x})|$ denotes the cardinality of the set $A(\underset{\sim}{x})$.

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

$$\tilde{d}_{iG}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } i = I(\underset{\sim}{x}), \\ 0 & \text{otherwise.} \end{cases}$$

Note that the Bayes selection rule d_G is dependent of the values of the parameters (α_i, μ_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 π_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 $\underset{\sim}{P}_j = (P_{1j}, \ldots, P_{kj})$. We assume that $\underset{\sim}{P}_j$, $j = 1, 2, \ldots$ are iid with a prior density $g(\underset{\sim}{p}) = \prod_{i=1}^{k} g_i(p_i)$, 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 $\underset{\sim}{X}_j = (X_{1j}, \ldots, X_{kj})$ denote the random observations at the jth stage, $j = 1, \ldots, n$. We also let $\underset{\sim}{X}_{n+1} \equiv \underset{\sim}{X} = (X_1, \ldots, X_k)$ denote the random observation at the present stage.

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

$$\begin{cases} E[X_i/N] = \mu_i \\ E[(X_i/N)^2] = \mu_i/N + (\alpha_i\mu_i + 1)\mu_i(N-1)/(N(\alpha_i + 1)) \equiv \mu_{i2} \text{ (say).} \end{cases} \tag{3.1}$$

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

$$\begin{cases} B_i = \mu_i - \mu_{i2} \\ A_i = \mu_{i2} - \mu_i N^{-1} + \mu_i^2 N^{-1} - \mu_i^2. \end{cases} \tag{3.2}$$

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

Let μ_{in} and μ_{i2n} be the moment estimators of μ_i and μ_{i2}, respectively, based on the n past observations at hand. That is,

$$\begin{cases} \mu_{in} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}/N, \\ \mu_{i2n} = \frac{1}{n}\sum_{j=1}^{n} (X_{ij}/N)^2. \end{cases} \tag{3.3}$$

Also, let

$$\begin{cases} A_{in} = \mu_{i2n} - \mu_{in}N^{-1} + \mu_{in}^2 N^{-1} - \mu_{in}^2, \\ B_{in} = \mu_{in} - \mu_{i2n}. \end{cases} \tag{3.4}$$

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

$$\alpha_{in} = \begin{cases} B_{in}/A_{in} & \text{if } A_{in} > 0, \\ \infty & \text{otherwise;} \end{cases} \tag{3.5}$$

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

We then propose an empirical Bayes selection rule $d_n^* = (d_{1n}^*, \dots, d_{kn}^*)$ for the selection problem under study, as follows:

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

$$A_n^*(\underset{\sim}{x}) = \{i | \varphi_{in}(x_i) = \max_{1\le j\le k} \varphi_{jn}(x_j)\} \tag{3.7}$$

and for each $i = 1, \dots, k$, let

$$d_{in}^*(\underset{\sim}{x}) = \begin{cases} |A_n^*(\underset{\sim}{x})|^{-1} & \text{if } i \in A_n^*(\underset{\sim}{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{3.8}$$

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

$$r(G, d_n^*) = C - \sum_{\underset{\sim}{x}\in\mathcal{X}} \left[\sum_{i=1}^{k} d_{in}^*(\underset{\sim}{x})\varphi_i(x_i)\right] f(\underset{\sim}{x}). \tag{3.9}$$

Remarks

1. Note that $\alpha_i = \infty \Longrightarrow \mathrm{Var}(P_i) = 0$, 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 μ_{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 $i = 1, \ldots, k$, $d_i(\underset{\sim}{x})$ 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 $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$, 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

$$E[r(G, d_n)] = C - \sum_{\underset{\sim}{x} \in \mathcal{X}} \left[\sum_{i=1}^{k} E[d_{in}(\underset{\sim}{x})] \varphi_i(x_i) \right] f(\underset{\sim}{x}) \tag{4.1}$$

and the expectation $E[d_{in}(\underset{\sim}{x})]$ is taken with respect to $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$. 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 β_n relative to the unknown prior distribution G if $E[r(G,d_n)] - r(G) \le O(\beta_n)$ 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 $a \le Y_i \le b$, $i = 1, \ldots, k$, then for each $t > 0$,

$$P\{\overline{Y} - \mu \ge t\} \le \exp\{-2nt^2/(b-a)^2\},$$

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 μ_i, μ_{i2}, μ_{in} and μ_{i2n} be as defined in (3.1) and (3.3), respectively. Then, for any $c > 0$,

a) $P\{\mu_{in} - \mu_i \le -c\} \le O(\exp(-2nc^2))$,

b) $P\{\mu_{in} - \mu_i \ge c\} \le O(\exp(-2nc^2))$,

c) $P\{\mu_{i2n} - \mu_{i2} \le -c\} \le O(\exp(-2nc^2))$ and

d) $P\{\mu_{i2n} - \mu_{i2} \ge c\} \le O(\exp(-2nc^2))$.

Proof: Note that under the framework of the statistical model under consideration, X_{ij}/N, $j = 1, \ldots, n$, are iid and $0 \le X_{ij}/N \le 1$. Then, $0 \le \mu_{in} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}/N \le 1$. Thus, $P\{\mu_{in} - \mu_i \ge c\} = 0$ if $\mu_i + c > 1$, and $P\{\mu_{in} - \mu_i \ge c\} \le \exp\{-2nc^2\}$ if $\mu_i + c \le 1$, 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.

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

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,

$$\begin{aligned}
&P\{[(\frac{1}{N} - 1)(\mu_{in} + \mu_i) - \frac{1}{N}](\mu_{in} - \mu_i) \leq -\frac{c}{2}\} \\
&\leq P\{\frac{1-2N}{N}(\mu_{in} - \mu_i) \leq -\frac{c}{2}\} \\
&= P\{\mu_{in} - \mu_i) \geq \frac{Nc}{2(2N-1)}\} \\
&\leq P\{\mu_{in} - \mu_i \geq \frac{c}{4}\}.
\end{aligned}$$

Thus,

$$
\begin{aligned}
&P\{A_{in} - A_i \leq c\} \\
&\leq P\{\mu_{i2n} - \mu_{i2} \leq -\frac{c}{2}\} + P\{\mu_{in} - \mu_i \geq \frac{c}{4}\} \\
&\leq P\{\mu_{i2n} - \mu_{i2} \leq -\frac{c}{4}\} + P\{\mu_{in} - \mu_i \geq \frac{c}{4}\} \\
&\leq O(\exp(-nc^2/8)),
\end{aligned}
$$

which follows from Lemma 4.2.

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

$$
\begin{aligned}
&0 \leq E[r(G, d_n^*)] - r(G) \\
&\leq \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i \in A(\underset{\sim}{x})} \sum_{j \in B(\underset{\sim}{x})} P\{\varphi_{in}(x_i) \leq \varphi_{jn}(x_j)\}. \qquad (4.2)
\end{aligned}
$$

Since the sample space $\mathcal{X}$ is finite and for each $\underset{\sim}{x} \in \mathcal{X}$, $|A(\underset{\sim}{x})| + |B(\underset{\sim}{x})| = k$, therefore, it suffices to evaluate the asymptotic behavior of the probability $P\{\varphi_{in}(x_i) \leq \varphi_{jn}(x_j)\}$ where $i \in A(\underset{\sim}{x})$, $j \in B(\underset{\sim}{x})$. Now, for each $\underset{\sim}{x} \in \mathcal{X}$, $i \in A(\underset{\sim}{x})$, $j \in B(\underset{\sim}{x})$,

$$
\begin{aligned}
&P\{\varphi_{in}(x_i) \leq \varphi_{jn}(x_j)\} \\
&= P\{\varphi_{in}(x_i) \leq \varphi_{jn}(x_j) \text{ and } (\alpha_{in} < \infty \text{ and } \alpha_{jn} < \infty)\} \qquad (4.3) \\
&+ P\{\varphi_{in}(x_i) \leq \varphi_{jn}(x_j) \text{ and } (\alpha_{in} = \infty \text{ or } \alpha_{jn} = \infty)\}.
\end{aligned}
$$

Let

$$a = \min\{A_i | i = 1, \ldots, k\} \qquad (4.4)$$

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 $\underset{\sim}{x} \in \mathcal{X},\ i \in A(\underset{\sim}{x}),\ j \in B(\underset{\sim}{x})$,

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

Proof: Note that

$$\begin{aligned} &P\{\varphi_{in}(x_i) \le \varphi_{jn}(x_j) \text{ and } (\alpha_{in} = \infty \text{ or } \alpha_{jn} = \infty)\} \\ &\le P\{\alpha_{in} = \infty\} + P\{\alpha_{jn} = \infty\} \\ &= P\{A_{in} \le 0\} + P\{A_{jn} \le 0\} \\ &= P\{A_{in} - A_i \le -A_i\} + P\{A_{jn} - A_j \le -A_j\} \\ &\le P\{A_{in} - A_i \le -a\} + P\{A_{jn} - A_j \le -a\} \\ &\le O(\exp(-na^2/8)), \end{aligned}$$

which is obtained from Lemma 4.3.

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

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

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

$$\begin{aligned} &P\{\varphi_{in}(x_i) \le \varphi_{jn}(x_j) \text{ and } (\alpha_{in} < \infty \text{ and } \alpha_{jn} < \infty)\} \\ &= P\{C_{in}(x_i)D_{jn} - C_{jn}(x_j)D_{in} \le 0\} \\ &\le P\{[C_{in}(x_i)D_{jn} - C_{jn}(x_j)D_{in}] - [C_i(x_i)D_j - C_j(x_j)D_i] \le -b\} \qquad (4.6) \\ &\le P\{C_{in}(x_i)D_{jn} - C_i(x_i)D_j \le -\frac{b}{2}\} + P\{C_{jn}(x_j)D_{in} - C_j(x_j)D_i \ge \frac{b}{2}\}. \end{aligned}$$

Now,

$$
\begin{aligned}
&P\{C_{in}(x_i)D_{jn} - C_i(x_i)D_j \le -\frac{b}{2}\} \\
&= P\{[C_{in}(x_i)D_{jn} - C_{in}(x_i)D_j] + [C_{in}(x_i)D_j - C_i(x_i)D_j] \le -\frac{b}{2}\} \qquad (4.7) \\
&\le P\{C_{in}(x_i)[D_{jn} - D_j] \le -\frac{b}{4}\} + P\{[C_{in}(x_i) - C_i(x_i)]D_j \le -\frac{b}{4}\}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&P\{C_{jn}(x_j)D_{in} - C_j(x_j)D_i \ge \frac{b}{2}\} \\
&\le P\{C_{jn}(x_j)(D_{in} - D_i) \ge \frac{b}{4}\} + P\{[C_{jn}(x_j) - C_j(x_j)]D_i \ge \frac{b}{4}\}. \qquad (4.8)
\end{aligned}
$$

Lemma 4.5.

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

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

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

d) $P\{[C_{in}(x_i) - C_i(x_i)]D_j \ge \frac{b}{4}\} \le O(\exp(-b(n)))$, 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)

$$
\begin{aligned}
&P\{C_{in}(x_i)[D_{jn} - D_j] \le -\frac{b}{4}\} \\
&= P\{(x_iA_{in} + B_{in}\mu_{in})(NA_{jn} + B_{jn} - NA_j - B_j) \le -\frac{b}{4}\} \\
&\le P\{(N+1)[N(A_{jn} - A_j) + (B_{jn} - B_j)] \le -\frac{b}{4}\} \\
&\qquad (\text{since } 0 \le x_iA_{in} + B_{in}\mu_{in} \le N+1) \\
&\le P\{A_{jn} - A_j \le -\frac{b}{8N(N+1)}\} + P\{B_{jn} - B_j < -\frac{b}{8(N+1)}\} \\
&\le O(\exp(-\frac{nb^2}{512N^2(N+1)^2})),
\end{aligned}
$$

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 $x_i > 0$,

$$P\{x_i(A_{in} - A_i) \leq -\frac{b}{8(N+1)}\}$$
$$\leq P\{A_{in} - A_i \leq -\frac{b}{8N(N+1)}\}$$
$$\leq O(\exp(-\frac{nb^2}{512N^2(N+1)^2})).$$

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 $i \in A(\underset{\sim}{x})$, $j \in B(\underset{\sim}{x})$,

$$P\{\varphi_{in}(x_i) \leq \varphi_{jn}(x_j) \text{ and } (\alpha_{in} < \infty \text{ and } \alpha_{jn} < \infty)\} \leq O(\exp(-b(n))), \quad (4.9)$$

where the expression at the right-hand-side of (4.9) is independent of the present observation $\underset{\sim}{x}$.

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

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

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

References

Abughalous, M. M. and Miescke, K. J. (1987). On selecting the largest success probability under unequal sample sizes. To appear in *J. Statist. Plann. Inference.*

Bechhofer, R. E. and Kulkarni, R. V. (1982). Closed adaptive sequential procedures for selecting the best of $k \geq 2$ Bernoulli populations. *Statistical Decision Theory and Related Topics-III* (Eds. S. S. Gupta and J. O. Berger), New York, Academic Press, **Vol. I**, 61–108.

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., 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 Liang, T. (1987). Empirical Bayes rules for selecting the best binomial populations. *Statistical Decision Theory and Related Topics-IV* (Eds. S. S. Gupta and J. O. Berger), Springer-Verlag, **Vol. I**, 213–224.

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

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.

Hoeffding, W. (1963). Probabilities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.

Kulkarni, R. V. and Jennison, C. (1986). Optimal properties of the Bechhofer-Kulkarni Bernoulli selection procedures. *Ann. Statist.*, **14**, 298–314.

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

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.

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

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY UNCLASSIFIED	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report #88-1	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Purdue University — 6b. OFFICE SYMBOL *(If applicable)*	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS *(City, State, and ZIP Code)* Department of Statistics West Lafayette, IN 47907	7b. ADDRESS *(City, State, and ZIP Code)*
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research — 8b. OFFICE SYMBOL *(If applicable)*	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-84-C-0167, NSF Grant DMS-8702620 N00014-88-K-0170 and NSF Grant DMS-860694
8c. ADDRESS *(City, State, and ZIP Code)* Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*
Selecting the Best Binomial Population: Parametric Empirical Bayes Approach

12. PERSONAL AUTHOR(S)
Shanti S. Gupta and TaChen Liang

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT *(Year, Month, Day)*	15. PAGE COUNT
Technical	FROM ____ TO ____	February 1, 1988	15

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

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 π_i has a 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 a denote the index of the selected population. Under the loss function $L(\underset{\sim}{p}, a) = p_{[k]} - p_a$, this statistical selection problem is studied via a parametric empirical Bayes approach. It is assumed that the binomial parameters p_i, i = 1,...,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

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 — 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

19. c, where n is the number of accumulated past experience (observations) at hand.