EMPIRICAL BAYES TWO-STAGE SELECTION PROCEDURES
FOR SELECTING THE BEST BERNOUlli TREATMENT USING
THE INVERSE BINOMIAL SAMPLING*

by
Shanti S. Gupta and TaChen Liang
Purdue University Wayne State University

Xun Lin
Purdue University

Technical Report #97-16C

Department of Statistics
Purdue University
West Lafayette, IN USA

September 1997
EMPIRICAL BAYES TWO-STAGE SELECTION PROCEDURES
FOR SELECTING THE BEST BERNOULLI TREATMENT USING
THE INVERSE BINOMIAL SAMPLING*

by

Shanti S. Gupta and TaChen Liang
Department of Statistics Department of Mathematics
Purdue University Wayne State University
West Lafayette, IN 47907 Detroit, MI 48202

Xun Lin
Purdue University
Department of Statistics
West Lafayette, IN 47907

Abstract

In this paper we investigate the problem of selecting the treatment with the largest probability of success from \(k(\geq 2)\) independent Bernoulli treatments. The selected treatment must also be better than a given control. We employ the empirical Bayes approach and develop a two-stage selection procedure. We prove that the proposed selection rule is asymptotically optimal at the rate of convergence of order \(O(\exp(-c^*n))\), for some positive constant \(c^*\), where \(n\) is the number of the historical data at hand. We also carry out a simulation study to investigate the performance of the proposed empirical Bayes selection procedure for small to moderate values of \(n\). The simulated results are provided in the paper.

AMS Classification: primary 62F07; secondary 62C12.

Keywords: Asymptotically optimal; empirical Bayes; two-stage selection procedure; rate of convergence; inverse binomial sampling.

*This research was supported in part by US Army Research Office, Grant DAAH04-95-1-0165 at Purdue University.
1. Introduction

Consider \( k \) independent Bernoulli treatments \( \pi_1, \ldots, \pi_k \), where for each \( i \), treatment \( \pi_i \) is characterized by the probability of success \( \theta_i \). Let \( \theta_{[1]} \leq \ldots \leq \theta_{[k]} \) denote the ordered values of the parameters \( \theta_1, \ldots, \theta_k \). It is assumed that the exact pairing between the ordered and the unordered parameters is unknown. A treatment \( \pi_i \) with \( \theta_i = \theta_{[k]} \) is called the best among the \( k \) underlying treatments. The Bernoulli treatment occurs in many fields, such as medicine, engineering and sociology. A number of statistical procedures have been studied to find the best Bernoulli treatment. Tamhane (1980) considered the problem of selecting the better of two Bernoulli treatments using a matched pair sample design. Sanchez (1987) investigated a modified least-failure sampling method for Bernoulli subset selection. Yang (1989) studied the selection procedures through a Bayes approach. Gupta and Liang (1986, 1988, 1989) developed empirical Bayes procedures for selecting the best Bernoulli treatment using binomial sampling design. In two recent papers, Gupta, Liang and Rau (1994, 1995) have specifically, investigated the empirical Bayes selection procedures for the Bernoulli and normal populations using two-stage procedures. A good review in this area is contained in Gupta and Panchapakesan (1996). For further references in the area, see also Gupta and Panchapakesan (1979) and Bechhofer, Santner and Goldsman (1996).

In this paper, we are concerned with the problem of selecting the best Bernoulli treatment compared with a specific standard using the inverse binomial sampling scheme. Unlike binomial sampling, in which the sample size (number of trials) from each treatment is fixed and therefore the pertaining random variable is always being bounded, the sample size in inverse binomial sampling is not fixed and its associated random observation is unbounded.

The formulation of the selection problem is described in Section 2. A Bayes two-stage selection procedure is derived in Section 3. We construct an empirical Bayes two-stage selection procedure in Section 4. The asymptotic optimality of the proposed empirical Bayes two-stage selection procedure is investigated in Section 5. We prove that the convergence rate of the proposed selection procedure is of exponential order, that is, \( O(\exp(-c^* n)) \), for some positive constant \( c^* \), where \( n \) is the number of the historical data at hand. In
Section 6, we carry out a simulation study to investigate the performance of the proposed selection procedure for small to moderate values of $n$. The simulated results are consistent with the rate of convergence obtained in Section 5.

2. Formulation of the Selection Problem

Consider $k$ independent Bernoulli treatments $\pi_1, \ldots, \pi_k$, with unknown success probabilities $\theta_1, \ldots, \theta_k$, respectively. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ be the ordered values of the parameters $\theta_1, \ldots, \theta_k$. We assume that the exact pairing between the ordered and unordered parameters is unknown. Any treatment associated with the largest success probability $\theta_{[k]}$ is defined as the best treatment. For a given standard $\theta_0 (0 < \theta_0 < 1)$, treatment $\pi_i$ is defined to be good if the corresponding $\theta_i > \theta_0$, and bad otherwise. Our goal is to select a treatment which is the best among the $k$ treatments and also good compared with the standard $\theta_0$. If there is no such treatment, we select none.

A two-stage selection procedure is defined as follows. First, from each of the $k$ treatments, a sequence of Bernoulli trials are carried out until a fixed number $m_1$ of successes are achieved. Let $X_i$ be the number of failures before observing the $m_1$-th success from $\pi_i$. Based on the observations $X = (X_1, \ldots, X_k)$, we decide whether the selection should be made immediately or not. If we decide to make the selection immediately, we may select a treatment from the $k$ treatments based on $X$. Or we may select none in which case the $k$ treatments are all excluded as bad. If we decide not to make the selection immediately, then based on $X$, one (and only one) treatment, say, $\pi_i$ is chosen. And from this chosen treatment $\pi_i$, a sequence of Bernoulli trials are carried out until $m_2$ additional successes are achieved. We let $Y_i$ be the number of failures before observing the $m_2$-th success at the second-stage sampling. Then based on $X$ and $Y_i$, we decide either to select treatment $\pi_i$ as the best treatment and consider it to be good, or select none and thus exclude all $k$ treatments as bad.

Let $\Omega = \{\varnothing = (\theta_1, \ldots, \theta_k) | 0 \leq \theta_i \leq 1, i = 1, \ldots, k\}$ be the parameter space. Let $a = (a_0, \ldots, a_k)$ be an action, where $a_i = 0, 1; i = 0, 1, \ldots, k$ and $\sum_{i=0}^{k} a_i = 1$. For each $i = 1, \ldots, k, a_i = 1$ means that treatment $\pi_i$ is selected as the best and also considered to be good compared with $\theta_0$. $a_0 = 1$ means that all the $k$ treatments are excluded as bad.
and none is selected. We define the termination action \( t \) as follows:

\[
t = \begin{cases} 
1, & \text{if the selection is made immediately after } X \text{ is observed,} \\
0, & \text{otherwise.}
\end{cases}
\]

When \( t = 0 \), let \( \Delta = (\Delta_1, \ldots, \Delta_k) \) be the identity action, where \( \Delta_i = 0, 1; i = 1, \ldots, k; \) and \( \sum_{i=1}^{k} \Delta_i = 1. \Delta_i = 1 \) means that the second stage sampling is taken from treatment \( \pi_i \).

The loss function is defined to be

\[
L(\theta, (a, t, \Delta)) = \max(\theta_{[k]}, \theta_0) - t \sum_{i=0}^{k} a_i \theta_i + kc_1 + (1 - t) \left\{ - \sum_{i=1}^{k} \Delta_i \left( a_i \theta_i + (1 - a_i) \theta_0 \right) + c_2 \right\},
\]

(2.1)

where \( c_1 > 0 \) is the cost of sampling from each of the \( k \) treatments at the first stage and \( c_2 \) is the cost of the sampling at the second stage.

Conditional on \( \theta_i \), we have \( X_i \sim \text{NB}(m_1, \theta_i) \) and \( Y_i \sim \text{NB}(m_2, \theta_i) \). Here, \( \text{NB}(\cdot, \cdot) \) denotes the negative binomial distribution with the two parameters. Given \( \theta_i \), let \( f_i(x|\theta_i) \) and \( g_i(y|\theta_i) \) be the conditional probability functions of \( X_i \) and \( Y_i \), respectively. That is,

\[
f_i(x|\theta_i) = \binom{m_1 + x - 1}{x} \theta_i^{m_1} (1 - \theta_i)^x, \quad x = 0, 1, \ldots
\]

and

\[
g_i(y|\theta_i) = \binom{m_2 + y - 1}{y} \theta_i^{m_2} (1 - \theta_i)^y, \quad y = 0, 1, \ldots
\]

It is assumed that for each \( i = 1, \ldots, k, \theta_i \) is a realization of a random variable \( \Theta_i \) which has a Beta distribution with probability density function \( h_i(\theta_i|\alpha_i, \mu_i) \), where

\[
h_i(\theta_i|\alpha_i, \mu_i) = \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i \mu_i) \Gamma(\alpha_i (1 - \mu_i))} \theta_i^{\alpha_i - 1} (1 - \theta_i)^{\alpha_i (1 - \mu_i) - 1}, \quad 0 < \theta_i < 1.
\]

Both \( \alpha_i \) and \( \mu_i \) are unknown, \( 0 < \mu_i < 1 \) and \( \alpha_i > 0 \). Note that \( E(\Theta_i) = \mu_i \) and \( \text{Var}(\Theta_i) = \frac{\mu_i (1 - \mu_i)}{\alpha_i + 1} \). It is also assumed that \( \Theta_1, \ldots, \Theta_k \) are mutually independent.

Denote \( Y = (Y_1, \ldots, Y_k) \). Let \( \mathcal{X} \) and \( \mathcal{Y} \) be the sample spaces generated by \( X \) and \( Y \), respectively. A two-stage selection procedure, in general, consists of the following rules:
a. Stopping rule $\tau$: For each $x \in \mathcal{X}$, $\tau(x)$ is the probability of terminating the sampling after observing $x$ and making a selection decision immediately based on $x$.

b. Identity rule $\delta = (\delta_1, \ldots, \delta_k)$: For each $x \in \mathcal{X}$, $\delta_i(x)$ is the probability of taking the second-stage inverse binomial sample from treatment $\pi_i$ when we decide not to make a selection decision immediately after observing $x$. Note that $\delta$ satisfies $\sum_{i=1}^{k} \delta_i(x) = 1$ for each $x \in \mathcal{X}$.

c. First-stage selection rule $d_1 = (d_{10}, \ldots, d_{1k})$: For each $x \in \mathcal{X}$ and $i = 1, \ldots, k$, let $d_{1i}(x)$ be the probability of selecting treatment $\pi_i$ as the best from among the $k$ treatments and also good compared with $\theta_0$, and let $d_{10}(x)$ be the probability of excluding all the $k$ treatments and selecting none. For all $x \in \mathcal{X}$, $\sum_{i=0}^{k} d_{1i}(x) = 1$.

d. Second-stage selection rule $d_2 = (d_{20}, \ldots, d_{2k})$. For each $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, when the decision of going to take the second-stage sample from $\pi_i$ is made, $d_{2i}(x, y)$ is the probability of selecting $\pi_i$ as the best from among the $k$ treatments and also considered to be good compared with $\theta_0$, for $i = 1, \ldots, k$. Note that $d_{2i}(x, y)$ depends on $\mathcal{Y}$ only through $Y_i$ since there are no second-stage observations from any other treatments. Therefore, we denote $d_{2i}(x, y_i) = d_{2i}(x, y), 1 \leq i \leq k$. Let $d_{20}(x, y) = \sum_{i=1}^{k} \delta_i(x)(1 - d_{2i}(x, y_i))$ be the probability of excluding all the $k$ treatments and selecting none, based on $x$ and $y$.

Under the preceding statistical model, the Bayes risk of the two-stage selection procedure $(\tau, \delta, d_1, d_2)$ is denoted by $R(\tau, \delta, d_1, d_2)$.

We have

$$R(\tau, \delta, d_1, d_2) = C - \sum_{x \in \mathcal{X}} \tau(x) \left( \sum_{i=0}^{k} d_{1i}(x) \varphi_i(x_i) \right) f(x) + \sum_{x \in \mathcal{X}} (1 - \tau(x)) \left( c_2 - \theta_0 + \sum_{i=1}^{k} \delta_i(x) \left( \sum_{y_i=0}^{+\infty} d_{2i}(x, y_i)(\theta_0 - \psi_i(x_i + y_i)) f_i(y_i|x_i, \alpha_i, \mu_i) \right) \right) f(x)$$

$$= \sum_{x \in \mathcal{X}} \tau(x) \left\{ \begin{array}{l} \sum_{i=0}^{k} d_{1i}(x)(\theta_0 - \varphi_i(x_i)) - c_2 + \theta_0 \\
- \sum_{i=1}^{k} \delta_i(x) \left( \sum_{y_i=0}^{+\infty} d_{2i}(x, y_i)(\theta_0 - \psi_i(x_i + y_i)) f_i(y_i|x_i, \alpha_i, \mu_i) \right) \end{array} \right\} f(x)$$
\[\sum_{\bar{x} \in \mathcal{X}} \left\{ c_2 - \theta_0 + \sum_{i=1}^{k} \delta_i(x) \left( \sum_{y_i=0}^{+\infty} d_{2i}(x, y_i) (\theta_0 - \psi_i(x_i + y_i)) f_i(y_i | x_i, \alpha_i, \mu_i) \right) \right\} f(x) + C.\]

(2.3)

where \( \varphi_i(x_i) = E(\Theta_i | X_i = x_i) = \frac{\alpha_i \mu_i + m_i}{\alpha_i + m_i + x_i} \) is the posterior mean of \( \Theta_i \) given \( X_i = x_i, i = 1, \ldots, k \), \( \varphi_0(x_0) = \theta_0; \psi_i(x_i + y_i) = \frac{\alpha_i \mu_i + m_i + m_2}{\alpha_i + m_1 + m_2 + x_i + y_i} \) is the posterior mean of \( \Theta_i \) given \( X_i = x_i \) and \( Y_i = y_i \) for each \( i = 1, \ldots, k \). Also \( C = \int_{\Omega} \max(\theta_0, \theta_0) dH(\theta) + kc_1 \) and \( H(\theta) \) is the joint distribution of \( \Theta = (\Theta_1, \ldots, \Theta_k) \); \( f(x) = \prod_{i=1}^{k} f_i(x_i) \) where

\[ f_i(x_i) = \int_{0}^{1} f_i(x_i | \theta_i) h_i(\theta_i | \alpha_i, \mu_i) d\theta_i \]

\[ = \left( \frac{m_1 + x_i - 1}{x_i} \right) \frac{\Gamma(\alpha_i) \Gamma(m_1 + \alpha_i + \mu_i) \Gamma(x_i) + \alpha_i (1 - \mu_i))}{\Gamma(\alpha_i \mu_i) \Gamma(\alpha_i (1 - \mu_i)) \Gamma(m_1 + x_i + \alpha_i)} \]

is the marginal probability function of \( X_i \).

\[ f_i(y_i | x_i, \alpha_i, \mu_i) = \frac{f_i(x_i, y_i)}{f_i(x_i)}, \text{ where } f_i(x_i, y_i) = \int_{0}^{1} f_i(x_i | \theta_i) g_i(y_i | \theta_i) h_i(\theta_i | \alpha_i, \mu_i) d\theta_i \]

Note that \( f_i(y_i | x_i, \alpha_i, \mu_i) \) is the marginal conditional probability function of \( Y_i \) given \( X_i = x_i \) and \( (\alpha_i, \mu_i) \). A direct computation yields

\[ f_i(y_i | x_i, \alpha_i, \mu_i) = \left( \frac{m_2 + y_i - 1}{y_i} \right) \]

\[ \times \frac{\Gamma(m_1 + x_i + \alpha_i) \Gamma(m_1 + m_2 + \alpha_i \mu_i) \Gamma(x_i + y_i + \alpha_i (1 - \mu_i))}{\Gamma(m_1 + m_2 + x_i + y_i + \alpha_i) \Gamma(m_1 + \alpha_i \mu_i) \Gamma(x_i + \alpha_i (1 - \mu_i))}. \]

3. Derivation of a Bayes Two-stage Selection Procedure

We derive a Bayes two-stage selection procedure as follows.

a. First-stage selection rule

For each \( \bar{x} \in \mathcal{X} \), let \( I(\bar{x}) = \{ i \mid \varphi_i(x_i) = \max_{0 \leq j \leq k} \varphi_j(x_j), i = 0, 1, \ldots, k \} \). Define \( i^* = \min\{ i \mid i \in I(\bar{x}) \} \). Then we have a first-stage selection rule \( \bar{d}^B = (d_{10}^B, d_{11}^B, \ldots, d_{1k}^B) \) as follows:
\[
\begin{aligned}
    \begin{cases}
        d^B_{1i^*}(x) = 1, \\
        d^B_{1j}(x) = 0, \quad \text{for } j \neq i^*.
    \end{cases}
\end{aligned}
\quad (3.1)
\]

b. Second-stage selection rule

We define a second-stage selection rule \( d^B_2 = (d^B_{20}, d^B_{21}, \ldots, d^B_{2k}) \) as follows: for \( i = 1, \ldots, k, \)

\[
    d^B_{2i}(x, y) = d^B_{2i}(x, y_i) = \begin{cases} 
        1, & \text{if } \psi_i(x_i + y_i) > \theta_0; \\
        0, & \text{otherwise},
    \end{cases}
\]

and

\[
    d^B_{20}(x, y) = \sum_{i=1}^{k} \delta^B_i(x_i)(1 - d^B_{2i}(x, y_i))
\]

where \( \delta^B(x) = (\delta^B_1(x), \ldots, \delta^B_k(x)) \) is the identity rule defined below.

c. Identity rule

For each \( i = 1, \ldots, k, \) and \( x \in \mathcal{X}, \) define

\[
    T_i(x) = \sum_{y_i=0}^{\infty} d^B_{2i}(x, y_i)(\theta_0 - \psi_i(x_i + y_i))f_i(y_i|x_i, \alpha_i, \mu_i).
\]

(3.3)

Let \( J(x) = \{ j | T_j(x) = \min_{1 \leq j \leq k} T_i(x) \} \) and \( j^* = \min\{ j | j \in J(x) \}. \)

We then define an identity rule \( \delta^B = (\delta^B_1, \ldots, \delta^B_k) \) as follows:

\[
    \delta^B_j(x) = \begin{cases} 
        1, & \text{if } j = j^*; \\
        0, & \text{otherwise}.
    \end{cases}
\]

(3.4)

d. Stopping rule

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

\[
    Q(x) = \sum_{i=0}^{k} d^B_{1i}(x)(\theta_0 - \varphi_i(x_i)) - c_2 - \sum_{i=1}^{k} \delta^B_i(x)T_i(x).
\]

(3.5)

The stopping rule \( \tau^B \) is defined to be
\[ \tau^B(\bar{z}) = \begin{cases} 
1, & \text{if } Q(\bar{z}) \leq 0; \\
0, & \text{otherwise.} 
\end{cases} \tag{3.6} \]

Next we will show that the selection procedure \((\tau^B, \delta^B, d_1^B, d_2^B)\) is a Bayes two-stage selection procedure.

**Theorem 3.1** The two-stage selection procedure \((\tau^B, \delta^B, d_1^B, d_2^B)\) is a Bayes two-stage selection procedure.

**Proof:** Let \((\tau, \delta, \delta_1, \delta_2)\) be any two-stage selection procedure. It suffices to show \(R(\tau, \delta, d_1, d_2) - R(\tau^B, \delta^B, d_1^B, d_2^B) \geq 0\). We have

\[
R(\tau, \delta, d_1, d_2) - R(\tau^B, \delta^B, d_1^B, d_2^B) = I + II + III, \tag{3.7}
\]

where

\[
\begin{align*}
I &= R(\tau, \delta, d_1, d_2) - R(\tau, \delta, d_1^B, d_2^B), \\
II &= R(\tau, \delta, d_1^B, d_2^B) - R(\tau, \delta^B, d_1^B, d_2^B), \\
III &= R(\tau, \delta^B, d_1^B, d_2^B) - R(\tau^B, \delta^B, d_1^B, d_2^B).
\end{align*} \tag{3.8}
\]

From (2.2),

\[
I = \sum_{\bar{z} \in \mathcal{X}} \tau(\bar{z}) \left( \sum_{i=0}^{k} (d_{1i}(\bar{z}) - d_{1i}(\bar{z})) \varphi_i(x_i) \right) f(\bar{z})
\]

\[
+ \sum_{\bar{z} \in \mathcal{X}} (1 - \tau(\bar{z})) \left\{ \sum_{i=1}^{k} \delta_i(\bar{z}) \left( \sum_{y_i=0}^{+\infty} (d_{2i}(\bar{z}, y_i) - d_{2i}^B(\bar{z}, y_i))(\theta_0 - \psi_i(x_i + y_i)) \right) \times f_i(y_i|x_i, \alpha_i, \mu_i) \right\} f(\bar{z}) \tag{3.9}
\]

By the definition of \(d_{2i}^B\) and \(d_{1i}\), for each \(i = 1, \ldots, k\),

\[
\sum_{i=0}^{k} (d_{1i}(\bar{z}) - d_{1i}(\bar{z})) \varphi_i(x_i) = \max_{0 \leq i \leq k} (\varphi_i(x_i)) - \sum_{i=1}^{k} d_{1i}(\bar{z}) \varphi_i(x_i) \geq 0,
\]

and by the definition of \(d_{2i}^B\), we have \((d_{2i}(\bar{z}, y_i) - d_{2i}^B(\bar{z}, y_i))(\theta_0 - \psi_i(x_i + y_i)) \geq 0\). Therefore, \(I \geq 0\) since all the other terms in (3.9) are nonnegative.
From (2.2) again, since
\[\sum_{i=1}^{k} (\delta_i(z) - \delta_i^B(z))T_i(z) = \sum_{i=1}^{k} \delta_i(z)T_i(z) - \min_{1 \leq i \leq k} T_i(z) \geq 0,\] (3.10)
we have
\[II = \sum_{z \in \mathcal{X}} (1 - \tau(z)) \left( \sum_{i=1}^{k} (\delta_i(z) - \delta_i^B(z))T_i(z) \right) f(z) \geq 0.\] (3.11)

From (2.3), by the definition of \(\tau^B\),
\[III = \sum_{z \in \mathcal{X}} (\tau(z) - \tau^B(z))Q(z) f(z) \geq 0.\] (3.12)

The proof of Theorem 3.1 is completed by combining (3.7), (3.9), (3.11) and (3.12).

4. The Construction of an Empirical Bayes Two-Stage Selection Procedure

4.1. Empirical Bayes Framework

The Bayes two-stage selection procedure \((\tau^B, \delta^B, q^B_1, q^B_2)\) defined in Section 3 depends on the unknown parameters \((\alpha_i, \mu_i), i = 1, \ldots, k\). Since the parameters are unknown, it is impossible to implement the Bayes two-stage selection procedure for the selection problem in practice. In the empirical Bayes framework, it is generally assumed that there are some past observations when the present selection is to be made. At time \(j = 1, \ldots, n\), let \(X_{ij}\) be the number of failures before observing the \(m_1\)-th success from \(\pi_i\). Let \(\Theta_j = (\Theta_{1j}, \ldots, \Theta_{kj})\) be a random vector where \(\Theta_{ij}\) stands for the probability of success from treatment \(\pi_i\) at time \(j\). We assume that \(\Theta_j, j = 1, \ldots, n\), are i.i.d with prior density \(h(\Theta_j) = \prod_{i=1}^{k} h_i(\Theta_{ij}|\alpha_i, \mu_i)\). Therefore, conditional on \(\Theta_{ij} = \theta_{ij}, X_{ij}|\theta_{ij} \sim NB(m_1, \theta_{ij})\) and \(X_{ij}\) has a marginal probability function \(f_i(x)\). Let \(X_j = (X_{ij}, \ldots, X_{kj})\) be the random observations of the first-stage sampling taken at time \(j = 1, \ldots, n\). Let \(X_{n+1} = X = (X_1, \ldots, X_k)\) be the current (present) observations taken at the first stage.

4.2. Certain Useful Empirical Bayes Estimates

In order to construct an empirical Bayes two-stage selection procedure, we first introduce certain properties related to the parameters \((\alpha_i, \mu_i)\). Then based on these properties, the empirical Bayes estimators are derived. We then construct an empirical Bayes two-stage selection procedure in the next subsection.
For each $i = 1, \ldots, k$ and $j = 1, 2, \ldots, n$, we have

$$E\left(\frac{\left(m_1 + X_{ij} - 2\right)}{m_1 + X_{ij} - 1}\right) = E\left(\mathbb{E}\left[\frac{\left(m_1 + X_{ij} - 2\right)}{m_1 + X_{ij} - 1} \mid \Theta_{ij}\right]\right) = E\Theta_{ij} = \mu_i, \quad (4.1)$$

and

$$E\left(\frac{\left(m_1 + X_{ij} - 3\right)}{m_1 + X_{ij} - 1} I(X_{ij} \geq 2)\right) = E\left(\mathbb{E}\left[\frac{\left(m_1 + X_{ij} - 3\right)}{m_1 + X_{ij} - 1} I(X_{ij} \geq 2) \mid \Theta_{ij}\right]\right)$$

$$= E(1 - \Theta_{ij})^2 = \frac{(\alpha_i(1 - \mu_i) + 1)(1 - \mu_i)}{\alpha_i + 1} \equiv \nu_i \quad (4.2)$$

Let $C_i = \nu_i - (1 - \mu_i)^2$ and $D_i = 1 - \mu_i - \nu_i$, for $i = 1, \ldots, k$. From (4.2), we have,

$$\begin{cases} C_i = \frac{(1 - \mu_i)\mu_i}{\alpha_i + 1}, \\ D_i = \frac{\alpha_i(1 - \mu_i)\mu_i}{\alpha_i + 1}. \quad (4.3) \end{cases}$$

Note that for $i = 1, \ldots, k$, $\alpha_i = D_i/C_i$. Since $\alpha_i > 0$ and $0 < \mu_i < 1$, both $C_i$ and $D_i$ are positive.

For $i = 1, \ldots, k$, based on $(X_{i1}, \ldots, X_{in})$, let,

$$\begin{cases} \mu_{in} = \frac{1}{n} \sum_{j=1}^{n} \frac{\left(m_1 + X_{ij} - 2\right)}{m_1 + X_{ij} - 1}, \\ \nu_{in} = \frac{1}{n} \sum_{j=1}^{n} \frac{\left(m_1 + X_{ij} - 3\right)}{m_1 + X_{ij} - 1} I(X_{ij} \geq 2), \quad (4.4) \end{cases}$$

and

$$\begin{cases} C_{in} = \nu_{in} - (1 - \mu_{in})^2, \\ D_{in} = 1 - \mu_{in} - \nu_{in}. \quad (4.5) \end{cases}$$

Then we have

$$D_{in} = \frac{1}{n} \sum_{j=1}^{n} \left(1 - \frac{\left(m_1 + X_{ij} - 2\right)}{m_1 + X_{ij} - 1} - \frac{\left(m_1 + X_{ij} - 3\right)}{m_1 + X_{ij} - 1} I(X_{ij} \geq 2)\right)$$

$$= \frac{1}{n} \sum_{j=1}^{n} \left(1 - \frac{m_1 - 2}{m_1 + X_{ij} - 2} - \frac{(X_{ij} - 1)(X_{ij} - 2)}{(m_1 + X_{ij} - 2)(m_1 + X_{ij} - 3)} I(X_{ij} \geq 2)\right)$$

$$\geq 0.$$
However, $C_{in}$ may be nonpositive while $C_i > 0$. From $\alpha_i = D_i/C_i$, we can see that $\alpha_i \to +\infty$ as $C_i \to 0$. So we define

$$\alpha_{in} = \begin{cases} 
D_{in}/C_{in}, & \text{if } C_{in} > 0, \\
\infty, & \text{otherwise}.
\end{cases} \quad (4.6)$$

We can also see that $\lim_{\alpha_i \to +\infty} \varphi_i(x_i) = \mu_i$ and $\lim_{\alpha_i \to +\infty} \psi_i(x_i + y_i) = \mu_i$. Therefore, we define empirical Bayes estimators $\varphi_{in}(x_i)$ and $\psi_{in}(x_i + y_i)$ for the posterior means $\varphi_i(x_i)$ and $\psi_i(x_i + y_i)$, respectively, as follows:

$$\begin{align*}
\varphi_{in}(x_i) &= \begin{cases} 
\frac{\alpha_{in}\mu_i + m_1}{\alpha_{in} + x_i + m_1}, & \text{if } C_{in} > 0; \\
\mu_i, & \text{otherwise},
\end{cases} \\
\psi_{in}(x_i + y_i) &= \begin{cases} 
\frac{\alpha_{in}\mu_i + m_2}{\alpha_{in} + x_i + y_i + m_1 + m_2}, & \text{if } C_{in} > 0; \\
\mu_i, & \text{otherwise}.
\end{cases} \quad (4.7)
\end{align*}$$

Let $\varphi_{on}(x_0) = \theta_0$. Since $\lim_{\alpha_i \to +\infty} f_i(y_i|x_i, \alpha_i, \mu_i) = (m_2 + y_i - 1)\mu_i^{m_2} (1 - \mu_i)^{y_i}$, we define an empirical Bayes estimator for the marginal conditional probability function $f_i(y_i|x_i, \alpha_i, \mu_i)$:

$$f_{in}(y_i|x_i, \alpha_{in}, \mu_{in}) = \begin{cases} 
f_i(y_i|x_i, \alpha_{in}, \mu_{in}), & \text{if } C_{in} > 0; \\
(m_2 + y_i - 1)^{m_2} \mu_{in}^{m_2} (1 - \mu_{in})^{y_i}, & \text{otherwise}.
\end{cases} \quad (4.8)$$

### 4.3 The Proposed Empirical Bayes Two-Stage Selection Procedure

We now propose an empirical Bayes two-stage selection procedure $(\tau^{*n}, \delta^{*n}, d_1^{*n}, d_2^{*n})$ as follows:

a. **Empirical Bayes first-stage selection rule** $d_1^{*n} = (d_{10}^{*n}, d_{11}^{*n}, \ldots, d_{1k}^{*n})$.

Let $I_n(x) = \{i|\varphi_{in}(x) = \max_{0 \leq j \leq k} \varphi_{jn}(x_j), i = 0, 1, \ldots, k\}$. Define $i_n^* = \min\{i|i \in I_n(x)\}$. Then we define $d_1^{*n}$ as follows:

$$\begin{cases} 
d_{1i_{in}^*}^{*n}(x) = 1, \\
d_{1j}^{*n}(x) = 0, \text{ for } j \neq i_n^*.
\end{cases} \quad (4.9)$$

b. **Empirical Bayes second-stage selection rule** $d_2^{*n} = (d_{20}^{*n}, d_{21}^{*n}, \ldots, d_{2k}^{*n})$.

For $i = 1, \ldots, k$, let
\[ d_{2i}(\mathbf{x}, y) = d_{20}(\mathbf{x}, y_i) = \begin{cases} 1, & \text{if } \psi_{in}(x_i + y_i) > \theta_0; \\ 0, & \text{otherwise}, \end{cases} \]  

(4.10)

and 

\[ d_{20}(\mathbf{x}, y) = \sum_{i=1}^{k} \delta_{i}^{*n}(\mathbf{x})(1 - d_{20}(\mathbf{x}, y_i)) \]

where \( \delta^{*n} = (\delta_{1}^{*n}, \ldots, \delta_{k}^{*n}) \) is the empirical Bayes identity rule defined below.

c. **Empirical Bayes identity rule** \( \delta^{*n} = (\delta_{1}^{*n}, \ldots, \delta_{k}^{*n}) \)

For \( i = 1, \ldots, k \), define 

\[ T_{in}(\mathbf{x}) = \sum_{y_i=0}^{+\infty} d_{2i}^{*n}(\mathbf{x}, y_i)(\theta_0 - \psi_{in}(x_i + y_i))f_{in}(y_i|x_i, \alpha_{in}, \mu_{in}) \]  

(4.11)

Let \( J_n(\mathbf{x}) = \{ j | T_{jn}(\mathbf{x}) = \min_{1 \leq i \leq k} T_{in}(\mathbf{x}), j = 1, \ldots, k \} \) and \( j^{*} = \min \{ j | j \in J_n(\mathbf{x}) \} \). Then the empirical Bayes identity rule \( \delta^{*n} \) is

\[ \delta_{j}^{*n}(\mathbf{x}) = \begin{cases} 1, & \text{if } j = j^{*}, \\ 0, & \text{otherwise}. \end{cases} \]  

(4.12)

d. **Empirical Bayes Stopping Rule**

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

\[ Q_{n}(\mathbf{x}) = \sum_{i=0}^{k} d_{1i}^{*n}(\mathbf{x})(\theta_0 - \varphi_{in}(x_i)) - c_2 - \sum_{i=1}^{k} \delta_{i}^{*n}(\mathbf{x})T_{in}(\mathbf{x}) \]  

(4.13)

and 

\[ \tau^{*n}(\mathbf{x}) = \begin{cases} 1, & \text{if } Q_{n}(\mathbf{x}) \leq 0, \\ 0, & \text{otherwise}. \end{cases} \]  

(4.14)

5. **Asymptotic Optimality of** \( (\tau^{*n}, \delta^{*n}, d_{1n}^{*n}, d_{2n}^{*n}) \)

Consider the empirical Bayes two-stage selection procedure \( (\tau^{*n}, \delta^{*n}, d_{1n}^{*n}, d_{2n}^{*n}) \) constructed in Section 4. Let \( R(\tau^{*n}, \delta^{*n}, d_{1n}^{*n}, d_{2n}^{*n}) \) be the conditional Bayes risk given \( (X_1, X_2, \ldots, X_n) \) and \( E_n R(\tau^{*n}, \delta^{*n}, d_{1n}^{*n}, d_{2n}^{*n}) \) the Bayes risk of the empirical Bayes two-stage selection procedure \( (\tau^{*n}, \delta^{*n}, d_{1n}^{*n}, d_{2n}^{*n}) \), respectively, where \( E_n \) is the expectation
taken with respect to \((X_1, \ldots, X_n)\). Note that \(R(\tau^*, \delta^*, d_1^*, d_2^*) - R(\tau^B, \delta^B, d_1^B, d_2^B) \geq 0\) since \((\tau^B, \delta^B, d_1^B, d_2^B)\) is the Bayes two-stage selection procedure. Therefore, we have
\[
E_n R(\tau^*, \delta^*, d_1^*, d_2^*) - R(\tau^B, \delta^B, d_1^B, d_2^B) \geq 0.
\]
We use the nonnegative regret risk
\[
E_n R(\tau^*, \delta^*, d_1^*, d_2^*) - R(\tau^B, \delta^B, d_1^B, d_2^B)
\]
as a measure of the performance of the empirical Bayes two-stage procedure \((\tau^*, \delta^*, d_1^*, d_2^*)\). Following (3.7) – (3.12), we have
\[
R(\tau^*, \delta^*, d_1^*, d_2^*) - R(\tau^B, \delta^B, d_1^B, d_2^B) = I_n + II_n + III_n
\]
where
\[
0 \leq I_n = R(\tau^*, \delta^*, d_1^*, d_2^*) - R(\tau^*, \delta^*, d_1^*, d_2^*)
\]
\[
= \sum_{\bar{x} \in \mathcal{X}} \tau^*(\bar{x}) (d_{1i}^*(\bar{x}) \varphi_{i*}(x_{i*}) - d_{n*}^*(\bar{x}) \varphi_{i*}(x_{i*})) f(\bar{x})
\]
\[
+ \sum_{\bar{x} \in \mathcal{X}} (1 - \tau^*(\bar{x})) \left\{ \sum_{i=1}^{k} \delta_{i}^*(\bar{x}) \left( \sum_{y_i=0}^{+\infty} (d_{2i}^*(\bar{x}, y_i) - d_{2j}^*(\bar{x}, y_j)) \right) \times (\theta_0 - \psi_i(x_i + y_i)) f_i(y_i|\bar{x}, \alpha_i, \mu_i) \right\} f(\bar{x})
\]
\[
= I_{n,1} + I_{n,2},
\]
\[
0 \leq II_n = R(\tau^*, \delta^*, d_1^B, d_2^B) - R(\tau^*, \delta^*, d_1^B, d_2^B)
\]
\[
= \sum_{\bar{x} \in \mathcal{X}} (1 - \tau^*(\bar{x})) (\delta_{i*}^*(\bar{x}) T_{i*}^*(\bar{x}) - \delta_{j*}^*(\bar{x}) T_{j*}^*(\bar{x})) f(\bar{x})
\]
and
\[
0 \leq III_n = R(\tau^*, \delta^*, d_1^B, d_2^B) - R(\tau^B, \delta^*, d_1^B, d_2^B)
\]
\[
= \sum_{\bar{x} \in \mathcal{X}} (\tau^*(\bar{x}) - \tau^B(\bar{x})) Q(\bar{x}) f(\bar{x}).
\]

### 5.1 Preliminary Analysis

To investigate the asymptotic optimality of \((\tau^*, \delta^*, d_1^*, d_2^*)\), we first do some preliminary analysis. Define for \(i = 1, \ldots, k,\)
\[
L_{1i} = \begin{cases} 
\max\{N|\frac{\alpha_i \mu_i + m_1}{\alpha_i + m_1 + N} > \theta_0, & N \text{ is a nonnegative integer}\}, \\
-1, & \text{if no such nonnegative integer exists.}
\end{cases}
\]
and
\[
L_{2i} = \begin{cases} 
\max\{N|\frac{\alpha_i \mu_i + m_1 + m_2}{\alpha_i + m_1 + m_2 + N} > \theta_0, & N \text{ is a nonnegative integer}\}, \\
-1, & \text{if no such nonnegative integer exists.}
\end{cases}
\]
From (5.5) and (5.6), we can immediately see that for $1 \leq i \leq k$,

$$\begin{aligned}
&\begin{cases}
\text{If } L_{1i} \geq 0, & \text{then } \varphi_i(x_i) > \theta_0 \text{ for } 0 \leq x_i \leq L_{1i}, \\
\text{If } L_{1i} = -1, & \text{then } \varphi_i(x_i) \leq \theta_0 \text{ for all } x_i = 0,1,\ldots
\end{cases} \\
\quad \text{(5.7)}
\end{aligned}$$

and

$$\begin{aligned}
&\begin{cases}
\text{If } L_{2i} \geq 0, & \text{then } \psi_i(x_i + y_i) > \theta_0 \text{ for } 0 \leq (x_i + y_i) \leq L_{2i}; \\
\text{If } L_{2i} = -1, & \text{then } \psi_i(x_i + y_i) \leq \theta_0 \text{ for all } x_i(1) = 0,1,\ldots
\end{cases} \\
\quad \text{(5.8)}
\end{aligned}$$

Note that both $L_{1i}$ and $L_{2i}$ are finite whole numbers given $\alpha_i$ and $\mu_i$, and we always have $L_{2i} \geq L_{1i}$, for $i = 1,\ldots,k$.

For $i = 1,\ldots,k$, let

$$S_{1i} = \begin{cases}
\{x \in X| 0 \leq x_i \leq L_{1i}\}, & \text{if } L_{1i} \geq 0; \\
\phi, & \text{otherwise.}
\end{cases} \quad \text{(5.9)}$$

$$S_{2i} = \begin{cases}
\{(x,y) \in X \times Y| 0 \leq x_i + y_i \leq L_{2i}\}, & \text{if } L_{2i} \geq 0; \\
\phi, & \text{otherwise.}
\end{cases} \quad \text{(5.10)}$$

Then from (5.7) and (5.9), for each $x \in S_{1i}, \varphi_i(x_i) - \theta_0 > 0$. Similarly, from (5.8) and (5.10), for $(x,y) \in S_{2i}, \psi_i(x_i + y_i) - \theta_0 > 0$.

Also, for $i = 1,\ldots,k$, let

$$S_{3i} = \{x \in X|^T_i(x) < 0\}. \quad \text{(5.11)}$$

From (3.2) and (3.3), we have, for $i = 1,\ldots,k$,

$$\begin{aligned}
&\begin{cases}
T_i(x) = 0, & \text{if } x_i > L_{2i}; \\
T_i(x) < 0, & \text{if } x_i \leq L_{2i} \text{ and } L_{2i} \geq 0.
\end{cases} \\
\quad \text{(5.12)}
\end{aligned}$$

Next, for $j = 1,\ldots,k$, let

$$D_{1j} = \{x \in X| \theta_0 - \varphi_j(x_j) > 0\} \quad \text{(5.13)}$$

and

$$D_{2j} = \{(x,y) \in X \times Y| \theta_0 - \psi_j(x_j + y_j) > 0\} \quad \text{(5.14)}$$

Since $\varphi_j(x_j) \to 0$ as $x_j \to +\infty$, and $\psi_j(x_j + y_j) \to 0$ as $(x_i + y_i) \to +\infty$, we can see that both $D_{1j}$ and $D_{2j}$ are always nonempty.
(I) Analysis for \( I_n \)

We have, from (5.2),

\[
I_{n,1} = \sum_{\mathcal{X}} \tau^{*n}(x) (d_{i*}^{B}(x) \varphi_i(x) - d_{i*}^{*n}(x) \varphi_{i*}(x)) f(x)
\]

\[
= \sum_{i=1}^{k} \sum_{\mathcal{X}} \tau^{*n}(x) I\{i^* = i, i^*_n = 0\} (\varphi_i(x_i) - \theta_0) f(x)
\]

\[
+ \sum_{j=1}^{k} \sum_{\mathcal{X}} \tau^{*n}(x) I\{i^* = 0, i^*_n = j\} (\theta_0 - \varphi_j(x_j)) f(x)
\]

(5.15)

\[
+ \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\mathcal{X}} \tau^{*n}(x) I\{i^* = i, i^*_n = j\} (\varphi_i(x_j) - \varphi_j(x_j)) f(x)
\]

\[
= I_{n,1,1} + I_{n,1,2} + I_{n,1,3}
\]

Let \( b_1 = \min_{1 \leq i \leq k} \min_{0 \leq x_i \leq L_{1i}, \varphi_i(x_i) - \theta_0 > 0} \{\varphi_i(x_i) - \theta_0\} \). Note that when \( S_{1i} \neq \phi \) and \( 0 \leq x_i \leq L_{1i}, \varphi_i(x_i) - \theta_0 > 0 \). Since \( L_{1i} \) is finite, minimum operation in \( b_1 \) is taken on a finite set. Therefore, \( b_1 > 0 \).

Since \( 0 \leq \tau^{*n}(x) \leq 1 \) and \( |\varphi_i(x_i) - \theta_0| \leq 1 \) for all \( x \in \mathcal{X} \),

\[
E_n I_{n,1,1} \leq \sum_{i=1}^{k} \sum_{\mathcal{X}} P_n\{i^* = i, i^*_n = 0\} f(x)
\]

\[
= \sum_{i=1}^{k} \sum_{\mathcal{X} \in S_{1i}} P_n\{i^* = i, i^*_n = 0\} f(x)
\]

(5.16)

\[
= \sum_{i=1}^{k} \sum_{x \in S_{1i}} P_n\{\varphi_i(x_i) \geq \varphi_l(x_i), \forall l \neq i, \varphi_i(x_i) > \theta_0, \varphi_{ln}(x_i) \leq \theta_0, \forall l \neq 0\} f(x)
\]

\[
\leq \sum_{i=1}^{k} \sum_{x \in S_{1i}} P_n\{|\varphi_{ln}(x_i) - \varphi_i(x_i)| \geq \varphi_i(x_i) - \theta_0\} f(x)
\]

\[
\leq \sum_{i=1}^{k} \sum_{x \in S_{1i}} P_n\{|\varphi_{ln}(x_i) - \varphi_i(x_i)| \geq b_1\} f(x)
\]

Denote \( b_{2j} = \min_{x \in D_{1j}} \{\theta_0 - \varphi_j(x_j)\} \) and \( b_2 = \min_{1 \leq j \leq k} b_{2j} \). Since \( \varphi_j(x_j) \) is strictly de-
creasing as \(x_j\) increases, we have

\[
b_{2j} = \begin{cases} 
\theta_0 - \varphi_j(L_{1j} + 1), & \text{if } \theta_0 > \varphi_j(L_{1j} + 1); \\
\theta_0 - \varphi_j(L_{1j} + 2), & \text{if } \theta_0 = \varphi_j(L_{1j} + 1). 
\end{cases}
\]  

(5.17)

Therefore, \(b_2 = \min_{1 \leq j \leq k} b_{2j} > 0\).

\[
E_n I_{n,1,2} \leq \sum_{j=1}^{k} \sum_{\bar{x} \in D_{1j}} P_n\{i^* = 0, i_n^* = j\} f(\bar{x}) \\
\leq \sum_{j=1}^{k} \sum_{\bar{x} \in D_{1j}} P_n\{|\varphi_j(x_{jn}) - \varphi_j(x_j)| \geq \theta_0 - \varphi_j(x_j)\} f(\bar{x}) \\
\leq \sum_{j=1}^{k} \sum_{\bar{x} \in D_{1j}} P_n\{|\varphi_j(x_{jn}) - \varphi_j(x_j)| \geq b_2\} f(\bar{x})
\]  

(5.18)

Let \(b_3 = \frac{1}{2} \min_{1 \leq i \leq k} \min_{\bar{x} \in S_{1i} \cap S_{1j}} \{ |\varphi_i(x_i) - \varphi_j(x_j)| \neq 0 \}\)

and

\[
b_4 = \frac{1}{2} \min_{1 \leq i \leq k} \min_{\bar{x} \in S_{1i} \cap S_{1j}} \{ \varphi_i(x_i) - \varphi_j(x_j) \}. 
\]

We can see that \(b_3 > 0\) because the minimum operation is taken on a finite set. As for \(b_4\), since \(\varphi_j(x_j)\) is strictly decreasing as \(x_j\) increases, and \(\min_{0 \leq i \leq L_{1j}} \varphi_i(x_i) = \varphi_i(L_{1i}) > \theta_0, \max_{\bar{x} \in S_{1i}^j} \varphi_j(x_j) = \varphi_j(L_{1j} + 1) \leq \theta_0\), we have

\[
b_4 = \frac{1}{2} \min_{i \neq j} \min_{\bar{x} \in S_{1i} \cap S_{1j}} \{ \varphi_i(x_i) - \varphi_j(L_{1j} + 1) \} > 0.
\]

Thus,

\[
E_n I_{n,1,3} \leq \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\bar{x} \in S_{1i}} P_n\{i^* = i, i_n^* = j\} (\varphi_i(x_i) - \varphi_j(x_j)) f(\bar{x}) \\
= \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\bar{x} \in (S_{1i} \cap S_{1j})} P_n\{i^* = i, i_n^* = j\} (\varphi_i(x_i) - \varphi_j(x_j)) f(\bar{x})
\]  

(5.19)
\[
+ \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\bar{x} \in (S_{ij} \cap S_{ij}')} P_n\{i^* = i, i_n^* = j\}(\varphi_i(x_i) - \varphi_j(x_j))f(\bar{x})
\]
\[
\leq \sum_{i \neq j} \sum_{\bar{x} \in (S_{ij} \cap S_{ij}')} \{P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i)| \geq b_3\} + P_n\{|\varphi_{jn}(x_j) - \varphi_j(x_j)| \geq b_3\}\}f(\bar{x})
\]
\[
\leq \sum_{i \neq j} \sum_{\bar{x} \in (S_{ij} \cap S_{ij}')} \{P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i)| \geq b_4\} + P_n\{|\varphi_{jn}(x_j) - \varphi_j(x_j)| \geq b_4\}\}f(\bar{x})
\]

Let \(b_5 = \min_{1 \leq i \leq k} \min_{(\bar{x}, \bar{y}) \in S_{2i}} \{\psi_i(x_i + y_i) - \psi_0\}, b_{6i} = \min_{(\bar{x}, \bar{y}) \in S_{2i}} \{\theta_0 - \psi_i(x_i + y_i)\}, \) and \(b_6 = \min_{1 \leq i \leq k} \{b_{6i}\}.\)

Since for \(i = 1, \ldots, k, S_{2i}\) has only finite elements, \(b_5 > 0.\) As for \(D_{2i},\) since \(\psi_i(x_i + y_i)\) is strictly decreasing as \((x_i + y_i)\) increases,

we have

\[
b_{6i} = \begin{cases} 
\theta_0 - \psi_i(L_{2i} + 1), & \text{if } \theta_0 > \psi_i(L_{2i} + 1); \\
\theta_0 - \psi_i(L_{2i} + 2), & \text{if } \theta_0 = \psi_i(L_{2i} + 1).
\end{cases}
\]

Therefore, \(b_{6i} > 0\) for \(i = 1, \ldots, k.\) and \(b_6 = \min_{1 \leq i \leq k} b_{6i} > 0.\)

\[
E_n(I_{n,2}) = \sum_{\bar{x} \in X} E_n(1 - r^n(\bar{x})) \left( \sum_{i=1}^{k} \sum_{y_i=0}^{+\infty} \delta^*_i(\bar{x}) \delta^*_i(\bar{x}, y_i) - d^B_{2i}(\bar{x}, y_i) \right)
\times (\theta_0 - \psi_i(x_i + y_i))f_i(y_i|\bar{x}, \alpha_i, \mu_i)f(\bar{x})
\]
\[
\leq \sum_{\bar{x} \in X} \sum_{i=1}^{k} \sum_{y_i=0}^{+\infty} E_n(d^*_i(\bar{x}, y_i) - d^B_{2i}(\bar{x}, y_i))(\theta_0 - \psi_i(x_i + y_i))
\times f_i(y_i|\bar{x}, \alpha_i, \mu_i)f(\bar{x})
\]
\[
\leq \sum_{i=1}^{k} \sum_{(\bar{x}, \bar{y}) \in S_{2i}} P_n\{\psi_i(x_i + y_i) \geq \theta_0 \text{ and } \psi_{in}(x_i + y_i) \leq \theta_0\}
\times f_i(y_i|\bar{x}, \alpha_i, \mu_i)f(\bar{x})
\]
\[ + \sum_{i=1}^{k} \sum_{(x,y) \in D_{2i}} P_n \{ \psi_i(x_i + y_i) \geq \theta_0 \text{ and } \psi_{in}(x_i + y_i) \leq \theta_0 \} \times f_i(y_i|x, \alpha_i, \mu_i) f(x) \]
\[ \leq \sum_{i=1}^{k} \sum_{(x,y) \in S_{2i}} P_n \{ \psi_i(x_i + y_i) - \psi_{in}(x_i + y_i) > b_3 \} \times f_i(y_i|x, \alpha_i, \mu_i) f(x) \]
\[ \leq \sum_{i=1}^{k} \sum_{(x,y) \in D_{2i}} P_n \{ \psi_{in}(x_i + y_i) - \psi_i(x_i + y_i) > b_6 \} \times f_i(y_i|x, \alpha_i, \mu_i) f(x) \] (5.20)

(II) Analysis for \( II_n \).

From (5.11) and (5.12), for \( i = 1, \ldots, k \), we have

\[ S_{3i} = \{ x \in \mathcal{X} | T_i(x) < 0 \} = \{ x \in X | 0 \leq x_i \leq L_{2i} \text{ and } L_{2i} \geq 0 \} \] (5.21)

Let \( b_7 = \frac{1}{2} \min_{1 \leq i \leq k} \min_{0 \leq x_{i} \leq L_{2i}} \{ T_j(x) - T_i(x) | T_j(x) \neq T_i(x) \} \)

and

\[ b_8 = \frac{1}{2} \min_{1 \leq i \leq k} \min_{0 \leq x_{i} \leq L_{2i}} \min_{x_{j} > L_{2j}} \{ T_j(x) - T_i(x) \} \]

We can see that \( b_7 > 0 \) since the minimum is taken on a finite set. As for \( b_8 \), from (5.12), we have \( T_j(x) = 0 \) while \( T_i(x) < 0 \), therefore,

\[ b_8 = \frac{1}{2} \min_{1 \leq i \leq k} \min_{0 \leq x_{i} \leq L_{2i}} \{ |T_i(x)| \} > 0 \]
Since $|T_j(x) - T_i(x)| \leq 1$, we have

$$E_n(I_{II_n}) \leq \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{x \in \mathcal{X}} P_n\{j^* = i, j_n^* = j\}(T_j(x) - T_i(x))f(x)$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{x \in S_{3i}} P_n\{j^* = i, j_n^* = j\}(T_j(x) - T_i(x))f(x)$$

$$= \sum_{i=1}^{k} \sum_{j \neq i} \sum_{x \in (S_{3i} \cap S_{3j})^c} P_n\{j^* = i, j_n^* = j\}(T_j(x) - T_i(x))f(x)$$

$$+ \sum_{i=1}^{k} \sum_{j \neq i} \sum_{x \in (S_{3i} \cap S_{3j})^c} P_n\{j^* = i, j_n^* = j\}(T_j(x) - T_i(x))f(x)$$

$$\leq \sum_{i=1}^{k} \sum_{j \neq i} \sum_{0 \leq x_i \leq L_{2i}, \atop 0 \leq x_j \leq L_{2j}} P_n\{T_i(x) - T_j(x) < T_i(x)\} f(x) \text{ and } T_{in}(x) \geq T_{jn}(x)\} f(x)$$

$$+ \sum_{i=1}^{k} \sum_{j \neq i} \sum_{0 \leq x_i \leq L_{2i}, \atop x_j > L_{2j}} P_n\{T_i(x) - T_j(x) < T_i(x)\} f(x) \text{ and } T_{in}(x) \geq T_{jn}(x)\} f(x)$$

$$\leq \sum_{i=1}^{k} \sum_{j \neq i} \sum_{0 \leq x_i \leq L_{2i}, \atop x_j > L_{2j}} P_n\{|T_{in}(x) - T_i(x)| \geq \beta_7\} + P_n\{|T_{jn}(x) - T_j(x)| \geq \beta_7\} f(x)$$

$$+ \sum_{i=1}^{k} \sum_{j \neq i} \sum_{0 \leq x_i \leq L_{2i}, \atop x_j > L_{2j}} P_n\{|T_{in}(x) - T_i(x)| \geq \beta_7\} + P_n\{|T_{jn}(x) - T_j(x)| \geq \beta_7\} f(x)$$

$$= (5.22)$$

(III) Analysis for $III_n$

Denote $K(x) = \{i = 1, \ldots, k \mid 0 \leq x_i \leq L_{2i}\}$ and $S_4 = \bigcup_{i=1}^{k} S_{3i}$. Also define $b_9 = \min\{|Q(x)| \neq 0\}$. From (3.5) and (5.12), we know that for $x \in S_4^c$, $Q(x) = -c_2$, and for $x \in S_4$,

$$Q(x) = \sum_{i=0}^{k} d_{1i}^B(x)(\theta_i - \varphi_i(x_i)) - c_2 - \sum_{i=1}^{k} \delta_i^B(x)T_i(x)$$

$$= \sum_{i \in K(x)} d_{1i}^B(x)(\theta_i - \varphi_i(x_i)) - c_2 - \sum_{i \in K(x)} \delta_i^B(x)T_i(x). \tag{5.23}$$

Therefore,
\[ b_9 = \min_{\bar{x} \in S_4} \{|Q(\bar{x})| : Q(\bar{x}) \neq 0\} > 0. \]

Since \(|Q(\bar{x})| \leq 2 + C_2\), we have

\[
E_n(III_n) = \sum_{\bar{x} \in X} E_n(\tau^{*n}(\bar{x}) - T^B(\bar{x})) \ Q(\bar{x})f(\bar{x})
\]

\[
= \sum_{\bar{x} \in S_4} E_n(\tau^{*n}(\bar{x}) - \tau^B(\bar{x})) \ Q(\bar{x})f(\bar{x})
\]

\[
+ \sum_{\bar{x} \in S_4} E_n(\tau^{*n}(\bar{x}) - \tau^B(\bar{x})) \ Q(\bar{x})f(\bar{x})
\]

\[
\leq c_2 \sum_{\bar{x} \in S_4} P_n\{Q_n(\bar{x}) > 0 \ \text{and} \ Q(\bar{x}) = -c_2\}f(\bar{x})
\]

\[
+ (2 + c_2) \sum_{\bar{x} \in S_4} P_n\{\tau^{*n}(\bar{x}) \neq \tau^B(\bar{x}), \ \text{and} \ Q(\bar{x}) \neq 0\}f(\bar{x})
\]

\[
\leq c_2 \sum_{\bar{x} \in S_4} P_n\{|Q_n(\bar{x}) - Q(\bar{x})| \geq c_2\}f(\bar{x})
\]

\[
+ (2 + c_2) \sum_{\bar{x} \in S_4} P_n\{|Q_n(\bar{x}) - Q(\bar{x})| \geq b_9\}f(\bar{x}).
\]

Therefore, in order to investigate the asymptotic optimality of the empirical Bayes two-stage selection procedure \((\tau^{*n}, \delta^{*n}, d_1^{*n}, d_2^{*n})\), it suffices to study the asymptotic behavior of

\[
P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i)| \geq (b_1 \land b_2 \land b_3 \land b_4)\};
\]

\[
P_n\{|\psi_{in}(x_i + y_i) - \psi_i(x_i + y_i)| \geq (b_5 \land b_6)\};
\]

\[
P_n\{|T_{in}(\bar{x}) - T_i(\bar{x})| \geq (b_7 \land b_8)\};
\]

\[
P_n\{|Q_n(\bar{x}) - Q(\bar{x})| \geq (b_9 \land c_2)\}.
\]

We have

\[
\{|\varphi_{in}(x_i) - \varphi_i(x_i)| \geq (b_1 \land b_2 \land b_3 \land b_4)\}
\]

\[
\subset \{|\varphi_{in}(x_i) - \varphi_i(x_i)| \geq (b_1 \land b_2 \land b_3 \land b_4), C_{in} > 0\} \cup \{C_{in} \leq 0\}. \tag{5.25}
\]
Lemma 5.1  For $\epsilon > 0$, there exists a positive constant $q_{i1}(\epsilon, \alpha_i, \mu_i)$, such that for any $x \in \mathcal{X}$, when $C_{in} > 0, |\alpha_i - \alpha_{in}| \leq q_{i1}(\epsilon, \alpha_i, \mu_i)$ and $|\mu_i - \mu_{in}| \leq q_{i1}(\epsilon, \alpha_i, \mu_i)$, we have $|\varphi_{in}(x_i) - \varphi_i(x_i)| \leq \epsilon$ for all $x \in \mathcal{X}$. This property is called equi-continuity. In other words, $\varphi_{in}(x_i)$ is an equi-continuous function of $\alpha_i$ and $\mu_i$.

Proof: Since $\mu_{in} \leq \mu_i + |\mu_i - \mu_{in}|$ and $\alpha_{in} \leq \alpha_i + |\alpha_i - \alpha_{in}|$ we have

$$|\alpha_{in}\mu_{in} - \alpha_i\mu_i| \leq |\alpha_i - \alpha_{in}| \times |\mu_i - \mu_{in}| + \alpha_i|\mu_i - \mu_{in}| + \mu_i|\alpha_i - \alpha_{in}|.$$  (5.26)

Therefore,

$$|\varphi_{in}(x_i) - \varphi_i(x_i)|$$

$$= \left| \frac{\alpha_{in}\mu_{in} + m_1}{\alpha_{in} + m_1 + x_i} - \frac{\alpha_i\mu_i + m_1}{\alpha_i + m_1 + x_i} \right|$$

$$= \left| \frac{\alpha_i\alpha_{in}(\mu_{in} - \mu_i) + (m_1 + x_i)(\alpha_{in}\mu_{in} - \alpha_i\mu_i) + m_1(\alpha_i - \alpha_{in})}{(\alpha_{in} + m_1 + x_i)(\alpha_i + m_1 + x_i)} \right|$$

$$\leq \frac{\alpha_i\alpha_{in}|\mu_{in} - \mu_i| + m_1|\alpha_{in} - \alpha_i| + (m_1 + x_i)|\alpha_{in}\mu_{in} - \alpha_i\mu_i|}{(\alpha_{in} + m_1 + x_i)(\alpha_i + m_1 + x_i)}$$

$$+ \frac{\alpha_i(1 + x_i)|\mu_{in} - \mu_i| + \mu_i(m_1 + x_i)|\alpha_{in} - \alpha_i|}{(\alpha_{in} + m_1 + x_i)(\alpha_i + m_1 + x_i)}.$$  (5.27)

$$= \frac{\alpha_i|\mu_i - \mu_{in}|((\alpha_{in} + m_1 + x_i) + m_1|\alpha_i - \alpha_{in}|)}{(\alpha_{in} + m_1 + x_i)(\alpha_i + m_1 + x_i)}$$

$$+ \frac{(m_1 + x_i)|\alpha_i - \alpha_{in}| \cdot |\mu_i - \mu_{in}| + \mu_i(m_1 + x_i)|\alpha_i - \alpha_{in}|}{(\alpha_{in} + m_1 + x_i)(\alpha_i + m_1 + x_i)}$$

$$= \frac{\alpha_i \cdot |\mu_i - \mu_{in}|}{\alpha_i + m_1 + x_i} + \frac{m_1 + (m_1 + x_i)|\mu_i - \mu_{in}|}{(\alpha_{in} + m_1 + x_i)(\alpha_i + m_1 + x_i)}$$

$$\leq \frac{\alpha_i \cdot |\mu_i - \mu_{in}|}{\alpha_i + m_1} + \frac{(1 + |\mu_i - \mu_{in}| + \mu_i)}{(\alpha_i + m_1 + x_i)}$$

From the above we see that for any $\epsilon > 0$, there exists a positive constant $q_{i1}(\epsilon, \alpha_i, \mu_i)$ such that for any $x \in \mathcal{X}$, when $C_{in} > 0, |\alpha_i - \alpha_{in}| \leq q_{i1}(\epsilon, \alpha_i, \mu_i)$ and $|\mu_i - \mu_{in}| \leq q_{i1}(\epsilon, \alpha_i, \mu_i)$, we have $|\varphi_{in}(x_i) - \varphi_i(x_i)| \leq \epsilon$. 

21
From Lemma 5.1, when $C_{in} > 0$, there exists a positive constant $q_{i1} = q_{i1}((b_1 \land b_2 \land b_3 \land b_4), \alpha_i, \mu_i)$ such that

$$\{ |\varphi_{in}(x_i) - \varphi_i(x_i) | \geq (b_1 \land b_2 \land b_3 \land b_4), C_{in} > 0 \}$$

$$\subset \{ |\alpha_{in} - \alpha_i | \geq q_{i1}, C_{in} > 0 \} \cup \{ |\mu_{in} - \mu_i | \geq q_{i1}, C_{in} > 0 \}. \quad (5.28)$$

Hence,

$$P_n \{ |\varphi_{in}(x_i) - \varphi_i(x_i) | \geq (b_1 \land b_2 \land b_3 \land b_4) \} \leq P_n \{ |\alpha_{in} - \alpha_i | \geq q_{i1}, C_{in} > 0 \}$$

$$+ P_n \{ |\mu_{in} - \mu_i | \geq q_{i1}, C_{in} > 0 \} \quad (5.29)$$

$$+ P_n \{ C_{in} \leq 0 \}.$$

Similarly, when $C_{in} > 0$, there exists a positive constant $q_{i2} = q_{i2}((b_5 \land b_6), \alpha_i, \mu_i)$ such that

$$\{ |\psi_{in}(x_i + y_i) - \psi_i(x_i + y_i) | \geq (b_5 \land b_6), C_{in} > 0 \}$$

$$\subset \{ |\alpha_{in} - \alpha_i | \geq q_{i2}, C_{in} > 0 \} \cup \{ |\mu_{in} - \mu_i | \geq q_{i2}, C_{in} > 0 \}, \quad (5.30)$$

and

$$P_n \{ |\psi_{in}(x_i + y_i) - \psi_i(x_i + y_i) | \geq (b_5 \land b_6) \} \leq P_n \{ |\alpha_{in} - \alpha_i | \geq q_{i2}, C_{in} > 0 \}$$

$$+ P_n \{ |\mu_{in} - \mu_i | \geq q_{i2}, C_{in} > 0 \} \quad (5.31)$$

$$+ P_n \{ C_{in} \leq 0 \}.$$

Using the similar approach, we can obtain the following two lemmas.

**Lemma 5.2** For any $\epsilon > 0$, there exists a positive constant $q_{i3}(\epsilon, \alpha_i, \mu_i)$ such that when $C_{in} > 0, |\alpha_{in} - \alpha_i | \leq q_{i3}(\epsilon, \alpha_i, \mu_i)$ and $|\mu_{in} - \mu_i | \leq q_{i3}(\epsilon, \alpha_i, \mu_i)$, we have, for all $\tilde{x} \in \mathcal{X},$

$$|T_{in}(\tilde{x}) - T_i(\tilde{x}) | \leq \epsilon. \quad (5.32)$$

**Lemma 5.3** For any $\epsilon > 0$, there exists a positive constant $q_{4}(\epsilon, \alpha, \mu)$ such that when $C_{in} > 0, |\alpha_{in} - \alpha_i | \leq q_{4}(\epsilon, \alpha, \mu)$ and $|\mu_{in} - \mu_i | \leq q_{4}(\epsilon, \alpha, \mu)$, we have $|Q_n(\tilde{x}) - Q(\tilde{x}) | \leq \epsilon.$
From Lemma 5.2 and 5.3, if we let \( q_{i3} = q_{i3}(b_7 \land b_8, \alpha_i, \mu_i) \) and \( q_4 = q_4(b_9 \land c_2, \varphi, \eta) \), we can immediately obtain

\[
P_n \{ |T_{in}(\underline{x}) - T_i(\underline{x})| \geq (b_7 \land b_8) \} \\
\leq P_n \{ |\alpha_{in} - \alpha_i| \geq q_{i3}, C_{in} > 0 \} \\
+ P_n \{ |\mu_{in} - \mu_i| \geq q_{i3}, C_{in} > 0 \} \\
+ P_n \{ C_{in} \leq 0 \}
\]

(5.33)

and

\[
P_n \{ |Q_{n}(\underline{x}) - Q(\underline{x})| \geq (b_9 \land c_2) \} \\
\leq \sum_{i=1}^{k} P_n \{ |\alpha_{in} - \alpha_i| > q_4, C_{in} > 0 \} \\
+ \sum_{i=1}^{k} P_n \{ |\mu_{in} - \mu_i| > q_4, C_{in} > 0 \} \\
+ \sum_{i=1}^{k} P_n \{ C_{in} \leq 0 \}.
\]

(5.34)

From (5.29), (5.31), (5.33) and (5.34), it suffices to investigate the asymptotic behavior of \( P_n \{ |\alpha_{in} - \alpha_i| > \epsilon, C_{in} > 0 \} \), \( P_n \{ |\mu_{in} - \mu_i| > \epsilon, C_{in} > 0 \} \), and \( P_n \{ C_{in} \leq 0 \} \) for each \( i = 1, \ldots, k \), where \( \epsilon > 0 \).

5.2 Lemmas

In this section, we introduce some lemmas which are helpful to investigate the asymptotic behavior. The following lemma is from Hoeffding.

**Lemma 5.4** (Hoeffding Inequality) If random variables \( Z_1, \ldots, Z_n \) are i.i.d., such that \( a \leq Z_i \leq b, i = 1, \ldots, k \), then for any \( \epsilon > 0 \).

\[
P_n \left\{ \bar{Z} - \mu \geq \epsilon \right\} \leq \exp(-2n\epsilon^2/(b - a)^2),
\]

(5.35)

where

\[
\bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i \quad \text{and} \quad \mu = E(Z).
\]

**Proof:** This lemma is a special case of Theorem 1 of Hoeffding (1963).
Lemma 5.5 Let $\mu_i, \nu_i, \mu_{in}$ and $\nu_{in}$ be defined as in Section 4, respectively. Then for any $\epsilon > 0$,

\begin{align*}
\text{a)} \quad P_n\{\mu_{in} - \mu_i \leq -\epsilon\} &\leq O(\exp(-2n\epsilon^2)), \\
\text{b)} \quad P_n\{\mu_{in} - \mu_i \geq \epsilon\} &\leq O(\exp(-2n\epsilon^2)), \\
\text{c)} \quad P_n\{\nu_{in} - \mu_i \leq -\epsilon\} &\leq O(\exp(-2n\epsilon^2)), \\
\text{d)} \quad P_n\{\nu_{in} - \mu_i \geq \epsilon\} &\leq O(\exp(-2n\epsilon^2)).
\end{align*}

Proof: We only prove part b. The proof of the other inequalities are analogous.

For each $i = 1, \ldots, k$, and $j = 1, \ldots, n$, denote $Z_{ij} = (m_{1+X_{ij}-2})/(m_{1+X_{ij}-1})$. Then we can see that $Z_{ij} = (m_1 - 1)/(m_1 + X_{ij} - 1)$, from which we have $0 \leq Z_{ij} \leq 1$. For $i = 1, \ldots, k$, $(Z_{i1}, Z_{i2}, \ldots, Z_{in})$ are i.i.d. Therefore, from Lemma 5.4, we obtain the proof of part b.

Lemma 5.6 Let $C_i, D_i, C_{in}$ and $D_{in}$ be defined as in Section 4, respectively. Then for any $\epsilon > 0$, we have

\begin{align*}
\text{a)} \quad P_n\{C_{in} - C_i \leq -\epsilon\} &\leq O(\exp(-n\epsilon^2/8)), \\
\text{b)} \quad P_n\{C_{in} - C_i \geq \epsilon\} &\leq O(\exp(-n\epsilon^2/8)), \\
\text{c)} \quad P_n\{D_{in} - D_i \leq -\epsilon\} &\leq O(\exp(-n\epsilon^2/2)), \\
\text{d)} \quad P_n\{D_{in} - D_i \geq \epsilon\} &\leq O(\exp(-n\epsilon^2/2)).
\end{align*}

Proof: The techniques used to prove the four inequalities are similar. So we only give the proof of part a. We have

\begin{align*}
P_n\{C_{in} - C_i \leq -\epsilon\} &
= P_n\{(\nu_{in} - \nu_i) + (1 - \mu_{in})^2 - (1 - \mu_i)^2 \leq -\epsilon\} \\
&\leq P_n\{\nu_{in} - \nu_i \leq -\epsilon/2\} + P_n\{\mu_{in} - \mu_i \leq -\epsilon/2\}.
\end{align*}

(5.36)

Since $0 \leq \mu_{in} \leq 1$, $0 < \mu_i < 1$, $0 < (2 - \mu_{in} - \mu_i) < 2$,

\begin{align*}
P_n\{\mu_{in} - \mu_i \leq -\epsilon/2\} &\leq \frac{P_n\{(\mu_{in} - \mu_i)(2 - \mu_{in} - \mu_i) \leq -\epsilon/2\}}{\epsilon/4}.
\end{align*}
Thus,
\[ P_n\{C_{in} - C_i \leq -\epsilon\} \]
\[ \leq P_n\{(\mu_i - \mu_i) \leq -\epsilon/4\} + P_n\{(\nu_{in} - \nu_i) \leq -\epsilon/2\} \]  
(5.37)
\[ \leq O(\exp(-n\epsilon^2/8)). \]

**Corollary 5.7** From Lemma 5.6, we have
\[ P_n\{C_{in} \leq 0\} = O(\exp(-nC_i^2/8)). \]  
(5.38)

**Lemma 5.8** For \( \epsilon > 0 \),
\[ P_n\{|\alpha_{in} - \alpha_i| > \epsilon, C_{in} > 0\} = O\left(\exp\left(-\frac{n\epsilon^2C_i^2}{8}\min(1, \frac{1}{4(\alpha_i + \epsilon)^2})\right)\right). \]  
(5.39)

**Proof:**
\[ P_n\{|\alpha_{in} - \alpha_i| > \epsilon, C_{in} > 0\} \]
\[ = P_n\{\alpha_{in} - \alpha_i < -\epsilon, C_{in} > 0\} + P_n\{\alpha_{in} - \alpha_i > \epsilon, C_{in} > 0\}, \]  
(5.40)
where \( P_n\{\alpha_{in} - \alpha_i < -\epsilon, C_{in} > 0\} = 0 \), if \( \alpha_i - \epsilon \leq 0 \). When \( \alpha_i - \epsilon > 0 \), by the Bonferroni inequality, we have
\[ P_n\{\alpha_{in} - \alpha_i < -\epsilon, C_{in} > 0\} \]
\[ \leq P_n\{(D_{in} - D_i) - (C_{in} - C_i)(\alpha_i - \epsilon) < -\epsilon C_i\} \]
\[ \leq P_n\{D_{in} - D_i \leq -\epsilon C_i/2\} + P_n\{C_{in} - C_i > \epsilon C_i/(2(\alpha_i - \epsilon))\} \]  
(5.41)
\[ = O(\exp(-n\epsilon^2C_i^2/8)) + O(\exp(-n\epsilon^2C_i^2/(32(\alpha_i - \epsilon)^2))) \]
\[ = O\left(\exp\left(-\frac{n\epsilon^2C_i^2}{8}\min(1, \frac{1}{4(\alpha_i + \epsilon)^2})\right)\right). \]

Similarly, we have
\[ P_n\{\alpha_{in} - \alpha_i > \epsilon, C_{in} > 0\} \]
\[ \leq P_n\{(D_{in} - D_i) - (C_{in} - C_i)(\alpha_i + \epsilon) > \epsilon C_i\} \]
\[ \leq P_n\{D_{in} - D_i > \epsilon C_i/2\} + P_n\{C_{in} - C_i < -\epsilon C_i/(2(\alpha_i + \epsilon))\} \]  
(5.42)
\[ = O\left(\exp\left(-\frac{n\epsilon^2C_i^2}{8}\min(1, \frac{1}{4(\alpha_i + \epsilon)^2})\right)\right). \]
This completes the proof of Lemma 5.8.

Define \( e_1 = \min_{1 \leq i \leq k} q_{i1}, \ e_2 = \min_{1 \leq i \leq k} q_{i2}, \ e_3 = \min_{1 \leq i \leq k} q_{i3}, \ e_4 = \frac{1}{2} \min_{1 \leq i \leq k} \{ \alpha_i \}, \) and \( e^* = \min_{1 \leq j \leq 4} \{ e_j \}. \) We see that \( e^* > 0. \)

In the following theorem we prove the asymptotic optimality of the empirical Bayes two-stage procedure \((\tau^* n, \xi^* n, d_1^* n, d_2^* n)\).

**Theorem 5.1** Let \( \{(\tau^*_n, \xi^*_n, d^{*n}_1, d^{*n}_2)\}_{n=1}^{\infty} \) be the sequence of empirical Bayes two-stage selection procedures constructed in Section 4. Then,

\[
E_n R(\tau^*_n, \xi^*_n, d^{*n}_1, d^{*n}_2) - R(\tau^B, \xi^B, d^B_1, d^B_2) = O(\exp(-c^* n))
\]

where \( c^* = \min \left( \frac{e^2}{8}, \min_{1 \leq i \leq k} \frac{C_i^2}{8}, \min_{1 \leq i \leq k} \left( \frac{e^2 C_i^2}{8} \min(1, \frac{1}{\alpha_i}) \right) \right) > 0. \)

**Proof:** By the definition of \( e^* \) and \( C^* \), from Lemma (5.5) to (5.8), we have

\[
P_n \{ |\varphi_{in}(x_i) - \varphi_i(x_i)| \geq (b_1 \wedge b_2 \wedge b_3 \wedge b_4) \} = O(\exp(-c^* n)),
\]

\[
P_n \{ |\psi_{in}(x_i + y_i) - \psi_i(x_i + y_i)| \geq (b_5 \wedge b_6) \} = O(\exp(-c^* n)),
\]

\[
P_n \{ |T_{in}(z) - T_i(z)| \geq (b_7 \wedge b_8) \} = O(\exp(-c^* n)),
\]

\[
P_n \{ |Q_n(z) - Q(z)| \geq (b_9 \wedge c_2) \} = O(\exp(-c^* n)).
\]

We obtain

\[
E_n R(\tau^*_n, \xi^*_n, d^{*n}_1, d^{*n}_2) - R(\tau^B, \xi^B, d^B_1, d^B_2)
\]

\[
\leq O(\exp(-c^* n)) k^2 \sum_{x \in \mathcal{X}} f(x) + O(\exp(-c^* n)) \sum_{x \in \mathcal{X}} \sum_{i=1}^{k} \left[ \sum_{y_i=0}^{+\infty} f_i(y_i|x, \alpha_i, \mu_i) \right] f(x),
\]

\[
+ O(\exp(-c^* n)) \sum_{x \in \mathcal{X}} (2 + c_2) f(x)
\]

\[
= O(\exp(-c^* n)).
\]
This completes the proof of the theorem.

6. Simulation Study

A simulation study was carried out to investigate the performance of the proposed empirical Bayes two-stage selection procedures for small to moderate values of $n$.

For given past observations $(X_j = (X_{1j}, \ldots, X_{kj}), j = 1, \ldots, n)$, $R(\tau^n, \delta^n, \tilde{d}_1^n, \tilde{d}_2^n)$ is the associated conditional Bayes risk of the proposed selection procedure $(\tau^B, \delta^B, \tilde{d}_1^B, \tilde{d}_2^B)$. Then we use $R(\tau^n, \delta^n, \tilde{d}_1^n, \tilde{d}_2^n) - R(\tau^B, \delta^B, \tilde{d}_1^B, \tilde{d}_2^B)$ as an estimator of the difference $E_n R(\tau^n, \delta^n, \tilde{d}_1^n, \tilde{d}_2^n) - R(\tau^B, \delta^B, \tilde{d}_1^B, \tilde{d}_2^B)$.

We consider the following case in which $k = 3$. That is, there are 3 Bernoulli treatments $\pi_1, \pi_2$ and $\pi_3$, and we would like to make a two-stage selection using the inverse binomial sampling scheme proposed in Section 4.

The scheme of the simulation is described as follows:

1. For each $n$ and for each $i = 1, 2, 3$, generate independent random variables $X_{i1}, X_{i2}, \ldots, X_{in}$ as follows:

   \[
   \begin{cases}
   \text{for } j = 1, \ldots, n, \\
   \quad (a) \text{ first generate } \Theta_{ij} \text{ from a Beta distribution with density } h_i(\theta_i|\alpha_i, \mu_i) \\
   \quad (b) \text{ then generate } X_{ij} \text{ from a negative binomial } \text{NB}(m_1, \theta_{ij}) \text{ distribution.}
   \end{cases}
   \]

2. Based on the past observations $(X_j, j = 1, \ldots, n)$ and the present observations $X = (X_1, \ldots, X_k)$ and $Y = (Y_1, \ldots, Y_k)$, we construct the empirical Bayes two-stage selection procedure $(\tau^{*n}, \delta^{*n}, \tilde{d}^{*n}_1, \tilde{d}^{*n}_2)$ and compute the conditional difference

   \[
   D(n) = R(\tau^{*n}, \delta^{*n}, \tilde{d}^{*n}_1, \tilde{d}^{*n}_2) - R(\tau^B, \delta^B, \tilde{d}^B_1, \tilde{d}^B_2).
   \]

3. Repeat step (1) and (2) 5000 times. The average of the conditional difference on the 5000 repetitions which is denoted by $\overline{D}(n)$, is used as an estimator of the difference $E_n R(\tau^{*n}, \delta^{*n}, \tilde{d}^{*n}_1, \tilde{d}^{*n}_2) - R(\tau^B, \delta^B, \tilde{d}^B_1, \tilde{d}^B_2)$. The estimated standard error, denoted by $SE(\overline{D}(n))$, is also computed.

Tables (1) and (2) give the results of this simulation study on the performance of the proposed empirical Bayes two-stage selection procedures. In both cases we choose

27
\(\theta_0 = 0.7, c_1 = c_2 = 0.05. m_1 = m_2 = 20, \text{ and } \alpha_1 = \alpha_2 = \alpha_3 = 3.\) Furthermore, we use 
\(\mu_1 = \mu_2 = \mu_3 = 0.6\) in Table 1 and \(\mu_1 = 0.67 \mu_2 = 0.69 \text{ and } \mu_3 = 0.71\) in Table 2.

From these results, we see that \(\bar{D}(n)\) decreases to zero very rapidly in both cases, which coincides with theorem 5.1 that the convergence rate is \(O(\exp(-c^* n)), \ c^* > 0.\)
Table 1
Performance of the selection rule
when $\mu_1 = \mu_2 = \mu_3 = 0.6$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\bar{D}(n)$</th>
<th>$SE(\bar{D}(n))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.04200000</td>
<td>0.000456063</td>
</tr>
<tr>
<td>10</td>
<td>0.01767500</td>
<td>0.000045031</td>
</tr>
<tr>
<td>15</td>
<td>0.00244500</td>
<td>0.000003197</td>
</tr>
<tr>
<td>20</td>
<td>0.00076700</td>
<td>0.000001487</td>
</tr>
<tr>
<td>30</td>
<td>0.00110000</td>
<td>0.000003565</td>
</tr>
<tr>
<td>40</td>
<td>0.00049980</td>
<td>0.000000801</td>
</tr>
<tr>
<td>50</td>
<td>0.00051410</td>
<td>0.000000265</td>
</tr>
<tr>
<td>60</td>
<td>0.00042840</td>
<td>0.000000943</td>
</tr>
<tr>
<td>70</td>
<td>0.00029078</td>
<td>0.000000101</td>
</tr>
<tr>
<td>80</td>
<td>0.00019035</td>
<td>0.000000127</td>
</tr>
<tr>
<td>90</td>
<td>0.00016365</td>
<td>0.000000081</td>
</tr>
<tr>
<td>100</td>
<td>0.00016000</td>
<td>0.000000037</td>
</tr>
<tr>
<td>125</td>
<td>0.00009243</td>
<td>0.000000012</td>
</tr>
<tr>
<td>150</td>
<td>0.00008065</td>
<td>0.000000009</td>
</tr>
</tbody>
</table>

Table 2
Performance of the selection rule
when $\mu_1 = 0.67$, $\mu_2 = 0.69$, $\mu_3 = 0.71$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\bar{D}(n)$</th>
<th>$SE(\bar{D}(n))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.0576700</td>
<td>0.000332773</td>
</tr>
<tr>
<td>10</td>
<td>0.0124585</td>
<td>0.000080474</td>
</tr>
<tr>
<td>15</td>
<td>0.0075045</td>
<td>0.000082546</td>
</tr>
<tr>
<td>20</td>
<td>0.0024041</td>
<td>0.00001701</td>
</tr>
<tr>
<td>30</td>
<td>0.0005337</td>
<td>0.00000883</td>
</tr>
<tr>
<td>40</td>
<td>0.0008987</td>
<td>0.00000967</td>
</tr>
<tr>
<td>50</td>
<td>0.0009491</td>
<td>0.000003398</td>
</tr>
<tr>
<td>60</td>
<td>0.0004567</td>
<td>0.00000698</td>
</tr>
<tr>
<td>70</td>
<td>0.0004036</td>
<td>0.000001782</td>
</tr>
<tr>
<td>80</td>
<td>0.0003893</td>
<td>0.000000979</td>
</tr>
<tr>
<td>90</td>
<td>0.0001930</td>
<td>0.00000135</td>
</tr>
<tr>
<td>100</td>
<td>0.0001525</td>
<td>0.00000051</td>
</tr>
<tr>
<td>125</td>
<td>0.0000012</td>
<td>0.00000011</td>
</tr>
<tr>
<td>150</td>
<td>0.0000000</td>
<td>0.00000000</td>
</tr>
</tbody>
</table>
Graph for Table 1

$\mu_1 = \mu_2 = \mu_3 = 0.6$

Rate of Convergence for Case 1
Graph for Table 2

$\mu_1 = 0.67, \mu_2 = 0.69, \mu_3 = 0.71$

Rate of Convergence for Case 2
References


REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)

4. TITLE AND SUBTITLE

Empirical Bayes Two-Stage Selection Procedures for Selecting the Best Bernoulli Treatment Using the Inverse Binomial Sampling

5. FUNDING NUMBERS

Shanti S. Gupta, Ta Chen Liang and Xun Lin

DAAH04-95-01-0165

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)

Purdue University
1399 Mathematical Sciences Building
West Lafayette, IN 47907-1399

3. PERFORMING ORGANIZATION REPORT NUMBER

8. AUTHOR(S)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

In this paper we investigate the problem of selecting the treatment with the largest probability of success from \( k \geq 2 \) independent Bernoulli treatments. The selected treatment must also be better than a given control. We employ the empirical Bayes approach and develop a two-stage selection procedure. We prove that the proposed selection rule is asymptotically optimal at the rate of convergence of order \( O(\exp(-c^*n)) \), for some positive constant \( c^* \), where \( n \) is the number of the historical data at hand. We also carry out a simulation study to investigate the performance of the proposed empirical Bayes selection procedure for small to moderate values of \( n \). The simulated results are provided in the paper.

13. ABSTRACT (Maximum 200 words)

14. SUBJECT TERMS

Asymptotically optimal; empirical Bayes; two-stage selection procedure; rate of convergence; inverse binomial sampling

15. NUMBER OF PAGES

33

16. PRICE CODE

UL

17. SECURITY CLASSIFICATION OR REPORT

UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT

UNCLASSIFIED

20. LIMITATION OF ABSTRACT

UL

NSN 7540-01-280-5500

Enclosure 1

Standard Form 298 (Rev. 2-88)
Prepared by JNSI 7/8/98