EMPIRICAL BAYES RULES FOR SELECTING GOOD BINOMIAL POPULATIONS*

Shanti S. Gupta

Purdue University

and

TaChen Liang

University of Michigan

This paper deals with the problem of selecting good binomial populations compared with a standard or a control through the empirical Bayes approach. Two cases have been studied: one with the prior distribution completely unknown and the other with the prior distribution symmetrical about $p = 1/2$, but otherwise unknown. In each case, empirical Bayes rules are derived and their rates of convergence are shown to be of order $O(\exp(-cn))$ for some $c > 0$, where n is the number of accumulated post experiences at hand.

1. Introduction.

The empirical Bayes approach in statistical decision theory is appropriate when one is confronted repeatedly and independently with the same decision problem. 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 approach is due to Robbins (1956, 1964, 1983). Many such empirical Bayes rules have been shown

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

AMS 1980 Subject Classification. Primary 62F07; Secondary 62C12.

Key Words: Bayes rule, empirical Bayes rules, prior distribution, asymptotically optimal, rate of convergence.

to be asymptotically optimal in the sense that the risk for the n-th decision problem converges to the optimal Bayes risk which would have been obtained if the prior distribution was known and the Bayes rule with respect to this prior distribution was used.

Empirical Bayes rules have been derived for multiple decision problems by Deely (1965) for selecting a subset containing the best population. Van Ryzin (1970), Huang (1975), Van Ryzin and Susarla (1977) and Singh (1977) also studied other multiple decision problems by using the empirical Bayes approach. Recently, Gupta and Hsiao (1983) and Gupta and Leu (1983) studied empirical Bayes rules for selecting good populations with respect to a standard or a control with the underlying populations being uniformly distributed.

In this paper, we are concerned with the problem of selecting good binomial populations with respect to a control through the empirical Bayes approach. Two cases have been studied: one with the prior distribution completely unknown and the other with the prior distribution symmetrical about $p = 1/2$, but otherwise unknown. In each case, empirical Bayes rules are derived and their rates of convergence are shown to be of order $O(\exp(-c_i n))$ for some $c_i > 0$, $i=1,2$. For the case of the symmetrical prior distribution two smoothing methods are studied in order to improve the performance of the sequence of empirical Bayes rules.

2. Formulation of the Empirical Bayes Approach.

Let $\pi_0, \pi_1, \ldots, \pi_k$ denote $k + 1$ populations and let X_i be a random observation from π_i. Assume that $X_i \sim B(N_i, p_i)$, where $p_i \in (0,1)$ and N_i is fixed and known. Let π_0 be the control population. For each $i=1,\ldots,k$, population π_i is said to be good if $p_i \geq p_0$ and bad if $p_i < p_0$, where the control parameter p_0 is either known or unknown. Our goal is to derive some empirical Bayes rules to select all the good populations and exclude all the bad populations.

When the control parameter p_0 is known, the empirical Bayes framework can be formulated as follows:

(1) Let $\Omega = \{\underset{\sim}{p} | \underset{\sim}{p} = (p_1,\dots,p_k),\ p_i \in (0,1) \text{ for } i = 1,2,\dots,k\}$. For each $\underset{\sim}{p} \in \Omega$, $\subset$define $A(\underset{\sim}{p}) = \{i | p_i > p_0\}$, $B(\underset{\sim}{p}) = \{i | p_i < p_0\}$. That is, $A(\underset{\sim}{p})(B(\underset{\sim}{p}))$ is the set of indices of good (bad) populations.

(2) Let $A = \{a | a \subset \{1,2,\dots,k\}\}$ be the action space. When action a is taken, it means that population π_i is selected as a good population if $i \in a$, and excluded as a bad population if $i \notin a$.

(3) The loss function $L(\underset{\sim}{p},a)$ is defined as follows:

$$L(\underset{\sim}{p},a) = \sum_{i \in A(\underset{\sim}{p})-a} (p_i - p_0) + \sum_{i \in a-A(\underset{\sim}{p})} (p_0 - p_i) \tag{2.1}$$

where the first summation is the loss due to not selecting some good populations and the second summation is the loss due to selecting some bad populations.

(4) Let $dG(\underset{\sim}{p}) = \prod_{i=1}^{k} dG_i(p_i)$ be the prior distribution over the parameter space Ω, where $G_i(\cdot)$ are unknown for all $i=1,2,\dots,k$.

(5) For each i, let (X_{ij},P_{ij}), $j=1,2,\dots$, be pairs of random variables associated with population π_i, where X_{ij} is observable but P_{ij} is not observable. P_{ij} has distribution G_i. Conditional on $P_{ij} = p_{ij}$, $X_{ij}|p_{ij}$ is binomially distributed with parameters N_i and p_{ij}. For the case where the prior distributions G_i's are completely unknown, some additional observations $\underset{\sim}{Y}_{ij} = (Y_{ij1},\dots,Y_{ijn_i})$ from each population π_i, $i=1,2,\dots,k$, are assumed to be at hand, where $Y_{ijm}|p_{ij}$, $m=1,\dots,n_i$, are i.i.d., independent of $X_{ij}|p_{ij}$ and follow $B(1,p_{ij})$ distribution. Thus, in this case, the j-th stage observations are $\underset{\sim}{Z}_j = ((X_{1j},\underset{\sim}{Y}_{1j}),\dots,(X_{kj},\underset{\sim}{Y}_{kj}))$. For the second case where G_i's are assumed to be symmetric about $P = 1/2$, no additional data are needed for the construction of our empirical Bayes rule.

(6) Let $\underset{\sim}{X} = (X_1,\dots,X_k)$ be the present observation. Conditional on $\underset{\sim}{p} = (p_1,\dots,p_k)$, $\underset{\sim}{X}$ has joint probability function $f(\underset{\sim}{x}|\underset{\sim}{p}) = \prod_{i=1}^{k} f_i(x_i|p_i)$ where $f_i(x|p) = \binom{N_i}{x} p^x (1-p)^{N_i - x}$ for each $i=1,\dots,k$.

Finally, since we are interested in Bayes rules, we can restrict our attention to the nonrandomized rules.

(7) Let $D = \{d | d : \chi \to A$, being measurable$\}$ be the set of nonrandomized rules, where $\chi = \prod_{i=1}^{k} \{0,1,\ldots,N_i\}$, 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. When the control parameter p_0 is unknown, for the related framework, the indices in the associated notations should begin at 0 instead of at 1. In the sequel, (0) will be used to show this additional fact.

We now consider empirical Bayes decision rule $d_n(\underset{\sim}{x}, \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n)$ whose form depends on $\underset{\sim}{x}$ and $\underset{\sim}{Z}_j$, $j=1,\ldots,n$. Let $r(G,d_n)$ be the Bayes risk associated with decision rule $d_n(\underset{\sim}{x}, \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n)$. That is,

$$r(G,d_n) \equiv \sum_{\underset{\sim}{x}\in\chi} E \int_{\Omega} L(\underset{\sim}{p},d_n(\underset{\sim}{x},\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n))\ f(\underset{\sim}{x}|\underset{\sim}{p})\ dG(\underset{\sim}{p})$$

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

<u>Definition 2.1</u>. A sequence of decision rules $\{d_n(\underset{\sim}{x})\}_{n=1}^{\infty}$ is said to be asymptotically optimal (a.o.) relative to the prior distribution G if $r(G,d_n) \to r(G)$ as $n \to \infty$.

For constructing a sequence of a.o. rules, we first need to find the minimum Bayes risk and the associated Bayes rule, say d_G. From (2.1), the Bayes risk associated with decision rule d is

$$r(G,d) = \sum_{\underset{\sim}{x}\in\chi} \ \sum_{i\in d(\underset{\sim}{x})} \Delta_{iG}(\underset{\sim}{x}) \prod_{\substack{j=1\\ j\neq i}}^{k} f_j(x_j) + C, \tag{2.2}$$

where

$$\Delta_{iG}(\underset{\sim}{x}) = \begin{cases} p_0 f_i(x_i) - W_i(x_i) & \text{if } p_0 \text{ is known;} \\ W_0(x_0)f_i(x_i) - W_i(x_i)f_0(x_0) & \text{if } p_0 \text{ is unknown;} \end{cases} \tag{2.3}$$

$$f_i(x) = \int_0^1 f_i(x|p)\ dG_i(p);$$

$$W_i(x) = \int_0^1 pf_i(x|p)dG_i(p);$$

and

$$C = \sum_{\underset{\sim}{x}\varepsilon\chi} \sum_{i=1}^{k} \int_{\Omega} (p_i - p_0) I_{(p_0,1)}(p_i) f(\underset{\sim}{x}|\underset{\sim}{p}) dG(\underset{\sim}{p}).$$

Hence, the Bayes rule d_G can be obtained as follows:

$$d_G(\underset{\sim}{x}) = \{i|\Delta_{iG}(\underset{\sim}{x}) < 0\}. \tag{2.4}$$

Now, for each $i = (0), 1, \ldots, k$, and for each $n=1,2,\ldots$, let $W_{in}(x_i) \equiv W_{in}(x_i; (X_{i1}, \underset{\sim}{Y}_{i1}), \ldots, (X_{in}, \underset{\sim}{Y}_{in}))$ be an estimator of $W_i(x_i)$ and $f_{in}(x_i) \equiv f_{in}(x_i; (X_{i1}, \underset{\sim}{Y}_{i1}), \ldots, (X_{in}, \underset{\sim}{Y}_{in}))$ be an estimator of $f_i(x_i)$. Define

$$\Delta_{in}(\underset{\sim}{x}) = \begin{cases} W_{0n}(x_0)f_{in}(x_i) - W_{in}(x_i)f_{0n}(x_0) & \text{if } p_0 \text{ is unknown;} \\ p_0 f_{in}(x_i) - W_{in}(x_i) & \text{if } p_0 \text{ is known;} \end{cases} \tag{2.5}$$

and

$$d_n(\underset{\sim}{x}) = \{i|\Delta_{in}(\underset{\sim}{x}) < 0\}. \tag{2.6}$$

If $W_{in}(x) \xrightarrow{p} W_i(x)$ and $f_{in}(x) \xrightarrow{p} f_i(x)$ for all $x=0,1,\ldots,N_i$ where " $\xrightarrow{p}$ " means convergence in probability, then $\Delta_{in}(\underset{\sim}{x}) \xrightarrow{p} \Delta_{iG}(\underset{\sim}{x})$ for all $\underset{\sim}{x}\varepsilon\chi$. Therefore, from Corollary 2 of Robbins (1964), it follows that $r(G,d_n) \to r(G)$ as $n \to \infty$. So, the sequence of decision rules $\{d_n(\underset{\sim}{x})\}$ defined in (2.6) is asymptotically optimal for our selection problem. Hence, in the following, we have only to find sequences of estimators $\{W_{in}(x)\}$ and $\{f_{in}(x)\}$ possessing the above mentioned convergence property.

3. Case when the Prior Distribution is Completely Unknown.

Robbins (1964) and Samuel (1963), respectively, pointed out that there was no way of approximating $W_i(x)$ just by using the observations $(X_{i1},\ldots,X_{in})$. In order to remedy this deficiency, we take, at each stage, some more observations $(Y_{ij1},\ldots,Y_{ijn_i})$ in our model where n_i can be any positive integer. For simplicity, let $n_i = 1$ for all $i = (0), 1,\ldots,k$.

Estimation of $W_i(x)$ and $f_i(x)$. A usual estimator of $f_i(x)$ can be given as follows:

$$f_{in}(x) = \frac{1}{n} \sum_{j=1}^{n} I_{\{x\}}(X_{ij}) \quad \text{for } x = 0,1,\ldots,N_i. \tag{3.1}$$

Then $f_{in}(x)$ is an unbiased estimator of $f_i(x)$, and by the strong law of large numbers, $f_{in}(x) \to f_i(x)$ with probability 1 for each $x=0,1,\ldots,N_i$. Hence, $f_{in}(x) \overset{p}{\to} f_i(x)$ for all $x=0,1,\ldots,N_i$.

For the estimation of $W_i(x)$, we consider the following. Define

$$V_{ij}(x) = Y_{ij} I_{\{x\}}(X_{ij}). \tag{3.2}$$

Under the assumption (5) of Section 2, it is easy to see that $E[V_{ij}(x)] = W_i(x)$. We then define

$$W_{in}(x) = \frac{1}{n} \sum_{j=1}^{n} V_{ij}(x). \tag{3.3}$$

Since $V_{ij}(x)$, $i=1,2,\ldots$, are i.i.d. and bounded, it is easy to show that $W_{in}(x) \to W_i(x)$ with probability one for all $x=0,1,\ldots,N_i$. Now, let $\Delta_{in}(\underset{\sim}{x})$ and $d_n(\underset{\sim}{x})$ be defined as in (2.5) and (2.6), respectively. From the discussion of Section 2 and the construction of the sequence of decision rules $\{d_n\}_{n=1}^{\infty}$ through (2.5), (2.6), (3.1) and (3.3), we get the following result.

THEOREM 3.1. For our decision problem, the sequence of decision rules $\{d_n\}_{n=1}^{\infty}$ is asymptotically optimal relative to the prior distribution G.

Rate of Convergence of Empirical Bayes Rules $\{d_n\}$

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

Definition 3.1. The sequence of empirical Bayes rules $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal at least of order α_n relative to G if $r(G,d_n) - r(G) \leq O(\alpha_n)$ as $n \to \infty$ where $\lim_{n\to\infty} \alpha_n = 0$.

For each $i=1,\ldots,k$, define $S_i = \{\underset{\sim}{x}\in\chi|\Delta_{iG}(\underset{\sim}{x}) < 0\}$, $T_i = \{\underset{\sim}{x}\in\chi|\Delta_{iG}(\underset{\sim}{x}) > 0\}$. Let $\varepsilon_1 = \min\limits_{\substack{\underset{\sim}{x}\in S_i\\ 1\leq i\leq k}} (-\Delta_{iG}(\underset{\sim}{x}))$, $\varepsilon_2 = \min\limits_{\substack{\underset{\sim}{x}\in T_i\\ 1\leq i\leq k}} (\Delta_{iG}(\underset{\sim}{x}))$ and

$\varepsilon = \min(\varepsilon_1,\varepsilon_2)$. Since χ is a finite space, therefore $\varepsilon > 0$. Now, by the fact that $0 < f_j(x_j) \leq 1$ and $|\Delta_{iG}(\underset{\sim}{x})| \leq 1$, with straightforward calculations, one can obtain

$$
\begin{aligned}
&0 \leq r(G,d_n) - r(G) \\
(3.4)\qquad &\leq \sum_{i=1}^{k} \left\{ \sum_{\underset{\sim}{x}\in S_i} P\{\Delta_{in}(\underset{\sim}{x}) > 0\} + \sum_{\underset{\sim}{x}\in T_i} p\{\Delta_{in}(\underset{\sim}{x}) \leq 0\} \right\}.
\end{aligned}
$$

From (3.4), it suffices to consider the behavior of $P\{\Delta_{in}(\underset{\sim}{x}) > 0\}$ when $\underset{\sim}{x}\in S_i$ and that of $P\{\Delta_{in}(\underset{\sim}{x}) \leq 0\}$ when $\underset{\sim}{x}\in T_i$ as $n \to \infty$ for each $i=1,2,\ldots,k$.

Note that for each $\underset{\sim}{x} \in S_i$,

$$
\begin{aligned}
P\{\Delta_{in}(\underset{\sim}{x}) > 0\} &= P\{\Delta_{in}(\underset{\sim}{x}) - \Delta_{iG}(\underset{\sim}{x}) > -\Delta_{iG}(\underset{\sim}{x})\} \\
&\leq P\{\Delta_{in}(\underset{\sim}{x}) - \Delta_{iG}(\underset{\sim}{x}) > \varepsilon\}.
\end{aligned}
$$

Then, by (2.3), (2.5) and the fact that $0 \leq W_i(x_i), f_i(x_i), W_{in}(x_i), f_{in}(x_i) \leq 1$ and $p_0 \in (0,1)$, one can obtain the following inequalities:

$$
(3.5)\qquad P\{\Delta_{in}(\underset{\sim}{x}) > 0\} \leq P\{f_{in}(x_i) - f_i(x_i) > \tfrac{\varepsilon}{2}\} + P\{W_{in}(x_i) - W_i(x_i) < -\tfrac{\varepsilon}{2}\}
$$

when p_0 is known; and

$$P\{\Delta_{in}(\underset{\sim}{x}) > 0\} \leq P\{W_{0n}(x_0) - W_0(x_0) > \frac{\varepsilon}{4}\} + P\{f_{in}(x_i) - f_i(x_i) > \frac{\varepsilon}{4}\}$$

$$(3.6) \qquad + P\{W_{in}(x_i) - W_i(x_i) < -\frac{\varepsilon}{4}\} + P\{f_{0n}(x_0) - f_0(x_0) < -\frac{\varepsilon}{4}\}$$

when p_0 is unknown.

(3.5) and (3.6) show that it suffices to consider the behavior of $P\{|f_{in}(x_i) - f_i(x_i)| > \delta\}$ and $P\{|W_{in}(x_i) - W_i(x_i)| > \delta\}$ for some $\delta > 0$.

From (3.2) and (3.3), $W_{in}(x) - W_i(x) = \sum_{j=1}^{n} A_{ij}(x)/n$ where $A_{ij}(x) = Y_{ij} I_{\{x\}}(X_{ij}) - W_i(x)$. It is easy to see that $A_{ij}(x)$, $j=1,\ldots,n$, are i.i.d. with mean 0 and finite variance, say $\beta_i(x)$, since $|A_{ij}(x)| \leq 1$. Therefore, for $m \geq 2$,

$$E[A_{ij}^m(x)] \leq E[|A_{ij}(x)|^m] \leq E[|A_{ij}(x)|^2] = \beta_i(x) \leq \frac{1}{2}\,\beta_i(x)m!\,.$$

Let $B_n(x) = n\beta_i(x)$. Thus, by Bernstein's inequality (see Ibragimov and Linnik (1971), page 169), for any $\delta > 0$,

$$P\{|W_{in}(x) - W_i(x)| > \delta\}$$

$$(3.7) \quad \leq P\{|\sum_{j=1}^{n} A_{ij}(x)| > 2B_n^{1/2}(x)\min(\frac{1}{2}\,n^{1/2}\delta\beta_i^{-1/2}(x),\ \frac{1}{2}\,n^{1/2}\beta_i^{1/2}(x))\}$$

$$\leq 2\exp\{-\frac{n}{4}\min(\delta^2\beta_i^{-1}(x),\ \beta_i(x))\}.$$

Similarly, from (3.1),

$$f_{in}(x) - f_i(x) = \sum_{j=1}^{n} C_{ij}(x)/n \text{ where } C_{ij}(x) = I_{\{x\}}(X_{ij}) - f_i(x).$$

Also, $C_{ij}(x)$, $j=1,\ldots,n$, are i.i.d. with mean 0 and $|C_{ij}(x)| \leq 1$ and hence with finite variance, say $\alpha_i(x)$. Applying Bernstein's inequality again, we obtain

$$(3.8) \qquad P\{|f_{in}(x) - f_i(x)| > \delta\} \leq 2\exp\{-\frac{n}{4}\min(\delta^2\alpha_i^{-1}(x),\ \alpha_i(x))\}.$$

Thus, we take $\delta = \frac{\varepsilon}{4}$ if p_0 is unknown or take $\delta = \frac{\varepsilon}{2}$ if p_0 is known. Then, from (3.5) through (3.8), for each $\underset{\sim}{x} \in S_i$,

$$P\{\Delta_{in}(\underset{\sim}{x}) > 0\} \le O(\exp\{-\tfrac{n}{4}\min(\delta^2\alpha_i^{-1}(x_i),\ \alpha_i(x_i))\}) + O(\exp\{-\tfrac{n}{4}\min(\delta^2\beta_i^{-1}(x_i),\ \beta_i(x_i))\}). \tag{3.9}$$

Following an argument analogous to the above, we also get the conclusion given below:

For each $\underset{\sim}{x} \in T_i$, $i=1,\ldots,k$,

$$P\{\Delta_{in}(\underset{\sim}{x}) < 0\} \le O(\exp\{-\tfrac{n}{4}\min(\delta^2\alpha_i^{-1}(x_i),\ \alpha_i(x_i))\}) + O(\exp\{-\tfrac{n}{4}\min(\delta^2\beta_i^{-1}(x_i),\ \beta_i(x_i))\}). \tag{3.10}$$

Now, let $c_1 = \frac{1}{4}\min(b_1,b_2)$ where $b_1 = \min[\min\limits_{m\le i\le k}\ \min\limits_{0\le x\le N_i}(\delta^2\alpha_i^{-1}(x),\ \alpha_i(x))]$, $b_2 = \min[\min\limits_{m\le i\le k}\ \min\limits_{0\le x\le N_i}(\delta^2\beta_i^{-1}(x),\ \beta_i(x))]$, here $m = 1$ if p_0 is known and $m = 0$ if p_0 is unknown. It is clear that $c_1 > 0$ since $\beta_i(x) > 0$, $\alpha_i(x) > 0$ and χ is finite. Thus, we have the following theorem:

THEOREM 3.2. Let $\{d_n\}_{n=1}^{\infty}$ be the sequence of asymptotically optimal rules described in Theorem 3.1. Then $r(G,d_n) - r(G) \le O(\exp\{-c_1 n\})$ for some $c_1 > 0$.

An Alternative Empirical Bayes Rule. With the same framework as above, define

$$T_{ij} = X_{ij} + Y_{ij}. \tag{3.11}$$

Then, $T_{ij}|p_{ij} \sim B(N_i + 1,\ p_{ij})$. With $f_i(x|p) = \binom{N_i}{x}p^x(1-p)^{N_i-x}$, writing from (2.3),

$$f_i(x) = \int_0^1 f_i(x|p)dG_i(p) = f_i(x,N_i).$$

Then, from (2.3), following Robbins (1956), we see that

$$W_i(x) = \frac{x+1}{N_i+1} f_i(x+1, N_i+1).$$

Hence, let

$$W_{in}^0(x) = \frac{x+1}{N_i+1} \sum_{j=1}^{n} I_{\{x+1\}}(T_{ij}), \tag{3.12}$$

and define

$$\Delta_{in}^0(\underset{\sim}{x}) = \begin{cases} p_0 f_{in}(x_i) - W_{in}^0(x_i) & \text{if } p_0 \text{ is known,} \\ W_{0n}^0(x_0)f_{in}(x_i) - W_{in}^0(x_i)f_{0n}(x_0) & \text{if } p_0 \text{ is unknown;} \end{cases} \tag{3.13}$$

and

$$d_n^0(\underset{\sim}{x}) = \{i|\Delta_{in}^0(\underset{\sim}{x}) < 0\}. \tag{3.14}$$

Note that $W_{in}^0(x)$ is also an unbiased consistent estimator of $W_i(x)$. Therefore, following an argument analogous to that of (3.7), we can conclude that $r(G,d_n^0) - r(G) < 0(\exp(-c_2 n))$ for some $c_2 > 0$.

4. Case when $G_i(x)$ are Symmetrical about $p = 1/2$.

In this section, we suppose that there is sufficient information to tell us that $G_i(x)$ are symmetrical about $p = 1/2$ for all $i = (0), 1,\ldots,k$. Further, we also assume that N_i are even integers for all $i = (0), 1,\ldots,k$.

Estimation of $W_i(x)$ and $f_i(x)$. Under the above assumptions $f_i(x) = f_i(N_i - x)$ for all $x=0,1,\ldots,N_i$. Therefore, it is reasonable to use

$$(4.1) \quad f^1_{in}(x) \equiv f^1_{in}(N_i-x) = \begin{cases} \frac{1}{2n} \sum_{j=1}^{n} I_{\{x,N_i-x\}}(X_{ij}) & \text{for } x \neq \frac{N_i}{2}, \\ \frac{1}{n} \sum_{j=1}^{n} I_{\{x\}}(X_{ij}) & \text{for } x = \frac{N_i}{2} \end{cases}$$

to estimate $f_i(x)$.

For $W_i(x)$, $x=0,1,\ldots,N_i$ we will construct a sequence of consistent estimators $\{W^1_{in}(x)\}$, in terms of $f^1_{in}(y)$, $y=0,1,\ldots,N_i$, by using the observations $(X_{ij}, j=1,\ldots,n)$ only. The following lemma is very helpful for the above purpose.

Lemma 4.1. Suppose that the prior distribution $G_i(.)$ is symmetric about $p = 1/2$. Then

(a) $W_i(x) = \frac{x+1}{N_i-x} W_i(N_i-x-1)$ for each $x=0,1,\ldots,N_i-1$.

(b) $W_i(x) + W_i(N_i-x) = f_i(x) = f_i(N_i-x)$ for each $x=0,1,\ldots,N_i$.

(c) Furthermore, if N_i is an even integer, then, $W_i(\frac{N_i}{2}) = \frac{1}{2} f_i(\frac{N_i}{2})$.

Proof. Direct computation.

THEOREM 4.1. Suppose that $G_i(.)$ is symmetric about $p = 1/2$ and N_i is an even integer. Then, for each $x=0,1,\ldots,N_i$, $W_i(x)$ can be represented as a linear function of $f_i(y)$, $y=0,1,\ldots,N_i$.

Proof. It follows from Lemma 4.1 that for each $x=0,1,\ldots,N_i-1$ and $z = x - \frac{N_i}{2} + 1$,

$$(4.2) \quad W_i(\frac{N_i}{2} - z) = \frac{N_i+2-2z}{N_i+2z} f_i\ (\frac{N_i}{2} - z + 1) - \frac{N_i+2-2z}{N_i+2z} W_i\ (\frac{N_i}{2} - z + 1).$$

Then, by (4.2), Lemma 4.1 (b), (c) and by induction, the result follows.

By Theorem 4.1, for each $x=0,1,\ldots,N_i$,

$$W_i(x) = \sum_{y=0}^{N_i} \beta(N_i,x,y)\, f_i(y), \tag{4.3}$$

where the coefficients $\beta(N_i,x,y)$ depend on N_i, x and y. Also, the values of $\beta(N_i,x,y)$ can be obtained from Lemma 4.1 (c) and the iterative relation (4.2)

We then define

$$W^1_{in}(x) = \sum_{y=0}^{N_i} \beta(N_i,x,y)\, f^1_{in}(y) \tag{4.4}$$

where $f^1_{in}(y)$ is defined in (4.1).

Now, define

$$\Delta^1_{in}(\underset{\sim}{x}) = \begin{cases} W^1_{0n}(x_0)f^1_{in}(x_i) - W^1_{in}(x_i)f^1_{0n}(x_0) & \text{if } p_0 \text{ is unknown,} \\ p_0 f^1_{in}(x_i) - W^1_{in}(x_i) & \text{if } p_0 \text{ is known,} \end{cases} \tag{4.5}$$

and

$$d^1_n(\underset{\sim}{x}) = \{i \mid \Delta^1_{in}(\underset{\sim}{x}) < 0\}. \tag{4.6}$$

From (4.1), it is clear that $f^1_{in}(x) \to f_i(x)$ with probability 1 as $n \to \infty$ for each $x=0,1,\ldots,N_i$. Therefore, from (4.3) and (4.4), $W^1_{in}(x) \to W_i(x)$ with probability 1 as $n \to \infty$ for each $x=0,1,\ldots,N_i$. Thus we have the following theorem:

THEOREM 4.2. Suppose that the prior distributions $G_i(\cdot)$ are symmetrical about $p = 1/2$ and N_i are even integers for all $i = (0), 1,\ldots,k$. Then, the sequence of decision rules $\{d^1_n\}^{\infty}_{n=1}$ is asymptotically optimal relative to the prior distribution G.

Rate of Convergence of Empirical Bayes Rules $\{d_n^1\}$. We now consider the rate of convergence of the empirical Bayes rules $\{d_n^1\}$. Following the same discussion as given in (3.4) through (3.6), and using the fact that $f_{in}^1(x) \to f_i(x)$ with probability 1, it suffices to consider the behavior of $P\{W_{in}^1(x) - W_i(x) > \delta\}$ and $P\{W_{in}^1(x) - W_i(x) < -\delta\}$ as $n \to \infty$ for some $\delta > 0$, for each $x=0,1,\ldots,N_i$, $i = (0), 1,\ldots,k$.

From (4.3) and (4.4), for each $x=0,1,\ldots,N_i$,

$$\begin{aligned} P\{W_{in}^1(x) - W_i(x) > \delta\} &= P\left\{\sum_{y=0}^{N_i} \beta(N_i,x,y)\ [f_{in}^1(y)-f_i(y)] > \delta\right\} \\ &\leq \sum_{y=0}^{N_i} P\{\beta(N_i,x,y)\ [f_{in}^1(y) - f(y)] > \delta_1\} \end{aligned}$$

where $\delta_1 = \frac{\delta}{N_i+1}$. If $\beta(N_i,x,y) = 0$ for some $0 \leq y \leq N_i$, then

$P\{\beta(N_i,x,y)[f_{in}^1(y) - f_i(y)] > \delta_1\} = 0$. So, we assume $\beta(N_i,x,y) \neq 0$. When $\beta(N_i,x,y) > 0$, then

$$P\{\beta(N_i,x,y)[f_{in}^1(y)-f_i(y)] > \delta_1\} = P\{f_{in}^1(y) - f_i(y) > \delta_1/\beta(N_i,x,y)\}.$$

When $\beta(N_i,x,y) < 0$, then

$$P\{\beta(N_i,x,y)[f_{in}^1(y) - f_i(y)] > \delta_1\} = P\{f_{in}^1(y) - f_i(y) < \delta_1/\beta(N_i,x,y)\}.$$

In either case, the problem can be reduced to considering the convergence rate of $P\{|f_{in}^1(y) - f_i(y)| > \delta_2\}$ as $n \to \infty$ for some $\delta_2 > 0$. Similarly, for the convergence rate of $P\{W_{in}^1(x) - W_i(x) < -\delta\}$ where $x=0,1,\ldots,N_i$ and $\delta > 0$, we get a similar result. Therefore, by applying Bernstein's inequality and following an argument similar to that of (3.7), we conclude the following theorem:

THEOREM 4.3. Let $\{d_n^1\}_{n=1}^{\infty}$ be the sequence of decision rules defined in (4.6). Then, $\{d_n^1\}_{n=1}^{\infty}$ is asymptotically optimal at least of order $\exp\{-c_3 n\}$ relative to the prior distribution G for some $c_3 > 0$.

5. Smoothed Empirical Bayes Rules.

In this section, we again assume that $G_i(.)$ are symmetrical about $p = 1/2$ and N_i are even integers for all $i = (0), 1, \dots, k$. In Section 4, the marginal frequency functions $f_i(x)$, $x=0,1,\dots,N_i$, $i = (0), 1,\dots,k$, are estimated in terms of the empirical frequency functions $f_{in}^1(x)$, regardless of the properties associated with the marginal function $f_i(x)$. In this section, by considering some properties related to $f_i(x)$ and $W_i(x)$, two methods for obtaining smooth estimators of $f_i(x)$ and $W_i(x)$ are studied.

We first state the following lemma (without proof), which can be verified by direct computations.

Lemma 5.1. Suppose that $G_i(.)$ is symmetrical about $p = 1/2$ and N_i is an even integer. Then,

(a) $\quad f_i(x)\binom{N_i}{x}^{-1} \le f_i(y)\binom{N_i}{y}^{-1}$ for $0 \le y \le x \le N_i/2$.

(b) $\quad W_i(x)\binom{N_i}{x}^{-1} \le W_i(y)\binom{N_i}{y}^{-1}$ for $0 \le y \le x \le N_i/2$ and $N_i/2 \le x \le y \le N_i$.

(c) $\quad W_i(y) \le W_i(N_i-y)$ for $0 \le y \le N_i/2$.

<u>Procedure 1. Smoothing Based on $f_{in}^1(x)$.</u> For each $0 \le y \le N_i/2$, let

$$\tilde{f}_{in}(y) = \binom{N_i}{y} \max_{y \le x \le N_i/2} \min_{0 \le z \le x} \left\{ \sum_{a=z}^{x} f_{in}^1(a)\binom{N_i}{a}^{-1}/(x-z+1)\right\}, \tag{5.1}$$

and let $\tilde{f}_{in}(N_i-y) = \tilde{f}_{in}(y)$. Then, let

$$\tilde{W}_{in}(y) = \sum_{z=0}^{N_i} \beta(N_i,y,z)\tilde{f}_{in}(z) \quad \text{for } y=0,1,\dots,N_i. \tag{5.2}$$

Define

$$(5.3)\quad \tilde{\Delta}_{in}(\underset{\sim}{x}) = \begin{cases} p_0\tilde{f}_{in}(x_i)-\tilde{W}_{in}(x_i) & \text{if } p_0 \text{ is known,} \\ \tilde{W}_{0n}(x_0)\tilde{f}_{in}(x_i)-\tilde{W}_{in}(x_i)\tilde{f}_{0n}(x_0) & \text{if } p_0 \text{ is unknown.} \end{cases}$$

Finally, define the selection rule $\tilde{d}_n$ as follows:

$$(5.4)\qquad \tilde{d}_n(\underset{\sim}{x}) = \{i|\tilde{\Delta}_{in}(\underset{\sim}{x}) < 0\}.$$

<u>Asymptotic Optimality of $\{\tilde{d}_n\}$.</u> Note that $\tilde{f}_{in}(y)\binom{N_i}{y}^{-1}$, $y=0,1,\ldots,N_i$ are the isotonic estimators of $f_i(y)\binom{N_i}{y}^{-1}$, based on $f^1_{in}(x)\binom{N_i}{x}^{-1}$, $x=0,1,\ldots,N_i$, with equal weights. Since $f^1_{in}(x)$ is a strongly consistent estimator of $f_i(x)$ for all $x=0,1,\ldots,N_i$, then, by Theorem 2.2 of Barlow et al (1972), Lemma 4.1(b), (4.3) and the definition of $\tilde{W}_{in}(y)$, it not hard to see that $\tilde{f}_{in}(y)$ and $\tilde{W}_{in}(y)$ are strongly consistent estimators of $f_i(y)$ and $W_i(y)$, respectively.

Next, we consider the rate of convergence of the difference $r(G,\tilde{d}_n) - r(G)$. For each $0 \le y \le N_i$ and $\delta > 0$, by Theorem 2.1 of Barlow, et al (1972), we can obtain the following inequality.

$$(5.5)\quad P\{|\tilde{f}_{in}(y) - f_i(y)| > \delta\} \le \sum_{x=0}^{N_i} P\{|f^1_{in}(x) - f_i(x)| > \binom{N_i}{x}\binom{N_i}{y}^{-1}\delta(N_i+1)^{-1/2}\}.$$

Then, with a discussion similar to that given in Section 4, we can conclude that $r(G,\tilde{d}_n) - r(G) \le O(\exp\{-c_4 n\})$ for some $c_4 > 0$.

It is easy to see that the new estimators $\tilde{f}_{in}(y)$, $0 \le y \le N_i$, always satisfy the constraint of Lemma 5.1(a). However, one would also like to see whether the estimators $\tilde{W}_{in}(y)$, $0 \le y \le N_i$, satisfy the corresponding constraint or not. The following lemma is useful for this purpose.

LEMMA 5.2 Let $U(x)$, $h(x)$ be nonnegative functions defined on $\{0,1,\ldots,N\}$, where N is an even positive integer, which satisfy

(a) $\quad U(x) = \frac{x+1}{N-x}\, U(N-x-1) \quad$ for all $x=0,1,\ldots,N-1$.

(b) $U(x) + U(N-x) = h(x) = h(N-x)$ for all $x=0,1,\ldots,N$ and

(c) $U(x) \le U(N-x)$ for all $x=0,1,\ldots,N/2$.

Then

(d) $(x+1)h(x+1) \le (N-x)h(x)$ for all $x=0,1,\ldots,N/2-1$.

We note that (a), (b) and (d) of Lemma 5.2 do not imply (c), and the estimators $\tilde{W}_{in}(y)$, $0 \le y \le n_i$, do not always satisfy the required constraint. Lemma 5.2 suggests resmoothing based on $\tilde{W}_{in}(y)$.

Procedure 2. Resmoothing Based on $\tilde{W}_{in}(y)$. First, let $Q_{in}(N_i) = \tilde{W}_{in}(N_i)$ and for each $N_i/2 \le y \le N_i-1$, let

(5.6) $$Q_{in}(y) = [\tilde{W}_{in}(y)\binom{N_i}{y}^{-1} + \tilde{W}_{in}(N_i-y-1)\binom{N_i}{N_i-y-1}^{-1}]/2.$$

Step 1. For each $N_i/2 \le y \le N_i$, let

(5.7) $$Q^*_{in}(y) = \max_{N_i/2\le x\le y} \min_{x\le z\le N_i} \{ \sum_{a=x}^{z} Q_{in}(a)/(z-x+1)\}.$$

Step 2. Let $W^*_{in}(N_i) = Q^*_{in}(N_i)$ and for each $N_i/2 \le y \le N_i-1$, let

$$W^*_{in}(y) = Q^*_{in}(y)\binom{N_i}{y} \text{ and } W^*_{in}(N_i-y-1) = Q^*_{in}(y)\binom{N_i}{N_i-y-1}.$$

Then, let

(5.8) $$f^*_{in}(y) = W^*_{in}(y) + W^*_{in}(N_i-y) \quad \text{for } y=0,1,\ldots,N_i$$

and define

(5.9) $$\Delta^*_{in}(\underset{\sim}{x}) = \begin{cases} p_0 f^*_{in}(x_i)-W^*_{in}(x_i) & \text{if } p_0 \text{ is known,} \\ W^*_{0n}(x_0)f^*_{in}(x_i)-W^*_{in}(x_i)f^*_{0n}(x_0) & \text{if } p_0 \text{ is unknown.} \end{cases}$$

Finally, define the selection rule d_n^* as follows:

(5.10) $$d_n^*(\underset{\sim}{x}) = \{i \mid \Delta_{in}^*(\underset{\sim}{x}) < 0\}.$$

Remark. By Step 1 and Step 2 of Procedure 2, the estimators $W_{in}^*(y)$, $0 < y < N_i$, always satisfy the constraint of Lemma 5.1(b) and (c). Then, by Lemma 5.2, the estimators $f_{in}^*(y)$, $0 < y < N_i$, also satisfy the corresponding constraint.

Asymptotic Optimality of $\{d_n^*\}$. By Theorem 2.2 of Barlow et al. (1972) and the fact that $\tilde{W}_{in}(y)$, $0 < y < N_i$, are strongly consistent estimators of $W_i(y)$, $0 < y < N_i$, we conclude that $W_{in}^*(y)$, $0 < y < N_i$, are strongly consisent estimators of $W_i(y)$, $0 < y < N_i$. Then, by Lemma 4.1(b) and (5.8), $f_{in}^*(y)$, $0 < y < N_i$, are also consistent estimators of $f_i(y)$, $0 < y < N_i$. Therefore, the sequence of empirical Bayes selection rules $\{d_n^*\}$ is asymptotically optimal.

By Theorem 2.1 of Barlow, et al. (1972) and (5.8) we obtain, for $\delta > 0$,

$$P\{|f_{in}^*(y) - f_i(y)| > \delta\}$$

$$< p\{|W_{in}^*(y)-W_i(y)| > \delta/2\} + P\{|W_{in}^*(N_i-y)-W_i(N_i-y)| > \delta/2\}$$

$$< P\{\sum_{x=0}^{N_i} |W_{in}^*(x)\binom{N_i}{x}^{-1}-W_i(x)\binom{N_i}{x}^{-1}|^2 > \binom{N_i}{y}^{-2}\delta^2/4\}$$

(5.11) $$+ P\{\sum_{x=0}^{N_i} |W_{in}^*(x)\binom{N_i}{x}^{-1} - W_i(x)\binom{N_i}{x}^{-1}|^2 > \binom{N_i}{N_i-y}^{-2}\delta^2/4\}$$

$$< 2\, P\{\sum_{x=0}^{N_i} |\tilde{W}_{in}(x)\binom{N_i}{x}^{-1}-W_i(x)\binom{N_i}{x}^{-1}|^2 > \binom{N_i}{y}^{-2}\delta^2/4\}$$

$$< 2 \sum_{x=0}^{N_i} P\{|\tilde{W}_{in}(x)-W_i(x)| > \binom{N_i}{x}\binom{N_i}{y}^{-1}\delta(N_i+1)^{-1/2}/2\}.$$

Then by (4.3), (5.2) and (5.5), with a discussion similar to that given in Theorem 4.3, we conclude that $r(G,d_n^*)-r(G) \leq O(\exp\{-c_5 n\})$ for some $c_5 > 0$.

Acknowledgment.

The authors wish to thank a referee for very helpful comments and for pointing out the alternative empirical Bayes rule described in Section 3.

REFERENCES

Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). Statistical Inference under Order Restrictions. Wiley, New York.

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

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

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

Huang, W.T. (1975). Bayes approach to a problem of partitioning k normal populations. Bull. Inst. Math. Acad. Sinica **3** 87-97.

Ibragimov, I.A. and Linnik, Yu.V. (1971). Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen.

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

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

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

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

Singh, A.K. (1977). On slippage tests and multiple decision (selection and ranking) procedures. Ph.D. Thesis (Mimeo. Ser. No. 494), Department of Statistics, Purdue University, West Lafayette, Indiana.

Van Ryzin, J. (1970). On some nonparametric empirical Bayes multiple decision problems. Nonparametric Techniques in Statistical Inference, (ed. M.L. Puri), 585-603, Cambridge University Press.

Van Ryzin, J. and Susarla, V. (1977). On the empirical Bayes approach to multiple decision problems. Ann. Statist. 5 172-181.