Parametric Empirical Bayes Rules for Selecting the Most Probable Multinomial Event*

by

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

Technical Report #86-54

Department of Statistics
Purdue University

December 1986
Revised July 1987
Revised August 1988

* This research was partially supported by the Office of Naval Research Contract N00014-84-C-0167 and NSF Grant DMS-8606964 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Parametric Empirical Bayes Rules for Selecting
the Most Probable Multinomial Event*

by

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

Abstract

Consider a multinomial population with $k(\geq 2)$ cells and the associated probability vector $\underset{\sim}{p} = (p_1, \ldots, p_k)$. Let $p_{[k]} = \max_{1 \leq i \leq k} p_i$. A cell associated with $p_{[k]}$ is called the most probable event. We are interested in selecting the most probable event. Let i denote the index of the selected cell. Under the loss function $L(\underset{\sim}{p}, i) = p_{[k]} - p_i$, this statistical selection problem is studied via a parametric empirical Bayes approach. Two empirical Bayes selection rules are proposed. They are shown to be asymptotically optimal at least of order $0(\exp(-c_i n))$ for some positive constants c_i, $i = 1, 2$, where n is the number of accumulated past experiences (observations) at hand. Finally, for the problem of selecting the least probable event associated with $p_{[1]}$ under the loss $p_i - p_{[1]}$, two empirical Bayes selection rules are also proposed. The corresponding rates of convergence are found to be at least of order $0(\exp(-c_i n))$ for some positive constants c_i, $i = 3, 4$.

* This research was partially supported by the Office of Naval Research Contract N00014-84-C-0167 and NSF Grant DMS-8606964 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: Asymptotically optimal, Bayes rules, empirical Bayes rules, parametric empirical Bayes, Dirichlet prior, the most probable event, the least probable event.

1. Introduction

Consider a multinomial population with $k \geq 2$ cells and the associated probability vector $\underset{\sim}{p} = (p_1, \ldots, p_k)$ where $\sum_{i=1}^{k} p_i = 1$. Let $p_{[1]} \leq \ldots \leq p_{[k]}$ denote the ordered values of the parameters $p_1, \ldots, p_k$. It is assumed that the exact pairing between the ordered and the unordered parameters is unknown. Any event associated with $p_{[k]}$ is considered as the most probable event. A number of statistical procedures based on single samples or sequential sampling rules have been considered in the literature in the classical framework for selecting the most probable event. Bechhofer, Elmaghraby and Morse (1959) have considered a fixed sample procedure through the indifference zone approach. Gupta and Nagel (1967), Panchapakesan (1971) and, Gupta and Huang (1975) have studied this selection problem using a subset selection approach. Cacoullos and Sobel (1966), Alam (1971), Alam, Seo and Thompson (1971), Ramey and Alam (1979, 1980) and Bechhofer and Kulkarni (1984) have considered sequential selection procedures.

We now consider a situation in which one repeatedly deals with the same selection problem independently. In such instances, it is reasonable to formulate the component problem in the sequence as a Bayes decision problem with respect to an unknown (or partially known) prior distribution on the parameter space, and then, use the accumulated observations to improve the decision rule at each stage. This is the empirical Bayes approach due to Robbins (1956, 1964 and 1983).

Empirical Bayes rules have been derived for subset selection goals by Deely (1965). Recently, Gupta and Hsiao (1983) and Gupta and Leu (1988) have studied empirical Bayes rules for selecting good populations with respect to a standard or a control with the underlying populations being uniformly distributed. Gupta and Liang (1986, 1988) have studied empirical Bayes rules for the problem of selecting the best binomial population or selecting good binomial populations. Many such empirical Bayes procedures have been shown to be asymptotically optimal in the sense that the risk for the nth decision problem converges to the optimal Bayes risk which could have been obtained if the prior distribution was fully known and the Bayes procedure with respect to this prior distribution was used.

Note that the above mentioned empirical Bayes rules use the so-called nonparametric empirical Bayes approach. That is, one assumes that the form of the prior distribution is unknown. However, in many cases, an experimenter may have some prior information about the parameters of interest, and he would like to use this information to make appropriate decisions. Usually, it is suggested (for example, see Robbins (1964)), that the prior information be quantified through a class of subjectively plausible priors. In view of this situation, in this paper, it is assumed that the parameters of interest in a multinomial distribution follow some conjugate prior distribution with unknown hyperparameters. Under this statistical framework, two empirical Bayes selection rules are proposed. They are shown to be asymptotically optimal at least of order $0(\exp(-c_i n))$ for some positive constants c_i, $i = 1, 2$, where n is the number of accumulated past experiences (observations) at hand. Finally, for the problem of selecting the least probable event associated with $p_{[1]}$ under the loss $p_i - p_{[1]}$, two empirical Bayes selection rules are also proposed. The corresponding rates of convergence are found to be at least of order $0(\exp(-c_i n))$ for some positive constants c_i, $i = 3, 4$.

2. Formulation of the Problem under the Empirical Bayes Approach

Consider a multinomial population with $k(\geq 2)$ cells, where the cell π_i has probability p_i, $i = 1, \ldots, k$. Let X_i denote the observations that arise in the cell π_i based on $N(\geq 2)$ independent trials. Thus, for given $\underset{\sim}{p} = (p_1, \ldots, p_k), \underset{\sim}{X} = (X_1, \ldots, X_k)$ has the probability function

$$(2.1) \qquad f(\underset{\sim}{x}|\underset{\sim}{p}) = \frac{N!}{\prod_{i=1}^{k}(x_i!)} \prod_{i=1}^{k} p_i^{x_i},$$

where, $x_i = 0, 1, \ldots, N$ and $\sum_{i=1}^{k} x_i = N$.

For each $\underset{\sim}{p}$, let $p_{[1]} \leq \ldots \leq p_{[k]}$ denote the ordered parameters $p_1, \ldots, p_k$. It is assumed that there is no apriori knowledge about the exact pairing between the ordered and the unordered parameters. Any cell π_i associated with $p_{[k]}$ is considered as the most probable event. Our goal is to derive empirical Bayes rules to select the most probable event.

Let $\Omega = \{\underset{\sim}{p}|\underset{\sim}{p} = (p_1, \ldots, p_k),\ 0 < p_i < 1 \text{ and } \sum_{i=1}^{k} p_i = 1\}$ be the parameter space. It is assumed that $\underset{\sim}{p}$ has a Dirichlet prior distribution G with hyperparameters $\underset{\sim}{\alpha} = (\alpha_1, \ldots, \alpha_k)$, where all α_i are positive but unknown. That is, $\underset{\sim}{p}$ has a probability density function of the form

$$(2.2) \qquad g(\underset{\sim}{p}) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{k}\Gamma(\alpha_i)} \prod_{i=1}^{k} p_i^{\alpha_i - 1},\ 0 < p_i < 1,\ \sum_{i=1}^{k} p_i = 1,$$

where $\alpha_0 = \sum_{i=1}^{k} \alpha_i$.

Let $\mathcal{A} = \{i|i = 1, \ldots, k\}$ be the action space. When action i is taken, it means that the cell π_i is selected as the most probable event. For the parameter $\underset{\sim}{p}$ and action i, the loss function $L(\underset{\sim}{p}, i)$ is defined as

$$(2.3) \qquad L(\underset{\sim}{p}, i) = p_{[k]} - p_i,$$

the difference between the most probable and the selected event.

Let $\mathcal{X}$ be the sample space of $\underset{\sim}{X} = (X_1, \ldots, X_k)$. A selection rule $d = (d_1, \ldots, d_k)$ is a mapping from $\mathcal{X}$ into $[0,1]^k$ such that for each $\underset{\sim}{x} \epsilon \mathcal{X}$, the function $d(\underset{\sim}{x}) = (d_1(\underset{\sim}{x}), \ldots, d_k(\underset{\sim}{x}))$ is such that $0 \le d_i(\underset{\sim}{x}) \le 1,\ i = 1, \ldots, k$, and $\sum_{i=1}^{k} d_i(\underset{\sim}{x}) = 1$. Note that $d_i(\underset{\sim}{x}),\ i = 1, \ldots, k$ is the probability of selecting cell π_i as the most probable event given $\underset{\sim}{X} = \underset{\sim}{x}$.

Let D be the class of all selection rules as defined above. For each $d \epsilon D$, let $r(G, d)$ denote the associated Bayes risk. Then $r(G) = \inf_{d \epsilon D} r(G, d)$ is the minimum Bayes risk.

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

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

Consider the selection rule $d_G = (d_{1G}, \ldots, d_{kG})$ defined below: for each $i = 1, \ldots, k$,

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

where $|A|$ denotes the cardinality of the set A.

It should be noted that in (2.4) any selection rule $d = (d_1, \ldots, d_k)$ satisfying the condition $\sum_{i \in A(\underset{\sim}{x})} d_i(\underset{\sim}{x}) = 1$ is a Bayes selection rule.

A straightforward computation shows that the selection rule d_G is a randomized Bayes selection rule in the class D. Since the values of the hyperparameters $(\alpha_1, \ldots, \alpha_k)$ are unknown, it is impossible to apply this Bayes selection rule d_G for the selection problem at hand. As we mentioned above, we study this selection problem via empirical Bayes approach.

For each $j = 1, 2, \ldots$, let $\underset{\sim}{X}_j = (X_{1j}, \ldots, X_{kj})$ denote the random observations arising from N independent trials at stage j. Let $\underset{\sim}{P}_j = (P_{1j}, \ldots, P_{kj})$ denote the (random) parameters at stage j. Conditional on $\underset{\sim}{P}_j$, $\underset{\sim}{X}_j$ has a probability function of the form of (2.1). It is assumed that independent observations $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ are available, and $\underset{\sim}{P}_j$, $1 \leq j \leq n$, have the same prior probability density function of the form (2.2), though not observable. We also let $\underset{\sim}{X}_{n+1} = \underset{\sim}{X} = (X_1, \ldots, X_k)$ denote the present observations.

Two empirical Bayes selection rules are proposed depending on whether the value of the parameter α_0 is known or unknown. Note that α_0 is the sum of all the parameters α_i, $1 \leq i \leq k$. In the case that α_0 is known, the individual values of α_i, $1 \leq i \leq k$, are still unknown.

First, for each $i = 1, \ldots, k$, and each $n = 1, 2, \ldots$, we let

$$(2.6) \qquad \begin{cases} \bar{X}_i(n) = \frac{1}{n} \sum\limits_{j=1}^{n} X_{ij}, M_i(n) = \frac{1}{n} \sum\limits_{j=1}^{n} X_{ij}^2, Z_i(n) = [N\bar{X}_i(n) - M_i(n)]\bar{X}_i(n), \\ Y_i(n) = [M_i(n) - \bar{X}_i(n)]N - (N-1)(\bar{X}_i(n))^2. \end{cases}$$

When α_0 is known, let

$$(2.7) \qquad \hat{\alpha}_{in} = \alpha_0 \bar{X}_i(n) N^{-1},$$

and let

$$(2.8) \qquad A_n(\underset{\sim}{x}) = \{i | x_i + \hat{\alpha}_{in} = \max_{1 \leq j \leq k} (x_j + \hat{\alpha}_{jn})\}.$$

We then define an empirical Bayes selection rule $\tilde{d}_n = (\tilde{d}_{1n}, \ldots, \tilde{d}_{kn})$ as follows: for each $i = 1, \ldots, k$, $\underset{\sim}{x} \epsilon \mathcal{X}$,

$$\tilde{d}_{in}(\underset{\sim}{x}) = \begin{cases} |A_n(\underset{\sim}{x})|^{-1} & \text{if } i \epsilon A_n(\underset{\sim}{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$

Let $\mu_{i1} = E[\bar{X}_i(n)]$ and $\mu_{i2} = E[M_i(n)]$. Then, following a direct computation, we have $\mu_{i1} = N\alpha_i\alpha_0^{-1}$, $\mu_{i2} = N\alpha_i\alpha_0^{-1} + (N^2 - N)\alpha_i(\alpha_i + 1)\alpha_0^{-1}(\alpha_0 + 1)^{-1}$. Hence, $\alpha_i = L_{i1}L_{i2}^{-1}$, where $L_{i1} = (N\mu_{i1} - \mu_{i2})\mu_{i1}, L_{i2} = (\mu_{i2} - \mu_{i1})N - (N-1)\mu_{i1}^2$. Thus, $Z_i(n), Y_i(n)$, and $Z_i(n)/Y_i(n)$ are moment estimators of L_{i1}, L_{i2}, and $\alpha_i = L_{i1}L_{i2}^{-1}$, respectively. Note that L_{i1} and L_{i2} are both positive, which can be verified directly by the definition of μ_{i1} and μ_{i2}. Also, $Z_i(n) \geq 0$. However, it is possible that $Y_i(n) \leq 0$. Hence, for the case when α_0 is unknown, we first let

$$\Delta_{in}(x_i) = \begin{cases} x_i + Z_i(n)/Y_i(n) & \text{if } Y_i(n) > 0, \\ x_i & \text{otherwise.} \end{cases} \tag{2.10}$$

Also, let

$$A_n^*(\underset{\sim}{x}) = \{i | \Delta_{in}(x_i) = \max_{1 \leq j \leq k} \Delta_{jn}(x_j)\}. \tag{2.11}$$

We then define an empirical Bayes selection rule $d_n^* = (d_{1n}^*, \ldots, d_{kn}^*)$ as follows: for each $i = 1, \ldots, k$, $\underset{\sim}{x} \epsilon \mathcal{X}$,

$$d_{in}^*(\underset{\sim}{x}) = \begin{cases} |A_n^*(\underset{\sim}{x})|^{-1} & \text{if } i \epsilon A_n^*(\underset{\sim}{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

In the next section, we will study the optimality of the two sequences of empirical Bayes selection rules $\{\tilde{d}_n\}$ and $\{d_n^*\}$.

3. Asymptotic Optimality of Selection Rules $\{\tilde{d}_n\}$ and $\{d_n^*\}$

Consider an empirical Bayes selection rule $d_n(\underset{\sim}{x})$. Let $r(G, d_n)$ be the Bayes risk associated with the selection rule $d_n(\underset{\sim}{x})$. Then $r(G, d_n) - r(G) \geq 0$, since $r(G)$ is the minimum Bayes risk. The nonnegative difference is always used as a measure of optimality of the selection rule d_n.

Definition 3.1. A sequence of empirical Bayes rules $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal at least of order β_n relative to the prior distribution G if $r(G, d_n) - r(G) \leq 0(\beta_n)$ as $n \to \infty$, where $\{\beta_n\}$ is a sequence of positive values satisfying $\lim_{n\to\infty} \beta_n = 0$.

3.1. Asymptotic Optimality of $\{\tilde{d}_n\}$.

We first consider the case where α_0 is known. Note that $\hat{\alpha}_{in}$ is an unbiased estimator of α_i; also $\sum_{i=1}^{k} \hat{\alpha}_{in} = \alpha_0$ for each $n = 1, 2, \ldots$.

For each $\underset{\sim}{x} \epsilon \mathcal{X}$, let $A(\underset{\sim}{x})$ be as defined in (2.4) and let $B(\underset{\sim}{x}) = \{1, 2, \ldots, k\} \backslash A(\underset{\sim}{x})$. Thus, for each $\underset{\sim}{x} \epsilon \mathcal{X}$, $i \epsilon A(\underset{\sim}{x})$, $j \epsilon B(\underset{\sim}{x})$, $x_i + \alpha_i > x_j + \alpha_j$. Following straightforward computation, we can show

$$
\begin{aligned}
&0 \leq r(G, \tilde{d}_n) - r(G) \\
&\leq \sum_{\underset{\sim}{x} \epsilon \mathcal{X}} \sum_{i \epsilon A(\underset{\sim}{x})} \sum_{j \epsilon B(\underset{\sim}{x})} P\{x_i + \hat{\alpha}_{in} \leq x_j + \hat{\alpha}_{jn}\}.
\end{aligned}
\tag{3.1}
$$

Now, for $i \epsilon A(\underset{\sim}{x})$, $j \epsilon B(\underset{\sim}{x})$,

$$
\begin{aligned}
&P\{x_i + \hat{\alpha}_{in} \leq x_j + \hat{\alpha}_{jn}\} \\
&\quad = P\{\frac{1}{n} \sum_{m=1}^{n} [\frac{1}{N}(X_{im} - X_{jm}) - \frac{1}{\alpha_0}(\alpha_i - \alpha_j)] < -(x_i + \alpha_i - x_j - \alpha_j)\alpha_0^{-1}\}
\end{aligned}
$$

$$
\begin{aligned}
(3.2) \qquad & \leq P\{\frac{1}{n}\sum_{m=1}^{n}[\frac{1}{N}(X_{im}-X_{jm})-\frac{1}{\alpha_0}(\alpha_i-\alpha_j)] < -\varepsilon_{ij}\} \\
& \leq \exp\{-n2^{-1}\varepsilon_{ij}^2\} \\
& \leq \exp\{-nc_1\},
\end{aligned}
$$

where

$$
\begin{aligned}
\varepsilon_{ij} = \min_{x_i,x_j}\{&|x_i+\alpha_i-x_j-\alpha_j|\alpha_0^{-1}|, \quad x_i,x_j=0,1,\ldots,N, 0\leq x_i+x_j\leq N, \\
(3.3) \qquad & x_i+\alpha_i-x_j-\alpha_j\neq 0\} \\
& > 0 \quad \text{since } N \text{ is a finite number .}
\end{aligned}
$$

and

$$
(3.4) \qquad c_1 = 2^{-1}\min\{\varepsilon_{ij}^2|i,j=1,\ldots,k,\ i\neq j\} > 0.
$$

In (3.2), the second inequality is obtained using the fact that

$$
\begin{aligned}
& E[\frac{1}{N}(X_{im}-X_{jm})-\frac{1}{\alpha_0}(\alpha_i-\alpha_j)] = 0, \\
& -1-\frac{1}{\alpha_0}(\alpha_i-\alpha_j) \leq \frac{1}{N}(X_{im}-X_{jm})-\frac{1}{\alpha_0}(\alpha_i-\alpha_j) \leq 1-\frac{1}{\alpha_0}(\alpha_i-\alpha_j)
\end{aligned}
$$

and then making use of Theorem 2 of Hoeffding (1963).

By noting that $\mathcal{X}$ is a finite space, from (3.1) and (3.2), we have the following theorem.

Theorem 3.1. Let $\{\tilde{d}_n\}$ be the sequence of empirical Bayes selection rules defined in (2.9). Then $r(G,\tilde{d}_n) - r(G) \leq 0(\exp(-c_1 n))$ for some positive constant c_1.

3.2. Asymptotic Optimality of $\{d_n^*\}$.

For each $\underset{\sim}{x}\epsilon\mathcal{X}$, let $A(\underset{\sim}{x})$ and $B(\underset{\sim}{x})$ be as defined in the previous sections. For the selection rule d_n^*, one can obtain the following result

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

Since $\mathcal{X}$ is finite, we only need to consider the behavior of $P\{\Delta_{in}(x_i) \leq \Delta_{jn}(x_j)\}$ for each $\underset{\sim}{x}\epsilon\mathcal{X}$. Now

$$
\begin{aligned}
&P\{\Delta_{in}(x_i) \leq \Delta_{jn}(x_j)\} \\
(3.6) = &P\{\Delta_{in}(x_i) \leq \Delta_{jn}(x_j) \text{ and } (Z_i(n) \leq 0 \text{ or } Z_j(n) \leq 0 \text{ or } Y_i(n) \leq 0 \text{ or } Y_j(n) \leq 0)\} \\
&+ P\{\Delta_{in}(x_i) \leq \Delta_{jn}(x_j) \text{ and } (Z_i(n) > 0, Z_j(n) > 0, Y_i(n) > 0 \text{ and } Y_j(n) > 0)\}.
\end{aligned}
$$

Before we go further to study the associated asymptotic behaviors of the above probabilities appearing on the right hand side of (3.6), we need the following lemma.

Lemma 3.1. Let $b > 0$ be a constant. Then,

a) $P\{Z_i(n) - L_{i1} < -b\} \leq 0(\exp(-b_i n))$; b) $P\{Z_i(n) - L_{i1} > b\} \leq 0(\exp(-b_i n))$;

c) $P\{Y_i(n) - L_{i2} < -b\} \leq 0(\exp(-b_i n))$; d) $P\{Y_i(n) - L_{i2} > b\} \leq 0(\exp(-b_i n))$;

where $b_i = b^2[2N^4(N + \mu_{i1})^2]^{-1} > 0$.

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

Note that $Z_i(n) = [N\bar{X}_i(n) - M_i(n)]\bar{X}_i(n) \geq 0$. Hence, $P\{Z_i(n) - L_{i1} < -b\} = 0$ if $L_{i1} - b \leq 0$. So, we assume that $b > 0$ is small enough so that $L_{i1} - b > 0$. Then,

$$
\begin{aligned}
&P\{Z_i(n) - L_{i1} < -b\} \\
&= P\{N[(\bar{X}_i(n))^2 - \mu_{i1}^2] - [M_i(n)\bar{X}_i(n) - \mu_{i2}\mu_{i1}] < -b\} \\
&\leq P\{\bar{X}_i(n) - \mu_{i1} < -b(2N(N + \mu_{i1}))^{-1}\} \\
&\qquad + P\{\bar{X}_i(n) - \mu_{i1} > b(4N^2)^{-1}\} + P\{M_i(n) - \mu_{i2} > b(4\mu_{i1})^{-1}\} \qquad (3.7) \\
&\leq \exp\{-nb^2[2N^4(N + \mu_{i1})^2]^{-1}\} \\
&\qquad + \exp\{-nb^2[8N^4]^{-1}\} + \exp\{-nb^2[8N^4\mu_{i1}]^{-1}\} \\
&\leq 0(\exp(-nb_i)).
\end{aligned}
$$

Note that in (3.7), the first inequality is obtained from the fact that $0 \leq \bar{X}_i(n) \leq N, 0 \leq M_i(n) \leq N^2$ and an application of Bonferroni's inequality; the second inequality follows from an application of Theorem 2 of Hoeffding (1963) and the last inequality is obtained from the definition of b_i.

Hence, the proof of part a) is complete.

By the positivity of L_{i1} and L_{i2}, and by Lemma 3.1,

$$
\begin{aligned}
&P\{\Delta_{in}(x_i) \leq \Delta_{jn}(x_j) \text{ and } (Z_i(n) \leq 0 \text{ or } Z_j(n) \leq 0 \text{ or } Y_j(n) \leq 0 \text{ or } Y_j(n) \leq 0)\} \\
&\leq 0(\exp(-n \min(b_i, b_j))) \qquad (3.8) \\
&= 0(\exp(-nb_{ij})), \text{ where } b_{ij} = \min(b_i, b_j).
\end{aligned}
$$

Therefore, we then only need to consider the asymptotic behavior of $P\{\Delta_{in}(x_i) \leq \Delta_{jn}(x_j) \text{ and } (Z_i(n) > 0, Z_j(n) > 0, Y_i(n) > 0 \text{ and } Y_j(n) > 0)\}$.

Let $Q_{ij} = (x_i - x_j)L_{i2}L_{j2} + L_{i1}L_{j2} - L_{i2}L_{j1}$. Then $Q_{ij} > 0$ if $i \epsilon A(\underset{\sim}{x})$ and $j \epsilon B(\underset{\sim}{x})$. Therefore,

$$P\{\Delta_{in}(x_i) \le \Delta_{jn}(x_j) \text{ and } (Z_i(n) > 0, Z_j(n) > 0, Y_i(n) > 0 \text{ and } Y_j(n) > 0)\}$$

$$\le P\{(x_i - x_j)[Y_i(n)Y_j(n) - L_{i2}L_{j2}] < -Q_{ij}/3\} \tag{3.9}$$

$$+ P\{Z_i(n)Y_j(n) - L_{i1}L_{j2} < -Q_{ij}/3\}$$

$$+ P\{Y_i(n)Z_j(n) - L_{i2}L_{j1} > Q_{ij}/3\}.$$

With repeated applications of Bonferroni's inequality, we have the following inequalities:

$$P\{(x_i - x_j)[Y_i(n)Y_j(n) - L_{i2}L_{j2}] < -Q_{ij}/3\}$$

$$\le P\{Y_i(n) - L_{i2} < -Q_{ij}(6N^4)^{-1}\} + P\{Y_j(n) - L_{j2} < -Q_{ij}(6NL_{i2})^{-1}\} \tag{3.10.a}$$

if $x_i > x_j$;

$$P\{(x_i - x_j)[Y_i(n)Y_j(n) - L_{i2}L_{j2}] < -Q_{ij}/3\}$$

$$\le P\{Y_i(n) - L_{i2} > Q_{ij}(6N^4)^{-1}\} + P\{Y_j(n) - L_{j2} > Q_{ij}(6NL_{i2})^{-1}\} \tag{3.10.b}$$

if $x_i < x_j$;

$$P\{(x_i - x_j)[Y_i(n)Y_j(n) - L_{i2}L_{j2}] < -Q_{ij}/3\} = 0 \quad \text{if } x_i = x_j; \tag{3.10.c}$$

$$P\{Z_i(n)Y_j(n) - L_{i1}L_{j2} < -Q_{ij}/3\}$$

$$\le P\{Z_i(n) - L_{i1} < -Q_{ij}(6N^3)^{-1}\} + P\{Y_j(n) - L_{j2} < -Q_{ij}(6L_{i1})^{-1}\}; \tag{3.11}$$

and

$$P\{Y_i(n)Z_j(n) - L_{i2}L_{j1} > Q_{ij}/3\}$$

$$\leq P\{Y_i(n) - L_{i2} > Q_{ij}(6N^3)^{-1}\} + P\{Z_j(n) - L_{j1} > Q_{ij}(6L_{i2})^{-1}\}. \tag{3.12}$$

Then, by Lemma 3.1 and from Equations (3.9) through (3.12), we conclude that

$$P\{\Delta_{in}(x_i) \leq \Delta_{jn}(x_j) \text{ and } (Z_i(n) > 0, Z_j(n) > 0, Y_i(n) > 0 \text{ and } Y_j(n) > 0)\}$$

$$\leq 0(\exp(-na_{ij})) \text{ for some } a_{ij} > 0. \tag{3.13}$$

Now, from (3.6), (3.8) and (3.13), for each $\underset{\sim}{x}\epsilon\mathcal{X}$, $i\epsilon A(\underset{\sim}{x})$ and $j\epsilon B(\underset{\sim}{x})$,

$$P\{\Delta_{in}(x_i) \leq \Delta_{jn}(x_j)\} \leq 0(\exp(-n \min(b_{ij}, a_{ij}))). \tag{3.14}$$

Now, let $c_2 = \min_{i\neq j}\{\min(b_{ij}, a_{ij})\}$. Then $c_2 > 0$.

Based on the preceding, we have the following result.

Theorem 3.2. Let $\{d_n^*\}$ be the sequence of empirical Bayes selection rules defined in (2.12). Then $r(G, d_n^*) - r(G) \leq 0(\exp(-c_2 n))$ for some positive constant c_2.

Remark: Another selection problem related to the multinomial distribution is to select the least probable event; that is, to select the cell associated with $p_{[1]}$. If we consider the loss function

$$L(\underset{\sim}{p}, i) = p_i - p_{[1]}, \tag{3.15}$$

the difference between the selected and the least probable event, then under the statistical model described in Section 2, a uniformly randomized Bayes selection rule is $d_G = (d_{1G}, \ldots, d_{kG})$, where, for each $i = 1, \ldots, k$,

$$d_{iG} = d_{iG}(\underset{\sim}{x}) = \begin{cases} |\tilde{A}(\underset{\sim}{x})|^{-1} & \text{if } i \epsilon \tilde{A}(\underset{\sim}{x}), \\ 0 & \text{otherwise,} \end{cases} \tag{3.16}$$

and

$$\tilde{A}(\underset{\sim}{x}) = \{i | x_i + \alpha_i = \min_{1 \le j \le k} (\alpha_j + x_j)\}. \tag{3.17}$$

Let $\hat{\alpha}_{in}, \Delta_{in}(x_i)$ be defined as in (2.7) and (2.10), respectively. When α_0 is known, we let

$$\tilde{A}_n(\underset{\sim}{x}) = \{i | x_i + \hat{\alpha}_{in} = \min_{1 \le j \le k} (x_j + \hat{\alpha}_{jn})\}, \tag{3.18}$$

and define a randomized selection rule $\tilde{d}_n(\underset{\sim}{x}) = (\tilde{d}_{1n}(\underset{\sim}{x}), \ldots, \tilde{d}_{kn}(\underset{\sim}{x}))$ as follows:

$$\tilde{d}_{in}(\underset{\sim}{x}) = \begin{cases} |\tilde{A}_n(\underset{\sim}{x})|^{-1} & \text{if } i \epsilon \tilde{A}_n(\underset{\sim}{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{3.19}$$

When α_0 is unknown, we let

$$\tilde{A}^*_n(\underset{\sim}{x}) = \{i | \Delta_{in}(x_i) = \min_{1 \le j \le k} \Delta_{jn}(x_j)\}, \tag{3.20}$$

and define a randomized selection rule $d^*_n(\underset{\sim}{x}) = (d^*_{1n}(\underset{\sim}{x}), \ldots, d^*_{kn}(\underset{\sim}{x}))$ as follows:

$$d^*_{in}(\underset{\sim}{x}) = \begin{cases} |\tilde{A}^*_n(\underset{\sim}{x})|^{-1} & \text{if } i \epsilon \tilde{A}^*_n(\underset{\sim}{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{3.21}$$

Following a discussion analogous to that given earlier for the most probable event, we can see that $\{\tilde{d}_n\}$ and $\{d_n^*\}$ are both asymptotically optimal and have the following convergence rates:

$$0 \leq r(G, \tilde{d}_n) - r(G) \leq 0(\exp(-c_3 n)),$$

$$0 \leq r(G, d_n^*) - r(G) \leq 0(\exp(-c_4 n)),$$

for some positive constants c_3 and c_4, where $r(G)$ now denotes the minimum Bayes risk with respect to the loss function (3.15).

References

Alam, K. (1971). On selecting the most probable category. *Technometrics* **13**, 843–850.

Alam, K., Seo, K. and Thompson, J. R. (1971). A sequential sampling rule for selecting the most probable multinomial event. *Ann. Inst. Statist. Math.* **23**, 365–374.

Bechhofer, R. E., Elmaghraby, S. and Morse, N. (1959). A single- sample multiple-decision procedure for selecting the multinomial event which has the highest probability. *Ann. Math. Statist.* **30**, 102–119.

Bechhofer, R. E. and Kulkarni, R. V. (1984). Closed sequential procedures for selecting the multinomial events which have the largest probabilities. *Commun. Statist.-Theor. Meth.* **A13(24)**, 2997–3031.

Cacoullos, T. and Sobel, M. (1966). An inverse-sampling procedure for selecting the most probable event in a multinomial distribution. *Multivariate Analysis* (Ed. P. R. Krishnaiah), Academic Press, New York, 423–455.

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 Huang, D. Y. (1975). On subset selection procedures for Poisson populations and some applications to the multinomial selection problems. *Applied Statistics* (Ed. R. P. Gupta), North-Holland, Amsterdam, 97–109.

Gupta, S. S. and Leu, L. Y. (1988). On Bayes and empirical Bayes rules for selecting good populations. To appear in Probability and Information Theory 3, a volume in the *Journal of Statistical Planning and Inference*, (Eds. M. Behara and D. Bierlein).

Gupta, S. S. and Liang, T. (1986). Empirical Bayes rules for selecting good binomial populations. *Adaptive Statistical Procedures and Related Topics* (Ed. J. Van Ryzin), IMS Lecture Notes-Monograph Series, Vol. 8, 110–128.

Gupta, S. S. and Liang, T. (1988). Empirical Bayes rules for selecting the best binomial population. *Statistical Decision Theory and Related Topics-IV* (Eds. S. S. Gupta and J. O. Berger), Vol. 1, 213–224, Springer-Verlag, New York.

Gupta, S. S. and Nagel, K. (1967). On selection and ranking procedures and order statistics from the multinomial distribution. *Sankhya Ser. B* **29**, 1–17.

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

Panchapakesan, S. (1971). On a subset selection procedure for the most probable event in a multinomial distribution. *Statistical Decision Theory and Related Topics* (Eds. S. S. Gupta and J. Yackel), Academic Press, New York, 275–298.

Ramey, Jr. J. T. and Alam, K. (1979). A sequential procedure for selecting the most probable multinomial event. *Biometrika* **66**, 171–173.

Ramey, Jr. J. T. and Alam, K. (1980). A Bayes sequential procedure for selecting the most probable multinomial event. *Commun. Statist.-Theor. Meth.* **A9(3)**, 265–276.

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.

BEFORE COMPLETING FORM

1. REPORT NUMBER Technical Report #86-54	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PARAMETRIC EMPIRICAL BAYES RULES FOR SELECTING THE MOST PROBABLE MULTINOMIAL EVENT		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Technical Report #86-54
7. AUTHOR(s) Shanti S. Gupta and TaChen Liang		8. CONTRACT OR GRANT NUMBER(s) N00014-84-C-0167 NSF DMS-8606964
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK, AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. REPORT DATE December 1986
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION, DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Asymptotically optimal, Bayes rules, empirical Bayes rules, parametric empirical Bayes, Dirichlet prior, the most probable event.

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

Consider a multinomial population with $k(\geq 2)$ cells and the associated probability vector $\underset{\sim}{p} = (p_1,\ldots,p_k)$. Let $p_{[k]} = \max_{1\leq i\leq k} p_i$. A cell associated with $p_{[k]}$ is called the most probable event. We are interested in selecting the most probable event. Let i denote the index of the selected cell. Under the loss function $L(\underset{\sim}{p},i) = p_{[k]} - p_i$, this statistical selection problem is studied via a parametric empirical Bayes approach. Two empirical Bayes selection rules are proposed. They are shown to be asymptotically optimal at least of order $O(\exp(-c_i n))$ for some

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

positive constants c_i, $i = 1,2$, where n is the number of accumulated past experiences (observations) at hand. Finally, for the problem of selecting the least probable event associated with $p_{[1]}$ under the loss $p_i - p_{[1]}$, two empirical Bayes selection rules are also proposed. The corresponding rates of convergence are found to be at least of order $O(\exp(-c_i n))$ for some positive constants c_i, $i = 3, 4$.