BAYES AND EMPIRICAL BAYES RULES FOR SELECTING FAIR MULTINOMIAL POPULATIONS*

by

Shanti S. Gupta
Department of Statistics
Purdue University
West Lafayette, IN 47907

and

TaChen Liang
Department of Mathematics
Wayne State University
Detroit, MI 48202

Technical Report #90-01C

Department of Statistics
Purdue University

January 1990
Revised April 1991
Revised October 1992

* This research was supported in part by the Office of Naval Research Contract N00014-88-K-0170 and NSF Grants DMS-8702620, DMS-8923071, and DMS-8717799 at Purdue University.

◇

Bayes and Empirical Bayes Rules For Selecting Fair Multinomial Populations[1]

SHANTI S. GUPTA
Department of Statistics
Purdue University
West Lafayette, IN 47907

TACHEN LIANG
Department of Mathematics
Wayne State University
Detroit, MI 48202

Abstract

This paper deals with the problem of selecting fair multinomial populations compared with a standard. Two selection procedures are investigated: the natural selection procedure of Gupta and Leu (1990) and an empirical Bayes simultaneous selection procedure. It is proved that the natural selection procedure is a Bayes procedure relative to a symmetric Dirichlet prior, and therefore it is admissible. For the empirical Bayes simultaneous selection procedure, the associated asymptotic optimality is investigated. It is shown that the proposed procedure is asymptotically optimal relative to a class of symmetric Dirichlet priors, with a rate of convergence of order $O(\exp(-\tau k + \ln k))$ for some positive constant τ, where k is the number of populations involved in the selection problem. Also, presented are results of a simulation study of the small sample performance of the empirical Bayes procedure.

[1] *AMS Subject Classification*: Primary 62F07; Secondary 62C12
Key words and Phrases: Fair Multinomial Population, Entropy Function, Gini-Simpson Index, Natural Selection Procedure, Empirical Bayes Selection Procedure, Asymptotically Optimal, Rate of Convergence.

1 Introduction

The concept of diversity within a population is of considerable importance in statistical theory and applications. The problem of measuring diversity arises in a variety of studies in ecology, sociology, econometrics, genetics and many other sciences. For a multinomial population with m cells, the index of diversity is a function of the corresponding probability parameter vector $\underline{p} = (p_1, \ldots, p_m)$. In practice, a Schur-convex or Schur-concave function of $\underline{p}$ may be appropriate. There are two measures of diversity of a multinomial population which have been commonly used. These are Shannon's entropy function and the Gini-Simpson index. The notion of the entropy function was introduced by Shannon (1948). The Gini-Simpson index was introduced by Gini (1912) and Simpson (1949). Both are Schur-concave function of $\underline{p}$.

In the literature, selection procedures using indices of diversity as selection criteria have been studied by many authors. Gupta and Huang (1976) studied the problem of selecting the population with the largest entropy function for binomial distributions. Gupta and Wong (1975) considered the problem of selecting a subset containing the population with the largest entropy for multinomial distributions. Dudewicz and Van der Meulen (1981) investigated a selection procedure based on a generalized entropy function. Alam, Mitra, Rizvi and Saxena (1986) studied selection procedures based on Shannon's entropy function and the Gini-Simpson index using the indifference zone approach. Rizvi, Alam and Saxena (1987) also considered a subset selection procedure based on certain other diversity indexes. Recently, Gupta and Leu (1990) and Liang and Panchapakesan (1991) have studied certain selection procedures based on the Gini-Simpson index.

In this paper, we are dealing with the problem of selecting fair multinomial populations compared with a standard level. Consider k independent multinomial populations $\pi_1, \ldots, \pi_k$. For each $i = 1, \ldots, k$, population π_i has m cells, and is characterized by the corresponding probability parameter vector $\underline{p}_i = (p_{i1}, \ldots, p_{im})$, where $0 \le p_{ij} \le 1$, $j = 1, \ldots, m$, and $\sum_{j=1}^{m} p_{ij} = 1$ for each $i = 1, \ldots, k$. Define

$$\theta_i = \Psi(\underline{p}_i) = \sum_{j=1}^{m}(p_{ij} - \frac{1}{m})^2. \tag{1.1}$$

We use θ_i as a measure of diversity (or uniformity) of population π_i. Note that since

$\theta_i = \sum_{j=1}^{m} p_{ij}^2 - \frac{1}{m}$, it is essentially equivalent to the Gini-Simpson index. Also note that $0 \leq \theta_i \leq 1 - \frac{1}{m}$. For a given constant θ_0, $0 < \theta_0 < 1 - \frac{1}{m}$, population π_i is said to be a fair population if $\theta_i \leq \theta_0$ and a bad population, otherwise. Our goal is to derive statistical selection procedures for selecting all fair populations while excluding all bad populations. Since a fair multinomial would imply equal cell probabilities analogous to a fair coin, the problem at hand is equivalent to selecting multinomial populations that are fairer (more homogeneous) than the standard. It should be noted that the problem of selecting fair multinomial populations has been considered by Gupta and Leu (1990) through a classical approach.

Let $\Omega = \{\underline{\underline{p}} = (\underline{p}_1, \ldots, \underline{p}_k) | \underline{p}_i = (p_{i1}, \ldots, p_{im}), 0 \leq p_{ij} \leq 1, j = 1, \ldots, m, \sum_{j=1}^{m} p_{ij} = 1$ for each $i = 1, \ldots, k\}$ be the parameter space and let $\mathcal{A} = \{s | s \subset \{1, \ldots, k\}\}$ be the action space. When action s is taken, it means that population π_i is selected as a fair population if $i \in s$ and excluded as a bad population if $i \notin s$. For $\underline{\underline{p}} \in \Omega$ and action $s \in \mathcal{A}$, the loss function $L(\underline{\underline{p}}, s)$ is defined to be:

$$L(\underline{\underline{p}}, s) = \sum_{i \in s} (\theta_i - \theta_0) I_{(\theta_0, 1-\frac{1}{m}]}(\theta_i) + \sum_{i \notin s} (\theta_0 - \theta_i) I_{[0, \theta_0]}(\theta_i). \quad (1.2)$$

In (1.2), the first summation is the loss due to selecting certain bad populations and the second summation is the loss due to not selecting certain fair populations.

The content of this paper consists of two parts. In Section 2, we investigate some optimal properties of the natural selection procedure of Gupta and Leu (1990). It is shown that, for the loss function $L(\underline{\underline{p}}, s)$ of (1.2), the natural selection procedure is Bayes relative to some symmetric Dirichlet prior, and therefore, it is admissible. Section 3 deals with this selection problem through a parametric empirical Bayes approach. An empirical Bayes selection procedure is proposed and the corresponding asymptotic optimality is investigated. It is shown that the proposed empirical Bayes selection procedure is asymptotically optimal relative to a class of symmetric Dirichlet priors. The rate of convergence of the proposed empirical Bayes selection procedure is also established, and shown to be of order $O(\exp(-\tau k + \ln k))$ for some positive constant τ, where k is the number of populations involved in the selection problem. Finally, a simulation study is carried out to investigate the small sample performance of the empirical Bayes selection rule. The results of this study are presented in the last section of the paper.

2 Optimality of Natural Selection Procedure

For each $i = 1, \ldots, k$, let $\underline{X}_i = (X_{i1}, \ldots, X_{im})$ be the random observation associated with population π_i, where X_{ij}, $1 \leq j \leq m$, are nonnegative integer random variables such that $0 \leq X_{ij} \leq N$ and $\sum_{j=1}^{m} X_{ij} = N$. Then, $\underline{X}_i$ has the probability function

$$f(\underline{x}_i | \underline{p}_i) = \frac{N!}{\prod_{j=1}^{m} (x_{ij}!)} \prod_{j=1}^{m} p_{ij}^{x_{ij}} \tag{2.1}$$

at point $\underline{x}_i = (x_{i1}, \ldots, x_{im})$ for which $0 \leq x_{ij} \leq N$, $1 \leq j \leq m$ and $\sum_{j=1}^{m} x_{ij} = N$. Let $\mathcal{X}_i$ be the sample space generated by $\underline{X}_i$. Let $\underline{\underline{X}} = (\underline{X}_1, \ldots, \underline{X}_k)$ and denote the corresponding observed value by $\underline{\underline{x}} = (\underline{x}_1, \ldots, \underline{x}_k)$. Also, let $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_k$ denote the sample space of $\underline{\underline{X}}$.

A selection procedure $d = (d_1, \ldots, d_k)$ is defined to be a mapping from the sample space $\mathcal{X}$ into the product space $[0,1]^k$. That is, for each $\underline{\underline{x}} = (\underline{x}_1, \ldots, \underline{x}_k) \in \mathcal{X}$, $d(\underline{\underline{x}}) = (d_1(\underline{\underline{x}}), \ldots, d_k(\underline{\underline{x}}))$, where $d_i(\underline{\underline{x}})$ is the probability of selecting population π_i as a fair population given $\underline{\underline{X}} = \underline{\underline{x}}$ is observed. We let D denote the class of all selection procedures defined in the above way.

For each $i = 1, \ldots, k$, $(\frac{X_{i1}}{N}, \ldots, \frac{X_{im}}{N})$ is the maximum likelihood estimator of $\underline{p}_i = (p_{i1}, \ldots, p_{im})$. From (1.1), it is natural and reasonable to estimate θ_i by $\hat{\theta}_i = \sum_{j=1}^{m} (\frac{X_{ij}}{N})^2 - \frac{1}{m}$. Gupta and Leu (1990) proposed a natural selection procedure $d^N = (d_1^N, \ldots, d_k^N)$ based on $\hat{\theta}_i$, $i = 1, \ldots, k$, which is equivalent to the following: For each $i = 1, \ldots, k$,

$$d_i^N(\underline{\underline{X}}) = \begin{cases} 1 & \text{if } \hat{\theta}_i \leq \delta, \\ 0 & \text{otherwise,} \end{cases} \tag{2.2}$$

where δ is a prespecified positive constant such that $0 < \delta < 1 - \frac{1}{m}$. Since this natural selection procedure is heavily dependent on the constant δ, we denote this procedure by $d^{N(\delta)} = (d_1^{N(\delta)}, \ldots, d_k^{N(\delta)})$.

In the following, it is assumed that for each $i = 1, \ldots, k$, the parameter vector $\underline{p}_i = (p_{i1}, \ldots, p_{im})$ is a realization of the random vector $\underline{P}_i = (P_{i1}, \ldots, P_{im})$. It is also assumed that $\underline{P}_1, \ldots, \underline{P}_k$ are iid with a common prior distribution G_α belonging to a class of symmetric Dirichlet distributions $\mathcal{C}$, where

$$\mathcal{C} = \{g_\alpha | g_\alpha(\underline{p}_i) = \frac{\Gamma(m\alpha)}{[\Gamma(\alpha)]^m} \prod_{j=1}^{m} p_{ij}^{\alpha-1}, \ 0 \le p_{ij} \le 1, \ j = 1, \ldots, m, \ \sum_{j=1}^{m} p_{ij} = 1\}. \quad (2.3)$$

For a prior distribution $G_\alpha \in \mathcal{C}$ and a selection procedure $d = (d_1, \ldots, d_k) \in D$, we denote the corresponding Bayes risk by $r(G_\alpha, d)$. From (1.2) and the statistical model described previously,

$$r(G_\alpha, d) = \sum_{i=1}^{k} r_i(G_\alpha, d_i) \quad (2.4)$$

where

$$r_i(G_\alpha, d_i) = \sum_{\underline{\underline{x}} \in \mathcal{X}} d_i(\underline{\underline{x}}) \left[\sum_{j=1}^{m} E[P_{ij}^2 | \underline{x}_i] - \frac{1}{m} - \theta_0 \right] \prod_{j=1}^{k} f(\underline{x}_j)$$
$$+ \int_{\Omega_i(\theta_0)} (\theta_0 - \theta_i) g_\alpha(\underline{p}_i) d\underline{p}_i \quad (2.5)$$

$f(\underline{x}_i) = \int f(\underline{x}_i | \underline{p}_i) g_\alpha(\underline{p}_i) d\underline{p}_i = \frac{N!}{\prod_{j=1}^{m}(x_{ij}!)} \frac{\Gamma(m\alpha)}{[\Gamma(\alpha)]^m} \frac{\prod_{j=1}^{m} \Gamma(\alpha + x_{ij})}{\Gamma(m\alpha + N)}$,
and $\Omega_i(\theta_0) = \{\underline{p}_i = (p_{i1}, \ldots, p_{im}) | \theta_i \le \theta_0\}$.

Since the second term of (2.5) does not depend on the selection procedure d, a Bayes selection procedure $d^{G_\alpha} = (d_1^{G_\alpha}, \ldots, d_k^{G_\alpha})$ can be obtained as follows: For each $i = 1, \ldots, k$,

$$d_i^{G_\alpha}(\underline{\underline{x}}) = \begin{cases} 1 & \text{if } \sum_{j=1}^{m} E[P_{ij}^2 | \underline{x}_i] - \frac{1}{m} \le \theta_0, \\ 0 & \text{otherwise.} \end{cases} \quad (2.6)$$

Then, we have the following theorem.

Theorem 2.1. For each positive constant δ such that $0 < \delta < 1 - \frac{1}{m}$ and $\frac{m\delta N^2 + mN - N}{mN(N+1)} > \theta_0$, for the loss function $L(\underline{\underline{p}}, s)$, the natural selection procedure $d^{N(\delta)}$ given in (2.2) is a Bayes procedure relative to some symmetric prior distribution G_α.

Proof: First, straightforward computation yields for each $i = 1, \ldots, k$ and $j = 1, \ldots, m$,

$$E[P_{ij}^2 | \underline{X}_i] = \frac{(X_{ij} + \alpha + 1)(X_{ij} + \alpha)}{(m\alpha + N + 1)(m\alpha + N)}.$$

and therefore

$$\sum_{j=1}^{m} E[P_{ij}^2|\underline{X}_i] - \frac{1}{m} = \frac{(m^2-m)\alpha + mN + m\sum_{j=1}^{m} X_{ij}^2 - N^2 - N}{m(m\alpha+N+1)(m\alpha+N)}. \tag{2.7}$$

Note that

$$\hat{\theta}_i \leq \delta \Longleftrightarrow \sum_{j=1}^{m} E[P_{ij}^2|\underline{X}_i] - \frac{1}{m} \leq H(\alpha)$$

where

$$H(\alpha) = \frac{(\delta + \frac{1}{m})mN^2 + (m^2-m)\alpha + mN - N^2 - N}{m(m\alpha+N+1)(m\alpha+N)}.$$

Thus, it suffices to prove that for given θ_0, $0 < \theta_0 < 1 - \frac{1}{m}$, there exists a positive α such that $H(\alpha) = \theta_0$. Since $H(\alpha)$ is decreasing in α, $H(0) = \frac{m\delta N^2 + mN - N}{mN(N+1)} > \theta_0$ by the assumption, $\lim_{\alpha\to\infty} H(\alpha) = 0 < \theta_0$ and $H(\alpha)$ is a continuous function of α on $[0,\infty)$, there exists a unique $\alpha \equiv \alpha(\theta_0) > 0$ such that $H(\alpha) = \theta_0$. This implies that the natural selection procedure $d^{N(\delta)}$ is the Bayes procedure relative to the symmetric Dirichlet prior $G_{\alpha(\theta_0)}$. The proof is now complete. □

The following corollary is a direct consequent of Theorem 2.1.

Corollary 2.1. For each positive constant δ such that $0 < \delta < 1 - \frac{1}{m}$, $\frac{m\delta N^2 + mN - N}{mN(N+1)} > \theta_0$, the natural selection procedure is admissible for the loss function $L(\underline{\underline{p}}, s)$.

3 An Empirical Bayes Selection Procedure

We assume that the hyperparameter α of the symmetric Dirichlet prior G_α is unknown. In this situation, it is not possible to apply the Bayes selection procedure d^{G_α} for the problem at hand. Thus, the empirical Bayes approach is employed here.

For each $i = 1, \ldots, k$, let $W_i = \sum_{j=1}^{m} X_{ij}^2$ and let w_i denote the observed value of W_i. Under the preceding statistical model, $W_1, \ldots, W_k$ are iid random variables such that $\frac{N^2}{m} \leq W_i \leq N^2$. It follows from straightforward computations that

$$\mu_2 \equiv E[W_1] = N[1 + \frac{(N-1)(\alpha+1)}{m\alpha+1}], \tag{3.1}$$

and therefore,

$$\alpha = \frac{N^2 - \mu_2}{m\mu_2 - N(m+N-1)}. \tag{3.2}$$

From (3.2) and (2.7), for each $i = 1, \ldots, k$,

$$\begin{aligned}
&\sum_{j=1}^{m} E[P_{ij}^2 | \underline{x}_i] - \frac{1}{m} \\
&\quad = \{(m^2 - m)(N^2 - \mu_2) + [m\mu_2 - N(m+N-1)][mN + mw_i - N^2 - N]\} \\
&\qquad \times (m\mu_2 - mN - N^2 + N)/[m(N-1)(m\mu_2 - N^2)N(m\mu_2 - N^2 - m + 1)] \\
&\quad \equiv Q_i(\mu_2 | w_i),
\end{aligned} \tag{3.3}$$

where $w_i = \sum_{j=1}^{m} x_{ij}^2$.

Note that for w_i(or $\underline{x}_i$) being kept fixed, $Q_i(\mu_2|w_i)$ is increasing in μ_2.

Define, $\hat{\mu}_2 = \frac{1}{k}\sum_{i=1}^{k} W_i$. We will use $\hat{\mu}_2$ to estimate μ_2 and use $Q_i(\hat{\mu}_2|w_i)$ to estimate $Q_i(\mu_2|w_i)$. However, by the definition of μ_2 (see (3.1)), $N + \frac{N(N-1)}{m} < \mu_2 < N^2$ and $Q_i(\mu_2|w_i)$ tends to zero as μ_2 tends to $N + \frac{N(N-1)}{m}$. Also, it is possible that $\hat{\mu}_2 \leq N + \frac{N(N-1)}{m}$. Thus we define

$$\varphi_i^*(w_i) = \begin{cases} Q_i(\hat{\mu}_2|w_i) & \text{if } \hat{\mu}_2 > N + \frac{N(N-1)}{m}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

We now propose an empirical Bayes selection procedure $d^* = (d_1^*, \ldots, d_k^*)$ as follows: For each $i = 1, \ldots, k$,

$$d_i^*(\underline{x}) = \begin{cases} 1 & \text{if } \varphi_i^*(w_i) \leq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

The following example describes the empirical Bayes selection procedure.

Example: Suppose that $k = 10$, $m = 5$, $N = 20$ and $\theta_0 = 0.1$. Let the data $(X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5})$, $i = 1, \ldots, 10$, from each of the 10 populations, be as listed below.

i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}
1	3	5	4	7	1
2	4	1	2	8	5
3	10	4	3	2	1
4	5	0	0	10	5
5	6	0	5	6	3
6	8	8	2	1	1
7	13	1	1	1	4
8	7	8	2	3	0
9	3	2	2	9	4
10	12	2	3	1	2

Using $W_i = \sum_{j=1}^{m} X_{ij}^2$ and (3.4), the values of the empirical Bayes estimators $\varphi_i^*(W_i)$ are as follows:

i	1	2	3	4	5
W_i	100	110	130	150	106
$\varphi_i^*(W_i)$	0.04610	0.06566	0.09085	0.11603	0.06063

i	6	7	8	9	10
W_i	134	188	126	114	162
$\varphi_i^*(W_i)$	0.09588	0.16388	0.08581	0.07070	0.13114

Comparing the values of $\varphi_i^*(W_i)$ with $\theta_0 = 0.1$, the empirical Bayes rule selects populations π_1, π_2, π_3, π_5, π_6, π_8 and π_9 as fair populations. According to this rule π_4, π_7, and π_{10} are not selected.

In the following, we will investigate the asymptotic performance of the empirical Bayes selection procedure d^* for the case where k, the number of populations involved in the selection problem under study, is sufficiently large.

Since $d^{G_\alpha} = (d_1^{G_\alpha}, \ldots, d_k^{G_\alpha})$ is the Bayes selection procedure, for the empirical Bayes selection procedure $d^* = (d_1^*, \ldots, d_k^*)$, $r_i(G_\alpha, d_i^*) - r_i(G_\alpha, d_i^{G_\alpha}) \geq 0$ for each $i = 1, \ldots k$ and therefore, $r(G_\alpha, d^*) - r(G_\alpha, d^{G_\alpha}) = \sum_{i=1}^{k} [r_i(G_\alpha, d_i^*) - r_i(G_\alpha, d_i^{G_\alpha})] \geq 0$. This nonnegative regret value $r(G_\alpha, d^*) - r(G_\alpha, d^{G_\alpha})$ will be used as a measure of performance of the empirical Bayes selection procedure d^*.

Definition 3.1. A selection procedure $d = (d_1, \ldots, d_k) \in D$ is said to be asymptotically optimal of order $\{\beta_k\}$ relative to the prior distribution G if

$$r(G, d) - r(G, d^G) = O(\beta_k)$$

where $\{\beta_k\}$ is a sequence of positive numbers such that $\lim_{k\to\infty} \beta_k = 0$.

For each $i = 1, \ldots, k$ and for the fixed μ_2, $Q_i(\mu_2|w_i)$, which is defined in (3.3), can be viewed as a function of w_i. It is clear that $Q_i(\mu_2|w_i)$ is increasing in w_i. Let

$$A_i = \{w_i | w_i = w_i(\underline{x}_i) = \sum_{j=1}^{m} x_{ij}^2,\ \underline{x}_i \in \mathcal{X}_i,\ Q_i(\mu_2|w_i) < \theta_0\}$$

$$B_i = \{w_i | w_i = w_i(\underline{x}_i) = \sum_{j=1}^{m} x_{ij}^2,\ \underline{x}_i \in \mathcal{X}_i,\ Q_i(\mu_2|w_i) > \theta_0\}$$

From the statistical model under consideration, $Q_i(\mu_2|w) = Q_j(\mu_2|w) = Q(\mu_2|w)$ (say) for all $i, j = 1, \ldots, k$. Thus, $A_1 = \cdots = A_k = A$ (say) and $B_1 = \cdots = B_k = B$ (say).

Let $h(w)$ denote the common marginal probability function of the iid random variables $W_i = \sum_{j=1}^{m} X_{ij}^2$, $i = 1, \ldots, k$. From (2.5), (2.6) and (3.5), straightforward computation yields

$$\begin{aligned} 0 &\le r_i(G_\alpha, d_i^*) - r_i(G_\alpha, d_i^{G_\alpha}) \\ &\le \sum_{w_i \in A} P_i\{\varphi_i^*(w_i) > \theta_0 | W_i = w_i\} h(w_i) \\ &\quad + \sum_{w_i \in B} P_i\{\varphi_i^*(w_i) \le \theta_0 | W_i = w_i\} h(w_i) \end{aligned} \tag{3.6}$$

where the probability measure P_i is computed with respect to $(W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_k)$.

Thus, it suffices to investigate the asymptotic behavior of $P_i\{\varphi_i^*(w_i) \le \theta_0 | W_i = w_i\}$ for $w_i \in B$ and $P_i\{\varphi_i^*(w_i) > \theta_0 | W_i = w_i\}$ for $w_i \in A$.

Lemma 3.1. For each $c > 0$ and for sufficiently large k,

(a) $P_i\{\hat{\mu}_2 - \mu_2 < -c | W_i = w_i\} \leq O(\exp(-kc^2N^{-4}(1-\frac{1}{m})^{-2}/2))$.

(b) $P_i\{\hat{\mu}_2 - \mu_2 > c | W_i = w_i\} \leq O(\exp(-kc^2N^{-4}(1-\frac{1}{m})^{-2}/2))$.

(c) $P_i\{\hat{\mu}_2 \leq N + \frac{N(N-1)}{m} | W_i = w_i\} \leq O(\exp(-k(\mu_2 - N - (N-1)N/m)^2N^{-4}(1-\frac{1}{m})^{-2}/2))$.

Note that the above upper bounds are independent of the value w_i.

Proof: (a). Let $\hat{\mu}_2(i) = \frac{1}{k-1}\sum_{\substack{j=1\\j\neq i}}^{k} W_j$. Then,

$$P_i\{\hat{\mu}_2 - \mu_2 < -c | W_i = w_i\} = P_i\{\hat{\mu}_2(i) - \mu_2 < -\frac{kc}{k-1} + \frac{\mu_2 - w_i}{k-1}\}.$$

Note that $\frac{N^2}{m} \leq w_i \leq N^2$ for all $w_i = w_i(\underline{x}_i) = \sum_{j=1}^{m} x_{ij}^2$. Thus for k sufficiently large $-\frac{kc}{k-1} + \frac{\mu_2 - w_i}{k-1} \leq -\frac{c}{2}$. Hence, we obtain, for k sufficiently large, that

$$\begin{aligned}
&P_i\{\hat{\mu}_2 - \mu_2 < -c | W_i = w_i\} \\
&\leq P_i\{\hat{\mu}_2(i) - \mu_2 < -\frac{c}{2}\} \\
&\leq \exp\{-kc^2N^{-4}(1-\frac{1}{m})^{-2}/2\},
\end{aligned}$$

where the last inequality follows from Theorem 2 of Hoeffding (1963). Note that the upper bound is independent of w_i.

The proof of part (b) is similar to that of part (a). By letting $c = \mu_2 - [N + \frac{N(N-1)}{m}]$, then $c > 0$. Thus, the result of part (c) follows directly from part (a). □

Lemma 3.2. For $w_i \in A$,

$$P_i\{\varphi_i^*(w_i) > \theta_0 | W_i = w_i\} \leq O(\exp(-k(Q_i^{-1}(\theta_0|w_i) - \mu_2)^2N^{-4}(1-\frac{1}{m})^{-2}/2))$$

where $Q_i^{-1}(\cdot|w_i)$ is the inverse function of $Q_i(\cdot|w_i)$.

Proof: From (3.4) and the fact that $\theta_0 > 0$,

$$P_i\{\varphi_i^*(w_i) > \theta_0 | W_i = w_i\} = P_i\{Q_i(\hat{\mu}_2|w_i) > \theta_0 | W_i = w_i\}. \qquad (3.7)$$

Now, for each fixed $w_i \in A$, $Q_i(\mu|w_i)$ is strictly increasing in μ for $N + \frac{N(N-1)}{m} < \mu < N^2$ and $Q_i(\mu_2|w_i) < \theta_0$. Thus $\mu_2 < Q_i^{-1}(\theta_0|w_i)$. Then,

$$\begin{aligned} Q_i(\hat{\mu}_2|w_i) > \theta_0 &\Longleftrightarrow \hat{\mu}_2 > Q_i^{-1}(\theta_0|w_i) \\ &\Longleftrightarrow \hat{\mu}_2 - \mu_2 > Q_i^{-1}(\theta_0|w_i) - \mu_2 > 0. \end{aligned} \tag{3.8}$$

Combining (3.7) and (3.8), by Lemma 3.1 (b), we obtain, for $w_i \in A$,

$$\begin{aligned} &P_i\{\varphi_i^*(w_i) > \theta_0|W_i = w_i\} \\ &\quad = P_i\{\hat{\mu}_2 - \mu_2 > Q_i^{-1}(\theta_0|w_i) - \mu_2|W_i = w_i\} \\ &\quad \le O(\exp(-k(Q_i^{-1}(\theta_0|w_i) - \mu_2)^2 N^{-4}(1 - \frac{1}{m})^{-2}/2)). \end{aligned}$$

Thus, the proof of this lemma is complete. □

Lemma 3.3. For $w_i \in B$,

$$\begin{aligned} P_i\{\varphi_i^*(w_i) \le \theta_0|W_i = w_i\} &\le O(\exp(-k(Q_i^{-1}(\theta_0|w_i) - \mu_2)^2 N^{-4}(1 - \frac{1}{m})^{-2}/2)) \\ &\quad + O(\exp(-k(\mu_2 - N - (N-1)N/m)^2 N^{-4}(1 - \frac{1}{m})^{-2}/2)). \end{aligned}$$

Proof:

$$\begin{aligned} &P_i\{\varphi_i^*(w_i) \le \theta_0|W_i = w_i\} \\ &= P_i\{\varphi_i^*(w_i) \le \theta_0,\ \hat{\mu}_2 \le N + \frac{N(N-1)}{m}|W_i = w_i\} \\ &\quad + P_i\{\varphi_i^*(w_i) \le \theta_0,\ \hat{\mu}_2 > N + \frac{N(N-1)}{m}|W_i = w_i\}. \end{aligned} \tag{3.9}$$

From Lemma 3.1 (c),

$$\begin{aligned} &P_i\{\varphi_i^*(w_i) \le \theta_0,\ \hat{\mu}_2 \le N + \frac{N(N-1)}{m}|W_i = w_i\} \\ &\le O(\exp(-k(\mu_2 - N - (N-1)N/m)^2 N^{-4}(1 - \frac{1}{m})^{-2}/2)). \end{aligned} \tag{3.10}$$

From (3.4) and an argument analogous to that given in the proof of Lemma 3.2, we have

$$\begin{aligned} &P_i\{\varphi_i^*(w_i) \le \theta_0,\ \hat{\mu}_2 > N + \frac{N(N-1)}{m}|W_i = w_i\} \\ &= P_i\{Q_i(\hat{\mu}_2|w_i) \le \theta_0,\ \hat{\mu}_2 > N + \frac{N(N-1)}{m}|W_i = w_i\} \\ &\le P_i\{\hat{\mu}_2 - \mu_2 \le Q_i^{-1}(\theta_0|w_i) - \mu_2|W_i = w_i\} \\ &= O(\exp(-k(Q_i^{-1}(\theta_0|w_i) - \mu_2)^2 N^{-4}(1 - \frac{1}{m})^{-2}/2)) \end{aligned} \tag{3.11}$$

where $Q_i^{-1}(\theta_0|w_i) - \mu_2 < 0$ since $w_i \in B$. Thus, the lemma follows from (3.9)–(3.11) $\square$

Let $\tau_1 = \min\{(Q_i^{-1}(\theta_0|w_i) - \mu_2)^2 N^{-4}(1-\frac{1}{m})^{-2}/2 | w_i \in A_i \cup B_i\}$. Then, $\tau_1 > 0$ since $Q_i^{-1}(\theta_0|w_i) - \mu_2 \neq 0$ for all $w_i \in A_i \cup B_i$ and $A_i \cup B_i$ is a finite set. Then $\tau \equiv \min(\tau_1, (\mu_2 - N - (N-1)N/m)^2 N^{-4}(1-\frac{1}{m})^{-2}/2)) > 0$.

The following theorem describes the asymptotic optimality property of the empirical Bayes selection procedure $d^* = (d_1^*, \ldots, d_k^*)$.

Theorem 3.1. Let $d^* = (d_1^*, \ldots, d_k^*)$ be the empirical Bayes selection procedure defined through (3.4) and (3.5). Suppose that the prior is a member of the class $\mathcal{C}$ of symmetric Dirichlet distributions given in (2.3). Then

(a) For each $i = 1, \ldots, k$, $r_i(G_\alpha, d_i^*) - r_i(G_\alpha, d_i^{G_\alpha}) \leq O(\exp(-\tau k))$, and

(b) $r(G_\alpha, d^*) - r(G_\alpha, d^{G_\alpha}) \leq O(\exp(-\tau k + \ln k))$

where τ is the positive constant defined previously.

Proof: Part (b) is a direct result of part (a). Thus, we need to prove part (a) only. From (3.6), and Lemmas 3.2 and 3.3,

$$\begin{aligned} & r_i(G_\alpha, d_i^*) - r_i(G_\alpha, d_i^{G_\alpha}) \\ & \leq O(\exp(-\tau k)) \sum_{w_i \in A \cup B} h(w_i) \\ & = O(\exp(-\tau k)). \end{aligned}$$

The theorem now follows. $\square$

4 Small Sample Performance: A Simulation Study

A Monte Carlo study was designed to investigate the small sample performance of the proposed empirical Bayes selection rule d^*. The simulation scheme used is as follows:

(1) For given α and m, generate $k-1$ random vectors $\underline{p}_i = (p_{i1}, \ldots, p_{im})$, $i = 1, \ldots, k-1$, according to the symmetric Dirichlet distribution

$$g_\alpha(\underline{p}) = \frac{\Gamma(m\alpha)}{[\Gamma(\alpha)]^m} \prod_{i=1}^{m} p_i^{\alpha-1}$$

where $\underline{p} = (p_1, \ldots, p_m)$, $0 \le p_i \le 1$ and $\sum_{i=1}^{m} p_i = 1$.

(2) For each N and each $i = 1, \ldots, k-1$, generate random vector $\underline{X}_i = (X_{i1}, \ldots, X_{im})$ according to the multinomial distribution

$$f(\underline{x}|\underline{p}_i) = \frac{N!}{\prod_{j=1}^{m} x_j!} p_{ij}^{x_j},$$

where $0 \le x_j \le N$ and $\sum_{j=1}^{m} x_j = N$. Note that under the given statistical model,

$$r_1(G_\alpha, d_1^*) = r_2(G_\alpha, d_2^*) = \ldots = r_k(G_\alpha, d_k^*), \text{ and}$$
$$r_1(G_\alpha, d_1^{G_\alpha}) = r_2(G_\alpha, d_2^{G_\alpha}) = \ldots = r_k(G_\alpha, d_k^{G_\alpha}).$$

Therefore,

$$r(G_\alpha, d^*) - r(G_\alpha, d^{G_\alpha}) = k[r_k(G_\alpha, d_k^*) - r_k(G_\alpha, d_k^{G_\alpha})].$$

Hence, it suffices to generate $r_k(G_\alpha, d_k^*) - r(G_\alpha, d_k^{G_\alpha})$ only.

(3) Let $\underline{x}_k$ be an observed vector obtained from a population with probability function $f(\underline{x}) = \int f(\underline{x}|\underline{p}) g_\alpha(\underline{p}) d\underline{p}$. Use $\underline{X}_1, \ldots, \underline{X}_{k-1}$ and $\underline{x}_k$ to compute $\varphi_k^*(w_k)$ and $d_k^*(\underline{X}_1, \ldots, \underline{X}_{k-1}, \underline{x}_k)$ according to (3.3), (3.4) and (3.5). Then compute

$$\begin{aligned} & r_k(G_\alpha, d_k^* | \underline{X}_1, \ldots, \underline{X}_{k-1}) - r_k(G_\alpha, d_k^{G_\alpha}) \\ \equiv & \sum_{\underline{x}_k} d_k^*(\underline{X}_1, \ldots, \underline{X}_{k-1}, \underline{x}_k)[Q_k(\mu_2|w_k) - \theta_0] f(\underline{x}_k) \\ & \quad - \sum_{\underline{x}_k} d_k^{G_\alpha}(\underline{X}_1, \ldots, \underline{X}_{k-1}, \underline{x}_k)[Q_k(\mu_2|w_k) - \theta_0] f(\underline{x}_k) \\ \equiv & D_k(\underline{X}_1, \ldots, \underline{X}_{k-1}). \end{aligned}$$

The quantity $D_k(\underline{X}_1, \ldots, \underline{X}_{k-1})$ can be viewed as the difference between the conditional Bayes risk of the empirical Bayes selection rule d_k^* given $(\underline{X}_1, \ldots, \underline{X}_{k-1})$ and the minimum Bayes risk $r_k(G_\alpha, d_k^{G_\alpha})$.

(4) The process was repeated 500 times. The average of the differences based on the 500 repetitions denoted by $\overline{D}_k$ is used as an estimator of the difference $r_k(G_\alpha, d_k^*) - r_k(G_\alpha, d_k^{G_\alpha})$. We then use $k\overline{D}_k$ as an estimator of the overall difference $r(G_\alpha, d^*) - r(G_\alpha, d^{G_\alpha})$.

Tables 1–3 list simulation results on the small sample performance of the empirical Bayes selection rule d^* for selected values of the parameters k, m, α and N. $SE(\overline{D}_k)$ denotes the estimated standard error $\overline{D}_k$.

In each of the three tables, the results indicate that the estimated overall difference $k\overline{D}_k$ decreases as k becomes large. However, the rate of convergence to zero depends on the value of the parameter α. For $\alpha = 4$ we see that for $k \geq 50$, $\overline{D}_k$ is very close to zero. This indicates that the asymptotic optimality of the empirical Bayes selection rule d^* may also hold for moderate sample sizes.

Acknowledgments

This research was supported in part by the office of Naval Research Contract N00014-88-K-0170 and NSF Grants DMS-8702620, DMS-8923071 and DMS-8717799 at Purdue University. The authors are grateful to a referee for helpful suggestions to improve the presentation of the article.

Table 1. Small Sample Performance of d^*
$\alpha = 2$, $m = 5$, $N = 10$ and $\theta_0 = 0.3$

k	$\overline{D}_k$	$SE(\overline{D}_k)$	$k\overline{D}_k$
10	3.6898(−5)	5.883(−6)	36.898(−5)
20	0.7199(−5)	1.254(−6)	14.398(−5)
30	0.4678(−5)	0.595(−6)	14.034(−5)
40	0.2892(−5)	0.479(−6)	11.568(−5)
50	0.2297(−5)	0.430(−6)	11.485(−5)
60	0.1871(−5)	0.390(−6)	11.226(−5)
70	0.0595(−5)	0.223(−6)	4.165(−5)
80	0.0851(−5)	0.266(−6)	6.808(−5)
90	0.0595(−5)	0.223(−6)	5.355(−5)
100	0.0425(−5)	0.189(−6)	4.25(−5)

Table 2. Small Sample Performance of d^*
$\alpha = 3$, $m = 5$, $N = 15$ and $\theta_0 = 0.3$

k	D_k	$SE(\overline{D}_k)$	$k\overline{D}_k$
10	0.1980(−5)	0.483(−6)	1.980(−5)
20	0.0715(−5)	0.250(−6)	1.430(−5)
30	0.0043(−5)	0.008(−6)	0.129(−5)
40	0.0070(−5)	0.031(−6)	0.280(−5)
50	0.0014(−5)	0.004(−6)	0.070(−5)
60	0.0011(−5)	0.004(−6)	0.066(−5)
70	0.0007(−5)	0.003(−6)	0.049(−5)
80	0.0004(−5)	0.002(−6)	0.032(−5)
90	0.0004(−5)	0.002(−6)	0.036(−5)
100	0.0001(−5)	0.001(−6)	0.010(−5)

Table 3. Small Sample Performance of d^*
$\alpha = 4$, $m = 5$, $N = 15$ and $\theta_0 = 0.3$

k	$\overline{D}_k$	$SE(\overline{D}_k)$	$k\overline{D}_k$
10	0.0235(−5)	0.090(−6)	0.235(−5)
20	0.0036(−5)	0.017(−6)	0.072(−5)
30	0.0003(−5)	0.001(−6)	0.009(−5)
40	0.0010(−5)	0.010(−6)	0.040(−5)
50	0.	0.	0.
60	0.	0.	0.
70	0.	0.	0.
80	0.	0.	0.
90	0.	0.	0.
100	0.	0.	0.

An entry such as 3.6898 (-5) stands for 3.6898×10^{-5}.

REFERENCES

Alam, K., Mitra, A., Rizvi, M. H. and Saxena, K. M. L. (1986). Selection of the most diverse multinomial population. *American Journal of Mathematical and Management Sciences*, Vol. **6**, 65–86.

Dudewicz, E. J. and Van der Meulen, E. C. (1981). Selection procedures for the best binomial population with generalized entropy goodness. *Tamkang Journal of Mathematics*, **12**, 206–208.

Gini, C. (1912). Variabilita e Mutabilitá. *Studi Economico-aguridici della facotta di Giurisprudegza dell*, Universite di Cagliari, III, Part II.

Gupta, S. S. and Huang, D. Y. (1976). On subset selection procedures for the entropy function associated with the binomial populations. *Sankhyā*, **38A**, 153–173.

Gupta, S. S. and Leu, L. Y. (1990). Selecting the fairest of $k(\geq 2)$ m-sided dice. *Communications in Statistics - Theory and Methods*, **A(19)**, 2159–2177.

Gupta, S. S. and Wong, W. Y. (1975). Subset selection procedures for finite schemes in information theory. *Colloquia Mathematica Societatis János Bolya*, **16**, 279–291.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of American Statistical Association*, **58** 13–30.

Liang, T. and Panchapakesan, S. (1991). An empirical Bayes procedure for selecting the most homogeneous multinomial population according to the Gini-Simpson index. *Proceedings of the 1990 Taipei Symposium in Statistics June 28–30, 1990*, ed. by M. T. Chao and P. E. Cheng, Institute of Statistical Science, Academia Sinica, 447–460.

Rizvi, M. H., Alam, K. and Saxena, K. M. L. (1987). Selection procedure for multinomial populations with respect to diversity indices. *Contributions to the Theory and Application of Statistics*, A Volume in Honor of Herbert Solomon, ed. by Alan E. Gelfand, Academic Press, Inc. 485–509.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423, 623–656.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, **163**, 688.