EMPIRICAL BAYES APPROACH TO SELECTION PROBLEMS WITH SPECIAL REFERENCE TO MULTINOMIAL POPULATIONS

by

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

Technical Report #93-54C

Department of Statistics
Purdue University

November 1993

EMPIRICAL BAYES APPROACH TO SELECTION PROBLEMS WITH SPECIAL REFERENCE TO MULTINOMIAL POPULATIONS*

by

Shanti S. Gupta** and TaChen Liang

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 empirical Bayes methodology for ranking and selection problems with specific reference to multinomial populations. Two kinds of empirical Bayes selection rules are considered. One is to incorporate information from past data to improve the current decision. The other is to incorporate information from each other as to simultaneously improve the decision for each of the component problem under study. Certain important selection problems regarding multinomial populations, including the selection of the most probable event within a multinomial population, the selection of the most homogeneous population from among k multinomial populations, the selection of homogeneous populations compared with a control, are considered. The empirical Bayes methodology is discussed through these selection problems.

Key words and phrases: Asymptotically optimal, Bayes rule, empirical Bayes rule, entropy function, Gini-Simpson index, most probable event, population homogeneity, rate of convergence.

1. Introduction

*This research was supported in part by NSF Grant DMS–8923071 at Purdue University.

**Professor Gupta's research effort was also partly supported by Fachbereich Mathematik of the University of Darmstadt while he was an invited guest Professor there for two weeks in October 1993. Thanks are due to Professor Dr. Lehn for this invitation.

denote the ordered values of the parameters $\theta_1, \ldots, \theta_k$. The population associated with $\theta_{[k]}$ is called the best population. For a given standard θ_0, a population π_i is said to be good if $\theta_i \geq \theta_0$, and bad otherwise. In many practical situations, an experimenter may be interested in the selection of the best population and/or the selection of all good populations. These problems are known as selection and ranking problems. The formulation of selection and ranking procedures has been accomplished generally using the indiffrence zone approach (see Bechhofer (1954)) or the subset selection approach(see Gupta (1956, 1965)). A discussion of their differences and various modifications that have taken place since then can be found in Gupta and Panchapakesan (1979).

In many situations, an experimenter may have some prior information about the parameters of interest and he would like to use this information to make an appropriate decision. If the information at hand can be quantified into a single prior distribution, one would like to apply a Bayes procedure since it achieves the minimum of Bayes risks among a class of decision procedures. Some contributions to selection and ranking problems using a Bayesian approach have been made by Deely and Gupta (1968, 1988), Bickel and Yahav (1977), Chernoff and Yahav (1977), Goel and Rubin (1977), Gupta and Hsu (1978), Miescke (1979), Gupta and Miescke (1984), Gupta and Yang (1985), Berger and Deely (1988), and Fong (1989, 1990, 1992), among many others.

The empirical Bayes approach in statistical decision theory is typically appropriate when one is confronted repeatedly and independently with the same decision problem. In such instances, it is reasonable to formulate the component problem with respect to an unknown prior distribution on the parameter space. One then uses information borrowed from other sources to improve the decision procedure for each component. This approach is due to Robbins (1956, 1964). Empirical Bayes procedures have been derived for multiple decision problems by Deely (1965). Recently, Gupta and Hsiao (1983), Gupta and Liang (1986, 1988, 1989a, b, 1991, 1993a, b), Liang and Panchapakesan (1991) and Gupta and Hande (1993) have investigated empirical Bayes procedures for several selection problems. Many such empirical Bayes procedures have been shown to be asymptotically optimal in the sense that the component Bayes risk will converge to the optimal Bayes risk which would have been obtained if the prior distribution were fully known, and the Bayes procedure with respect to this prior distribution was used.

The present paper is concerned with the selection and ranking problem using the empirical Bayes approach. Two kinds of empirical Bayes procedures will be considered. One is to incorporate information from accumulated past data to improve the current decision. The other is to incorporate information from each other so as to simultaneously improve the decision for each of the component problems under study. The paper is organized in the following way. We briefly introduce the Bayes selection problem in Section 2. The empirical Bayes principle is addressed in Section 3. In the later part of the paper, special reference to multinomial populations is made. In Sections 4 and 5, we consider certain important selection problems for multinomial population(s). The empirical Bayes methodology is discussed through these selection problems.

2. Bayes Selection Problems and Procedures

Let $\theta_i \in \Theta \subset \mathbb{R}$ denote the unknown characteristic of interest associated with the population π_i, $i = 1, \ldots, k$. Let $X_1, \ldots, X_k$ be random variables representing the k populations $\pi_1, \ldots, \pi_k$, respectively, with X_i having the probability density function $f_i(x|\theta_i)$. It is assumed that given $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$, $\underset{\sim}{X} = (X_1, \ldots, X_k)$ have a joint probability density function $\underset{\sim}{f}(\underset{\sim}{x}|\underset{\sim}{\theta}) = \prod_{i=1}^{k} f_i(x_i|\theta_i)$, where $\underset{\sim}{x} = (x_1, \ldots, x_k)$. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ denote the ordered values of θ_i's. The population associated with $\theta_{[k]}$ is called a best population. For a given standard θ_0, a population π_i is said to be good if $\theta_i \geq \theta_0$ and bad otherwise. Let $\Omega = \{\underset{\sim}{\theta}|\theta_i \in \Theta,\ i = 1, \ldots, k\}$ denote the parameter space. Also, it is assumed that the value of the parameter θ_i is a realization of a random variable Θ_i having a prior distribution G_i and $\Theta_1, \ldots, \Theta_k$ are mutually independent. Hence $\underset{\sim}{\Theta} = (\Theta_1, \ldots, \Theta_k)$ have a joint prior distribution $\underset{\sim}{G}(\underset{\sim}{\theta}) = \prod_{i=1}^{k} G_i(\theta_i)$ on the parameter $\underset{\sim}{\theta}$ over the parameter space Ω.

In many situations, an experimenter is interested in identifying the best population or selecting the more promising subset of the k populations for further experimentation. For a specified selection goal, an action is a subset of the set $\{1, \ldots, k\}$. When action $S \subset \{1, \ldots, k\}$ is taken, it means that population π_i is included in the selected subset if $i \in S$. Let $\mathcal{A}$ denote the action space. For each $\underset{\sim}{\theta} \in \Omega$ and $S \in \mathcal{A}$, let $L(\underset{\sim}{\theta}, S)$ denote the loss incurred when $\underset{\sim}{\theta}$ is the true state of nature and the action S is taken. A decision procedure d is defined to be a mapping from $\mathcal{X} \times \mathcal{A}$ into [0,1] such that $\sum_{S \in \mathcal{A}} d(\underset{\sim}{x}, S) = 1$ for all $\underset{\sim}{x} \in \mathcal{X}$, where $\mathcal{X}$ is the sample space of $\underset{\sim}{X}$. $d(\underset{\sim}{x}, S)$ can be viewed as the probability of taking action S when $\underset{\sim}{X} = \underset{\sim}{x}$ is observed.

Let $\mathcal{D}$ be the class of all decision procedures. For each $d \in \mathcal{D}$, let $r(\underset{\sim}{G}, d)$ denote the associated Bayes risk. Then, $r(\underset{\sim}{G}) \equiv \inf_{d \in \mathcal{D}} r(\underset{\sim}{G}, d)$ is the minimum Bayes risk. An optimal decision procedure, denoted by $d_{\underset{\sim}{G}}$, is obtained if $d_{\underset{\sim}{G}}$ has the property that $r(\underset{\sim}{G}, d_{\underset{\sim}{G}}) = r(\underset{\sim}{G})$. Such a procedure is called Bayes with respect to $\underset{\sim}{G}$. Under some regularity conditions,

$$r(\underset{\sim}{G}, d) = \int_{\mathcal{X}} \sum_{S \in \mathcal{A}} d(\underset{\sim}{x}, S) \left[\int_{\Omega} L(\underset{\sim}{\theta}, S) d\underset{\sim}{G}(\underset{\sim}{\theta} | \underset{\sim}{x}) \right] \underset{\sim}{f}(\underset{\sim}{x}) d\underset{\sim}{x}$$

where $\underset{\sim}{G}(\underset{\sim}{\theta}|\underset{\sim}{x})$ is the joint posterior distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{X} = \underset{\sim}{x}$, $\underset{\sim}{f}(\underset{\sim}{x}) = \overset{k}{\underset{i=1}{\pi}} f_i(x_i)$ and $f_i(x_i) = \int_{\Theta} f_i(x_i|\theta_i) dG_i(\theta_i)$ is the marginal probability density function of X_i.

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

$$\Delta_{\underset{\sim}{G}}(\underset{\sim}{x}, S) = \int_{\Omega} L(\underset{\sim}{\theta}, S) d\underset{\sim}{G}(\underset{\sim}{\theta}|\underset{\sim}{x}),$$
$$A(\underset{\sim}{x}) = \{S \in \mathcal{A} | \Delta_{\underset{\sim}{G}}(\underset{\sim}{x}, S) = \min_{S' \in \mathcal{A}} \Delta_{\underset{\sim}{G}}(\underset{\sim}{x}, S')\}.$$

Then, the Bayes decision procedure $d_{\underset{\sim}{G}}$ clearly satisfies that $\sum_{S \in A(\underset{\sim}{x})} d_{\underset{\sim}{G}}(\underset{\sim}{x}, S) = 1$.

It should be noted that the Bayes decision procedures vary for different selection problems and goals, and depend on the loss function chosen. Also, the Bayes decision procedure is very sensitive to the prior distribution which is obtained through quantifying prior information into a single prior distribution.

3. Empirical Bayes Selection Procedures

In this section, we continue with the general setup of the early section. However, we assume only the existence of a prior distribution $\underset{\sim}{G}(\underset{\sim}{\theta}) = \overset{k}{\underset{i=1}{\pi}} G_i(\theta_i)$ on $\underset{\sim}{\theta}$ over Ω; the form of the prior distributions G_i, $i = 1, \ldots, k$, are either unknown or partially known.

We use the empirical Bayes approach. Two kinds of empirical Bayes procedures will be considered. One is to incorporate information from the accumulated past data to improve the current decision. The other is to incorporate information from each other so as to simultaneously improve the decision for each of the component decision problems.

Incorporating Information from Past Observations

According to the usual empirical Bayes framework, for each $i = 1, \dots, k$, let X_{ij} denote the random observation taken from π_i at stage j. Let Θ_{ij} denote the random characteristic of π_i at stage j. Given $\Theta_{ij} = \theta_{ij}, X_{ij}$ has the conditional probability density function $f_i(x|\theta_{ij})$. Let $\underset{\sim}{X}_j = (X_{1j}, \dots, X_{kj})$, and $\underset{\sim}{\Theta}_j = (\Theta_{1j}, \dots, \Theta_{kj})$. Suppose that independent observations $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n$ are available and $\underset{\sim}{\Theta}_j, j = 1, \dots, n$, are mutually independent and have the same prior distribution $\underset{\sim}{G}$, though $\underset{\sim}{\Theta}_j$ are not observable. Also, let $\underset{\sim}{X} = (X_1, \dots, X_k)$ denote the present random observation.

Consider an empirical Bayes decision procedure $d_n((\underset{\sim}{x}; \underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n), S) \equiv d_n(\underset{\sim}{x}, S)$, which is a function of the present observation $\underset{\sim}{x}$ and the past random observations $\underset{\sim}{X}_1, \dots,$ $\underset{\sim}{X}_n$. Let $r(\underset{\sim}{G}, d_n)$ be the conditional Bayes risk associated with the empirical Bayes procedure d_n, conditional on the past observations $(\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n)$. That is,

$$r(\underset{\sim}{G}, d_n) = \int_{\mathcal{X}} \sum_{S \in \mathcal{A}} d_n(\underset{\sim}{x}, S) \int_{\Omega} L(\underset{\sim}{\theta}, S) d\underset{\sim}{G}(\underset{\sim}{\theta}|\underset{\sim}{x}) \underset{\sim}{f}(\underset{\sim}{x}) d\underset{\sim}{x}.$$

Also, let $E[r(\underset{\sim}{G}, d_n)]$ be the overall Bayes risk of the empirical Bayes procedure d_n. That is,

$$E[r(\underset{\sim}{G}, d_n)] = \int_{\mathcal{X}} \sum_{S \in \mathcal{A}} E[d_n(\underset{\sim}{x}, S)] \int_{\Omega} L(\underset{\sim}{\theta}, S) d\underset{\sim}{G}(\underset{\sim}{\theta}|\underset{\sim}{x}) \underset{\sim}{f}(\underset{\sim}{x}) d\underset{\sim}{x},$$

where the expectation E is taken with respect to $(\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n)$. Note that $r(\underset{\sim}{G}, d_n) - r(\underset{\sim}{G}) \geq 0$ since $r(\underset{\sim}{G})$ is the minimum Bayes risk among the class of all decision procedures $\mathcal{D}$. Hence $E[r(\underset{\sim}{G}, d_n)] - r(\underset{\sim}{G}) \geq 0$. Either of the two non-negative difference can be used as a measure of optimality of the empirical Bayes procedure d_n. A sequence of empirical Bayes procedures $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal relative to the prior distribution $\underset{\sim}{G}$ if $E[r(\underset{\sim}{G}, d_n)] - r(\underset{\sim}{G}) \to 0$ as $n \to \infty$. The problem concerned here is to construct empirical Bayes procedures possessing the desired asymptotic optimality. Gupta and Liang (1986, 1988, 1989a, b), Gupta and Hande (1993) and Liang and Panchapakesan (1991) have investigated several empirical Bayes procedures for certain selection problems under this empirical Bayes framework.

Incorporating Information from Other Components

We now consider the case where it is assumed that the k prior distributions $G_1, \dots, G_k$ are identical, but there is no past observation available. Under this assumption, the empirical Bayes idea can still be employed. We may incorporate information from each of the k populations to make an appropriate decision for the concerned selection problem. Let d_k be a decision procedure constructed under such consideration (the detailed methods will be discussed later through some examples), and let $r(\underset{\sim}{G}, d_k)$ denote the corresponding Bayes risk. Since $r(\underset{\sim}{G})$ is the minimum Bayes risk, $r(\underset{\sim}{G}, d_k) - r(\underset{\sim}{G}) \geq 0$. An empirical Bayes procedure d_k is said to be asymptotically optimal if $r(\underset{\sim}{G}, d_k) - r(\underset{\sim}{G}) \to 0$ as $k \to \infty$. One may desire to construct empirical Bayes procedures having such asymptotic optimality. Gupta and Liang (1991, 1993a,b) have studied several empirical Bayes selection problems using this empirical Bayes approach.

Approaches for Constructing Empirical Bayes Procedures

There are three main approaches for constructing empirical Bayes procedures, according to how much we know about the prior distribution $\underset{\sim}{G}$, namely, nonparametric empirical Bayes, parametric empirical Bayes and hierarchical empirical Bayes, respectively.

For the nonparametric empirical Bayes approach, one assumes that the form of the prior distribution $\underset{\sim}{G}$ is completely unknown. In this situation, one may either use the information obtained from other sources (may be either from the past data or from the other components) to estimate the prior distribution $\underset{\sim}{G}$, then do a Bayesian analysis based on the estimated prior or represent the Bayes procedure in terms of the unknown prior, and then use the data to estimate the behavior of the Bayes decision procedure directly. Gupta and Hsiao (1983) and Gupta and Liang (1986, 1988, 1991) have studied some selection problems using the nonparametric empirical Bayes approach.

For the parametric empirical Bayes approach, it is assumed that the prior distribution $\underset{\sim}{G}$ is a member of some parameter family Γ and is indexed by some unknown parameter(s), say λ. Hence the prior distribution is denoted by $\underset{\sim}{G}_\lambda$. Suppose now an estimate $\hat{\lambda}$ depending on the data can be found and we denote the prior distribution associated with $\hat{\lambda}$ by $\underset{\sim}{G}_{\hat{\lambda}}$. Note that $\underset{\sim}{G}_{\hat{\lambda}}$ is also a member of the family Γ. We use $\underset{\sim}{G}_{\hat{\lambda}}$ to estimate the unknown prior $\underset{\sim}{G}_\lambda$. We then follow the usual Bayesian analysis and derive the Bayes procedure $d_{\underset{\sim}{G}_{\hat{\lambda}}}$ with respect to the estimated prior distribution $\underset{\sim}{G}_{\hat{\lambda}}$. Using this line of parametric empirical Bayes approach, Gupta and Liang (1989a,b) have studied empirical Bayes selection procedure for selecting the most probable event in a multinomial distribution and for selecting the best population from among k binomial populations.

For the hierarchical empirical Bayes approach, it is assumed that the prior distribution of component i belongs to some parameter family Γ and is indexed by a parameter (or parameters) λ_i and the λ_i's are assumed to be iid, follow a hierarchical prior distribution. This hierarchical prior distribution may be either known or indexed by an unknown parameter (or parameters). In the latter case, the unknown parameter(s) should be estimated. One then follows a hierarchical Bayesian analysis. A decision procedure derived through this framework is called a hierarchical empirical Bayes procedure. Gupta and Liang (1993a) have investigated a hierarchical empirical Bayes selection procedure for sampling inspection.

4. Selecting the Most Probable Multinomial Event

Consider a multinomial population with $k(\geq 2)$ cells, where the cell π_i has probability p_i, $i = 1, \ldots, k$. Let $p_{[1]}, \leq \cdots \leq p_{[k]}$ denote the ordered values of the p_i's. It is assumed that the exact pairing between the ordered and the unordered parameters is unknown. Any cell associated with $p_{[k]}$ is considered as the most probable event. In the literature, a number of statistical procedures have been considered for selecting the most probable event. For example, see Bechhofer, Elmaghraby and Morse (1959), Gupta and Nagel (1967), Panchapakesan (1971), Cacoullos and Sobel (1966), Alam (1971), Ramey and Alam (1979, 1980), Bechhofer and Kulkarni (1984) and Chen (1986, 1988), among many others. In the following, our goal is to derive empirical Bayes rules to select the most probable event.

A Bayes Selection Rule

For each $i = 1, \ldots, k$, 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 a multinomial distribution with the probability function

$$f(\underset{\sim}{x}|\underset{\sim}{p}) = \frac{N!}{\prod_{j=1}^{k}(x_j!)} \prod_{j=1}^{k} p_j^{x_j}, \quad x_j = 0, 1, \ldots, N \quad \text{and} \quad \sum_{j=1}^{k} x_j = N. \tag{4.1}$$

Let $\Omega = \{\underset{\sim}{p} | 0 < p_i < 1,\ i = 1, \ldots, k,\ \sum_{j=1}^{k} p_j = 1\}$ be the parameter space. It is assumed that $\underset{\sim}{p}$ is a realization of a random parameter vector $\underset{\sim}{P} = (P_1, \ldots, P_k)$, which has a Dirichlet prior distribution G with hyperparameters $\underset{\sim}{\alpha} = (\alpha_1, \ldots, \alpha_k)$ and is denoted by $D_k(\alpha_1, \ldots, \alpha_k)$, where all α_i's are positive but unknown. That is, $\underset{\sim}{P}$ has a probability density of the form

$$g(\underset{\sim}{p}) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} p_i^{\alpha_i - 1}, \quad 0 < p_i < 1,\ \sum_{j=1}^{k} p_j = 1 \tag{4.2}$$

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

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

$$L(\underset{\sim}{p}, i) = p_{[k]} - p_i, \tag{4.3}$$

the difference between the most probable and the selected event.

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

For decision rule $\underset{\sim}{d}$, let $r(G, \underset{\sim}{d})$ denote the corresponding Bayes risk. Straightforward computation yields that

$$r(G, \underset{\sim}{d}) = C - \sum_{\underset{\sim}{x} \in \mathcal{X}} d_i(\underset{\sim}{x}) \varphi_i(\underset{\sim}{x}) f(\underset{\sim}{x}) \tag{4.4}$$

where

$$\varphi_i(\underset{\sim}{x}) = E[P_i | \underset{\sim}{X} = \underset{\sim}{x}] = \frac{x_i + \alpha_i}{N + \alpha_0},$$

$$f(\underset{\sim}{x}) = \int_\Omega f(\underset{\sim}{x} | \underset{\sim}{p}) dG(\underset{\sim}{p}): \text{ the marginal probability function of } \underset{\sim}{X},$$

and

$$C = \int_\Omega p_{[k]} dG(\underset{\sim}{p}).$$

From (4.4), a Bayes selection rule $\underset{\sim}{d}_G = (d_{G1}, \ldots, d_{Gk})$ can be obtained as follows. For each $\underset{\sim}{x} \in \mathcal{X}$, let

$$A(\underset{\sim}{x}) = \{i | x_i + \alpha_i = \max_{1 \le j \le k} (x_j + \alpha_j)\}. \tag{4.5}$$

Then for each $i = 1, \ldots, k$, define

$$d_{Gi}(\underset{\sim}{x}) = \begin{cases} |A(\underset{\sim}{x})|^{-1} & \text{if } i \in A(\underset{\sim}{x}), \\ 0 & \text{otherwise,} \end{cases} \tag{4.6}$$

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

Note that the Bayes selection rule $\underset{\sim}{d}_G$ defined through (4.6) is a randomized rule. A non-randomized Bayes selection rule, say $\underset{\sim}{d}^*_G = (d^*_{G1}, \ldots, d^*_{Gk})$, is also given below. Let $i^* = i^*(\underset{\sim}{x}) = \min A(\underset{\sim}{x})$. Then for each $i = 1, \ldots, k$, let

$$d^*_{Gi}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } i = i^*, \\ 0 & \text{otherwise.} \end{cases} \tag{4.7}$$

The minimum Bayes risk for the concerned selection problem is $r(G, \underset{\sim}{d}_G)(= r(G, \underset{\sim}{d}^*_G))$.

Since the values of the hyperparameters $(\alpha_1, \ldots, \alpha_k)$ are unknown, it is not possible to apply the Bayes selection rules for the selection problem at hand. In the following, empirical Bayes approach is employed.

Incorporating Information from Past Observations

According to the usual empirical Bayes framework, it is assumed that there are marginally iid random observations $\underset{\sim}{X}_j = (X_{j1}, \ldots, X_{jk})$, $j = 1, \ldots, n$, with marginal probability function $f(\underset{\sim}{x})$ available when the current decision is made. We also let $\underset{\sim}{X}_{n+1} = \underset{\sim}{X} = (X_1, \ldots, X_k)$ denote the present observation. Empirical Bayes rules are proposed depending on whether the value of the parameter α_0 is known or unknown. Note that $\alpha_0 = \sum_{j=1}^{k} \alpha_j$. In the case where α_0 is known, the individual values of α_i, $i = 1, \ldots, k$, are still unknown.

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

$$\overline{X}_i(n) = \frac{1}{n} \sum_{j=1}^{n} X_{ij}, \; M_i(n) = \frac{1}{n} \sum_{j=1}^{n} X_{ij}^2,$$
$$Z_i(n) = [N\overline{X}_i(n) - M_i(n)]\overline{X}_i(n),$$
$$Y_i(n) = [M_i(n) - \overline{X}_i(n)]N - (N-1)(\overline{X}_i(n))^2.$$

α_0 Known Case

Let $\hat{\alpha}_{in} = \alpha_0 \overline{X}_i(n) N^{-1}$ and let $A_n(\underset{\sim}{x}) = \{i | x_i + \hat{\alpha}_{in} = \max_{1 \le j \le k}(x_j + \hat{\alpha}_{jn})\}$. Then define empirical Bayes selection rule $\underset{\sim}{d}_n = (d_{n1}, \ldots, d_{nk})$ and $\underset{\sim}{d}_n^* = (d_{n1}^*, \ldots, d_{nk}^*)$, respectively as follows:

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

$$d_{ni}(\underset{\sim}{x}) = \begin{cases} |A_n(\underset{\sim}{x})|^{-1} & \text{if } i \in A_n(\underset{\sim}{x}) \\ 0 & \text{otherwise,} \end{cases} \tag{4.8}$$

$$d_{ni}^*(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } i = \min A_n(\underset{\sim}{x}) \\ 0 & \text{otherwise.} \end{cases} \tag{4.9}$$

Note that $\underset{\sim}{d}_n$ is a randomized rule while $\underset{\sim}{d}_n^*$ is a nonrandomized rule. The selection rule $\underset{\sim}{d}_n$ has been considered by Gupta and Liang (1989b).

α_0 Unknown Case

Let $\mu_i = E[\overline{X}_i(n)]$ and $\lambda_i = E[M_i(n)]$. Then, a direct computation shows that $u_i = N\alpha_i\alpha_0^{-1}$, $\lambda_i = N\alpha_i\alpha_0^{-1} + (N^2 - N)\alpha_i(\alpha_i + 1)\alpha_0^{-1}(\alpha_0 + 1)^{-1}$. Hence, $\alpha_i = L_{i1}L_{i2}^{-1}$, where $L_{i1} = (N\mu_i - \lambda_i)$, $L_{i2} = (\lambda_i - \mu_i)N - (N-1)\mu_i^2$. Thus, $Z_i(n)$, $Y_i(n)$ and $Z_i(n)/Y_i(n)$ are moment estimators of L_{i1}, L_{i2} and α_i, respectively. Note that L_{i1} and L_{i2} are both positive. Also, $Z_i(n) \ge 0$. However, it is possible that $Y_i(n) \le 0$. So first define

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

and let

$$\tilde{A}_n(\underset{\sim}{x}) = \{i | \Delta_{in}(x_i) = \max_{1 \le j \le k} \Delta_{jn}(x_j)\}.$$

Then define empirical Bayes selection rules $\underset{\sim}{\tilde{d}}_n = (\tilde{d}_{n1}, \ldots, \tilde{d}_{nk})$ and $\underset{\sim}{\tilde{d}}_n^* = (\tilde{d}_{n1}^*, \ldots, \tilde{d}_{nk}^*)$ as follows:

For each $i = 1, \ldots, k$ and $\underset{\sim}{x} \in \mathcal{X}$, define

$$\tilde{d}_{ni}(\underset{\sim}{x}) = \begin{cases} |\tilde{A}_n(\underset{\sim}{x})|^{-1} & \text{if } i \in A_n(\underset{\sim}{x}), \\ 0 & \text{otherwise,} \end{cases} \tag{4.10}$$

and

$$\tilde{d}_{ni}^*(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } i = \min \tilde{A}_n(\underset{\sim}{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

The selection mle $\tilde{\underset{\sim}{d}}_n$ has been studied by Gupta and Liang (1989b).

Asymptotic Optimality

Let $r(G, \underset{\sim}{d}_n)$ be the Bayes risk associated with the empirical Bayes selection rule $\underset{\sim}{d}_n$. Since $r(G, \underset{\sim}{d}_G)$ is the minimum Bayes risk, $r(G, \underset{\sim}{d}_n) - r(G, \underset{\sim}{d}_G) \geq 0$ for all n. This regret risk is always used as a measure of performance of the selection rule $\underset{\sim}{d}_n$.

Definition 4.1. A sequence of empirical Bayes rules $\{\underset{\sim}{d}_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal at least of order c_n relative to the prior distribution G if $r(G, \underset{\sim}{d}_n) - r(G, \underset{\sim}{d}_G) = O(c_n)$, where $\{c_n\}$ is a sequence of positive numbers such that $\lim_{n \to \infty} c_n = 0$.

Gupta and Liang (1989b) have studied the two empirical Bayes selection rules $\underset{\sim}{d}_n$ and $\tilde{\underset{\sim}{d}}_n$, defined in (4.8) and (4.10), respectively, and established the corresponding asymptotic optimality. Along a similar line, the asymptotic optimality of $\underset{\sim}{d}_n^*$ and $\tilde{\underset{\sim}{d}}_n^*$ can also be established.

Theorem 4.1. (Gupta and Liang (1989b)). For the empirical Bayes selection rules $\underset{\sim}{d}_n$, $\underset{\sim}{d}_n^*$, $\tilde{\underset{\sim}{d}}_n$ and $\tilde{\underset{\sim}{d}}_n^*$ defined previously we have,

$$
\begin{aligned}
r(G, \underset{\sim}{d}_n) - r(G, \underset{\sim}{d}_G) &= O(\exp(-c_1 n)), \\
r(G, \underset{\sim}{d}_n^*) - r(G, \underset{\sim}{d}_G) &= O(\exp(-c_2 n)) \\
r(G, \tilde{\underset{\sim}{d}}_n) - r(G, \underset{\sim}{d}_G) &= O(\exp(-c_3 n)) \\
r(G, \tilde{\underset{\sim}{d}}_n) - r(G, \underset{\sim}{d}_G) &= O(\exp(-c_4 n))
\end{aligned}
$$

for some positive constants c_i, $i = 1, 2, 3, 4$.

Remark 4.1. Another selection problem related to the multinomial population is to select the least probable event; that is, to select the cell associated with $p_{[1]}$. By using the loss $L(\underset{\sim}{p}, i) = p_i - p_{[1]}$, Gupta and Liang (1989b) have studied two empirical Bayes selection rules according to whether α_0 is known or unknown, and establish the associated asymptotic optimality.

Remark 4.2. Gupta and Hande (1993) have generalized the result of Gupta and Liang (1989b) for a more general loss and established the exponential rate of convergence for their proposed empirical Bayes rule. They also considered the problem of selecting the most (least) probable event and simultaneously estimating the probability associated with the selected cell in a decision-theoretic framework. A sequence of empirical Bayes rules is constructed and is proved to be asymptotically optimal of order $O(n^{-1})$.

5. Selection from Several Multinomial Populations

The multinomial distribution provides a model for studying the diversity within a population which is categorized into several cells according to a qualitative characteristic. Such studies arise in ecology, sociology, genetics, economics and other disciplines. Diversity in ecological contexts has been discussed by Pielou (1975) and Patil and Taillie (1982). There are two measure of diversity of multinomial population which have been commonly used. These are Shannon's entropy function and Gini-Simpson index.

Selection from several multinomial populations has been earlier discussed by a few authors. Gupta and Wong (1975) have considered selection in terms of the Shannon entropy function using the subset selection approach. Alam, Mitra, Rizvi and Lal Saxena (1986) have considered selection in terms of the entropy function as well as the Gini-Simpson index using the indifference zone approach. Rizvi, Alam and Lal Saxena (1987) also considered a subset selection rule based on certain other diversity indexes. Recently, Gupta and Leu (1990), Liang and Panchapakesan (1991) and Gupta and Liang (1993b) have studied certain selection rules based on Gini-Simpson index.

Let $\pi_1, \ldots, \pi_k$ be k multinomial populations, each having m cells with associated probability vector $\underset{\sim}{p}_i = (p_{i1}, \ldots, p_{im})$, $i = 1, \ldots, k$. Define

$$\theta_i = \sum_{j=1}^{m} (p_{ij} - \frac{1}{m})^2 = \sum_{j=1}^{m} p_{ij}^2 - \frac{1}{m}, \tag{5.1}$$

which is essentially equivalent to Gini-Simpson index. θ_i is used as a measure of homogeneity of population π_i. Note that $0 \leq \theta_i \leq 1 - \frac{1}{m}$. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ denote the ordered values of θ_i's. The population associated with $\theta_{[1]}$ is called the most homogeneous population. For a given constant θ_0, $0 < \theta_0 < 1 - \frac{1}{m}$, population π_i is said to be homogeneous if $\theta_i \leq \theta_0$ and nonhomogeneous if $\theta_i > \theta_0$. Our goal is to develop selection rules to select the most homogeneous population and/or to select all homogeneous populations.

5.1. Selecting the Most Homogeneous Multinomial Population

For each $i = 1, \ldots, k$, let $\underset{\sim}{X}_i = (X_{i1}, \ldots, X_{im})$ be the cell frequencies in a sample of N_i trials from population π_i. Then, given $\underset{\sim}{p}_i$, $\underset{\sim}{X}_i$ has the probability function

$$f_i(\underset{\sim}{x}_i|\underset{\sim}{p}_i) = \frac{N_i!}{\prod_{j=1}^{m} x_{ij}!} \prod_{j=1}^{m} p_{ij}^{x_{ij}} \tag{5.2}$$

where x_{ij}, $j = 1, \ldots, m$, are nonnegative integers such that $\sum_{j=1}^{m} x_{ij} = N_i$. Given $\underset{\sim}{p}_i$, $\underset{\sim}{X}_i$, $i = 1, \ldots, k$, are assumed to be conditionally independent. We also assume that $\underset{\sim}{p}_i$ is a realization of a random vector $\underset{\sim}{P}_i = (P_{i1}, \ldots, P_{im})$ which has a Dirichlet prior distribution $D_m(\alpha_{i1}, \ldots, \alpha_{im})$ with probability density given by

$$g_i(\underset{\sim}{p}_i) = \frac{\Gamma(\alpha_i)}{\prod_{j=1}^{m} \Gamma(\alpha_{ij})} \prod_{j=1}^{m} p_{ij}^{\alpha_{ij}-1}, \qquad \alpha_{ij} > 0 \quad \text{and} \quad \alpha_i = \sum_{j=1}^{m} \alpha_{ij}. \tag{5.3}$$

and $\underset{\sim}{P}_1, \ldots, \underset{\sim}{P}_k$ are mutually independent. Let $G(\underset{\sim}{p})$ denote the joint distribution of $\underset{\sim}{P}_1, \ldots, \underset{\sim}{P}_k$, where $\underset{\sim}{p} = (\underset{\sim}{p}_1, \ldots, \underset{\sim}{p}_k)$. Let $\mathcal{A} = \{i | i = 1, \ldots, k\}$ be the action space. Action i corresponds to selecting π_i as the most homogeneous population. For parameter $\underset{\sim}{p}$ and action i, the loss function $L(\underset{\sim}{p}, i)$ is defined as

$$L(\underset{\sim}{p}, i) = \theta_i - \theta_{[1]}. \tag{5.4}$$

A selection rule $\underset{\sim}{d} = (d_1, \ldots, d_k)$ is defined as a mapping from $\mathcal{X}$, the sample space of $\underset{\sim}{X} = (\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_k)$, into $[0,1]^k$ such that $0 \leq d_i(\underset{\sim}{x}) \leq 1$, $i = 1, \ldots, k$, and $\sum_{i=1}^{k} d_i(\underset{\sim}{x}) = 1$ for each $\underset{\sim}{x} \in \mathcal{X}$. That is, $d_i(\underset{\sim}{x})$ is the probability of selecting π_i as the most homogeneous population given $\underset{\sim}{X} = \underset{\sim}{x}$.

A Bayes Selection Rule

The Bayes risk $r(G, \underset{\sim}{d})$ of selection rule $\underset{\sim}{d}$ can be written as

$$r(G, \underset{\sim}{d}) = \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i=1}^{k} d_i(\underset{\sim}{x}) \varphi_i(\underset{\sim}{x}_i) f(\underset{\sim}{x}) - \int_{\Omega} \theta_{[1]} dG(\underset{\sim}{p}) - \frac{1}{m}, \tag{5.5}$$

where

$$\varphi_i(\underset{\sim}{x}_i) = E\left[\sum_{j=1}^{m} P_{ij}^2 | \underset{\sim}{X}_i = \underset{\sim}{x}_i\right] : \text{ the posterior expectation of } \sum_{j=1}^{m} P_{ij}^2 \text{ given } \underset{\sim}{X}_i = \underset{\sim}{x}_i,$$

$$f(\underset{\sim}{x}) = \prod_{i=1}^{k} f_i(\underset{\sim}{x}_i),$$

$$f_i(\underset{\sim}{x}_i) = \int f_i(\underset{\sim}{x}_i|\underset{\sim}{p})g_i(\underset{\sim}{p})d\underset{\sim}{p}: \text{ the marginal probability function of } \underset{\sim}{X}_i,$$

and $\Omega = \{\underset{\sim}{p}|\underset{\sim}{p} = (\underset{\sim}{p}_1, \ldots, \underset{\sim}{p}_k)\}$ is the parameter space.

Let

$$I(\underset{\sim}{x}) = \{i|\varphi_i(\underset{\sim}{x}_i) = \min_{1\le j\le k} \varphi_j(\underset{\sim}{x}_j)\}. \tag{5.6}$$

Then a randomized Bayes selection rule $\underset{\sim}{d}_G = (d_{G1}, \ldots, d_{Gk})$, which minimizes $r(G, \underset{\sim}{d})$ among the class of all selection rules, is given by

$$d_{Gi}(\underset{\sim}{x}) = \begin{cases} |I(\underset{\sim}{x})|^{-1} & \text{if } i \in I(\underset{\sim}{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{5.7}$$

A nonrandomized Bayes selection rule $\underset{\sim}{d}_G^* = (d_{G1}^*, \ldots, d_{Gk}^*)$ is given by

$$d_{Gi}^*(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } i = \min I(\underset{\sim}{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{5.8}$$

The minimum Bayes risk is $r(G, \underset{\sim}{d}_G)$.

Expressions for $\varphi_i(\underset{\sim}{x}_i)$

It is well known that the posterior distribution of $\underset{\sim}{P}_i$ given that $\underset{\sim}{X}_i = \underset{\sim}{x}_i$ is the Dirichlet distribution $D_m(x_{i1} + \alpha_{i1}, \ldots, x_{im} + \alpha_{im})$. Thus a straightforward computation yields that

$$\varphi_i(\underset{\sim}{x}_i) = \frac{(N_i + \alpha_i) + \sum_{j=1}^{m}(x_{ij} + \alpha_{ij})^2}{(N_i + \alpha_i + 1)(N_i + \alpha_i)}. \tag{5.9}$$

Also, let $\mu_{ij} = E(X_{ij})$, $\lambda_{ij} = E(X_{ij}^2)$. Then, after computation, we obtain

$$\mu_{ij} = N_i\alpha_{ij}/\alpha_i, \tag{5.10}$$

$$\lambda_{ij} = \mu_{ij} + N_i(N_i - 1)\alpha_{ij}(\alpha_{ij} + 1)(\alpha_i + 1)^{-1}\alpha_i^{-1}, \tag{5.11}$$

and hence,

$$\alpha_i\{\lambda_{ij} - \mu_{ij} - N_i^{-1}(N_i - 1)\mu_{ij}^2\} = N_i\mu_{ij} - \lambda_{ij} > 0, \tag{5.12}$$

the positivity being verified from (5.10) and (5.11). Summing both sides of (5.12) and noting that $\sum_{j=1}^{m} \mu_{ij} = N_i$, we obtain

$$\alpha_i = B_i/A_i, \tag{5.13}$$

where

$$\begin{cases} A_i = N_i \sum_{j=1}^{m} \lambda_{ij} - N_i^2 - (N_i - 1)\sum_{j=1}^{m} \mu_{ij}^2 > 0, \\ B_i = N_i^3 - N_i \sum_{j=1}^{m} \lambda_{ij} > 0. \end{cases} \tag{5.14}$$

Now, use (5.10) and (5.13) to substitute for α_{ij} and α_i in (5.9). This yields

$$\varphi_i(\underset{\sim}{x}_i) \equiv \frac{D_i(\underset{\sim}{x}_i)}{C_i} \tag{5.15}$$

where

$$\begin{cases} D_i(\underset{\sim}{x}_i) = (A_iB_i + N_iA_i^2)^2 + \sum_{j=1}^{m}(A_ix_{ij} + N_i^{-1}B_i\mu_{ij})^2, \\ C_i = (B_i + N_iA_i + A_i)(B_i + N_iA_i). \end{cases} \tag{5.16}$$

Incorporating Information from Past Observations

It is now assumed that the hyperparameters α_{ij} of the Dirichlet priors are unknown and certain past observations are available. Thus, empirical Bayes approach is employed. For each $i = 1, \ldots, k$, let $\underset{\sim}{X}_{i\ell} = (X_{i1\ell}, \ldots, X_{im\ell})$, $\ell = 1, \ldots, n$, denote the past data from π_i. Assume that $\underset{\sim}{X}_{i\ell}, \ell = 1, \ldots, n$, are iid with marginal probability function $f_i(\underset{\sim}{x}_i)$ and $\underset{\sim}{X}_{i\ell}$, $\ell = 1, \ldots, n$, $i = 1, \ldots, k$ are mutually independent. Let

$$\overline{X}_{ij}(n) = \frac{1}{n}\sum_{\ell=1}^{n} X_{ij\ell},\ M_{ij}(n) = \frac{1}{n}\sum_{\ell=1}^{n} X_{ij\ell}^2,\ M_i(n) = \sum_{j=1}^{m} M_{ij}(n)$$

$$A_{in} = N_iM_i(n) - N_i^2 - (N_i - 1)\sum_{j=1}^{m}[\overline{X}_{ij}(n)]^2$$

$$A_{in}^{+} = \max(0, A_{in})$$

$$B_{in} = N_i^3 - N_iM_i(n)$$

Then, $A_{in} \longrightarrow A_i$, $A_{in}^{+} \longrightarrow A_i$ and $B_{in} \longrightarrow B_i$ almost everywhere. Finally, define for each $i = 1, \ldots, k$,

$$C_{in} = (B_{in} + A_{in}^{+} N_i)(B_{in} + A_{in}^{+} N_i + A_{in}^{+})$$
$$D_{in}(\underset{\sim}{x}_i) = (A_{in}^{+} B_{in} + N_i A_{in}^{+2})^2 + \sum_{j=1}^{m} \{A_{in}^{+} x_{ij} + N_i^{-1} B_{in} \overline{X}_{ij}(n)\}^2$$
$$\varphi_{in}(\underset{\sim}{x}_i) = D_{in}(\underset{\sim}{x}_i)/C_{in}.$$

Liang and Panchapakesan (1991) proposed an empirical Bayes selection rule $\underset{\sim}{d}_n^* = (d_{n1}^*, \ldots, d_{nk}^*)$ defined as follows.

$$d_{ni}^*(\underset{\sim}{x}) = \begin{cases} |I_n(\underset{\sim}{x})|^{-1} & \text{if } i \in I_n(\underset{\sim}{x}), \\ 0 & \text{otherwise}, \end{cases} \tag{5.17}$$

where

$$I_n(\underset{\sim}{x}) = \{i | \varphi_{in}(\underset{\sim}{x}_i) = \min_{1 \le j \le k} \varphi_{jn}(\underset{\sim}{x}_j)\}. \tag{5.18}$$

Note that $\underset{\sim}{d}_n^*$ is a randomized selection rule. A nonrandomized selection rule, say $\underset{\sim}{d}_n^{**} = (d_{n1}^{**}, \ldots, d_{nk}^{**})$ can be obtained as follows.

$$d_{ni}^{**}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } i = \min I_n(\underset{\sim}{x}), \\ 0 & \text{otherwise}. \end{cases} \tag{5.19}$$

Asymptotic Optimality

Liang and Panchapakesan (1991) have studied the empirical Bayes selection rule $\underset{\sim}{d}_n^*$ and established the corresponding asymptotic optimality. The asymptotic optimality of the empirical Bayes selection rule $\underset{\sim}{d}_n^{**}$ can also be established in a similar way. We state the results as follows.

Theorem 5.1. (Liang and Panchapakesan (1991)) Let $\{\underset{\sim}{d}_n^*\}$ and $\{\underset{\sim}{d}_n^{**}\}$ be sequences of empirical Bayes selection rules defined through (5.17) and (5.19), respectively. Then

$$r(G, \underset{\sim}{d}_n^*) - r(G, \underset{\sim}{d}_G) = O(\exp(-c_5 n)),$$
$$r(G, \underset{\sim}{d}_n^{**}) - r(G, \underset{\sim}{d}_G) = O(\exp(-c_6 n)).$$

for some positive constants c_5 and c_6.

5.2. Selecting Homogeneous Multinomial Populations

In this section, we discuss the problem of selecting homogeneous multinomial populations compared with a standard θ_0. Let $\mathcal{A} = \{s | s \subset \{1, 2, \ldots, k\}\}$ be the action space. When action s is taken, it means that population π_i is selected as a homogeneous population if $i \in s$ and excluded as a non-homogeneous population if $i \notin S$. For parameter $\underset{\sim}{p}$ and action s, the loss function is defined to be

$$L(\underset{\sim}{p}, s) = \sum_{i \in S} (\theta_i - \theta_0) I_{(\theta_0, 1 - \frac{1}{m})}(\theta_i) + \sum_{i \notin S} (\theta_0 - \theta_i) I_{[0, \theta_0]}(\theta_i). \tag{5.20}$$

A selection rule $\underset{\sim}{d} = (d_1, \ldots, d_k)$ is defined to be a mapping from the sample space $\mathcal{X}$ into the product space $[0,1]^k$, such that for each $\underset{\sim}{x} \in \mathcal{X}$, $d_i(\underset{\sim}{x})$ is the probability of selecting population π_i as a homogeneous population. In the following, it is assumed that the random probability vectors $\underset{\sim}{P}_1, \ldots, \underset{\sim}{P}_k$ are mutually independent, with a common but unknown Dirichlet prior distribution $D_m(\beta_1, \ldots, \beta_m)$.

For selection rule $\underset{\sim}{d}$, the associated Bayes risk $r(G, \underset{\sim}{d})$ is given by

$$r(G, \underset{\sim}{d}) = \sum_{i=1}^{k} r_i(G, d_i), \tag{5.21}$$

where

$$\begin{aligned} r_i(G, d_i) = &\sum_{\underset{\sim}{x} \in X} d_i(\underset{\sim}{x}) \left[\sum_{j=1}^{m} E[P_{ij}^2 | \underset{\sim}{x}_i] - \frac{1}{m} - \theta_0 \right] f(\underset{\sim}{x}) \\ &+ \int (\theta_0 - \theta_i) I_{(0, \theta_0)}(\theta_i) g(\underset{\sim}{p}_i) d\underset{\sim}{p}_i. \end{aligned} \tag{5.22}$$

Thus, A Bayes selection rule $\underset{\sim}{d}_G = (d_{G1}, \ldots, d_{Gk})$ is obtained as follows:

$$d_{Gi}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \sum_{j=1}^{m} E[P_{ij}^2 | \underset{\sim}{X}_i = \underset{\sim}{x}_i] \leq \frac{1}{m} + \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.23}$$

The minimum Bayes risk is $r(G, \underset{\sim}{d}_G) = \sum_{i=1}^{k} r_i(G, d_{Gi})$. Also, as seen from the preceding section,

$$E\left[\sum_{j=1}^{m} P_{ij}^2 | \underset{\sim}{X}_i = \underset{\sim}{x}_i \right] = \frac{D(\underset{\sim}{x}_i)}{C} \tag{5.24}$$

where

$$
\begin{cases}
C = (B + NA + A)(B + NA) \\
D(\underset{\sim}{x}_i) = (AB + NA^2)^2 + \sum_{j=1}^{m} (Ax_{ij} + N^{-1}B\mu_j)^2 \\
A = N\sum_{j=1}^{m} \lambda_j - N^2 - (N-1)\sum_{j=1}^{m} \mu_j^2 > 0, \\
B = N^3 - N\sum_{j=1}^{m} \lambda_j > 0, \\
\mu_j = E[X_{1j}] \text{ and } \lambda_j = E[X_{1j}^1].
\end{cases}
\tag{5.25}
$$

Incorporating Information from Among k Components

We assume that the hyperparameters of the Dirichlet prior distribution $D_m(\tau_1, \ldots, \tau_m)$ are unknown. In this situation, it is not possible to apply the Bayes selection procedure $\underset{\sim}{d}_G$. By adopting the idea of empirical Bayes, we incorporate information from each of the k multinomial populations and develop empirical Bayes selection rules.

For each $j = 1, \ldots, m$, let $\hat{\mu}_j = \frac{1}{k}\sum_{i=1}^{k} X_{ij}$, $\hat{\lambda}_j = \frac{1}{k}\sum_{i=1}^{k} X_{ij}^2$. Then, $E\hat{\mu}_j = \mu_j$ and $E\hat{\lambda}_j = \lambda_j$. We use $\hat{\mu}_j$ and $\hat{\lambda}_j$ to estimate μ_j and λ_j, respectively. In (5.25), replacing μ_j and λ_j by the corresponding estimators $\hat{\mu}_j$ and $\hat{\lambda}_j$, respectively, we obtain the following estimators.

$$
\begin{cases}
\hat{A} = N\sum_{j=1}^{m} \hat{\lambda}_j - N^2 - (N-1)\sum_{j=1}^{m} \hat{\mu}_j^2, \\
\hat{B} = N^3 - N\sum_{j=1}^{m} \hat{\lambda}_j, \\
\hat{A}^+ = \max(0, \hat{A}), \\
\hat{C} = (\hat{B} + (N+1)\hat{A}^+)(\hat{B} + N\hat{A}^+), \\
\hat{D}(\underset{\sim}{x}_i) = (\hat{A}^+\hat{B} + N\hat{A}^{+2})^2 + \sum_{j=1}^{m} (\hat{A}^+ x_{ij} + N^{-1}\hat{B}\hat{\mu}_j)^2.
\end{cases}
\tag{5.26}
$$

Now, define

$$
\hat{\varphi}_i(\underset{\sim}{x}_i) = \frac{\hat{D}(\underset{\sim}{x}_i)}{\hat{C}}.
\tag{5.27}
$$

We then obtain an empirical Baye selection rule $\underset{\sim}{\hat{d}} = (\hat{d}_1, \ldots, \hat{d}_k)$ given as follows: For each $i = 1, \ldots, k$, and $\underset{\approx}{x} \in \mathcal{X}$,

$$
\hat{d}_i(\underset{\approx}{x}) = \begin{cases} 1 & \text{if } \hat{\varphi}_i(\underset{\sim}{x}_i) \le \frac{1}{m} + \theta_0, \\ 0 & \text{otherwise.} \end{cases}
\tag{5.28}
$$

Note: Gupta and Liang (1993b) studied an empirical Bayes selection rule for this selection problem under the assumption that $\tau_1 = \tau_2 = \ldots = \tau_m$ which means that the prior is a symmetric Dirichlet distribution.

Asymptotic Optimality

The Bayes risk of the empirical Bayes selection rule $\hat{\underset{\sim}{d}}$ is given by $r(G, \hat{\underset{\sim}{d}})$ where

$$r(G, \hat{\underset{\sim}{d}}) = \sum_{i=1}^{k} r_i(G, \hat{d}_i)$$

and $r_i(G, \hat{d}_i)$ is given in (5.22). Note that $r_i(G, \hat{d}_i) \geq r_i(G, d_{Gi})$ since d_{Gi} is the Baye rule for component i. Therefore, $r(G, \hat{\underset{\sim}{d}}) - r(G, \underset{\sim}{d}_G) \geq 0$. This nonnegative regret risk is used as a measure of performance of the selection rule $\hat{\underset{\sim}{d}}$.

Theorem 5.2. Let $\hat{\underset{\sim}{d}}$ be the empirical Bayes selection rule constructed precedingly. Then, for each i,

$$r_i(G, \hat{d}_i) - r_i(G, d_{Gi}) = O(\exp(-ck))$$

for some positive constant c and therefore,

$$r(G, \hat{\underset{\sim}{d}}) - r(G, \underset{\sim}{d}_G) = O(\exp(-ck + \ell n\ k)).$$

References

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

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

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.*, **25**, 16–39.

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.

Berger, J. O. and Deely, J. J. (1988). A Bayesian approach to ranking and selection of related means with alternatives to AOV methodology. *J. Amer. Statist. Assoc.*, **83**, 364–373.

Bickel, P. J. and Yahav, J. A. (1977). On selecting a subset of good populations. *Statistical Decision Theory and Related Topics — II* (Eds. S. S. Gupta and D. S. Moore), Academic Press, New York, 37–55.

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.

Chen, P. (1986). On the least favorable configuration in multinomial selection problems. *Communications in Statistics: Theory and Methods*, **15**, 367–385.

Chen, P. (1988). An integrated formulation for selecting the most probable multinomial cell. *Ann. Inst. Statist. Math.*, **40**, 615–625.

Chernoff, H. and Yahav, J. A. (1977). A subset selection problem employing a new criterion. *Statistical Decision Theory and Related Topics — II* (Eds. S. S. Gupta and D. S. Moore), Academic Press, New York, 93–119.

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.

Deely, J. J. and Gupta, S. S. (1968). On the properties of subset selection procedures. *Sankhyā*, **A30**, 37–50.

Deely, J. J. and Gupta, S. S. (1988). Hierarchical Bayesian selection procedures for the best binomial population. Technical Report #88–21, Department of Statistics, Purdue University, West Lafayette, Indiana.

Fong, D. K. H. (1989). A Bayesian approach to the comparison of two means. *Commun. in Statist. Theory and Methods*, **18(10)**, 3785–3800.

Fong, D. K. H. (1990). Ranking and estimation of related means in two-way models — A Bayesian approach. *J. Statist. comput. Simul.*, **34**, 107–117.

Fong, D. K. H. (1992). A Bayesian approach to the estimation of the largest normal mean. *J. Statist. Comput. Simul.*, **40**, 119–133.

Goel, P. K. and Rubin, H. (1977). On selecting a subset containing the best population – A Bayesian approach. *Ann. Statist.*, **5**, 969–983.

Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Ph.D. Thesis (Mimeo. Ser. No. 150), Inst. of Statist., University of North Carolina, Chapel Hill, North Carolina.

Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics*, **7**, 225–245.

Gupta, S. S. and Hande, S. N. (1993). Single-sample Bayes and empirical Bayes rules for ranking and estimating multinomial probabilities, *J. Statist. Plann. Inference*, **35**, 367–382.

Gupta, S. S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. *J. Statist. Plann. Inference*, **8**, 87–101.

Gupta, S. S. and Hsu, J. C. (1978). On the performance of some subset selection procedures. *Commun. Statist.-Simula. Computa.*, **B7(6)**, 561–591.

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

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 populations. *Statistical Decision Theory and Related Topics-IV* (Eds. S. S. Gupta and J. O. Berger), Springer–Verlag, Vol. I, 213–224.

Gupta, S. S. and Liang, T. (1989a). Selecting the best binomial population: parametric empirical Bayes approach. *J. Statist. Plann. Inference*, **23**, 21–31.

Gupta, S. S. and Liang, T. (1989b). Parametric empirical Bayes rules for selecting the most probable multinomial event. Chapter 18 in *Contribution to Probability and Statistics:*

Essays in Honor of Ingram Olkin, Springer–Verlag, 318–328.

Gupta, S. S. and Liang, T. (1991). On the asymptotic optimality of certain empirical Bayes simultaneous testing procedures for Poisson populations. *Statistics & Decisions*, **9**, 263–283.

Gupta, S. S. and Liang, T. (1993a). On empirical Bayes selection rules for sampling inspection. To appear in *J. Statist. Plann. Inference.*

Gupta, S. S. and Liang, T. (1993b). Bayes and Empirical Bayes rules for selecting fair multinomial populations. *Statistics and Probability: A Raghu Raj Bahadur Festschrift*, (Eds. J. K. Ghosh, ets.), Wiley Eastern Limited, 265–277.

Gupta, S. S. and Miescke, K. J. (1984). On two–stage Bayes selection procedures. *Sankhyā* **B46**, 123–134.

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

Gupta, S. S. and Panchapakesan, S. (1979). *Multiple Decision Procedures.* Wiley, New York.

Gupta, S. S. and Wong, W.-Y. (1975). Subset selection procedures for finite schemes in information theory. Colloquia Mathematica Societatis Janos Bolya. 16: Topics in Information Theory (eds. I. Csiszar and P. Elias), 279–291.

Gupta, S. S. and Yang, H. M. (1985). Bayes–P^* subset selection procedures for the best populations. *J. Statist. Plann. Inference* **12**, 213–233.

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 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.

Miescke, K. J. (1979). Bayesian subset selection for additive and linear loss functions. *Commun. Statist.-Theor. Meth.*, **A8(12)**, 1205–1226.

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.

Patil, G. P. and Taillie, C. (1982). Diversity as a concept and its measurement. *Journal of the American Statistical Association*, **77**, 548–567.

Pielou, E. C. (1975). Ecological Diversity. Wiley Interscience, New York.

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 probably multinomial event. *Commun. Statist.-Theor. Meth.*, **A9(3)**, 265–276.

Rizvi, M. H., Alam, K. and Lal Saxena, K. M. (1987). Selection procedure for multinomial populations with respect to diversity indices. Contributions to the Theory and Application of Statistics (ed. A. E. Gelfand), Academic Press, New York, 485–510.

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.