Simultaneous Selection For Homogeneous Multinomial Populations Based on Entropy Function: An Empirical Bayes Approach.*

by

Shanti S. Gupta** and Lii-Yuh Leu
Purdue University National Central University

TaChen Liang
Wayne State University

Technical Report #94-1C

Department of Statistics
Purdue University

January 1994

*

Simultaneous Selection For Homogeneous Multinomial Populations Based on Entropy Function: An Empirical Bayes Approach.*

by

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

and

Lii-Yuh Leu
Graduate Institute of Statistics
National Central University
Chung-Li, Taiwan, R.O.C.

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

Abstract

Consider k independent m-cell multinomial populations $\pi_1, \ldots, \pi_k$, where π_i has the associated cell probability vector $\underline{p}_i = (p_{i1}, \ldots, p_{im})$, $i = 1, \ldots, k$. The entropy $\theta_i \equiv \theta(\underline{p}_i) = -\sum_{j=1}^{m} p_{ij} \log p_{ij}$ is used as a measure of homogeneity of the population π_i. A Bayes selection rule relative to the Dirichlet prior distribution $D_m(\alpha_1, \ldots, \alpha_m)$ is obtained for selecting all homogeneous populations compared with a standard θ_0. When the values of the parameters $\alpha_1, \ldots, \alpha_m$ are unknown, an empirical Bayes selection rule is proposed and its corresponding asymptotic optimality is investigated. It is shown that the proposed empirical Bayes selection rule is asymptotically optimal relative to the class of Dirichlet prior distributions, and the associated regret risk converges to zero with a convergence rate of order $O(\exp(-\tau k + \ln k))$ for some positive constant $\tau \equiv \tau(\alpha_1, \ldots, \alpha_m)$.

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

Key words and phrases: asymptotically optimal, empirical Bayes, entropy function, homogeneous population, rate of convergence, simultaneous selection.

---

*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 Technische Hochschule Darmstadt, while he was visiting there in October 1993, thanks to Professor Dr. Lehn for the invitation.

1. Introduction

The concept of diversity within a population is of considerable importance in statistical theory and applications. The multinomial distribution provides a model for studying the diversity within a population which is categorized into several classes according to a qualitative characteristic. Such studies arise in ecology, sociology, genetic, economics and many other sciences. Diversity in ecological contexts has been discussed by Pielou (1975) and Patil and Taillie (1982). 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.

In the literature, selection procedures using indices of diversity as selection criterion have been studied by many authors. Gupta and Huang (1976) studied the problem of selecting the population with the largest entropy for binomial populations. 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. Jeyaratnam and Panchapakesan (1989) developed entropy based subset selection procedure for Bernoulli populations. Gupta and Leu (1990) studied certain selection procedures based on the Gini-Simpson index. Recently, Liang and Panchapakesan (1991) and Gupta and Liang (1993) developed empirical Bayes procedures for selecting the most homogeneous multinomial populations and for selecting fair multinomial populations in terms of Gini-Simpson index.

In this paper, we deal with the problem of selecting homogeneous (defined later) multinomial populations according to the entropies using a parametric empirical Bayes approach. Let $\pi_1, \ldots, \pi_k$ be k independent multinomial populations, each having m cells with associated probability vector $\underline{p}_i = (p_{i1}, \ldots, p_{im}), i = 1, \ldots, k$. The entropy associated with π_i is $\theta_i \equiv \theta(\underline{p}_i) = -\sum_{j=1}^{m} p_{ij} \log p_{ij}$. Note that $0 < \theta_i \leq \log m$ for all $\underline{p}_i$; also, $\theta_i = \log m$ when $p_{i1} = \ldots = p_{im} = \frac{1}{m}$, and $\theta_i \approx 0$ when one of the cell probabilities, say p_{ij}, is very close to 1 and all the other p_{il}, $l \neq j$, are close to 0. For a given constant θ_0

such that $0 < \theta_0 < \theta^* = \log\ m, \pi_i$ is said to be homogeneous if $\theta_i \geq \theta_0$ and nonhomogeneous, otherwise. Our goal is to derive statistical selection procedures for selecting all homogeneous populations while excluding all nonhomogeneous populations.

The paper is organized as follows. In Section 2, we derive a Bayes procedure $\underset{\sim}{d}_G$ for selecting all homogeneous multinomial populations according to Shannon's entropy under a Dirichlet prior distribution G for the cell probability vector. When the hyperparameters of the prior are unknown, the Bayes procedure $\underset{\sim}{d}_G$ cannot be implemented. In such a situation, we study this selection problem through a parametric empirical Bayes approach in Section 3. We incorporate information from the k populations to construct estimators for the unknown hyperparameters and obtain a "Bayes" selection procedure $\underset{\sim}{d}^*$ relative to the estimated prior distribution. We investigate the corresponding asymptotic optimality of the selection procedure $\underset{\sim}{d}^*$ in Section 4. It is shown that the associated regret risk of the selection procedure $\underset{\sim}{d}^*$ converges to zero with a rate of convergence of order $O(\exp(-\tau k + \ln\ k))$ for some positive constant τ depending on the values of the hyperparameters.

2. A Bayes Selection Procedure

In order to derive an empirical Bayes selection procedure, as a first step, we obtain a Bayes selection procedure under a known prior distribution. Let $\Omega = \{\underset{\sim}{p} = (\underset{\sim}{p}_1, \ldots, \underset{\sim}{p}_k) | \underset{\sim}{p}_i = (p_{i1}, \ldots, p_{im}), i = 1, \ldots, k\}$ be the parameter space. Let $\underset{\sim}{X}_i = (X_{i1}, \ldots, X_{im})$ be the cell frequencies arising from a sample of N independent trials from population $\pi_i, i = 1, \ldots, k$. Then, given $\underset{\sim}{p}_i, \underset{\sim}{X}_i$ has a multinomial distribution Multi $(N, \underset{\sim}{p}_i)$ with the probability function

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

at point $\underset{\sim}{x}_i = (x_{i1}, \ldots, x_{im})$ for which $0 \leq x_{ij} \leq N, 1 \leq j \leq m$ and $\sum_{j=1}^{m} x_{ij} = N$. Given the $\underset{\sim}{p}_1, \ldots, \underset{\sim}{p}_k$, the random vectors $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_k$ are assumed to be conditionally independent. We also assumed that $\underset{\sim}{p}_i$ is a realization of a random probability vector $\underset{\sim}{P}_i = (P_{i1}, \ldots, P_{im})$, and $\underset{\sim}{P}_1, \ldots, \underset{\sim}{P}_k$ are *iid*, having a Dirichlet prior distribution G with hyperparameters

$(\alpha_1, \ldots, \alpha_m)$; namely, the corresponding probability density function is given by,

$$g(\underset{\sim}{p}_i|\alpha_1, \ldots, \alpha_m) = \frac{\Gamma(\alpha_0)}{\prod\limits_{j=1}^{m} \Gamma(\alpha_j)} \prod_{j=1}^{m} p_{ij}^{\alpha_j - 1}, \tag{2}$$

where $\alpha_j > 0, j = 1, \ldots, m$ and $\alpha_0 = \sum\limits_{j=1}^{m} \alpha_j$. We denote the Dirichlet prior having hyperparameters $(\alpha_1, \ldots, \alpha_m)$ by $D_m(\alpha_1, \ldots, \alpha_m)$. Let $A = \{\underset{\sim}{a}|\underset{\sim}{a} = (a_1, \ldots, a_k), a_i = 0, 1, i = 1, \ldots, k\}$ be the action space. When action $\underset{\sim}{a}$ is taken, it means that population π_i is selected as a homogeneous population if $a_i = 1$ and excluded as nonhomogeneous if $a_i = 0$. For the parameter vector $\underset{\sim}{p}$ and action $\underset{\sim}{a}$, the loss function $L(\underset{\sim}{p}, \underset{\sim}{a})$ is defined to be

$$L(\underset{\sim}{p}, \underset{\sim}{a}) = \sum_{i=1}^{k} a_i(\theta_0 - \theta_i) I_{[0,\theta_0)}(\theta_i) + \sum_{i=1}^{k} (1 - a_i)(\theta_i - \theta_0) I_{[\theta_0, \theta^*]}(\theta_i). \tag{3}$$

In (3), the first summation is the loss due to selecting some nonhomogeneous populations while the second summation is the loss of not selecting some homogeneous populations.

Let $\mathcal{X}$ be the sample space of the random vector $\underset{\approx}{X} = (\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_k)$. A selection procedure $\underset{\sim}{d} = (d_1, \ldots, d_k)$ is defined to be a mapping from the sample space $\mathcal{X}$ into the product space $[0, 1]^k$, so that, for each $\underset{\approx}{x} = (\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_k) \in \mathcal{X}, \underset{\sim}{d}(\underset{\approx}{x}) = (d_1(\underset{\approx}{x}), \ldots, d_k(\underset{\approx}{x}))$, and $d_i(\underset{\approx}{x})$ is viewed as the probability of selecting population π_i as a homogeneous population given $\underset{\approx}{X} = \underset{\approx}{x}$ is observed. We let $\mathcal{C}$ denote the class of all selection procedures defined in the above way.

Bayes Risk of $\underset{\sim}{d}$ Relative to G

Under the preceding statistical model and the loss function $L(\underset{\sim}{p}, \underset{\sim}{a})$ in (3), by interchanging the order of summation and integration, the Bayes risk of a selection procedure $\underset{\sim}{d} = (d_1, \ldots, d_k)$, denoted by $r(G, \underset{\sim}{d})$, can be obtained as follows.

$$\begin{aligned} r(G, \underset{\sim}{d}) = \sum_{\underset{\approx}{x} \in \mathcal{X}} \sum_{i=1}^{k} d_i(\underset{\approx}{x}) \{\theta_0 - E[\theta_i | \underset{\approx}{X} = \underset{\approx}{x}]\} \prod_{i=1}^{k} f_i(\underset{\sim}{x}_i) \\ + \sum_{i=1}^{k} \int_{\Omega_i(\theta_0)} (\theta_i - \theta_0) g(\underset{\sim}{p}_i | \alpha_1, \ldots, \alpha_m) d\underset{\sim}{p}_i, \end{aligned} \tag{4}$$

where $\Omega_i(\theta_0) = \{\underset{\sim}{p}_i | \theta_i = \theta(\underset{\sim}{p}_i) > \theta_0\}, i = 1, \ldots, k,$

$f_i(\underset{\sim}{x}_i) = \int f_i(\underset{\sim}{x}_i|\underset{\sim}{p}_i) g(\underset{\sim}{p}_i|\alpha_1, \ldots, \alpha_m) d\underset{\sim}{p}_i$

: the marginal probability function of the random vector $\underset{\sim}{X}_i = (X_{i1}, \ldots, X_{im})$,

$E[\theta_i|\underset{\approx}{X} = \underset{\approx}{x}]$: the posterior expectation of θ_i given $\underset{\approx}{X} = \underset{\approx}{x}$

$$\begin{aligned} &= -\sum_{j=1}^{m} E[P_{ij} \;\log\; P_{ij}|\underset{\approx}{X} = \underset{\approx}{x}] \\ &= -\sum_{j=1}^{m} E[P_{ij} \;\log\; P_{ij}|\underset{\sim}{X}_i = \underset{\sim}{x}_i] \end{aligned} \tag{5}$$

since $\underset{\sim}{P}_i$ is independent of $\underset{\sim}{X}_l, l \neq i$, under the preceding assumptions. Therefore, a Bayes selection procedure $\underset{\sim}{d}_G = (d_{1G}, \ldots, d_{kG})$ can be obtained as follows: For each $\underset{\approx}{x} \epsilon \mathcal{X}$, $i = 1, \ldots, k$,

$$d_{iG}(\underset{\approx}{x}) = \begin{cases} 1 & \text{if } E[\theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i] \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Note that from (5) and (6), d_{iG} depends on $\underset{\approx}{x}$ only through $\underset{\sim}{x}_i$, and therefore, is denoted by $d_{iG}(\underset{\sim}{x}_i)$. The minimum Bayes risk among the class $\mathcal{C}$ is $r(G, \underset{\sim}{d}_G) = \sum_{i=1}^{k} r_i(G, d_{iG})$, where

$$\begin{aligned} r_i(G, d_{iG}) &= \sum_{\underset{\sim}{x}_i} d_{iG}(\underset{\sim}{x}_i) \;\{\theta_0 - E[\theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i]\;\} f_i(\underset{\sim}{x}_i) + C_i \\ \text{and } C_i &= \int_{\Omega_i(\theta_0)} (\theta_i - \theta_0) g(\underset{\sim}{p}_i|\alpha_1, \ldots, \alpha_m) d\underset{\sim}{p}_i. \end{aligned} \tag{7}$$

An Expression for $E[\theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i]$

It is known that the posterior distribution of $\underset{\sim}{P}_i = (P_{i1}, \ldots, P_{im})$ given $\underset{\sim}{X}_i = \underset{\sim}{x}_i$ is the Dirichlet distribution $D_m(x_{i1} + \alpha_1, \ldots, x_{im} + \alpha_m)$. Thus, for each component $j = 1, \ldots, m$, the posterior distribution of P_{ij} given $\underset{\sim}{X}_i = \underset{\sim}{x}_i$ follows a $Beta(x_{ij} + \alpha_j, N + \alpha_0 - x_{ij} - \alpha_j)$ distribution and the associated posterior probability density function is denoted by $h_{ij}(p_{ij}|x_{ij} + \alpha_j, N + \alpha_0 - x_{ij} - \alpha_j)$.

Now, since $0 \leq p_{ij} \leq 1$, $-p_{ij} \;\log\; p_{ij} = \sum_{y=1}^{\infty} p_{ij}(1 - p_{ij})^y / y$. Therefore,

$$
\begin{aligned}
&E[-P_{ij}\log P_{ij}|\underset{\sim}{X}_i=\underset{\sim}{x}_i] \\
=&\sum_{y=1}^{\infty}\frac{1}{y}\int_0^1 p(1-p)^y\ h_{ij}(p|x_{ij}+\alpha_j,N+\alpha_0-x_{ij}-\alpha_j)dp \\
=&\sum_{y=1}^{\infty}\frac{\Gamma(N+\alpha_0)\Gamma(x_{ij}+\alpha_j+1)\Gamma(N+\alpha_0-x_{ij}-\alpha_j+y)}{y\Gamma(x_{ij}+\alpha_j)\Gamma(N+\alpha_0-x_{ij}-\alpha_j)\Gamma(N+\alpha_0+y+1)} \\
=&\sum_{y=1}^{\infty}\left[\frac{x_{ij}+\alpha_j}{y(N+\alpha_0+y)}\prod_{z=0}^{y-1}\left(\frac{N+\alpha_0-x_{ij}-\alpha_j+z}{N+\alpha_0+z}\right)\right] \\
=&\sum_{y=1}^{\infty}q_{ijy}(\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i) \\
\equiv&q_{ij}(\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i),
\end{aligned}
\tag{8}
$$

where $q_{ijy}(\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i)=\frac{x_{ij}+\alpha_j}{y(N+\alpha_0+y)}\prod_{z=0}^{y-1}\left(\frac{N+\alpha_0-x_{ij}-\alpha_j+z}{N+\alpha_0+z}\right)$.

Note that for each $y=1,2,\ldots,q_{ijy}(\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i)$ is continuous in α_l for $\alpha_l\geq 0$, $\ell=1,\ldots,m$. Also, $\sum_{y=1}^{\infty}q_{ijy}\ (\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i)$ converges uniformly for each $(\alpha_1,\ldots,\alpha_m)$ $\epsilon[0,\infty)^m$. Therefore, $q_{ij}(\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i)$ is continuous in $(\alpha_1,\ldots,\alpha_m)$ for each $(\alpha_1,\ldots,\alpha_m)$ $\epsilon[0,\infty)^m$.

Now,

$$
\begin{aligned}
E[\theta_i|\underset{\sim}{X}_i=\underset{\sim}{x}_i]&=\sum_{j=1}^{m}E[-P_{ij}\ \log\ P_{ij}|\underset{\sim}{X}=\underset{\sim}{x}_i] \\
&=\sum_{j=1}^{m}\ q_{ij}(\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i) \\
&=q_i(\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i).
\end{aligned}
\tag{9}
$$

Therefore, $q_i(\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i)$ is continuous in $(\alpha_1,\ldots,\alpha_m)$ for each $(\alpha_1,\ldots,\alpha_m)\epsilon[0,\infty)^m$. Also, the following limits exist: For $\alpha_l\geq 0, l=1,\ldots,m$,

(a) $q_i(\alpha_1,\ldots,\alpha_{j-1},0,\alpha_{j+1},\ldots,\alpha_m|\underset{\sim}{x}_i)=\lim_{\alpha_j\downarrow 0}q_i(\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i)$.

(b) $q_i(\alpha_1,\ldots,\alpha_{j-1},\infty,\ \alpha_{j+1},\ldots,\alpha_m|\underset{\sim}{x}_i)\ =\lim_{\alpha_j\to\infty}q_i(\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i)$.

Finally, for the later use, we define $q_i(\infty, \ldots, \infty | \underset{\sim}{x}_i)$ by

$$\begin{aligned} q_i(\infty, \ldots, \infty | \underset{\sim}{x}_i) &\equiv \lim_{\alpha \to \infty} \; q_i(\alpha, \ldots, \alpha | \underset{\sim}{x}_i) \\ &= \sum_{j=1}^{m} \lim_{\alpha \to \infty} \; q_{ij}(\alpha, \ldots, \alpha | \underset{\sim}{x}_i) \\ &= \sum_{y=1}^{\infty} \frac{1}{y} \, (\frac{m-1}{m})^y \\ &= \log \; m. \end{aligned}$$

Alternative Expression for $E[\theta_i | \underset{\sim}{X}_i = \underset{\sim}{x}_i]$ in terms of Moments

First note that $\underset{\sim}{X}_i = (X_{i1}, \ldots, X_{im}), i = 1, \ldots, k$, are *iid*, and

$$\mu_j \equiv E[X_{1j}] = \frac{N\alpha_j}{\alpha_0} \tag{10}$$

$$\lambda_j \equiv E[X_{1j}^2] = \mu_j + \frac{N(N-1)\alpha_j(\alpha_j \; + \; 1)}{\alpha_0(\alpha_0 + 1)} \tag{11}$$

Now, using (10) to substitute for α_j in (11), we obtain

$$\alpha_0[N(\lambda_j - \mu_j) - (N-1)\mu_j^2] = N[N\mu_j - \lambda_j] > 0. \tag{12}$$

since $N\mu_j - \lambda_j = \frac{N(N-1)\alpha_j}{\alpha_0}[1 - \frac{\alpha_j+1}{\alpha_0+1}] > 0$. Summing both sides of the equality in (12) for $j = 1$ to m and noting that $\sum_{j=1}^{m} \mu_j = N$, we obtain

$$\alpha_0 = \frac{N(N^2 - \sum_{j=1}^{m} \lambda_j)}{A} > 0 \tag{13}$$

$$\alpha_j = \frac{\mu_j(N^2 - \sum_{j=1}^{m} \lambda_j)}{A} > 0 \tag{14}$$

where $A = N \sum_{j=1}^{m} \lambda_j - N^2 - (N-1) \sum_{j=1}^{m} \mu_j^2 > 0.$

By substituting (13) and (14) into (8) and (9), we see that $E[-P_{ij} \log P_{ij}|\underset{\sim}{X}_i = \underset{\sim}{x}_i]$ and $E[\theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i]$ can also be viewed as functions of the moments μ_j and $\lambda_j, j = 1\ldots, m$. We denote such functions by

$$\begin{aligned} E[-P_{ij} \log P_{ij}|\underset{\sim}{X}_i = \underset{\sim}{x}_i] &\equiv q_{ij}(\alpha_1, \ldots, \alpha_m|\underset{\sim}{x}_i) \\ &\equiv Q_{ij}(\mu_1, \ldots, \mu_m, \lambda_1, \ldots, \lambda_m|\underset{\sim}{x}_i), \end{aligned} \tag{15}$$

and

$$\begin{aligned} E[\theta_i|\underset{\sim}{X}_i = x_i] &\equiv q_i(\alpha_1, \ldots, \alpha_m|\underset{\sim}{x}_i) \\ &\equiv Q_i(\mu_1, \ldots, \mu_m, \lambda_1, \ldots, \lambda_m|\underset{\sim}{x}_i). \end{aligned} \tag{16}$$

Also, $Q_i(\mu_1, \ldots, \mu_m, \lambda_1, \ldots, \lambda_m|\underset{\sim}{x}_i)$ is continuous in μ_j and $\lambda_j, j = 1, \ldots, m$.

3. An Empirical Bayes Simultaneous Selection Procedure

It is assumed that the hyperparameters of the Dirichlet prior distribution $D_m(\alpha_1, \ldots, \alpha_m)$ are unknown. In such a situation, it is not possible to implement the Bayes selection procedure $\underset{\sim}{d}_G$ for the concerned selection problem. In the following, by incorporating information from the k populations, we first construct estimators for the unknown parameters α_j, μ_j and $\lambda_j, j = 1, \ldots, m$. Then, by mimicking the behavior of the Bayes selection procedure, we derive an empirical Bayes simultaneous selection procedure.

Define, for each $j = 1, \ldots, m$,

$$\begin{cases} \hat{\mu}_j = \frac{1}{k}\sum_{i=1}^{k} X_{ij}, \quad \hat{\lambda}_j = \frac{1}{k}\sum_{i=1}^{k} X_{ij}^2 \qquad \text{and} \\ \hat{A} = N\sum_{j=1}^{m} \hat{\lambda}_j - N^2 - (N-1)\sum_{j=1}^{m} \hat{\mu}_j^2. \end{cases} \tag{17}$$

Note that $E[\hat{\mu}_j] = \mu_j$, $E[\hat{\lambda}_j] = \lambda_j$, $j = 1, \ldots, m$, and $\hat{\mu}_j$, $\hat{\lambda}_j$ and $\hat{A}$ are consistent estimators of μ_j, λ_j and A, respectively. We may estimate α_j by substituting the unknown parameters by the corresponding estimators according to (14). However, though $A > 0$, it is possible that $\hat{A} \le 0$. Using (10) and (11),

$$\begin{aligned} A &\equiv N\sum_{j=1}^{m} \lambda_j - N^2 - (N-1)\sum_{j=1}^{m} \mu_j^2 \\ &= \frac{N^3}{\alpha_0} - \frac{N}{\alpha_0} - \sum_{j=1}^{m} \frac{N(N-1)\alpha_j(\alpha_j+1)}{\alpha_0^2(\alpha_0+1)} \end{aligned}$$

which converges to zero as α_0 tends to infinity. Therefore, as $\hat{A} \le 0$, it may indicate that the value of α_0 may be very large, and therefore, at least one of the α_j's is very large since $\alpha_0 = \sum_{j=1}^{m} \alpha_j$ and $\alpha_j > 0, j = 1, \ldots, m$. Based on these discussions, we derive an empirical Bayes simultaneous selection procedure as follows.

When $\hat{A} > 0$, we estimate $Q_i(\mu_1, \ldots, \mu_m, \lambda_1, \ldots, \lambda_m | \underset{\sim}{x}_i)$ by $Q_i(\hat{\mu}_1, \ldots, \hat{\mu}_m, \hat{\lambda}_1, \ldots, \hat{\lambda}_m | \underset{\sim}{x}_i)$.

When $\hat{A} \le 0$, we estimate $\alpha_j, j = 1, \ldots, m$, by $\hat{\alpha}_j$, where

$$\hat{\alpha}_j = \begin{cases} \infty & \text{if either } (N^2 - \sum_{i=1}^{m} \hat{\lambda}_i)\hat{\mu}_j > 0 \text{ or } \hat{\lambda}_j = N^2, \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

When $\hat{A} \le 0$, according to the preceding definition (18), either all the α_j's equal to infinity or exact one of the $\hat{\alpha}_j$'s equals to infinity and all the others equal to zero. Now, we estimate $q_i(\alpha_1, \ldots, \alpha_m | \underset{\sim}{x}_i)$ by $q_i(\hat{\alpha}_1, \ldots, \hat{\alpha}_m | \underset{\sim}{x}_i)$.

Empirical Bayes Simultaneous Selection Procedure $\underset{\sim}{d}^* = (d_1^*, \ldots, d_k^*)$

We propose an empirical Bayes simultaneous selection procedure $\underset{\sim}{d}^*$ as follows: For each $i = 1, \ldots, k, \underset{\sim}{x} \epsilon \mathcal{X}$, define

$$d_i^*(\underset{\sim}{x}) = \begin{cases} 1 & \text{if either } (\hat{A} > 0 \text{ and } Q_i(\hat{\mu}_1, \ldots, \hat{\mu}_m, \hat{\lambda}_1, \ldots, \hat{\lambda}_m | \underset{\sim}{x}_i) \ge \theta_0) \\ & \quad \text{or } (\hat{A} \le 0 \text{ and } q_i(\hat{\alpha}_1, \ldots, \hat{\alpha}_m | \underset{\sim}{x}_i) \ge \theta_0), \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

The Bayes risk of the selection procedure $\underset{\sim}{d}^*$ is

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

where

$$\begin{aligned} r_i(G, d_i^*) &= \sum_{\underset{\sim}{x} \epsilon \mathcal{X}} d_i^*(\underset{\sim}{x}) \{\theta_0 - E[\theta_i | \underset{\sim}{X}_i = \underset{\sim}{x}_i]\} \prod_{j=1}^{k} f_j(\underset{\sim}{x}_j) + C_i \\ &= \sum_{\underset{\sim}{x}_i} E_i[d_i^*(\underset{\sim}{X}(i), \underset{\sim}{x}_i)] \ \{\theta_0 - E[\theta_i | \underset{\sim}{X}_i = \underset{\sim}{x}_i] \ \} f_i(\underset{\sim}{x}_i) + C_i, \end{aligned} \tag{20}$$

where $\underset{\sim}{X}(i) = (\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_{i-1}, \underset{\sim}{X}_{i+1}, \ldots, \underset{\sim}{X}_k)$ and the expectation E_i is computed with respect to the probability measure generated by $\underset{\sim}{X}(i)$.

An Illustrative Example: Suppose that $k = 10$, $m = 5$, $N = 20$, $\theta^* = \ln 5$ and the control level $\theta_0 = 1.40$. The data $(X_{i1}, \ldots, X_{i5})$, $i = 1, \ldots, 10$, from each of the 10 multinomial populations are listed as follows.

i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}
1	7	3	3	4	3
2	3	3	1	10	3
3	0	2	10	5	3
4	4	4	2	9	1
5	8	1	3	5	3
6	1	3	2	12	2
7	2	1	2	2	13
8	12	2	1	2	3
9	2	10	2	1	5
10	11	2	3	0	4

We estimate the parameters μ_j, λ_j, $j = 1, \ldots, m$, A and α_j, $j = 1, \ldots, m$ based on the data. The corresponding values of the estimators are:

$$\begin{array}{lllll}
\hat{\mu}_1 = 5.0, & \hat{\mu}_2 = 3.1, & \hat{\mu}_3 = 2.9, & \hat{\mu}_4 = 5.7, & \hat{\mu}_5 = 3.3 \\
\hat{\lambda}_1 = 41.2, & \hat{\lambda}_2 = 15.7, & \hat{\lambda}_3 = 14.5, & \hat{\lambda}_4 = 47.7, & \hat{\lambda}_5 = 12.7 \\
\hat{A} = 594.4, & \text{and} & & & \\
\hat{\alpha}_1 = 2.256, & \hat{\alpha}_2 = 1.399, & \hat{\alpha}_3 = 1.309, & \hat{\alpha}_4 = 2.572, & \hat{\alpha}_5 = 1.489.
\end{array}$$

Accordingly, the values of the empirical Bayes estimators $Q_i(\underset{\sim}{X}_i) \equiv Q_i(\hat{\mu}_1, \ldots, \hat{\mu}_5, \hat{\lambda}_1, \ldots, \hat{\lambda}_5 | \underset{\sim}{X}_i)$ are as follows:

i	1	2	3	4	5
$Q_i(\underset{\sim}{X}_i)$	1.488	1.371	1.361	1.398	1.419
i	6	7	8	9	10
$Q_i(\underset{\sim}{X}_i)$	1.296	1.254	1.283	1.386	1.312

Comparing the values $Q_i(\underset{\sim}{X}_i)$ with $\theta_0 = 1.40$, the empirical Bayes procedure $\underset{\sim}{d}^*$ selects populations π_1 and π_5 as homogeneous populations.

4. Asymptotic Optimality

Since $\underset{\sim}{d}_G = (d_{1G}, \ldots, d_{kG})$ is the Bayes selection procedure, for the empirical Bayes selection procedure $\underset{\sim}{d}^* = (d_1^*, \ldots, d_k^*)$, $r_i(G, d_i^*) - r_i(G, d_{iG}) \geq 0$ for each component $i =$

$1,\ldots,k$, and therefore, $r(G,\underset{\sim}{d}^*) - r(G,\underset{\sim}{d}_G) = \sum_{i=1}^{k}[r_i(G,d_i^*) - r_i(G,d_{iG})] \geq 0$. This non-negative regret risk $[r(G,\underset{\sim}{d}^*) - r(G,\underset{\sim}{d}_G)]$ is used as a measure of performance of the empirical Bayes selection procedure $\underset{\sim}{d}^*$.

Definition 4.1. A selection procedure $\underset{\sim}{d} = (d_1,\ldots,d_k) \in \mathcal{C}$ is said to be asymptotically optimal of order $\{\varepsilon_k\}$ relative to the prior distribution G if $r(G,\underset{\sim}{d}) - r(G,\underset{\sim}{d}_G) = O(\varepsilon_k)$ where $\{\varepsilon_k\}$ is a sequence of positive values such that $\lim_{k\to\infty} \varepsilon_k = 0$.

Since $0 < \theta_0, \theta_i <\ \log\ m \equiv \theta^*$, $|\theta_0 - E[\theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i]| < \theta^*$. According to (7) and (20), for each $i = 1,\ldots,k$,

$$
\begin{aligned}
0 \leq & r_i(G,d_i^*) - r_i(G,d_{iG}) \\
= & \sum_{\underset{\sim}{x}_i} E_i[d_i^*(\underset{\approx}{X}) - d_{iG}(\underset{\sim}{x}_i)|\underset{\sim}{X}_i = \underset{\sim}{x}_i]\{\theta_0 - E[\theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i]\} f_i(\underset{\sim}{x}_i) \\
\leq & \theta^* \sum_{\underset{\sim}{x}_i \in A_i} P_i\ \{d_i^*(\underset{\approx}{X}) = 0,\ d_{iG}(\underset{\sim}{x}_i) = 1|\underset{\sim}{X}_i = \underset{\sim}{x}_i\}\ f_i(\underset{\sim}{x}_i) \\
& + \theta^* \sum_{\underset{\sim}{x}_i \in B_i} P_i\ \{d_i^*(\underset{\approx}{X}) = 1,\ d_{iG}(\underset{\sim}{x}_i) = 0|\underset{\sim}{X}_i = \underset{\sim}{x}_i\}\ f_i(\underset{\sim}{x}_i),
\end{aligned}
\tag{21}
$$

where $A_i = \{\underset{\sim}{x}_i|E[\theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i] > \theta_0\}$, $B_i = \{\underset{\sim}{x}_i|E[\theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i] < \theta_0\}$, and P_i is the probability measure generated by $\underset{\approx}{X}(i)$. Note that $A_1 = \ldots = A_k, B_1 = \ldots = B_k$ since $(\underset{\sim}{X}_i, \underset{\sim}{P}_i)$, $i = 1,\ldots,k$ are *iid* random vectors. Also, recall that

$$
\begin{aligned}
E[\theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i] &= q_i(\alpha_1,\ldots,\alpha_m|\underset{\sim}{x}_i) \\
&= Q_i(\mu_1,\ldots,\mu_m,\lambda_1,\ldots,\lambda_m|\underset{\sim}{x}_i).
\end{aligned}
$$

Let $\underset{\sim}{\lambda} = (\lambda_1,\ldots,\lambda_m), \hat{\underset{\sim}{\lambda}} = (\hat{\lambda}_1,\ldots,\hat{\lambda}_m)$, $\underset{\sim}{\mu} = (\mu_1,\ldots,\mu_m)$ and $\hat{\underset{\sim}{\mu}} = (\hat{\mu}_1,\ldots,\hat{\mu}_m)$. For each $\underset{\sim}{x}_i \in A_i, Q_i(\underset{\sim}{\mu},\underset{\sim}{\lambda}|\underset{\sim}{x}_i) > \theta_0$. Also, $A = N\sum_{j=1}^{m}\lambda_j - N^2 - (N-1)\sum_{j=1}^{m}\mu_j^2 > 0$. According to (19), for each $\underset{\sim}{x}_i \in A_i$,

$$
\begin{aligned}
& P_i\ \{d_i^*(\underset{\approx}{X}) = 0,\ d_{iG}(\underset{\sim}{X}_i) = 1|\underset{\sim}{X}_i = \underset{\sim}{x}_i\} \\
\leq & P_i\ \{Q_i(\hat{\underset{\sim}{\mu}},\hat{\underset{\sim}{\lambda}}|\underset{\sim}{x}_i) - Q_i(\underset{\sim}{\mu},\underset{\sim}{\lambda}|\underset{\sim}{x}_i) < \theta_0 - Q_i(\underset{\sim}{\mu},\underset{\sim}{\lambda}|\underset{\sim}{x}_i) \text{ or } \hat{A} - A \leq -A|\underset{\sim}{X}_i = \underset{\sim}{x}_i\}.
\end{aligned}
\tag{22}
$$

Since both $Q_i(\underset{\sim}{\mu},\underset{\sim}{\lambda}|\underset{\sim}{x}_i)$ and A are continuous functions of $\underset{\sim}{\mu}$ and $\underset{\sim}{\lambda}$ for $\mu_j \geq 0, \lambda_j \geq 0, j = 1,\ldots,m$, there exists positive constant $w(\underset{\sim}{\mu},\underset{\sim}{\lambda},\underset{\sim}{x}_i)$ such that

$$\begin{aligned}&\{|Q_i(\hat{\mu},\hat{\lambda}|x_i)-Q_i(\mu,\lambda|x_i)|>|\theta_0-Q_i(\mu,\lambda|x_i)|\}\ \cup\ \{\ |\hat{A}-A|\geq A\}\\&\subset\ \{|\hat{\mu}_j-\mu_j|>w(\mu,\lambda,x_i)\text{ or }|\hat{\lambda}_j-\lambda_j|>w(\mu,\lambda,x_i),\text{ for some }j=1,\ldots,m\ \}.\end{aligned}\tag{23}$$

Combining (22) and (23) yields that for each $x_i \in A_i$,

$$\begin{aligned}&P_i\ \{d_i^*(X)=0,d_{iG}(X_i)=1|X_i=x_i\ \}\\\leq&P_i\ \{\ |\hat{\mu}_j-\mu_j|>w(\mu,\lambda,x_i)\text{ or }|\hat{\lambda}_j-\lambda_j|>w(\mu,\lambda,x_i)\text{ for some }j=1,\ldots,m|X_i=x_i\ \}\\\leq&\sum_{j=1}^{m}[P_i\ \{|\hat{\mu}_j-\mu_j|>w(\mu,\lambda,x_i)|X_i=x_i\ \}+P_i\ \{|\hat{\lambda}_j-\lambda_j|>w(\mu,\lambda,x_i)|X_i=x_i\ \}].\end{aligned}\tag{24}$$

Similarly, for each $x_i \in B_i, Q_i(\mu,\lambda|x_i) < \theta_0$, and there exists positive constant $w(\mu,\lambda,x_i)$ such that

$$\begin{aligned}&P_i\{d_i^*(X)=1,d_{iG}(X_i)=0|X_i=x_i\}\\\leq&\sum_{j=1}^{m}[P_i\{|\hat{\mu}_j-\mu_j|>w(\mu,\lambda,x_i)|X_i=x_i\}+P_i\{|\hat{\lambda}_j-\lambda_j|>w(\mu,\lambda,x_i)|X_i=x_i\}].\end{aligned}\tag{25}$$

Thus, it suffices to evaluate the behaviors of $P_i\{|\hat{\mu}_j-\mu_j|>w(\mu,\lambda,x_i)|X_i=x_i\}$ and $P_i\{|\hat{\lambda}_j-\lambda_j|>w(\mu,\lambda,x_i)|X_i=x_i\}$ for each $j=1,\ldots,m$, and $i=1,\ldots,k$.

Let $\hat{\mu}_j(i)=\frac{1}{k-1}\sum\limits_{\substack{\ell=1\\\ell\neq i}}^{k}X_{\ell j}$. Then, $\hat{\mu}_j=\frac{k-1}{k}\hat{\mu}_j(i)+\frac{X_{ij}}{k}$. Thus,

$$\begin{aligned}&P_i\{|\hat{\mu}_j-\mu_j|>w(\mu,\lambda,x_i)|X_i=x_i\}\\&=P_i\{\hat{\mu}_j-\mu_j<-w(\mu,\lambda,x_i)|X_i=x_i\}+P_i\ \{\hat{\mu}_j-\mu_j>w(\mu,\lambda,x_i)|X_i=x_i\ \}\\&=P_i\{\hat{\mu}_j(i)-\mu_j<[-kw(\mu,\lambda,x_i)-x_{ij}+\mu_j]/(k-1)|X_i=x_i\}\\&\qquad+P_i\{\hat{\mu}_j(i)-\mu_j>[kw(\mu,\lambda,x_i)-x_{ij}+\mu_j]/(k-1)|X_i=x_i\}\\&=P_i\{\hat{\mu}_j(i)-\mu_j<[-kw(\mu,\lambda,x_i)-x_{ij}+\mu_j]/(k-1)\}\\&\qquad+P_i\{\hat{\mu}_j(i)-\mu_j>[kw(\mu,\lambda,x_i)-x_{ij}+\mu_j]/(k-1)\}\end{aligned}\tag{26}$$

since $X(i)$ and X_i are independent.

When k is sufficiently large, for fixed μ,λ, $[-kw(\mu,\lambda,x_i)-x_{ij}+\mu_j]\ /(k-1) < -\frac{1}{2}w(\mu,\lambda,x_i)$ and $[kw(\mu,\lambda,x_i)\ -x_{ij}+\mu_j]/(k-1) > \frac{1}{2}w(\mu,\lambda,x_i)$. Hence, for k being sufficiently large and for fixed μ and λ, (26) and (27) together yield that

$$
\begin{aligned}
&P_i\{|\hat{\mu}_j - \mu_j| > w(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i) | \underset{\sim}{X}_i = \underset{\sim}{x}_i\} \\
\leq &P_i\{\hat{\mu}_j(i) - \mu_j < -\frac{1}{2} w(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i)\} + P_i\{\hat{\mu}_j(i) - \mu_j > \frac{1}{2} w(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i)\} \\
\leq &\exp\left\{-(k-1)\frac{w^2(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i)}{4N^2}\right\} + \exp\left\{-\frac{(k-1)w^2(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i)}{4N^2}\right\},
\end{aligned} \tag{28}
$$

where the last inequality follows from Theorem 2 of Hoeffding (1963).

Similarly, we can also obtain the following inequality

$$
\begin{aligned}
&P_i\{|\hat{\lambda}_j - \lambda_j| > w(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i) | \underset{\sim}{X}_i = \underset{\sim}{x}_i\} \\
\leq &\exp\{-\frac{(k-1)w^2(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i)}{4N^4}\} + \exp\left\{-\frac{(k-1)w^2(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i)}{4N^4}\right\}
\end{aligned} \tag{29}
$$

Let $\tau = \frac{1}{4N^4} \min \ \{w^2(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i) | \underset{\sim}{x}_i \in A_i \cup B_i\}$

and $c = 2 \ \max \left\{\exp(\frac{w^2(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i)}{4N^2}) | \underset{\sim}{x}_i \in A_i \cup B_i\right\}$.

Then, $\tau \equiv \tau(\underset{\sim}{\mu}, \underset{\sim}{\lambda}) > 0$ and $c \equiv c(\underset{\sim}{\mu}, \underset{\sim}{\lambda}) < \infty$ since $A_i \cup B_i$ is a finite set. By the definitions of τ and c and (28) and (29), we obtain, for k being sufficiently large,

$$
\begin{aligned}
&P_i\{|\hat{\mu}_j - \mu_j| > w(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i) | \underset{\sim}{X}_i = \underset{\sim}{x}_i\} \leq \ c \ \exp\,(-\tau k), \\
&P_i\{|\hat{\lambda}_j - \lambda_j| > w(\underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{x}_i) | \underset{\sim}{X}_i = \underset{\sim}{x}_i\} \leq \ c \ \exp\,(-\tau k).
\end{aligned} \tag{30}
$$

By combining (21)-(30) together, we obtain an asymptotic optimality of the empirical Bayes selection procedure $\underset{\sim}{d}^*$, which is stated as follows.

Theorem 4.1 Let $\underset{\sim}{d}^*$ be the empirical Bayes simultaneous selection procedure defined in Section 3. Then under the Dirichlet prior $D_m(\alpha_1, \ldots, \alpha_m)$, $\underset{\sim}{d}^*$ is asymptotically optimal, and $r(G, \underset{\sim}{d}^*) - r(G, \underset{\sim}{d}_G) = O(\exp\,(-\tau \ k + \ell n \ k))$, where $\tau \equiv \tau(\underset{\sim}{\mu}, \underset{\sim}{\lambda})$ is the positive constant defined before.

Proof: Through (21)-(30), we can obtain that for each $i = 1 \ldots k$,

$$
\begin{aligned}
& r_i(G, d_i^*) - r_i(G, d_{iG}) \\
\le & \theta^* \sum_{\underset{\sim}{x}_i \in A_i} P_i \, \{d_i^*(\underset{\approx}{X}) = 0,\ d_{iG}(\underset{\sim}{x}_i) = 1 | \underset{\sim}{X}_i = \underset{\sim}{x}_i\} \ f_i(x_i) \\
& + \theta^* \sum_{\underset{\sim}{x}_i \in B_i} P_i \, \{d_i^*(\underset{\approx}{X}) = 1,\ d_{iG}(\underset{\sim}{x}_i) = 0 | \underset{\sim}{X}_i = \underset{\sim}{x}_i\} \ f_i(x_i) \\
\le & \theta^* 2m \ c \ \exp(-\tau k) \sum_{\underset{\sim}{x}_i \in A_i} f_i(\underset{\sim}{x}_i) + \theta^* \ 2m \ c \ \exp(-\tau k) \sum_{\underset{\sim}{x}_i \in B_i} f_i(\underset{\sim}{x}_i) \\
\le & 2m\theta^* \ c \ \exp(-\tau k).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
r(G, \underset{\sim}{d}^*) - r(G, \underset{\sim}{d}_G) \le & \sum_{i=1}^{k} 2m\theta^* c \quad \exp(-\tau k) \\
= & O(\exp(-\tau \ k + \ell \ n \ k)).
\end{aligned}
$$

References

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 Management Sciences*, **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.

Gupta, S.S. and Huang, D.Y. (1976). On subset selection procedures for the entropy function associated with the binomial populations. *Sankhyá*, **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 Liang, T. (1993). Bayes and empirical Bayes rules for selecting fair multinomial populations. *Statistics and Probability: A Raghu Raj Bahadur Festsehrift*, (Eds. J.K. Ghosh, et al), Wiley Eastern Limited, 265-277.

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.

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

Jeyaratnam, S. and Panchapakesan, S. (1989). Entropy based subset selection from Bernoulli populations. In *Proceedings of the International Conference on Computing and Information, ICCI' 89*, (Eds. R. Janicki and W.W. Koczkodaj), Canadian Scholar's Press, Toronto, **2**, 202-204.

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* (Eds. M.T. Chao and P.E. Cheng), Institute of Statistical Science, Academia Sinica, 447-460.

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

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