SINGLE-STAGE BAYES AND EMPIRICAL BAYES RULES FOR RANKING AND ESTIMATING MULTINOMIAL PROBABILITIES

by

Shanti S. Gupta and Sayaji N. Hande
Purdue University

Technical Report # 91-14

Department of Statistics
Purdue University

March 1991
Revised May 1991
Revised February 1992

Single-Sample Bayes and Empirical Bayes Rules for Ranking and Estimating Multinomial Probabilities*

Shanti S. Gupta and Sayaji N. Hande
Department of Statistics, Purdue University

Abstract

The problem of selecting the most probable cell, more generally the problem of partitioning k multinomial cells according to the values of the cell probabilities is considered. Bayes rules are obtained for the general ranking problem under a general class of loss functions. A sequence of parametric empirical Bayes selection rules is proposed and shown to be asymptotically optimal of order $O(e^{-cn})$ for some positive constant c. Also the problem of selecting the most (least) probable cell and simultaneously estimating the associated probability of the selected cell is considered. For this problem also a sequence of parametric empirical Bayes rules is proposed and shown to be asymptotically optimal of order $O(n^{-1})$.

Key Words: Selection and ranking, Simultaneous selection and estimation, Bayes rules, Empirical Bayes rules.
AMS 1985 subject classification: 62F07, 62C12.

1 Introduction

Selection and ranking problems arise in many practical situations where the tests of homogeneity do not provide the answer the experimenter wants. In this paper, we study several problems of selection and ranking of cell probabilities of a multinomial distribution. We consider a multinomial population with k (≥ 2) cells. Let $\Pi_1, \Pi_2, \ldots, \Pi_k$ be the cells with associated probability vector $\underset{\sim}{p} = (p_1, \ldots, p_k)$, where $\sum_{i=1}^{k} p_i = 1$, $p_i \geq 0$, for $i = 1, \ldots, k$. Let $p_{[1]} \leq p_{[2]} \ldots \leq p_{[k]}$ denote the ordered values of the parameters. The cell associated with $p_{[k]}$ is called the most probable cell and the cell associated with $p_{[1]}$ is called the least probable cell.

Single-stage and sequential selection rules for multinomial distribution have earlier been studied in the literature. For selecting the most (least) probable cell, Bechhofer, Elmaghraby and Morse (1959) have considered a single-sample procedure through the indifference zone approach. Gupta and Nagel (1967), Panchapakesan (1971) and Gupta and Huang (1975) have studied this selection problem using the subset selection approach. Sequential selection procedures have been investigated by Cacoullos and Sobel (1966), Alam (1971), Alam, Seo and Thompson (1971), Ramey and Alam (1979, 1980), Bechhofer and Kulkarni (1984), among others.

*This research was supported in part by NSF Grants DMS-8702620, DMS-8923071, and DMS-8717799 at Purdue University.

In this paper we first consider the general problem of partitioning the k cells into r non-empty, mutually exclusive subsets, $S_1, S_2, \ldots, S_r$, where S_1 is of size t_1, S_2 is of size t_2 and so on. Based on statistical observations we would like to partition the k cells into these r subsets, so that S_1 is the set of the t_1 least probable cells, S_2 is the set of the next t_2 least probable cells ... and S_r is the set of the t_r most probable cells. When $r = 2$ and $t_2 = 1$, this reduces to the problem of selecting the most probable cell.

In many situations, an experimenter may have some prior information about the parameter of interest, and would like to use that information for making an appropriate decision. In such cases, one may wish to use the Bayesian approach. In Section 2, the problem is described and a Bayes rule is derived, assuming that the unknown parameter $\underset{\sim}{p}$ has a conjugate prior. In Section 3, it is assumed that the prior is partially known. In this case we consider the situation where one is repeatedly dealing with the same selection problem independently. In such instances, at each stage, one would like to use the past information to derive a rule which is close to a Bayes rule. This approach is known as the empirical Bayes and is due to Robbins (1956, 1964). Empirical Bayes rules have been derived for subset selection goals by Deely (1965), for selecting good populations by Gupta and Hsiao (1983) and by Gupta and Leu (1983). Gupta and Liang (1986, 1988) have considered the problem of selecting good populations and the problem of selecting the best binomial population using the empirical Bayes approach.

Gupta and Liang (1989) have derived a sequence of parametric empirical Bayes rules for selecting the most (least) probable cell. Their loss function for selecting the most probable cell is $p_{[k]} - p_i$ and for selecting the least probable cell is $p_i - p_{[1]}$. In Section 3, we generalize their results and prove that the sequence of empirical Bayes rules proposed is asymptotically optimal of order $O(e^{-cn})$ for a more general loss function, where c is some positive constant and where n denotes the number of past observations.

In Section 4, we consider the problem of selecting the most (least) probable multinomial cell and simultaneously estimating the probability associated with the selected cell in a decision-theoretic framework. A Bayes rule is derived when the unknown parameter $\underset{\sim}{p}$ has a conjugate prior. Then a sequence of parametric empirical Bayes rules is constructed and is proved to be asymptotically optimal of order $O(n^{-1})$.

2 Bayes Rules

Let X_i denote the number of counts in the cell Π_i based on N independent trials, for $i = 1, 2, \ldots, k$. Then the distribution of $\underset{\sim}{X} = (X_1, X_2, \ldots, X_k)$ is given by the probability function,

$$f(\underset{\sim}{x}|\underset{\sim}{p}) = \begin{cases} \frac{N!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k p_i^{x_i} & \text{if } 0 \le x_i \le N \ \forall \ i \text{ and } \sum_{i=1}^k x_i = N; \\ 0 & \text{otherwise.} \end{cases}$$

Let $p_{[1]} \le p_{[2]} \cdots \le p_{[k]}$ be the ordered values of the parameters $p_1, p_2, \ldots, p_k$. We assume that there is no prior knowledge about the exact pairing between the ordered and the unordered parameters. Our goal is to partition the k cells into r mutually exclusive subsets, $S_1, S_2, \ldots, S_r$ such that S_1 is the set of the t_1 cells associated with the probabilities $p_{[1]}, p_{[2]}, \ldots, p_{[t_1]}$, S_2 is the set of the cells associated with the probabilities $p_{[t_1+1]}, p_{[t_1+2]}, \ldots, p_{[t_1+t_2]}$ and S_r is the set of the cells associated with the probabilities $p_{[k-t_r+1]}, p_{[k-t_r+2]}, \ldots, p_{[k]}$. Here $t_1, t_2, \ldots, t_r$ are positive integers fixed in advance, such that $\sum_{i=1}^{r} t_i = k$.

Let $\mathcal{X}$ be the sample space of the random vector $\underset{\sim}{X} = (X_1, X_2, \ldots, X_k)$, and let

$$\mathcal{A} = \left\{ (S_1, S_2, \ldots, S_r) : \; S_i \cap S_j = \emptyset, \; \forall i \neq j \text{ and } |S_i| = t_i, \; S_i \subset \{1, 2, \ldots, k\} \; \forall i \right\}$$

be the action space and

$$\Omega = \left\{ (p_1, p_2, \ldots, p_k) : \textstyle\sum_{i=1}^{k} p_i = 1, p_i \geq 0 \;\; \forall \;\; i \right\}$$

be the parameter space.

We assume that the prior distribution of the parameter $\underset{\sim}{P}$ follows a Dirichlet distribution, $G = G(.|\underset{\sim}{\alpha})$, with the hyperparameters $\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, where $\alpha_1, \alpha_2, \ldots, \alpha_k$ are positive constants. Let $\alpha_0 = \sum_{i=1}^{k} \alpha_i$. The density of the parameter vector $\underset{\sim}{P}$ is given by

$$g(\underset{\sim}{p}) = \begin{cases} \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} p_i^{\alpha_i - 1} & \text{if } p_i \geq 0 \text{ for } i = 1, 2, \ldots, k \text{ and } \sum_{i=1}^{k} p_i = 1; \\ 0 & \text{otherwise.} \end{cases}$$

Let $L(.,.) : \Omega \times \mathcal{A} \longrightarrow \mathbf{R}$ be the loss function. We assume that the loss function is non-negative, permutation invariant and "increasing". Mathematically, we write these conditions as:

[C1] $L(\underset{\sim}{p}, a) \geq 0 \qquad \forall \quad \underset{\sim}{p} \in \Omega$ and $a \in \mathcal{A}$.

[C2] $L(\pi \underset{\sim}{p}, \pi a) = L(\underset{\sim}{p}, a) \qquad \forall \quad \underset{\sim}{p} \in \Omega; \quad a \in \mathcal{A}$ and for each permutation π.

Here $\pi(a) = (\pi S_1, \pi S_2, \ldots, \pi S_r)$ and for any $S \subset \{1, 2, \ldots, k\}$, $S = \{i_1, i_2, \ldots, i_q\}$ then $\pi S = \{\pi(i_1), \pi(i_2), \ldots, \pi(i_q)\}$.

[C3] Let $\underset{\sim}{p} = (p_1, p_2, \ldots, p_k)$ and $p_i < p_j$; $a \in \mathcal{A}$, $a = (S_1, S_2, \ldots, S_r)$ and $j \in S_{r_1}, i \in S_{r_2}$, $r_1 < r_2$. Let π_{ij} be the permutation which interchanges the ith and jth co-ordinates then,

$$L(\underset{\sim}{p}, \pi_{ij} a) < L(\underset{\sim}{p}, a).$$

[C4] There exists a positive constant η, such that for every $a \in \mathcal{A}$

$$\int L^{1+\eta}(\underset{\sim}{p}, a) g(\underset{\sim}{p}) d\underset{\sim}{p} < \infty.$$

Condition [C1] says that the loss function is non-negative; [C2] says it is permutation invariant; [C3] implies that the loss function is "monotone". The $(1+\eta)$-integrability in [C4] is used later to prove the asymptotic optimality of the proposed empirical Bayes selection rules.

Let $\underset{\sim}{X} = \underset{\sim}{x}$ be the observed vector. Now we prove that any decision rule that "ranks" the cell Π_i according to the rank of $x_i + \alpha_i$ in $x_1 + \alpha_1, x_2 + \alpha_2, \ldots, x_k + \alpha_k$ is a Bayes rule.

Theorem 2.1 *:*

Let $\underset{\sim}{X}$ have a multinomial distribution with parameters $(N, \underset{\sim}{p})$. For the loss function as given in C1-C4 above and for the Dirichlet prior $G(.|\underset{\sim}{\alpha})$ on $\underset{\sim}{P}$, we consider the decision rule, $\delta(\underset{\sim}{x}) = (S_1(\underset{\sim}{x}), S_2(\underset{\sim}{x}), \ldots, S_r(\underset{\sim}{x}))$, which is such that for every $r_2 > r_1$, and every $i \in S_{r_1}$ and $j \in S_{r_2}$, $x_i + \alpha_i \leq x_j + \alpha_j$. Then $\delta(\underset{\sim}{x})$ is a Bayes rule.

Proof: Let B be the class of non-randomized decision rules defined as follows:

$$B = \left\{ \delta(\underset{\sim}{x}) = (S_1(\underset{\sim}{x}), \ldots, S_r(\underset{\sim}{x})) : \begin{array}{c} \text{for every } 1 \le r_1 < r_2 \le r \\ \min_{i \in S_{r_2}(\underset{\sim}{x})}(x_i + \alpha_i) \ge \max_{j \in S_{r_1}(\underset{\sim}{x})}(x_j + \alpha_j), \quad \forall \underset{\sim}{x} \in \mathcal{X} \end{array} \right\}.$$

To find the Bayes rules, it is enough to consider the non-randomized rules. First we prove that if a non-randomized rule δ is not in B, then there is a rule in B which has smaller Bayes risk than the Bayes risk of δ. Then we prove that all decision rules in the class B have the same Bayes risk, and this would complete the proof of the theorem.

For a decision rule δ, let $r(\delta)$ be the associated Bayes risk and $r(\delta|\underset{\sim}{x})$ be the posterior expected loss for given $\underset{\sim}{X} = \underset{\sim}{x}$. Then

$$r(\delta|\underset{\sim}{x}) = \int L(\underset{\sim}{p}, \delta(\underset{\sim}{x}))g(\underset{\sim}{p}|\underset{\sim}{x})d\underset{\sim}{p} \quad \text{and} \quad r(\delta) = E[r(\delta|\underset{\sim}{X})],$$

where $g(\underset{\sim}{p}|\underset{\sim}{x})$ is the posterior density of $\underset{\sim}{P}$ for given $\underset{\sim}{X} = \underset{\sim}{x}$.

We assume that the decision rule δ is not in the class B. Thus, there exist $\underset{\sim}{x}^0 = (x_1^0, x_2^0, \ldots, x_k^0) \in \mathcal{X}$ and $r_1, r_2, \quad 1 \le r_1 < r_2 \le r$ such that $x_i^0 + \alpha_i > x_j^0 + \alpha_j$ for some $i \in S_{r_1}(\underset{\sim}{x}^0)$ and $j \in S_{r_2}(\underset{\sim}{x}^0)$.

Next we construct a decision rule δ' whose Bayes risk is smaller than the Bayes risk of δ. Let $\delta'(\underset{\sim}{x}) = (S_1'(\underset{\sim}{x}), S_2'(\underset{\sim}{x}), \ldots, S_r'(\underset{\sim}{x}))$

$$\begin{aligned} S_m'(\underset{\sim}{x}^0) &= S_m(\underset{\sim}{x}^0) \quad \forall\ m \ne r_1 \quad \text{and} \quad m \ne r_2 \\ S_{r_1}'(\underset{\sim}{x}^0) &= (S_{r_1}(\underset{\sim}{x}^0) - \{i\}) \bigcup \{j\} \\ S_{r_2}'(\underset{\sim}{x}^0) &= (S_{r_2}(\underset{\sim}{x}^0) - \{j\}) \bigcup \{i\}, \end{aligned}$$

and $\delta'(\underset{\sim}{x}) = \delta(\underset{\sim}{x}) \quad \forall\ \underset{\sim}{x} \ne \underset{\sim}{x}^0$.

To prove that the decision rule δ' has smaller Bayes risk than the decision rule δ, it is enough to prove that, $r(\delta|\underset{\sim}{x}) - r(\delta'|\underset{\sim}{x}) \ge 0$ for all $\underset{\sim}{x}$. For $\underset{\sim}{x} \ne \underset{\sim}{x}^0$, $\delta(\underset{\sim}{x}) = \delta'(\underset{\sim}{x})$ thus, $r(\delta|\underset{\sim}{x}) = r(\delta'|\underset{\sim}{x})$. The posterior density of $\underset{\sim}{P}$ is given by,

$$\begin{aligned} g(\underset{\sim}{p}|\underset{\sim}{x}) &= h(\underset{\sim}{x}) \prod_{t=1}^{k} p_t^{x_t + \alpha_t - 1}, \\ \text{where} \quad h(\underset{\sim}{x}) &= \frac{\Gamma(\alpha_0 + N)}{\prod_{t=1}^{k} \Gamma(x_t + \alpha_t)}, \end{aligned}$$

and the posterior expected loss of the rule δ is given by,

$$r(\delta|\underset{\sim}{x}) = h(\underset{\sim}{x}) \int L(\underset{\sim}{p}, \delta(\underset{\sim}{x})) \prod_{t=1}^{k} p_t^{x_t + \alpha_t - 1} d\underset{\sim}{p}.$$

Hence

$$\begin{aligned} r(\delta|\underset{\sim}{x}^0) - r(\delta'|\underset{\sim}{x}^0) &= h(\underset{\sim}{x}^0) \int_{p_i > p_j} [L(\underset{\sim}{p}, \delta(\underset{\sim}{x}^0)) - L(\underset{\sim}{p}, \delta'(\underset{\sim}{x}^0))] \prod_{t=1}^{k} p_t^{x_t^0 + \alpha_t - 1} d\underset{\sim}{p} \\ &\quad + h(\underset{\sim}{x}^0) \int_{p_j > p_i} [L(\underset{\sim}{p}, \delta(\underset{\sim}{x}^0)) - L(\underset{\sim}{p}, \delta'(\underset{\sim}{x}^0))] \prod_{t=1}^{k} p_t^{x_t^0 + \alpha_t - 1} d\underset{\sim}{p}. \end{aligned}$$

Since L is invariant ([C2]), interchanging the variables p_i and p_j inside the integral, we get,

$$h(\underset{\sim}{x}^0)\int_{p_j>p_i}[L(\underset{\sim}{p},\delta(\underset{\sim}{x}^0)) - L(\underset{\sim}{p},\delta'(\underset{\sim}{x}^0))]\prod_{t=1}^{k}p_t^{x_t^0+\alpha_t-1}\,d\underset{\sim}{p}$$
$$= -h(\underset{\sim}{x}^0)\int_{p_i>p_j}[L(\underset{\sim}{p},\delta(\underset{\sim}{x}^0)) - L(\underset{\sim}{p},\delta'(\underset{\sim}{x}^0))][\prod_{t\neq i,t\neq j}p_t^{x_t^0+\alpha_t-1}]p_j^{x_i^0+\alpha_i-1}p_i^{x_j^0+\alpha_j-1}\,d\underset{\sim}{p}.$$

This implies

$$r(\delta|\underset{\sim}{x}^0) - r(\delta'|\underset{\sim}{x}^0)$$
$$= h(\underset{\sim}{x}^0)\int_{p_i>p_j}[L(\underset{\sim}{p},\delta(\underset{\sim}{x}^0)) - L(\underset{\sim}{p},\delta'(\underset{\sim}{x}^0))][p_i^{x_i^0+\alpha_i-1}p_j^{x_j^0+\alpha_j-1} - p_i^{x_j^0+\alpha_j-1}p_j^{x_i^0+\alpha_i-1}]$$
$$\prod_{t\neq i\neq j}p_t^{x_t^0+\alpha_t-1}d\underset{\sim}{p}\ .$$

Since $p_i > p_j$, it follows from [C3] that

$$L(\underset{\sim}{p},\delta(\underset{\sim}{x}^0)) - L(\underset{\sim}{p},\delta'(\underset{\sim}{x}^0)) > 0\ .$$

Also for $x_i^0 + \alpha_i > x_j^0 + \alpha_j$, and $p_i > p_j$ we have

$$p_i^{x_i^0+\alpha_i-1}p_j^{x_j^0+\alpha_j-1} - p_i^{x_j^0+\alpha_j-1}p_j^{x_i^0+\alpha_i-1} > 0.$$

From the above inequalities we get,

$$r(\delta|\underset{\sim}{x}^0) - r(\delta'|\underset{\sim}{x}^0) > 0.$$

Hence $r(\delta) > r(\delta')$. Since there are only a finite number of elements in $\mathcal{X}$, by the same method we can obtain a decision rule in B which has smaller Bayes risk than δ. It is straightforward to see that all the rules in B have the same Bayes risk. And this completes the proof. □

Remark 2.1 :
It should be noted that the Bayes rule is not necessarily unique.

Remark 2.2 :
The class of decision rules B, described in the theorem above, contains all the non-randomized Bayes rules.

3 Empirical Bayes rules

Now consider a situation in which one repeatedly deals with the same ranking problem independently, and assume that the prior is partially known. In such situations, one can use the empirical Bayes approach of Robbins (1956, 1964). Using this approach Gupta and

Liang (1989) have derived parametric empirical Bayes rules for selecting the most (least) probable cell. Their loss function for selecting the most probable cell is $p_{[k]} - p_i$ and for selecting the least probable cell is $p_i - p_{[1]}$. These loss functions do satisfy all the conditions mentioned in Section 2.

We assume that the prior is still a Dirichlet prior, but the hyperparameters $\alpha_1, \alpha_2, \ldots, \alpha_k$ are unknown. We consider two cases, (1) α_0 is known and (2) α_0 is unknown.

Let $\underset{\sim}{X}_j = (X_{1j}, X_{2j}, \ldots, X_{kj})$ denote the observable random vector at stage j and $\underset{\sim}{P}_j$ be the unobservable probability vector, for $j = 1, 2, 3, \ldots\ldots$. Here X_{ij} represents the number of counts in the ith cell at jth stage. Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n$ be the past available observations, and $\underset{\sim}{X}_{n+1} = \underset{\sim}{X} = (X_1, X_2, \ldots, X_k)$ denote the present observation.

Notice that

$$EX_{ij} = \frac{N\alpha_i}{\alpha_0}, \quad \forall\ i = 1, 2, \ldots, k \text{ and } \forall\ j = 1, 2, \ldots$$

For the case when α_0 is known, define

$$\hat{\alpha}_{in} = N^{-1}\, \alpha_0\, \bar{X}_{in}, \tag{1}$$

where, $\bar{X}_{in} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$. Thus for every i, $E\hat{\alpha}_{in} = \alpha_i$.

When α_0 is not known, to estimate $\alpha_1, \alpha_2, \ldots, \alpha_k$ we need to consider the higher moments. Following the notations in Gupta and Liang (1989), define

$$\bar{X}_{in} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}, \quad M_{in} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}^2, \quad Z_{in} = [N\bar{X}_{in} - M_{in}]\bar{X}_{in}, \tag{2}$$

$$Y_{in} = [M_{in} - \bar{X}_{in}]N - (N-1)\bar{X}_{in}^2. \tag{3}$$

Let $\mu_{i1} = E\bar{X}_{in} = N\alpha_i\alpha_0^{-1}$, $\mu_{i2} = EM_{in} = N\alpha_i\alpha_0^{-1} + N(N-1)\alpha_i(\alpha_i+1)\alpha_0^{-1}(\alpha_0+1)^{-1}$. Then, Z_{in}, Y_{in} and Z_{in}/Y_{in} are the moment estimators of L_{i1}, L_{i2}, and $\alpha_i = L_{i1}/L_{i2}$, where $L_{i1} = (N\mu_{i1} - \mu_{i2})\mu_{i1}$ and $L_{i2} = (\mu_{i2} - \mu_{i1})N - (N-1)\mu_{i1}^2$.

For the case when $\alpha_0, \alpha_1, \ldots, \alpha_k$ are unknown, define

$$\hat{\alpha}'_{in} = \begin{cases} Z_{in}/Y_{in} & \text{if } Y_{in} > 0; \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Now, for Case 1, we propose the sequence of the empirical Bayes rules $\{\delta_n\}_1^\infty$ which "ranks" the ith cell according to the rank of $x_i + \hat{\alpha}_{in}$ in $x_1 + \hat{\alpha}_{1n}, x_2 + \hat{\alpha}_{2n}, \ldots, x_k + \hat{\alpha}_{kn}$. Let

$$\delta_n(\underset{\sim}{x}) = (S_{1n}(\underset{\sim}{x}), S_{2n}(\underset{\sim}{x}), \ldots, S_{rn}(\underset{\sim}{x})),$$

where $S_{1n}(\underset{\sim}{x})$ contains the cells corresponding to $(\underset{\sim}{x} + \hat{\alpha})_{[1]}, (\underset{\sim}{x} + \hat{\alpha})_{[2]}, \ldots, (\underset{\sim}{x} + \hat{\alpha})_{[t_1]}$, and so on. Here $(\underset{\sim}{x} + \hat{\alpha})_{[1]}, (\underset{\sim}{x} + \hat{\alpha})_{[2]}, \ldots, (\underset{\sim}{x} + \hat{\alpha})_{[k]}$ are the ordered values of $x_1 + \hat{\alpha}_{1n}, x_2 + \hat{\alpha}_{2n}, \ldots, x_k + \hat{\alpha}_{kn}$. In case of ties we use randomization. For Case 2, we propose the sequence of the empirical Bayes rules $\{\delta'_n\}_1^\infty$, which ranks the ith cell according to the rank of $x_i + \hat{\alpha}'_{in}$ in $x_1 + \hat{\alpha}'_{1n}, x_2 + \hat{\alpha}'_{2n}, \ldots, x_k + \hat{\alpha}'_{kn}$. The optimality of a sequence of empirical Bayes rules can be judged by considering how large its Bayes risk is as compared to the minimum Bayes risk at the nth stage. In this connection, we define,

Definition 3.1 :- *A sequence of empirical Bayes rules* $\{\delta_n\}$ *is said to be asymptotically optimal at least of order* β_n *relative to the prior distribution* G *if*

$$r(G,\delta_n) - r(G) \leq O(\beta_n) \quad as \quad n \to \infty,$$

where $r(G)$ *is the minimum Bayes risk,* $r(G,\delta)$ *is the Bayes risk of the rule* δ*, and* $\{\beta_n\}$ *is a sequence of positive numbers such that* $\lim_{n\to\infty}\beta_n = 0$.

Theorem 3.1 :
The sequence of empirical Bayes rules $\{\delta_n\}_1^\infty$ *defined above is asymptotically optimal of order* $O(e^{-cn})$ *for some positive constant c.*

For proving the theorem we need the following lemma due to Hoeffding (1963).

Lemma 3.1 *(Hoeffding):*
If $Y_1, Y_2, \ldots, Y_n$ *are the independent random variables such that for each* i *there exists real numbers* a_i *and* b_i *such that* $P(a_i \leq Y_i \leq b_i) = 1$ *then for any* $t > 0$

$$P(\bar{Y} - \mu \ \geq \ t) \leq e^{-2n^2t^2(\sum_{i=1}^n (b_i-a_i)^2)^{-1}},$$

where $\bar{Y} = \sum_{i=1}^n Y_i/n$ *and* $\mu = E\bar{Y}$.

Proof of Theorem 3.1: Let $\underset{\approx}{x} = (\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_{n+1})$. For $1 \leq l \neq t \leq k$, define

$$A_{lt}^n = \left\{ \underset{\approx}{x} : \begin{array}{c} x_l + \hat{\alpha}_{ln} \geq x_t + \hat{\alpha}_{tn} \\ x_l + \alpha_l < x_t + \alpha_t \end{array} \right\}.$$

Let δ_G be a Bayes rule and let $r(G)$ be its Bayes risk. Then

$$\begin{aligned} & r(G,\delta_n) - r(G) \\ = & \sum_{\underset{\approx}{x}\in\mathcal{X}^{n+1}} \int [L(\underset{\sim}{p}_{n+1}, \delta_n(\underset{\sim}{x}_{n+1})) - L(\underset{\sim}{p}_{n+1}, \delta_G(\underset{\sim}{x}_{n+1}))] \prod_{j=1}^{n+1} f(\underset{\sim}{x}_j|\underset{\sim}{p}_j)g(\underset{\sim}{p}_j)d\underset{\sim}{p}_j \\ \leq & \sum_{\underset{\approx}{x}\in\mathcal{X}^{n+1}} \int |L(\underset{\sim}{p}_{n+1}, \delta_n(\underset{\sim}{x}_{n+1})) - L(\underset{\sim}{p}_{n+1}, \delta_G(\underset{\sim}{x}_{n+1}))| I_{(\cup_{l\neq t}A_{lt}^n)}(\underset{\approx}{x}) \prod_{j=1}^{n+1} \ f(\underset{\sim}{x}_j|\underset{\sim}{p}_j)g(\underset{\sim}{p}_j) \ d\underset{\sim}{p}_j, \end{aligned}$$

where $I_D(.)$ denotes an indicator function of the set D.

Now using the Holder's inequality we have

$$\begin{aligned} & r(G,\delta_n) - r(G) \\ & \leq \left(\sum_{\underset{\approx}{x}\in\mathcal{X}^{n+1}} \int |L(\underset{\sim}{p}_{n+1}, \delta_n(\underset{\approx}{x})) - L(\underset{\sim}{p}_{n+1}, \delta_G(\underset{\sim}{x}_{n+1}))|^{1+\eta} \prod_{j=1}^{n+1} f(\underset{\sim}{x}_j|\underset{\sim}{p}_j)g(\underset{\sim}{p}_j) \ d\underset{\sim}{p}_j \right)^{\frac{1}{1+\eta}} \\ & \times \left(\sum_{\underset{\approx}{x}\in\mathcal{X}^{n+1}} \int I_{\cup_{l\neq t}A_{lt}^n}(\underset{\approx}{x}) \prod_{j=1}^{n+1} f(\underset{\sim}{x}_j|\underset{\sim}{p}_j)g(\underset{\sim}{p}_j)d\underset{\sim}{p}_j \right)^{\frac{\eta}{1+\eta}}. \end{aligned}$$

Also,

$$\sum_{\underset{\sim}{x}\in\mathcal{X}^{n+1}} \int |L(\underset{\sim}{p}_{n+1}, \delta_n(\underset{\sim}{x}_{n+1})) - L(\underset{\sim}{p}_{n+1}, \delta_G(\underset{\sim}{x}_{n+1}))|^{1+\eta} \prod_{j=1}^{n+1} f(\underset{\sim}{x}_j|\underset{\sim}{p}_j) g(\underset{\sim}{p}_j)\ d\underset{\sim}{p}_j$$
$$\le 2^{1+\eta} \int [\sup_a L(\underset{\sim}{p}, a)]^{1+\eta} g(\underset{\sim}{p}) d\underset{\sim}{p}.$$

The supremum is taken over all $a \in \mathcal{A}$. Since $\mathcal{A}$ is finite and by the Assumption [C4] on the loss function, the right hand side of the above inequality is finite, say m_0.

$$\begin{aligned}\sum_{\underset{\sim}{x}\in\mathcal{X}^{n+1}} \int I_{\cup_{l\neq t} A_{lt}^n}(\underset{\sim}{x}) \prod_{j=1}^{n+1} f(\underset{\sim}{x}_j|\underset{\sim}{p}_j)\ g(\underset{\sim}{p}_j)\ d\underset{\sim}{p}_j &\le \sum_{l\neq t} \sum_{\underset{\sim}{x}\in\mathcal{X}^{n+1}} \int I_{A_{lt}^n}(\underset{\sim}{x}) \prod_{j=1}^{n+1} f(\underset{\sim}{x}_j|\underset{\sim}{p}_j) g(\underset{\sim}{p}_j)\ d\underset{\sim}{p}_j \\ &= \sum_{l\neq t} P(A_{lt}^n).\end{aligned}$$

Let

$$\epsilon' = \min\{|x_l + \alpha_l - x_t - \alpha_t| : |x_l + \alpha_l - x_t - \alpha_t| \neq 0\}.$$

Then to prove the theorem, it is enough to show that each term in $\sum_{l\neq t} P(A_{lt}^n)$ is $O(e^{-cn})$. Without loss of generality, let us assume that $l = 1$ and $t = 2$. Then,

$$\begin{aligned} & P(A_{12}^n) \\ = \ & P(X_{n+1,1} + \hat{\alpha}_{1n} \ge X_{n+1,2} + \hat{\alpha}_{2n} \ \text{and}\ X_{n+1,1} + \alpha_1 < X_{n+1,2} + \alpha_2) \\ \le \ & P(\hat{\alpha}_{2n} - \hat{\alpha}_{1n} - \alpha_2 + \alpha_1 < -\epsilon) \qquad (\epsilon = \epsilon'/2) \\ = \ & P\left(\frac{1}{n}\sum_{j=1}^{n}[\frac{(X_{2j} - X_{1j})}{N} - \frac{1}{\alpha_0}[\alpha_2 - \alpha_1]] < -\epsilon\right) \\ \le \ & O(e^{-cn}).\end{aligned}$$

The last inequality follows from Hoeffding's result (Lemma 3.1). Here $c = 2\epsilon^2[2/N + (\alpha_1 + \alpha_2)/\alpha_0]^{-1}$. This proves the theorem. □

Now we establish the asymptotic optimality of the sequence of empirical Bayes rules $\{\delta_n'\}_1^\infty$.

Theorem 3.2 *:*
The sequence of empirical Bayes rules $\{\delta_n'\}_1^\infty$ is asymptotically optimal of order $O(e^{-cn})$ for some positive constant c.

We need the following *lemma* due to Gupta and Liang (1989).

Lemma 3.2 *:*
Let b be a positive constant. Then,
a)$P(Z_{in} - L_{i1} < -b) \le 3e^{-b_i n} = O(e^{-b_i n})$; *b)$P(Z_{in} - L_{i1} > b) \le 3e^{-b_i n} = O(e^{-b_i n})$;*
c)$P(Y_{in} - L_{i2} < -b) \le 3e^{-b_i n} = O(e^{-b_i n})$; *d)$P(Y_{in} - L_{i2} > b) \le 3e^{-b_i n} = O(e^{-b_i n})$;*
where $b_i = b^2[2N^4(N + \mu_{i1})^2]^{-1} > 0$.

Note that it follows from the above lemma that for a given $\epsilon > 0$, there exists a positive constant c_1 such that $P(|Z_{in} - L_{i1}| > \epsilon) \leq O(e^{-c_1 n})$ and $P(|Y_{in} - L_{i2}| > \epsilon) \leq O(e^{-c_1 n})$.

Before proving the theorem we also need the following lemma.

Lemma 3.3 *For every $i = 1, 2, \ldots, k$ and for every $\epsilon > 0$, there exists a constant $c > 0$ such that*

$$P(\hat{\alpha}'_{in} - \alpha_i > \epsilon) \leq O(e^{-cn}) \ \textit{and} \ P(\hat{\alpha}'_{in} - \alpha_i < -\epsilon) \leq O(e^{-cn}).$$

Proof: We show that $P(\hat{\alpha}'_{in} - \alpha_i > \epsilon) \leq O(e^{-cn})$ for some $c > 0$. The same technique can be used on the other term to prove that it is also of order $O(e^{-cn})$.

Note that, by the definition of $\hat{\alpha}'_{in}$,

$$\begin{aligned}
& P(\hat{\alpha}'_{in} - \alpha_i > \epsilon) \\
= \ & P(\hat{\alpha}'_{in} - \alpha_i > \epsilon, \ Y_{in} > 0) + P(\hat{\alpha}'_{in} - \alpha_i > \epsilon, \ Y_{in} \leq 0) \\
\leq \ & P(Z_{in} - (\alpha_i + \epsilon)Y_{in} > 0) + P(Y_{in} \leq 0) \\
= \ & P\left((Z_{in} - L_{i1}) - (\alpha_i + \epsilon)(Y_{in} - L_{i2}) > [(\alpha_i + \epsilon)L_{i2} - L_{i1}]\right) + P(Y_{in} \leq 0) \\
= \ & P\left(Z_{in} - L_{i1} > \frac{(\alpha_i + \epsilon)L_{i2} - L_{i1}}{2}\right) \\
& + P\left((\alpha_i + \epsilon)(Y_{in} - L_{i2}) < -\frac{[(\alpha_i + \epsilon)L_{i2} - L_{i1}]}{2}\right) + P(Y_{in} \leq 0).
\end{aligned}$$

Since $[(\alpha_i + \epsilon)L_{i2} - L_{i1}]/2 > 0$, the result follows from Lemma 3.2. □

Now we proceed to prove Theorem 3.2.

Proof of Theorem 3.2:

As in Theorem 3.1, it suffices to prove that for each $1 \leq t, \ l \leq k$, $t \neq l$,

$$P(X_{n+1,l} + \alpha_l \geq X_{n+1,t} + \alpha_t \ \text{ and } \ X_{n+1,l} + \hat{\alpha}'_{ln} < X_{n+1,t} + \hat{\alpha}'_{tn}) \ \leq \ O(e^{-cn}).$$

Without loss of generality, let us assume $l = 1$ and $t = 2$. Then we only need to prove that $P(X_1 + \alpha_1 \ \geq \ X_2 + \alpha_2, \quad X_1 + \hat{\alpha}'_{1n} \ < \ X_2 + \hat{\alpha}'_{2n})$ is of order $O(e^{-cn})$ for some positive constant c. Let $\epsilon' = \min\{|x_l + \alpha_l - x_t - \alpha_t| : |x_l + \alpha_l - x_t - \alpha_t| \neq 0\}$, then

$$\begin{aligned}
& P(X_1 + \alpha_1 \ \geq \ X_2 + \alpha_2 \quad \text{and} \quad X_1 + \hat{\alpha}'_{1n} \ < \ X_2 + \hat{\alpha}'_{2n}) \\
\leq \ & P\left([\hat{\alpha}'_{2n} - \hat{\alpha}'_{1n}] \ - [\alpha_2 - \alpha_1] > \epsilon\right) \qquad\qquad (\epsilon = \epsilon'/2) \\
\leq \ & P\left(\hat{\alpha}'_{2n} - \alpha_2 > \frac{\epsilon}{2}\right) + P\left(\hat{\alpha}'_{1n} - \alpha_1 < -\frac{\epsilon}{2}\right).
\end{aligned}$$

Now the result follows from Lemma 3.3. This completes the proof of the theorem. □

4 Simultaneous Selection and Estimation

In this section we consider the problem of selecting the most (least) probable multinomial cell and simultaneously estimating the probability associated with the selected cell. Cohen and Sackrowitz (1988) have considered the problem of selecting the population associated with the largest mean and simultaneously estimating the mean of the selected population. They derived results for the case where $k = 2$ and the distributions associated with the populations are normal or uniform. Gupta and Miescke (1990a) extended their results for the case $k \geq 3$. Recently Gupta and Miescke (1990b) solved the problem of selection and estimation for the best binomial population under general decision-theoretic framework.

First we consider the problem of selecting the best cell, the cell associated with $p_{[k]} = \max p_i$ and simultaneously estimating the probability associated with the selected cell. Action space for this decision problem is represented by

$$\mathcal{A} = \{(s,p) : s \in \{1,2,\ldots,k\}, p \in [0,1]\}.$$

An action $a = (s,p)$ represents the selection of sth cell and p as the estimate of the probability of the selected cell. We assume that the loss function is of the following form,

$$L(\underset{\sim}{p},(s,p)) = L_1(\underset{\sim}{p},s) + L_2(p_s,p), \tag{5}$$

where $L_1(\underset{\sim}{p},s)$ denotes the loss due to selecting the cell Π_s as the best cell and $L_2(p_s,p)$ denotes the loss due to estimating the cell probability p_s by p. As in the previous section we assume that the probability vector $\underset{\sim}{P}$ has a Dirichlet prior. As has been pointed out by Gupta and Miescke (1990a, 1990b), the decision theoretic treatment leads to "selection after estimation" rather than "estimation after selection". This phenomenon also holds in the present case for the general loss function (5).

Lemma 4.1 :
Let $\hat{p}_i(\underset{\sim}{x})$ minimize $E[L_2(P_i,\hat{p}_i)|\underset{\sim}{X} = \underset{\sim}{x}]$ for $i = 1,2,\ldots,k$. Let $s^*(\underset{\sim}{x})$ which minimizes $E[L_1(\underset{\sim}{P},s) + L_2(P_s,\hat{p}_s)|\underset{\sim}{X} = \underset{\sim}{x}]$ for $s = 1,2,\ldots,k$. Then the Bayes rules at $\underset{\sim}{X} = \underset{\sim}{x}$ is

$$\delta(\underset{\sim}{x}) = (s^*(\underset{\sim}{x}), \hat{p}_{s^*(\underset{\sim}{x})}(\underset{\sim}{x})).$$

In general, it is hard to find the Bayes rule for a general loss function. We consider the following specific loss function. Let $L_1(\underset{\sim}{p},s) = p_{[k]} - p_s$ and $L_2(p_s,p) = c_0(p_s - p)^2$, where c_0 is a positive known constant. For each $i = 1,2,\ldots,k$, it is easy to see that $\hat{p}_i(\underset{\sim}{x}) = (\alpha_i + x_i)/(\alpha_0 + N)$. Let a_i' be the posterior expected loss if the ith population is selected and if $\hat{p}_i(\underset{\sim}{x})$ is used to estimate the probability associated with that cell. Hence

$$a_i' = c_0 \frac{(\alpha_0 + N - \alpha_i - x_i)(\alpha_i + x_i)}{(\alpha_0 + N)^2(\alpha_0 + N + 1)} - \frac{\alpha_i + x_i}{\alpha_0 + N} + C,$$

where $C = E\{\max_{1 \leq j \leq k} P_j | \underset{\sim}{X} = \underset{\sim}{x}\}$. Let $a_i = a_i' - C$. The above lemma and the discussion leads to the following theorem.

Theorem 4.1 :
The Bayes decision rule selects the cell π_i for which $a_i = \min_j a_j$. The Bayes estimate of the probability of the ith cell is given by $\hat{p}_i(\underset{\sim}{x}) = (\alpha_i + x_i)/(\alpha_0 + N)$.

In many situations we may not know the parameters of the prior but we may have data or information from past experience. To derive a sequence of empirical Bayes rules we need to get the estimates of the parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$. We define the estimates as in Section

3. Let $\{\delta_n^{se_1}\}_1^\infty$; $\delta_n^{se_1}(\underset{\sim}{x}) = (s_n^*(\underset{\sim}{x}), \hat{p}_{s_n^*(\underset{\sim}{x})})$, be the sequence of empirical Bayes rules defined as follows:
For the case when α_0 is known we define, $s_n^*(\underset{\sim}{x}) \in \{1, 2, \ldots, k\}$ which minimizes

$$\hat{a}_{in} = c_0 \frac{(\alpha_0 + N - \hat{\alpha}_{in} - x_i)(\hat{\alpha}_{in} + x_i)}{(\alpha_0 + N)^2(\alpha_0 + N + 1)} - \frac{\hat{\alpha}_{in} + x_i}{\alpha_0 + N}$$

and $\hat{p}_{in}(\underset{\sim}{x}) = (\hat{\alpha}_{in} + x_i)/(\alpha_0 + N)$. For the case when α_0 is not known we define, $s_n^*(\underset{\sim}{x}) \in \{1, 2, \ldots, k\}$ which minimizes

$$\hat{a}_{in} = c_0 \frac{(\hat{\alpha}'_{0n} + N - \hat{\alpha}'_{in} - x_i)(\hat{\alpha}'_{in} + x_i)}{(\hat{\alpha}'_{0n} + N)^2(\hat{\alpha}'_{0n} + N + 1)} - \frac{\hat{\alpha}'_{in} + x_i}{\hat{\alpha}'_{0n} + N}$$

and $\hat{p}_{in}(\underset{\sim}{x}) = (\hat{\alpha}'_{in} + x_i)/(\hat{\alpha}'_{0n} + N)$, where $\hat{\alpha}'_{0n} = \sum_{i=1}^k \hat{\alpha}'_{in}$.

The optimality of a sequence of empirical Bayes rules can be judged by considering how large its Bayes risk is as compared to the minimum Bayes risk at the nth stage. In this connection, we define,

Definition 4.1 :- *A sequence of empirical Bayes rules* $\{\delta_n\}_1^\infty$; $\delta_n = (s_n(\underset{\sim}{x}), p_{s_n}(\underset{\sim}{x}))$ *is said to be asymptotically optimal at least of order* β_n *relative to the prior distribution* G *if*

$$r(G, \delta_n) - r(G) \leq O(\beta_n) \quad \text{as} \quad n \to \infty,$$

where $r(G)$ *is the minimum Bayes risk,* $r(G, \delta)$ *is the Bayes risk of the rule* δ*, and* $\{\beta_n\}$ *is a sequence of positive numbers such that* $\lim_{n\to\infty} \beta_n = 0$.

We have the following result about the asymptotic optimality of the sequence of the empirical Bayes rules.

Theorem 4.2 :
The sequence of empirical Bayes rules $\{\delta_n^{se_1}\}$ *defined above is asymptotically optimal of order* $O(n^{-1})$.

For proving this theorem we need the following lemmas. For sake of simplicity of notation, we write $\hat{p}_{s_n^*(\underset{\sim}{x})}$ for $\hat{p}_{s_n^*(\underset{\sim}{x})}(\underset{\sim}{x})$ and $\hat{p}_{s^*(\underset{\sim}{x})}$ for $\hat{p}_{s^*(\underset{\sim}{x})}(\underset{\sim}{x})$.

Lemma 4.2

$$E\left[I_{(s_n^*(\underset{\sim}{X}) = s^*(\underset{\sim}{X}))}[(\hat{p}_{s_n^*(\underset{\sim}{X})} - p_{s^*(\underset{\sim}{X})})^2 - (\hat{p}_{s^*(\underset{\sim}{X})} - p_{s^*(\underset{\sim}{X})})^2]\right] = E\left[I_{(s_n^*(\underset{\sim}{X}) = s^*(\underset{\sim}{X}))}(\hat{p}_{s_n^*(\underset{\sim}{X})} - \hat{p}_{s^*(\underset{\sim}{X})})^2\right], \quad (6)$$

where the above expectation E *is taken with respect to* $\underset{\approx}{X} = (\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_{n+1})$.

Proof: The proof follows from straightforward computations. □

Lemma 4.3 *For each* $y = 0, 1, \ldots, N$

$$\int_0^1 tP\left(\frac{\hat{\alpha}'_{in} + y}{\hat{\alpha}'_{0n} + N} - \frac{\alpha_i + y}{\alpha_0 + N} > t\right) dt \leq O(n^{-1}), \quad (7)$$

and

$$\int_0^1 tP\left(\frac{\hat{\alpha}'_{in} + y}{\hat{\alpha}'_{0n} + N} - \frac{\alpha_i + y}{\alpha_0 + N} < -t\right) dt \leq O(n^{-1}). \quad (8)$$

Proof: For each $y = 0, 1, \ldots, N$, and $t \in (0, 1)$,

$$
\begin{aligned}
& P\left(\frac{\hat{\alpha}'_{in} + y}{\hat{\alpha}'_{0n} + N} - \frac{\alpha_i + y}{\alpha_0 + N} > t\right) \\
= \;& P\left(\hat{\alpha}'_{in}(\alpha_0 + N) + y(\alpha_0 - \hat{\alpha}'_{0n}) - \alpha_i(\hat{\alpha}'_{0n} + N) > t(\hat{\alpha}'_{0n} + N)(\alpha_0 + N)\right) \\
\leq \;& P\left((\hat{\alpha}'_{in} - \alpha_i)(\alpha_0 + N) - (y + \alpha_i)(\hat{\alpha}'_{0n} - \alpha_0) > tN(\alpha_0 + N)\right) \\
\leq \;& P\left(\hat{\alpha}'_{in} - \alpha_i > \frac{tN}{2}\right) + P\left(\hat{\alpha}'_{0n} - \alpha_0 < -\frac{tN(\alpha_0 + N)}{2(y + \alpha_i)}\right) \\
\leq \;& P\left(\hat{\alpha}'_{in} - \alpha_i > \frac{tN}{2}\right) + \Sigma_{l=1}^{k} P\left(\hat{\alpha}'_{ln} - \alpha_l < -\frac{tN(\alpha_0 + N)}{2k(y + \alpha_i)}\right),
\end{aligned}
$$

where the last inequality is obtained by the definition of $\hat{\alpha}'_{0n}$, α_0 and an application of the Bonferroni inequality. Let $c' = \min\{N/2, [N(\alpha_0 + N)]/[2k(y + \alpha_i)]\}$ and m' such that $0 < m' < \min_i L_{i2}$. We consider the term $P(|\hat{\alpha}'_{ln} - \alpha_l > c't)$. Then as in Lemma 3.3,

$$
\begin{aligned}
P\left(|\hat{\alpha}'_{ln} - \alpha_l| > c't\right) &\leq P\left(|Z_{ln} - \alpha_l Y_{ln}| > c'm't\right) + P(Y_{ln} < m') \\
&\leq 12e^{-c''t^2 n} + 3e^{-c''(m' - L_{i2})^2 n},
\end{aligned}
$$

where $c'' = \min_l \min\{[2N^4(N+\mu_{l1})^2]^{-1},\ (2^{-1}c'm')^2[2N^4(N+\mu_{l1})^2]^{-1},\ (2^{-1}\alpha^{-1}c'm')^2[2N^4(N+\mu_{l1})^2]^{-1}\}$.

Since for any $C'_0 > 0$, $\int_0^1 te^{-C'_0 nt^2} = O(n^{-1})$. It follows that

$$
\int_0^1 tP\left(\frac{\hat{\alpha}'_{in} + y}{\hat{\alpha}'_{0n} + N} - \frac{\alpha_i + y}{\alpha_0 + N} > t\right) dt \leq O(n^{-1}).
$$

Similarly it can be shown that

$$
\int_0^1 tP\left(\frac{\hat{\alpha}'_{in} + y}{\hat{\alpha}'_{0n} + N} - \frac{\alpha_i + y}{\alpha_0 + N} < -t\right) dt \leq O(n^{-1}).
$$

This completes the proof of the lemma. □

Proof of Theorem 4.2: Let $\{\delta_n^{se_1}\}_{n=1}^{\infty}$ be the sequence of empirical Bayes rules, $\delta_n^{se_1}(\underset{\sim}{x}) = (s_n^*(\underset{\sim}{x}), p_{s_n^*(\underset{\sim}{x})}(\underset{\sim}{x}))$ and δ be the Bayes rule defined in Lemma 4.1.

$$
\begin{aligned}
r(\delta_n^{se_1}) - r(\delta) \;=\;& c_0 E\left[I_{(s_n^*(\underset{\sim}{X}) = s^*(\underset{\sim}{X}))}\left[(\hat{p}_{s_n^*(\underset{\sim}{X})} - p_{s^*(\underset{\sim}{X})})^2 - (\hat{p}_{s^*(\underset{\sim}{X})} - p_{s^*(\underset{\sim}{X})})^2\right]\right] \\
& + E\left[I_{(s_n^*(\underset{\sim}{X}) \neq s^*(\underset{\sim}{X}))}\left[c_0(\hat{p}_{s_n^*(\underset{\sim}{X})} - p_{s^*(\underset{\sim}{X})})^2 - c_0(\hat{p}_{s^*(\underset{\sim}{X})} - p_{s^*(\underset{\sim}{X})})^2 + p_{s^*(\underset{\sim}{X})} - p_{s_n^*(\underset{\sim}{x})}\right]\right] \\
\leq\;& c_0 \sum_{i=1}^{k} E(\hat{p}_{in} - \hat{p}_i)^2 + (c_0 + 1)P(s_n^*(\underset{\sim}{X}) \neq s^*(\underset{\sim}{X})), \qquad \text{(by Lemma 4.2)}.
\end{aligned}
$$

As we proved in Section 3 we can prove that $P(s_n^*(\underset{\sim}{X}) \neq s^*(\underset{\sim}{X})) \leq O(e^{-c_2 n})$ for some positive constant c_2. To prove the theorem we need to prove that $E(\hat{p}_{in} - \hat{p}_i)^2$ is of order $O(n^{-1})$ for $i = 1, 2, \ldots, k$.

Now we consider two cases: Case (1) α_0 known, and Case (2) α_0 not known.

Case 1:

In this case,

$$\hat{p}_{in} = \frac{\hat{\alpha}_{in} + X_i}{\alpha_0 + N}, \quad \hat{p}_i = \frac{\alpha_i + X_i}{\alpha_0 + N} \quad \text{and} \quad \hat{\alpha}_{in} = N^{-1}\alpha_0 \bar{X}_{in}$$

for $i = 1, 2, \ldots, k$. Hence

$$\begin{aligned} E(\hat{p}_{in} - \hat{p}_i)^2 &= \frac{1}{(\alpha_0 + N)^2} E(\hat{\alpha}_{in} - \alpha_i)^2 \\ &= \frac{\alpha_0^2}{N^2(\alpha_0 + N)^2} E(\bar{X}_{in} - EX_{i1})^2 \\ &= O(n^{-1}). \end{aligned}$$

This completes the proof for this case.

Case 2:

For the case when α_0 is not known, as in the proof of the Case 1,

$$\begin{aligned} E[\hat{p}_{in} - \hat{p}_i]^2 &= \sum_{y=0}^{N} E\left[\frac{\hat{\alpha}'_{in} + y}{\hat{\alpha}'_{0n} + N} - \frac{\alpha_i + y}{\alpha_0 + N}\right]^2 P(X_i = y) \\ &\le \sum_{y=0}^{N} \left[2\int_0^1 tP\left(\frac{\hat{\alpha}_{in} + y}{\hat{\alpha}_{0n} + N} - \frac{\alpha_i + y}{\alpha_0 + N} < -t\right) dt + 2\int_0^1 tP\left(\frac{\hat{\alpha}_{in} + y}{\hat{\alpha}_{0n} + N} - \frac{\alpha_i + y}{\alpha_0 + N} > t\right) dt\right] \\ &\le O(n^{-1}), \qquad \text{(by Lemma 4.3).} \end{aligned}$$

This completes the proof of the theorem. □

The problem of selecting the least probable cell and simultaneously estimating the probability associated with the selected cell can be treated similarly. Let the action $a = (s, p)$ represent the selection of sth cell and p as the estimate of the probability associated with the selected cell. As before let the loss function be of the following form,

$$L(\underset{\sim}{p}, (s, p)) = L_1(\underset{\sim}{p}, s) + L_2(p_s, p), \tag{9}$$

where $L_1(\underset{\sim}{p}, s)$ denotes the loss due to selecting the cell Π_s as the least probable cell and $L_2(p_s, p)$ denotes the loss due to estimating the cell probability p_s as p. Lemma 4.1 holds true in this case. We consider the specific loss function

$$L(\underset{\sim}{p}, (s, p)) = p_s - p_{[1]} + c_0(p_s - p)^2.$$

Let b'_i be the posterior expected loss if the ith population is selected and if $\hat{p}_i$ is used to estimate the probability associated with that cell. Hence

$$b'_i = c_0 \frac{(\alpha_0 + N - \alpha_i - x_i)(\alpha_i + x_i)}{(\alpha_0 + N)^2(\alpha_0 + N + 1)} + \frac{\alpha_i + x_i}{\alpha_0 + N} - C',$$

where $C' = E\{\min_{1 \le j \le k} P_j | \underset{\sim}{X} = \underset{\sim}{x}\}$. Let $b_i = b'_i + C'$. The following theorem is analogous to Theorem 4.1.

Theorem 4.3 *:*
The Bayes decision rule selects the cell π_i for which $b_i = \min_j b_j$. The Bayes estimate of the probability of the ith cell is given by $\hat{p}_i(\underset{\sim}{x}) = \frac{\alpha_i + x_i}{\alpha_0 + N}$.

In the situation when we do not know the parameters of the prior but we may have the past experience, to derive the sequence of empirical Bayes rules, we need to get the estimates of the parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$. We define the estimates as in Section 3. Let $\{\delta_n^{se_2}\}_1^\infty$; $\delta_n^{se_2}(\underset{\sim}{x}) = (s_n^{**}(\underset{\sim}{x}), \hat{p}_{s_n^{**}(\underset{\sim}{x})})$, be the sequence of empirical Bayes rules. For the case when α_0 is known, let $s_n^{**}(\underset{\sim}{x}) \in \{1, 2, \ldots, k\}$ which minimizes

$$\hat{b}_{in} = c_0 \frac{(\alpha_0 + N - \hat{\alpha}_{in} - x_i)(\hat{\alpha}_{in} + x_i)}{(\alpha_0 + N)^2(\alpha_0 + N + 1)} + \frac{\hat{\alpha}_{in} + x_i}{\alpha_0 + N}$$

and $\hat{p}_{in} = (\hat{\alpha}_{in} + x_i)/(\alpha_0 + N)$. For the case when α_0 is not known, let $s_n^{**}(\underset{\sim}{x}) \in \{1, 2, \ldots, k\}$ which minimizes

$$\hat{b}_{in} = c_0 \frac{(\hat{\alpha}'_{0n} + N - \hat{\alpha}'_{in} - x_i)(\hat{\alpha}'_{in} + x_i)}{(\hat{\alpha}'_{0n} + N)^2(\hat{\alpha}'_{0n} + N + 1)} + \frac{\hat{\alpha}'_{in} + x_i}{\hat{\alpha}'_{0n} + N}$$

and $\hat{p}_{in} = (\hat{\alpha}'_{in} + x_i)/(\hat{\alpha}'_{0n} + N)$.

Then we have the following result about the asymptotic optimality of the sequence of the empirical Bayes rules.

Theorem 4.4 *:*
The sequence of empirical Bayes rules $\{\delta_n^{se_2}\}$ defined above is asymptotically optimal of order $O(n^{-1})$.

Acknowledgment:

The authors would like to thank Professor TaChen Liang for his valuable suggestions which have been very helpful in the preparation of this paper.

References

[1] Alam, K. (1971) *"On selecting the most probable category"* Technometrics, Vol. 13, 843-50.

[2] Alam, K., K. Seo and J. R. Thompson (1971) *"A Sequential Sampling rule for the most probable multinomial event"* Ann. Inst. Stat. Math., Vol. 23, 365-74.

[3] Bechhofer, R. E., S. A. Elmaghraby and N. Morse (1959) *"A Single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability"* Ann. Math. Statist., Vol. 30, 102-19.

[4] Bechhofer, R. E. and R. V. Kulkarni (1984) *"Closed sequential procedure for selecting the multinomial events which have the largest probabilities"* Comm. Stat. -Theor. Meth., Vol. 13 (24), 2997-31.

[5] Cacoullos, T. and M. Sobel (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, 423-55.

[6] Cohen, A. and H. B. Sackrowitz (1988) *"A decision theory formulation for population selection followed by estimating the mean of the selected population"* Statistical Decision Theory and Related Topics - IV (Eds. S. S. Gupta and J. O. Berger), Vol. 2, Springer Verlag, New York, 33-36.

[7] 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.

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

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

[10] Gupta, S. S. and L. Y. Leu (1983) *"On Bayes and empirical Bayes rule for selecting good population"* Technical Report # 83-37, Department of Statistics, Purdue University.

[11] Gupta, S. S. and T. C. Liang (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, 110-28.

[12] Gupta, S. S. and T. C. Liang (1988) *"Empirical Bayes rules for selecting the best binomial population"* Statistical Decision Theory and Related Topics (Eds. S.S. Gupta, J. O. Berger), Springer-Verlag, Vol. 1, 213-24.

[13] Gupta, S. S. and T. C. Liang (1989) *"Parametric empirical Bayes rules for selecting the most probable multinomial event"* Contributions to Probability and Statistics, Essays in Honor of Ingram Olkin (Ed. Gleser et. al.), Springer-Verlag, 318-28.

[14] Gupta, S. S. and K. J. Miescke (1990a) *"On finding the largest normal mean and estimating the selected mean"* Sankhya B, Vol. 52, 144-57.

[15] Gupta, S. S. and K. J. Miescke (1990b) *"On combining selection and estimation in the search for the largest binomial parameter"* Technical Report # 90-33, Department of Statistics, Purdue University. To appear in the Proceedings of the International Workshop on Linear Models, held in Tampere, Finland, Aug 6-8, 1990.

[16] Gupta, S. S. and K. Nagel (1967) *"On selection and ranking procedures and order statistics from the multinomial distribution"* Sankhya B, Vol. 29, 1-34.

[17] Hoeffding, W. (1963) *"Probability inequalities for sums of bounded random variables"* J. Amer. Statist. Assoc., Vol. 58, 13-30.

[18] 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, 275-98.

[19] Ramey, J. and K. Alam (1979) *"A sequential procedure for selecting the most probable multinomial event"* Biometrika, Vol. 67, 171-73.

[20] Ramey, J. and K. Alam (1980) *"A Bayes sequential procedure for selecting the most probable multinomial event"* Comm. Stat. -Theor. Meth. A, Vol. 9(3), 265-76.

[21] Robbins, H. (1956) *"An empirical Bayes approach to statistical decision problems"* Proc. Third Berkeley Symp. Math. Stat. Probab., Vol. 1, University of California Press, 157-63.

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