Selecting the t-best Cells in a Multinomial Distribution

by

Subir Kumar Bhandari and Arup Bose
Indian Statistical Institute and Purdue University

Technical Report #87-19

Department of Statistics
Purdue University

May 1987
Revised January 1989

*

SELECTING THE t-BEST CELLS IN A MULTINOMIAL DISTRIBUTION

by

Subir Kumar Bhandari
Stat.-Math. Division
Indian Statistical Institute
203 B. T. Road, Calcutta 700035
India

and

Arup Bose
Department of Statistics
Purdue University
West Lafayette, IN 47907
USA

ABSTRACT

The problem of selecting the t-best cells in a multinomial distribution with t+k cells, $k > 1, 2 \leq$ t is considered under the fixed sample-size indifference zone approach. The least favourable configuration is derived for the usual procedure of selection, for large values of N (the sample size). The result settles Conjecture I (for large N) and Conjecture IV of Chen and Hwang (Commun. Statist. - Theory Meth. 13 (10), 1289–1298, 1984) in the affirmative.

Key Words and Phrases: Least favourable configuration, slippage configuration.

AMS 1985 Subject Classification: Primary 62H30 Secondary 62C99

1. INTRODUCTION

Consider a multinomial distribution with k_0 cells. Let $p_1 \geq p_2 \geq \ldots \geq p_t \geq p_{t+1} \geq \ldots \geq p_{t+k}$ with $t + k = k_0$ be the cell probabilities arranged in a descending order (the order among the cells being unknown). Based on a sample of size N, we wish to choose the t cells with the largest probabilities. The procedure is to select those t cells which have the highest count in the sample, with ties broken by randomization. The preference zone is $D(t, k_0, b) = \{(p_1, \ldots, \ldots, p_{t+k}) : p_t \geq p_{t+1} + b\}$ where b is a known constant in $(0, 1/t)$. For a given sample size N, any probability vector p in $D(t, k_0, b)$ which minimizes the probability of correct selection is called a least favourable configuration (LFC) over the preference zone $D(t, k_0, b)$. The derivation of LFC is important to find the minimum value of the sample size N required to achieve a given level of probability of correct selection.

A direct application of Theorem 3.1 of Chen (1986) shows that the LFC is of the form $p = (\alpha_1, \ldots, \alpha_t, \delta_1, \ldots, \delta_q, \delta_{q+1}, 0, \ldots, 0)$ where

$$\alpha_1 = \alpha_2 = \ldots = \alpha_t = \alpha, \delta_1 = \ldots = \delta_q, \alpha - b = \delta_1,$$
$$\delta_q \geq \delta_{q+1}, q + 1 \leq k \quad \ldots \quad \ldots \quad \ldots \quad \ldots \qquad (1.1)$$

This theorem narrows down the search for the LFC but does not identify it completely.

For the case $t = 1$, examples of Chen and Hwang (1984) show that the slippage configuration is sometimes the LFC and in some other cases it is not. Bhandari and Bose (1987) proved that the LFC for large values of N is given by $(\alpha, \delta, 0, \ldots, 0)$ and it is not the slippage configuration.

The case $t = k_0 - 1$ can be viewed as the problem of choosing the cell with the lowest probability. Alam and Thompson (1972) showed that in this case the slippage configuration is always the LFC, whatever be the value of N. Bhandari and Bose (1985) have dealt with a more general class of indifference zones and derived the LFC for large N. They have also provided a simple proof of the result of Alam and Thompson.

Our aim in this paper is to derive the LFC for large values of N when $1 < t < k_0 - 1$. (It should be remarked here that there does not seem to exist a general solution which works for every N. This belief stems from the fact that for $t = 1$, the slippage configuration is sometimes the LFC and sometimes not). It turns out that for $1 < t < k_0 - 1$, the LFC, for large N, is the configuration (2.1). This straightway settles Conjecture IV of Chen and Hwang (1984) and settles their Conjecture I in the affirmative for large values of N.

2. NOTATIONS AND PRELIMINARIES

Suppose we have N observations from a multinomial population with probability vector p.

$$\text{Let} X_i = \text{ Number of observations in the ith cell, } 1 \leq i \leq t.$$

$$Y_j = \text{Number of observations in the } (t+j) \text{ th cell, } 1 \leq j \leq k.$$

$$X = (X_1, \ldots, X_t),\ Y = (Y_1, \ldots, Y_k).$$

Let $L = (L_1, \ldots, L_t)$, $M = (M_1, \ldots, M_k)$ where L_i

$1 \leq i \leq t$ and M_j, $1 \leq j \leq k$ are non-negative integers

$$\text{Let } L_0 = \min_{1 \leq i \leq t} L_i, \qquad M_0 = \max_{1 \leq i \leq R} M_i.$$

$$\text{Let } u = \text{ Number of indices } \ell \text{ such that } L_\ell = L_0$$

$$v = \text{ Number of indices } \ell \text{ such that } M_\ell = M_0.$$

$$\begin{aligned} \text{Let } a(L,M) &= 1 \quad \text{if } L_0 > M_0 \\ &= \binom{u+v}{u}^{-1} \quad \text{if } L_0 = M_0 \\ &= 0 \text{ otherwise} \end{aligned}$$

For any probability vector p, let $\phi(p)$ denote the probability of correct selection (PCS) under p. By definition of our selection procedure, given in section 1,

$$\phi(p) = \Sigma\ a\ (L,M)\ P_p\ (X = L,\ Y = M)$$

where the sum is over all L, M such that $\Sigma L_i + \Sigma M_j = N$.

The least favourable configuration is that configuration p which minimizes $\phi(p)$ over $p \epsilon D(t, k_0, p)$. By Theorem 3.1 of Chen (1986), the search for LFC can be confined to vectors p of the form (1.1). We will denote such vectors by $p(\alpha, q)$.

$$\text{Let } S(L,M) = a\ (L,M)\ t/u \quad \text{if } L_t = L_0$$
$$= 0 \text{ otherwise.}$$

It is easy to see that

$$\phi(p(\alpha, q)) = \sum_{\Sigma L_i + \Sigma M_j = N} S(L,M) P_{p(\alpha,q)}(X = L,\ Y = M)$$

For the individual probability terms which are to appear in various summations, define $F(L,M) = \frac{N!}{\pi L_i! \pi M_j!}\ \pi\ \alpha_i^{L_i} \pi\ \delta_j^{M_j}$.

A natural conjecture would be that the slippage configuration is the LFC for all values of N. But as we have mentioned earlier, there does not exist a general form of the LFC which holds for all values of N. This poses serious problems for minimum sample size determination.

We will show later that the LFC for sufficiently large values of N is given by:

$$p = (\alpha, \ldots, \alpha, \delta, 0, 0, \ldots, 0) \text{ with } \delta = (1 - bt)/(t+1)$$
$$\alpha = (1 + b)/(t+1) \tag{2.1}$$

3. THE MAIN RESULT

In this section, we will prove that for sufficiently large N, the *LFC* is the configuration (2.1). Then we will discuss our results vis-a-vis some of Chen and Hwang's conjectures. The basic idea of the proof runs as follows:

To prove that the configuration (2.1) is the *LFC*, it is enough to show that the integral $w.r.t.\alpha$, of the directional derivative of ϕ $w.r.t.$ $(\alpha_1, \alpha_2, \ldots, \alpha_t, \delta_1, \ldots, \delta_{q+1})$ along the direction $(\underbrace{-1,\ldots,-1}_{t},\underbrace{-1,\ldots,-1}_{q},t+q,\ 0,\ldots,0)$ from α_0 to $\frac{1+qb}{t+q}$ is positive for suitable values of α_0. i.e. $\int_{\alpha_0}^{(1+qb)/(t+q)} D(p(\alpha,q))d\alpha > 0$ where

$$D = -\left(\sum_{i=1}^{k}\frac{\partial\phi}{\partial\alpha_i}+\sum_{i=1}^{q}\frac{\partial\phi}{\partial\alpha_i}\right)+(t+q)\frac{\partial\phi}{\partial\delta_{q+1}}.$$

Let $B_1(N) = \Sigma\ \{F(L,M) : L_t = M_{i_1}$ or $L_t = M_{i_1} \pm 1$ for some i_1,

$$L_i \geq L_t \geq M_j \forall i,j,i \neq i_1\}$$

$$B_2(N) = \Sigma\ \{F(L,M) : L_t = L_1 - 1, M_j + 1 < L_t < L_{i_1} - 1\ \forall i,j,i \neq 1\}$$

We will show that one of the terms in the integral of D is integral of $(t-1)\ B_2\ (N)/2$. The other terms are dominated by integral of $c\ B_1(N)$ for some constant c independent of N. Using Stirling's approximation for factorials, the above sums can be written as approximate Riemann sums of the N^{th} power of two functions f_1 and f_2 respectively. The partition lengths of these sums are small and hence they are approximate integrals. Thus the N^{th} root of these sums converge respectively to max f_1 and max f_2 by the well known result from analysis that the L_p norms converge to L_∞ norm as $p \to \infty$. By comparing these maximums, we will show that $\int B_1(N)/\int B_2(N) \to 0$ as $N \to \infty$. This will prove our result.

Let $f_1(\alpha,r) = r^{-2r}(1-2r)^{-(1-2r)}\alpha^r(1-2\alpha+a)^{1-2r}(\alpha-a)^r$

where $0 \leq r \leq (t+1)^{-1}$,

$\dfrac{1+(q+1)b}{t+q+1} \leq \alpha \leq \dfrac{1+qb}{t+q}$ and $a \geq b > 0$

$f_2\ [\alpha,\ r] = r^{-2r}(1-2r)^{-(1-2r)}\alpha^{2r}(1-2\alpha)^{1-2r}$

where $0 \leq r \leq t^{-1}$,

$\dfrac{1+(q+1)b}{t+q+1} \leq \alpha_0 \leq \alpha \leq \dfrac{1+qb}{t+q}$, $1/t > b > 0$ and α_0 is fixed.

Lemma 3.1:

(i) max $f_1 < 1$

(ii) f_2 attains maximum at $\alpha = r$ and max $f_2 = 1$

Proof: (i) $f_1(\alpha, r) = (\alpha/r)^r[(\alpha - a)/r]^r[(1 - 2\alpha + a)/(1 - 2r)]^{1-2r}$

By the inequality $(1 + x)^n \leq 1 + nx$ which holds for $0 \leq n \leq 1$, we have

$$f_1(\alpha, r) \leq [1 + (\alpha - r)]\,[1 + (\alpha - r) - a]\,[1 - 2(\alpha - r) + a]\,.$$

Putting $\lambda = \alpha - r$, the above bound reduces to

$$I(\lambda) = 1 - [2\lambda^2(1 + \lambda) + \lambda^2 + 3a(-\lambda)(1 + \lambda) + a^2(1 + \lambda)]$$

Note that $-1 < \lambda < 1$.

If $-1 < \lambda < 0, I(\lambda)$ is clearly less than 1.

If $0 \leq \lambda \leq 1$, then

$$I(\lambda) = 1 - \left[(3\lambda/2 - a)^2 + \lambda(3\lambda/2 - a)^2 + (3\lambda^2/4 - \lambda^3/4)\right] < 1.$$

Hence (i) is proved.

(ii) $\frac{\partial \log\ f_2}{\partial \alpha} = 2r/\alpha - 2(1 - 2r)/(1 - 2\alpha)$ which vanishes iff $\alpha = r$.

$$\frac{\partial^2 \log\ f_2}{\partial \alpha^2} = -2r/\alpha^2 + (-4)(1 - 2r)/(1 - 2\alpha)^2 < 0.$$

Thus for fixed r, f_2 attains its maximum at $\alpha = r$ (which is a possible value since $b < 1/t$) and obviously at this point, $f_2 = 1$. This proves the lemma.

We now state and prove our main theorem.

Theorem 3.2: For all large N, the LFC is the configuration given by (2.1).

Proof: As we have already remarked, by Theorem 3.1 of Chen (1986), it is enough to find the point of minimum of ϕ on the set

$$\Omega = \{(\underbrace{\alpha, \ldots, \alpha}_{t}, \underbrace{\delta, \ldots, \delta}_{q}, \delta_{q+1}, 0, \ldots, 0) : \alpha - b = \delta, \delta \geq \delta_{q+1}, q + 1 \leq k\}.$$

The directional derivative ϕ along $(\underbrace{-1,\ldots,-1}_{t}\underbrace{-1,\ldots,-1}_{q},t+q,0,\ldots,0)$ is given by

$$D = -\sum_{i=1}^{t}\frac{\partial\phi}{\partial\alpha_i} - \sum_{j=1}^{q}\frac{\partial\phi}{\partial\delta_j} + (t+q)\frac{\partial\phi}{\partial\delta_{q+1}}$$

$$= \sum_{i=1}^{t}\left(V - \frac{\partial\phi}{\partial\alpha_i}\right) + \sum_{j=1}^{q}\left(V - \frac{\partial\phi}{\partial\delta_i}\right) - (t+q)\left(V - \frac{\partial\phi}{\partial\delta_{q+1}}\right)$$

where $V = \sum\limits_{\sum_{i=1}^{t} L_i + \sum_{j=1}^{q+1} M_j = N-1} a(L,M)F(L,M)$ and it is to

be remembered that all the expressions occurring are evaluated at $p(\alpha,q)$. Henceforth, we will write S for $S(L,M)$ and F for $F(L,M)$ and unless mentioned otherwise, $\sum_{i=1}^{t} L_i + \sum_{j=1}^{q+1} M_j = N-1$.

V can be split up into

$$V = \sum_{i_1=1}^{t-1}\frac{1}{2}\{F : M_j + 1 < L_t < L_i \ \forall i \neq i_1, \forall_j \text{ and } L_t = L_{i_1}\}$$

$$+ \sum \{F : M_j + 1 < L_t < L_i \ \forall_{i,j}\} + K_1 \text{ (say)}$$
$$= V_1 + V_2 + K_1 \text{ (say)}$$

Split up $\frac{\partial\phi}{\partial\alpha_p}, 1 \leq p \leq t-1$ as

$$\frac{\partial\phi}{\partial\alpha_p} = \sum \{F : M_j + 1 < L_t < L_i\}$$
$$+ \sum \{F : M_j + 1 < L_t < L_i \ \forall i \neq p,\ L_t = L_p\}$$
$$+ \sum_{i_1 \neq p}\sum \frac{1}{2}\{F : L_t = L_{i_1}, M_j + 1 < L_t < L_i \ \forall i \neq i_1\}$$
$$+ K_{p_2} \text{ (say)}$$
$$= J_{p1} + J_{p2} + J_{p3} + K_{p_2} \text{ (say)}.$$

Clearly, $V_2 = J_{p_1}$,

$$V_1 - J_{p_3} = \frac{1}{2} \sum \{F : L_t = L_p, M_j + 1 < L_t < L_i \; \forall i \neq p\}$$
$$= \frac{1}{2} J_{p2}$$
$$\text{Thus } V - \frac{\partial \phi}{\partial \alpha_p} = -\frac{1}{2} J_{p2} + K_1 - K_{p_2} \ldots \quad \ldots \quad \ldots \tag{3.1}$$

Split up $\frac{\partial \phi}{\partial \alpha_t}$ as

$$\frac{\partial \phi}{\partial \alpha_t} = \sum \{F : M_j + 1 < L_t < L_i - 1 \; \forall i, j\}$$
$$+ \sum_{i_1=1}^{t-1} \sum \frac{1}{2} \{F : M_j + 1 < L_t = L_{i_1} - 1, L_t < L_i - 1 \; \forall i \neq i_1\}$$
$$+ K_{t_2}$$
$$= J_{t1} + J_{t2} + K_{t2} \text{ (say)}.$$

Note that $V_2 \geq J_{t1} + 2 \; J_{t2}$.

$$\text{Hence } V - \frac{\partial \phi}{\partial \alpha_t} \geq V_1 + J_{t2} + K_1 - K_{t2} \quad \ldots \quad \ldots \quad \ldots \tag{3.2}$$

Further, since the derivatives etc. are finally evaluated at points of the form $p(\alpha, q)$, so that $\alpha_1 = \alpha_2 = \ldots = \alpha_t = \alpha$, it is clear that

$$V_1 = \frac{1}{2} \sum_{p=1}^{t-1} J_{p2} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \tag{3.3}$$

Thus, adding (3.1) and (3.2) and using (3.3),

$$D \geq J_{t2} + tK_1 - \sum_{j=1}^{t} K_{j2} + \sum_{j=1}^{q} \left(V - \frac{\partial \phi}{\partial \delta_i} \right) - (t+q) \left(V - \frac{\partial \phi}{\partial \delta_{q+1}} \right)$$
$$= J_{t2} + E \text{ (say)}.$$

$$\text{Note that } J_{t2} = \frac{(t-1)}{2} \sum \{F : M_j + 1 < L_t < L_i - 1 \; \forall i, j, \; i \neq 1, L_t = L_1 - 1\}.$$

We will now prove that $\int_{\alpha_0}^{\alpha^*} E d\alpha$ is negligible compared to $\int_{\alpha_0}^{\alpha^*} J_{t2} \, d\alpha$ where $\alpha^* = (1+qb)/(t+q)$. This will prove the theorem.

We have to take care only of the terms in E which are negative. We first examine the term $\sum_{j=1}^{t} K_{j2}$. An examination of the expression for $\frac{\partial \phi}{\partial \alpha_p}, 1 \leq p \leq t-1$ shows that

$$\begin{aligned} K_{p_2} &\leq c \sum \{SF : M_j + 1 = L_t \text{ or } M_j = L_t \text{ for some } j\} \\ &+ c \sum \{SF : M_j + 1 < L_t = L_{i_1} = L_{i_2} \text{ for some } i_1, i_2 \text{ and for all } j\} \\ &= c(B_1 + B_2) \text{ (say).} \end{aligned}$$

[Here c is a constant independent of α, δ and N].

Similarly,

$$\begin{aligned} K_{t_2} &\leq c \sum \{SF : M_j \pm 1 = L_t \text{ or } M_j = L_t\} \\ &+ c \sum \{SF : M_j + 1 < L_t = L_{i_1} - 1 = L_{i_2} - 1 \text{ for some } i_1 \neq i_2 \neq t\} \\ &= c\ (B_3 + B_4) \text{ (say).} \end{aligned}$$

For $1 \leq p \leq q+1$,

$$\begin{aligned} \frac{\partial \phi}{\partial \delta_p} &= V_1 + V_2 + \sum \{SF : M_j + 1 < L_t = L_{i_1} = L_{i_2} \text{ for some } i_1, i_2 \text{ and for all } j\} \\ &+ \sum \{SF : M_p + 1 < L_t, M_j + 1 = L_t \text{ or } M_j = L_t \text{ for some } j\} \\ &+ \sum \{S'F : M_p + 1 = L_t\} \text{ (for some } S', 0 \leq S' \leq S) \\ &= V_1 + V_2 + B_5 + B_6 + B_7 \text{ (say).} \end{aligned}$$

Hence $V - \frac{\partial \phi}{\partial \delta_p} = K_1 - (B_5 + B_6 + B_7)$.

Note that in the final expression for D, all K_1 terms cancel out.

Thus it remains to show that $\sum_{i=1}^{7} \int_{\alpha_0}^{\alpha^*} B_i d\alpha$ is negligible compared to $\int_{\alpha_0}^{\alpha^*} J_{t2}\, d\alpha$. Clearly,

$$\begin{aligned}
J_{t2} \geq\ & \frac{1}{2} \sum \{F : M_j + 1 < L_t < L_i - 1\ \forall i, j, i \neq 1 \text{ and } L_t = L_1 - 1\} \\
&= \Sigma_1\ P(X_1 = L_1, X_t = L_1 - 1 | X_1 + X_t = 2L_1 - 1) \\
&\times\ P\ (\sum_{i=2}^{t-1} X_i + \sum_{j=1}^{k} Y_j = N - (2L_1 - 1)) \\
&\times\ \Sigma_2\ P(X_i = L_i, i = 2, \ldots, t-1, Y_j = M_j, j = 1, \ldots k) \\
&\times\ \eta\ (L_1, \ldots, L_{t-1}, M_1, \ldots, M_k) \\
&\times \left[P \left(\sum_{i=2}^{t-1} X_i + \sum_{j=1}^{k} Y_j = N - (2L_1 - 1) \right) \right]^{-1}
\end{aligned}$$

where $\eta = 1$ if $M_j + 1 < L_1 - 1 < L_i - 1\ \forall i \neq 1,\ t\ \forall j$ and $L_t = L_1 - 1$

$= 0$ otherwise.

and Σ_1 denotes the summation over L_1 such that $3 \leq L_1 \leq Nt^{-1}$ and, for fixed L_1, Σ_2 denotes summation over all L_i's and M_j's such that $\sum_{i=2}^{t-1} L_i + \sum_{j=1}^{k} M_j = N - (2L_1 - 1)$.

Thus $J_{t2} \geq \frac{1}{2}\ \Sigma^* P(X_1 = Nr, X_t = Nr - 1 | X_1 + X_t = 2Nr - 1) \times f(r, \delta_{q+1}, N)$

$$\times\ P \left(\sum_{i=2}^{t-1} X_i + \sum_{j=1}^{k} Y_j = N - (2Nr - 1) \right)$$

where $*$ denotes summation over all $r, 0 < r \leq t^{-1}$ and Nr is an integer.

From the relations $\alpha - b = \delta$, and $t\alpha + q\delta + \delta_{q+1} = 1$, it follows that

$$[1 + (q+1)b](t + q + 1)^{-1} \leq \alpha \leq (1 + qb)(t + q)^{-1}.$$

$$P(X_1 = Nr, X_t = Nr - 1 | X_1 + X_t = Nr - 1) \cdot P \left(\sum_{i=2}^{t-1} X_i + \sum_{j=1}^{k} Y_j = N - (2Nr - 1) \right)$$

$= \dfrac{N!\ \alpha^{2Nr-1}(1 - 2\alpha)^{N-(2Nr-1)}}{(Nr)!(Nr - 1)!(N - (2Nr - 1))!}$ and using Stirling's approximation

$\geq c\ N^{-1} \left[r^{-2r}(1 - 2r)^{-(1-2r)} \alpha^{2r} (1 - 2\alpha)^{1-2r} \right]^N$ where c is a constant independent of α, δ_{q+1} and N.

Recalling the definition of f_2,

$$\int_{\alpha_0}^{\alpha^*} J_{t2} \geq c \int_{\alpha_0}^{\alpha^*} \left[\Sigma^* \frac{f_2^N(\alpha, r) f(r, \delta_{q+1}, N)}{N} \right].$$

We shall show later than in open neighbourhoods arbitrary close to the point, where f_2 attains its maximum, $f(r, \delta_{q+1}, N)$ remains bounded away from zero.

It is not difficult to show that if $f(y, x)$ is a continuous positive function bounded by 1 on $D = [a_1, b_1] \times [a_2, b_2]$ then for a sequence of partitions $\{\pi_n\}$ of $[a_2, b_2]$ such that $||\pi_n|| \to 0$,

$$[\int_{a_1}^{b_1} \sum_{x_i, n \in \pi_n} f^n(y, x_i, n)(x_{i+1,n} - x_i, n)]^{1/n} \longrightarrow \sup_{(y,x) \in D} |f(y, x)|$$

Let D be a neighbourhood of the above form, of a point where f_2 attains its maximum. Then

$$[\int_{\alpha_0}^{\alpha^*} (\frac{(\sum^* f_2^N(\alpha, r) f(r, \delta_{q+1}, N))}{N})]^{1/N}$$

$$\geq c^{1/N} [\int_{a_1}^{b_1} \sum_{x_{in}} f_2^N(\alpha, x_i, n) f(x_{i+1}, n - x_{in})]^{1/N}$$

where $x_{i,N} = i/N$ such that $Ni \leq Nt^{-1}$ and by the fact mentioned, the right side of the above inequality converges to max f_2. This shows that,

$$\frac{\lim}{N \to \infty} \left(\int_{\alpha_0}^{\alpha^*} J_{t2} \right)^{1/N} \geq max\ f_2 = 1 \text{ (and in fact uniformly over } \alpha_0).$$

Proceeding in a similar way

$$\overline{\lim_{N \to \infty}} \left(\int_{\alpha_0}^{\alpha^*} B_i \right)^{\frac{1}{N}} \leq max\ f_{1i} < 1$$

$\forall\ i = 1, 3, 6, 7$ and f_{1i} is f_1 with different a for different i.

Hence $\int B_1, \int B_3, \int B_6, \int B_7$ are negligible with respect to $\int J_{t2}$.

$$F(L_1, \ldots, L_t, M_1, \ldots, M_k) = \frac{N!}{\pi L_i! \pi M_j!}^{\alpha^{\Sigma L_i}} (\alpha - b)^{\Sigma M_j}\ \delta_{q+1}^{M_{q+1}}$$

Let $B_8 = \Sigma\ \{SF : M_{j+1} < L_t = L_1 - 1 = L_2 - 2 < L_i - 1\ \forall j \text{ and } i \neq 1\}$

$$\text{Now } \frac{F(L_t, L_t, L_3, \ldots L_t, M_1, \ldots, M_k)}{F(L_t, L_t + 1, L_3, \ldots, L_{t-1}, L_t - 1, M_1, \ldots, M_k)} = \frac{L_t + 1}{L_t} \leq 2$$

This shows that $B_2 = B_5 \leq B_1 + c\ B_8$.

So we need to show that $\int_{\alpha_0}^{\alpha^*} B_4$ and $\int_{\alpha_0}^{\alpha^*} B_8$ are negligible with respect to $\int_{\alpha_0}^{\alpha^*} J_{t2}$.

$$\{F : F \text{ is an element of } B_4 \text{ or } B_8\} \subseteq \{F : F \text{ is in } J_{t2}\}.$$

Again as $\max_{\substack{0<r\leq t-1 \\ 0<\alpha_0\leq\alpha\leq(1+qb)/(t+q)<1/t}} f_2(\alpha, r) = 1$ is attained at $r = \alpha \geq \alpha_0 \geq [1 + (q + 1)b]/(t + q + 1)$, we can take those L_t for which $L_t/N \geq \varepsilon$, for some fixed preassigned small $\varepsilon > 0$.

Let $B_m^* = \sum \{SF : M_j + 1 < L_t = L_1 - 1 = L_2 - m < L_i - 1 \text{ for all } j \text{ and } i \neq 1, 2\}$

Now for large N, with $L_t \geq \varepsilon N$,

$F(L_t + 1, L_t + 2, L_3, \ldots, L_t, M_1, \ldots, M_k)/F(L_t, L_t + 4, L_3, \ldots, L_{t-1}, L_t - 1, M_1, \ldots, M_k)$

$= (L_t + 3)(L_t + 4)/(L_t + 1)\ L_t \leq (1 + \varepsilon_1)$ for some pre-assigned small $\varepsilon_1 > 0$.

Hence $B_8 = B_2^* \leq B_1 + (1 + \varepsilon_1)\ B_5^* + \varepsilon_2\ J_{t2}$.

In a similar way we can prove

$$B_{3m-1}^* \leq B_1 + (1 + \varepsilon_1)\ B_{3m+2}^* + \varepsilon_2\ J_{t2} \text{ for all } 1 \leq m \leq \ell$$

and small $\varepsilon_1, \varepsilon_2 > 0$, with ℓ large (but fixed) and for sufficiently large N depending on ℓ and $\varepsilon_1, \varepsilon_2$.

$$\begin{aligned} \text{So } B_8 &\leq B_1 + (1 + \varepsilon_1)\ B_1 + (1 + \varepsilon_1)^2 B_1 + \ldots + (1 + \varepsilon_1)^{m-2} B_1 \\ &\quad + (1 + \varepsilon_1)^{m-1} B_{3m-1}^* + \varepsilon_2 J_{t2} \\ &\leq \varepsilon_1^{-1}(1 + \varepsilon_1)^{m-1} B_1 + (1 + \varepsilon_1)^{m-1} B_{3m-1}^* + \varepsilon_2\ J_{t2}\ \forall\ 1 \leq m \leq \ell. \end{aligned}$$

$$\begin{aligned} \text{Hence } B_8 &\leq \varepsilon_1^{-1}(1 + \varepsilon_1)^{\ell-1} B_1 + \ell^{-1}(1 + \varepsilon_1)^{\ell-1} \sum_{m=1}^{\ell} B_{3m-1}^* + \varepsilon_2 J_{t2} \\ &\leq \varepsilon_1^{-1}(1 + \varepsilon_1)^{\ell-1} B_1 + \ell^{-1}(1 + \varepsilon_1)^{\ell-1} J_{t2} + \varepsilon_2\ J_{t2} \end{aligned}$$

$$\text{Hence } \int B_8 \Big/ \int J_{t2} \le \varepsilon_1^{-1}(1+\varepsilon_1)^{\ell-1} \int B_1 \Big/ \int J_{t2} \; + \ell^{-1}(1+\varepsilon_1)^{\ell-1} + \varepsilon_2$$

Choosing ℓ such that $2\ell^{-1} \le \varepsilon$ and ε_1 such that $(1+\varepsilon_1)\ell \le 2, \varepsilon_2 \le \varepsilon$, we have, $\int B_8$ is negligible with respect to $\int J_{t2}$. Similarly $\int B_4$ is negligible with respect $\int J_{t2}$. Thus $\int B_i, i = 1, 2, \ldots, 7$ are all negligible with respect to $\int J_{t2}$ for large N.

It now remains to be shown that $f(r, \delta_{q+1}, N)$ remains bounded away from 0 in open neighbourhoods arbitrarily close to the points where f_2 attains its maximum.

By the SLLN and Lemma 3.1 (ii), it suffices to show that one pair of points with $\alpha = r$ is contained in the region given by

$$\begin{aligned}
\{(\alpha, r) : & EN^{-1}\left(Y_\ell \Big| \sum_{i=2}^{t-1} X_i + \sum_{j=1}^{k} Y_j = N - (2L_1 - 1)\right) \le r \text{ for all} \\
& 1 \le \ell \le k \text{ and } EN^{-1}\left(X_\ell \Big| \sum_{i=2}^{t-1} X_i + \sum_{j=1}^{k} Y_j = N - (2L_1 - 1)\right) \ge r \\
& \text{for all } \ell, 2 \le \ell \le t-1\} \\
& = \{(\alpha, r) = \frac{\alpha - b}{1 - 2\alpha}(1 - 2r) \le r \text{ and } \frac{\alpha}{1 - 2\alpha}(1 - 2r) \ge r\}.
\end{aligned}$$

It is obvious that $\alpha = r$ belongs to this set. Hence the proof of Theorem 3.2 is complete.

Remark 3.3: Chen and Hwang (1984) set down some conjectures regarding the LFC for choosing the t-best cells in the same framework as ours. We quote their conjectures here. We need the following definitions.

$$p(\ell) = \left(\underbrace{\alpha, \ldots \alpha}_{t}, \underbrace{\delta, \ldots \delta}_{\ell}, 0, 0, \ldots 0\right) \text{ where } \delta = \alpha - b.$$

$$p(t, k_0, b) = \left(\underbrace{\alpha, \ldots, \alpha}_{t}, \underbrace{\delta, \ldots, \delta}_{k}\right) \text{ where } \delta = \alpha - b.$$

Conjecture I: For any sample size N and any preference zone $D(t, k_0, b)$ one of the $p(\ell)'s, \ell = 1, 2, \ldots, k$ is a LFC over $D(t, k_0, b)$.

Conjecture IV: For each $t = 1, 2, \ldots \forall k_0 \geq t + 2$ and any $b \epsilon (0, 1/t)$, there exists N_o such that $p(t, k_0, b)$ is not LFC over $D(t, k_0, b)$.

Theorem 3.2 shows that Conjecture IV is true. In fact Theorem 3.2 shows that for all sufficiently large N, the LFC is of the form $(\alpha, \ldots, \alpha, \delta, 0, \ldots, 0)$ with $\delta = \alpha - b$. It also shows that Conjecture I is true for large values of N. In fact $p(1)$ is the LFC for all large values of N.

Remark 3.4: The problem of determining the minimum sample size N is not taken care of by Theorem 3.2. For this we need an approximate value of $N = N_0$ so that the LFC is (2.1) for all $N \geq N_0$. But getting hold of N_0 seems to be a difficult problem.

Acknowledgement

The authors are grateful to an anonymous referee who read an earlier version of the paper and whose suggestions have improved the presentation of our results.

References

Alam, K. and Thompson, J. R. (1972): On selecting the least probable multinomial event. *Ann. Math. Statist.* **43**, 1981–1990.

Bhandari, S. K. and Bose, A. (1986): On selecting the least likely event. **Tech. Rep. No. 34/85**, Stat.-Math. Division, Indian Statistical Institute, Calcutta.

Bhandari, S. K. and Bose, A. (1987): On selecting the most likely event. *Journal of Statistical Planning and Inference* (to appear).

Chen, P. (1986): On the least favourable configuration in multinomial selection problems. *Commun. Statist. Theor. Meth.* **15(2)**, 367–385.

Chen, R. W. and Hwang, F. K. (1984): Some theorems, counter examples, and conjectures in multinomial selection theory. *Commun. Statist. Theor. Meth.* **13(10)**, 1289–1298.