Journal of Statistical Planning and Inference 4 (1980) 391–402.

MINIMAX SUBSET SELECTION FOR THE MULTINOMIAL DISTRIBUTION*

Roger L. BERGER
Florida State University, Department of Statistics, Tallahassee, FL 32306, USA

Received 2 April 1979; revised manuscript received 28 August 1980
Recommended by S.S. Gupta

Abstract: Let $(X_1, \ldots, X_k)$ be a multinomial vector with unknown cell probabilities $(p_1, \ldots, p_k)$. A subset of the cells is to be selected in a way so that the cell associated with the smallest cell probability is included in the selected subset with a preassigned probability, P^*. Suppose the loss is measured by the size of the selected subset, S. Using linear programming techniques, selection rules can be constructed which are minimax with respect to S in the class of rules which satisfy the P^*-condition. In some situations, the rule constructed by this method is the rule proposed by Nagel (1970). Similar techniques also work for selection in terms of the largest cell probability.

AMS 1970 Subject Classifications: Primary 62F07; Secondary 62C05.

Key words: Linear Programming; Minimax Subset Selection; Expected Subset Size.

1. Introduction

In this paper, subset selection problems for the multinomial distribution are considered. In these problems the aim is to select a non-empty subset of the cells which contains the cell with the highest or lowest cell probability. The goal is to find a selection rule which includes the highest or lowest cell probability with probability at least equal to a preassigned number, P^*. Having satisfied this minimum requirement, the goal is to find a rule which effectively excludes non-best cells. This leads to the use of the number of cells selected or the number of non-best cells selected as a measure of the loss to the experimenter. Minimax rules for these losses are considered and the main result of this paper is that minimax rules can be constructed by solving the appropriate linear programming problem. The rule obtained in this way in some situations corresponds to a particularly simple and easy to implement rule proposed and studied by Nagel (1970). This rule can also be shown to have another optimal property if the cell probabilities are in a slippage configuration.

The subset selection problem for multinomial distributions has been previously considered by Gupta and Nagel (1967), Nagel (1970), Panchapakesan (1971) and Gupta and Huang (1975). Minimax subset selection rules have been recently

* This research was supported in part by U.S. Army Research Office Grants No. DAAG29-76-G-0238 and DAAG29-79-C-0158.

investigated in a general setting by Berger (1979) and in a location parameter setting by Bjørnstad (1978).

Section 2 contains the necessary notation for a formulation of the problem. In Section 3, the problem of choosing the smallest cell probability is considered and the fact that minimax selection rules can be constructed using linear programming methods is proven. Examples of rules constructed in this way are given. A particular rule which arises as the solution to the problem is further considered in Section 4. This rule is found to have a certain optimality property if the parameters are in a slippage configuration. In Section 5, the analogous results for the problem of selecting the largest cell probability are outlined.

2. Notation and formulation

Let $\boldsymbol{X} = (X_1, \ldots, X_k)$ be a multinomial random vector with $\sum_{i=1}^k X_i = N$. Let $\boldsymbol{p} = (p_1, \ldots, p_k)$ be the unknown cell probabilities with $\sum_{i=1}^k p_i = 1$. The ordered cell probabilities will be denoted by $p_{[1]} \leq \cdots \leq p_{[k]}$. The goal of the experimenter is to select a subset of the cells including the *best* cell, the cell associated with $p_{[1]}$. A *correct selection*, CS, is the selection of any subset which contains the best cell. If for a particular parameter value, $\boldsymbol{p}$, more than one of the cells is tied with the smallest p_i, one of these cells will be considered to be tagged and a CS occurs if the tagged cell is selected. This assumption is not of essential importance and can be dropped without effecting any of the results. But it eliminates arguments involving limits of parameter values.

A selection rule will be denoted by $\boldsymbol{\phi}(\boldsymbol{x}) = (\phi_1(\boldsymbol{x}), \ldots, \phi_k(\boldsymbol{x}))$ where $\phi_i(\boldsymbol{x})$ is the probability of including the ith cell in the selected subset having observed $\boldsymbol{X} = \boldsymbol{x}$. The $\phi_i(\boldsymbol{x})$ are called the *individual selection probabilities*. To insure that a nonempty subset is always selected, only rules which satisfy $\sum_{i=1}^k \phi_i(\boldsymbol{x}) \geq 1$ for all $\boldsymbol{x}$ will be considered. This is a necessary and sufficient that a randomization scheme exists, with these individual selection probabilities, which always selects a nonempty subset. The first requirement a selection rule must satisfy is that it has a certain minimum probability of selecting the best cell. Let P^*, $1/k < P^* < 1$, be a preassigned fixed number. Only rules which satisfy the P^*-condition, viz.,

$$\inf_{\boldsymbol{p}} P_{\boldsymbol{p}}(\mathrm{CS} \mid \boldsymbol{\phi}) \geq P^*, \tag{2.1}$$

will be considered. The set of rules which satisfy the P^*-condition will be denoted by $\mathscr{D}_{P^*}$.

Two losses which are commonly used in subset selection problems are the size of the selected subset, S, and the number of non-best cells selected, S'. The risk using a rule $\boldsymbol{\phi}$ when $\boldsymbol{p}$ is the true parameter is then the expected subset size, $E_{\boldsymbol{p}}(S \mid \boldsymbol{\phi})$, or the expected number of non-best cells selected, $E_{\boldsymbol{p}}(S' \mid \boldsymbol{\phi})$. Let

$$S(\boldsymbol{x}) = \sum_{i=1}^{k} \phi_i(\boldsymbol{x}). \tag{2.2}$$

Then $S(\boldsymbol{x})$ is the expected subset size having observed $\boldsymbol{X}=\boldsymbol{x}$ and $E_{\boldsymbol{p}}(S \mid \boldsymbol{\phi}) = E_{\boldsymbol{p}}(S(\boldsymbol{X}) \mid \boldsymbol{\phi})$.

A selection rule, $\boldsymbol{\phi}^*$, is said to be *minimax with respect to S* if $\boldsymbol{\phi}^* \in \mathscr{D}_{P^*}$ and

$$\inf_{\mathscr{D}_{P^*}} \sup_{\boldsymbol{p}} E_{\boldsymbol{p}}(S \mid \boldsymbol{\phi}) = \sup_{\boldsymbol{p}} E_{\boldsymbol{p}}(S \mid \boldsymbol{\phi}^*). \tag{2.3}$$

Replace S by S' in (2.3) to define minimaxity with respect to S'. Berger (1979) investigated minimaxity with respect to S and S' in a general setting. In Section 3, selection rules which are minimax with respect to S and S' will be constructed.

3. Construction of minimax rules

In Theorem 3.1, the following class of selection rules will be shown to be minimax with respect to S and S'. Let $\mathscr{D}^*$ be the class of selection rules which satisfy: (i) $\phi_i(\boldsymbol{x}) = \delta_i(x_i)$ for $1 \le i \le k$, i.e., the selection or rejection of the ith cell depends only on the number of observations in the ith cell; (ii) $\delta_i(m) \ge \delta_i(n)$ if $0 \le m \le n \le N$ and $1 \le i \le k$, i.e., the probability of selecting the ith cell decreases as the number of observations in the ith cell increases; (iii) $S(\boldsymbol{x}) = \sum_{i=1}^k \delta_i(x_i)$ is a Schur concave function of $\boldsymbol{x}$ and (iv) $E_{\boldsymbol{p}_0}(\delta_i(X_i)) = P^*$ for $1 \le i \le k$ where $\boldsymbol{p}_0 = (1/k, \ldots, 1/k)$.

Remark 3.1. The requirement in (i) that the decision to include or exclude the ith cell depends only on X_i may seem to be a waste of the information contained in the other X_j's. But since $\sum_{i=1}^k X_i = N$, if X_i is 'large' then the other X_j's must be small and if X_i is 'small', some of the other X_j's must be large. So information about the other X_j's is contained in X_i and the requirement is not as counterintuitive as it would be in a problem in which the X_j's are independent, e.g., the normal means problem (see Gupta (1965)). In fact, some rules of this form have been shown to be admissible by Berger (1979b). The rules considered by Nagel (1970) had property (i) as well as (ii) and (iv).

Remark 3.2. Condition (ii) seems reasonable since larger values of X_i indicate larger values of p_i.

Remark 3.3. See Proschan and Sethuraman (1977) and Nevius, Proschan and Sethuraman (1977) for a discussion of Schur functions in statistics. In a sense, $\boldsymbol{x}$ majorizes $\boldsymbol{y}$ if the coordinates of $\boldsymbol{x}$ are more spread out than the coordinates of $\boldsymbol{y}$. The more spread out the cell frequencies are, the easier it should be to decide which is the best cell and the smaller the subset that need be selected. This is an interpretation of condition (iii).

Remark 3.4. By Theorem 4.1 of Berger (1979), condition (iv) is a necessary condition if a rule is to be minimax with respect to S or S'.

Remark 3.5. Gupta and Nagel (1967) proposed and studied the following selection rule for this problem: select the ith cell if $x_i \le x_{\min} + C$ where $x_{\min} = \min(x_1, \ldots, x_k)$ and C is a non-negative integer chosen so that the P^*-condition is satisfied. In general, this rule is not in $\mathcal{D}^*$ since neither (i) nor (iii) are true. In the special case of $k=2$, however, it can be verified that the Gupta–Nagel rule is in $\mathcal{D}^*$ since $X_2 = N - X_1$.

Theorem 3.1. *If $\boldsymbol{\phi} \in \mathcal{D}^*$, then $\boldsymbol{\phi}$ is minimax with respect to S and S'.*

Proof. The first fact to be verified is that if $\boldsymbol{\phi} \in \mathcal{D}^*$, then $\boldsymbol{\phi}$ satisfies the P^*-condition. Let $\mathcal{P}_i = \{\boldsymbol{p}\colon p_j \ge p_i \ge 0 \text{ for all } j \ne i \text{ and } \sum_{j=1}^k p_j = 1\}$. Then,

$$\inf_{\boldsymbol{p}} P_{\boldsymbol{p}}(\mathrm{CS} \mid \boldsymbol{\phi}) = \inf_{1 \le i \le k} \inf_{\mathcal{P}_i} P_{\boldsymbol{p}}\,(\text{select } i\text{th cell} \mid \boldsymbol{\phi}) = \inf_{1 \le i \le k} \inf_{\mathcal{P}_i} E_{\boldsymbol{p}}(\phi_i(\boldsymbol{X})).$$

Let Y have a binomial distribution with parameters N and p_i. Then $E_{\boldsymbol{p}}(\phi_i(\boldsymbol{X})) = E_{p_i}(\delta_i(Y))$. Since the binomial distribution has monotone likelihood ratio and since δ_i is a nonincreasing function, by Lemma 2, Lehmann (1959, p.74), $E_{p_i}(\delta_i(Y))$ is nonincreasing in p_i. So

$$\inf_{\mathcal{P}_i} E_{\boldsymbol{p}}(\phi_i(\boldsymbol{X})) = \inf_{\mathcal{P}_i} E_{p_i}(\delta_i(Y)) = E_{p_i = 1/k}(\delta_i(Y)) = E_{\boldsymbol{p}_0}(\delta_i(X_i)) = P^*$$

by condition (iv). Thus

$$\inf_{\boldsymbol{p}} P_{\boldsymbol{p}}(\mathrm{CS} \mid \boldsymbol{\phi}) = \inf_{1 \le i \le k} P^* = P^*$$

and $\boldsymbol{\phi}$ satisfies the P^*-condition.

Now (2.3) will be verified. By Theorem 3.1 of Berger (1979), the minimax value is kP^*. so it suffices to show that $\sup_{\boldsymbol{p}} E_{\boldsymbol{p}}(S \mid \boldsymbol{\phi}) = kP^*$ for any $\boldsymbol{\phi} \in \mathcal{D}^*$. But, since $S(\boldsymbol{x})$ is Schur concave, by Application 4.2(a) of Nevius, Proschan and Sethuraman (1977) $E_{\boldsymbol{p}}(S(\boldsymbol{X}) \mid \boldsymbol{\phi})$ is Schur concave in $\boldsymbol{p}$ and hence takes on the maximum value at $\boldsymbol{p}_0 = (1/k, \ldots, 1/k)$.

$$E_{\boldsymbol{p}_0}(S(\boldsymbol{X}) \mid \boldsymbol{\phi}) = \sum_{i=1}^k E_{\boldsymbol{p}_0} \phi_i(\boldsymbol{X}) = \sum_{i=1}^k E_{\boldsymbol{p}_0}(\delta_i(X_i)) = kP^*$$

by (iv). So (2.3) is verified.

Finally, by Remark 3.3 of Berger (1979), if $\boldsymbol{\phi}$ is minimax with respect to S then $\boldsymbol{\phi}$ is also minimax with respect to S'.

The fact used in the preceding proof that $E_{\boldsymbol{p}}(S(\boldsymbol{X}) \mid \boldsymbol{\phi})$ is Schur concave in $\boldsymbol{p}$ is also a consequence of Example 2 in Rinott (1973).

Other authors have provided bounds on $E(S)$ for the rules they have proposed. For example, Gupta and Huang (1975) give an upper bound for $E(S)$ when the parameters are in slippage configuration. But the result of Theorem 3.1 is stronger in that minimaxity considers all parameter configurations and the exact upper bound of kP^* for $E(S)$ is achieved.

$\mathcal{D}^*$ is a wide class of selection rules which are minimax. Finding one rule in $\mathcal{D}^*$ which has an additional optimality property may be accomplished by solving the following linear programming problem. Consider $\boldsymbol{\delta} = (\delta_1(0), \ldots, \delta_1(N), \delta_2(0), \ldots, \delta_2(N), \ldots, \delta_k(0), \ldots, \delta_k(N))$ as the solution vector for which we wish to solve. Condition (ii) provides kN linear constraints on the solution. Condition (iii) provides additional linear constraints on the solution. For example, since $(6, 2, 0)$ majorizes $(4, 3, 1)$, we must have $\delta_1(6) + \delta_2(2) + \delta_3(0) \leq \delta_1(4) + \delta_2(3) + \delta_3(1)$. Finally, condition (iv) provides k additional linear constraints on the solution. Subject to these constraints we wish to minimize $E_{\boldsymbol{p}'}(S \mid \boldsymbol{\delta})$, a linear function of the coordinates of $\boldsymbol{\delta}$, for some $\boldsymbol{p}' \neq \boldsymbol{p}_0$. The parameter $\boldsymbol{p}'$ could be some particular parameter value in which the experimenter is particularly interested. As an example, this problem was solved for $P^* = 0.4\,(0.1)\,0.9$ and $\boldsymbol{p}' = (0.1, 0.45, 0.45)$ and $\boldsymbol{p}' = (0.2, 0.4, 0.4)$ in the case where $k = 3$ and $N = 6$. The resulting minimax rules are shown in Table 1. In finding these solutions, the additional constraint was made that the solution was permutation invariant, i.e., $\delta_i(m) = \delta_j(m)$ for all $1 < i, j \leq k$ and all $0 \leq m \leq N$. These solutions were obtained using the NYBLPC computer program. This program uses the criss-cross method

Table 1

Minimax selection rule for $p_{[1]}$ which minimizes $E_{\boldsymbol{p}'}(S)$; $\boldsymbol{p}' = (0.1, 0.45, 0.45)$, $N = 6$, $k = 3$; $\delta(x) = P(\text{select } i\text{th cell} \mid X_i = x)$

x \ P^*	0.9	0.8	0.7	0.6	0.5	0.4
0	1.000	1.000	0.886	0.759	0.633	0.506
1	1.000	1.000	0.886	0.759	0.633	0.506
2	1.000	1.000	0.886	0.759	0.633	0.506
3	1.000	0.545	0.443	0.380	0.316	0.253
4	0.002	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000
$E_{\boldsymbol{p}'}(S \mid \delta)$	2.489	2.206	1.929	1.654	1.378	1.102

$\boldsymbol{p}' = (0.2, 0.4, 0.4)$

x \ P^*	0.9	0.8	0.7	0.6	0.5	0.4
0	1.000	1.000	0.778	0.667	0.556	0.500
1	1.000	1.000	0.778	0.667	0.556	0.500
2	1.000	1.000	0.778	0.667	0.556	0.500
3	1.000	0.545	0.778	0.667	0.556	0.272
4	0.002	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000
$E_{\boldsymbol{p}'}(S \mid \delta)$	2.625	2.336	2.042	1.750	1.458	1.168

as developed by Dr. Stanley Zionts. The computations were done on the CDC CYBER 74 at the Florida State University Computing Center.

4. A Simple rule

An examination of Table 1 reveals that, in all cases when P^* is large, the optimal rule obtained by solving the linear programming problem, as outlined in Section 3, has the following simple form;

$$\phi_i^*(\boldsymbol{x}) = \delta^*(x_i) = \begin{cases} 1 & x_i < t \\ \alpha & x_i = t. \\ 0 & x_i > t \end{cases} \tag{4.1}$$

Here t and α are chosen so that condition (iv) (Section 3) is satisfied. Namely, let Y have a binomial distribution with parameters $1/k$ and N. Then t is the integer which satisfies $P(Y<t) \le P^* < P(Y \le t)$ and $\alpha = (P^* - P(Y<t))/P(Y=t)$. This rule is analogous to a rule proposed by Nagel (1970) for selecting the largest cell probability. This rule is appealing for its simplicity and because the constants t and α, needed for implementation, can be easily obtained from a binomial table. If $P^* < P(Y<N/k) + (1/k)P(Y=N/k)$, the rule, $\boldsymbol{\phi}^*$, may not satisfy $\sum_{i=1}^k \phi_i^*(\boldsymbol{x}) \ge 1$ for all $\boldsymbol{x}$ and may select an empty subset for some $\boldsymbol{x}$. It is assumed that $P^* \ge P(Y<N/k) + (1/k)P(Y=N/k)$ so $\boldsymbol{\phi}^*$ corresponds to a rule which always selects non-empty subsets. In this section the rule $\boldsymbol{\phi}^*$ is studied more closely. First, in Theorem 4.1, $\boldsymbol{\phi}^*$ is shown to be minimax if P^* is sufficiently large. Then, in Lemmas 4.1 and 4.2 and Theorem 4.2, $\boldsymbol{\phi}^*$ is shown to have the following optimality property if the parameters are in a slippage configuration. Suppose all the cell probabilities are equal except for the ith. $\boldsymbol{\phi}^*$ minimizes the probability of selecting the ith cell if its cell probability is larger than all the rest and maximizes the probability of selecting the ith cell if its cell probability is smaller than all the rest among all minimax rules.

Theorem 4.1. *Let Y have a binomial distribution with parameters $1/k$ and N. Then, if*

$$P^* \ge P(Y<N/2) + \tfrac{1}{2}P(Y=N/2), \tag{4.2}$$

$\boldsymbol{\phi}^*$ is minimax with respect to S and S'.

Proof. By Theorem 3.1 it suffices to show $\boldsymbol{\phi}^* \in \mathcal{D}^*$. Conditions (i), (ii) and (iv) are obviously satisfied by the definition of $\boldsymbol{\phi}^*$. It remains to show that (iii) is true, i.e., $S^*(\boldsymbol{x}) = \sum_{i=1}^k \delta^*(x_i)$ is a Schur concave function of $\boldsymbol{x}$.

To see that S^* is Schur concave, suppose $\boldsymbol{x}$ majorizes $\boldsymbol{y}$ where $\sum_{i=1}^k y_i = N = \sum_{i=1}^k x_i$. By the result of Hardy, Littlewood and Polya (1952, page 47), we may assume, without loss of generality, that $\boldsymbol{x}$ and $\boldsymbol{y}$ differ in two coordinates only, say,

$x_i > y_i \geq y_j > x_j$ where $x_i + x_j = y_i + y_j$. Since the constants t and α are chosen so that $P(Y<t)+\alpha P(Y=t)=P^*$, eq. (4.2) implies that either $t>N/2$ or $t=N/2$ and $\alpha \geq \frac{1}{2}$. (Recall that t is defined to be an integer so, if N is odd, t must be greater than $N/2$.) Since all of the coordinates of $\boldsymbol{x}$ and $\boldsymbol{y}$ are equal, except the ith and jth, $S^*(\boldsymbol{y})-S^*(\boldsymbol{x})=\delta^*(y_i)+\delta^*(y_j)-\delta^*(x_i)-\delta^*(x_j)$. Since $\sum_{l=1}^{k} y_l = N$ and $y_l \geq 0$, $1 \leq l \leq k$, $y_j \leq N/2 \leq t$. If $y_j < t$ then $\delta^*(y_j)=\delta^*(x_j)=1$ and $\delta^*(y_i) \geq \delta^*(x_i)$ so $S^*(\boldsymbol{y})-S^*(\boldsymbol{x}) \geq 0$. If $t=y_j$ then $t=N/2=y_j=y_i$ and $S^*(\boldsymbol{y})-S^*(\boldsymbol{x})=\alpha+\alpha-0-1 \geq 0$ since $\alpha \geq \frac{1}{2}$ if $t=N/2$. Thus S^* is Schur concave.

Values of the lower bound given in (4.2) are tabled in Table 2 for $k=2(1)\,10$ and $N=1(1)\,20$. Table 2 reveals that, for small values of k and N, this lower bound is reasonably small. But as k or N increases, the lower bound converges to one. So Theorem 4.1 shows that $\boldsymbol{\phi}^*$ is minimax only for moderate values of k and N. A fact about binomial probabilities, which is proven in Theorem A.1, implies that if N is an odd number then the expression given in (4.2) is the same for both N and $N+1$. (Note that the distribution of Y, as well as the bounds in the probability statements in (4.2), changes as N varies.) That is why one column suffices for each consecutive odd-even pair in Table 2.

Now the behavior of $\boldsymbol{\phi}^*$ when the parameter is in a slippage configuration will be examined. For the remainder of this section let $\boldsymbol{p}'=(p,\ldots,p,\ q,\ p,\ldots,p)$ where $(k-1)p+q=1$ and q is the ith coordinate. The following result will be proven.

Theorem 4.2. *Suppose $\boldsymbol{\phi}^*$ is minimax. Then among all minimax rules, $\boldsymbol{\phi}$, $\boldsymbol{\phi}^*$ maximizes $P_{\boldsymbol{p}'}$* (select ith cell $\mid \boldsymbol{\phi}$) *if $q<p$ and $\boldsymbol{\phi}^*$ minimizes $P_{\boldsymbol{p}'}$* (select ith cell $\mid \boldsymbol{\phi}$) *if $q>p$.*

The proof of Theorem 4.2 will be accomplished via the following two lemmas.

Lemma 4.1. *Suppose $q<p$. Among all rules, $\boldsymbol{\phi}$, which satisfy $P_{\boldsymbol{p}_0}$(select ith cell $\mid \boldsymbol{\phi}) \leq P^*$, $\boldsymbol{\phi}^*$ maximizes $P_{\boldsymbol{p}'}$* (select ith cell $\mid \boldsymbol{\phi}$).

Table 2

Lower bound for P^*; if $P^* \geq$ table entry, $\boldsymbol{\phi}^*$ is minimax

k \ N	1, 2	3, 4	5, 6	7, 8	9, 10	11, 12	13, 14	15, 16	17, 18	19, 20
2	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
3	0.6667	0.7407	0.7901	0.8267	0.8552	0.8779	0.8965	0.9118	0.9245	0.9352
4	0.7500	0.8438	0.8965	0.9294	0.9511	0.9657	0.9757	0.9827	0.9876	0.9911
5	0.8000	0.8960	0.9421	0.9667	0.9804	0.9883	0.9930	0.9958	0.9974	0.9984
6	0.8333	0.9259	0.9645	0.9824	0.9910	0.9954	0.9976	0.9987	0.9993	0.9996
7	0.8571	0.9446	0.9767	0.9898	0.9955	0.9979	0.9991	0.9996	0.9998	0.9999
8	0.8750	0.9570	0.9839	0.9938	0.9975	0.9990	0.9996	0.9998	0.9999	1.0000
9	0.8889	0.9657	0.9885	0.9960	0.9986	0.9995	0.9998	0.9999	1.0000	1.0000
10	0.9000	0.9720	0.9914	0.9973	0.9991	0.9997	0.9999	1.0000	1.0000	1.0000

Proof. For any $\mathbf{p}$ and $\boldsymbol{\phi}$, $P_{\mathbf{p}}$ (select ith cell $|\,\boldsymbol{\phi}) = E_{\mathbf{p}}\phi_i(\mathbf{X})$. By the Neyman–Pearson Lemma, among all rules which satisfy $E_{\mathbf{p}_0}\phi_i(\mathbf{X}) \le P^*$, a rule which maximizes $E_{\mathbf{p}}\phi_i(\mathbf{X})$ is given by

$$\phi_i(\mathbf{x}) = \begin{cases} 1 & \dfrac{\binom{N}{\mathbf{x}} q^{x_i} p^{N-x_i}}{\binom{N}{\mathbf{x}} k^{-N}} > t, \\ & \qquad\qquad = t, \\ 0 & \qquad\qquad < t, \end{cases} \tag{4.3}$$

where t and α are chosen so that $E_{\mathbf{p}_0}\phi_i(\mathbf{X}) = P^*$. Using the fact that $q < p$, it can be seen that (4.3) is equivalent to (4.1).

Lemma 4.2. *Let $q > p$. Among all rules, $\boldsymbol{\phi}$, which satisfy $P_{\mathbf{p}_0}$* (select ith cell $|\,\boldsymbol{\phi}) \ge P^*$, *$\boldsymbol{\phi}^*$ minimizes $P_{\mathbf{p}}$* (select ith cell $|\,\boldsymbol{\phi}$).

Proof. The proof follows the same lines as that of Lemma 4.1.

Proof of Theorem 4.2. By Theorem 4.1 of Berger (1979), every rule, $\boldsymbol{\phi}$, which is minimax with respect to S or S' must satisfy $P_{\mathbf{p}_0}$ (select ith cell $|\,\boldsymbol{\phi}) = P^*$. So Theorem 4.2 follows from Lemmas 4.1 and 4.2.

Seal (1955, Section 4) examines a slippage problem, analogous to the one considered in Theorem 4.2, for the normal means selection problem. He shows, 'approximately,' that a certain rule has the property that $\boldsymbol{\phi}^*$ has in the multinomial problem. 'Approximately' is Seal's term. It refers to the fact that he used an asymptotic argument. But his result can be proved exactly using the Neyman–Pearson Lemma as in Lemmas 4.1 and 4.2.

5. Selection in terms of the largest cell probability

In some problems the experimenter might be interested in selecting a subset of the cells including the cell associated with $p_{[k]}$, the largest cell probability. In this problem a correct selection, CS, is the selection of any subset including the cell associated with $p_{[k]}$. In this section, results analogous to those found in Sections 3 and 4 will be briefly outlined.

The following class of selection rules can be shown to be minimax with respect to S and S'. Let $\mathscr{D}_*$ be the class of selection rules which satisfy (i), (ii'), (iii) and (iv) where (i), (iii) and (iv) are as in Section 3 and (ii') $\delta_i(m) \le \delta_i(n)$ if $0 \le m \le n \le N$ and $1 \le i \le k$, i.e., the probability of selecting the ith cell increases as the number of observations in the ith cell increases.

Theorem 5.1. *If $\boldsymbol{\phi} \in \mathcal{D}_*$, then $\boldsymbol{\phi}$ is minimax with respect to S and S'.*

Proof. This proof follows the lines of the proof of Theorem 3.1 where now (ii') is used to show that if $\boldsymbol{\phi} \in \mathcal{D}_*$, then $\boldsymbol{\phi}$ satisfies the P^*-condition.

Rules in $\mathcal{D}_*$ can be constructed by solving a linear programming problem. The $\boldsymbol{\delta}$ which minimizes $E_{\boldsymbol{p}'}(S \mid \boldsymbol{\delta})$ for some $\boldsymbol{p}' \neq \boldsymbol{p}_0$, subject to the constraint $\boldsymbol{\delta} \in \mathcal{D}_*$ can be solved for. This was done for $P^* = 0.4(0.1)0.9$ and $\boldsymbol{p}' = (0.8, 0.1, 0.1)$ and $\boldsymbol{p}'' = (0.6, 0.2, 0.2)$ in the case where $k = 3$ and $N = 6$. The same rule was found to be optimal for both $\boldsymbol{p}'$ and $\boldsymbol{p}''$. The resulting minimax rule is presented in Table 3.

A particularly simple rule, which arose as the solution to the linear programming problem when P^* was large, is the rule proposed by Nagel (1970):

$$\phi_{*_i}(\boldsymbol{x}) = \delta_*(x_i) = \begin{cases} 1 & \text{if } x_i > t, \\ \alpha & \text{if } x_i = t, \\ 0 & \text{if } x_i < t, \end{cases} \tag{5.1}$$

where α and t are chosen so that condition (iv) is satisfied. The following two theorems, whose proofs are omitted since they are similar to the proofs of Theorems 4.1 and 4.2, describe some properties of $\boldsymbol{\phi}_*$.

Theorem 5.2. *Let Y have a binomial distribution with parameters $1/k$ and N. Then, if*

$$P^* \geq P(Y > 1) + \tfrac{1}{2}P(Y = 1), \tag{5.2}$$

$\boldsymbol{\phi}_*$ is minimax with respect to S and S'.

Theorem 5.3. *Let $\boldsymbol{p}' = (p, \ldots, p, q, p, \ldots, p)$ where $(k-1)p + q = 1$ and q is the ith coordinate. Suppose $\boldsymbol{\phi}_*$ is minimax. Then among all minimax rules, $\boldsymbol{\phi}$, $\boldsymbol{\phi}_*$ maximizes $P_{\boldsymbol{p}'}$ (select ith cell $\mid \boldsymbol{\phi}$) if $q > p$ and $\boldsymbol{\phi}_*$ minimizes $P_{\boldsymbol{p}'}$ (select ith cell $\mid \boldsymbol{\phi}$) if $q < p$.*

Table 3

Minimax selection rule for $p_{[k]}$ which minimizes $E_{\boldsymbol{p}'}(S)$ and $E_{\boldsymbol{p}''}(S)$; $\boldsymbol{p}' = (0.8, 0.1, 0.1)$; $\boldsymbol{p}'' = (0.6, 0.2, 0.2)$, $N = 6$, $k = 3$; $\delta(x) = P(\text{select } i\text{th cell} \mid X_i = x)$

x \ P^*	0.9	0.8	0.7	0.6	0.5	0.4
0	0.000	0.000	0.000	0.000	0.000	0.000
1	0.954	0.574	0.448	0.384	0.320	0.253
2	1.000	1.000	0.897	0.768	0.639	0.507
3	1.000	1.000	0.897	0.768	0.639	0.507
4	1.000	1.000	0.897	0.768	0.639	0.507
5	1.000	1.000	0.897	0.768	0.680	0.747
6	1.000	1.000	1.000	1.000	1.000	1.000
$E_{\boldsymbol{p}'}(S \mid \delta)$	1.904	1.635	1.446	1.276	1.122	1.025
$E_{\boldsymbol{p}''}(S \mid \delta)$	2.433	2.121	1.852	1.594	1.341	1.112

Table 4

Lower bound for P^*; if $P^* \geq$ table entry, ϕ_* is minimax

N \	2	3	4	5	6	7	8	9	10
1	0.2500	0.1667	0.1250	0.1000	0.0833	0.0714	0.0625	0.0556	0.0500
2	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111	0.1000
3	0.6875	0.4815	0.3672	0.2960	0.2477	0.2128	0.1865	0.1660	0.1495
4	0.8125	0.6049	0.4727	0.3856	0.3248	0.2803	0.2463	0.2196	0.1981
5	0.8906	0.7037	0.5649	0.4675	0.3972	0.3446	0.3039	0.2717	0.2455
6	0.9375	0.7805	0.6440	0.5412	0.4642	0.4051	0.3589	0.3218	0.2914
7	0.9648	0.8390	0.7108	0.6068	0.5256	0.4618	0.4110	0.3697	0.3357
8	0.9805	0.8829	0.7664	0.6645	0.5814	0.5144	0.4600	0.4154	0.3782
9	0.9893	0.9155	0.8123	0.7148	0.6318	0.5630	0.5061	0.4587	0.4189
10	0.9941	0.9393	0.8498	0.7584	0.6770	0.6076	0.5490	0.4996	0.4576
11	0.9968	0.9566	0.8803	0.7960	0.7174	0.6483	0.5889	0.5381	0.4944
12	0.9983	0.9692	0.9050	0.8282	0.7533	0.6855	0.6259	0.5742	0.5293
13	0.9991	0.9782	0.9248	0.8557	0.7850	0.7192	0.6601	0.6080	0.5622
14	0.9995	0.9846	0.9406	0.8791	0.8131	0.7497	0.6916	0.6395	0.5933
15	0.9997	0.9892	0.9532	0.8988	0.8377	0.7772	0.7205	0.6689	0.6225
16	0.9999	0.9924	0.9633	0.9156	0.8594	0.8019	0.7470	0.6962	0.6500
17	0.9999	0.9947	0.9712	0.9296	0.8783	0.8242	0.7712	0.7215	0.6757
18	1.0000	0.9963	0.9774	0.9415	0.8948	0.8441	0.7934	0.7450	0.6998
19	1.0000	0.9974	0.9824	0.9514	0.9092	0.8619	0.8136	0.7666	0.7223
20	1.0000	0.9982	0.9863	0.9596	0.9217	0.8778	0.8319	0.7866	0.7433

For fixed values of N and P^*, the lower bound in eq. (5.2) is a decreasing function of k. The lower bound in eq. (5.3) is tabulated in Table 4 for $N = 1(1)\ 20$ and $k = 2(1)\ 10$.

Appendix

Theorem A.1. *Let m be a positive integer and p a fixed probability. Let Y have a binomial distribution with parameters p and $N = 2m - 1$. Let Z have a binomial distribution with parameters p and $N + 1 = 2m$. Then*

$$P(Y < N/2) + \tfrac{1}{2}P(Y = N/2) = P(Y < N/2) = P\left(Z < \frac{N+1}{2}\right) + \tfrac{1}{2}P\left(Z = \frac{N+1}{2}\right). \tag{A.1}$$

Proof. The first equality in (A.1) is true since N is odd.

Let X be a Bernoulli random variable with success probability p which is independent of Y. Z has the same distribution as $Y + X$. Let $q = 1 - p$. Then the following equalities are true.

$$P\left(Y + X < \frac{N+1}{2}\right) = P(Y + X \leq m - 1) = P(Y < m - 1) + P(Y = m - 1, X = 0) \tag{A.2}$$

$$\begin{aligned}P(Y=m-1, X=1) &= P(Y=m-1)P(X=1)\\ &= \left[\binom{N}{m-1}p^{m-1}q^{m}\right]p\\ &= \left[\binom{N}{m}p^{m}q^{m-1}\right]q\\ &= P(Y=m)P(X=0)\\ &= P(Y=m, X=0). \end{aligned} \tag{A.3}$$

Using (A.3), eq. (A.4) can be obtained.

$$\begin{aligned}\tfrac{1}{2}P\left(Y+X=\frac{N+1}{2}\right) &= \tfrac{1}{2}P(Y+X=m)\\ &= \tfrac{1}{2}[P(Y=m-1, X=1)+P(Y=m, X=0)]\\ &= P(Y=m-1, X=1). \end{aligned} \tag{A.4}$$

Using eqs. (A.2) and (A.4), the last equality in (A.1) can be obtained in this way.

$$\begin{aligned}P(Y<N/2) &= P\left(Y\le\frac{N-1}{2}\right)=P(Y\le m-1)\\ &= P(Y<m-1)+P(Y=m-1)\\ &= P(Y<m-1)+P(Y=m-1, X=0)+P(Y=m-1, X=1)\\ &= P\left(Y+X<\frac{N+1}{2}\right)+\tfrac{1}{2}P\left(Y+X=\frac{N+1}{2}\right)\\ &= P\left(Z<\frac{N+1}{2}\right)+\tfrac{1}{2}P\left(Z=\frac{N+1}{2}\right). \end{aligned}$$

References

Berger, R.L. (1979). Minimax subset selection for loss measured by subset size. *Ann. Statist.* 7, 1333–1338.

Berger, R.L. (1979b). A minimax and admissible subset selection rule for the least probable multinomial cell. *Tech. Report* M552. Dept. of Statistics, Florida State University.

Bjørnstad, J.F. (1978). The subset selection problem. II. On the optimality of some subset selection procedures. Mimeograph series No. 78-27. Dept. of Statistics, Purdue University.

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

Gupta, S.S. and D.Y. Huang, (1975). On subset selection procedures for Poisson populations and some applications to the multinomial selection problems. In: R.P. Gupta, ed., *Applied Statistics*. North Holland, Amsterdam.

Gupta, S.S. and K. Nagel (1967). On selection and ranking procedures and order statistics from the multinomial distribution. *Sankhya Ser. B* 29, 1–34.

Hardy, G.H., J.E. Littlewood and G. Polya (1959). *Inequalities*. Cambridge University Press, London.

Lehmann, E.L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

Nagel, K. (1970). On subset selection rules with certain optimality properties. Ph.D. Thesis, Purdue University.

Nevius, S.E., F. Proschan and J. Sethuraman (1977). Schur functions is statistics II. Stochastic majorization. *Ann. Statist.* 5, 263–274.

NYBLPC, State University of New York at Buffalo Computing Center Press.

Panchapakesan, S. (1971). On a subset selection procedure for the most probable event in a multinomial distribution. In: S.S. Gupta and J. Yackel, eds., *Statistical Decision Theory and Related Topics*. Academic Press.

Proschan, F. and J. Sethuraman (1977). Schur functions in statistics I. The preservation theorem. *Ann. Statist.* 5, 256–263.

Rinott, Y. (1973). Multivariate majorization and rearrangement inequalities with some applications to probability and statistics. *Israel J. Math.* 15, 60–77.

Seal, K.C. (1955). On a class of decision procedures for ranking the means of normal populations. *Ann. Math. Statist.* 26, 387–398.