Ann. Inst. Statist. Math.
Vol. 40, No. 3, 615–625 (1988)

AN INTEGRATED FORMULATION FOR SELECTING THE MOST PROBABLE MULTINOMIAL CELL

PINYUEN CHEN

Department of Mathematics, Syracuse University, 200 Carnegie, Syracuse, NY 13244-1150, U.S.A.

(Received June 9, 1986; revised June 23, 1987)

Abstract. We refer to the two classical approaches to multinomial selection as the indifference zone approach and the subset selection approach. This paper integrates these two approaches by separating the parameter space into two disjoint subspaces: the preference zone (PZ) and the indifference zone (IZ). In the PZ we insist on selecting the best (most probable) cell for a correct selection (CS_1) but in the IZ we define any selected subset to be correct (CS_2) if it contains the best cell. We then propose a single stage procedure R to achieve the selection goals CS_1 and CS_2 simultaneously with certain probability requirements. It is shown that both the probability of a correct selection under PZ, $P(CS_1|PZ)$, and the probability of a correct selection under IZ, $P(CS_2|IZ)$, satisfy some monotonicity properties and the least favorable configuration in PZ and the worst configuration in IZ can be found by these properties.

Key words and phrases: Indifference zone approach, indifference zone, least favorable configuration, most probable cell, multinomial distribution, subset selection formulation, worst configuration.

1. Introduction

This paper considers an integrated approach to ranking and selection of the multinomial cells. The two major approaches which we refer to as the indifference zone approach and subset selection approach are well known and have been treated in Bechhofer *et al.* (1959) and Gupta and Nagel (1967), respectively. In the present paper these two approaches are combined in a meaningful manner to form a new integrated approach for selecting among multinomial cells. The basic idea is to introduce both a preference zone (PZ) and an indifference zone (IZ) (these are defined in Section 2 below); in the former zone PZ our goal is to find and select precisely the cell with the largest cell probability and in the latter zone IZ our goal is to select a random-size subset of size at least two which contains the best cell. The proposed composite procedure R for selecting the best cell makes use of the data in a fixed sample size procedure to determine which part of the procedure should be used. This is a meaningful approach to the problem since when the parameters are close together (in the IZ) we can avoid an unnecessarily large sample size by adopting the weaker goal of selecting a subset containing the best. This integrated formulation has been used by Chen and Sobel (1987*a*) for selecting from normal populations in terms of the means, and by Chen and Sobel (1987*b*) for selecting the most probable multinomial cell by using an inverse sampling procedure. Formally we set up two different probability requirements (i.e., P^*-conditions), one for the PZ and the other for the IZ; the required constants to specify our composite procedure R are then determined so as to satisfy both P^*-condition. In Section 2 we write out the goal of selecting the largest cell probability in a former manner, give the proposed procedure R, define the concept of a correct decision explicitly and discuss the infimum of the probability of a correct selection. In Section 3, we present the table of the probability of a correct selection under the least favorable configuration, the probability of a correct selection under the worst configuration and the corresponding expected selected subset size.

2. Selecting the largest cell probability

A multinomial distribution with k cells is given; let the ordered values of the cell probabilities be denoted by

$$p_{[1]} \leq \cdots \leq p_{[k-1]} < p_{[k]}\,, \tag{2.1}$$

where we assumed that $p_{[k-1]}$ is strictly less than $p_{[k]}$ in order that the best cell should be well-defined. If we let δ denote the ratio $p_{[k]}/p_{[k-1]}$, then by (2.1), $\delta > 1$. The set of parameter vectors $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$ for which $\delta \geq \delta^*$ (where $\delta^* > 1$ is specified) will be called the preference zone (PZ); the complementary set of $\boldsymbol{p}$ for which $1 < \delta < \delta^*$ will be called the indifference zone (IZ).

We define our goal in two parts, according to whether the true parameter $\boldsymbol{p}$ is in the PZ or the IZ, as follows:

$$\text{Goal}\begin{cases} \text{For} \quad \boldsymbol{p} \in PZ & \text{we want to select the best with probability at least } P_1^*\,, \\ \text{For} \quad \boldsymbol{p} \in IZ & \text{we want to select a subset containing the best with probability at least } P_0^*\,. \end{cases} \tag{2.2}$$

Thus in the PZ a correct decision (CD) is the same as a correct selection of the best and the selection of a subset of size at least two is always incorrect. However, in the IZ we can make a correct decision in two ways: either by selecting the best or by selecting a random subset of size at least two containing the best. Hence the goal in (2.2) can be restated as

$$P(CD) \geq \begin{cases} P_1^* & \text{for} \quad \delta \geq \delta^*, \\ P_0^* & \text{for} \quad 1 < \delta < \delta^*, \end{cases} \tag{2.3}$$

where both P_1^* and P_0^* (as well as δ^*) are specified.

Let $f_i (i = 1,\ldots,k)$ be the observed frequencies in the i-th cell π_i and let $\sum_{i=1}^{k} f_i = n$. The ordered values of f's are defined by

$$f_{[1]} \leq f_{[2]} \leq \cdots \leq f_{[k]}. \tag{2.4}$$

Then the procedure R for selecting goal (2.2) is defined as follows.

PROCEDURE R. If $f_{[k]} - f_{[k-1]} > c$, then we select the cell that gives rise to the largest frequency $f_{[k]}$.

(2.5) If $f_{[k]} - f_{[k-1]} \leq c$, then we select a random-size subset consisting of all those cells π with frequencies $f_i > f_{[k-1]} - d$ $(i = 1,\ldots,k)$.

It should be noted that we have three constants (c, d and n) to determine and only two conditions (1.1) and (1.2) to determine them. Hence we can regard any one of them as fixed and determine the other two. If n is fixed, then our formulation is closer to the "Subset Selection" approach; if one of the others is fixed and we determine n (and the remaining constant), then our formulation is closer to the "Indifference Zone" approach.

We need some notations before we discuss the infimum of the probability of a correct decision ($P(CD)$) under the procedure R. Let $F(p_1, p_2,\ldots,p_k; f_1, f_2,\ldots,f_k)$ denote the probability of a multinomial distribution with the cell probabilities $p_1, p_2,\ldots,p_k$ and their corresponding frequencies $f_1, f_2,\ldots,f_k$ where $\sum_{i=1}^{k} f_i = n$. To derive the probability of a correct selection $P(CS_1)$ for $\boldsymbol{p} \in PZ$, we first note that a CS_1 in the PZ can occur only if we select the best cell and no others, i.e., if and only if

$$P(CS_1 \mid PZ) = \sum_{\substack{f_k - c > f_i \\ i=1,2,\ldots,k}} F(p_{[1]},\ldots,p_{[k]}; f_1,\ldots,f_k). \tag{2.6}$$

In the expression $F(p_1, p_2,\ldots,p_k; f_1,\ldots,f_k)$, we fix all the p's and f's except p_i, p_j, f_i and f_j. Now we define

$$x = p_i, \quad w(x) = 1 - x - \sum_{\substack{\alpha=1 \\ \alpha \neq i,j}}^{k} p_\alpha , \tag{2.7}$$

then F becomes a function of three variables x, f_i and f_j. Denote this function by $g_n(x, f_i, f_j)$, i.e.,

$$g_n(x, f_i, f_j) = n! \frac{x^{f_i}(w(x))^{f_j}}{f_i!\, f_j!} \prod_{\alpha \neq i,j} \frac{p_\alpha^{f_\alpha}}{f_\alpha!} . \tag{2.8}$$

The partial derivative of g_n in (2.8) with respect to x is

$$\frac{\partial g_n(x, f_i, f_j)}{\partial x} \tag{2.9}$$

$$= \begin{cases} n[g_{n-1}(x, f_i - 1, f_j) - g_{n-1}(x, f_i, f_j - 1)] & \text{for } f_i, f_j \geq 1 , \\ -n g_{n-1}(x, f_i, f_j - 1) & \text{for } f_i = 0 \text{ and } f_j \neq 0 , \\ n g_{n-1}(x, f_i - 1, f_j) & \text{for } f_j = 0 \text{ and } f_i \neq 0 , \\ 0 & \text{for } f_i = f_j = 0 . \end{cases}$$

To find the least favorable configuration (LFC) when $\boldsymbol{p} \in PZ$, we need the following two lemmas on the monotonicity properties of $P(CS_1 | PZ)$. The results that we obtain are quite similar to those of Kesten and Morse (1959) where the authors found the LFC for the procedure proposed by Bechhofer *et al.* (1959). In our formulation c is at least 0 and thus any tie for the first place is not possible. Hence randomization among the cells tied for the first place is not required in our proposed procedure R. Therefore, the result in Kesten and Morse (1959) is not applicable here.

LEMMA 2.1. *Keeping the sum* $p_{[i]} + p_{[k]}$, $1 \leq i < k$, *constant, the* $P(CS_1 | PZ)$ *as given in* (2.6) *decreases as we pass from the configuration* $(p_{[1]}, \ldots, p_{[i]}, \ldots, p_{[k]})$ *to* $(p_{[1]}, \ldots, p_{[i]} + \varepsilon, \ldots, p_{[k]} - \varepsilon)$ *where* $0 < \varepsilon \leq p_{[k]}$.

PROOF. Let $x = p_{[i]}$ and $w(x) = 1 - x - \sum_{\substack{\alpha \neq i \\ \alpha \neq k}} p_{[\alpha]}$. For a typical term $g_n(x, f_i, f_k) = F(p_{[1]}, \ldots, p_{[k]}; f_1, \ldots, f_k)$ in $P(CS_1 | PZ)$ of (2.6), the positive term in the partial derivative of $g_n(x, f_i, f_k)$ with respect to x is, by (2.9), $n g_{n-1}(x, f_i - 1, f_k)$. It is clear that $g_n(x, f_i - 1, f_{k+1})$ must also be in $P(CS_1 | PZ)$ since $f_i - 1 < f_i < f_k - c < f_k + 1 - c$. The partial derivative of $g_n(x, f_i - 1, f_k + 1)$ has a negative term (by (2.9)) $n g_{n-1}(x, f_i - 1, f_k)$ which cancels the positive term in the derivative of the typical term $g_n(x, f_i, f_k)$ in

$P(CS_1|PZ)$. Thus $P(CS_1|PZ)$ is a non-increasing function in $x = p_{[i]}$. This completes the proof of the lemma.

LEMMA 2.2. *Keeping the sum* $p_{[i]} + p_{[j]}$, $1 \le i < j < k$, *constant, the* $P(CS_1|PZ)$ *as given in* (2.6) *decreases as we pass from the configuration* $(p_{[1]},\dots,p_{[i]},\dots,p_{[j]},\dots,p_{[k]})$ *to* $(p_{[1]},\dots,p_{[i]} - \varepsilon,\dots,p_{[j]} + \varepsilon,\dots,p_{[k]})$ *where* $0 < \varepsilon \le p_{[i]}$.

PROOF. Let $x = p_{[i]}$ and $w(x) = 1 - x - \sum_{\alpha \ne i,j} p_{[\alpha]}$. For a typical term $g_n(x, f_i, f_j) = F(p_{[1]},\dots,p_{[k]}; f_1,\dots,f_k)$ in $P(CS_1|PZ)$ of (2.6), we consider the following two situations:

(1) If $f_i + 1 < f_k - c$, then $g_n(x, f_i + 1, f_j - 1)$ is in $P(CS_1|PZ)$. Thus the negative term in $\partial g_n(x, f_i, f_j)/\partial x$, i.e., $-ng_{n-1}(x, f_i, f_j - 1)$, can be cancelled with the positive term $ng_{n-1}(x, f_n, f_i - 1)$ in the derivative of the term $g_n(x, f_i + 1, f_j - 1)$ which is also in $P(CS_1|PZ)$.

(2) If $f_i + 1 = f_k - c$, then we investigate the term $g_n(x, f_j, f_i)$. It is in $P(CS_1|PZ)$ since $g_n(x, f_i, f_j)$ is in $P(CS_1|PZ)$ and f_i and f_j satisfy the conditions that $f_i < f_k - c$ and $f_j < f_k - c$. The positive term $ng_n(x, f_j - 1, f_i)$ in $\partial g_n(x, f_j, f_i)/\partial x$ is at least $ng_{n-1}(x, f_i, f_j - 1)$ since $f_i = f_k - c - 1 > f_j - 1$ and $p_{[i]} \le p_{[j]}$.

Thus in both situations (1) and (2), we can always find a positive term in the derivative of $P(CS_1|PZ)$ that cancels the negative term in the derivative of the typical term from $P(CS_1|PZ)$. Hence $P(CS_1|PZ)$ is a non-increasing function in $x = p_{[i]}$. This completes the proof of the lemma.

The overall minimum of $P(CS_1|PZ)$ has to be at a configuration which can't be changed to one with a smaller probability by using the above two lemmas. We have the following theorem.

THEOREM 2.1. *Under procedure R the LFC for the* $P(CS_1|PZ)$ *is given by the configuration of the type*:

$$(0,\dots,0,s,p,p,\dots,p,\delta^* p), \quad s \le p\,. \tag{2.10}$$

PROOF. In the PZ, consider an arbitrary $\boldsymbol{p}$, i.e.,

$$p_{[1]} \le p_{[2]} \le \cdots \le p_{[k]} \quad \text{where} \quad p_{[k]}/p_{[k-1]} \ge \delta^*\,. \tag{2.11}$$

We apply Lemma 2.1 to $p_{[k-1]}$ and $p_{[k]}$. By moving $p_{[k-1]}$ upward to $p_{[k]}$ and keep all the other p's fixed, we cannot increase $P(CS_1|PZ)$. However, the ratio $p_{[k]}/p_{[k]}$ is at least δ^*. Thus $P(CS_1|PZ)$ under (2.11) is minimized when $p_{[k]}/p_{[k-1]} = \delta^*$. Now we work on the p's which are less than the new $p_{[k-1]}$. By applying Lemma 2.2 to $p_{[1]}$ and $p_{[k-2]}$, and can move either $p_{[1]}$ to

0 or $p_{[k-2]}$ to $p_{[k-1]} = p_{[k]}/\delta^*$. If $p_{[1]}$ reaches 0 first, we apply Lemma 2.2 again to $p_{[2]}$ and $p_{[k-1]}$. If $p_{[k-2]}$ reaches $p_{[k-1]}$ first, we apply Lemma 2.2 to $p_{[1]}$ and $p_{[k-3]}$. Repeating the above argument and each time applying Lemma 2.2, we finally reach a configuration of the type (2.10) for which $P(CS_1|PZ)$ is a minimum among configurations in (2.11). This completes the proof of the theorem.

Now we restrict the true parameter $\boldsymbol{p}$ to be an arbitrary configuration in the IZ and look for the worst configuration (WC) in the IZ. The results of Lemmas 2.3, 2.4 and Theorem 2.2 below are similar in trend and nature of results to those of Gupta and Nagel (1967). Here we use two constants c and d (rather than just one constant in Gupta and Nagel (1967)) in our procedure in order to satisfy our composite formulation. Thus the results of Gupta and Nagel (1967) cannot be applied directly here. Again we consider two lemmas which are analogous to Lemmas 2.1 and 2.2. The result of $fP(CS_2|IZ)$ is a sum of two parts, the first of which (denoted by P_1) is exactly the same as in (2.6). The second part (denoted by P_2) assumes that the largest cell frequency $f_{[k]}$ is less than or equal to $f_{[k-1]} + c$ and the frequency f_k of the cell associated with the cell probability $p_{[k]}$ is larger than $f_{[k-1]} - d$. Thus we can write the $P(CS_2|IZ)$ as follows:

$$(2.12)\qquad \begin{aligned} P(CS_2|IZ) &= P_1 + P_2 \\ &= \sum_{\substack{f_k - c > f_i \\ i=1,\dots,k-1}} F(p_{[1]},\dots,p_{[k]}; f_1,\dots,f_k) \\ &\quad + \sum_{\substack{f_{[k]}-f_{[k-1]}\le c \\ f_k + d > f_{[k-1]}}} F(p_{[1]},\dots,p_{[k]}; f_1,\dots,f_k)\,. \end{aligned}$$

LEMMA 2.3. *Keeping the sum* $p_{[i]} + p_{[k]}$, $1 \le i < k$, *constant, the* $P(CS_2|IZ)$ *as given in* (2.12) *decreases as we pass from the configuration* $(p_{[1]},\dots,p_{[i]},\dots,p_{[k]})$ *to* $(p_{[1]},\dots,p_{[i]}+\varepsilon,\dots,p_{[k]}-\varepsilon)$ *where* $0 < \varepsilon \le p_{[k]}$.

PROOF. Let $x = p_{[i]}$ and $w(x) = 1 - x - \sum_{i\neq 1,k} p_{[\alpha]}$. We only have to consider the derivatives of the terms in P_2 since the result of Lemma 2.1 is applicable to P_1. For a typical term $g_n(x, f_i, f_k) = F(p_{[1]},\dots,p_{[k]}; f_1,\dots,f_k)$ in P_2, the negative part of the derivative is $ng_{n-1}(x, f_i - 1, f_k)$. It is clear that the term $g_n(x, f_i - 1, f_k + 1)$ must be either in P_1 (when $f_k + 1 > f_i + c$, $i = 1,\dots,k-1$), or in P_2 since $f_k + 1 + d > f_k + d > f_{[k-1]}$. Moreover, the term $g_n(x, f_i - 1, f_k + 1)$ has not been used to cancel the negative part in the derivative in P_1 since the only term in $P(CS_2|IZ)$ that gives $-ng_{n-1}(x, f_i - 1, f_k)$ in its derivative is $g_n(x, f_i, f_k)$ which must be in P_2. The positive term in the derivative of the term $g_n(x, f_i - 1, f_k + 1)$ is $ng_{n-1}(x, f_i - 1, f_k)$ which cancels the negative term in the derivative of the typical term

$g_n(x, f_i, f_k)$ in P_2. Thus $P(CS_2|IZ)$ is a non-decreasing function of x. This completes the proof of the lemma.

LEMMA 2.4. *Keeping the sum* $p_{[i]} + p_{[j]}$, $1 \le i \le j < k$, *constant,* $P(CS_2|IZ)$ *as given in* (2.12) *decreases as we pass from the configuration* $(p_{[1]}, \ldots, p_{[i]}, \ldots, p_{[j]}, \ldots, p_{[k]})$ *to* $(p_{[1]}, \ldots, p_{[i]} - \varepsilon, \ldots, p_{[j]} + \varepsilon, \ldots, p_{[k]})$ *where* $0 < \varepsilon \le p_{[i]}$.

PROOF. Let $x = p_{[i]}$ and $w(x) = 1 - \sum_{\alpha \neq i,j} p_{[\alpha]}$. For a typical term $g_n(x, n_i, n_j) = F(p_{[1]}, p_{[2]}, \ldots, p_{[k]}; f_1, f_2, \ldots, f_k)$ in (2.12), we consider the following two situations:

(1) If the term $g_n(x, n_i + 1, n_j - 1)$ is in $P(CS_2|IZ)$, then the negative part of the derivative of $g_n(x, n_i, n_j)$ can be cancelled with the positive part of the derivative of $g_n(x, n_i + 1, n_j - 1)$.

(2) If the term $g_n(x, n_i + 1, n_j - 1)$ is not in $P(CS_2|IZ)$, then $g_n(x, n_j - 1, n_i + 1)$ is not in $P(CS_2|IZ)$ either. We also know that $n_i + 1 \ge n_j$. The positive part in the derivative of $g_n(x, n_j, n_i)$, namely, $ng_{n-1}(x, n_j - 1, n_i)$ is in the derivative of $P(CS_2|IZ)$ and cannot be cancelled with the negative part in the derivative of any term in $P(CS_2|IZ)$ since $g_n(x, n_j - 1, n_i + 1)$ is not in $P(CS_2|IZ)$. Thus we can use $ng_{n-1}(x, n_j - 1, n_i)$ to cancel the negative part in the derivative of the term $g_n(x, n_i, n_j)$ since $g_{n-1}(x, n_i, n_j - 1) \le g_{n-1}(x, n_j - 1, n_i)$ where $n_i \ge n_j - 1$.

Thus in either situation, the negative part from the derivative of the typical term will always be cancelled or exceeded by a positive term. Thus $P(CS_2|IZ)$ is a non-decreasing function of x. This completes the proof of the lemma.

The overall minimum of $P(CS_2|IZ)$ has to be at a configuration which cannot be changed to one with a smaller probability by using the above two lemmas. Hence we have the following theorem on the WC of procedure R.

THEOREM 2.2. *Under procedure* R, *the* WC *for the* $P(CS_2|IZ)$ *as is given by the configuration of the type*:

$$(0, 0, \ldots, 0, s, p, \ldots, p), \quad s \le p. \tag{2.13}$$

PROOF. By applying Lemmas 2.3 and 2.4, the proof of this theorem is analogous to that of Theorem 2.1 and so is omitted here.

For the procedure R, the size S of the selected subset is a random variable which can take on integer values from 1 to k. The desired result for $E(S)$ can be written as

$$(2.14)\qquad E(S) = \sum_{f_{[k]}-c>f_{[k-1]}} F(p_1, p_2, \ldots, p_k; f_1, \ldots, f_k) + \sum_{f_{[k]}-c\le f_{[k-1]}} F(p_1, p_2, \ldots, p_k; f_1, \ldots, f_k) B(f_1, \ldots, f_k)\,,$$

where $B(f_1, f_2, \ldots, f_k)$ = number of f_i's $> f_{[k-1]} - d$. The first sum is the expected value of S when the first part of the procedure is used and the second sum is the expected value of S when the second part of the procedure is used.

3. Tables and remarks

Table 1 gives the probability of a correct selection $P(CS_1 | PZ)$ under the LFC for $k = 2, 3, 4$; $c = 1, 2, 3$ and $\delta^* = 2.0$ and 5.0. Since the LFC is in the form (2.10), the overall minimum can be found by

$$(3.1)\qquad \min_{\boldsymbol{P}} P(CS_1 | \boldsymbol{P} \in PZ) = \min_{r=2,\ldots,k} (\min P(CS_1 | (0, \ldots, 0, s, p, \ldots, p, \delta^* p)))\,,$$

where $s = 1 - \delta^* p - (r-2)p$, and r is the number of positive cell proba-

Table 1. $P(CS_1 | LFC)$: The minimum probability of a correct selection in the preference zone.

$\delta^* = 2.0$

	$k=2$			$k=3$			$k=4$		
n \ c	1	2	3	1	2	3	1	2	3
6	.6804	.3512	.3512	.3438	.2266	.1094	.2406	.1331	.4100
7	.5706	.5706	.2634	.4316	.2266	.1445	.2683	.1393	.0705
8	.7514	.4682	.4682	.4658	.3086	.1445	.3285	.1599	.0773
9	.6503	.6503	.3772	.4692	.3462	.2129	.3533	.2108	.0911
10	.7869	.5593	.5593	.5308	.3513	.2488	.3706	.2364	.1291
11	.7110	.7110	.4726	.5564	.4154	.2543	.3991	.2538	.1513
12	.8223	.6315	.6315	.5635	.4456	.3147	.4343	.2805	.1663
13	.7587	.7587	.5520	.6080	.4542	.3462	.4550	.3163	.1886
14	.8505	.6898	.6898	.6287	.5041	.3552	.4740	.3387	.2208
15	.7970	.7970	.6184	.6366	.5292	.4062	.4964	.3587	.2422

$\delta^* = 5.0$

6	.9377	.7368	.7368	.7703	.6109	.4516	.6533	.4888	.2742
7	.9042	.9042	.6698	.8386	.6792	.5199	.7257	.5379	.3815
8	.9693	.8652	.8652	.8766	.7703	.5882	.7922	.6201	.4323
9	.9520	.9520	.8217	.8982	.8202	.6966	.8274	.7042	.5164
10	.9845	.9303	.9303	.9289	.8499	.7570	.8593	.7511	.6118
11	.9755	.9755	.9044	.9442	.8931	.7940	.8889	.7931	.6681
12	.9921	.9637	.9637	.9557	.9154	.8498	.9119	.8335	.7183
13	.9873	.9873	.9489	.9678	.9318	.8799	.9285	.8865	.7687
14	.9959	.9809	.9809	.9750	.9501	.9017	.9427	.8905	.8118
15	.9934	.9934	.9726	.9802	.9610	.9272	.9542	.9110	.8432

bilities. The probability $P(CS_1|(0,\dots,0,s,p,\dots,p,\delta^*p))$ of k cell multinomial distribution is obviously equal to $P(CS_1|(s,p,\dots,p,\delta^*p))$ of r cell multinomial distribution for fixed n. Thus, we only have to consider the cases where r is less than k, when the problem is already solved for all smaller values of k for the same n and same δ^*. Hence, we only consider vectors of the type $(s,p,\dots,p,\delta^*p)$. For s and δ^* fixed, we have $p = (1-s)/(k-2+\delta^*)$. When we let s run from 0 to $1/(k-1+\delta^*)$ with equal increment of 10^{-3}, we can detect the $\boldsymbol{P}$ that minimizes $P(CS_1|\boldsymbol{P} \in PZ)$. Numerical results showed that the minimum always took place at one end of the interval in question, i.e., for $s = 0$ or for $s = p$.

In Table 2, we provide the values of the probability of a correct selection $P(CS_2|IZ)$ under the worst configuration for $k = 2,3,4$; $c = 1,2,3$ and $d = 2,4$. We use the technique in analogy with the one we described above to find the WC. In the present case, we let s run from 0 to $1/k$ with equal increment of 10^{-3}. Numerical results showed that the minimum of $P(CS_2|IZ)$ no longer took place at either end of the interval. We use $*$ to denote those cases in which the minimum occurs in the interior of the

Table 2. $P(CS_2|WC)$: The minimum probability of a correct selection in the indifference zone and $E(S|WC)$: The expected subset size under WC.

$d = 2$

	$k=2$			$k=3$			$k=4$		
n \ c	1	2	3	1	2	3	1	2	3
6	.6563 (1.313)	.8906 (1.781)	.8906 (1.781)	.6563 (2.317)	.8134 (2.546)	.8906 (2.811)	.6748 (3.139)	.7737 (3.622)	.8396 (3.886)
7	.7734 (1.547)	.7734 (1.547)	.9375 (1.875)	.6214 (2.056)	.7734 (2.633)	.8327 (2.748)	.6214 (3.077)	.7819 (3.538)	.8242 (3.738)
8	.6367 (1.273)	.8555 (1.711)	.8555 (1.711)	.6000 (2.056)	.7452 (2.492)	.8476 (2.799)	.6001 (2.762)	.7452 (3.465)	.8008 (3.649)
9	.7461 (1.492)	.7461 (1.492)	.9102 (1.820)	.6911 (2.073)	.7423 (2.381)	.8277 (2.637)	.5599 (2.701)	.7012 (3.312)	.7714 (3.628)
10	.6231 (1.246)	.8281 (1.656)	.8281 (1.656)	.5759 (1.888)	.7537 (2.421)	.7820 (2.592)	.5759 (2.725)	.6820 (3.232)	.7516 (3.530)
11	.7256 (1.451)	.7256 (1.451)	.8867 (1.773)	.5628 (1.923)	.6932 (2.314)	.8013 (2.638)	.5628 (2.648)	.6902 (3.190)	.7399 (3.442)
12	.6128 (1.226)	.8062 (1.612)	.8062 (1.612)	.6128 (1.918)	.6945 (2.240)	.7789 (2.493)	.5451 (2.448)	.6840 (3.103)	.7317 (3.373)
13	.7095 (1.419)	.7095 (1.419)	.8666 (1.733)	.5464 (1.775)	.7095 (2.266)	.7439 (2.464)	.5143 (2.424)	.6840 (2.987)	.7132 (3.337)
14	.6049 (1.209)	.7880 (1.576)	.7880 (1.576)	.5357 (1.823)	.6568 (2.181)	.7648 (2.505)	.5278 (2.444)	.6250 (2.945)	.6952 (3.258)
15	.6964 (1.393)	.6964 (1.393)	.8491 (1.698)	.6032 (1.810)	.6597 (2.130)	.7420 (2.377)	.5349 (2.386)	.6313 (2.921)	.6865 (3.202)

Table 2. (continued).

$d = 4$

	$k=3$			$k=4$		
n \ c	1	2	3	1	2	3
6	.6563 (2.399)	.8745* (2.646)	.8906 (2.893)	.7997 (3.197)	.8820 (3.681)	.9643 (3.944)
7	.7174 (2.152)	.7734 (2.728)	.9277* (2.844)	.7174 (3.230)	.9095 (3.692)	.9479 (3.892)
8	.6367 (2.088)	.8332* (2.575)	.8555 (2.882)	.6854 (2.980)	.8476 (3.703)	.9500 (3.888)
9	.7137* (2.208)	.7461 (2.515)	.8953* (2.797)	.7231 (2.911)	.8255 (3.569)	.9194 (3.892)
10	.6231 (2.016)	.8048* (2.570)	.8281 (2.741)	.6719 (2.938)	.8434* (3.508)	.9066 (3.825)
11	.6411 (1.970)	.7256 (2.940)	.8709* (2.775)	.6411 (2.907)	.7976 (3.502)	.9056 (3.783)
12	.6128 (2.075)	.7779 (2.396)	.8062 (2.694)	.6695 (2.758)	.7779 (3.472)	.8773 (3.767)
13	.6369 (1.911)	.7095 (2.440)	.8463* (2.639)	.6369 (2.712)	.7938* (3.364)	.8667 (3.742)
14	.6047 (1.876)	.7572 (2.324)	.7880 (2.671)	.6078 (2.707)	.7572 (3.312)	.8652 (3.670)
15	.6325 (1.973)	.6964 (2.293)	.8274* (2.596)	.6259* (2.677)	.7392 (3.295)	.8404 (3.631)

interval $[0, 1/k]$. In the same table, we also provide the values of $E(S)$ under WC. $E(S)$ is generally considered as a criterion of the efficiency of a selection procedure which satisfies a specific probability requirement $P(CS_2|IZ) \geq P^*$.

Remark 3.1. In Table 2, $P(CS_2|WC)$ and $E(S|WC)$ values for $k = 2$ are identical for all the D values, since the d value in the proposed procedure R does not play a role in this case for any number of observations.

Remark 3.2. An analogous procedure for selecting the least probable cell will not be similar in nature. The ratios cannot be even used to measure the cell probabilities as it was pointed out in Alam and Thompson (1972) where the authors considered a selection procedure for the least probable cell in PZ. An appropriate measure under both PZ and IZ and its properties for selecting the least probable cell under our integrated formulation is worthy of further investigation.

REFERENCES

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

Bechhofer, R. E., Elmaghraby, S. A. and Morse, N. (1959). A single sample multiple-decision procedure for selecting the multinomial event which has the largest probability, *Ann. Math. Statist.*, **30**, 102–119.

Chen, P. and Sobel, M. (1987*a*). An integrated formulation for selecting the t best of k normal populations, *Comm. Statist. A—Theory Methods*, **16**(1), 121–146.

Chen, P. and Sobel, M. (1987*b*). A new formulation for the multinomial selection problem, *Comm. Statist. A—Theory Methods*, **16**(1), 147–180.

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

Kesten, H. and Morse, N. (1959). A property of the multinomial distribution, *Ann. Math. Statist.*, **30**, 120–127.