On Selection and Ranking Procedures and Order Statistics from the Multinomial Distribution*

by

Shanti S. Gupta and Klaus Nagel

Purdue University

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series Number 77

June 1966

*This research was supported in part by Contract AF 33(657)11737 with the Aerospace Research Laboratories. Reproduction in whole or in part permitted for any purposes of the United States Government.

On Selection and Ranking Procedures and Order Statistics from the Multinomial Distribution*

by

Shanti S. Gupta and Klaus Nagel

Purdue University

1. Introduction and Summary

The purpose of the present paper is to consider selection problems for the multinomial distribution. In some situations the experimenter may be interested in selecting a subset containing the cell with the highest probability based on a set of observations. The main problem is to define a selection rule which selects a small, non-empty subset such that the probability of including the cell with the largest cell-probability is at least equal to a preassigned number P^*.

Let $p_1, p_2, \ldots, p_k$ be the unknown cell-probabilities in the multinomial distribution with $\sum_1^k p_i = 1$. Let $x_1, x_2, \ldots, x_k$ be the respective observations in the k cells of the distribution with $\sum_1^k x_i = N$. Let the ordered cell-probabilities be given by

$$p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[k]} . \tag{1.1}$$

The pairing of the ordered $p_{[i]}$ and the ordered or unordered x_i is not known. The goal of the experimenter is to select a subset containing the cell corresponding to $p_{[k]}$. A correct selection $\{CS\}$ is defined as the selection

*This research was supported in part by Contract AF 33(657)11737 with the Aerospace Research Laboratories. Reproduction in whole or in part permitted for any purposes of the United States Government.

of any subset of the k cells which contains the cell with the largest probability. In the case of a tie one of the cells with the largest value is considered ''tagged'' and the selection is correct if this ''tagged'' population is in the selected subset. This restriction becomes necessary and meaningful for considering the infimum of the probability of a correct selection where the worst configuration is the limit of configurations with $p_{[k]} > p_{[k-1]}$.

It may be pointed out that the above subset selection formulation for the multinomial distribution is different from the ''indifference zone'' selection of the best cell given by Bechhofer, Elmagharaby and Morse (1959) and Kesten and Morse (1959). In the latter formulation due to Bechhofer, an ''indifference zone'' in the parameter space is specified, the number of observations needed is tabulated and the final decision is the selection of a single population which is asserted to be the best population. The present formulation follows along the lines discussed, for example, for other discrete distributions by Gupta and Sobel (1960), Gupta (1966), Nagel (1966).

In the subset selection problem, the size or the number of cells in the selected subset is a random variable which takes values 1 to k inclusive. One meaningful criterion of the efficiency of such a selection procedure is the smallness of the expected value of the size of the selected subset. Thus the evaluations of the expected size or the expected proportion in the selected subset become relevant and are given in this paper.

The solution for the problem of selecting a subset containing the cell with the smallest probability cannot be obtained from the previous solution for the largest case. Therefore, similar investigations have been carried out for the smallest cell-probability selection. The procedures proposed in this paper depend on the largest or the smallest order statistic in a multinomial sample and hence the first two moments of these order statistics have been evaluated and are given in this paper.

Section 2 of the paper describes the selection rule for the maximum cell-probability case and discusses the infimum of the probability of a correct selection. It is shown that the infimum takes place for configurations of the type $(0,0,\ldots,0,q,p,\ldots,p)$ where $q \leq p$ and the number of zeros in the above configuration is not known. Formulae are derived for the expected proportion in the selected subset. Section 3 deals with the selection problem for the minimum cell-probability case. A selection rule is proposed and its efficiency is studied by deriving appropriate expressions for the expected proportion. The infimum of the probability of a correct selection is studied and from the numerical evaluations it appears that the minimum takes place when all cell probabilities are equal. Section 4 of this paper discusses the order statistics of the multinomial distributions and formulae for the first and second moments of largest and the smallest are given.

Table 1A gives the expected proportion, the probability of a correct selection, the probability of selecting any of the cells with smaller probabilities when the cell probabilities for the configuration $p, p,\ldots,pA$, where $A \geq 1$. From this table for $A = 1$ one can obtain the infimum of the probability of correct selection and the configuration for which it is obtained. Table 1B gives the minimum D to satisfy the condition that the probability of a correct selection $\geq P^*$ (a given number). Tables 2A and 2B have same entries as Table 1A and 1B except that theydeal with the minimum probability case. Table 3 gives the expected value and the variance of the largest and smallest order statistic in a multinomial distribution with cell-probabilities $p,p,\ldots,pA$ where $A \geq 1$. Table 4 has the same values as Table 3 for the configuration $p/A,p,\ldots,p$ where $A \geq 1$.

2. Selecting the Subset Containing the Cell with the Largest Probability

Let $x_i (i = 1,2,\ldots,k)$ be the observed numbers in the ith cell and let $\Sigma x_i = N$. Then the rule R for selecting the subset to contain the cell with the largest probability is as follows.

R: Select the cell with observed x_i iff

$$x_i \geq x_{max} - D \tag{2.1}$$

where $x_{max} = \max(x_1, x_2, \ldots, x_k)$ and D is a given non-negative integer.

It is clear that the above rule selects a non-empty subset of random size. For $D \geq N$, the rule selects all the cells.

Using the rule R, the probability of a correct selection is given by

$$P\{CS|R\} = F(k,N,D;p_{[1]},\ldots,p_{[k]}) = \sum_{\substack{\Sigma\nu_i = N \\ \nu_i \leq \nu_k + D \\ i=1,2,\ldots,k}} \frac{N!}{\nu_1! \cdots \nu_k!} p_{[1]}^{\nu_1} \cdots p_{[k]}^{\nu_k} \tag{2.2}$$

To find out for which vector $(p_{[1]},\ldots,p_{[k]})$, the expression (2.2) attains its minimum, we use a similar method as Kesten and Morse (1959).

In (2.2), we put $p_{[i]} + p_{[j]} = q$, $i < j$, and try to minimize the right hand side as a function of $p_{[j]}$. We rewrite (2.2) as

$$P\{CS|R\} = \sum_{\substack{\Sigma\nu_t = N \\ \nu_{max} \leq \nu_k + D}} \frac{N!}{\prod_{t=1}^{k} \nu_t!} p_{[j]}^{\nu_j} (q - p_{[j]})^{\nu_i} \prod_{\ell \neq i,j} p_{[\ell]}^{\nu_\ell} . \tag{2.3}$$

Now putting $\nu_i + \nu_j = m$ and summing over m, (2.3) becomes

$$(2.4) \quad P\{CS|R\} = \sum_{m=0}^{N} \binom{N}{m} \sum_{\substack{\Sigma\nu_\ell = N-m \\ \ell \neq i,j}} \frac{(N-m)!}{\prod_{\ell \neq i,j} \nu_\ell!} \prod_{\ell \neq i,j} p_{[\ell]}^{\nu_\ell} \cdot$$

$$\cdot \left[\sum_{\substack{\nu_j \le \nu_k + D \\ m-\nu_j \le \nu_k + D}} \binom{m}{\nu_j} p_{[j]}^{\nu_j} (q - p_{[j]})^{m-\nu_j} \right]$$

or

$$(2.5) \quad P\{CS|R\} = \sum_{m=\nu}^{N} \binom{N}{m} \sum_{\substack{\Sigma\nu_\ell = N-m \\ \ell \neq i,j}} \frac{(N-m)!}{\prod_{\ell \neq i,j} \nu_\ell!} \prod_{\ell \neq i,j} p_{[\ell]}^{\nu_\ell} q^m \cdot$$

$$\cdot \left[\sum_{\substack{\nu_j \le \nu_k + D \\ m-\nu_j \le \nu_k + D}} \binom{m}{\nu_j} r^{\nu_j} (1-r)^{m-\nu_j} \right], \quad \text{where } r = \frac{p_{[j]}}{q} \ge \frac{1}{2}$$

Now there are two different cases:

Case 1. $j \neq k$. Then the summation inside the square brackets goes from $m - (\nu_k + D)$ to $\nu_k + D$ provided $m - (\nu_k + D) \le \nu_k + D$, otherwise this sum is zero. Thus the expression inside the square brackets can be written as the difference of two incomplete Beta function as follows:

$$(2.6) \quad \frac{m!}{(\nu_k + D)!(m - \nu_k - D - 1)!} \int_r^1 \left[x^{\nu_k + D} (1-x)^{m-\nu_k-D-1} - x^{m-\nu_k-D-1} (1-x)^{\nu_k + D} \right] dx.$$

Now the integrand in (2.6) is non-negative for $x \geq \frac{1}{2}$ and since $r \geq \frac{1}{2}$, (2.6) is a decreasing function of r. Since the right hand side of (2.5) is a linear combination of decreasing functions of r with non-negative coefficients, the probability of a correct selection is a decreasing function of r. This means if we keep the sum $p_{[i]} + p_{[j]}$ fixed, the probability of a correct selection is a decreasing function of $p_{[j]}$. Hence we have the following lemma.

<u>Lemma 1</u>. Keeping the sum $p_{[i]} + p_{[j]}$ $(1 \leq i < j < k)$, constant, the probability of a correct selection as given by (2.5) decreases as we pass from the configuration $(p_{[1]}, \ldots, p_{[i]}, \ldots, p_{[j]}, \ldots, p_{[k]})$ to $(p_{[1]}, \ldots, p_{[i]} - \epsilon, \ldots, p_{[j]} + \epsilon, \ldots, p_{[k]})$ where $0 < \epsilon \leq p_{[i]}$.

Remark: It may be pointed out that the result is true even if the order is disturbed in the new configuration.

Case 2: $j = k$. In this case the summation inside the square brackets in (2.5) extends from $[\frac{m-D+1}{2}]$ to m and this can be written as

$$(2.7) \qquad 1 - (m-\alpha+1)\binom{m}{\alpha-1} \int_r^1 t^{\alpha-1}(1-t)^{m-\alpha} dt, \quad \alpha = [\frac{m-D+1}{2}]$$

which shows that it is an increasing function of r or $p_{[k]}$. Hence using the same argument as in Lemma 1, we have the following lemma.

<u>Lemma 2</u>. Keeping the sum $p_{[i]} + p_{[k]}$, $1 \leq i < k$, constant, the probability of a correct selection as given by (2.5) decreases as we pass from the configuration $(p_{[1]}, \ldots, p_{[i]}, \ldots, p_{[k]})$ to $(p_{[1]}, \ldots, p_{[i]} + \epsilon, \ldots, p_{[k]} - \epsilon)$ where $0 < \epsilon \leq p_{[k]}$. [The remark given at the end of Lemma 1 is also true for Lemma 2.]

The overall minimum of the probability of a correct selection has to be at a configuration which cannot be changed to one with a smaller probability by using the procedures of Lemma 1 or Lemma 2.

Hence, we have the following theorem:

<u>Theorem 1</u>: Let $p_{[i]}$ be the ith ordered cell-probability. Denote by μ the smallest integer such that $p_{[\mu]} > 0$ and let ν be the largest integer such that $p_{[\nu]} < p_{[k]}$. For a configuration minimizing the probability of a correct selection which is given by the function $F(k,N,D;(p_{[1]},\dots,p_{[k]}))$ defined in (2.2), the following relations must hold:

$$\text{(A)} \qquad \mu \geq \nu .$$

Furthermore, if $\mu = k-1$ then

$$\text{(B)} \qquad \mu > \nu .$$

Proof:

(A) Assume the minimum of (2.2) is obtained for a configuration with $\mu < \nu$. Then by using Lemma 1 with $i = \mu$, $j = \nu$ a worse configuration can be constructed, a contradiction.

(B) Assume for a worse configuration $\mu = \nu = k-1$. Then by Lemma 2 with $i = k-1$ a worse configuration can be constructed which again leads to a contradiction.

According to this theorem the worst configuration is of the type

$$(2.8) \qquad (0,\dots,0,s,p,\dots,p), \quad s \leq p .$$

Let r be the number of positive $p_{[i]}$'s. The overall minimum then can be found as

$$(2.9) \qquad \min_{\underline{p}} F(k,n,D;\underline{p}) = \min_{r=2,\dots,k} \Big(\min_{\frac{1}{r} \leq p \leq \frac{1}{r-1}} F(k,n,D;(0,\dots,0,s,p,\dots,p))$$

where $s = 1-(r-1)p$.

For configuration of type (2.8), (2.2) can be rewritten as

$$(2.10)\quad F(k,N,d;(0,\dots,0,s,p,\dots,p)) = \sum_{\substack{\Sigma\nu_\ell=N \\ \nu_\ell\le \nu_k+0 \\ \ell=k-r+1,\dots,k}} \frac{N!}{\nu_{k-r+1}!\cdots\nu_k!}\, s^{\nu_{k-r+1}}\, p^{N-\nu_{k-r+1}}$$

which obviously is equal to

$$F(r,N,D;\ (s,p,\dots,p)).$$

Hence it is not necessary 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 D. Hence we have to consider only vectors of the type

$$(s,p,\dots,p), \qquad s = 1-(k-1)p$$

for which (2.2) becomes

$$(2.11)\quad F(k,N,D;(s,p,\dots,p)) = \sum_{\substack{\Sigma\nu_i=N \\ \nu_i\le \nu_k+D \\ i=1,2,\dots,k}} \frac{N!}{\nu_1!\cdots\nu_k!}\, s^{\nu_1}\, p^{N-\nu_1} =$$

$$= \sum_{\nu_1=0}^{N}\binom{N}{\nu_1}\sum_{\nu_k\ge\nu_1+D}\binom{N-\nu_1}{\nu_k}\left\{\sum_{\substack{\sum_{i=2}^{k-1}\nu_i=N-\nu_1-\nu_k \\ \nu_i\le\nu_k+D \\ i=2,\dots,k-1}} \frac{(N-\nu_1-\nu_k)!}{\nu_2!\cdots\nu_{k-1}!}\right\} s^{\nu_1}p^{N-\nu_1}$$

where $\frac{1}{k}\le p\le\frac{1}{k-1}$.

It should be noted that expression in (2.11) is connected with the coefficients $A_r(m,t)$ given in Gupta (1963). (2.11) is a polynomial in p, the minimum of which can be found by differentiation. This was done numerically and it turned out that the minimum usually took place at one end of the interval in question, i.e. for $p = \frac{1}{k}$ or for $p = \frac{1}{k-1}$. For the parameter values $D = 0(1)4$, $k=2(1)10$ and $N = 2(1)15$, it happened only once that the minimum was attained in the interior, namely, for the case $k=3$, $N=6$ and $D=4$. In all the other cases the worst configuration for k,N,D can be found in the tables giving the probabilities of a correct selection in slippage configuration with $A = 1$, in the following way: Look for the smallest value of $F(r,N,D;(\frac{1}{r},\ldots,\frac{1}{r}))$ for all $r \leq k$. Let r^* be the corresponding r. Then the worst configuration with one exception, is found to be $(0,\ldots,0,\frac{1}{r^*},\ldots,\frac{1}{r^*})$. Hence we are able to evaluate the minimum D value which, for fixed N,k and P^*, guarantees that the probability of a correct selection is $\geq P^*$. This is done by following the procedure given above and by consulting the tables of the probability of a correct selection (Table 1A) for the equal probability case. These minimum values of D are given in Table 1B.

Expected Size (Proportion) in the Selected Subset

For the procedure R, the size S of the selected subset is a random variable which can take on only integer values from 1 to k inclusive. For any fixed values of N, k and P^* the expected size of the selected subset is a function of the true configuration $\underline{p} = (p_1,p_2,\ldots,p_k)$ and this function in analogy with the power function of tests of hypotheses can be regarded as a criterion of the efficiency of any procedure which satisfies the same probability requirement $\inf P\{CS\} \geq P^*$. It is easy to see that the expected size $= k$ (expected proportion) is given by

$$(2.12) \quad E(S) = \sum_{i=1}^{k} P(\text{cell } i \text{ is included in the subset})$$

$$= \sum_{\Sigma\nu_i=N} \frac{N!}{\nu_1!\cdots\nu_k!} p_1^{\nu_1} \cdots p_k^{\nu_k} B_{\underline{\nu}}$$

where $B_{\underline{\nu}}$ = number of ν_i's $\geq \nu_{max} - D$. For any D whether it satisfies the P^* condition or not, the above function can be computed as follows. Count the number $B_{\underline{\nu}}$ by considering all possible outcomes $\nu_1, \nu_2, \ldots, \nu_k$, $\Sigma\nu_i = N$ i.e. by considering all partitions of N which satisfy $\nu_i \geq \nu_{max} - D$, one can evaluate the expected proportion in the selected subset. For the configuration with $\underline{p} = (p, p, \ldots, pA)$ $A \geq 1$, (2.12) simplifies as follows.

$$(2.13) \qquad E(S) = p^N \sum_{\Sigma\nu_i=N} \frac{N!}{\nu_1!\cdots\nu_k!} A^{\nu_k} B_{\underline{\nu}} .$$

Using (2.13), values of expected proportion were computed for different values of $A \geq 1$ for fixed values of N, k and D and are given in Table 1A.

<u>Probability of a Correct Selection and the Probability of Selecting a Non-Best Cell. in the Slippage Configuration</u>

For the configuration $(p, \ldots, p, Ap)$, we have

$$(2.14) \qquad P\{CS|R\} = p^N \sum_{\substack{\Sigma\nu_i=N \\ \nu_i \leq \nu_k + D \\ i=1,2,\ldots,k}} \frac{N!}{\nu_1!\cdots\nu_k!} A^{\nu_k} .$$

From (2.13) and (2.14) one can obtain the probability of selecting any fixed of the non-best equal-probability cell by the relation

$$E(S) = P\{CS|R\} + (k-1)\, P\{\text{Selecting a fixed non-best cell}|R\} \tag{2.15}$$

Table 1A also gives the values of the probability of a correct selection and the probability of selecting any fixed non-best cell. By consulting these one can determine the minimum N, which for fixed P^* and fixed number k of cells in some slippage configurations will make the probability of a correct selection $\geq P^*$ and keep the expected proportion $\leq \gamma$ where γ is a preassigned number between 0 and 1. In Section 5 we illustrate this numerically.

3. Selecting a Subset Containing the Cell with the Smallest Probability

Using the same notation as in Section 2, the rule T for selecting a subset to contain the cell with the smallest probability is as follows.

T: Select the cell with observed x_i iff

$$x_i \leq x_{min} + C \tag{3.1}$$

where $x_{min} = \min(x_1, x_2, \ldots, x_k)$ and C is a given non-negative integer. For $C \geq N$, the rule T selects all the cells in the selected subset. The probability of a correct selection is given by

$$P\{CS|T\} = G(k,N,C;\, p_{[1]},\ldots,p_{[k]}) \tag{3.2}$$

$$= \sum_{\substack{\Sigma\nu_\ell = N \\ \nu_{min} \geq \nu_1 - C}} \frac{N!}{\nu_1! \cdots \nu_k!}\, p_{[1]}^{\nu_1} \cdots p_{[k]}^{\nu_k}\,.$$

Proceeding as before by putting $p_{[i]} + p_{[j]} = q$, $i < j$, we obtain

$$(3.3)\quad P\{CS|T\} = \sum_{m=0}^{N} \binom{N}{m} \sum_{\Sigma\nu_{\ell}=N-m} \frac{(N-m)!}{\prod_{\ell \neq i,j} \nu_{\ell}!} \prod_{\ell \neq i,j} p_{[\ell]}^{\nu_{\ell}} q^{m} \Bigg[\sum_{\substack{\nu_i \geq \nu_1 - C \\ m-\nu_i \geq \nu_1 - C}} \binom{m}{\nu_i} r^{\nu_i} (1-r)^{\nu_i} \Bigg]$$

where $r = p_{[i]}/q \leq 1/2$.

Again, there are two different cases

Case 1: $i \neq 1$. By slight modifications of the arguments in the corresponding case in Section 2, we prove the lemma.

Lemma 3. Keeping the sum $p_{[i]} + p_{[j]}$, $1 < i < j \leq k$, constant, the probability of a correct selection in using the rule T as given by (3.3) decreases as we pass from the configuration $(p_{[1]},\ldots,p_{[i]},\ldots,p_{[j]},\ldots,p_{[k]})$ to $(p_{[1]},\ldots,p_{[i]}-\epsilon,\ldots,p_{[j]}+\epsilon,\ldots,p_{[k]})$ where $0 < \epsilon \leq p_{[i]}$.

Remark: Lemma 3 is true even if the order is distubed in the configuration.

Case 2: $i = 1$. Slight modification of the arguments in Section 2, corresponding to the case $j = k$, leads to the following lemma.

Lemma 4: Keeping the sum $p_{[1]} + p_{[j]}$, $1 < j \leq k$, constant, the probability of a correct selection in using the rule T as given by (3.3) decreases as we pass from the configuration $(p_{[1]},\ldots,p_{[j]},\ldots,p_{[k]})$ to $(p_{[1]}+\epsilon,\ldots,p_{[j]}-\epsilon,\ldots,p_{[k]})$ where $0 < \epsilon \leq p_{[j]}$.

The same remark as at the end of Lemma 3 also holds for Lemma 4.

In obtaining the overall minimum of the probability of a correct selection, we look for the configurations for which this probability cannot be decreased by using the procedures of Lemma 3 or 4. Hence we have the theorem.

Theorem 2.

In using the procedure T, the probability of a correct selection given by $G(k,N,C;p_{[1]},\ldots,p_{[k]})$ in (3.2), is minimized at a configuration $(p_{[1]},\ldots,p_{[k]})$

given by

$$(3.4) \qquad p_{[\ell]} = \begin{cases} p, & \ell=1,\dots,k-1 \\ 1-(k-1)p, & \ell=k \end{cases} .$$

Proof: Assume the minimum takes place for a configuration which is not of the type defined in (3.4). Provided $k > 2$ there exists a smallest positive integer $\mu < k$ such that $p_{[\mu]} > p_{[1]}$. Applying Lemma 3 with $i = \mu$ and $j = k$, leads to a configuration with smaller probability, a contradiction.

For $k = 2$ any configuration is of the type described in (3.4), and, furthermore, by Lemma 4 the worst configuration is: (1/2, 1/2). This completes the proof of the theorem.

As in the previous section the problem is now reduced to minimization of a polynomial in one variable. On substituting $(p,\dots,p,q)$, $q = 1-(k-1)p$, in (3.3) gives

$$(3.5) \qquad G(k,N,C;p,\dots,p,q) = \sum_{\substack{\Sigma\nu_\ell = N \\ \nu_\ell \geq \nu_{min}-C}} \frac{N!}{\nu_1!\cdots\nu_k!} p^{N-\nu_k} q^{\nu_k} ,$$

Numerical evaluation of (3.5) for $0 < p \leq 1/k$, $k = 2(1)10$, $N = 2(1)15$ and $C = 0(1)4$ showed that the overall minimum actually takes place for $p = q = \frac{1}{k}$ i.e. the worst configuration is $(\frac{1}{k}, \frac{1}{k},\dots,\frac{1}{k})$.

Expressions for the expected size of the selected subset in using T can be obtained in the same manner as those for R. For the configurations $(p/A,p,\dots,p)$ with $A \geq 1$, which include the slippage configurations, tables were computed for the expected proportion, the probability of a correct selection and the probability of selecting any fixed cell with probability p. These values are given in Table 2A. Using Table 2A for the case $A = 1$, we have computed the minimum value of C such that the probability that the rule T selects the cell with the smallest probability is at least equal to a preassigned number P^*. These minimum values of C are given in Table 2B.

4. Order Statistics from the Multinomial Distribution

Both procedures R and T are closely related to the largest and the smallest order statistics from the multinomial distribution. In considering the correct selection, the random variables of interest are $X_{max} - X$ or $X_{min} - X$ where X corresponds to the cell with the largest or smallest probability and X_{max} and X_{min} are the largest or the smallest order statistics from the remaining $(k-1)$ random variables. Hence, it is of interest to obtain the moments of X_{max} and X_{min}. The moments of these order statistics can be written as follows

$$(4.1) \qquad E(X^j_{max}) = \sum_{\Sigma\nu_\ell = N} \frac{N!}{\nu_1! \cdots \nu_k!} \, p_{[1]}^{\nu_1} \cdots p_{[k]}^{\nu_k} \, \nu^j_{max}$$

where $\nu_{max} = \max(\nu_1, \nu_2, \ldots, \nu_k)$. By replacing X_{max} and ν_{max} by X_{min} and ν_{min}, one obtains the formulae for the moments of the smallest order statistic. If we use the configuration $(p,p,\ldots,p,pA)$ with $A \geq 1$, (4.1), reduces to

$$(4.2) \qquad E(X^j_{max}) = \sum_{\Sigma\nu_\ell = N} \frac{N!}{\nu_1! \cdots \nu_k!} \, p^N A^{\nu_k} \, \nu^j_{max} \; .$$

Using (4.2), we have evaluated the expected value and the variance of the largest order statistic for the configuration $(p,p,\ldots,p,pA)$, $A \geq 1$. We have also considered the configuration $(p/A,p,\ldots,p)$, $A \geq 1$ and evaluated the corresponding mean and variance for X_{max} by using an expression similar to (4.2). These moments are given in Table 3A and 4A, respectively. Tables 3B and 4B, respectively, give the corresponding values for X_{min} for the same two configurations by using the appropriate formulae. It should be pointed out that the worst configuration for rule T is of the type $(p,p,\ldots,pA)$, $A \geq 1$. Again the worst configuration for R reduces to one of the type $(p/A,p,\ldots,p)$, $A \geq 1$. This is one reason why we chose these configurations.

5. Examples to Illustrate the Use of Tables 1A and 2A

In some problems the experimenter may wish to design an experiment i.e., he wishes to determine the minimum N (this is equivalent linear cost per observation model) such that for configuration $(p,\ldots,p,pA)$ with $A \geq A_0$, the probability of a correct selection $\geq P^*$ and the expected proportion in the selected subset $\leq \gamma$, $0 < \gamma < 1$ where P^* and γ are preassigned and when the rule R is used. Table 1A can be used to solve this problem. As an example, let $P^* = .95$, $\gamma = .4$, $k = 3$ and $A_0 = 3.0$ then from Table 1A, we find that the pairs (N,D) that satisfy the above conditions are (14,0) and (12,1). Hence the minimum value of N is 12. Similarly Table 2B can be used to design experiments when rule T is used.

References

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

2. Gupta, S.S. and Sobel, M. (1960). Selecting a subset containing the best of several binomial populations. Contributions to Probability and Statistics Ch. XX, Stanford University Press.

3. Gupta, S.S. (1963). On a selection and ranking procedure for gamma populations. Ann. Inst. Stat. Math. 14, 199-216.

4. Gupta, S.S. (1966). Selection and ranking procedures and order statistics for the binomial distribution. Proceedings of the International Symposium on Classical and Contagious Discrete Distributions, held in Montreal, August 1963. Statistical Publishing Society, Calcutta and Pergamon Press.

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

6. Nagel, K. (1966). On selecting a subset containing the best of several discrete distributions. Dept. of Statistics, Purdue University, Mimeo Series Number 66. Submitted for publication.

Table 1B

$P^* = .75$

For given values of k,N and P*, the following table gives the minimum D such that $\inf_{\underline{p}} P\{CS|R\} \geq P^*$.

k \ N	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	0	1	2	1	2	1	2	3	2	3	2	3	2	3
3,...,10	1	2	2	2	2	2	3	3	3	3	3	3	3	3

$P^* = .90$

k \ N	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	2	3	2	3	4	3	4	3	4	5	4	5	4	5
3,...,10	2	3	3	3	4	4	4	4	4	5	5	5	5	5

Table 2B

$P^* = .75$

For given values of k,N and P^*, the following table gives the minimum C such that $\inf_{\underline{p}} P\{CS|T\} \geq P^*$.

N \ k	2	3	4	5	6	7	8	9	10
2	0	1	1	1	1	1	0	0	0
3	1	2	1	1	1	1	1	1	1
4	2	2	2	1	1	1	1	1	1
5	1	2	2	2	1	1	1	1	1
6	2	2	2	2	2	1	1	1	1
7	1	2	2	2	2	2	1	1	1
8	2	2	2	2	2	2	2	1	1
9	3	2	2	2	2	2	2	2	1
10	2	3	2	2	2	2	2	2	2
11	3	3	3	2	2	2	2	2	2
12	2	3	3	3	2	2	2	2	2
13	3	3	3	3	3	2	2	2	2
14	2	3	3	3	3	3	2	2	2
15	3	3	3	3	3	3	2	2	2

$P^* = .90$

N \ k	2	3	4	5	6	7	8	9	10
2	2	2	1	1	1	1	1	1	1
3	2	3	2	2	1	1	1	1	1
4	2	3	2	2	2	2	1	1	1
5	2	3	2	2	2	2	2	1	1
6	4	3	3	2	2	2	2	2	2
7	3	3	3	3	2	2	2	2	2
8	4	4	3	3	3	2	2	2	2
9	3	4	3	3	3	3	2	2	2
10	4	4	4	3	3	3	3	2	2
11	5	4	4	4	3	3	3	3	2
12	4	4	4	4	3	3	3	3	3
13	5	5	4	4	4	3	3	3	3
14	4	5	4	4	4	4	3	3	3
15	5	5	5	4	4	4	3	3	3