A PROPERTY OF THE MULTINOMIAL DISTRIBUTION

By Harry Kesten[1] and Norman Morse[2]

Cornell University

1. Introduction. The purpose of this paper is to prove a property of the multinomial distribution which is fundamental to the choice of sample size for the selection procedure described in the preceding paper [1] of this issue. As in [1], we let $p_{[1]} \leqq p_{[2]} \leqq \cdots \leqq p_{[k]}$ denote the ranked multinomial probabilities and let $\phi_k = \phi(p_{[1]}, \cdots, p_{[k]})$ be the probability of a correct selection when the selection procedure in section 4 of [1] is used. We wish to prove that for any integers $N \geqq 1$, $k \geqq 2$ and for any number $1 < \theta^* < \infty$, ϕ_k is minimized among all configurations with $p_{[k]} \geqq \theta^* p_{[i]}$ $(i = 1, \cdots, k - 1)$ by the configuration $p_{[1]} = p_{[2]} = \cdots = p_{[k-1]} = p_{[k]}/\theta^* = 1/(\theta^* + k - 1)$. This configuration is called "least favorable" because of this property.

The theorem on least favorable configurations, proved below in section 3, merely assembles the results of the preceding lemmas in a rather obvious way. The main ideas of the result are contained in the two lemmas of section 2. Note that Lemma 1, proved below for $k \geqq 3$, is not needed to prove the theorem for the case $k = 2$.

We should like to thank Dr. Milton Sobel for his valuable comments and suggestions.

2. The lemmas. It was found convenient to deviate in the following respect from the notation used in [1]. Let p_k denote the *largest* of the k probabilities, but let the other $(k - 1)$ probabilities $p_1, \cdots, p_{k-1}$ be unranked among themselves. Let E_i be the event associated with p_i $(i = 1, \cdots, k)$, and let y_{iN} be the number of occurrences of event E_i after N observations $(i = 1, \cdots, k)$; of course we have $0 \leqq y_{iN} \leqq N$ $(i = 1, \cdots, k)$ and $\sum_{i=1}^{k} y_{iN} = N$. (For notational convenience we use the same symbol for a chance variable and its observed value.) In correspondence to the notation just given, the probability ϕ_k of correctly choosing event E_k is in this paper a function of the p_i's as defined above rather than a function of the $p_{[i]}$'s as it is in [1]. It is sufficient to restrict our attention to configurations with $p_k \geqq \theta^* \max(p_1, \cdots, p_{k-1})$.

Suppose $k \geqq 3$ and any $k - 2$ of the p_i's *including* p_k, are held fixed. For notational convenience, we take the two unfixed probabilities to be p_1 and p_2, and we arbitrarily call p_1 the one which is equal to or greater than the other, so that we have $p_1 \geqq p_2$. Note that since the sum $(p_1 + p_2)$ is fixed, ϕ_k can be

Received February 5, 1958.

[1] Research supported by the Office of Naval Research.

[2] Research supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 18(600)-331. Reproduction in whole or in part is permitted for any purpose of the United States Government.

regarded as a function of p_1 only, say $\psi_1(p_1)$, with $(p_1 + p_2)/2 \leqq p_1 \leqq (p_1 + p_2)$.

LEMMA 1. *Let* $k \geqq 3$, *and let* $(p_1 + p_2), p_3, \cdots, p_k$ *be fixed as above. Then* $\psi_1(p_1)$ *is a non-increasing function of* p_1, *where* p_1 *lies in the interval* $(p_1 + p_2)/2 \leqq p_1 \leqq (p_1 + p_2)$. [Hence ϕ_k is minimized over all vectors with fixed $p_3, \cdots, p_k$ and such that $p_k \geqq \theta^* \max(p_1, \cdots, p_{k-1})$ by taking $p_1 = \min(p_1 + p_2, p_k/\theta^*)$.]

PROOF OF LEMMA 1. ϕ_k can be rewritten

$$\phi_k = \sum \frac{1}{s} \frac{N!}{w_N!\, y_{3N}!, \cdots, y_{kN}!} q^{w_N} p_3^{y_{3N}}, \cdots, p_k^{y_{kN}} \cdot \sum_{y_{1N}=w_N-y_{kN}}^{y_{kN}} \binom{w_N}{y_{1N}} r^{y_{1N}}(1-r)^{y_{2N}} S(y_{1N}, y_{2N}, y_{kN}) \tag{1}$$

where $w_N = y_{1N} + y_{2N}$, $q = (p_1 + p_2)$, $r = p_1/q$; the outer summation is over all $(k-1)$-vectors $(w_N, y_{3N}, \cdots, y_{kN})$ such that $w_N + \sum_{i=3}^{k} y_{iN} = N$ and $y_{kN} \geqq y_{iN}$ $(i = 3, \cdots, k-1)$; $(s-1)$ is the number of y_{iN}'s $(i = 3, \cdots, k-1)$ which equal y_{kN}; and $S(y_{1N}, y_{2N}, y_{kN})$ is defined as follows:

$$S(y_{1N}, y_{2N}, y_{kN}) \begin{aligned} &= 1 &&\text{if } \max(y_{1N}, y_{2N}) < y_{kN} \\ &= s/(s+1) &&\text{if } \max(y_{1N}, y_{2N}) = y_{kN} \text{ and } y_{1N} \neq y_{2N} \\ &= s/(s+2) &&\text{if } y_{1N} = y_{2N} = y_{kN} \end{aligned} \tag{2}$$

The inner summation in (1) insures that $y_{kN} \geqq y_{iN}$ for $i = 1, 2$. Note that s is actually a function of $(y_{3N}, \cdots, y_{kN})$.

To prove Lemma 1 it is sufficient to prove the monotonicity property for the inside sum in (1) with w_N and y_{kN} fixed, for then ϕ_k will be termwise monotonic. Using (2), it is easy to see that for w_N and y_{kN} fixed the inside sum in (1) is equal to either

$$\Pr\{w_N - y_{kN} < y_{1N} < y_{kN}\} + \frac{s}{s+1} \Pr\{y_{1N} = w_N - y_{kN}\} + \frac{s}{s+1} \Pr\{y_{1N} = y_{kN}\} \tag{3}$$

for the case $w_N < 2y_{kN}$, or

$$\frac{s}{s+2} \Pr\{y_{1N} = y_{2N} = y_{kN}\} \tag{4}$$

for the case $w_N = 2y_{kN}$. Since the inner sum is empty for $w_N > 2y_{kN}$, no other cases have to be considered.

Now for $w_N \leqq 2a$, define the function

$$f(a;r) = \sum_{x=w_N-a}^{a} \frac{w_N!}{x!(w_N - x)!} r^x (1-r)^{w_N - x}, \tag{5a}$$

which we rewrite as the difference of incomplete Beta functions:

$$\text{(5b)} \quad f(a; r) = \frac{w_N!}{a!(w_N - a - 1)!} \int_r^1 [x^a(1 - x)^{w_N - a - 1} - x^{w_N - a - 1}(1 - x)^a] \, dx.$$

For the case $w_N > 2a$, we take $f(a; r)$ to be zero. (Also individual terms in (5a) are zero if $x < 0$ or $x > w_N$.) The proof of Lemma 1 can now be completed by noting that expression (3) equals

$$\text{(6)} \qquad \frac{1}{s+1} f(y_{kN} - 1; r) + \frac{s}{s+1} f(y_{kN}; r),$$

and that expression (4) equals

$$\text{(7)} \qquad \frac{s}{s+2} f(y_{kN}; r).$$

Since the integrand in (5b) is non-negative for the case $w_N \leqq 2a + 1$ and $\frac{1}{2} \leqq r \leqq 1$, it follows that $f(a; r)$ is a non-increasing function of r. The expressions (6) and (7) are therefore non-increasing functions of r, and this proves the lemma.

Starting with any configuration for which $p_k \geqq \theta^* p_i$ $(i = 1, \cdots, k - 1)$, one can, by repeated application of Lemma 1 arrive at a new configuration which has $(k - g)$ of the p_i equal to zero, $(g - 2)$ of the p_i equal to p_k/θ^*, and the one remaining p_i in the closed interval $[0, p_k/\theta^*]$. Moreover, the probability of a correct selection for this configuration is at most as large as that for the original configuration. The purpose of the second lemma is to permit further reduction of the probability of a correct selection by starting with this new configuration and making changes of a second type.

To be more precise, we assume that

$$\text{(8)} \qquad \begin{aligned} p_k &= \theta^* p_{k-1} = \cdots = \theta^* p_{k-g+2} = \frac{\theta^*(1 - p_{k-g+1})}{\theta^* + g - 2} \\ 0 &\leqq p_{k-g+1} \leqq p_{k-g+2} = p_k/\theta^* \\ p_{k-g} &= \cdots = p_1 = 0, \end{aligned}$$

where g is an integer $2 \leqq g \leqq k$, $\theta^* > 1$, $p_i \geqq 0$ $(i = 1, \cdots, k)$, and $\sum_{j=1}^k p_i = 1$. Note that under these conditions $0 \leqq p_{k-g+1} \leqq 1/(\theta^* + g - 1)$. Let $p_{k-g+1} = p$ to simplify notation. It follows from (8) that ϕ_k may be regarded as a function of p only, say $\psi_2(p)$.

LEMMA 2. *Under the conditions* (8), $\psi_2(p)$ *is a non-increasing function of* p, *where* p *lies in the interval* $0 \leqq p \leqq 1/(\theta^* + g - 1)$. [Hence ϕ_k is minimized under (8) by the configuration $p_1 = \cdots = p_{k-g} = 0$, $p_{k-g+1} = p_{k-g+2} = \cdots = p_{k-1} = p_k/\theta^* = 1/(\theta^* + g - 1)$.]

PROOF OF LEMMA 2. Consider a multinomial problem in which only the $g - 1$ events $E_{k-g+2}, \cdots, E_k$ are involved; suppose the probabilities of these events are given by $1/(\theta^* + g - 2), \cdots, 1/(\theta^* + g - 2)$, $\theta^*/(\theta^* + g - 2)$, respectively. Let M observations be taken, and consider the quantities $y_{k-g+2, M}, \cdots, y_{kM}$. But now let an integer $0 \leqq c \leqq N$ be given which is to correspond to the event E_{k-g+1}, and use the selection procedure of [1] to choose among the g events $E_{k-g+1}, \cdots, E_k$. That is, the constant c plays the same role in the selection procedure that the chance quantity $y_{k-g+1,M}$ ordinarily would; otherwise the selection is made in the usual way. Let $Q_M(c)$ be the probability of choosing event E_k when this procedure is used.

ϕ_k can now be rewritten as

$$\phi_k = \psi_2(p) = \sum_{y=0}^{N} \binom{N}{y} p^y (1-p)^{N-y} Q_{N-y}(y). \tag{9}$$

Note that $Q_{N-y}(y)$ is independent of p, so that $\psi_2(p)$ is merely a linear combination of binomial terms. One may rewrite (9) as

$$\psi_2(p) = \sum_{y=0}^{N} [Q_{N-y}(y) - Q_{N-y+1}(y-1)] \sum_{m=y}^{N} \binom{N}{m} p^m (1-p)^{N-m}, \tag{10}$$

where we set $Q_{N+1}(-1) = 0$. Since $\sum_{m=y}^{N} \binom{N}{m} p^m (1-p)^{N-m}$ increases with p for all $0 < y \leqq N$ (i.e., the binomial distribution shifts to the right if the probability of a "success" increases), it clearly suffices for proving the lemma to show that $Q_{N-y}(y)$ is non-increasing in y.

From the definition of $Q_M(c)$, it is clear that

$$Q_{N-y-1}(y+1) \leqq Q_{N-y-1}(y). \tag{11}$$

In order to prove $Q_{N-y-1}(y+1) \leqq Q_{N-y}(y)$, therefore, one needs only the inequality $Q_{N-y-1}(y) \leqq Q_{N-y}(y)$. Hence it is sufficient for proving the lemma to prove the following

Assertion. Let $0 \leqq c \leqq N$ be given; then

$$Q_M(c) \leqq Q_{M+1}(c) \tag{12}$$

for all integers M.

To prove the assertion, note that $Q_M(c)$ is the probability that E_k is chosen after M observations; and that if an $(M+1)$st observation were taken, the vector $(y_{k-g+2,M+1}, \cdots, y_{k,M+1})$ could lead to any one of g possible decisions. Likewise, $Q_{M+1}(c)$ is the probability under the same selection procedure that E_k is chosen after $(M+1)$ observations, and this event might have arisen from any of a number of sequences, some of which would have led to E_k, some of which would have led to $E_{k-1}, \cdots$, some of which would haved led to E_{k-g+1}, had the selection been made after the Mth observation. In short, for fixed c, let $R(i,j) = Pr$ {choose E_i after M observations and choose E_j after $(M+1)$ observations}, where $i, j = k-g+1, \cdots, k$. Then we have

$$\begin{aligned} Q_{M+1}(c) &= \sum_{j=k-g+1}^{k} R(j,k), \\ Q_M(c) &= \sum_{j=k-g+1}^{k} R(k,j); \end{aligned} \tag{13}$$

also

$$Q_{M+1}(c) - Q_M(c) = \sum_{j=k-g+1}^{k} [R(j, k) - R(k, j)]. \tag{14}$$

Thus it suffices for proving Lemma 2 to show that each of the terms

$$R(j, k) - R(k, j) \geqq 0 \qquad (j = k - g + 1, \cdots, k - 1).$$

In the following, the subscript M is dropped from the cumulative sums y_{iM}; it is understood that the symbols y_i $(i = k - g + 1, \cdots, k)$ stand for cumulative sums after M observations. Write $m = \max(c, y_{k-g+2}, \cdots, y_k)$. m is said to occur s times if exactly s of the numbers $c, y_{k-g+2}, \cdots, y_k$ are equal to m and the other y_i's are less than m. The detailed expressions below for $R(j, k)$ and $R(k, j)$ apply only for $j = k - g + 2, \cdots, k - 1$. It is, of course, necessary to show $R(k - g + 1, k) - R(k, k - g + 1) \geqq 0$ as well, however we do not write out the expressions here. The argument for this case is much the same as is set out below for the other j, and is in fact somewhat simpler. In the selection procedure which defines $Q_{M+1}(c)$, the only way one can choose E_{k-g+1} after having chosen E_k on the previous observation is through the randomization process. Thus the last two parts of $R(k, k - g + 1)$ become zero. This also has the effect of strengthening the desired inequality.

One has, then, for $j = k - g + 2, \cdots, k - 1$,

$$\begin{aligned} R(j, k) = \sum_{s=1}^{g} \Big[& \frac{1}{s^2} \Pr\{y_i = y_k = m; m \text{ occurs } s \text{ times}; y_i < m - 1 \text{ for } \\ & \quad y_i \neq m; E_i \text{ occurs at the } (M+1)\text{st observation, for} \\ & \quad \text{some } i \text{ such that } y_i \neq m\} \\ & + \frac{1}{s(s+1)} \Pr\{y_j = y_k = m; m \text{ occurs } s \text{ times}; y_{i_1} = \cdots = y_{i_u} \\ & \quad = m - 1; y_i < m - 1 \text{ if } y_i \leqq m - 1 \text{ and} \\ & \quad i \neq i_1, \cdots, i_u; E_{i_h} \text{ occurs at the } (M+1)\text{st observa-} \\ & \quad \text{tion for some } h = 1, \cdots, u\} \\ & + \frac{1}{s(s+1)} \Pr\{y_j = m = y_k + 1; m \text{ occurs } s \text{ times}; E_k \text{ occurs} \\ & \quad \text{at the } (M+1)\text{st observation}\} \\ & + \frac{1}{s} \Pr\{y_j = y_k = m; m \text{ occurs } s \text{ times}; E_k \text{ occurs at the} \\ & \quad (M+1)\text{st observation}\} \Big] \end{aligned} \tag{15}$$

Similarly,

$$R(k, j) = \sum_{s=1}^{g} \Bigg[\frac{1}{s^2} \Pr \{ y_j = y_k = m; m \text{ occurs } s \text{ times}; y_i < m - 1 \text{ for } y_i \neq m; E_i \text{ occurs at the } (M + 1)\text{st observation, for some } i \text{ such that } y_i \neq m \}$$

$$+ \frac{1}{s(s+1)} \Pr \{ y_j = y_k = m; m \text{ occurs } s \text{ times}; y_{i_1} = \cdots = y_{i_u} = m - 1; y_i < m - 1 \text{ if } y_i \leqq m - 1 \text{ and } i \neq i_1, \cdots, i_u; E_{i_h} \text{ occurs at the } (M + 1)\text{st observation for some } h = 1, \cdots, u \} \tag{16}$$

$$+ \frac{1}{s(s+1)} \Pr \{ y_k = m = y_j + 1; m \text{ occurs } s \text{ times}; E_j \text{ occurs at the } (M + 1) \text{ observation} \}$$

$$+ \frac{1}{s} \Pr \{ y_j = y_k = m; m \text{ occurs } s \text{ times}; E_j \text{ occurs at the } (M + 1)\text{st observation} \} \Bigg]$$

Each of the probabilities $R(j, k)$ and $R(k, j)$ has been divided into four parts, and the parts are now considered in turn. The first two parts of (15) are identical with the first two parts of (16), hence these parts contribute nothing to $R(j, k) - R(k, j)$. As for the third parts, a typical term from the third part of (15) may be written

$$\frac{1}{s(s+1)} \frac{M!}{(m!)^s (m-1)!\, y_{j_1}! \cdots y_{j_{g-s-1}}!} (\theta^* p_1)^{m-1} p_1^{M-m+1} \cdot \theta^* p_1, \tag{17}$$

where $p_1 = 1/(\theta^* + g - 2)$. Similarly a corresponding term from the third part of (16) may be found merely by switching the role of k and j; this term may be written

$$\frac{1}{s(s+1)} \frac{M!}{(m!)^s (m-1)!\, y_{j_1}! \cdots y_{j_{g-s-1}}!} (\theta^* p_1)^{m} p_1^{M-m} \cdot p_1, \tag{18}$$

which is equal to (17). Since the terms in the third parts of (15) and (16) can be put into one-one correspondence, with the corresponding terms being equal, the third parts contribute nothing to $R(j, k) - R(k, j)$. One only has to consider the fourth parts of (15) and (16) therefore.

A typical term from the fourth part of (15) may be written

$$\frac{1}{s} \frac{M!}{(m!)^s y_{j_1}!, \cdots, y_{j_{g-s}}!} (\theta^* p_1)^{m} p_1^{M-m} \cdot \theta^* p_1, \tag{19}$$

while the corresponding term from the fourth part of (16) may be written

$$\frac{1}{s}\frac{M!}{(m!)^s y_{j_1}!, \cdots, y_{j_{g-s}}!}(\theta^* p_1)^m p_1^{M-m} \cdot p_1. \tag{20}$$

Expression (19) is at least as large as expression (20), since $\theta^* > 1$. Thus the fourth part of (15) is at least as large as the fourth part of (16) and we have $R(j, k) - R(k, j) \geqq 0$. The lemma then follows.

3. Proof of the property of the least favorable configuration.

THEOREM. *Let* $1 < \theta^* < \infty$ *be given, and suppose* $p_k \geqq \theta^* p_i$ $(i = 1, \cdots, k-1)$. *Then*

$$\phi(p_1, \cdots, p_k) \geqq \phi\left(\frac{1}{\theta^* + k - 1}, \cdots, \frac{1}{\theta^* + k - 1}, \frac{\theta^*}{\theta^* + k - 1}\right). \tag{21}$$

PROOF OF THEOREM. Let a vector of probabilities $(p_1, \cdots, p_k)$ be given which satisfies the hypotheses. By applying Lemma 1 repeatedly, one can produce a configuration $(p_1', \cdots, p_k')$ such that $\phi(p_1', \cdots, p_k') \leqq \phi(p_1, \cdots, p_k)$, and which has the following properties:

$$\text{(22)}\quad \begin{array}{lll} \text{(a)} & p_k' = p_k & \\ \text{(b)} & p_i' = p_k/\theta^* & \text{for } h-2 \text{ of the } p_i' \\ \text{(c)} & p_i' = 0 & \text{for } k-h \text{ of the } p_i' \\ \text{(d)} & 0 \leqq p_i' \leqq p_k/\theta^* & \text{for the one remaining } p_i', \end{array}$$

where h is an integer $2 \leqq h \leqq k$. If $p_i' = p_k/\theta^*$ for the i in (d) of (22), then the p_i' can be renumbered so that the configuration is in the form of (24), and we can proceed immediately to apply Lemma 2 with $g = h + 1$. If that $p_i' < p_k/\theta^*$, however, we can without loss of generality, renumber the p_i' for $(i < k)$ and rewrite (22) as follows:

$$\text{(23)}\quad \begin{array}{ll} \text{(a)} & p_k' = p_k \\ \text{(b)} & p_{k-1}' = p_{k-2}' = \cdots = p_{k-h+2}' = p_k/\theta^* \\ \text{(c)} & p_{k-h+1}' = 1 - \sum_{i=k-h+2}^{k} p_i' < p_k/\theta^* \\ \text{(d)} & p_{k-h}' = p_{k-h-1}' = \cdots = p_1' = 0. \end{array}$$

One can now apply Lemma 2 to the p_i' by substituting h for the g of that lemma. The result is a new configuration $(p_1'', \cdots, p_k'')$ which yields a probability of correct selection at most as large as the one corresponding to $(p_1', \cdots, p_k')$, and which has

$$\text{(24)}\quad \begin{array}{ll} \text{(a)} & p_k'' = \theta^* p_{k-1}'' = \cdots = \theta^* p_{k-h+1}'' \\ \text{(b)} & p_{k-h}'' = p_{k-h-1}'' = \cdots = p_1'' = 0. \end{array}$$

Another application of Lemma 2, with $g = h + 1$, to this new configuration will yield still another configuration whose probability of correct selection is again not larger than the one corresponding to $(p_1'', \cdots, p_k'')$, and for which property (24) holds with h replaced by $h + 1$.

Repeated applications of Lemma 2 lead to a configuration where (24) holds with $h = k$, and this configuration certainly yields $\phi_k \leqq \phi(p_1, \cdots, p_k)$ since the probability of correct selection did not increase at each step. But (24) with $h = k$ is the configuration whose probability of correct selection is given in the right side of (21). Hence the theorem is proved.

REFERENCE

[1] R. E. Bechhofer, S. Elmaghraby, and N. Morse, "A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 102–119.