SOME PROBLEMS IN STATISTICAL INFERENCE ON MULTINOMIAL POPULATIONS

By

Agnes Andreassian

Submitted in Partial Fulfillment for the Requirements
of the Degree Master of Science
in the Mathematics Department of the
American University of Beirut
Beirut, Lebanon.
1963

INFERENCE ON MULTINOMIAL POPULATIONS

Agnes Andreassian

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to Professor Mary Hanania Regier who guided me in the preparation and writing of the thesis with stimulating criticism and advice.

The Arts and Sciences Research Committee of the American University of Beirut provided the funds which made possible the lengthy computations involved in the preparation of Table 1. Mrs. A.B. Darakjian, Mr. K.Umar, and Mr. A.Tülkiyan assisted in the computations; and Miss M.Jabbour typed the thesis. To all these I wish to express my sincere thanks.

CONTENTS

INTRODUCTION

Problems of statistical inference involving multinomial distributions are very commonly met in practice. Many of these problems cannot readily be solved by the use of standard procedures, on account of the complicated form of the multinomial probability function, and so for large samples approximating distributions are used. There are a number of papers suggesting various approximations that may be used in connection with finding moments and testing the hypothesis of equal class probabilities. A summary of these results is given in Chapter I.

Another problem of importance is determination of the appropriate sample size, so that we do not take a sample unnecessarily large and yet are confident that the sample is large enough for our results to satisfy some specified requirements. Such a problem is discussed by Bechhofer, Elmaghraby, and Morse (1959) in which it is required to select the event with the highest class probability such that the probability of a correct selection is greater than or equal to P^* whenever the true ratio of the largest to the second largest of the class probabilities is greater than or equal to t^*, where P^* and t^* have been specified by the experimenter. It is this kind of problem that I intend to discuss in the following pages. In Chapter II, I consider the problem of determining the appropriate sample size for obtaining, with a specified confidence coefficient, a confidence region that has a volume less than or equal to some prescribed number. In Chapter III, the problem is to classify the events in a trinomial population according to their class probabilities. The approach taken is analogous to that used by Bechhofer, Elmaghraby, and Morse (1959). Two numbers, t and P, are specified by the experimenter and we wish to determine the sample size necessary to guarantee that the probability of correct classification is greater than or equal to P when the ratio of two consecutive values of the ranked class probabilities is greater than or equal to t. In obtaining the smallest sample size that guarantees such a requirement use is made of a property of the trinomial distribution the proof of which is given in Chapter IV.

CHAPTER I

REVIEW OF RELATED LITERATURE

In most problems of statistical inference involving multinomial distributions, we need to use some approximation, since for large samples, computations using the multinomial distribution become extremely laborious. There are several papers that treat this problem.

One such approximation was obtained by Johnson (1960). Considering a population of k classes with probabilities p_j ($j = 1, \ldots, k$) and a sample of size N yielding relative frequencies f_j ($j = 1, \ldots, k$), he obtains the approximating density

$$g(f_1, \ldots, f_k) = \Gamma(N-1) \prod_{j=1}^{k} [f_j^{(N-1)p_j - 1} / \Gamma((N-1)p_j)],$$

where $\sum_{j=1}^{k} p_j = 1$, $f_j = \frac{n_j}{N}$, and $\sum_{j=1}^{k} n_j = N$.

This approximation gives the correct first and second moments and product moments for the joint distribution of the relative frequencies and the correct range of variation. Also it is self-consistent in that distributions of subsets of the relative frequencies and distributions of sums of subsets of the relative frequencies derived from the approximation are of the same form as the initial function $g(f_1, \ldots, f_k)$. However, the approximation has some shortcomings. It does not give correct multinomial moments of order higher than two; the self-consistency of the approximation does not extend to the conditional distributions of some relative frequencies given others; and the maximum likelihood estimators of p_j in the exact model and the approximation differ by as much as $1/(2N)$.

With increasing sample size, the joint distribution of the standardized multinomial variables tends to a degenerate multivariate normal distribution with zero means, unit variances, and if $p_1 = p_2 = \ldots = p_k = 1/k$, covariance equal to $-1/(k-1)$. Therefore, an alternative approximation would be to consider this approximate multivariate normal distribution as the joint distribution of the standardized deviates from the sample mean. This was done by Johnson and Young (1960).

To test the equality of probabilities in a multinomial distribution against the alternative that the largest observed frequency is significant, using the multivariate normal approximation to the multinomial, one gets the result that the observed proportions are asymptotically distributed with the multivariate normal distribution with means $\frac{1}{k}$, variances $(k-1)/k^2N$ and covariances $-1/k^2N$. If t_j $(j = 1, \ldots, k)$ are the corresponding standardized variables, for the event $t_j \geq t^*$ where t^* is sufficiently large:

$$\Pr[\max t_j \geq t^*] \doteq \frac{k}{\sqrt{2\pi}} \int_{t^*}^{\infty} e^{-\frac{1}{2}t^2}\, dt$$

Korzelka (1956) has found that this gives a satisfactory approximation even for small values of N, but that there is a decrease in accuracy for increasing k.

Johnson and Young (1960) suggest the criteria $R = \frac{\min n_j}{\max n_j}$ and $W = \frac{\max n_j - \min n_j}{N}$ as alternatives to the standard chi-square test for the hypothesis of equal class probabilities. These ratios are simpler to calculate than is chi-square, though they cannot be safely used if the p_j's are not equal. The authors use both of the above mentioned approximations to the multinomial distribution to derive approximate significance limits for these criteria.

Weiss (1962) proposes a sequential test for the hypothesis of equal class probabilities which at each stage of the sampling process observes whether or not the observation falls in the set of categories that contain the largest number of earlier observations. If the observation falls in that set of categories at too many stages, it rejects the hypothesis. A comparison of this sequential test with the standard chi-square test for the binomial case shows that when the hypothesis is true the sequential test requires a smaller average number of observations than the fixed sample size test.

There are several papers that investigate the problem of moments in the multinomial distribution.

Greenwood and Glasgow (1950) derive the mean and variance for the maximum and minimum observed frequencies for the binomial case and the special trinomial case with $p_1 = p_2$. However, generalization of their method for the trinomial case in which p_1, p_2, and p_3 are not equal, seems to be unsatisfactory.

Kozelka (1956) derives the mean and the variance of the largest observed proportion for the general trinomial case by making use of the moment generating function technique, but he points out that an extension of this technique even to the case $k = 4$ presents difficulties.

McCarthy (1947) considers a certain class of box problems which is equivalent to the following. In a multinomial population of k classes we are interested in sampling until we obtain for the first time at least n_{j1} observations from the $j1^{th}$ class, at least n_{j2} observations from $j2^{th}$ class, ..., and at least n_{js} observations from the js^{th} class, where j1, j2, ..., js represent the numbers of that set of s classes ($1 \leq s \leq k$) which first satisfy the stated condition. The total number of observations to be taken is a random variable $N_s[n_1(p_1), \ldots, n_k(p_k)]$ the distribution of which assumes a very complicated multinomial form for $k > 2$. The author finds the exact moments of $N_1[n_1(p_1), n_2(p_2)]$ and $N_2[n_1(p_1), n_2(p_2)]$ and using these moments he derives approximate values for the mean and variance for any k and any set of n_j's and p_j's when $s = 1$ or $s = k$. He also derives an approximate formula for the mean of this random variable for any k and $2 \leq s \leq k - 1$, when $p_j = 1/_k$, $n_j = n$ $(j = 1,\ldots,k)$. Since it is very difficult to obtain exact values, there is no effective analytic approach to evaluate errors. A few isolated cases are considered by a combination of computational, graphical, and analytical methods and some precautions are given in order to minimize the error.

The problem of cumulants of multinomial distributions has been investigated by Wishart (1949). In his paper, he derives cumulant recurrence relations for the general multinomial form and for the corresponding form that is derived as a generalization of the negative binomial.

Rao (1957-58 and 1958) has two papers about maximum likelihood estimation for the multinomial distribution. He first proves the consistency of the maximum likelihood estimates for the hypothetical frequencies of the multinomial distribution and then establishes the consistency of the maximum likelihood estimate for a parameter occurring in the specification of the hypothetical frequencies. He also discusses properties of the "maximum likelihood equation estimates" which he defines as those roots of the likelihood equation that provide the maximum of $\sum_{j=1}^{k} f_j \log p_j(m)$, where f_j denote the observed relative frequencies and $p_j(m)$ are the hypothetical frequencies, when the paremeter m is restricted to the roots of $\sum_{j=1}^{k} \frac{f_j}{p_j} \frac{dp_j}{d_m} = 0$. In the first paper (1957-58) he considers the finite multinomial distribution and in the second paper (1958) he extends those results to the case of a multinomial distribution with an infinite number of cells.

Finally, there is a paper by Bechhofer, Elmaghraby, and Morse (1959) which describes a single-sample multiple-decision procedure for selecting the multinomial event that is associated with the largest class probability. It is required to have

$$\Pr[\text{correct selection} \mid t_{k,k-1} \geq t] \geq P$$

where $t_{k,k-1} = \frac{q_k}{q_{k-1}}$, $q_1 \leq q_2 \leq \cdots \leq q_k$ denoting the ranked probabilities of the k categories and where $t(> 1)$ and P have been specified by the experimenter such that t is the samllest value of the ratio $t_{k,k-1}$ that is considered worth detecting and P is the smallest acceptable value of the probability of correct selection.

The authors derive a large sample approximation to the probability of a correct selection in what they call "the least favorable configuration",[1] where

$$q_1 = q_2 = \cdots = q_{k-1} = \frac{1}{t+k-1} \quad \text{and} \quad q_k = \frac{t}{t+k-1} .$$

---

[1] Kesten and Morse (1959) prove that the probability of a correct selection in the configuration given by $\frac{q_k}{q_j} = t$ is less than or equal to the probability of a correct selection when $\frac{q_k}{q_j} \geq t$ $(j = 1, \ldots, k-1)$.

The approximation turns out to be a (k-1)-variate normal distribution for which tables are available. They also indicate the smallest value of sample size which will guarantee a specified probability of a correct selection for a particular value of t.

CHAPTER II

CHOICE OF SAMPLE SIZE FOR A CONFIDENCE REGION OF PRESCRIBED VOLUME

In problems of interval estimation, with a fixed confidence coefficient, the confidence interval can be made shorter by using a larger sample. Thus if the length of the desired confidence interval is determined ahead of the sampling process, we may be able to decide what sample size to take in order to obtain a confidence interval of length less than or equal to the prescribed length.

When the distribution involves several parameters, it would be desirable if we could prescribe the length of the confidence interval for each parameter separately and then find the smallest sample size that satisfies these requirements simultaneously. To do this in general, for the case of a multinomial population of $k+1$ classes, we would need an expression for each of the k semi-axes of the confidence ellipsoid in terms of the class probabilities p_i $(i = 1, \ldots, k)$ and sample size N, say $f_j(N, p_1, \ldots, p_k)$, $j = 1, \ldots, k$. We would then require that $f_j(N, p_1, \ldots, p_k) \leq a_j$ $(j = 1, \ldots, k)$ where the a_j are the prescribed numbers. If for the j^{th} inequality $(j = 1, \ldots, k)$, the maximum value of the smallest sample size that satisfies the inequality, N_j^o, is found corresponding to the different possible sets of p_i's, we would then obtain

$$N_1^o,\ N_2^o,\ \ldots,\ N_k^o\ .$$

Then $N = \max N_j^o$ $(j = 1, \ldots, k)$ would guarantee that all the conditions given by $f_j(N, p_1, \ldots, p_k) \leq a_j$ $(j = 1, \ldots, k)$ are satisfied for any possible set of class probabilities. But the expressions $f_j(N, p_1, \ldots, p_k)$ turn out to be complicated functions of $p_1, \ldots, p_k$ even for the case of $k = 3$ and there does not seem to be any straightforward way of getting such expressions for cases where we have more than three parameters. However, if we have no special interest in making one confidence interval shorter than any other, as is the case in most problems, we can use as will be shown below the lengths of the confidence intervals to guide us in the determination of a desirable volume of the confidence region. Having decided on such a volume we can determine what sample size to take to guarantee that with a given confidence coefficient our confidence region will be smaller than or equal to the desired size of the region.

It is a known fact (Mood, 1950) that the maximum likelihood estimators

$\hat{p}_i = \frac{N_i}{N}$ $\left(i = 1, \ldots, k\right.$ where N_i are the number of occurrences of the event $\left.E_i\right)$ for the parameters p_i of a multinomial distribution of $k + 1$ classes from samples of size N are, for large samples, approximately distributed by the multivariate normal distribution with means p_i and with coefficients $N\sigma^{ij}$ of the quadratic form where

$$\sigma^{ij} = \frac{\delta_{ij}}{p_i} + \frac{1}{p_{k+1}} \qquad i, j = 1, \ldots, k$$

$$\delta_{ij} = 1 \text{ if } i = j \text{ and } \delta_{ij} = 0 \text{ if } i \neq j$$

Hence, the approximate large-sample distribution of the estimators is given by: $g(\hat{p}_1, \ldots, \hat{p}_k)$

$$= \left(\frac{1}{2\pi}\right)^{k/2} \sqrt{\frac{N^k}{\prod_{i=1}^{k+1} p_i}}\; e^{-\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} N\left(\frac{\delta_{ij}}{p_i} + \frac{1}{p_{k+1}}\right)(\hat{p}_i - p_i)(\hat{p}_j - p_j)}$$

and the large sample variances and covariances are given by $\frac{p_i}{N}(1 - p_i)$ and $\frac{-p_i p_j}{N}$ respectively which are also the exact variances and covariances for any sample size.

It is also known that the quadratic form of a k-variate normal distribution has the chi-square distribution with k degrees of freedom and that the accuracy of our approximation is not impaired by substituting the maximum likelihood estimates for the parameters in the coefficients of the quadratic form. That is, for the case of a multinomial distribution, the quantity

$$v = \sum_{i=1}^{k} \sum_{j=1}^{k} N\left(\frac{\delta_{ij}}{\hat{p}_i} + \frac{1}{\hat{p}_{k+1}}\right)(\hat{p}_i - p_i)(\hat{p}_j - p_j)$$

is approximately distributed like chi-square with k degrees of freedom. Thus, if $K_{\varepsilon,k}$ is the upper ε point of the chi-square distribution corresponding to a confidence coefficient $1 - \epsilon$ and k degrees of freedom,

$$\Pr[v < K_{\epsilon,k}] = 1 - \epsilon$$

determines a confidence region in the parameter space the boundary of which is given by the equation of an ellipsoid in a k-dimensional space with its center at $(\hat{p}_1, \ldots, \hat{p}_k)$. Thus, the equation of the ellipsoid is given by

$$\sum_{i=1}^{k} \sum_{j=1}^{k} N \left(\frac{\delta_{ij}}{\hat{p}_i} + \frac{1}{\hat{p}_{k+1}} \right) (\hat{p}_i - p_i)(\hat{p}_j - p_j) = K_{\epsilon,k}$$

It is possible to determine the sample size such that the volume of the confidence region we set up will be smaller than or equal to some preassigned number. While limiting the volume does not limit the length of each confidence interval separately, it does limit the product of their lengths and thus provides a kind of control on the over-all precision of the estimating process. Since our parameters in the multinomial distribution are all population proportions, one would usually desire the same precision in the estimation of each parameter, as measured by the length of its confidence interval. Thus we shall choose the sample size in order that the volume of the confidence region we get may be less than or equal to that of the sphere whose diameter is the average length of the confidence intervals we expect.

Cramér (1946) proves that the k-dimensional volume of the domain bounded by the ellipsoid $Q = c^2$, where Q represents a definite positive quadratic form of matrix $\underset{\sim}{A}$, is given by

$$V = \frac{\pi^{k/2}}{\Gamma(\frac{k}{2}+1)} \cdot \frac{c^k}{\sqrt{A}}$$

where A is the determinant $|\underset{\sim}{A}|$.

In our problem, it is clear from $g(\hat{p}_1, \ldots, \hat{p}_k)$ that $A = \frac{N^k}{\prod_{i=1}^{k+1} \hat{p}_i}$. Thus,

$$V = \frac{\pi^{k/2}}{\Gamma(\frac{k}{2}+1)} \cdot \frac{c^k \left(\prod_{i=1}^{k+1} \hat{p}_i\right)^{1/2}}{N^{k/2}} \quad \text{where} \quad c^2 = K_{\varepsilon,k}$$

Now, if V_o represents the desired volume, we would like to know what sample size to take so that

$$\frac{\pi^{k/2}}{\Gamma(\frac{k+2}{2})} \cdot \frac{c^k \left(\prod_{i=1}^{k+1} \hat{p}_i\right)^{1/2}}{N^{k/2}} \leq V_o$$

Since the $\hat{p}_i$ depend on the sample, we cannot predict what these will be, but if we would like to choose N so that the above inequality will be satisfied no matter what the maximum likelihood estimates turn out to be, we can consider those values of $\hat{p}_i$ which would require the largest N and then we would be sure that the same sample size would satisfy the inequality for any other values of $\hat{p}_i$.

Thus we wish to find those values of $\hat{p}_i$ which would maximize $\prod_{i=1}^{k+1} \hat{p}_i$ and thus demand the largest value of N.

Since $\sum_{i=1}^{k+1} \hat{p}_i = 1$ is fixed, $\prod_{i=1}^{k+1} \hat{p}_i$ will be maximized when all the $\hat{p}_i$ are equal to $\frac{1}{k+1}$. And so we must choose N such that

$$N^{k/2} \geq \frac{\pi^{k/2} c^k}{(k+1)^{\frac{k+1}{2}} V_o \Gamma(\frac{k+2}{2})},$$

or

$$N \geq \frac{\pi c^2}{(k+1)^{\frac{k+1}{k}} [V_o \Gamma(\frac{k+2}{2})]^{2/k}}$$

where $c^2 = K_{\varepsilon,k}$, $k + 1$ is the number of classes, and V_o is the desired volume.

As an example, consider a trinomial distribution where we would like to estimate p_1 and p_2 with confidence coefficient 0.95, and where we wish to have the product of the lengths of the two confidence intervals $\leq \frac{1}{9}$. Thus we want to determine sample size such that the area of the ellipse $\leq \frac{\pi}{36}$. We have

$$c^2 = K_{0.05,2} = 5.99$$

$$k = 2$$

$$V_o = A_o = \frac{\pi}{36}.$$

Therefore, $N \geq \frac{\pi(5.99)}{(3)^{3/2} \cdot \frac{\pi}{36} \Gamma(2)}$,

or $N \geq 41.47$

Now, if we take a sample of size 42, using the maximum likelihood estimates, we can set up a confidence region of area $\leq \frac{\pi}{36}$ with confidence coefficient 0.95.

CHAPTER III

CLASSIFYING TRINOMIAL EVENTS ACCORDING TO THEIR CLASS PROBABILITIES

In certain problems on multinomial populations we may not be interested in estimates of the class probabilities but only in the ranks of the class probabilities. In this chapter a single-sample procedure for classifying the events in a trinomial population is considered.

Let $\underset{\sim}{X}_j = (X_{1j}, X_{2j}, X_{3j})$ be independent vector observations from a trinomial population with an unknown probability vector $\underset{\sim}{p} = (p_1, p_2, p_3)$, where p_i is the probability of the event E_i, $0 \leq p_i \leq 1$, $\sum_{i=1}^{3} p_i = 1$, and $X_{ij} = 1$ if the event E_i occurs on the j^{th} observation and $X_{ij} = 0$ if it does not $(i = 1, 2, 3;$ $j = 1, \ldots, N$ where N is the number of vector observations). Let $q_1 \leq q_2 \leq q_3$ denote the ranked probabilities p_1, p_2, p_3 and $T_{2,1} = \frac{q_2}{q_1}$ and $T_{3,2} = \frac{q_3}{q_2}$. Let $Y_{iN} = \sum_{j=1}^{N} X_{ij}$ be the number of observations falling in the i^{th} class $(i = 1, 2, 3)$; let $u_1 \leq u_2 \leq u_3$ denote the ranked values of Y_{iN} and F_i be the event associated with u_i.

The problem is to associate each event, E_i, with one of the ranked probabilities, q_i. Before experimentation starts, the experimenter should specify the smallest value, t, of the ratios $T_{2,1}$ and $T_{3,2}$ that is worth detecting and the smallest acceptable value, P, of the probability of correctly classifying the events when $T_{2,1}$ and $T_{3,2}$ are greater than or equal to t. That is, it is required that

$$\Pr[\text{correct classification} \mid T_{k,k-1} \geq t] \geq P \quad (k = 2,3)$$

To achieve the above goal, take a random sample of N vector observations and compute u_1, u_2, and u_3.

(1) If $u_1 < u_2 < u_3$ associate each F_i with q_i $(i = 1, 2, 3)$.

(2) If $u_1 < u_2$ but $u_2 = u_3$, associate F_1 with q_1 and use a random device to select one of F_2 and F_3 to be associated with q_2 and the other with q_3.

(3) Use a similar procedure if $u_1 = u_2$ and $u_2 < u_3$.

(4) If $u_1 = u_2 = u_3$ select as the event associated with q_1 one of F_1, F_2, and F_3, and associated with q_2 one of the remaining two events, always using a random device.

With this procedure for classifying the three events, our problem reduces to the proper choice of sample size so that the probability of a correct classification $\geq P$ for all possible $q_1 \leq q_2 \leq q_3$ for which $T_{k,k-1} \geq t$ $(k = 2, 3)$.

Thus we consider that set of probabilities q_1, q_2, q_3 which for any given N and t minimizes the probability of a correct classification when $T_{k,k-1} \geq t$ $(k = 2, 3)$. This will be called the least favorable configuration of the q_i's. It is proved in the next chapter that the least favorable configuration of the q_i's in a trinomial distribution is given by $T_{k,k-1} = t$ $(k = 2, 3)$. That is, $\frac{q_3}{q_2} = \frac{q_2}{q_1} = t$ and since $q_1 + q_2 + q_3 = 1$,

$$q_1 = \frac{1}{1 + t + t^2}, \quad q_2 = \frac{t}{1 + t + t^2}, \quad q_3 = \frac{t^2}{1 + t + t^2}$$

This means that if we choose N large enough to make the probability of correct classification with the least favorable configuration greater than or equal to P, we will be sure of at least that probability for any configuration of the q_i's with $T_{k,k-1} \geq t$ $(k = 2, 3)$.

For any fixed N the exact probability of correct classification is given by

$$Q = Q(q_1, q_2, q_3) = \Sigma\, S(u_1, u_2, u_3) . \frac{N!}{u_1!u_2!u_3!} q_1^{u_1} q_2^{u_2} q_3^{u_3}$$

where the summation is taken over all vectors $\underset{\sim}{u}_N = (u_1, u_2, u_3)$ such that $\sum_{i=1}^{3} u_i = N$ and $u_1 \leq u_2 \leq u_3$, and

$$\begin{aligned} S(u_1, u_2, u_3) &= 1 \quad \text{if} \quad u_1 < u_2 < u_3 \\ &= \tfrac{1}{2} \quad \text{if} \quad u_1 < u_2 \text{ and } u_2 = u_3; \text{ or } u_1 = u_2 \\ &\qquad\qquad \text{and } u_2 < u_3. \\ &= \tfrac{1}{6} \quad \text{if} \quad u_1 = u_2 = u_3. \end{aligned}$$

If any two q_i's are equal, either way of classifying then is considered correct. Similarly, if all three q_i's are equal, any way of classification is considered correct.

Table 1 gives exact probabilities of correct classification with the least favorable configuration for $t = 1.1(.1)1.4(.2)2.0(.5)3.0$ and $N = 1(1)30$.

TABLE 1

EXACT PROBABILITY OF CORRECT CLASSIFICATION[1]

N	t=1.1	t=1.2	t=1.3	t=1.4	t=1.6	t=1.8	t=2.0	t=2.5	t=3.0
1	0.1828	0.1978	0.2118	0.2248	0.2481	0.2682	0.2857	0.3205	0.3462
2	.1883	.2087	.2277	.2454	.2769	.3037	.3265	.3698	.3995
3	.1937	.2216	.2480	.2632	.3198	.3609	.3965	.4646	.5100
4	.1986	.2300	.2602	.2893	.3415	.3873	.4265	.5005	.5488
5	.2028	.2387	.2736	.3070	.3686	.4223	.4681	.5546	.7166
6	.2066	.2459	.2857	.3232	.3916	.4505	.4999	.5896	.7556
7	.2100	.2536	.2964	.3373	.4121	.4763	.5300	.6262	.8511
8	.2083	.2604	.3067	.3511	.4318	.4998	.5560	.6540	.8760
9	.2164	.2670	.3169	.3646	.4508	.5231	.5820	.6827	.8961
10	.2188	.2730	.3332	.3764	.4675	.5449	.6032	.7052	.9046
11	.2221	.2790	.3341	.3880	.4816	.5627	.6251	.7275	.9106
12	.2248	.2847	.3434	.3998	.4965	.5811	.6448	.7464	.9117
13	.2274	.2901	.3517	.4104	.5110	.5980	.6628	.7642	.9122
14	0.2299	0.2954	0.3599	0.4210	0.5256	0.6142	0.6818	0.7799	0.9111

TABLE 1 - Continued.

N	t=1.1	t=1.2	t=1.3	t=1.4	t=1.6	t=1.8	t=2.0	t=2.5	t=3.0
15	0.2323	0.3005	0.3677	0.4410	0.5395	0.6285	0.6865	0.7921	0.9080
16	.2347	.3056	.3754	.4412	.5535	.6440	.7086	.8078	.9162
17	.2254	.2994	.3730	.4426	.5614	.6548	.7217	.8198	.9181
18	.2358	.3123	.3877	.4583	.5781	.6665	.7362	.8316	.9251
19	.2329	.3116	.3897	.4631	.5875	.6819	.7482	.8415	.9303
20	.2327	.3166	.3978	.4735	.6008	.6954	.7614	.8516	.9380
21	.2356	.3190	.4014	.4785	.6090	.7048	.7714	.8603	.9443
22	.2395	.3253	.4102	.4895	.6206	.7170	.7831	.8688	.9531
23	.2355	.3221	.4101	.4918	.6186	.7247	.7906	.8762	.9576
24	.2426	.3321	.4219	.5064	.6406	.7375	.8151	.8839	.9683
25	.2417	.3325	.4250	.5093	.6452	.7436	.8078	.8904	.9743
26	.2424	.3365	.4294	.5160	.6562	.7646	.8148	.8961	.9820
27	.2456	.3410	.4347	.5223	.6634	.7730	.8250	.9040	.9899
28	.2481	.3458	.4423	.5306	.6718	.7813	.8318	.9072	.9940
29	.2469	.3468	.4439	.5348	.6787	.7954	.8375	.9122	0.9985
30	0.2504	0.3526	0.4517	0.5440	0.6883	0.8049	0.8449	0.9075	1.0000

[1] For a given value of t, with increasing sample size we expect to get higher probabilities of correct classification. There are a number of entries contradictory to this, indicating possible mistakes in computation. Lack of time and funds has prevented re-computing these probabilities.

Even for moderately large values of N, computing the exact probability of classification becomes extremely laborious. Thus, we consider next an approximation to the probability of correct classification.

Using an approach analogous to that used by Bechhofer, Elmaghraby, and Morse (1959) in their paper "A Single-Sample Multiple-Decision Procedure for Selecting the Multinomial Event Which Has The Highest Probability", we let

$$W_i = 2 \arcsin \sqrt{u_{i+1}/N} - 2 \arcsin \sqrt{u_i / N} \qquad i = 1, 2$$

The probability of correct classification is given by

$$Q = \Pr[u_3 - u_2 > 0,\ u_2 - u_1 > 0] + \tfrac{1}{2} \Pr[u_3 - u_2 > 0, u_2 - u_1 = 0]$$

$$+ \tfrac{1}{2} \Pr[u_3 - u_2 = 0,\ u_2 - u_1 > 0] + \tfrac{1}{6} \Pr[u_3 - u_2 = 0,\ u_2 - u_1 = 0].$$

Because of the equality signs, the last three terms become negligible for large N and Q can be approximated by

$$Q^\circ = \Pr[u_3 - u_2 \geq 0,\ u_2 - u_1 \geq 0]$$

or $$Q^\circ = \Pr[W_1 \geq 0,\ W_2 \geq 0].$$

We need to find $E(W_i)$, $V(W_i)$ and $\mathrm{Cov}(W_1, W_2)$ in the least favorable configuration.

If we expand $2 \arcsin \sqrt{u_i/N}$ around the point $\frac{u_i}{N} = q_i$, we obtain

$$2 \arcsin \sqrt{u_i/N} = f_i + f_i'(\frac{u_i}{N} - q_i) + \frac{1}{2} f_i''(\frac{u_i}{N} - q_i)^2$$
$$+ \frac{1}{6} f_i'''(\frac{u_i}{N} - q_i)^3 + \frac{1}{24} f_i^{iv}(\frac{u_i}{N} - q_i)^4 + 0(\frac{1}{N^5})$$

where $f_i = 2 \arcsin \sqrt{q_i}$

$$f_i' = \frac{1}{\sqrt{q_i(1 - q_i)}}$$

$$f_i'' = \frac{2q_i - 1}{2\sqrt{q_i^3(1 - q_i)^3}}$$

$$f_i''' = \frac{8q_i^2 - 8q_i + 3}{4\sqrt{q_i^5(1 - q_i)^5}}$$

$$f_i^{iv} = \frac{3(2q_i - 1)(8q_i^2 - 8q_i + 5)}{8\sqrt{q_i^7(1 - q_i)^7}}$$

We find

$$E(2 \arcsin \sqrt{u_i/N}) = f_i + f_i'' \frac{q_i(1 - q_i)}{2N} + f_i''' \frac{q_i(1-q_i)(1-2q_i)}{6N^2}$$
$$+ f_i^{iv}[\frac{q_i^2(1 - q_i)^2}{8N^2} + \frac{q_i(1 - q_i)(1 - 6q_i + 6q_i^2)}{24N^3}] + 0(\frac{1}{N^4})$$

Since $E(W_i) = E(2 \arcsin \sqrt{u_i + 1/N}) - E(2 \arcsin \sqrt{u_i/N})$ $(i=1,2)$, we find, in the least favorable configuration,

$$E(W_1) = a_1 + \frac{a_2}{N} + 0(\frac{1}{N^2})$$

where $a_1 = 2 \arcsin \sqrt{\frac{t}{1 + t + t^2}} - 2 \arcsin \sqrt{\frac{1}{1 + t + t^2}}$

$$a_2 = \frac{-(1 - t + t^2)}{4\sqrt{t(1 + t^2)}} - \frac{1 - t - t^2}{4\sqrt{t(1 + t)}}$$

$$E(W_2) = a_3 + \frac{a_4}{N} + o(\frac{1}{N^2})$$

where $a_3 = 2 \arcsin \sqrt{\frac{t^2}{1 + t + t^2}} - 2 \arcsin \sqrt{\frac{t}{1 + t + t^2}}$

$$a_4 = \frac{-(1 + t - t^2)}{4t \sqrt{1 + t}} + \frac{1 - t + t^2}{4 \sqrt{t(1+t^2)}}$$

The variance of W_i $(i = 1, 2)$ is given by

$$V(W_i) = V(2 \arcsin \sqrt{u_i + 1/N}) + V(2 \arcsin \sqrt{u_i/N})$$

$$- 2 \operatorname{Cov}(2 \arcsin \sqrt{u_i + 1/N}, 2 \arcsin \sqrt{u_i/N})$$

But $V(2 \arcsin \sqrt{u_i/N}) = E(2 \arcsin \sqrt{u_i/N})^2 - [E(2 \arcsin \sqrt{u_i/N})]^2$

$$= \frac{1}{N} + \frac{1}{N^2} [\frac{4q_i^2 - 4q_i + 3}{8q_i(1-q_i)}] + o(\frac{1}{N^3})$$

and

$$\operatorname{Cov}(2 \arcsin \sqrt{u_i/N}, 2 \arcsin \sqrt{u_j/N})$$

$$= E[(2 \arcsin \sqrt{u_i/N})(2 \arcsin \sqrt{u_j/N})] - E(2 \arcsin \sqrt{u_i/N})$$

$$E(2 \arcsin \sqrt{u_j/N})$$

$$= - \frac{1}{N} \frac{q_i q_j}{\sqrt{q_i q_j (1-q_i)(1-q_j)}} + \frac{1}{N^2}[\frac{q_i q_j (2q_i-1)(2q_j-1)}{8\sqrt{q_i q_j (1-q_i)^3 (1-q_j)^3}}$$

$$+ \frac{q_i q_j (2q_i - 1)^2}{4\sqrt{q_i^3 q_j (1-q_i)^3 (1-q_j)}} + \frac{q_i q_j (2q_j - 1)^2}{4\sqrt{q_i q_j^3 (1-q_i)(1-q_j)^3}}$$

$$- \frac{q_j (8q_i^2 - 8q_i + 3}{8\sqrt{q_i q_j (1-q_i)^3 (1-q_j)}} - \frac{q_i (8q_j^2 - 8q_j + 3)}{8\sqrt{q_i q_j (1-q_i)(1-q_j)^3}}] + o(\frac{1}{N^3})$$

Thus, in the least favorable configuration,

$$V(W_1) = \frac{b_1}{N} + \frac{b_2}{N^2} + O(\frac{1}{N^3})$$

where
$$b_1 = 2 + \frac{2}{\sqrt{(1+t)(1+t^2)}}$$

$$b_2 = \frac{(1+t+t^2)^2(2+t+t^2) + (1-t-t^2)(1-t+t^2)}{4t(1+t)(1+t^2)\sqrt{(1+t)(1+t^2)}}$$

$$+ \frac{3(1+t+t^2)^2\ (2+t+t^2)}{8t(1+t)(1+t^2)} - 1$$

and $V(W_2) = \frac{b_3}{N} + \frac{b_4}{N^2} + O(\frac{1}{N^3})$

where
$$b_3 = 2 + \frac{2t^2}{\sqrt{t(1+t)(1+t^2)}}$$

$$b_4 = \frac{(1+t+t^2)^2(1+t+2t^2) - t^2(1+t-t^2)(1-t+t^2)}{4(1+t)(1+t^2)\sqrt{t(1+t)(1+t^2)}}$$

$$+ \frac{3(1+t+t^2)^2\ (1+t+2t^2)}{8t^2(1+t)(1+t^2)} - 1$$

Finally,

$$\begin{aligned} Cov(W_1,W_2) &= E(W_1\ W_2) - E(W_1)\ E(W_2) \\ &= Cov(2 \arcsin \sqrt{u_2/N},\ 2 \arcsin \sqrt{u_3/N}) \\ &- V(2 \arcsin \sqrt{u_2/N}) - Cov(2 \arcsin \sqrt{u_1/N}, \\ &\qquad 2 \arcsin \sqrt{u_3/N}) \\ &+ Cov(2 \arcsin \sqrt{u_1/N},\ 2 \arcsin \sqrt{u_2/N}) \end{aligned}$$

which in the least favorable configuration gives the result:

$$\mathrm{Cov}(W_1, W_2) = \frac{c_1}{N} + \frac{c_2}{N^2} + O\left(\frac{1}{N^3}\right)$$

where $$c_1 = -1 - t\sqrt{\frac{t}{(1+t)(1+t^2)}} - \sqrt{\frac{1}{(1+t)(1+t^2)}} + \frac{\sqrt{t}}{1+t}$$

$$c_2 = \frac{t^2(1-t+t^2)(1+t-t^2) - (1+t+t^2)^2(1+t+2t^2)}{8(1+t^2)(1+t)\sqrt{t(1+t^2)(1+t)}}$$

$$- \frac{(1-t-t^2)(1-t+t^2) + (1+t+t^2)^2(2+t+t^2)}{8t(1+t^2)(1+t)\sqrt{(1+t^2)(1+t)}}$$

$$+ \frac{(1+t+t^2)^2}{8t(1+t)\sqrt{t}} + \frac{(1-t-t^2)(1+t-t^2)}{8(1+t)^3\sqrt{t}} + \frac{1}{2} - \frac{3(1+t+t^2)^2}{8t(1+t^2)}$$

As $N \to \infty$, it will be shown below, that the joint distribution of the standardized variables approaches a bivariate normal distribution with zero means, unit variances, and a correlation coefficient equal to

$$\frac{\mathrm{Cov}(W_1, W_2)}{\sqrt{V(W_1)}\sqrt{V(W_2)}}$$

It is shown by Cramer (1946, pp. 418-419) that in the limit the variables given by

$$z_i = \frac{u_i - Nq_i}{\sqrt{Nq_i}} \quad (i = 1, 2, 3)$$

have a singular normal distribution, with zero means and covariance matrix

$$\begin{pmatrix} 1 - q_1 & -\sqrt{q_1 q_2} & -\sqrt{q_1 q_3} \\ -\sqrt{q_1 q_2} & 1 - q_2 & -\sqrt{q_2 q_3} \\ -\sqrt{q_1 q_3} & -\sqrt{q_2 q_3} & 1 - q_3 \end{pmatrix}$$

Thus, as $N \to \infty$, the distribution of $\frac{u_1}{N}, \frac{u_2}{N}, \frac{u_3}{N}$ approaches $N(\underset{\sim}{\rho}, \underset{\sim}{\Sigma})$ where $\underset{\sim}{\rho} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}$ and

$$\underset{\sim}{\Sigma} = \begin{pmatrix} \frac{q_1(1-q_1)}{N} & \frac{-q_1 q_2}{N} & \frac{-q_1 q_3}{N} \\ \frac{-q_1 q_2}{N} & \frac{q_2(1-q_2)}{N} & \frac{-q_2 q_3}{N} \\ \frac{-q_1 q_3}{N} & \frac{-q_2 q_3}{N} & \frac{q_3(1-q_3)}{N} \end{pmatrix}$$

Now, since W_1 and W_2 are continuous functions of $\frac{u_i}{N}$ $(i = 1, 2, 3)$ and have continuous first partial derivatives in the open interval $0 < \frac{u_i}{N} < 1$, and since as $N \to \infty$ the probability that $\frac{u_i}{N} = 0$ or $\frac{u_i}{N} = 1$ approaches zero, we can apply the lemma given by Rao (1952, p.207) to conclude that as $N \to \infty$ the distribution of W_1 approaches that of $N(a_1, \frac{b_1}{N})$ and the distribution of W_2 approaches that of $N(a_2, \frac{b_2}{N})$. Again using the lemma, we can conclude that any linear combination of W_1 and W_2 approaches a normal distribution as $N \to \infty$. Now, by a theorem given by Anderson (1958, p.37) we conclude that W_1, W_2 have, in the limit, a bivariate normal distribution with mean vector $\begin{pmatrix} E(W_1) \\ E(W_2) \end{pmatrix}$ and covariance matrix $\begin{pmatrix} V(W_1) & Cov(W_1 W_2) \\ Cov(W_1 W_2) & V(W_2) \end{pmatrix}$.

Therefore, for given t and N, an approximation to the probability of correct classification is obtained by

$$Q^\circ = \int_{\frac{-A}{\sqrt{C}}}^{\infty} \int_{\frac{-B}{\sqrt{D}}}^{\infty} g(t_1,t_2)\, dt_1\, dt_2$$

where

$$A = a_1 + \frac{a_2}{N}$$

$$B = a_3 + \frac{a_4}{N}$$

$$C = \frac{b_1}{N} + \frac{b_2}{N^2}$$

$$D = \frac{b_3}{N} + \frac{b_4}{N^2}$$

$$E = \frac{c_1}{N} + \frac{c_2}{N^2}$$

and $g(t_1,t_2)$ is the bivariate normal density function with zero means, unit variances, and correlation $\frac{E}{\sqrt{BD}}$.

Since $\frac{-A}{\sqrt{C}}$ and $\frac{-B}{\sqrt{D}}$ happen to be negative numbers, the approximate probability of correct classification can be found using the fact that

$$Q^\circ = F\left(\frac{A}{\sqrt{C}}\right) + F\left(\frac{B}{\sqrt{D}}\right) + F'\left(\frac{A}{\sqrt{C}}, \frac{B}{\sqrt{D}}\right) - 1$$

where $F(x) = \int_{-\infty}^{x} \frac{e^{\frac{-z^2}{2}}}{\sqrt{2\pi}}\, dz$ is the area under the standardized normal curve

and $F'(x,y) = \int_{x}^{\infty} \int_{y}^{\infty} g(t_1, t_2)\, dt_1\, dt_2$ is the volume under a bivariate normal surface (Elderton et al. 1930).

Table 2 gives such approximate probabilities of correct classification with the least favorable configuration for $t = 1.1(.1)1.4(.2)2.0(.5)3.0)$ and $N = 5(5)50(10)100(25)$ 250(50)500(100)1000(200)2000(500)5000, 6000(2000)12000.

TABLE 2

APPROXIMATE PROBABILITIES OF CORRECT CLASSIFICATION

N	t=1.1	t=1.2	t=1.3	t=1.4	t=1.6	t=1.8	t=2.0	t=2.5	t=3.0
5	0.2018	0.2399	0.2746	0.3111	0.3738	0.4278	0.4718	0.5535	0.6075
10	.2209	.2746	.3272	.3767	.4696	.5436	.6051	.7030	.7545
15	.2327	.3004	.3684	.4305	.5435	.6292	.6951	.7917	.8382
20	.2422	.3273	.4021	.4787	.6081	.6946	.7592	.8490	.8855
25	.2544	.3463	.4333	.5178	.6502	.7446	.8088	.8870	.9162
30	.2619	.3685	.4646	.5533	.6922	.7872	.8444	.9144	.9377
35	.2720	.3854	.4928	.5859	.7293	.8200	.8753	.9330	.9531
40	.2798	.4022	.5180	.6143	.7610	.8471	.8959	.9477	.9642
45	.2901	.4193	.5374	.6421	.7866	.8683	.9145	.9584	.9722
50	.2953	.4334	.5590	.6661	.8088	.8863	.9286	.9678	.9781
60	.3084	.4618	.5992	.7053	.8463	.9168	.9509	.9788	.9871
70	.3219	.4874	.6323	.7415	.8769	.9369	.9643	.9862	.9918
80	.3354	.5125	.6595	.7746	.9000	.9525	.9743	.9908	.9951
90	.3463	.5349	.6878	.7994	.9174	.9631	.9812	.9940	.9968
100	.3574	.5567	.7125	.8228	.9327	.9716	.9863	.9959	.9979
125	.3825	.6047	.7658	.8678	.9572	.9845	.9934	.9984	.9993
150	.4050	.6403	.8071	.8996	.9728	.9917	.9968	.9994	.9998
175	.4249	.6790	.8384	.9237	.9828	.9953	.9984	.9998	0.9999
200	.4476	.7085	.8646	.9411	.9888	.9973	.9992	0.9999	1.0000
225	.4674	.7364	.8861	.9554	.9927	.9985	.9996	1.0000	......
250	.4844	.7607	.9052	.9659	.9952	.9991	.9998	......	......
300	.5181	.8029	.9314	.9789	.9977	.9997	0.9999	......	......
350	0.5485	0.8349	0.9504	0.9869	0.9990	0.9999	1.0000	......	......

TABLE 2 - Continued

N	t=1.1	t=1.2	t=1.3	t=1.4	t=1.6	t=1.8	t=2.0	t=2.5	t=3.0.
400	0.5728	0.8630	0.9642	0.9918	0.9995	1.0000	.....	.....	.....
450	.5993	.8838	.9739	.9949	.9998	.....	.....	.....	.....
500	.6225	.9021	.9805	.9968	0.9999	.....	.....	.....	.....
600	.6645	.9301	.9897	.9987	1.0000	.....	.....	.....	.....
700	.6993	.9502	.9942	.9994	.....	.....	.....	.....	.....
800	.7300	.9633	.9967	.9998	.....	.....	.....	.....	.....
900	.7565	.9738	.9982	0.9999	.....	.....	.....	.....	.....
1000	.7812	.9806	.9989	1.0000	.....	.....	.....	.....	.....
1200	.8207	.9894	.9997	.....	.....	.....	.....	.....	.....
1400	.8531	.9942	0.9999	.....	.....	.....	.....	.....	.....
1600	.8788	.9968	1.0000	.....	.....	.....	.....	.....	.....
1800	.8998	.9982	.....	.....	.....	.....	.....	.....	.....
2000	.9166	.9989	.....	.....	.....	.....	.....	.....	.....
2500	.9469	.9998	.....	.....	.....	.....	.....	.....	.....
3000	.9656	0.9999	.....	.....	.....	.....	.....	.....	.....
3500	.9775	1.0000	.....	.....	.....	.....	.....	.....	.....
4000	.9851	.....	.....	.....	.....	.....	.....	.....	.....
4500	.9902	.....	.....	.....	.....	.....	.....	.....	.....
5000	.9934	.....	.....	.....	.....	.....	.....	.....	.....
6000	.9970	.....	.....	.....	......	.....	......	.....	.....
8000	.9994	......	.....	.....	......	.....	.....	.....	.....
10000	0.9999	......	.....	.....	......	.....	.....	.....	.....
12000	1.0000	.....	.....	.....	......	.....	.....	.....	.....

Thus, in a problem where we wish to classify trinomial events according to their class probabilities such that the probability of correct classification is greater than or equal to P when the true ratio $\frac{q_k}{q_{k-1}}$ (k = 2,3) is greater than or equal to t, where P and t are specified by the experimenter, to determine the sample size that will guarantee this requirement, we consult one of the above two tables. If the values of P and t that have been specified are found in the table of the exact probabilities of correct classification we certainly use that table. Otherwise we make use of the table of the approximate probabilities of correct classification.

To illustrate the use of these tables consider a trinomial population in which we are interested in classifying the events such that

$$Pr[\text{correct classification} \mid T_{k,k-1} \geq 1.6] \geq 0.90 \quad (k=2,3)$$

We would like to know what sample size to take to achieve this goal. Table 2 indicates that with a sample of size 80 we will be sure of at least that probability. Thus we take a sample of size 80 and compute u_1, u_2, and u_3. Using the procedure described at the beginning of this chapter, we can now associate each F_i with one of q_i (i = 1, 2, 3). Since $q_1 \leq q_2 \leq q_3$ we have thus classified the events according to their class probabilities.

Table 3 compares the approximate and the exact probabilities of correct classification for N = 5(5)30.

TABLE 3

COMPARISON OF EXACT AND APPROXIMATE PROBABILITIES OF CORRECT CLASSIFICATION

N		t=1.1	t=1.2	t=1.3	t=1.4	t=1.6	t=1.8	t=2.0	t=2.5	t=3.0
5	Exact	0.203	0.239	0.274	0.307	0.369	0.422	0.468	0.555	0.717
	Approx.	0.202	0.240	0.275	0.311	0.374	0.428	0.472	0.554	0.608
10	Exact	0.219	0.273	0.333	0.376	0.468	0.545	0.603	0.705	0.905
	Approx.	0.221	0.275	0.327	0.377	0.470	0.544	0.605	0.703	0.754
15	Exact	0.232	0.300	0.368	0.441	0.540	0.628	0.686	0.792	0.908
	Approx	0.233	0.300	0.368	0.430	0.544	0.629	0.695	0.792	0.838
20	Exact	0.233	0.317	0.398	0.474	0.601	0.695	0.761	0.852	0.938
	Approx.	0.242	0.327	0.402	0.479	0.608	0.695	0.759	0.849	0.886
25	Exact	0.242	0.332	0.425	0.509	0.645	0.744	0.808	0.890	0.974
	Approx.	0.254	0.346	0.433	0.518	0.650	0.745	0.809	0.887	0.916
30	Exact	0.250	0.353	0.452	0.544	0.688	0.805	0.845	0.908	1.000
	Approx.	0.262	0.368	0.465	0.553	0.692	0.787	0.844	0.914	0.938

We would expect our approximation to improve with larger values of N. However, even though comparison of the exact and the approximate probabilities of correct classification indicates satisfactory results in general, it fails to show consistent improvement of the approximation with increasing sample size. This is probably due to mistakes in computation which time does not permit to check.

Use of the same approach for cases where we have more than three classes will lead to similar approximations, but since the variables W_i are not identically distributed, lack of appropriate tables of the multivariate normal distribution prevents the possibility of making similar tables for the probability of correct classification for given values of t and N.

CHAPTER IV

THE CONFIGURATION OF CLASS PROBABILITIES LEAST FAVORABLE TO CORRECT CLASSIFICATION

Let $q_1 \leq q_2 \leq q_3$ denote the ranked class probabilities in a trinomial population and let $Q = Q(q_1,q_2,q_3)$ be the probability of correct classification of the events while using the procedure described in the previous chapter. We will show that among all configurations of $\underset{\sim}{q} = (q_1,q_2,q_3)$ for which $T_{k,k-1} \geq t$ where $T_{k,k-1} = \frac{q_k}{q_{k-1}}$ $(k = 2,3)$ the one that minimizes the probability of correct classification is given by $T_{k,k-1} = t$; that is, $q_3 = tq_2$ and $q_2 = tq_1$. The method of proof to be used is analogous to that used in a paper by Kesten and Morse (1959) where they find that configuration of the q_i's which minimizes the probability of correct selection of the event with the highest class probability.

We shall first consider the simpler case of a binomial population. Using the same notation as above we wish to show that among all configurations of $\underset{\sim}{q} = (q_1,q_2)$ for which $T_{2,1} = \frac{q_2}{q_1} \geq t$, the one that minimizes the probability of correct classification is given by $q_2 = tq_1$.

For any fixed sample size N, the probability of correct classification is given by

$$\Sigma\, S(u_1, u_2) \binom{N}{u_1} q_1^{u_1} q_2^{u_2}$$

where $q_1 + q_2 = 1$, and the summation is taken over all vectors $\underset{\sim}{u}_N = (u_1, u_2)$ such that $u_1 + u_2 = N$ and $u_2 \geq u_1$, and

$$\begin{aligned} S(u_1, u_2) &= 1 \quad \text{if} \quad u_1 < u_2 \\ &= \tfrac{1}{2} \quad \text{if} \quad u_1 = u_2. \end{aligned}$$

Since the sum is empty for $u_1 > u_2$, the sum above represents $Pr[u_1 < u_2] + \frac{1}{2} Pr\,[u_1 = u_2]$.

For $u_1 \leq u_2$, define

$$f(u_2, q_2) = \sum_{x=o}^{u_2} \binom{N}{x} q_1^{x} q_2^{N-x}$$

which can be rewritten as the incomplete beta function (Rao, 1952),

$$f(u_2, q_2) = \frac{N!}{u_2!(N-u_2-1)!} \int_o^{q_2} x^{N-u_2-1} (1 - x)^{u_2} dx$$

Thus, $f(u_2, q_2)$ is an increasing function of q_2.

Now, the probability of correct classification can be written as

$$\tfrac{1}{2} f(u_2 - 1, q_2) + \tfrac{1}{2} f(u_2, q_2)$$

which is the sum of two terms each of which is an increasing function of q_2, and so the probability of correct classification is minimized by minimizing q_2. Thus in the binomial case, among all configurations of $\underset{\sim}{q} = (q_1, q_2)$ for which $\frac{q_2}{q_1} \geq t$, the configuration that minimizes q_2, namely $q_2 = tq_1$, will minimize the probability of correct classification. Applying the condition that $q_1 + q_2 = 1$ gives the result that $q_1 = \frac{1}{1+t}$ and $q_2 = \frac{t}{1+t}$.

In a trinomial population, the probability of correctly classifying the events can be written as

$$Q = \Sigma\ S(u_1, u_2, u_3) \quad \frac{N!}{u_1!\ u_2!\ u_3!}\ q_1^{u_1}\ q_2^{u_2}\ q_3^{u_3}$$

$$= \Sigma\ \frac{N!}{(u_1+u_3)!u_2!}\ (q_1 + q_3)^{u_1+u_3}\ q_2^{u_2}.\ \sum_{u_1=o}^{u_2} \frac{(u_1+u_3)!}{u_1!\ u_3!}\ \left(\frac{q_1}{q_1+q_3}\right)^{u_1}.$$

$$\left(\frac{q_3}{q_1+q_3}\right)^{u_3}.\ S(u_1, u_2, u_3)$$

where the outer summation is over all $[(u_1 + u_3), u_2]$ such that $u_1 + u_2 + u_3 = N$ and $u_1 \leq u_2 \leq u_3$, and

$$\begin{aligned} S(u_1, u_2, u_3) &= 1 \quad \text{if} \quad u_1 < u_2 < u_3 \\ &= \tfrac{1}{2} \quad \text{if} \quad u_1 < u_2 \quad \text{but} \quad u_2 = u_3 \text{ ; also} \\ &\qquad \text{if} \quad u_1 = u_2 \quad \text{and } u_2 < u_3 \\ &= \tfrac{1}{6} \quad \text{if} \quad u_1 = u_2 = u_3. \end{aligned}$$

If q_2 is held fixed, $q_1 + q_3$ is also fixed. Therefore Q can be considered as a function of only one variable. Considering it first as a function of q_1 we will show that this probability is a non-increasing function of q_1, and so the probability of correct classification is minimized by maximizing q_1. On the other hand, considering Q as a function of q_3, we find that it is an increasing function of q_3. Thus the probability of correct classification is minimized by minimizing q_3. Now, since we should have $\frac{q_2}{q_1} \geq t$ and $\frac{q_3}{q_2} \geq t$, we can maximize q_1 by putting $q_1 = \frac{q_2}{t}$ and we can minimize q_3 by putting $q_3 = q_2 t$. Since we must also have $q_1 + q_2 + q_3 = 1$, the three conditions:

$$q_1 = q_2/t$$

$$q_3 = q_2 t$$

$$q_1 + q_2 + q_3 = 1$$

give us the values of q_1, q_2, and q_3 in terms of t. The values of the q_i that we so find represent that configuration of $\underset{\sim}{q} = (q_1, q_2, q_3)$, among all configurations in which $\frac{q_k}{q_{k-1}} \geq t$ $(k = 2,3)$, which minimizes the probability of correct classification. This will be called the least favorable configuration.

It is sufficient to prove that the inside sum in the expression for Q above is a non-increasing function of q_1 and an increasing function of q_3 since the probability of correct classification consists of the sum of terms each of which is proved to be a non-increasing function of q_1 and an increasing function of q_3.

Consider the inner sum for fixed q_2 and some fixed value of u_2, say a. Thus $u_1 + u_3 = N - a$ is also fixed. Now, define

$$g(a) = \sum_{x=o}^{a} \frac{(N - a)!}{x!(N - a - x)!} (1 - r)^{N-a-x} r^x$$

where
$$r = \frac{q_1}{q_1 + q_3} .$$

This can be rewritten as the incomplete beta function (Rao, 1952):

$$g(a) = \frac{(N - a)!}{a! (N - 2a - 1)!} \int_o^{1-r} x^{N-2a-1} (1 - x)^a dx$$

This is an increasing function of $1 - r = \frac{q_3}{q_1+q_3}$ and so an increasing function of q_3, but it is a non-increasing function of r and therefore a non-increasing function of q_1.

(1) If the fixed value of u_2 is given by $a = \frac{N}{3}$, the inside sum in Q can be written as

$$Pr[u_1 < a] + \frac{1}{6} Pr[u_1 = a] = \frac{5}{6} g(a - 1) + \frac{1}{6} g(a).$$

(2) If $a < \frac{N}{3}$, we have

$$Pr[u_1 < a] + \frac{1}{2} Pr[u_1 = a] = \frac{1}{2} g(a - 1) + \frac{1}{2} g(a).$$

(3) If $a > \frac{N}{3}$, we have

$$Pr[u_1 < N-2a] + \frac{1}{2} Pr[u_1 = N - 2a] = \frac{1}{2} g(N - 2a - 1)+ \frac{1}{2} g(N-2a).$$

Therefore, if we use values of q_1 and q_3 that minimize g, we will be minimizing the inner sum and so each term of the sum involved in the probability of correct classification, as we vary u_2, will have been minimized.

This condition together with $q_1 + q_2 + q_3 = 1$ gives the result that the least favorable configuration is given by

$$q_1 = \frac{1}{1+t+t^2}, \quad q_2 = \frac{t}{1+t+t^2}, \quad \text{and} \quad q_3 = \frac{t^2}{1+t+t^2}.$$

It is this configuration of the q_i's that has been used in the preparation of the tables of the previous chapter that give the exact and the approximate probabilities of correct classification for various values of t and N.

REFERENCES

Books

Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons, Inc., 1958.

Cramér, H. *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946.

Mood, A.M. *Introduction to the Theory of Statistics*. New York; McGraw-Hill Book Company, Inc., 1950.

Rao, C.R. *Advanced Statistical Methods in Biometric Research.* New York: John Wiley and Sons, Inc., 1952.

Articles and Periodicals

Bechhofer, R., Elmaghraby, S., and Morse, N. "A Single-Sample Multiple-Decision Procedure for Selecting the Multinomial Event Which Has The Highest Probability.", *Ann. Math. Stat.*, 30 (1959), pp.102-119.

Elderton, E.M., *et al.* "Tables for Determining the Volume of a Bi-variate Normal Surface", *Biometrika*, 22 (1930), pp.1-35.

Greenwood, R.E., and Glasgow, M.O. "Distribution of Maximum and Minimum Frequencies in a Sample Drawn from a Multinomial Distribution", *Ann. Math. Stat.*, 21 (1950), pp.416-424.

Johnson, N.L. "An Approximation to the Multinomial Distribution: Some Properties and Applications", *Biometrika*, 47 (1960), pp.93-102.

Johnson, N.L., and Young, D.H. "Some Applications of Two Approximations to the Multinomial Distribution", *Biometrika*, 47 (1960), pp.463-469.

Kesten, H., and Morse, N. "A Property of the Multinomial Distribution", *Ann. Math. Stat.*, 30 (1959), pp.120-127.

Korzelka, R.M. "Approximate Upper Percentage Points for Extreme Values in Multinomial Sampling", *Ann. Math. Stat.*, 27 (1956), pp.507-512.

McCarthy, P.J. "Approximate Solutions for Means and Variances in a Certain Class of Box Problems", Ann. Math. Stat., 18 (1947), pp.349-383.

Rao, C.R. "Maximum Likelihood Extimation for the Multinomial Distribution", Sankhya, 18 (1957-58), pp.139-148.

________. "Maximum Likelihood Estimation for the Multinomial Distribution with Infinite Number of Cells", Sankhya, 20 (1958), pp.211-218.

Weiss, L. "A Sequential Test of the Equality of Probabilities in a Multinomial Distribution", J. Am. Stat. Assoc., 57 (1962), pp.769-774.

Wishart, J. "Cumulants of Multivariate Multinomial Distributions", Biometrika, 36 (1949), pp.47-58.