The Annals of Mathematical Statistics
1972, Vol. 43, No. 1, 142-152

ESTIMATING THE SIZE OF A MULTINOMIAL POPULATION

BY LALITHA SANATHANAN

University of Illinois at Chicago Circle

1. Introduction and summary. This paper deals with the problem of estimating the number of trials of a multinomial distribution, from an incomplete observation of the cell totals, under constraints on the cell probabilities. More specifically let $(n_1, \cdots, n_k)$ be distributed according to the multinomial law $M(N; p_1, \cdots, p_k)$ where N is the number of trials and the p_i's are the cell probabilities, $\sum_{i=1}^{k} p_i$ being equal to 1. Suppose that only a proper subset of $(n_1, \cdots, n_k)$ is observable, that $N, p_1, \cdots, p_k$ are unknown and that N is to be estimated. Without loss of generality, $(n_1, \cdots, n_{l-1})$, $l \leqq k$ may be taken to be the observable random vector. For fixed N, $(n_1, \cdots, n_{l-1}, N - n)$ has the multinomial distribution $M(N; p_1, \cdots, p_l)$ where n denotes $\sum_{i=1}^{l-1} n_i$ and p_l denotes $1 - \sum_{i=1}^{l-1} p_i$. If the parameter space is such that N can take any non-negative integral value and each p_i can take any value between 0 and 1, such that $\sum_{i=1}^{l-1} p_i < 1$ then, clearly, the only inference one can make about N is that $N > n$. In specific situations, it might, however, be possible to postulate constraints of the type

$$p_i = f_i(\theta), \qquad i = 1, \cdots, l \tag{1.1}$$

where $\theta = (\theta_1, \cdots, \theta_r)$ is a vector of r independent parameters and f_i are known functions. This may lead to estimability of N. The problem of estimating N in such a situation is studied here.

The present investigation is motivated by the following problem. Experiments in particle physics often involve visual scanning of film containing photographs of particles (occurring, for instance, inside a bubble chamber). The scanning is done with a view to counting the number N of particles of a predetermined type (these particles will be referred to as events). But owing to poor visibility caused by such characteristics as low momentum, the distribution and configuration of nearby track patterns, etc., some events are likely to be missed during the scanning process. The question, then, is: How does one get an estimate of N? The usual procedure of estimating N is as follows. Film containing the N (unknown) events is scanned separately by w scanners (ordered in some specific way) using the same instructions. For each event E let a w-vector $Z(E)$ be defined, such that the jth component Z_j of $Z(E)$ is 1 if E is detected by the jth scanner and is 0 otherwise. Let $\mathcal{S}$ be the set of 2^w w-vectors of 1's and 0's and let I_0 by the vector of 0's. Let x_I be the number

Received September 18, 1969; revised June 3, 1971.

of events E whose $Z(E) = I$. For $I \in \mathcal{I} - \{I_0\}$, the x_I's are observed. A probability model is assumed for the results of the scanning process. That is, it is assumed that there is a probability p_I that $Z(E)$ assumes the value I and that these p_I's are constrained by equations of the type (1.1) (These constraints vary according to the assumptions made about the scanners and events, thus giving rise to different models. An example of $p_I(\theta)$ would be $E(v^{\Sigma_{j=1}^{w} I_j}(1-v)^{w-\Sigma_{j=1}^{w} I_j})$ where I_j is the jth component of I and expectation is taken with respect to the two-parameter beta density for v. This is the result of assuming that all scanners are equally efficient in detecting events, that the probability v that an event is seen by any scanner is a random variable and that the results of the different scans are locally independent. For a discussion of various models, see Sanathanan (1969), Chapter III. N is then estimated using the observed x_I's and the constraints on the p_I's, provided certain conditions (e.g., the minimum number of scans required) are met.

The following formulation of the problem of estimating N, however, leads to some systematic study including a development of the relevant asymptotic distribution theory for the estimators. The $Z(E)$'s may be regarded as realizations of N independent identically distributed random variables whose common distribution is discrete with probabilities p_I at I (In particle counting problems, it is usually true that the particles of interest are sparsely distributed throughout the film on account of their Poisson distribution with low intensity. Thus in spite of the factors affecting their visibility outlined earlier, the events can be assumed to be independent.). The joint distribution of the x_I's is, then, multinomial $M(N; p_I, I \in \mathcal{I})$. The problem of estimating N is now in the form stated at the beginning of this section. Since the estimate depends on the constraints provided for the p_I's, it is important to test the "fit" on the model selected. The conditional distribution of the x_I's $(I \neq I_0)$ given x is multinomial $M(x; p_I/p(I \neq I_0))$ where x is defined as $\sum_{I \neq I_0} x_I$ and p as $\sum_{I \neq I_0} p_I$. The corresponding χ^2 goodness of fit test may therefore be used to test the adequacy of a model in question.

Various estimators of N are considered in this paper and among them is, of course, the maximum likelihood estimator of N. Asymptotic theory for maximum likelihood estimation of the parameters of a multinomial distribution has been developed before for the case where N is known but not for the case where N is unknown. Asymptotic theory related to the latter case is developed is Section 4. The result on the asymptotic joint distribution of the relevant maximum likelihood estimators is stated in Theorem 2.

A second method of estimation considered is that of maximizing the likelihood based on the conditional probability of observing $(n_1, \cdots, n_{l-1})$, given n. This method is called the conditional maximum likelihood (C.M.L.) method. The C.M.L. estimator of N is shown (Theorem 2) to be asymptotically

equivalent to the maximum likelihood estimator. Section 5 contains an extension of these results to the situation involving several multinomial distributions. This situation arises in the particle scanning context when the detected events are classified into groups based on some factor like momentum which is related to visibility of an event, and a separate scanning record is available for each group.

A third method of estimation considered is that of equating certain linear combinations of the cell totals (presumably chosen on the basis of some criterion) to their respective expected values. Asymptotic theory for this method is given in Section 6. This discussion is motivated by a particular case which is applicable to some models in the particle scanning problem, using a criterion based on the method of moments for the unobservable random variable, given by the number of scanners detecting an event (Discussion of the particular case can be found in Sanathanan (1969) Chapter III.).

In the next section we give some definitions and a preliminary lemma.

2. Likelihood. The observation of $(n_1, \cdots, n_{l-1})$ yields the likelihood function

$$L(N; \theta) = (N!/(n_1! \cdots n_l! (N - n)!))(p_1(\theta))^{n_1} \cdots (p_{l-1}(\theta))^{n_{l-1}}(p_l(\theta))^{N-n}$$

where $p_i(\theta) = f_i(\theta)$ of (1.1), $i = 1, \cdots, l$.
$L(N; \theta)$ may also be written as $L(N; \theta) = L_1(N; p_l(\theta))L_2(\theta)$ where

$$L_1(N; p_l(\theta)) = (N!/(n! (N - n)!))(1 - p_l(\theta))^n(p_l(\theta))^{N-n}, \tag{2.1}$$

$$L_2(\theta) = (n_1! /(n! \cdots n_{l-1}!))(q_1(\theta))^{n_1} \cdots (q_{l-1}(\theta))^{n_{l-1}}$$

with $q_i(\theta) = p_i(\theta)/(1 - p_l(\theta))$, $i = 1, \cdots, l - 1$.

It is easily seen that L_1 is the likelihood based on the probability of n and hence L_2 is the likelihood based on the conditional probability of $(n_1, \cdots, n_{l-1})$ given n. The following lemma is known (see e.g. Chapman (1951)).

LEMMA 1. *For any given* p, $\hat{N} = [n/(1 - p_l)]$ (*greatest integer* $\leqq n/(1 - p_l)$) *maximizes* $L_1(N; p_l)$, *where* $L_1(N; p_l)$ *is defined in* (2.1). *If* $1 - p_l = n/N'$ *for some integer* N', *then* $\hat{N}$ *and* $\hat{N} - 1$ *both maximize* $L_1(N; p_l)$. *Otherwise* $\hat{N}$ *is the unique maximum.*

3. Maximum and conditional maximum likelihood estimates of N. Two estimates of N arise naturally. The first is, of course, the maximum likelihood estimate to be denoted as $\hat{N}_U$ and defined by the condition that there exists a value $\hat{\theta}_U$ of θ such that $(N, \theta) = (\hat{N}_U, \hat{\theta}_U)$ maximizes $L(N; \theta)$ over all admissible values of $(N; \theta)$. The second estimate which will be called the conditional maximum likelihood estimate to be denoted by $\hat{N}_C$ is defined by the condition that $N = \hat{N}_C$ maximizes $L_1(N; \hat{p}_C)$ where $\hat{p}_C = p_l(\hat{\theta}_C)$ and $\hat{\theta}_C$ is the value of θ which maximizes $L_2(\theta)$. That is, we first make a conditional inference about p_l based solely on $L_2(\theta)$ and then infer about N on the basis of $L_1(N; p_l)$ with

p_l replaced by its estimate. By Lemma 1, except when $p = n/N'$ for some integer N', $\hat{N}_C = [n/(1 - \hat{p}_C)]$. In the exceptional case, let us define $\hat{N}_C$ to be $(n/(1 - \hat{p}_C))$ and thus $\hat{N}_C$ is well defined. The problem then is to get $\hat{\theta}_C$ from which $\hat{p}_C$ may be calculated. But L_2 involves only θ and not N and hence maximizing L_2 is simpler than maximizing L. This is especially helpful when we have a combined likelihood involving several multinomial populations.

The appropriateness of inference based on conditional distributions is a topic which has been discussed widely in the literature (Fisher (1956), Cox (1958a), Cox (1958b), Bartlett (1956), Welch (1956)). No special effort is therefore made here to justify the estimates $\hat{\theta}_C$ and $\hat{N}_C$.

Some properties of $\hat{N}_C$ and $\hat{N}_U$ are discussed in the next section. In particular, their asymptotic distributions are derived under certain regularity conditions, and are shown to be the same. Thus they are asymptotically equivalent. This equivalence can also be seen heuristically as follows.

Assume that the partial derivatives $\partial p_i/\partial\theta_j$ exist for all i and j. Since $\theta = \hat{\theta}_U$ maximizes $L(\hat{N}_U; \theta)$, $\partial \log L/\partial\theta_j$ at $(\hat{N}_U, \hat{\theta}_U) = 0$. Therefore,

$$(3.1) \qquad -(n/(1 - \hat{p}_U) - (\hat{N}_U - n)/\hat{p}_U)p_{l,j}(\hat{\theta}_U) + L_{2,j}(\hat{\theta}_U) = 0\,, \quad j = 1, \cdots, r$$

where $\hat{p}_U = p_l(\hat{\theta}_U)$, $p_{l,j}(\hat{\theta}_U)$ denotes $\partial p_i/\partial\theta_j$ at $\hat{\theta}_U$ and $L_{2,j}(\hat{\theta}_U)$ denotes $\partial \log L_2/\partial\theta_j$ at $\hat{\theta}_U$. If $\hat{N}_U = [n/(1 - \hat{p}_U)]$ is replaced by $n/(1 - \hat{p}_U)$ in (3.1) the first term becomes 0 so that we have $L_{2,j}(\hat{\theta}_U) = 0$, $j = 1, \cdots, r$ yielding the same procedure as that used in getting $\hat{N}_C$.

4. Asymptotic distributions of $\hat{N}_C$ and $\hat{N}_U$. We now proceed to derive the asymptotic distributions of $\hat{N}_C$ and $\hat{N}_U$. The following assumption is made throughout.

A1. At every admissible value of θ, the functions $p_i(\theta)$ admit continuous first-order partial derivatives.

The notation $\longrightarrow_{\text{a.s.}}$ is used to denote almost sure convergence, $\longrightarrow_p$ to denote convergence in probability, $\longrightarrow_l$ to denote convergence in law, and $\mathcal{N}(0, \Sigma)$ to denote a normal random vector with mean vector 0 and covariance matrix Σ.

Let $p_i(\theta)$, $L(N; \theta)$ and $L_2(\theta)$ be denoted by p_i^0, L^0 and L_2^0 respectively when $(N, \theta) = (N_0, \theta_0)$ and by $\hat{p}_i$, $\hat{L}$ and $\hat{L}_2$ respectively when $(N, \theta) = (\hat{N}, \hat{\theta})$. Similarly let the partial derivatives of $p_i(\theta)$, $\log L(N; \theta)$ and $\log L_2(\theta)$ with respect to θ_j be denoted by $p_{i,j}^0$, L_j^0, $L_{2,j}^0$ respectively when $(N, \theta) = (N_0, \theta_0)$ and by $\hat{P}_{i,j}$, $\hat{L}_j$ and $\hat{L}_{2,j}$ respectively when $(N, \theta) = (\hat{N}, \hat{\theta})$.

THEOREM 1. *Let N_0 be the true value of N and $\theta_0 = (\theta_{01}, \cdots, \theta_{0r})$ be the true value of θ. Let $\hat{N}$ and $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_r)$ be the estimates of N_0 and θ_0 respectively such that as $N_0 \longrightarrow \infty$,*

(i) $\hat{\theta} \longrightarrow_{\text{a.s.}} \theta_0$;

(ii) $N_0^{-\frac{1}{2}}(\hat{N} - n/(1 - \hat{p}_l)) \longrightarrow_{\text{a.s.}} 0$;

(iii) $N_0^{-\frac{1}{2}}\hat{L}_j \to_{\text{a.s.}} 0\,, \qquad j = 1, \cdots, r\,.$

Let Σ^{-1} *be the* $(r+1) \times (r+1)$ *matrix given by*

$$\Sigma^{-1} = \begin{bmatrix} A & \mathbf{a}_0' \\ \mathbf{a}_0 & \mathbf{a}_{00} \end{bmatrix}$$

where $A = (a_{ij})$ *defined by* $a_{ij} = \sum_{s=1}^{l} (p_s)^{-1} p_{s,i}^0 p_{s,j}^0$, $i, j = 1, \cdots, r$; $\mathbf{a}_0 = (a_{10}, \cdots, a_{r0})$ *defined by* $a_{i0} = -(p_l^0)^{-1} p_{l,i}^0$, $i = 1, \cdots, r$, *and* $a_{00} = (1 - p_l^0)/(p_l^0)$. *Then,* $(N_0^{\frac{1}{2}}(\theta - \theta_0), N_0^{-\frac{1}{2}}(\hat{N} - N_0))$ *is asymptotically* $\mathcal{N}(0, \Sigma)$.

A is the usual information matrix for the multinomial distribution $M(N; p_1(\theta), \cdots, p_l(\theta))$ when N is known. (See Rao (1965) pp. 295–299 for a derivation of A.) That Σ is nonsingular is assumed implicitly in the above theorem.

PROOF. Throughout the proof of this theorem, unless otherwise specified, the subscript i will range from 1 to l and the subscripts j and m from 1 to r. Let n_l denote $N_0 - n$. We, then, have

$$N_0^{-\frac{1}{2}} \textstyle\sum_i (\hat{p}_i)^{-1} n_i \hat{p}_{i,j} = -N_0^{-\frac{1}{2}} (\hat{p}_l)^{-1} (\hat{N} - N_0) \hat{p}_{l,j} + N_0^{-\frac{1}{2}} \hat{L}_j\,.$$

Since $\sum_i p_{i,j} = 0$, it follows that

$$\text{(4.1)} \quad N_0^{-\frac{1}{2}} \textstyle\sum_i (\hat{p}_i)^{-1} (n_i - N_0 p_i^0) \hat{p}_{i,j} = N_0^{\frac{1}{2}} \sum_i (\hat{p}_i)^{-1} (\hat{p}_i - p_i^0) \hat{p}_{i,j} - N_0^{-\frac{1}{2}} (\hat{p}_l)^{-1} (\hat{N} - N_0) \hat{p}_{l,j} + N_0^{-\frac{1}{2}} \hat{L}_j\,.$$

If we use the mean value representation

$$\hat{p}_i - p_i^0 = \textstyle\sum_m (\hat{\theta}_m - \theta_{0m}) p_{i,m}(\theta^i)\,, \qquad \theta_m{}^i \in (\hat{\theta}_m, \theta_{0m})$$

(4.1) can be written as

$$\text{(4.2)} \qquad \textstyle\sum_m b_{j,m} N_0^{\frac{1}{2}} (\hat{\theta}_m - \theta_{0m}) + b_{j,r+1} N_0^{-\frac{1}{2}} (\hat{N} - N_0) = N_0^{-\frac{1}{2}} \sum_i d_{ij} y_i - N_0^{-\frac{1}{2}} \hat{L}_j$$

where for $j = 1, \cdots, r$ and for m from 1 to $r+1$ $b_{j,m} \to_{\text{a.s.}} \sigma^{j,m}$, $\sigma^{j,m}$ being the (j, m)th element of Σ^{-1},

$$\text{(4.3)} \qquad d_{ij} \to_{\text{a.s.}} (p_i^0)^{-1} p_{i,j}^0$$

and

$$y_i = n_i - N_0 p_i^0\,.$$

Since $N_0^{-\frac{1}{2}} y_i$ has a limiting distribution

$$\text{(4.4)} \qquad N_0^{-\frac{1}{2}} \textstyle\sum_i d_{ij} y_i - z_j \to_p 0$$

where

$$\text{(4.5)} \qquad z_j = N_0^{-\frac{1}{2}} \textstyle\sum_i (p_i^0)^{-1} p_{i,j}^0 y_i = N_0^{-\frac{1}{2}} \sum_i (p_i^0)^{-1} p_{i,j}^0 n_i\,.$$

Thus using (iii) in the statement of the theorem,

$$\text{(4.6)} \qquad \textstyle\sum_m b_{j,m} N_0^{\frac{1}{2}} (\hat{\theta}_m - \theta_{0m}) + b_{j,r+1} N_0^{-\frac{1}{2}} (\hat{N} - N_0) - z_j \to_p 0\,.$$

Now consider the following equation

$$(4.7)\qquad N_0^{-\frac{1}{2}}(\hat{N}(1-\hat{p}_l)-n) = N_0^{-\frac{1}{2}}(\hat{N}-N_0)(1-\hat{p}_l) - N_0^{\frac{1}{2}}(\hat{p}_l - p_l^0) - N_0^{-\frac{1}{2}}(n - N_0(1-p_l^0))\,.$$

Using the mean value representation for $\hat{p}_l - p_l^0$, dividing (4.7) by p_l^0 and using condition (ii), we have

$$(4.8)\qquad \sum_m b_{r+1,m}N_0^{\frac{1}{2}}(\hat{\theta}_m - \theta_{0m}) + b_{r+1,r+1}N_0^{-\frac{1}{2}}(\hat{N}-N_0) - z_{r+1} \to_p 0$$

where

$b_{r+1,m} \to_{\text{a.s.}} \sigma^{r+1,m}$, $m = 1, \cdots, r+1$ and $z_{r+1} = N_0^{-\frac{1}{2}}(p_l^0)^{-1}(n - N_0(1-p_l^0))$. Let $U' = (N_0^{\frac{1}{2}}(\theta - \theta_0), N_0^{-\frac{1}{2}}(\hat{N}-N_0))$ and $Z' = (z_1, \cdots, z_{r+1})$. Then by inverting the relations (4.6) and (4.8), we have

$$(4.9)\qquad U - \Sigma Z \to_p 0\,.$$

Next, we show that $Z \to_l \mathcal{N}(0, \Sigma^{-1})$.

For $m = 1, \cdots, N_0$ let $V_m = (V_{m,1}, \cdots, V_{m,r+1})$ be the random vector such that

(i) when the mth trial results in the ith category, i ranging from 1 to $l-1$, $V_{m,j}$ takes the value $(p_i^0)^{-1}p_{i,j}^0$, $j = 1, \cdots, r$ and $V_{m,r+1}$ takes the value 1.

(ii) when the mth trial results in the lth category, $V_{m,j}$ takes the value $(p_l^0)^{-1}p_{l,j}^0$, $j = 1, \cdots, r$ and $V_{m,r+1}$ takes the value $-(p_l^0)^{-1}(1-p_l^0)$

Then

$$(4.10)\qquad Z' = N_0^{-\frac{1}{2}}\sum_{m=1}^{N_0} V_m\,.$$

From the definition of V_m, it may be easily deduced that each V_m has 0 as its mean vector and Σ^{-1} as its covariance matrix. Also the V_m's are identically distributed. Therefore, by the Central Limit Theorem

$$(4.11)\qquad Z \to_l \mathcal{N}(0, \Sigma^{-1})\,.$$

Thus by (4.9) $U \to_l \mathcal{N}(0, \Sigma)$ and thus the proof of Theorem 1 is complete.

THEOREM 2. *Assume* A1 *and*

A2. *Given a* $\delta > 0$, *it is possible to find an* $\varepsilon > 0$ *such that*

$$\inf_{|\theta_* - \theta_0| > \delta} \sum_{i=1}^{l-1} q_i(\theta_0)\log(q_i(\theta_0)/q_i(\theta_*)) > \varepsilon$$

where $q_i(\theta) = p_i(\theta)/(1-p_l(\theta))$, $i = 1, \cdots, l-1$.

Then

I. $(\hat{N}_C/N_0, \hat{\theta}_C, \hat{p}_C) \to_{\text{a.s.}} (1, \theta_0, p_l^0)$ *as* $N_0 \to \infty$;

II. $(\hat{N}_U/N_0, \hat{\theta}_U, \hat{p}_U) \to_{\text{a.s.}} (1, \theta_0, p_l^0)$ *as* $N_0 \to \infty$;

III. $(N_0^{\frac{1}{2}}(\hat{\theta}_C - \theta_0), N_0^{-\frac{1}{2}}(\hat{N}_C - N_0))$ *and* $(N_0^{\frac{1}{2}}(\hat{\theta}_U - \theta_0), N_0^{-\frac{1}{2}}(\hat{N}_U - N_0))$ *are both asymptotically* $\mathcal{N}(0, \Sigma)$ *where* Σ *is as defined in Theorem* 1.

PROOF. By the law of large numbers,

$$n/N_0 \to_{\text{a.s.}} 1 - p_l^0 \quad \text{as} \quad N_0 \to \infty \tag{4.12}$$

and so $n \to_{\text{a.s.}} \infty$ as $N_0 \to \infty$.

But $\hat{\theta}_C$ is the maximum likelihood estimate of θ_0 based on an observation from $M(n; q_1(\theta_0), \cdots, q_{l-1}(\theta_0))$ and so by A2, $\hat{\theta}_C \to_{\text{a.s.}} \theta_0$ as $n \to \infty$ (for proof see pages 295–296 in Rao (1965)) and hence by (4.12)

$$\hat{\theta}_C \to_{\text{a.s.}} \theta_0 \quad \text{as} \quad N_0 \to \infty \,. \tag{4.13}$$

We also have $\hat{p}_C \to p_l^0$ as $N_0 \to \infty$ by the continuity of $p_l(\theta)$. $\hat{N}_C = [n/(1 - \hat{p}_C)]$ implies $(1/N_0)(\hat{N}_C - n/(1 - \hat{p}_C)) \to_{\text{a.s.}} 0$ as $N_0 \to \infty$ and so $\hat{N}_C/N_0 \to_{\text{a.s.}} 1$ as $N_0 \to \infty$. Thus statement I of the theorem is true. Statement II can be proved as follows:

It is enough to prove that $\sum_{i=1}^{l-1}(n_i/n) \log (q_i(\hat{\theta}_U)/(n_i/n)) \to_{\text{a.s.}} 0$ since this together with A2 would imply $\hat{\theta}_U \to_{\text{a.s.}} \theta_0$ (for proof see pages 292, 293 and 295 in Rao (1965)).

Consider L_1 and L_2 as defined in (2.1). Log $L_1 = 0$ at the point $(N, p_l) = (n, 0)$ and is $\leqq 0$ at all other points. Since $\hat{\theta}$ maximizes $\log [L(\hat{N}_U; \theta)]/n$, we have

$$\begin{aligned} &\sum_{i=1}^{l-1} (n_i/n) \log q_i(\hat{\theta}_U) + \text{a negative number} \\ &\qquad \geqq \sup_\theta [\sum_{i=1}^{l-1} (n_i/n) \log q_i(\theta) + \log [L_1(N; p_l(\theta))]/n] \\ &\qquad \geqq \sum_{i=1}^{l-1} (n_i/n) \log q_i(\theta_0) + \log [L_1(N_0; p_l(\theta_0))]/n \,. \end{aligned}$$

Also, by an inequality in information theory

$$\sum_{i=1}^{l-1} (n_i/n) \log q_i(\hat{\theta}_U) \leqq \sum_{i=1}^{l-1} (n_i/n) \log (n_i/n) \,.$$

Combining the above two inequalities we have

$$\begin{aligned} 0 &\geqq \sum_{i=1}^{l-1} (n_i/n) \log (q_i(\hat{\theta}_U)/(n_i/n)) \\ &\geqq \sum_{i=1}^{l-1} (n_i/n) \log (q_i(\theta_0)/(n_i/n)) + \log [L_1(N_0; p_l(\theta_0))]/n \,. \end{aligned}$$

As $N_0 \to \infty$, $n_i/n \to_{\text{a.s.}} q_i(\theta_0)$. Also, using the normal approximation to the binomial probability $L_1(N_0; p_l(\theta_0))$ and (4.12) it is seen that $\log [L_1(N_0; p_l(\theta_0))]/n \to_{\text{a.s.}} 0$. Hence the result.

We will now show that both $(\hat{N}_C, \hat{\theta}_C)$ and $(\hat{N}_U, \hat{\theta}_U)$ satisfy conditions (ii) and (iii) of Theorem 1. $\hat{N}_C = [(n/(1 - \hat{p}_C)]$ implies $|\hat{N}_C - n/(1 - \hat{p}_C)| < 1$ and therefore condition (ii) of Theorem 1 is satisfied by $(\hat{N}_C, \hat{\theta}_C)$ and similarly by $(\hat{N}_U, \hat{\theta}_U)$. That condition (iii) of Theorem 1 is satisfied by $(\hat{N}_U, \hat{\theta}_U)$ is obvious from the definition of $(\hat{N}_U, \hat{\theta}_U)$. Coming to $(\hat{N}_C, \hat{\theta}_C)$, by definition $\theta = \hat{\theta}_C$ maximizes $L_2(\theta)$ and since the partial derivatives are assumed to exist, we have

$$L_{2,j}(\hat{\theta}_C) = 0 \,, \qquad j = 1, \cdots, r \,.$$

Therefore

$$N_0^{-\frac{1}{2}}L_j(\hat{N}_C; \hat{\theta}_C) = N_0^{-\frac{1}{2}}L_{1,j}(\hat{N}_C; \hat{p}_C)$$
$$= N_0^{-\frac{1}{2}}p_{l,j}(\hat{\theta}_C)((\hat{N}_C - n)/\hat{p}_C - n/(1 - \hat{p}_C)), \qquad i = 1, \cdots, r.$$

By (4.13) and A1, $p_{l,j}(\hat{\theta}_C) \to_{\text{a.s.}} p_{l,j}^0$ as $N_0 \to \infty$. Also since $\hat{p}_C \to_{\text{a.s.}} p_l^0$, $(\hat{N}_C - n)/\hat{p}_C - n/(1 - \hat{p}_C)$ is almost surely bounded in the limit. Therefore $N_0^{-\frac{1}{2}}L_j(\hat{N}_C; \hat{\theta}_C) \to_{\text{a.s.}} 0$. Thus condition (iii) of Theorem 1 is satisfied by $(\hat{N}_C; \hat{\theta}_C)$ and this completes the proof of Theorem 2.

In finite samples, $\hat{N}_U$ need not be equal to $\hat{N}_C$. The following inequality may, however, be derived.

THEOREM 3. $\hat{N}_U \leqq \hat{N}_C$.

Proof is omitted.

Theory developed in this section is extended in the next section to the case of independent observations from several multinomial populations some of whose parameters are interrelated.

5. Case of s populations. Suppose we have observations $0_t = (n_{t1}, \cdots, n_{tl-1})$, $t = 1, \cdots, s$ such that 0_t's are independent and 0_t is the vector of the first $l - 1$ cell totals from a multinomial population $M_t(N_t; p_{t1}, \cdots, p_{tl})$. Assume as before that N_t's are unknown and that there are independent parameters $\tau_1, \cdots, \tau_v$ such that $p_{ti}(\tau) = f_{ti}(\tau)$, $i = 1, \cdots, l$; $t = 1, \cdots, s$; where $\tau = (\tau_1, \cdots, \tau_v)$. Define n_t to be $\sum_{i=1}^{l-1} n_{ti}$ and p_{tl} to be $1 - \sum_{i=1}^{l-1} p_{ti}$. Let N denote the vector $(N_1, N_2, \cdots, N_s)$. Let $N_0 = (N_{01}, \cdots, N_{0s})$ and $\tau_0 = (\tau_{01}, \cdots, \tau_{0v})$ be the true values of N and τ respectively. The unconditional and conditional maximum likelihood estimates of $(N_0; \tau_0)$ can be defined analogously as in the single population case.

Theorems 1 and 2 may be extended to the case of independent observations from s populations. Some assumptions about the relative sizes of the populations as they become large are necessary. More specifically, let $N_{0t} \to \infty$ in such a way that $\lim_{N_T \to \infty} N_{0t}/N_T = c_t$, $t = 1, \cdots, s$ where c_t is any real number between 0 and 1 such that $\sum_{t=1}^{s} c_t = 1$ and N_T is defined as $\sum_{t=1}^{s} N_{0t}$. Then Theorem 1 may be extended as follows.

THEOREM 4. *Let the $p_{ti}(\tau)$'s admit first order partial derivatives which are continuous at every admissible value of τ. Let $\hat{N} = (\hat{N}_1, \cdots, \hat{N}_s)$ and $\hat{\tau} = (\hat{\tau}_1, \cdots, \hat{\tau}_v)$ be estimates of N_0 and τ_0 respectively such that*

(i) $\hat{\tau} \to_{\text{a.s.}} \tau_0$
(ii) $< N_{0t}^{-\frac{1}{2}}(\hat{N}_t - n_t/(1 - \hat{p}_{tl})) > \to_{\text{a.s.}} 0$, *and*
(iii) $N_T^{-\frac{1}{2}}\hat{L}_j \to_{\text{a.s.}} 0$, $j = 1, \cdots, v$

where the symbol $\langle e_t \rangle$ denotes the vector $(e_1, \cdots, e_s)$ and the notations for partial derivatives are obvious extensions of those used earlier.

Then

$$(N_T^{\frac{1}{2}}(\hat{\tau} - \tau_0),\ N_{01}^{-\frac{1}{2}}(\hat{N}_1 - N_{01}),\ \cdots,\ N_{0s}^{-\frac{1}{2}}(\hat{N}_s - N_{0s}))$$

is asymptotically $\mathcal{N}(0, \bar{\Sigma})$ *where* $\bar{\Sigma}^{-1} = (\sigma^{ij})$ *is given by*

$$\begin{aligned}
\sigma^{j,m} &= \textstyle\sum_{t=1}^{s} c_t \sum_{i=1}^{l} (p_{ti}^0)^{-1} p_{ti,j}^0 p_{ti,m}^0\,, && j = 1, \cdots, v;\ m = 1, \cdots, v \\
\sigma^{j,v+t} &= -(p_{tl}^0)^{-1} c_t^{\frac{1}{2}} p_{tl,j}^0\,, && j = 1, \cdots, v;\ t = 1, \cdots, s \\
\sigma^{v+t,v+u} &= (p_{tl}^0)^{-1} \delta_{t,u}(1 - p_{tl}^0)\,, && t = 1, \cdots, s;\ u = 1, \cdots, s
\end{aligned}$$

where $\delta_{t,u} = 1$ *if* $t = u$, $\delta_{t,u} = 0$ *otherwise*. Proof is analogous to that of Theorem 1 (referred to as P1 henceforth) and therefore details are omitted. Condition (iii) of Theorem 1 is replaced here by the condition $N_T^{-\frac{1}{2}}\hat{L}_j \to_{\text{a.s.}} 0$, $j = 1, \cdots, v$. But

$$N_T^{-\frac{1}{2}}\hat{L}_j = \textstyle\sum_{t=1}^{s} N_T^{-\frac{1}{2}}\hat{L}_{t,j} = \sum_{t=1}^{s} (N_{0t}/N_T)^{\frac{1}{2}} N_{0t}^{-\frac{1}{2}}\hat{L}_{t,j}\,.$$

$\hat{L}_{t,j}$ here corresponds to $\hat{L}_j$ of P1. Hence using exactly the same arguments, we get

$$\begin{aligned}
&\textstyle\sum_{t=1}^{s} N_{0t}/N_T \sum_{m=1}^{v} b_{t,j,m} N_T^{\frac{1}{2}}(\hat{\tau}_m - \tau_{0m}) \\
&\qquad + \textstyle\sum_{t=1}^{s} (N_{0t}/N_T)^{\frac{1}{2}} b_{j,v+t}(\hat{N}_t - N_{0t}) N_{0t}^{-\frac{1}{2}} - z_j \to_p 0
\end{aligned}$$

where $\sum_{t=1}^{s} c_t b_{t,j,m} \to_{\text{a.s.}} \sigma^{j,m}, j = 1, \cdots, v;\ m = 1, \cdots, v;$

$$\begin{aligned}
c_t^{\frac{1}{2}} b_{j,v+t} &\to_{\text{a.s.}} \sigma^{j,v+t}\,, && j = 1, \cdots, v;\ t = 1, \cdots, s \text{ and} \\
z_j &= \textstyle\sum_{t=1}^{s} c_t^{\frac{1}{2}} N_{0t}^{-\frac{1}{2}} \sum_{i=1}^{l} (p_{ti}^0)^{-1} p_{ti,j}^0 n_{ti}\,, && j = 1, \cdots, v.
\end{aligned}$$

Condition (ii) of Theorem 1 is replaced here by conditions of the same type for the s populations. Hence, just as in P1, we have

$$\textstyle\sum_{m=1}^{v} (N_{0t}/N_T)^{\frac{1}{2}} b_{v+t,m} N_T^{\frac{1}{2}}(\hat{\tau}_m - \tau_{0m}) + b_{v+t,v+t} N_{0t}^{-\frac{1}{2}}(\hat{N}_t - N_{0t}) - z_{v+t} \to_p 0\,, \quad t = 1, \cdots, s$$

where

$$\begin{aligned}
c_t^{\frac{1}{2}} b_{v+t,m} &\to_{\text{a.s.}} \sigma^{v+t,m}\,, && m = 1, \cdots, v;\ t = 1, \cdots, s; \\
b_{v+t,v+t} &\to_{\text{a.s.}} \sigma^{v+t,v+t}\,, && t = 1, \cdots, s
\end{aligned}$$

and

$$z_{v+t} = (p_{tl}^0)^{-1} N_{0t}^{\frac{1}{2}}(n_t/N_{0t} - (1 - p_{tl}^0))\,, \qquad t = 1, \cdots, s.$$

Further arguments as in P1, with the additional fact that $N_{0t}/N_T \to c_t$ as $N_T \to \infty$, $t = 1, \cdots, s$ yield the result stated in Theorem 4.

Theorem 2 may also be extended analogously. In the extension, assumptions A1 and A2 are each to be replaced by assumptions of the same type separately for the s populations.

6. Other methods of estimation. Among other possible methods of estimation, one that is sometimes computationally convenient is the following. Let r be the number of unknown θ's. Based on some criterion, $(r + 1)$ different random

variables are formed by taking linear combinations of $n_1, \cdots, n_{l-1}$ and then equated to their expected values. Let

$$X_i = \sum_{j=1}^{l-1} h_{i,j} n_j , \qquad i = 1, \cdots, r+1 \tag{6.1}$$

where $h_{i,j}$'s are given constants and let $N_0 \Pi_i = E(X_i)$. The equations are given by

$$N \Pi_i (\theta) = X_i , \qquad i = 1, \cdots, r+1. \tag{6.2}$$

(6.2) is a set of $r+1$ equations in $r+1$ unknowns $N, \theta_1, \cdots, \theta_r$. Let us assume that a solution for (θ, N) exists and is unique. Denote this by $(\hat{\theta}_M, \hat{N}_M)$. Since $X_i/X_j \to_{\text{a.s.}} \Pi_i/\Pi_j, j = 2, \cdots, r+1$ under suitable regularity conditions, $\hat{\theta}_M \to_{\text{a.s.}} \theta_0$. The asymptotic distribution of $H = (N_0^{\frac{1}{2}}(\hat{\theta}_M - \theta_0), N_0^{-\frac{1}{2}}(\hat{N}_M - N_0))$ may be derived as follows:
By (6.2), we have $\hat{N}_M \Pi_i (\hat{\theta}_M) = X_i$, $i = 1, \cdots, r+1$ and

$$N_0^{-\frac{1}{2}}(\hat{N}_M - N_0) \Pi_i (\hat{\theta}_M) + N_0^{\frac{1}{2}} \Pi_i (\hat{\theta}_M) - N_0^{\frac{1}{2}} \Pi_i (\theta_0) = N_0^{-\frac{1}{2}}(X_i - N_0 \Pi_i (\theta_0)) .$$

Using the mean value representation for $\Pi_i (\hat{\theta}_M) - \Pi_i(\hat{\theta}_0)$, the consistency of $\hat{\theta}_M$, assumption A1 and the fact that for $i = 1, \cdots, r+1$, $K_i = N_0^{-\frac{1}{2}}(X_i - N_0 \Pi_i (\theta_0))$ has a limiting distribution, we have

$$H - \sum_L^{-1} K \to_p 0 \tag{6.3}$$

where

$$K = (K_1, \cdots, K_{r+1}) , \qquad \sum_L = (\sigma_L{}^{j,m}) ,$$
$$\sigma_L{}^{j,m} = \partial \Pi_j (\theta_0)/\partial \theta_m , \qquad j = 1, \cdots, r+1; \; m = 1, \cdots r;$$

and

$$\sigma_L{}^{j,r+1} = \Pi_j (\theta_0) , \qquad j = 1, \cdots, r+1.$$

From (6.1) it is easy to see that

$$K \to_l \mathcal{N}(0, \sum_R) \tag{6.4}$$

where

$$\sum_R = (\sigma_R{}^{j,m}) ,$$
$$\sigma_R{}^{j,m} = \sum_{i=1}^{l} p_i(\theta_0)(h_{j,i} - \Pi_j (\theta_0))(h_{m,i} - \Pi_m(\theta_0)) ,$$

$j = 1, \cdots, r+1$; $m = 1, \cdots, r+1$ with $h_{j,l}$ defined as 0.
(6.3) and (6.4) imply $H \to_l \mathcal{N}(0, \sum_L^{-1} \sum_R \sum_L^{-1})$.

Acknowledgment. The author wishes to thank Professor David Wallace for suggesting this problem and for his helpful criticisms. She also wishes to thank the referee for many suggestions which were helpful in the revision.

REFERENCES

BARTLETT, M. S. (1956). Comment on Sir Ronald Fisher's paper. *J. Royal Statist. Soc. Ser. B* **18** 295-296.

CHAPMAN, D. G. (1951). Some properties of the hypergeometric distribution with applications to zoological sample censuses. *Univ. California Publ. Statist.* **1** 131–160.

COX, D. R. (1958a). Some problems connected with statistical inference. *Ann. Math. Statist.* **29** 357–372.

COX, D. R. (1958b). The regression analysis of binary sequences. *J. Royal Statist. Soc. Ser. B* **20** 215–242.

FISHER, R. A. (1956). *Statistical Methods and Scientific Inference*. Oliver E. Boyd, Edinburgh.

RAO, C. R. (1965). *Linear Statistical Inference* Wiley, New York.

SANATHANAN, L. P. (1969). Estimating population size in the particle scanning context. Ph. D. dissertation, University of Chicago.

WELCH, B. L. (1956). Note on some criticisms made by Sir Ronald Fisher. *J. Royal Statist. Soc. Ser. B* **34** 28–35.