SOME PROPERTIES OF THE HYPERGEOMETRIC DISTRIBUTION WITH APPLICATIONS TO ZOÖLOGICAL SAMPLE CENSUSES

BY
DOUGLAS G. CHAPMAN

UNIVERSITY OF CALIFORNIA PUBLICATIONS IN STATISTICS
Volume 1, No. 7, pp. 131–160

One of 30 copies printed
on 100% rag paper

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY AND LOS ANGELES
1951

SOME PROPERTIES OF THE HYPERGEOMETRIC DISTRIBUTION WITH APPLICATIONS TO ZOÖLOGICAL SAMPLE CENSUSES

BY

DOUGLAS G. CHAPMAN

One of 30 copies printed on 100% rag paper

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY AND LOS ANGELES
1951

University of California Publications in Statistics
Editors: M. Loève, G. M. Kuznets, E. L. Lehmann, J. Neyman

Volume 1, No. 7, pp. 131–160

Submitted by editors May 22, 1950
Issued April 20, 1951
Price, 50 cents

University of California Press
Berkeley and Los Angeles
California

Cambridge University Press
London, England

PRINTED IN THE UNITED STATES OF AMERICA

SOME PRPOERTIES OF THE HYPERGEOMETRIC DISTRIBUTION WITH APPLICATIONS TO ZOÖLOGICAL SAMPLE CENSUSES

BY

DOUGLAS G. CHAPMAN

In this paper* certain aspects of the problem of sampling without replacement from a finite population are treated; such sampling involves the use of the hypergeometric distribution. The results will be applied to a problem that arises in many zoölogical studies, *viz.*, the determination of the total size of the population under consideration. In such studies, it is necessary to estimate and to compare population sizes in order to formulate plans, or to evaluate the results, for either extermination or conservation programs. Since a total census is usually impractical, some sampling approach to the problem must be undertaken. The practical considerations which usually obtain in such a sample census will be kept in mind throughout this paper.

In the first section of the paper, the theory of the various modified χ^2 tests and estimates, developed by Neyman [**13**],† is adapted for use in connection with the hypergeometric distribution. The second section of the paper is devoted to the question of point estimates of the population size, while in the last section the use of the χ^2 tests will be discussed in relation to the problems that arise in the comparison of populations.

1. χ^2 Estimates and tests for parameters of the hypergeometric distribution

The use of various modified forms of the classical χ^2 for certain estimation and test procedures is based on certain asymptotic properties of these χ^2 estimates and tests. It is necessary to consider in what sense it is possible to formulate asymptotic properties for sampling without replacement, from a finite population. In particular, if the population size is the parameter that is being studied, the usual approach—that is, allowing the population size to increase indefinitely—is obviously an unsatisfactory procedure. We shall use a method of attack formulated by David [**3**].

Consider a sequence of systems of populations Π_{ki} of kN_i individuals ($i = 1, 2, \cdots, r$) ($k = 1, 2, \cdots$). Each population is sub-classified into classes Π_{kij}, ($j = 1, 2, \cdots, v_i$) each containing kt_{ij} individuals. A random sample is drawn from each Π_{ki} without replacement, of size kn_i; then s_{ijk} will denote the number of individuals of the sample that belong to the subclass Π_{kij}.

Further, let

$$p_{ij} = \frac{t_{ij}}{N_i} \qquad \text{and} \qquad q_{ijk} = \frac{s_{ijk}}{kn_i} .$$

The s_{ijk} and hence also the q_{ijk} are random variables.

* This research was supported in part by the Office of Naval Research and was prepared while the author, now at the University of Washington, was a Research Assistant in the Statistical Laboratory, University of California, Berkeley.

† Boldface arabic numerals in square brackets indicate similarly numbered references on page 159.

Concerning the p_{ij}'s, the assumptions made by Neyman [**13**], pp. 239–240, will be used. These in brief are that

$$(1) \qquad p_{ij} = f_{ij}(\theta_1, \theta_2, \cdots, \theta_m) > 0\,, \qquad \sum_{j=1}^{v_i} f_{ij} \underset{\theta}{\equiv} 1, \qquad (i = 1, 2, \cdots, r),$$

where $\theta_1 \cdots \theta_m$ are unknown parameters. It is assumed that the functional forms f_{ij} are known and further that these functions have continuous first and second partial derivatives with respect to all the θ's. Since the properties of various tests are to be studied as k tends to infinity, it is possible to assume that N_i and t_{ij} take on any rational values within some range and hence are essentially continuous variables over this range. The N_i are parameters which may be known or unknown.

With these general conditions, the situation is completely analogous to that described by Neyman for the determination of a class of best asymptotically normal (BAN) estimates. A BAN estimate of a parameter is one that is consistent, asymptotically normal (AN) and has minimum asymptotic variance of all consistent AN estimates. Neyman has postulated a fourth desirable though not essential requirement, namely, that the estimate, a function of the random variables, have continuous partial derivatives with respect to these variables. Reference will also be made to various results of Cramér [**2**], chapter 32, concerning "efficient" and "asymptotically efficient" estimates. Cramér defines efficiency of an unbiased estimate in terms of a particular lower bound for the variance of the estimate. From this it follows that any "efficient estimate" certainly has minimum variance among all unbiased estimates but the converse may not be true. While in general, a wider definition would be more desirable, so that "efficient estimates" are equivalent to "minimum variance unbiased estimates," the results obtained by Cramér are sufficient for this paper. It may also be pointed out that these definitions of "bestness" and "efficiency" are relative to the known or assumed functional forms f_{ij} and to the distribution functions of the random variables.

In place of the multinomial distribution in Neyman's case, in this paper the random variables s_{ijk} have the following generalized hypergeometric distribution:

$$P(E_k \in W) = \sum \prod_{i=1}^{r} \frac{(kN_i - kn_i)\,!(kn_i)\,!}{(kN_i)\,!} \prod_{j=1}^{v_i} \frac{(kt_{ij})\,!}{s_{ijk}\,!(kt_{ij} - s_{ijk})\,!},$$

where E_k denotes the event point $(s_{11k} \cdots s_{rv_rk})$. W is any subset of the sample space (Euclidean space of $\sum_{i=1}^{r} v_i$ dimensions) and the summation is for all points which fall in W.

Using a superscript zero to denote the true values of the unknown θ's and p's, it is known that

$$E(q_{ijk}) = p_{ij}^0\,,$$

$$E(q_{ijk} - p_{ij}^0)^2 = \frac{(N_i - n_i)\, p_{ij}^0\,(1 - p_{ij}^0)}{kn_i\,(N_i - \frac{1}{k})}\,.$$

We may then state

LEMMA I. *If a_{ij} ($i = 1, 2, \cdots, r, j = 1, 2, \cdots, v_i$) are any fixed numbers, then the variance of the variable*

$$X_k = \sum_{i=1}^{r} \sum_{j=1}^{v_i} a_{ij}(q_{ijk} - p_{ij}^0) \tag{2}$$

is given by

$$\sigma_k^2 = \sum_{i=1}^{r} \frac{\sigma_i^2}{kn_i}\left(\frac{N_i - n_i}{N_i - \frac{1}{k}}\right), \tag{3}$$

where

$$\sigma_i^2 = \sum_{j=1}^{v_i} (a_{ij} - a_{i.})^2 p_{ij}^0,$$

and

$$a_{i.} = \sum_{j=1}^{v_i} a_{ij} p_{ij}^0.$$

Lemma I is obtained by straightforward algebra. It is further seen that as k tends to infinity, $k\sigma_k^2$ tends to

$$\sum_{i=1}^{r} \frac{\sigma_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right).$$

It is now desirable to consider two cases, according to whether the N_i are known or unknown. In case the N_i are known, in virtue of the theorem of David [**3**], it is possible to use the remaining results of Neyman's theorems with modifications only of notation. We state merely the results concerning the various BAN estimates and note that the tests may be similarly duplicated.

LEMMA 2. *If ϑ is a BAN estimate of θ, with continuous partial derivatives in the random variables up to the second order, it is representable in the form*

$$\vartheta = \theta + X_k + Y_k \sigma_k,$$

where X_k and σ_k are defined by formula (2) *and* (3), *with the a_{ij} depending on the* p's *but independent of the* q's *and of k, and Y_k is a random variable tending in probability to zero as $k \to \infty$.*

Here θ stands for any one of $\theta_1, \theta_2, \cdots, \theta_m$.

THEOREM 1. *Under the general conditions prescribed, BAN estimates of the parameters $\theta_1 \cdots \theta_m$ may be obtained by minimizing either*

$$\chi_1^2 = k \sum_{i=1}^{r} \sum_{j=1}^{v_i} \frac{n_i(q_{ijk} - p_{ij})^2}{p_{ij}(1 - \frac{n_i}{N_i})},$$

or

$$\chi_2^2 = k \sum_{i=1}^{r} \sum_{j=1}^{v_i} \frac{n_i(q_{ijk} - p_{ij})^2}{q_{ijk}(1 - \frac{n_i}{N_i})}, \tag{4}$$

under either of the restrictions

(5) $$F_j(p_{11}, \cdots, p_{rv_r}) = 0\,,$$

(6) $$F_j^* \equiv F_j(q_{11k}, \cdots, q_{rv_rk}) + \sum_{i=1}^{r} (p_{ij} - q_{ijk}) \left(\frac{\partial F}{\partial p_{ij}}\right)_{p_{ij} = q_{ijk}} = 0\,.$$

The equations F_j are obtained from the original set of equations (1), relating the p's and θ's by elimination of the θ's. The set of equations (6) represents essentially a linear approximation to the true restrictions.

The methods of estimation (and the corresponding test procedures) outlined above, will be useful for many problems of sampling from a finite population of known size but unknown composition. The alternative situation where some characteristics but not the total size are known is the sample census problem.

χ^2 tests and estimates will now be derived for the case that N_i are unknown but some of the t_{ij} are known. The algebra is simplified by assuming that $v_i = 2$ and t_{i1} $(= t_i)$ are known, while t_{i2} $(= N_i - t_i)$ are unknown. The difference between this situation and that treated above is caused by the additional unknown parameter in the denominator of the expression for χ^2.

It is still possible to use proofs parallel to Neyman's. However, the problem will be approached differently: this new approach may cast some additional light on the general class of BAN estimates and some of the lemmas may have some interest in themselves. Since the lemmas are more general than required to prove the theorems needed concerning hypergeometric tests and estimates, the notation required will be indicated in the statements of the lemmas and will not refer to that developed at the beginning of this section. We shall proceed to prove these lemmas.

LEMMA 3. *If X_i $(i = 1, 2, \cdots, n)$ are independent random variables distributed $N(m_i, \sigma_i)$, where the σ_i are known and where the m_i are linear functions of m independent but unknown parameters $\theta_1, \theta_2, \cdots, \theta_m$ $(m < n)$, then the maximum likelihood (ML) estimates of the θ's and m's are normally distributed and efficient.*

PROOF. Since the θ's are independent parameters, it is possible to find m linear equations of the form

$$\theta_1 = a_{i0} + \sum_{j=1}^{n} a_{ij} m_j \qquad (i = 1, 2, \cdots, m)\,,$$

expressing the θ's in terms of the m's.

Then the new random variables $Z_1, Z_2, \cdots, Z_m$ defined by the transformation

$$Z_i = a_{i0} + \sum_{j=1}^{n} a_{ij} X_j\,, \qquad (i = 1, 2, \cdots, m)\,,$$

are normally distributed with means θ_i and known variances and covariances. The lemma follows from example 2, p. 495 of Cramér [**2**] (generalized to m dimensions).

LEMMA 4. *If X_{ik} $(i = 1, 2, \cdots, n)$ $(k = 1, 2, \cdots)$ are independent random variables distributed $N(m_i, \sigma_i/\sqrt{k})$ where the σ_i are known and where the m_i are functions of m independent but unknown parameters $\theta_1, \theta_2, \cdots, \theta_m$ $(m < n)$, and if these functions are continuous with continuous first and second partial derivatives with respect to the*

θ's, then as k tends to infinity the ML estimates of the m's tend to be normally distributed and efficient.

Proof. Let $(m_1^0, m_2^0, \cdots, m_n^0)$ denote the true parameter point. It follows from the assumptions on the parameters and on the relations connecting the m's and the θ's that it is possible to find n–m equations.

$$F_j(m_1, m_2, \cdots, m_n) = 0\,, \qquad (j = 1, 2, \cdots, n-m)\,, \tag{7}$$

such that the functions F_j possess continuous partial derivatives up to the second order with respect to all the variables m_i.

A set of linear functions in the m's may then be defined as follows:

$$F_j^0(m_1, m_2, \cdots, m_n) = F_j(m_1^0, m_2^0, \cdots, m_n^0) + \sum_{i=1}^{n} \left(\frac{\partial F_j}{\partial m_i}\right)_0 (m_i - m_i^0),$$

where $\left(\frac{\partial F_j}{\partial m_i}\right)_0$ indicates that the partial derivatives are evaluated at the point $(m_1^0, m_2^0, \cdots, m_n^0)$.

Let

$$\rho^2 = \sum_{i=1}^{n} (m_i - m_i^0)^2.$$

It is known that given ϵ there exists a $\rho(\epsilon)$ such that, for $\rho < \rho(\epsilon)$

$$|F_j - F_j^0| < \rho(\epsilon)\,, \qquad (j = 1, 2, \cdots, n-m)\,. \tag{8}$$

By lemma 3 the ML estimates of the m's obtained under the set of restrictions

$$F_j^0(m_1, m_2, \cdots, m_n) = 0\,, \qquad (j = 1, 2, \cdots, n-m) \tag{9}$$

are normally distributed and efficient for all values of k. The efficiency of the estimates is relative to the distribution involving the functions F_j^0.

These ML estimates of the m's are determined by minimizing

$$W = \sum_{i=1}^{n} \frac{(X_{ik} - m_i)^2}{\sigma_i^2}\,, \tag{10}$$

subject to the restrictions of formula (9).

W, of formula (10), represents a weighted distance (squared) from the observed point $(X_{1k}, X_{2k}, \cdots, X_{nk})$ to a point on the hyperplane $F_j^0 = 0$ $(j = 1, 2, \cdots, n-m)$. The estimation procedure is equivalent to finding a point $m_1^*, m_2^*, \cdots, (m_m^*)$ on the hyperplane of the m's at a minimum weighted distance from the observed point, *i.e.*, the equivalent to solving the equations

$$\frac{X_{ik} - m_i^*}{\sigma_i^2} = \sum_{j=1}^{n-m} \lambda_i^0 \left(\frac{\partial F_j}{\partial m_i}\right)_{m_i = m_i^0} = t_i^0\,, \qquad (i = 1, 2, \cdots, n)\,, \tag{11}$$

where

$$\sum_{i=1}^{n} (t_i^0)^2 = \sum_{i=1}^{n} \left[\frac{X_{ik} - m_i^*}{\sigma_i^2}\right].$$

Again, the estimation of $m_1, m_2, \cdots, m_n$ under the original restrictions (7) is equivalent to finding a point $(m_1^*, m_2^*, \cdots, m_n^*)$ on the hypersurface defined by equations (7) at a minimum distance from the observed point, i.e., the equivalent to solving the equations

$$\text{(12)} \qquad \frac{X_{ik} - m_i^*}{\sigma_i^2} = \sum_{j=1}^{n-m} \lambda_j \left(\frac{\partial F_j}{\partial m_i}\right)_{m_i = m_i^*} = t_i\,, \qquad (i = 1, 2, \cdots, n)\,,$$

where

$$\sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} \left[\frac{X_{ik} - m_i^*}{\sigma_i^2}\right].$$

The conditions of continuity placed on the functions F ensure that the solutions of (11) and (12) have the same limiting distribution, provided it is shown that m_i^* converges to m_i^0.

To show this, consider the hyperellipsoid

$$E_{\left(\frac{\rho}{2}\right)}: \sum_{i=1}^{n} \frac{(X_{ik} - m_i^0)^2}{\sigma_i^2} = \frac{\rho}{2}$$

in the X-space. If the observed point $(X_{1k}, X_{2k}, \cdots, X_{nk})$ falls within this hyperellipsoid, it is certainly true that the estimated point $(m_1^*, m_2^*, \cdots, m_n^*)$ lies within the hyperellipsoid

$$E(\rho): \sum_{i=1}^{n} \frac{(X_{ik} - m_i^0)^2}{\sigma_i^2} = \rho\,.$$

For any point $(X_{1k}, X_{2k}, \cdots X_{nk})$ within the first hyperellipsoid

$$\sum_{i=1}^{n} \frac{(X_{ik} - m_i^0)^2}{\sigma_i^2} < \frac{\rho}{2}\,,$$

while for any point $(Y_1 \cdots Y_n)$ outside the second hyperellipsoid $E(\rho)$

$$\sum_{i=1}^{n} \frac{(X_{ik} - Y_i)^2}{\sigma_i^2} > \frac{\rho}{2}\,.$$

Hence if $(m_1^*, m_2^*, \cdots, m_n^*)$ is the point that lies on the hypersurface and minimizes W it must lie in the second hyperellipsoid, $E(\rho)$.

From the distribution of the X's, it follows that, given η, $E(\rho)$ an integer $k(\rho)$ can be found such that for $k > k(\rho)$

$$\text{(13)} \qquad P\{X_k \in E(\rho)\} > 1 - \eta\,.$$

Consequently, the asymptotic distribution of the ML estimates is the same whether the equations $F_j = 0$ or $F_j^0 = 0$ are used in their determination.

It remains to demonstrate that the efficiency is the same, at least asymptotically in both cases. In virtue of a theorem of Cramér [2], p. 495, this will be demonstrated if it is shown that, given ϵ, for k sufficiently large that

$$\left| E\left(\frac{\partial \log f_i}{\partial m_\alpha} \frac{\partial \log f_j}{\partial m_\beta}\right) - E\left(\frac{\partial \log f_i^0}{\partial m_\alpha} \frac{\partial \log f_j^0}{\partial m_\beta}\right) \right| < \epsilon, \tag{14}$$

$$(i, j = 1, 2, \cdots, n), (\alpha, \beta = 1, 2, \cdots, n-m),$$

where f_j, f_j^0 symbolize the distribution functions of the X's involving the restrictions F_j and F_j^0 respectively, and differentiation is performed with respect to the $n-m$ independent parameters.

Let $X_k = (X_{1k}, X_{2k}, \cdots, X_{nk})$, $m^0 = (m_1^0, m_2^0, \cdots, m_n^0)$ and write

$$E\left(\frac{\partial \log f_i^0}{\partial m_\alpha} \frac{\partial \log f_j^0}{\partial m_\beta}\right) = E\left(\frac{\partial \log f_i^0}{\partial m_\alpha} \frac{\partial \log f_j^0}{\partial m_\beta} \Big/ X_k \epsilon E(\rho)\right) \cdot P\{X_k \epsilon E(\rho)\}$$

$$+ E\left(\frac{\partial \log f_i^0}{\partial m_\alpha} \frac{\partial \log f_j^0}{\partial m_\beta} \Big/ X_k \epsilon' E(\rho)\right) \cdot P\{X_k \epsilon' E(\rho)\}.$$

The existence of a k_0 such that $k > k_0$ implies (14), follows from (8) and (13).

LEMMA 5. *If $X^{(k)}$ $(k = 1, 2, 3, \cdots)$ are independent random variables distributed $N(m, \sigma/\sqrt{k})$, and if $f(x)$ is a continuous function of x with continuous first and second partial derivatives then*

$$\frac{\sqrt{k}\, f(X^{(k)}) - f(m)}{\sigma \cdot \tau},$$

where

$$\tau = \left(\frac{df}{dx}\right)_{x=m}$$

is asymptotically $N(0, 1)$.

PROOF. Writing the Taylor expansion for $f(x)$ about m

$$f(x) = f(m) + (x - m)\left(\frac{df}{dx}\right)_{x=m} + \frac{(x-m)^2}{2}\left(\frac{d^2 f}{dx^2}\right)_{x=m+\theta(x-m)}, \quad (0 \leqq \theta \leqq 1),$$

it follows that for all k

$$\frac{\sqrt{k}\, f(X^{(k)}) - f(m)}{\sigma \cdot \tau} = \frac{\sqrt{k}\,(X^{(k)} - m)}{\sigma} + R_k.$$

Since $X^{(k)}$ tends in probability to m as k tends to infinity the lemma is proved if it is shown that at the same time R_k tends in probability to zero.

But

$$R_k = \frac{1}{2}\left[\frac{\sqrt{k}\,(X^{(k)} - m)}{\sigma}\right]\left[\left(\frac{d^2 f}{dx^2}\right)_{x=m+\theta(X^{(k)}-m)}\right] \cdot (X^{(k)} - m). \tag{15}$$

In denoting the three terms in (15) by U, V_k, W_k, it is observed that U is $N(0, 1)$ and hence given ϵ, $\eta < 1$ there exists an M such that

$$P(|U| < M) > 1 - \frac{\eta}{3}.$$

Further let

$$\left(\frac{d^2 f}{dx^2}\right)_{x=m} = C\,;$$

from the conditions on f, δ_1 exists such that

$$\left|\frac{d^2 f}{dx^2}\right| < 2\,|C| + 1 \qquad \text{for} \qquad |x - m| < \delta_1\,.$$

Set

$$\delta = \min\left(\delta_1, \frac{\epsilon}{M(2|C| + 1)}\right);$$

there exists a K such that for $k > K$,

$$P(|X_k - m| < \delta) > 1 - \frac{\eta}{3},$$

and hence

$$P(|X_k - m| < \delta / |U| < M) > 1 - \frac{\eta}{3}.$$

Now

$$\begin{aligned} P(|R_k| < \epsilon) &> P(|U| < M) \cdot P(|V_k| < 2C \mid |U| < M) \\ &\quad \cdot P(|W_k| < \mid |U| < M, |V_k| < 2C) \\ &> \left(1 - \frac{\eta}{3}\right)\left(1 - \frac{\eta}{3}\right)\left(1 - \frac{\eta}{3}\right) > 1 - \eta\,. \end{aligned}$$

Combining lemmas 4 and 5, it follows that the ML estimates of the θ's of lemma 4 are also asymptotically normal and efficient.

Since the F^0 equations can be formed only if the true parameter point is known, they cannot be used to simplify the procedure of actually determining estimates of the unknown parameters. However, following Neyman's procedure, it is possible to introduce a system of linear equations with known coefficients, *viz.*,

$$\begin{aligned} F_j^*(m_1, m_2, \cdots, m_n) &= F_j(x_{1k}, x_{2k}, \cdots, x_{nk}) \\ &\quad + \sum_{i=1}^{n} (m_i - x_{ik})\left(\frac{\partial F_j}{\partial m_i}\right)_{m_i = x_{ik}} = 0\,, \\ &(j = 1, 2, \cdots, n\text{–}m)\,. \end{aligned} \tag{16}$$

LEMMA 6. *Under the same assumptions as those of lemma 4, the ML estimates of the m's (and θ's) calculated under the restrictions $F_j^* = 0$ have the same limiting distribution as the estimates calculated under the restrictions $F_j = 0$.*

PROOF. These "modified" ML estimates are solutions of the equations

$$\frac{x_i - m_i}{\sigma_i^2} + \sum_{j=1}^{n-m} \lambda_j \left(\frac{\partial F_j}{\partial m_i}\right)_{m_i = x_{ik}} = 0 \,, \qquad (i = 1, 2, \cdots, n) \,, \tag{17}$$

and of the equations (16). The λ_j's are Lagrange multipliers.

Since the x_{ik} converge in probability to m_i^0 $(i = 1, 2, \cdots, n)$ as k tends to infinity, it follows, because of the continuity conditions imposed upon the functions F_j, that the linear equations (16) and (17) converge to the set of linear equations by which $(m_1^*, m_2^*, \cdots, m_n^*)$ were determined. Consequently, the solutions of the linear equations also converge in probability.

LEMMA 7. *If X_i $(i = 1, 2, 3, \cdots)$ is a sequence of random variables converging in probability to X_0 and if $f_i(x, \theta)$ is the distribution function of k sample values of X_i, depends on an unknown parameter θ, and if further, it is assumed that f_0 is continuous with continuous first partial derivatives in x and θ, and that $\frac{\partial f(x_i, \theta)}{\partial \theta} = 0$ has a unique solution θ_i for each x_i, then θ_i converges in probability to θ_0.*

PROOF. Under the assumptions made on f_0 the solution of the equation

$$\frac{\partial f(x, \theta)}{\partial \theta} = 0$$

is a continuous function of x (Graves [5], p. 138). That is, given ϵ, there exists δ_ϵ, such that

$$|x_i - x_0| < \delta \qquad \text{implies} \qquad |\theta_i - \theta_0| < \epsilon \,. \tag{18}$$

But, by the assumptions on the sequence $\{X_i\}$, it is known that there exists an i_0 such that $i > i_0$ implies

$$P\{|X_i - X_0| < \delta\} > 1 - \zeta \,,$$

and hence

$$P\{|\theta_i - \theta_0| < \epsilon\} > 1 - \zeta \,.$$

If, further, we generalize the conditions of lemma 7 slightly to consider a sequence of functions of a parameter and a random variable (denoted by $\{X_i(\theta)\}$), which converge to a random variable, x_0, where the distributions $f_i(x, \theta)$ and $f_0(x)$—the latter may be independent of θ—are such as to satisfy the conditions of the lemma, then the sequence θ_i will still converge in probability and further there exists a random variable θ_0, to which the random variables θ_i converge. This limiting distribution of θ_i is determined by the distribution function $f_0(x)$.

Returning now to the setup described at the beginning of this section, a lemma is proved concerning the hypergeometric distribution as a preliminary step to applying the results of lemmas 3 to 7.

LEMMA 8. *As k tends to infinity,*

$$\frac{\sqrt{kn_i}\,(q_i - p_i)}{\sqrt{q_i(1 - q_i)\left(1 - \frac{n_i q_i}{t_i}\right)}}$$

is asymptotically $N(0, 1)$.

PROOF. On writing

$$\frac{\sqrt{kn_i}\,(q_i - p_i)}{\sqrt{q_i(1 - q_i)\left(1 - \frac{n_i q_i}{t_i}\right)}} = \frac{\sqrt{kn_i}\,(q_i - p_i)}{\sqrt{p_i(1 - p_i)\left(1 - \frac{n_i p_i}{t_i}\right)}} \frac{\sqrt{p_i(1 - p_i)\left(1 - \frac{n_i p_i}{t_i}\right)}}{\sqrt{q_i(1 - q_i)\left(1 - \frac{n_i q_i}{t_i}\right)}}, \tag{19}$$

one sees that, of the two terms on the right hand side of (19), the first is asymptotically $N(0, 1)$ by the theorem of David [**3**] and the second tends to 1 in probability. Hence, by using a theorem of Cramér [**2**], p. 254, the lemma is proved.

We are now in a position to prove the theorem needed concerning BAN estimates that will be useful in sampling from a finite population.

THEOREM 2. *Under the general conditions outlined on pp.* 2, 3, *and* 6, *BAN estimates of the parameters* N_i $(i = 1, 2, \cdots, r)$ *are obtained by minimizing*

$$\chi_3^2 = \sum_{i=1}^{r} \frac{n_i(q_i - p_i)^2}{q_i(1 - q_i)\left(1 - \frac{n_i q_i}{t_i}\right)}, \tag{20}$$

under either of the restrictions (5) *or* (6).

PROOF. Combining lemmas 7 and 8 with lemma 4, in case the restrictions (4) are imposed and also lemma 6 in case the modified restrictions (5) are used, it is seen that this minimization yields asymptotically efficient and asymptotically normal estimates of the p's.

But the N's are functions of the p's which satisfy the conditions of lemma 5 and hence the estimates of the N's are asymptotically efficient and asymptotically normal.

It may be observed that the above lemmas may be applied together with the known facts concerning the asymptotic distribution of multinomial random variables to prove the theorems on BAN estimates originally formulated by Neyman [**13**]. However, this approach does not yield such general results as Neyman obtained in his theorems 1 and 2. These lemmas together with the Central Limit Theorem may also be used to derive some of the results obtained by Gurland [**7**] concerning BAN estimates for a class of continuous distributions.

The theorem needed in testing hypotheses concerning population sizes is a direct consequence of a theorem proved by Cramér [**2**], p. 506 and also pp. 426–434. The theorem will be formulated without proof.

As is customary, let Ω denote the whole parameter space and let H be the subset of the parameter space specified by the hypothesis to be tested. It is assumed that the restrictions that the parameters be elements of H (or Ω) is equivalent to the specification of a number of equations which the parameters must satisfy in H (or

Ω). Let m be the number of restrictions on the parameters under H and μ ($<m$), the number of restrictions under Ω.

THEOREM 3. In the limit as k tends to infinity

$$\chi_q^2 = k \sum_{i=1}^{r} \frac{n_i(q_i - p_i')^2}{q_i(1-q_i)\left(1 - \frac{n_i q_i}{t_i}\right)} - k \sum_{i=1}^{r} \frac{n_i(q_i - p_i'')^2}{q_i(1-q_i)\left(1 - \frac{n_i q_i}{t_i}\right)}, \tag{21}$$

and

$$\chi_p^2 = k \sum_{i=1}^{r} \frac{n_i(q_i - p_i')^2}{p_i'(1-p_i')\left(1 - \frac{n_i p_i'}{t_i}\right)} - k \sum_{i=1}^{r} \frac{n_i(q_i - p_i'')^2}{p_i''(1-p_i'')\left(1 - \frac{n_i p_i''}{t_i}\right)}, \tag{22}$$

where p_i', p_i'', are BAN estimates of the p's as defined in theorem 2, under H and Ω respectively, are both distributed according to the χ^2 distribution with $m - \mu$ degrees of freedom.

Denote the denominators of (21) and (22) by σ_q^2 and σ_p^2. Theorem 3 remains true with χ^2 defined as follows:

$$\chi_{p+q}^2 = k \sum_{i=1}^{r} \frac{n_i(q_i - p_i')^2}{a\sigma_q^2 + b\sigma_p^2} - k \sum_{i=1}^{r} \frac{n_i(q_i - p_i'')^2}{a\sigma_q^2 + b\sigma_p^2}, \qquad (a + b = 1). \tag{23}$$

It may be pointed out that different BAN estimates of the p's may be used in the numerators and denominators of (22) and (23).

2. Population estimates: single sample census

In a sample census to determine the size (and possibly other characteristics) of an animal population, a certain number of members of the population are tagged (or marked) on one or more occasions and simultaneously or subsequently a sample of the population is obtained and the number of tagged and untagged members noted. The notation used with respect to the hypergeometric distribution will now be given specific meanings in relation to such a tagging and sampling experiment. In this section, the case of the single sample obtained subsequent to the tagging will be considered.

The following symbols will be used:

N = the total population size,

t = the number of members of the population tagged,

n = the number of members of the population subsequently sampled,

s = the number of tagged individuals in the sample.

Such an approach to the problem of a zoological census on the basis of sampling experiments has been used with increasing frequency in recent years to determine the sizes of such populations as ducks, lizards, hares, tsetse flies and of several species of fish, notably salmon and sardine. For a discussion of various practical considerations involved in such programs and of certain results thereof, reference may be made to papers of Lincoln [**11**], Jackson [**9**], Green and Evans [**6**], Dice [**4**], and Ricker [**14**]. A fuller list of such references may be found in Ricker's paper, and in a paper of Schaeffer [**15**]. It will be assumed throughout this paper that the sampling process is random, and proceeding from this assumption various methods of analysis of the data from such experiments are discussed. Whether or not this assumption of randomness is valid in actual field experiments is discussed in the papers of Green and Evans [**6**] and in a paper by Howard [**8**]. It should be pointed out that the randomness assumed is only that the probability that an individual is sampled is independent of whether it is tagged or untagged—neither the tagging nor the sampling need be random with respect to any other characteristics of the population.

In most fields where sample censuses have been attempted, a satisfactory tagging program has been found to be much more difficult than the sampling part of the experiment. It will, therefore, be useful throughout the remainder of the paper, to restrict attention to the case $t < n$. Usually, t will be much smaller than n, though this restriction is not imposed.

The random variable s has the distribution

$$P(s; N, n, t) = \frac{n!(N-n)!(N-t)!t!}{s!(n-s)!N!(N-n-t+s)!(t-s)!}$$

(frequently $P(s; N)$ or $P(s)$ will be used in place of $P(s; N, n, t)$ if there is no ambiguity as to the parameters involved).

By a method completely analagous to that used for the parameter p of the binomial distribution, the ML estimate of N, given s, may be found.

In considering

$$q(N) = \frac{P(s; N)}{P(s; N-1)} = \frac{(N-t)(N-n)}{N(N-n-t-s)}.$$

$$q(N) = \frac{1 + \dfrac{nt}{N^2 - N(n+t)}}{1 + \dfrac{Ns}{N^2 - N(n+t)}},$$

it is seen that

$$Ns \gtrless nt \rightarrow q(N) \lessgtr 1.$$

Hence, the value of N (denoted by N_0) which maximizes $P(s; N)$ is the integer satisfying the double inequality,

$$\frac{nt}{s} - 1 < N_0 \leqq \frac{nt}{s}.$$

Since N_0 is undefined when $s = 0$, the moments of N_0 do not exist. In actual experiments, this limitation may be ignored because $P(s = 0)$ is negligible, or through the use of other information to establish an upper bound for N. In order to study the moments of N_0, ML estimate will be modified slightly as follows:

$$N_0 = 2nt, \qquad (s = 0).$$

As n increases and approaches N, nt/s tends to N. More generally, under the arrangement specified at the beginning of the first section, as k tends to infinity with n and N fixed, N_0 tends to N in probability. The ML estimate is therefore consistent.

It is easy to show that N_0 is also the estimate of N obtained by any of the procedures outlined in theorem 2 of section **1**. The asymptotic properties of N_0 are known from the general theory. Its properties for fixed finite sample sizes will now be studied and on the basis of these properties some other estimates of N will be derived.

The properties to be studied are bias and the variance about the true parameter point N. The bias of an estimate N^* of N, is defined to be

$$b^*(N) = E(N^*) - N\,.$$

The variance of N^* about the true parameter point N denoted $V(N^*)$ is

$$V(N^*) = E(N^* - N)^2 = E(N^{*2}) - 2Nb^*(N) - N^2\,.$$

It is convenient to use a method of Stephan [16] and to expand the reciprocals of $s + a$, $(s + a)^2$, where a is any non-negative number in a series of inverse factorials. Details of such expansions may be found in Whittaker and Watson [17], Milne-Thomson [12], or Jordan [10].

The expansions with remainder term are

$$(24)\qquad \frac{1}{s+a} = \frac{1}{s+1} + \frac{1-a}{(s+1)(s+2)} + \frac{(1-a)(2-a)}{(s+1)(s+2)(s+3)} + \cdots + \frac{(1-a)(2-a)\cdots(n-1-a)}{(s+1)(s+2)\cdots(s+n)} + R_n\,,$$

$$(25)\qquad \frac{1}{(s+a)^2} = \frac{1}{(s+1)(s+2)} + \frac{3-2a}{(s+1)(s+2)(s+3)} + \frac{11-12a+3a^2}{(s+1)(s+2)(s+3)(s+4)} + \cdots + \frac{A_{n-1}}{(s+1)(s+2)\cdots(s+n)} + R_n'\,,$$

where

$$R_n = \frac{(1-a)(2-a)\cdots(n-a)}{(s+1)(s+2)\cdots(s+n)(s+a)}\,,$$

$$A_{n-1} = (-1)^n \sum_{j=1}^{n-1} \frac{\prod_{k=1}^{n-1}(a-k)}{a-j}\,, \qquad a \neq 1, 2, \cdots (n-1)$$

$$= (-1)^n(a-1)(a-2)\cdots(a-k+1)(a-k-1)\cdots(a-n+1)\,,$$
$$(a = k, \text{integer} \leqq n)\,,$$

$$(26)\qquad R_n' = (-1)^n \left[\frac{A_n}{(s+1)(s+2)\cdots(s+n)(s+a)} - \frac{\prod_{i=1}^{n}(a-i)}{(s+1)(s+2)\cdots(s+n)(s+a)^2} \right].$$

R_n and R_n' tend to zero as n increases and consequently the series converge for values of $s > -a$. The expectations of the terms of these series may be evaluated directly or by the use of the equation

$$E\left(\frac{1}{(s+1)(s+2)\cdots(s+i)}\right) = \int_0^1 \cdots \int_0^{u_{i-2}} \int_0^{u_{i-1}} \sum_{s=0}^{t} P(s) du_1 \cdots du_i$$

$$= \frac{(N-n)!(N-t)!}{N!(N-n-t)!} \int_0^1 \cdots \int_0^{u_{i-2}} \int_0^{u_{i-1}} F(-n, -t, N-n-t-1, u_i) du_1 \cdots du_i,$$

where $F(a, b, c, x)$ is the hypergeometric function.
Thus

$$(27)\qquad E\left(\frac{1}{s+1}\right) = \frac{N+1}{(n+1)(t+1)} \left[1 - \frac{(N-n)!(N-t)!}{(N+1)!(N-n-t-1)!}\right].$$

Stephan [**16**] using the following notation

$$\mu_i = E\left[\frac{1}{(s+1)(s+2)\cdots(s+i)}\right], \qquad (i = 1, 2, \cdots n),$$

and

$$s_i = 1 - \sum_{s=0}^{i-1} P(s|N+i, n+i, t+i), \qquad (i = 1, 2, \cdots n),$$

showed that the following simple recurrence formula holds:

$$(28)\qquad \mu_i = \frac{s_i(N+i)}{s_{i-1}(n+i)(t+i)} \cdot u_{i-1}, \qquad (i \geqq 2).$$

Examining equation (27), it is found that the second term in the square brackets is negligible for n and t sufficiently large relative to N. This immediately suggests an alternative estimate for N, namely,

$$N_1 = \frac{(n+1)(t+1)}{s+1} - 1.\ddagger$$

‡ In practice, the whole number immediately less than $\frac{(n+1)(t+1)}{s+1}$ or even $\frac{nt}{s+1}$ will be the estimate. The above form is more convenient for mathematical purposes.

Stirling's formula may be used to find the values of n and t, in terms of N, such that

$$|E(N_1) - N| \leqq 1, \tag{29}$$

therefore for such values of n, t and N, N_1 is an essentially unbiased estimate of N. Since

$$|E(N_1) - N| = \left| \frac{(N-n)!(N-t)!}{N!(N-n-t-1)!} \right|,$$

(29) is true if

$$\frac{(N-n)!(N-t)!}{(N+1)!(N-n-t-1)!} \leqq \frac{1}{N+1}. \tag{30}$$

Then, using Stirling's formula, (29) is true if

$$\begin{aligned} (N - n + \tfrac{1}{2}) \log (N-n) + (N - t + \tfrac{1}{2}) \log (N-t) + |\delta| \\ \leqq \log \frac{1}{N+1} + (N+\tfrac{3}{2}) \log (N+1) + (N-n-t-\tfrac{1}{2}) \log (N-n-t-1), \end{aligned} \tag{31}$$

where

$$|\delta| < \left| \frac{1}{12N} - \frac{1}{12(N-n)} \right| + \left| \frac{1}{12(N-t)} - \frac{1}{12(N-n-t-1)} \right|.$$

After $(2N - n - t - 1) \log N$ is subtracted from each side the first and last two logarithmic terms may be expanded in a Maclaurin series and equation (31) becomes

$$\log N + \frac{1}{N} + |\delta| \leqq \frac{nt}{N} + \frac{n+t+1}{N} + \frac{1}{N^2}\left[\frac{n^2}{2} + \frac{t^2}{2} + \frac{nt}{2}(n+t+1) - \frac{3}{2}\right]$$
$$+ \text{ positive terms in higher powers of } 1/N.$$

If the values of n and t are such that $n + t + 1 < N/2$, then $|\delta| < 1/3N$, and (29) will certainly hold if $nt/N > \log N$.

The solutions of this equation for a range of values of N is given in terms of expected tag ratio in table 1 below.

TABLE 1

SAMPLE SIZE REQUIRED IN ORDER THAT $b_1(N)$ BE CERTAINLY LESS THAN 1

N	10^4	10^5	10^6	10^7	10^8	10^9
$\frac{nt}{N}$	9.2	11.5	13.8	16.1	18.4	20.7

For the sake of compactness and also because $E(s) = nt/N$, nt/N has been tabulated rather than nt. This table may have some value in the design of sampling programs of this type.

In a similar manner, it may be shown that, provided

$$\frac{nt}{N} > \log\frac{100}{\epsilon}\,,$$

$$|E(N_1) - N| < \frac{\epsilon}{100} N\,;$$

that is, the bias is less than ϵ per cent of the true value of N.

If nt/N is very small then N_1 may have a considerable negative bias. Table 2 below shows $E(n_1)$ for various values n, t, and N in this case. However, it will be shown subsequently that, in sample censuses for which nt/N is very small, the variance of the estimate of N will be so large as to make the estimate of little value.

TABLE 2

EXPECTED VALUE OF N_0, N_1, AND N_2 IN TERMS OF THE TRUE VALUE OF N

N	Census Size	$\frac{nt}{N}$	N_0	N_1	N_2
	$n = 1{,}000$ $t = 100$	10	$1.114N$	$1.0000N$	$0.912N$
10^4	$n = 5{,}000$ $t = 100$	50	$1.010N$	$1.0000N$	$0.990N$
	$n = 5{,}000$ $t = 500$	250	$1.002N$	$1.0000N$	$0.998N$
	$n = 1{,}000$ $t = 100$	1	$1.221N$	$0.6385N$	$0.007N$
10^5	$n = 5{,}000$ $t = 100$	5	$1.348N$	$1.0000N$	$0.810N$
	$n = 5{,}000$ $t = 500$	25	$1.041N$	$1.0000N$	$0.962N$
	$n = 10{,}000$ $t = 1{,}000$	100	$1.009N$	$1.0000N$	$0.991N$
	$n = 10{,}000$ $t = 100$	1	$1.221N$	$0.6385N$	$0.006N$
	$n = 5{,}000$ $t = 500$	2.5	$1.556N$	$0.9199N$	$0.554N$
10^6	$n = 10{,}000$ $t = 1{,}000$	10	$1.129N$	$1.0000N$	$0.901N$
	$n = 20{,}000$ $t = 1{,}000$	20	$1.052N$	$1.0000N$	$0.951N$

These results indicate that any other estimate which takes the form of the reciprocal of a linear function of s must be biased when N_1 is unbiased unless the expectations of all terms, after the first, in the series (24), are negligible. In particular, setting $a = 0$, the factorial series becomes

$$\frac{1}{s} = \frac{0!}{s+1} + \frac{1!}{(s+1)(s+2)} + \cdots + \frac{(n-1)!}{(s+1)(s+2)\cdots(s+n)} + R_n\,,$$

with

$$R_n = \frac{n!}{s(s+1)\cdots(s+n)}.$$

Then writing $1 - P(0) = s(0)$,

$$E\left(\frac{nt}{s}\right) = nt\,E\left[\frac{1}{s+1} + \frac{1}{(s+1)(s+2)} + \frac{2}{(s+1)(s+2)(s+3)} + R_3/s \neq 0\right] s_0 + 2nt(1 - s_0)$$

$$= nt\left[E\left(\frac{(n+1)(t+1)}{s+1} \Big/ s \neq 0\right)\right]\left[1 + \frac{s_2(N+1)}{s_1(n+1)(t+1)} + \frac{2s_3(N+1)(N+2)}{s_2(n+1)(n+2)(t+1)(t+2)}\right] s_0 + nt\,E\left(\frac{R_3}{s \neq 0}\right) s_0 + 2nt(1 - s_0).$$

If $\frac{nt}{N} > 10$, the ratios $\frac{s_2}{s_1}\ \frac{s_3}{s_2}\ \frac{s_0}{1}$ are almost 1, and hence for large n and t we may write:

$$E\left(\frac{nt}{s}\right) \geqq N\left[1 + \frac{N}{nt} + 2\left(\frac{N}{nt}\right)^2\right]. \tag{32}$$

Equation (32) indicates that N_0 has a positive bias for a range of values of n and t for which N_1 is essentially unbiased. While the approximation (32) is no longer valid for smaller values of nt/N, some calculations have been made which show that N_0 still has a considerable positive bias for such values. This is illustrated in Table 2, which shows $E(N_0)$ for comparison with $E(N_1)$.

A formula in series form for the variance of N_1 may be set up using equations (25), (27), and (28). This series is

$$V(N_1) = \frac{s_2(N+1)(N+2)(n+1)(t+1)}{s_1(n+2)(t+2)} + \frac{s_3(N+1)(N+2)(N+3)(n+1)(t+1)}{s_2(n+2)(t+2)(n+3)(t+3)} + 2\frac{s_4(N+1)(N+2)(N+3)(N+4)(n+1)(t+1)}{s_3(n+2)(t+2)(n+3)(t+3)(n+4)(t+4)} + E(R_4') - 2Nb_1(N) - N^2.$$

If it is assumed that $P(0)$ is negligible and that the ratios

$$\frac{s_2}{s_1},\ \frac{s_3}{s_2},\ \frac{s_4}{s_3} \qquad \text{and} \qquad \frac{nt}{(n+3)(t+3)}$$

are approximately unity, then the following approximate formula is derived:

$$V(N_1) \doteq N^2\left[\frac{N}{nt} + 2\left(\frac{N}{nt}\right)^2 + 6\left(\frac{N}{nt}\right)^3\right]. \tag{33}$$

If, for example, $nt/N = 10$, the standard deviation of estimates of N about the true value is approximately $0.33N$. Sample census programs in which the expected number of tagged members in the sample is much smaller than 10 may fail to give even the order of magnitude of the population correctly.

TABLE 3

STANDARD ERROR OF THE ESTIMATE OF N BY VARIOUS FORMULAE FOR DIFFERENT SAMPLE AND POPULATION SIZES

N	Size of Census	$\frac{nt}{N}$	$\epsilon(N_1-N)^2$	$\epsilon(N_0-N)^2$
	$n = 1{,}000$ $t = 100$	10	$0.3320N$	$0.4706N$
10^4	$n = 5{,}000$ $t = 100$	50	$0.1033N$	$0.1041N$
10^5	$n = 1{,}000$ $t = 100$	1	$0.4653N$	$0.6724N$
	$n = 5{,}000$ $t = 100$	5	$0.5474N$	$0.8827N$
	$n = 5{,}000$ $t = 500$	25	$0.2017N$	$0.2241N$
10^6	$n = 10{,}000$ $t = 100$	1	$0.4639N$	$0.6724N$
	$n = 5{,}000$ $t = 500$	2.5	$0.5571N$	$1.3591N$
	$n = 10{,}000$ $t = 1{,}000$	10	$0.3364N$	$0.3992N$
	$n = 20{,}000$ $t = 1{,}000$	20	$0.2333N$	$0.2766N$
			$\epsilon(N_1-N)^2$	$\epsilon(N_2-N)^2$
10^4	$n = 5{,}000$ $t = 500$	250	$0.0459N$	$0.0457N$
10^5	$n = 10{,}000$ $t = 1{,}000$	100	$0.0954N$	$0.0940N$

The variance of N_0 about the true parameter point N is even larger than that of N_1, for if a is set equal to zero in equation (21), with the same approximations as above,

$$V(N_0) \doteq \frac{N^3}{nt} + \frac{N^4}{(nt)^2} + R^1, \qquad R^1 > 0\,.$$

The variance of various values of N, n, and t is tabulated in table 3, above, in terms of N, N_0, and N_1. For smaller values of n, t and nt/N, these have been computed directly; for larger values, they have been evaluated by means of the series approximation. Also tabulated for certain values of these parameters is the variance of a new estimate

$$N_2 = \frac{(n+2)(t+2)}{s+2}.$$

If n and t are sufficiently large that the bias of N_0 is negligible, then an examination of formulas (24) and (25) suggests trying $a = 2$. The approximate formulae for the expectation and variance of N_2 are:

$$E(N_2) \doteq (N+1)\left[\frac{(n+2)(t+2)}{(n+1)(t+1)} - \frac{N+2}{(n+1)(t+1)}\right] = N\left(1 - \frac{N}{nt}\right),$$

$$V(N_2) \doteq N^2\left[\frac{N}{nt} - \frac{N^2}{nt} - \frac{N^3}{nt}\right].$$

The estimates considered so far are essentially modifications of the ML estimate. It is desirable to consider the problem from a wider viewpoint. Barankin [1] has given a set of necessary and sufficient conditions that the class of unbiased estimates of any parameter be non-empty: further, if the class of unbiased estimates is non-empty, a method is provided to determine that member of the class which has minimum variance at some pre-determined point in the parameter space.

To apply these methods, it is necessary to define

$$\Pi_{N_i}(s) = \frac{P(s; N_i)}{P(s; N_0)},$$

where N_0 is the chosen point at which the "best" estimate is to have minimum variance and N_i is any other possible value of N. Then, in order that there be an unbiased estimate of N, it is necessary and sufficient that there exist a finite constant c such that, for every set of m functions $\Pi_{N_1}, \Pi_{N_2}, \cdots, \Pi_{N_m}$ and every set of real numbers $a_1, \cdots, a_m$, $(m = 1, 2, \cdots)$,

$$\left|\sum_{i=1}^{m} a_i N_i\right|^2 \leqq c \sum_{s=0}^{t} \left|\sum_{i=1}^{m} a_i \Pi_{N_i}(s)\right|^2 P(s; N_0). \tag{34}$$

For any N_0

$$\lim_{N_i \to \infty} \sum_{s=0} \Pi_{N_i}(s) = \frac{P(0; N_i)}{P(0; N_0)};$$

that is, Π_{N_i} is bounded as N_i tends to infinity. Since the left hand side of (34) is unbounded as N_i tends to infinity, no finite constant c exists such that (34) holds. Consequently, if the parameter space of N is the whole set of integers $1, 2, \cdots$, there is no unbiased estimate of N.

On the other hand, even if the parameter space is bounded, there exists no unbiased estimate of N, if we consider the set of equations (in the a's)

$$\begin{array}{l} a_1\Pi_{N_1}(0) + a_2\Pi_{N_2}(0) \cdots + a_m\Pi_{N_m}(0) = 0 \\ a_1\Pi_{N_1}(1) + a_2\Pi_{N_2}(1) \cdots + a_m\Pi_{N_m}(1) = 0 \\ \quad \cdot \qquad\quad \cdot \qquad\quad \cdot \qquad\quad \cdot \qquad \cdot\ \cdot \\ \quad \cdot \qquad\quad \cdot \qquad\quad \cdot \qquad\quad \cdot \qquad \cdot\ \cdot \\ \quad \cdot \qquad\quad \cdot \qquad\quad \cdot \qquad\quad \cdot \qquad \cdot\ \cdot \\ a_1\Pi_{N_1}(t) + a_2\Pi_{N_2}(t) \cdots + a_m\Pi_{N_m}(t) = 0\,, \qquad (m > t)\,. \end{array}$$

It is possible to find a set of values of $a_1 \cdots a_m$ not all zero such that these equations hold for any set of values $N_1, N_2, \cdots, N_m$. Consequently, the right hand side of (34) is zero, while in general the left hand side will not be.

It may be concluded, then, that any estimate will be biased for some range of values of N, n, and t, and that the most that can be attained in this direction is an estimate whose bias is negligible for practical purposes over a useful range of values of N.

In a paper to be published elsewhere a formula is derived which permits the determination of a class of lower bounds for the variance of an estimate N^* of N in terms of two functions which depend only upon the distribution and the bias of the estimate (and on N). These functions are:

$$\Psi_h(s;N) = \frac{P(s;N+h) - P(s;N)}{h\,P(s;N)}$$

$$\Delta_h b^*(N) = \frac{b^*(N+h) - b^*(N)}{h}\,.$$

Then $V(N^*)$ must satisfy the following inequality:

$$V(N^*) \geqq \frac{[1 + \Delta_h b^*(N)]^2}{\sigma_\psi^2}$$

for all integers h for which $P(s; N+h)$ is defined.

For $h = -1$, $\sigma_\psi{}^2$ may be evaluated directly.

$$\Psi_{-1}[s;N] = \frac{nt - Ns}{(N-n)(N-t)}\,,$$

and

$$\sigma^{-2}[\Psi_{-1}[s;N]] = \frac{(N-1)(N-n)(N-t)}{nt}\,.$$

This is a lower bound for the variance of estimates of N, the bias of which is negligible. The estimates suggested so far, N_1 (or N_2 when nt/N is sufficiently large) are of this class.

A better bound for this variance may be found by choosing some other value for h. However, for large values of h, the only feasible method of obtaining $\sigma_\psi{}^2$ is by numerical methods. These were tried for the particular values $N = 10^4$, $n = 10^3$, $t = 10^2$ for values of h over a considerable range. The maximum value of σ_ψ^{-2} so obtained was $0.1096\, N^2$, for $h = 2000$. On the other hand, for these values of N, n, and t, $V(N_1) = 0.1102\, N^2$. It appears that N_1 is at least close to being a minimum variance estimate for the range of the values of the parameters for which it is essentially unbiased.

It will frequently happen that the population falls into several sub-catagories, e.g., it may be classified according to age, sex or race, etc. It is natural to ask whether it is not more desirable to estimate the size of these sub-groups and then to add the estimated populations to get the total population size.

It may be more difficult to ensure that the condition of randomness is satisfied for each sub-group: that is, to ensure that the probability that an individual is obtained in the sample is independent of whether or not it is tagged, for each of the sub-groups whose size is to be estimated (for experimental evidence along such lines, reference is made to Howard [**8**]). However, aside from this, the estimate obtained by summing estimates of sub-groups is more variable than the population estimate obtained directly from the whole sample.

To demonstrate this, we consider the estimate of the total population obtained by estimating the males and females separately. Subscripts m and f on N, n, t, and s denote males and females respectively. Further, let $N_m = pN$, $N_f = (1 - p)N$. It may be noted that n_m, n_f, t_m, t_f are themselves random variables.

If the required randomness assumption holds for both males and females, then

$$E(n_m) = pn\,, \qquad E(n_f) = (1-p)n$$
$$E(t_m) = pt\,, \qquad E(t_f) = (1-p)t\,.$$

Now

$$N_m^* = \frac{(n_m+1)(t_m+1)}{s_m+1}\,, \qquad N_f^* = \frac{(n_f+1)(t_f+1)}{s_f+1}$$

are almost unbiased estimates of N_m, N_f, respectively, so that

$$N^* = N_m^* + N_f^*$$

is an almost unbiased estimate of N.

Using the fact that for a positive random variable x for which $E(x^n)$, $E(x^{-n})$ exist

$$E(x^n) \geqq [E(x^{-n})]^{-1}\,,$$

and assuming that N/nt is sufficiently large that the same approximations hold as were used in deriving (33) then

$$V(N^*) \geqq \frac{N^3}{nt}[p + (1-p)] + 2\frac{N^4}{(nt)^2}[1+1] + 6\frac{N^5}{(nt)^3}\left[\frac{1}{p} + \frac{1}{1-p}\right].$$

On the other hand

$$V(N_1) \doteq \frac{N^3}{nt} + 2\,\frac{N^4}{(nt)^2} + 6\,\frac{N^5}{(nt)^3} + R\,.$$

Using the expression for the remainder (26), a slight computation shows that

$$R < 24\,\frac{N^5}{(nt)^2}\,,$$

so that $V(N^*) > V(N_1)$ is immediate.

Rather than using estimates of the sub-populations to estimate the whole populations, the reverse is true. Thus, if N_1 is an unbiased estimate of the whole population size, N, then an unbiased estimate of N_m is

$$N_m^* = \frac{n_m + t_m - s_m}{n + t - s} \cdot \frac{(n+1)(t+1)}{s+1}\,, \tag{35}$$

for, assuming that (30) holds,

$$E(N_m^*) = \underset{s}{E}\left[E\left(\frac{n_m + t_m - s_m}{n+t-s} \cdot \frac{(n+1)(t+1)}{s+1} \Big/ s\right)\right]$$

$$= \underset{s}{E}\left(\frac{p(n+1)(t+1)}{s+1}\right) \doteq pN = N_m\,.$$

Further, for N/nt sufficiently large

$$V(N_m) = \underset{s}{E}\left[\left(\frac{p(1-p)}{n+t-s} + p^2\right)\left(\frac{(n+1)^2(t+1)^2}{(s+1)^2}\right)\right] - p^2N^2$$

$$= p^2V(N_1) + \frac{p(1-p)(n+1)^2(t+1)^2}{n+t+1}\,E\left[\frac{1}{(s+1)^2}\left(1 - \frac{s+1}{n+t+1} + \frac{(s+1)^2}{(n+t+1)^2}\cdots\right)\right]$$

$$\doteq p^2\left[\frac{N^3}{nt}\left(1 + 2\frac{N}{nt} + 6\left(\frac{N}{nt}\right)^2\right)\right] + \frac{p(1-p)}{n+t+1}N^2.$$

The first ratio of formula (35) is the estimate of the proportion of males. If there is evidence to show that one of the processes of the sample is not random in its selection by sex, the formula may easily be adjusted. If, for example, it is known that males predominate among those tagged but the sampling process is random in this respect, then the best estimate of the proportions of males is simply

$$N_m^{**} = \frac{n_m}{n} \cdot \frac{(n+1)(t+1)}{s+1} \doteq \frac{n_m(t+1)}{s+1}$$

The formula for the variance is simplified in this case; the variance is larger, however. Similar considerations and similar formulae apply to the estimation of the size of any other sub-group of the total population.

3. Comparison of population sizes

The comparison of populations in time or space is a test of various hypotheses concerning the unknown parameters, the population sizes. Theorem 3 provides the machinery for such tests.

The simplest case is the comparison of two populations, that is, the test of the hypothesis $N_1 = N_2 = N$, say, on the basis of information provided by sample censuses. The procedure consists of calculating N^* (the BAN estimate of N under the hypothesis tested) by means of formula (20) and then using the estimated value in one of (21), (22), or (23) to calculate χ^2. The hypothesis is rejected at the ϵ level of significance if $\chi^2 \geqq \chi^2_{\epsilon,\, k_1+k_2-1}$, where $k_1 + k_2$ is the total number of samples and $\chi^2_{\epsilon,\, k_1+k_2-1}$ is the ϵ significance point of the χ^2 distribution with $k_1 + k_2 - 1$ degrees of freedom.

If $k_1 + k_2$ independent taggings and samplings have been performed (subscripts i on n, t, and s will differentiate between these different censuses) a BAN estimate of N is found by minimizing

$$\chi^2 = \sum_{i=1}^{k_1+k_2} \frac{(n_i t_i)^2 \left(s_i - \dfrac{n_i t_i}{N}\right)^2}{s_i(n_i - s_i)(t_i - s_i)}.$$

The solution of this minimization problem, which may be of interest independently of the comparison problem, is

$$N^* = \frac{\displaystyle\sum_{i=1}^{k_1+k_2} \frac{(n_i t_i)^3}{s_i(n_i - s_i)(t_i - s_i)}}{\displaystyle\sum_{i=1}^{k_1+k_2} \frac{(n_i t_i)^2}{(n_i - s_i)\,(t_i - s_i)}}.$$

The asymptotic variance of N^* is

$$AV(N^*) = \frac{N}{\displaystyle\sum_{i=1}^{k_1+k_2} \frac{n_i t_i}{(N - n_i)(N - t_i)}}.$$

It is to be noted that all n_i and t_i are assumed to be known parameters, a condition frequently not fulfilled in multiple sample censuses.

If $k_1 = k_2 = 1$, *i.e.*, a single sample census is taken in each population, the estimate N^* may be written

$$N^* = \frac{(n_1 t_1)^3 s_2(n_2 - s_2)(t_2 - s_2) + (n_2 t_2)^3 s_1(n_1 - s_1)(t_1 - s_1)}{s_1 s_2[(n_1 t_1)^2(n_2 - s_2)(t_2 - s_2) + (n_2 t_2)^2(n_1 - s_1)(t_1 - s_1)]}.$$

This test procedure is based on the theorem giving asymptotic properties of these sums of squares. In their actual usage for finite samples and populations, an approximation is involved: the size of the critical region, the probability that the hypothesis $H: N_1 = N_2$ is rejected when it is true is not exactly equal to the predetermined value ϵ.

Since several alternative forms of χ^2 were defined in theorem 3, several alternative tests for the hypothesis $N_1 = N_2$ are available. It is of interest to ask which of these yield the closest approximation to the desired size of the critical region, or alternatively whether the degree of approximation may be improved by some further modification such as the standard "continuity correction" of the classical χ^2.

At the same time, it is desirable to study the power of these tests, that is,

$$P\,(H \text{ is rejected}/N_1 - N_2 = \delta \neq 0)$$

for different values of δ and of the other parameters.

The nature of N^* and of χ^2 and the complexity of the hypergeometric distribution place considerable obstacles in the way of a general approach. Alternatively extensive computations have been made for the case $k_1 = k_2 = 1$, giving the size of the critical region and the power of the tests for the various alternative tests for various values of N, n, and t.

Four alternative tests were considered based on the following four formulae (expressed in terms of n_1, n_2, t_1, t_2, s_1, s_2 and N^*):

$$\chi_p^2 = \sum_{i=1}^{2} \frac{\left(s_i - \dfrac{n_i t_i}{N^*}\right)^2}{n_i \dfrac{t}{N^*}\left(1 - \dfrac{t_i}{N^*}\right)\left(1 - \dfrac{n_i}{N^*}\right)},$$

$$\chi_q^2 = \sum_{i=1}^{2} \frac{n_i t_i \left(s_i - \dfrac{n_i t_i}{N^*}\right)^2}{s_i(n_i - s_i)(t_i - s_i)},$$

(36) $$\chi_{p+q}^2 = \sum_{i=1}^{2} \frac{2\left(s_i - \dfrac{n_i t_i}{N^*}\right)^2}{\dfrac{n_i t_i}{N^*}\left(1 - \dfrac{t_i}{N^*}\right)\left(1 - \dfrac{n_i}{N^*}\right) + s_i\left(1 - \dfrac{s_i}{n_i}\right)\left(1 - \dfrac{s_i}{t_i}\right)},$$

$$\chi_y^2 = \frac{1}{2}\sum_{i=1}^{2} \frac{\left(s_i + \lambda_i - \dfrac{n_i t_i}{N^{**}}\right)^2}{\dfrac{n_i t_i}{N^{**}}\left(1 - \dfrac{t_i}{N^{**}}\right)\left(1 - \dfrac{n_i}{N^{**}}\right)} + \frac{1}{2}\chi_p^2,$$

where

$$\begin{aligned} \lambda_i &= 1 \qquad s_i < \frac{n_i t_i}{N^*}, \\ & \qquad\qquad\qquad\qquad (i = 1, 2), \\ \lambda_i &= -1 \qquad s_i > \frac{n_i t_i}{N^*}, \end{aligned}$$

and N^{**} is the estimate based on the random variables $s_1 + \lambda_1$, $s_2 + \lambda_2$.

The choice of an unweighted average of σ_p^2 and σ_q^2 in formula (36) was purely on the basis of convenience. The fourth formula is the analogue of the "corrected" χ^2 formula of Yates [**18**]. The rejection regions of each test for the range of values of s with significant probability were determined; the various probabilities of s were evaluated by the direct formula (using Stirling's approximation for the factorials), except where comparisons showed that the normal approximation to the hypergeometric gave the required degree of accuracy.

TABLE 4

COMPARISON OF DIFFERENT χ^2 TESTS: EXACT SIZE OF CRITICAL REGION WHEN $\epsilon = 0.05$

N	Size of Census	$\frac{n_1t_1}{N}$	$\frac{n_1t_1}{n_2t_2}$	Size of Critical Region		
				χ^2_q	χ^2_p	χ^2_{p+q}
10^4	$n_1 = 1{,}000$ $n_2 = 1{,}000$ $t_1 = 100$ $t_2 = 100$	10	1	0.051	0.083	0.063
10^5	$n_1 = 10{,}000$ $n_2 = 10{,}000$ $t_1 = 500$ $t_2 = 500$	50	1	0.051	0.059	.....
10^6	$n_1 = 10{,}000$ $n_2 = 10{,}000$ $t_1 = 1{,}000$ $t_2 = 1{,}000$	10	1	0.051	0.087	.....
10^4	$n_1 = 1{,}000$ $n_2 = 1{,}000$ $t_1 = 100$ $t_2 = 200$	10	2	0.053	0.070	0.057
10^4	$n_1 = 1{,}000$ $n_2 = 1{,}000$ $t_1 = 100$ $t_2 = 500$	10	5	0.064	0.062	0.059
10^4	$n_1 = 1{,}000$ $n_2 = 5{,}000$ $t_1 = 100$ $t_2 = 500$	10	25	0.066	0.052	0.054

In table 4, above, are listed the exact probabilities with which various tests reject H, when it is true and when the desired significance level is 0.05. These probabilities, which represent the exact size of the critical regions, are tabulated for several values of the parameters.

It soon became apparent in the calculations that the test based on χ^2_v represented no clearcut gain over the simpler ones. Consequently, the required probabilities associated with this test were calculated in a few cases only.

Furthermore, it is evident, for the range of N considered, that the absolute magnitudes of N_1, N_2, n_1, n_2, t_1, t_2 play a lesser role than the relative sizes of $\frac{n_1t_1}{N_1}$ and $\frac{n_2t_2}{N_2}$.

For equal sample censuses where $n_1t_1 = n_2t_2$, the χ^2_q test appears to be the most satisfactory of the tests considered, relative to the goodness of the approximation of the size of the critical region. On the other hand, for widely disparate sample censuses, the critical region defined by the χ^2_p formula provides the best approximation. For moderate differences in size between the sample censuses, the χ^2_{p+q} test appears to be more satisfactory than either of the other two.

The power of different tests for various values of N_2 and of n_1, n_2, t_1, t_2 is shown in table 5 below.

No clear decision between the different tests is indicated on the basis of their power function. When the expected tag ratio is as small as 10, the tests are generally insensitive to large differences between N_1 and N_2. The power appears to be dependent on both the bias and the direction of the alternative hypothesis from the true hypothesis. These computations indicate that any choice between the different tests should be made, primarily, on the basis of how well the test approximates the desired critical region.

TABLE 5

POWER OF DIFFERENT χ^2 TESTS OF THE HYPOTHESIS $N_1 = N_2$ FOR VARIOUS VALUES OF N_2 AND FOR DIFFERENT CENSUS SIZES WHEN $N_1 = 10{,}000$, $\epsilon = 0.05$

Size of Census	Test Used	N_2					
		5,000	9,000	10,000	11,000	20,000	30,000
$n_1 = 1{,}000$ $t_1 = 100$ / $n_2 = 1{,}000$ $t_2 = 100$	χ^2_q	0.533	0.058	0.051	0.059	0.273	0.469
	χ^2_{p+q}	0.536	0.071	0.063	0.070	0.305	0.502
	χ^2_p	0.569	0.093	0.083	0.091	0.355	0.567
$n_1 = 1{,}000$ $t_1 = 100$ / $n_2 = 1{,}000$ $t_2 = 200$	χ^2_q	0.616	0.054	0.053	0.070	0.445	0.664
	χ^2_{p+q}	0.650	0.065	0.057	0.068	0.383	0.627
	χ^2_p	0.695	0.083	0.070	0.079	0.415	0.696
$n_1 = 1{,}000$ $t_1 = 100$ / $n_2 = 5{,}000$ $t_2 = 500$	χ^2_p	1.000	.....	0.052	.....	0.569	0.845
	χ^2_{p+q}	1.000	.....	0.054	.....	0.458	0.752
$n_1 = 5{,}000$ $t_1 = 500$ / $n_2 = 100$ $t_2 = 100$	χ^2_p	0.846	0.079	0.052	0.053	0.398	0.719
	χ^2_{p+q}	0.796	0.063	0.054	0.070	0.496	0.800

Another slightly more complex comparison problem will frequently arise: a comparison of ratios of populations. For example, a comparison of the ratio of the survivors of a migration to the original population will be useful in order to evaluate the effects on the species in question under different nature- or man-imposed regimes. In some situations, it may be possible to mark the population only once and sample it both before and after the migration. If the migration (or more generally any intervening event) involves a considerable period of time, or takes place over a large area, this procedure must be regarded as possibly unsatisfactory because the required randomness assumption may not hold. In particular, during the migration period, there may be a differential mortality between marked and unmarked members or tags may be lost over a period of time. In some cases, the comparison will involve different generations. In general, then there will be four separate censuses involving four populations: N_1, N_2, N_3, N_4. The hypothesis to be tested is that

$$\frac{N_3}{N_1} = \frac{N_4}{N_2} = r \,. \tag{37}$$

As before, the χ^2 test procedure requires, first, the estimation of the unknown parameters. Under the hypothesis tested these parameters reduce to three, namely,

N_1, N_2, and r. Minimizing χ_3^2 (formula (18)), under the restriction (37) is possible; however, a sixth degree equation in r is obtained.

A simpler procedure may be found in the modified χ^2 method, that is, by "linearizing" the restriction upon the parameters. It is more convenient to write

$$p_i = \frac{t_i}{N_i}, \qquad q_i = \frac{s_i}{n_i} \qquad (i = 1, 2, 3, 4),$$

whence equation (37) becomes

$$\frac{p_1 p_4}{t_1 t_4} = \frac{p_2 p_3}{t_2 t_3}. \tag{38}$$

Any of the p_i's may be expressed in terms of the remaining three p's. From the point of view of the asymptotic theory, the choice is arbitrary. However, a rule that is suggested by general considerations is: if q_k is the largest of the q's, equation (38) can be expressed so that p_k is in the denominator.

To set down the general procedure, p_4 is expressed in terms of $p_1 p_2 p_3$, to wit,

$$p_4 = \frac{t_1 t_4}{t_2 t_3} \frac{p_2 p_3}{p_1}.$$

The modified restriction on the p's is:

$$p_4 = \frac{t_1 t_4}{t_2 t_3}\left[\frac{q_2 q_3}{q_1 q_4} - (p_1 - q_1)\frac{q_2 q_3}{q_1^2} + (p_2 - q_2)\frac{q_3}{q_1} + (p_3 - q_3)\frac{q_2}{q_1}\right].$$

The minimization leads to three linear equations in p_1, p_2, and p_3, which may be written more shortly by letting

$$\frac{n_i}{q_i(1 - q_i)\left(1 - \frac{n_i q_i}{t_i}\right)} = b_i, \qquad (i = 1, 2, 3, 4); \qquad \frac{q_2 q_3}{q_1}\frac{t_1 t_4}{t_2 t_3} = c.$$

These three equations are:

$$p_1\left(b_1 q_1 - \frac{c^2 b_4}{q_1}\right) - p_2\left(\frac{c^2 b_4}{q_2}\right) - p_3\left(\frac{c^2 b_4}{q_3}\right) = b_1 q_1^2 + b_4 c q_4,$$

$$p_1\left(\frac{c^2 b_4}{q_1}\right) + p_2\left(b_2 q_2 - \frac{c^2 b_4}{q_2}\right) - p_3\left(\frac{c^2 b_4}{q_3}\right) = b_2 q_2^2 - b_4 c q_4,$$

$$p_1\left(\frac{c^2 b_4}{q_1}\right) - p_2\left(\frac{c^2 b_4}{q_2}\right) + p_3\left(b_3 q_3 - \frac{c^2 b_4}{q_3}\right) = b_3 q_3^2 - b_4 c q_4.$$

After the estimates of the parameters are determined from these three equations and from equation (38) the test procedure is similar to that outlined above.

Conclusion

While the properties of the hypergeometric distribution derived in this paper have been discussed primarily in their application to the census sample problem, it is evident that they may have application in many other problems that involve this distribution.

Furthermore, the procedure may be useful not only in zoological censuses but also in other fields, for example, in astronomy, oceanography, etc. and in dealing with human sampling. Many surveys are made to determine the total number of "customers" or "listeners," or populations of areas in intercensal years. The usual sampling scheme to determine such totals may be unsatisfactory since there is no assurance that the survey is random in respect to the characteristic in question. In many cases, a "tagging" procedure would in general be more expensive than the standard survey method and would give rise to estimates with greater variance. Its usefulness in such situations would then seem to be in the nature of a check procedure on the standard method; it may also be of value in some preliminary studies where the steps necessary to assure randomness in the standard procedure have not been determined.

Consequently, the methods and the formulae of this paper may be of interest in determining the size of subgroups of known populations as well as the problem for which it has been specifically designed: determining the size of unknown populations where a total count is not feasible.

[The author wishes to express his thanks to Professor J. Neyman for advice in the preparation of this paper.]

REFERENCES

[1] E. W. Barankin, "Locally best unbiased estimates," *Annals of Math. Stat.*, vol. 20 (1949), pp. 477–501.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[3] F. N. David, "Limiting distributions connected with certain methods of sampling human populations," *Stat. Res. Mem.*, vol. II (1938), pp. 69–90.

[4] L. R. Dice, "Methods for estimating populations of mammals," *Jour. of Wildlife Management*, vol. 5 (1941), pp. 398–407.

[5] L. M. Graves, *Theory of Functions of Real Variables*, McGraw-Hill, New York, 1946.

[6] R. C. Green and C. A. Evans, "Studies on a population cycle of snowshoe hares on the Lake Alexander area; 1. Gross annual census 1932–39," *Jour. of Wildlife Management*, vol. 4 (1940), pp. 220–238.

[7] J. Gurland, *Best Asymptotically Normal Estimates*. Unpublished thesis. University of California, 1948.

[8] R. Howard, "A study of the tagging method in the enumeration of sockeye salmon populations," *International Pacific Salmon Fisheries Commission Bulletin*, no. 2 (1949).

[9] C. H. N. Jackson, "The analysis of a tsetse-fly population, III," *Annals of Eugenics*, vol. 14 (1948), pp. 91–108 (contains a bibliography of earlier articles on this subject by the same author).

[10] C. Jordan, *Calculus of Finite Differences*, Budapest, 1939.

[11] F. C. Lincoln, "Calculating waterfowl abundance on the basis of banding returns," *U. S. Dept. Agric. Circ.* 118: 1–4 (1930).

[12] J. M. Milne-Thomson, *The Calculus of Finite Differences*, Macmillan, London, 1933.

[13] J. Neyman, "Contribution to the theory of the χ^2 test," *Proceedings of the Berkeley Symposium. on Mathematical Statistics and Probability, 1945–1946*, University of California Press, 1949, pp. 239–273.

[14] W. E. Ricker, "Some applications of statistical methods to fisheries problems," *Biometrics Bulletin I* (1945), pp. 73–79.

[15] M. B. Schaeffer, *The Employment of Marked Numbers in the Estimation of Animal Populations with some Considerations of Sampling Theory*. Unpublished manuscript (1948).

[16] F. F. Stephan, "The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate," *Annals of Math. Stat.*, vol. 16.

[17] E. J. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge University Press, Cambridge, 1940.

[18] F. Yates, "Contingency tables involving small numbers and the χ^2 test," *Supp. Jour. Roy. Stat. Soc.* 1 (1934), pp. 217–235.