STUDIES IN CAPTURE-RECAPTURE ANALYSIS

by

J. N. Darroch.

Thesis

submitted in support of the application

for the degree of

DOCTOR OF PHILOSOPHY

to the

University of Cape Town.

(1960)

PREFACE.

This thesis comprises three papers, two on the multiple-recapture census which have been published in Biometrika, and a third on the two-sample census which is being submitted for publication to the same journal.

As an introduction to the subject as a whole, the three separate introductions are best read in the order: last, first, second.

J.N.Darroch.

[*From* Biometrika, *Vol.* 45. *Parts* 3 *and* 4, *December* 1958.]
[*All rights reserved*] PRINTED IN GREAT BRITAIN

THE MULTIPLE-RECAPTURE CENSUS

I. ESTIMATION OF A CLOSED POPULATION

BY J. N. DARROCH

Department of Mathematics, University of Cape Town

1. INTRODUCTION

1·1. A primary classification of the many problems which can now be included under the heading of capture-recapture analysis is the one separating the census which uses multiple recaptures from the census which does not.

The best example of the latter type is the 'fisheries census', where the main catching is done commercially and the experimenter's job is to keep the population supplied with tagged individuals. In this case, recaptured fish are obviously not available to him for retagging and all he can hope for is that the captured tags are returned to him. Most of the paper by Chapman (1954) is devoted to this type of situation and contains some very simple estimates derived by the use of large-sample large-population Poisson approximations. Gulland (1955) shows how, if the catching is considered as a continuous-time process with constant effort, the natural and fishing mortality *rates* can be estimated from the behaviour of tagged fish alone, that is without any knowledge of total catches. It follows of course that with this knowledge estimates of population size are also available.

In the multiple-recapture census it is usually the experimenter who does both tagging and sampling and it is assumed that he employs a method of capture which does not kill the animal or affect its future behaviour. The experiment then comprises a sequence of samples $S_1, \ldots, S_s$, say, where the members of $S_1, \ldots, S_{s-1}$ are all tagged before being returned to the population, while the members of $S_2, \ldots, S_s$ are classified according to when, if at all, they have been captured before. The majority of papers have discussed this census and among them may be mentioned Bailey (1951); Chapman (1951, 1952); Craig (1953); Goodman (1953); Hammersley (1953); Leslie & Chitty (1951); Leslie (1952) and Moran (1952). In all of these papers except Goodman's s is a constant. Goodman sets up a model in which the number of samples sequentially depends on the total number of recaptured tags, which is stipulated beforehand. As far as the individual sample sizes are concerned, we notice that everywhere except in Hammersley's paper, each sample S_i is completed when one of its statistics attains a prescribed value. This statistic is usually simply the sample size, but in what has come to be known as the inverse sample census it is the number of tagged or the number of untagged individuals recovered. In this connexion see Bailey (1951) and Chapman (1952). It goes without saying that the theory of all these papers can be applied to the estimation of the number of classes in a population if the classes are of equal size and sampling is with replacement. The number of classes represented in a sample constitutes its size and a class is 'recaptured' when it is represented in a subsequent sample.

The latest extensions to the general problem have been made by Chapman who exploits the natural stratification of animal populations, with respect to type (sex, species) of individual (1955) and with respect to place (1956, with Junge).

1·2. In the present paper we treat the multiple-recapture census for which the number of samples s is fixed (except in §§ 4·4, 4·5 and 5·6). In §§ 3 and 4 the sample sizes are regarded as constants and in § 5 as binomial variables.

Most of the above-mentioned work on the multiple-recapture census has been applied to closed populations in which there is neither departure due to death or emigration, nor augmentation due to birth or immigration. The restriction to a closed population also prevails here, but in a second paper we shall take account of both departure and augmentation.

1·3. To extract all the information from a multiple-recapture experiment, tagging must be *differentiated* in order that the full 'history' of any individual can be inferred each time it is captured. This can be effected in two ways. Either an individual is given a numbered tag at its first appearance, or, each time it is captured it is given a new mark distinctive of that capture. For some purposes, however, *similar* tagging is sufficient, where all that is required is that each individual bears a mark after being captured. It need not be remarked when recaptured.

2. THE ALTERNATIVE MODELS

2·1. Let n be the total number of individuals in the population.

Let s be the total, fixed, number of samples taken.

Let u_i be the number of individuals caught in the ith sample but not otherwise, u_{ij} the number caught in the ith and jth samples but not otherwise and similarly u_{ijk}, etc.

Denoting a subset of the integers $1, 2, \ldots, s$ by ω, let

$$r = \sum_{\omega} u_{\omega} = \sum_{i} u_i + \sum_{i<j} u_{ij} + \ldots + u_{1,2,\ldots,s},$$

the total number of different individuals caught in the complete experiment.

Let a_i be the size of the ith sample. Then $a_i = \sum_{\omega \supset i} u_{\omega}$ where summation is over all subsets ω which include the integer i.

We derive two probability distributions for $\{u_{\omega}\}$.

Let the probability that any individual is caught in the ith sample be $p_i = 1 - q_i$. Thus we assume that all individuals are equally likely to be members of any given sample. Further, we shall assume that, for any individual, the events: caught in the ith sample, $i = 1, 2, \ldots, s$, are independent.

The probability of any individual escaping capture throughout the experiment is

$$\prod_{i} q_i = Q, \quad \text{say.}$$

The probability of being caught in the $i, \ldots, l$ samples and no others is

$$\frac{p_i}{q_i} \ldots \frac{p_l}{q_l} Q = P_{i,\ldots,l}, \quad \text{say.}$$

Clearly, the probability density of $\{u_{\omega}\}$ is multinomial, viz.

$$p[\{u_{\omega}\}] = \frac{n!}{(n-r)! \prod_{\omega} u_{\omega}!} Q^{n-r} \prod_{\omega} P_{\omega}^{u_{\omega}}, \tag{A}$$

where $0 \leqslant u_{\omega} \leqslant n$ subject to $0 \leqslant r = \sum_{\omega} u_{\omega} \leqslant n$. We notice that

$$Q^{n-r} \prod_{\omega} P_{\omega}^{u_{\omega}} = Q^n \prod_{\omega} \left(\frac{P_{\omega}}{Q}\right)^{u_{\omega}} = Q^n \prod_{i} \left(\frac{p_i}{q_i}\right)^{\sum_{\omega \supset i} u_{\omega}} = \left(\prod_{i} q_i\right)^n \prod_{i} \left(\frac{p_i}{q_i}\right)^{a_i} = \prod_{i} p_i^{a_i} q_i^{n-a_i}.$$

Therefore, (A) may also be written

$$p[\{u_\omega\}] = \frac{n!}{(n-r)!\prod_\omega u_\omega!}\prod_i p_i^{a_i} q_i^{n-a_i}. \qquad (A')$$

We now find $p[\{u_\omega\} \mid \{a_i\}]$, the conditional density of $\{u_\omega\}$ given $\{a_i\}$. It is obvious from first considerations, and is easily deduced from (A), that the a_i are independent binomial variables $B[n, p_i]$. Therefore

$$p[\{a_i\}] = \prod_i \binom{n}{a_i} p_i^{a_i} q_i^{n-a_i}$$

and

$$p[\{u_\omega\} \mid \{a_i\}] = \frac{n!}{(n-r)!\prod_\omega u_\omega!}\prod_i \binom{n}{a_i}^{-1}, \qquad (B)$$

where

$$\max\{a_i\} \leqslant r \leqslant \sum_i a_i \quad (\text{strictly, } \min(n, \sum_i a_i)),$$

$$0 \leqslant u_i \leqslant a_i, \quad 0 \leqslant u_{ij} \leqslant \min(a_i, a_j), \quad \text{etc.},$$

with the linear constraints on the u_ω

$$\sum_{\omega \supset i} u_\omega = a_i \quad (i = 1, 2, \ldots, s).$$

(B) is a generalized hypergeometric density.

2·2. Which of models (A) and (B) is appropriate to any given experiment?

In (A) the sample sizes a_i are random variables while the p_i are parameters. This model is therefore applicable when the effort put into the catching of every sample is fixed before the experiment begins since the p_i are then fixed, though unknown. (B), on the other hand, involves the a_i as parameters and should be used only when the experimenter is determined to catch no more and no less than a_i individuals at the ith sample; and he will only be able to do this when animals are fairly easily caught. In fact, if we had to generalize, we could say that (B) is likely to be appropriate when the main limiting factor on sample size is the trouble involved in marking animals and (A) when it is the difficulty in catching them.

Most previous work has been based on (B) and (A) is new. (Hammersley (1953) constructed a model in which the a_i are binomial variables but this model involves a flaw which invalidates the estimation based on it, as we shall show in paper II.) It has been customary to derive (B) as a chain of $s-1$ hypergeometric probabilities $P[S_i \mid S_1, S_2, \ldots, S_{i-1}]$, and this has led to its simplicity being obscured either by the notation employed, by considering only the terms involving n or by making sampling-with-replacement approximations.

As well as being the exact probability description of the capture-recapture experiment when the a_i are constants, (B) may also be regarded as a very useful device for eliminating the nuisance parameters p_i when the a_i are variables; it leaves only n to be estimated and provides a sufficient statistic for n, namely r. One feels intuitively that to estimate n *as if* the a_i are constants, when in fact they are not, is not a serious misrepresentation, and this feeling is strengthened by the discovery that the two models lead to the same estimate $\tilde{n}$ of n, and to the same asymptotic estimate of $\operatorname{var}(\tilde{n})$. Apart from demonstrating this, it may be wondered why there is any need to consider (A) at all. The main reason is that (A) is capable of generalizations which (B) is unable to accommodate and it is necessary to discuss (A) for the closed population before going on to these generalizations, some of which are the subject of paper II. Also, the ease with which a multinomial probability is manipulated gives it considerable advantage over a hypergeometric probability, even for a closed population.

3. ESTIMATION USING MODEL B

3·1. The fact that r is sufficient for n has the important implication that n can be estimated from similar tagging.

Regarding (B) as the likelihood $L(n)$ of n, and omitting constant terms,

$$\log L(n) = \sum_i \log (n-a_i)! - (s-1)\log n! - \log (n-r)!.$$

An equation for the maximum likelihood estimate $\tilde{n}$ of n can be found by equating $\Delta \log L(n)$ to zero. This involves an error of less than unity in the solution and is equivalent to the 'ratio method' of maximizing L, which equates $L(n)$ to $L(n-1)$. Since $\Delta \log n! = \log n$, $\tilde{n}$ must be one of the roots of

$$\prod_i (n-a_i) = n^{s-1}(n-r). \tag{1}$$

(1) has a single finite root greater than r which maximizes the likelihood, except when r takes one of its extreme values. (i) If $r = \sum_i a_i$, no individual is observed more than once and $\tilde{n}$ is infinite. (ii) If $r = \max\{a_i\} = a_m$ say, no individual is observed which does not appear in the mth sample and $\tilde{n} = r = a_m$. It is of course in the nature of the capture-recapture experiment that (i) and (ii) are extremely unlikely to occur.

(1) may also be obtained by equating r to its expected value ρ, say. For

$$n-\rho = E[n-r] = \frac{\prod_i (n-a_i)}{n^{s-1}}. \tag{2}$$

This follows from the identity in n and the a_i

$$\prod_i \binom{n}{a_i}^{-1} \sum_{r,\{u_\omega\}} \frac{n!}{(n-r)!\prod_\omega u_\omega!} = 1,$$

since

$$E[n-r] = \frac{\prod_i (n-a_i)}{n^{s-1}} \prod_i \binom{n-1}{a_i}^{-1} \sum_{r,\{u_\omega\}} \frac{(n-1)!}{(n-1-r)!\prod_\omega u_\omega!}$$

$$= \frac{\prod_i (n-a_i)}{n^{s-1}}.$$

Similarly,

$$E[(n-r)(n-r-1)] = \frac{\prod_i (n-a_i)(n-a_i-1)}{n^{s-1}(n-1)^{s-1}}, \tag{3}$$

with corresponding expressions for the higher factorial moments of $n-r$.

3·2. To apply maximum likelihood large-sample theory in finding the variance of an estimate, it is necessary that the following three conditions are fulfilled. (*a*) The sample size must be a constant. (*b*) The likelihood must consist merely of the product of the individual likelihoods for the separate sample members. (*c*) The range of summation of the random variables must be independent of the unknown parameters. Except for one model discussed in § 5·7, which is artificially constructed for the purpose, no other model of this paper satisfies these three conditions. In the present context r is the sample size, as distinct from the a_i which are the separate catch sizes, and (B) obviously does not satisfy (a) or (b). (It does satisfy condition (c) provided $n > \sum_i a_i$.) Model (A) breaks all three conditions.

We can, however, use the 'δ-technique' to find the asymptotic variance and bias of $\tilde{n}$. (1) may be regarded as defining $\tilde{n}$ as a function $\tilde{n}(r)$ of r. By (2), $\tilde{n}(\rho) = n$ and we may therefore expand $\tilde{n}$ about n as a Taylor series in powers of $r-\rho$. If we consider $\tilde{n}$ and r as continuous variables we can say that $d\tilde{n}/dr$ is finite and differentiable in the range

$$a_m \leqslant r \leqslant \sum_i a_i - 1.$$

Confining attention to this range, that is ignoring the possibility of (i) occurring, we have

$$\tilde{n} - n = (r-\rho)\left[\frac{d\tilde{n}}{dr}\right]_\rho + \tfrac{1}{2}(r-\rho)^2\left[\frac{d^2\tilde{n}}{dr^2}\right]_{r'}, \tag{4}$$

where r' lies between r and ρ. Differentiating (1)

$$\left[\frac{d\tilde{n}}{dr}\right]_\rho = \frac{1}{n-\rho}\left[\frac{1}{n-\rho} + \frac{s-1}{n} - \sum_i \frac{1}{n-a_i}\right]^{-1} = O\left(\frac{1}{n}\right)O(n) = O(1),$$

where $f(n) = O(\phi(n))$ means $|f(n)| < K\phi(n)$ as $n \to \infty$ *and* each $a_i \to \infty$ in such a way that the a_i/n, and hence also ρ/n, are constant. Also

$$\frac{d^2\tilde{n}}{dr^2} = O\left(\frac{1}{n}\right) \quad \text{and} \quad \frac{d^3\tilde{n}}{dr^3} = O\left(\frac{1}{n^2}\right).$$

Further, $\qquad \operatorname{var}(r) = \operatorname{var}(n-r) = E[(n-r)(n-r-1)] + (n-\rho) - (n-\rho)^2.$

Therefore, using (3)

$$\operatorname{var}(r) \sim (n-\rho)^2\left[\frac{1}{n-\rho} + \frac{s-1}{n} - \sum_i \frac{1}{n-a_i}\right] = O(n), \tag{5}$$

where $f(n) \sim \phi(n)$ denotes that $f(n) = \phi(n)[1 + O(n^{-1})]$. Making further use of factorial moments we find that

$$E[(r-\rho)^3] = O(n) \quad \text{and} \quad E[(r-\rho)^4] = O(n^2).$$

Squaring (4) and taking expected values

$$E[(\tilde{n}-n)^2] \sim \operatorname{var}(r)\left[\frac{d\tilde{n}}{dr}\right]_\rho^2.$$

(The error in replacing $E[(r-\rho)^2]$ over the restricted range by $\operatorname{var}(r)$ is $O(n^2c^n)$, $0 < c < 1$, which is $o(n^{-1})$.) Thus, for the limit process stated

$$E[(\tilde{n}-n)^2] \sim \left[\frac{1}{n-\rho} + \frac{s-1}{n} - \sum_i \frac{1}{n-a_i}\right]^{-1} = O(n). \tag{6}$$

Let $\beta = E[\tilde{n}] - n$, the positive bias of $\tilde{n}$. Then extending the Taylor series by one term and taking expected values, we find that

$$\beta \sim \tfrac{1}{2}\operatorname{var}(r)\left[\frac{d^2\tilde{n}}{dr^2}\right]_\rho = O(1),$$

whence

$$\beta \sim \frac{1}{2}\frac{\left[\frac{s-1}{n} - \sum_i \frac{1}{n-a_i}\right]^2 + \left[\frac{s-1}{n^2} - \sum_i \frac{1}{(n-a_i)^2}\right]}{\left[\frac{1}{n-\rho} + \frac{s-1}{n} - \sum_i \frac{1}{n-a_i}\right]^2}. \tag{7}$$

Since $\beta = O(1)$ and $E[(\tilde{n}-n)^2] = O(n)$, $E[(\tilde{n}-n)^2] \sim \operatorname{var}(\tilde{n})$. Thus, it makes no difference whether we speak of mean square error or of variance and (6) is equivalent to

$$\operatorname{var}(\tilde{n}) \sim \left[\frac{1}{n-\rho}+\frac{s-1}{n}-\sum_i \frac{1}{n-a_i}\right]^{-1}. \tag{8}$$

3·3. When $s = 2$, $r = a_1+a_2-u_{12}$ and (B), written in standard hypergeometric form, is

$$\frac{\binom{a_1}{u_{12}}\binom{n-a_1}{a_2-u_{12}}}{\binom{n}{a_2}}.$$

(1) is a linear equation with solution $\tilde{n} = a_1 a_2/u_{12}$, the familiar Peterson estimate.

Chapman (1951) showed that

$$n' = \frac{(a_1+1)(a_2+1)}{u_{12}+1} - 1$$

is preferable to $\tilde{n}$, since it is always finite and is almost unbiased. This could very nearly be inferred from (7). For that formula gives $(n-a_1)(n-a_2)/(a_1 a_2)$ as the approximate bias of $\tilde{n}$, which is estimated by $(a_1-u_{12})(a_2-u_{12})/u_{12}^2$, and

$$\tilde{n}-n' = \frac{(a_1-u_{12})(a_2-u_{12})}{u_{12}(u_{12}+1)}.$$

We notice that n' is the solution of $(n-a_1)(n-a_2) = (n+1)(n-r)$, but unfortunately it is not true that $\prod_i (n-a_i) = (n+1)^{s-1}(n-r)$ yields an almost unbiased estimate for general values of s.

Chapman (1952) showed how n' can be made the basis of almost unbiased s-sample estimation. We shall wait until § 5·3 to comment on his recommendations, as they can be more easily discussed for model (A) than for (B).

3·4. Turning attention now from point estimation to confidence interval estimation of n, we assume that r is approximately normally distributed about ρ. We have already noted that r has moments $\mu_2 = O(n)$ and $\mu_3 = O(n)$. Therefore

$$\gamma_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = O\left(\frac{1}{\sqrt{n}}\right).$$

Also, one finds that

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = O(n^{-1}).$$

The expected value of r, regarded as a function of n, is

$$\rho(n) = n - \frac{\prod_i (n-a_i)}{n^{s-1}},$$

and in this notation, the equation for $\tilde{n}$ is $\rho(\tilde{n}) = r$ or $\tilde{n} = \rho^{-1}(r)$, say. Let $\sigma^2(n)$ denote the variance of r, given by (5). Then

$$P[r-k\sigma(n) \leqslant \rho(n) \leqslant r+k\sigma(n)] = 1-\epsilon,$$

where $k = k(\epsilon)$ may be read from normal tables. The inequalities are approximately

$$r-k\sigma(\tilde{n}) \leqslant \rho(n) \leqslant r+k\sigma(\tilde{n})$$

or

$$r_1 \leqslant \rho(n) \leqslant r_2, \quad \text{say,}$$

or, since $\rho^{-1}(r)$ can be shown to be a monotone increasing function of r,

$$\rho^{-1}(r_1) \leqslant n \leqslant \rho^{-1}(r_2).$$

$\rho^{-1}(r_1)$ and $\rho^{-1}(r_2)$ (which are precisely the same as $\tilde{n}(r_1)$ and $\tilde{n}(r_2)$) may be regarded as first approximations to the solutions for n of $r-k\sigma(n) = \rho(n)$ and $r+k\sigma(n) = \rho(n)$, respectively. We may now, if we wish, proceed to better approximations n_1^* and n_2^* say, obtaining

$$n_1^* \leqslant n \leqslant n_2^*$$

as the $100(1-\epsilon)$ % confidence interval for n.

4. ONE INDIVIDUAL PER CAPTURE

4·1. When each sample is of size one, (B) is the obvious probability model to use, though (A) can be adapted for the purpose as we shall show in § 5·6. (B) is the basis of the present section and is equal to

$$\frac{1}{n^s}\frac{n!}{(n-r)!}\frac{1}{\prod\limits_{\omega} u_\omega !}, \tag{B_1}$$

where $\prod\limits_{\omega} u_\omega ! = 1$, since every $u_\omega = 0$ or 1 and therefore every $u_\omega ! = 1$.

Summing (B_1) over all values of $\{u_\omega\}$ such that $\sum\limits_i u_i = f_1$, $\sum\limits_{i<j} u_{ij} = f_2$, ..., we obtain

$$\frac{1}{n^s}\frac{n!}{(n-r)!}\frac{s!}{(1!)^{f_1}(2!)^{f_2}\dots f_1!f_2!\dots} \tag{9}$$

as the probability of not catching $n-r$ individuals and of catching f_x x times where $x = 1, 2, \dots, s$ and $\sum\limits_x f_x = r$, $\sum\limits_x xf_x = s$. The step from (B_1) to (9) can be made by considering the number of ways of distributing s balls in r cells in such a way that none is empty. The argument, which need not be included here, follows from putting $u_i = 1$ if the ith ball is alone, $u_{ij} = 1$ and $u_i = u_j = 0$ if it is with the jth and no other, etc.

Summing (9) over all values of $\{f_x\}$ such that $\sum\limits_x f_x = r$ and $\sum\limits_x xf_x = s$, we obtain (Jordan, 1947, p. 206)

$$\frac{1}{n^s}\frac{n!}{(n-r)!}\sigma_s^r \tag{10}$$

as the probability of catching r individuals with s samples, where $\sigma_s^r = \Delta^r(0^s)/r!$, a Stirling number of the second kind. (10) was found by Craig (1953) when considering the estimation of a population of butterflies.

4·2. For the purpose of estimating n little alteration is required to the general results of §§ 3·2, 3·4.

$\tilde{n}$ is the solution of
$$(n-1)^s = n^{s-1}(n-r), \tag{11}$$

which may be approximately written

$$e^{-s/n} = 1-r/n.$$

The appropriate limit process is now $n\to\infty$, $s\to\infty$ such that s/n is constant, and we find that

$$\operatorname{var}(\tilde{n}) \sim n[e^{s/n}-1-s/n]^{-1} = O(n),$$

$$\beta \sim \frac{1}{2}\frac{s^2}{n^2}[e^{s/n}-1-s/n]^{-2} = O(1).$$

Confidence interval estimation may be performed with

$$\rho(n) = n[1-e^{-s/n}], \quad \sigma^2(n) = ne^{-2s/n}[e^{s/n}-1-s/n].$$

4·3. Suppose now that instead of s being fixed and r variable, sampling is continued until a fixed number, r, of individuals have been caught. Then

$P[r-1$ individuals in $s-1$ samples and a new one at the sth sample$]$

$$= \frac{1}{n^{s-1}}\frac{n!}{(n-r+1)!}\sigma_{s-1}^{r-1}\frac{n-r+1}{n}$$

$$= \frac{1}{n^s}\frac{n!}{(n-r)!}\sigma_{s-1}^{r-1}, \tag{12}$$

where $s = r, r+1, \ldots$.

This model will be referred to as *inverse*, since the term *sequential* has already been given by Goodman to his census which we discuss in § 4·4. We remark that the maximum likelihood estimate of n remains the same as for the *direct* model (10).

Let $\phi(t)$ be the probability generating function of s. Now

$$\sum_{s=r}^{\infty}\sigma_s^r t^s = \frac{t^r}{(1-t)(1-2t)\ldots(1-rt)},$$

(Jordan, p. 175). It follows that

$$\phi(t) = \frac{(n-1)!}{(n-r)!}\frac{t^r}{(n-t)(n-2t)\ldots(n-(r-1)t)}.$$

Differentiating $\phi(t)$,

$$E[s] = n\sum_{k=0}^{r-1}\frac{1}{n-k}, \tag{13}$$

$$\operatorname{var}(s) = n\sum_{k=1}^{r-1}\frac{k}{(n-k)^2}. \tag{14}$$

The method of moments estimation equation, obtained by equating observed and expected s, is

$$\sum_{k=0}^{r-1}\frac{1}{n-k} = \frac{s}{n}. \tag{15}$$

As Craig pointed out, (15) is the exact maximum likelihood equation for the likelihood $\frac{1}{n^s}\frac{n!}{(n-r)!}$ and the solutions of (11) and (15) therefore differ by one at most. (15) has no solution $\tilde{n} \geq r$ when $s > r(1+\frac{1}{2}+\ldots+1/r) = s_0(r)$ say. As far as the method of moments interpretation of (15) is concerned, this is explained by the fact that, since $n \geq r$, $E[s] \leq s_0(r)$. That is, there is no expected value of s to which an observed value greater than $s_0(r)$ corresponds, and it is therefore meaningless to equate them. It is not likely that s will ever be greater than $s_0(r)$ in practice. $s_0(100)$, for instance, is 519. Before making 519 catches to obtain 100 individuals, the experimenter would be sure to doubt the randomness of his sampling or the correctness of his (necessary) information that $n \geq 100$.

For the likelihood (12), s is sufficient for n and ignoring the possibility that $s = r$ (which makes $\tilde{n} = \infty$), using (14) and the same technique as in § 3·2, we find that

$$\operatorname{var}(\tilde{n}) \sim n\left[\sum_{k=1}^{r-1}\frac{k}{(n-k)^2}\right]^{-1} = O(n),$$

$$\beta \sim n\sum_{k=1}^{r-1}\frac{k}{(n-k)^3}\left[\sum_{k=1}^{r-1}\frac{k}{(n-k)^2}\right]^{-2} = O(1),$$

where the limit process is now $n\to\infty$, $r\to\infty$ such that r/n is constant.

It is readily shown that s is normally distributed neglecting terms $O(n^{-\frac{1}{2}})$. Therefore, confidence interval estimation of n based on (13) and (14) proceeds as in § 3·4 except that $\tilde{n} = \tilde{n}(s)$ is now a monotone-decreasing function of s.

All of the formulae appearing in this subsection may of course be simplified by using integral approximations to the sums of reciprocal powers of $n-k$.

4·4. The sequential census of Goodman (1953) can be described as follows. Before the experiment is begun, an infinite sequence $\{a_i\}$ of sample sizes is postulated together with l, the number of tagged individuals to be recovered. Sampling stops at the completion of the sth sample, s being defined by

$$\sum_{i=1}^{s-1} a_i - r_{s-1} < l, \quad \sum_{i=1}^{s} a_i - r_s \geqslant l,$$

where r_s denotes the number of individuals observed in the first s samples.

When all $a_i = 1$, sampling stops as soon as $s - r = l$. We comment briefly on this particular case using the approach of the present section. We require

$P[r$ individuals in $r+l-1$ samples and a previously caught individual at the $(r+l)$th$]$

$$= \frac{1}{n^{r+l-1}} \frac{n!}{(n-r)!} \sigma^r_{r+l-1} \frac{r}{n}$$

$$= \frac{r}{n^{r+l}} \frac{n!}{(n-r)!} \sigma^r_{r+l-1}. \qquad (16)$$

Maximum likelihood estimation remains the same as before and r or s $(= r+l)$ is sufficient for n. There is, however, a minimum-variance unbiased estimate. Since

$$\sum_{r=1}^{n} \frac{r}{n^{r+l}} \frac{n!}{(n-r)!} \sigma^r_{r+l-1} = 1$$

is an identity in n for any positive integer l,

$$\sum_{r=1}^{n} \frac{r}{n^{r+l}} \frac{n!}{(n-r)!} \sigma^r_{r+l} = n.$$

Therefore $\sigma^r_{r+l}/\sigma^r_{r+l-1}$ is an unbiased estimate of n. Moreover, it is uniquely unbiased as is easily seen by induction on n, and because it is sufficient it has minimum variance (Rao, 1952).

Using more general methods, Goodman expressed the same estimate as $K(r,l)/K(r,l-1)$, where $K(r,0) = r$ and $K(r,l) = r\sum_{t=1}^{r} K(t,l-1)$. By observing that $\sigma^r_{r+l} = \Delta^{-1}[(r+1)\,\sigma^{r+1}_{r+l}]$ (Jordan, p. 171), and defining $\Delta^{-1}\phi(r) = \sum_{t=0}^{r-1}\phi(t) + \text{constant}$ (Jordan, p. 101), it follows that $K(r,l) = r\sigma^r_{r+l}$, which accounts for the equivalence of the two expressions for the unbiased estimate. Goodman made reference to tables facilitating the calculation of this estimate. He also provided another basis for estimation by showing that as $n \to \infty$, l remaining constant, the distribution of s^2/n tends to that of χ^2_{2l}.

4·5. Does the ratio of two Stirling numbers afford an unbiased estimate of n for the direct or inverse models? The answer is no, except in one unimportant instance: for the direct census with $s \geqslant n$, $\sigma^r_{s+1}/\sigma^r_s$ is unbiased and has minimum variance.

The application of the three models of this section to the estimation of the number of classes may be framed in the language of coupon collecting. (10), (12) and (16) correspond, respectively, to the collection ceasing when the number of coupons (s), the number of different kinds of coupon (r) and the number of 'swop' coupons ($s-r$) reach prescribed values.

5. ESTIMATION USING MODEL A

5·1. It will be seen from (A') that r and $\{a_i\}$ are jointly sufficient for n and $\{p_i\}$. Confidence interval estimation of n is therefore no longer a practical possibility.

Differentiating (A') with respect to p_i, we find that the maximum-likelihood estimate of p_i is

$$\tilde{p}_i = \frac{a_i}{\tilde{n}}.$$

Taking the first difference with respect to n of the logarithm of (A'),

$$\frac{\tilde{n}}{\tilde{n}-r} = \frac{1}{\tilde{Q}} = \frac{1}{\prod\limits_i \tilde{q}_i}.$$

Hence, $\tilde{n}$ is again the solution of (1).

The same estimation equations may also be deduced by the method of moments, for $E[a_i] = np_i$ and $E[n-r] = nQ$.

The derivation of formulae for the variance and bias of $\tilde{n}$ is much the same as for model (B). Writing $r = a_{s+1}$ and $Q = q_{s+1} = 1-p_{s+1}$ for convenience, the solution of (1) is $\tilde{n} = \tilde{n}[\{a_\alpha\}]$, $\alpha = 1, 2, \ldots, s+1$. $\tilde{n}[\{np_\alpha\}] = n$ and

$$\tilde{n}-n = \sum_\alpha (a_\alpha - np_\alpha)\frac{\partial \tilde{n}}{\partial a_\alpha} + \tfrac{1}{2}\sum_\alpha (a_\alpha - np_\alpha)^2 \frac{\partial^2 \tilde{n}}{\partial a_\alpha^2} + \sum_{\alpha<\beta} (a_\alpha - np_\alpha)(a_\beta - np_\beta)\frac{\partial^2 \tilde{n}}{\partial a_\alpha \partial a_\beta} + \ldots, \tag{17}$$

where all derivatives are evaluated at $\{a_\alpha\} = \{np_\alpha\}$. It can be shown that any derivative of $\tilde{n}$ of order k is $O(n^{-k+1})$ for the limit process $n\to\infty$, $\{p_\alpha\}$ constant. Also, that all multinomial moments of order $2l$ are $O(n^l)$ and of order $2l+1$ are $O(n^l)$. These two facts, combined with several pages of tedious algebra, lead to

$$\operatorname{var}(\tilde{n}) \sim n\left[\frac{1}{Q} + s - 1 - \sum_i \frac{1}{q_i}\right]^{-1} \tag{18}$$

and

$$\beta \sim \frac{1}{2}\,\frac{\left[s-1-\sum\limits_i \frac{1}{q_i}\right]^2 + \left[s-1-\sum\limits_i \frac{1}{q_i^2}\right]}{\left[\frac{1}{Q}+s-1-\sum\limits_i \frac{1}{q_i}\right]^2}. \tag{19}$$

5·2. When $s = 2$, $$n' = \frac{(a_1+1)(a_2+1)}{u_{12}+1} - 1 = \frac{u_1 u_2}{u_{12}+1} + r$$

is again an almost unbiased estimate of n. For

$$\sum_{u_1, u_2, u_{12}=0}^{n} \frac{n!}{(n-r)!\,u_1!\,u_2!\,u_{12}!}\,Q^{n-r}P_1^{u_1}P_2^{u_2}P_{12}^{u_{12}} = 1,$$

where $Q = q_1 q_2$, $P_1 = p_1 q_2$, $P_2 = q_1 p_2$, $P_{12} = p_1 p_2$, and it follows easily that

$$E[n'] = n - nq_1q_2(1-p_1p_2)^{n-1}. \tag{20}$$

Seeing that $(1-p_1p_2)^{n-1}$ is approximated well by $e^{-E[u_{12}]}$,

$$E[n'] = n-E[n-r]\,e^{-E[u_{12}]} \tag{21}$$

to a good approximation. The negative bias of n' will in general be small.

Consider now the conditional expectation of n' given a_1. This leads to a slightly different statement of the last result which we shall require in § 5·3. For this purpose write

$$n' = a_1+(a_1+1)\frac{u_2}{u_{12}+1}.$$

Given a_1, u_2 and u_{12} are independent binomial variables $B[n-a_1, p_2]$ and $B[a_1, p_2]$, respectively, and

$$E[u_2 \mid a_1] = (n-a_1)\,p_2, \quad E\left[\frac{1}{u_{12}+1} \,\Big|\, a_1\right] = \frac{1}{(a_1+1)\,p_2}[1-q_2^{a_1+1}].$$

Therefore

$$\begin{aligned} E[n' \mid a_1] &= a_1+(n-a_1)\,[1-q_2^{a_1+1}] \\ &= n-(n-a_1)\,q_2(1-p_2)^{a_1} \qquad (22) \\ &= n-E[n-r \mid a_1]\,e^{-E[u_{12}\mid a_1]} \qquad (23) \end{aligned}$$

to the same approximation as before. The expectation of (22) over a_1 plainly gives (20) again, but the inference we wish to make is that the bias may be neglected after taking only the conditional expectation and, what is more, that the difference between (22) and (20) is negligible.

From (18),

$$\begin{aligned} \operatorname{var}(n') &\sim n\left[\frac{1}{q_1q_2}+1-\frac{1}{q_1}-\frac{1}{q_2}\right]^{-1} \qquad (24) \\ &= n\frac{q_1q_2}{p_1p_2}, \\ &= n\frac{E[n-r]}{E[u_{12}]}. \qquad (25) \end{aligned}$$

5·3. In order to consider Chapman's recommendation (1952) for an s-sample unbiased estimate of n, let $a_{<k}$ denote the number of different individuals captured before the kth sample, $k = 2, 3, \ldots, s$. (Thus $a_{<2} = a_1$.) Further, let $a_{<k.k}$ denote the number of different individuals captured before and at the kth sample. Thus, for instance, if $s = 4$,

$$a_{<3.3} = u_{13}+u_{23}+u_{123}+u_{134}+u_{234}+u_{1234}.$$

Clearly,

$$a_{<k+1} = a_{<k}+a_k-a_{<k.k}$$

and

$$n-a_{<k+1}, \quad a_{<k}-a_{<k.k}, \quad a_k-a_{<k.k}, \quad a_{<k.k}$$

are distributed multinomially with parameters n and

$$q_1\cdots q_k, \quad (1-q_1\cdots q_{k-1})\,q_k, \quad (q_1\cdots q_{k-1})\,p_k, \quad (1-q_1\cdots q_{k-1})\,p_k.$$

Therefore, in the same way as for $s = 2$

$$n'_k = \frac{(a_{<k}+1)\,(a_k+1)}{a_{<k.k}+1}-1 \quad (k = 2, 3, \ldots, s),$$

is an almost unbiased estimate of n if, as we shall assume throughout this subsection, sampling is large enough. We notice that the covariance of any two of these estimates is negligible compared with their variances. For, if $l<k$

$$E[n'_l(n'_k-E[n'_k]) \mid a_{<l}, a_l, a_{<k}] = n'_l(E[n'_k \mid a_{<k}]-E[n'_k]),$$

which is negligible compared with $\operatorname{var}(n'_k)$. (Compare n times the difference between (21) and (23) with (25).)

More important, from the practical point of view, than knowing how to get maximum information from fixed effort is to know how the information increases with increased effort. Supposing that all $e_i = \bar{e}$ say,

$$I = \frac{\alpha^2}{2n} s(s-1)\,\bar{e}^2[1+\tfrac{1}{3}\alpha(s+1)\,\bar{e}+\ldots].$$

Thus, if the effort per sample is enlarged to $k\bar{e}$, the information is enlarged to more than k^2I. While if $\bar{e}$ is held constant and the number of samples increased from s to $s+1$, the information is multiplied by more than $(s+1)/(s-1)$; in other words, the information is roughly proportional to the number of different pairs of samples, as might be expected.

5·6. If s is large and each p_i is small, the probability that any sample size is one will be a small quantity of first order, while the probability that it is greater than one will be of second order. Therefore, in the limit as $s \to \infty$ and each $p_i \to 0$, sampling becomes 'continuous' and each 'sample' is of size zero or one. In this way, we obtain a valid description of the experiment for which only one individual is captured at a time.

To formulate this idea precisely, let Q_x denote the coefficient of z^x in $\prod\limits_{i=1}^{s}(q_i+p_iz)$; that is, let $Q_0 = Q$, $Q_1 = \sum\limits_i P_i$, $Q_2 = \sum\limits_{i<j} P_{ij}$, ... Then, summing (A) over values of $\{u_\omega\}$ such that $\sum\limits_i u_i = f_1$, $\sum\limits_{i<j} u_{ij} = f_2$, ..., the density of $\{f_x\}$, where $f_0 = n-r$, is

$$\frac{n!}{\prod\limits_{x=0}^{s} f_x!} \prod_{x=0}^{s} Q_x^{f_x}. \tag{27}$$

Consider the limit process: $s \to \infty$, $\max\{p_i\} \to 0$ such that $\prod\limits_i (1-p_i) = Q$ remains fixed and equal to $e^{-\lambda}$ say; that is to say, such that $\alpha \sum\limits_i e_i$ remains fixed and equal to λ. For this process, $\prod\limits_i (q_i+p_iz) \to e^{\lambda(z-1)}$ and therefore $Q_x \to e^{-\lambda}\lambda^x/x!$. Thus, the limit of (27) is

$$\frac{n!}{\prod\limits_{x=0}^{\infty} f_x!} \prod_{x=0}^{\infty} \left(e^{-\lambda}\frac{\lambda^x}{x!}\right)^{f_x}. \tag{28}$$

Let us now redefine s as the random variable $\sum\limits_{x=1}^{\infty} xf_x$, the total number of catches made. (28) can then be written

$$p[\{f_x\}] = \frac{n!}{(n-r)!}\,\frac{e^{-n\lambda}\lambda^s}{(1!)^{f_1}(2!)^{f_2}\ldots f_1!f_2!\ldots}. \tag{A_1}$$

Craig (1953) postulated this model as an alternative to (10) and discussed the estimation of λ and n. Although the sampling is now a continuous process, it is by no means necessary that it extends over only one interval of time. In practice, the experimenter will expend effort until he catches an individual and then will pause while marking and recording it and letting it return to the population. There is no conflict between this practice and (A_1).

The joint probability generating function of r and s for (A_1) is

$$E[t_1^r t_2^s] = E\left[\prod_{x=1}^{\infty}(t_1t_2^x)^{f_x}\right] = e^{-n\lambda}[1+t_1(e^{\lambda t_2}-1)]^n.$$

We notice that the marginal distribution of s is Poisson with parameter $n\lambda$ and therefore that the conditional probability of $\{f_x\}$ given s is

$$p[\{f_x\}\mid s] = \frac{1}{n^s}\frac{n!}{(n-r)!}\frac{s!}{(1!)^{f_1}(2!)^{f_2}\ldots f_1!f_2!\ldots},$$

which is (9) again. Thus, there are two routes from (A) to (9): either via (B) and (B_1) or via (A_1).

The marginal distribution of r is binomial $B[n, 1-e^{-\lambda}]$. The conditional probability of $\{f_x\}$ given r is therefore

$$p[\{f_x\}\mid r] = \frac{r!\,\lambda^s}{(e^{\lambda}-1)^r\,(1!)^{f_1}(2!)^{f_2}\ldots f_1!f_2!\ldots}. \tag{29}$$

This density was discussed by Craig from the point of view of the truncated Poisson distribution defined by the probability of a caught individual being caught x times being

$$\frac{e^{-\lambda}}{1-e^{-\lambda}}\frac{\lambda^x}{x!} \quad (x = 1, 2, \ldots),$$

but plainly cannot serve as a probability description of any possible way of conducting a capture-recapture experiment; for it demands that both the total effort and the total number of different individuals be fixed in advance, which is impossible.

5·7. We notice that (29) satisfies conditions (a), (b) and (c) given in § 3·2 for the applicability of large-sample maximum likelihood theory. This is also true of the counterpart of (29) for the general case. From (A) we have

$$p[\{u_\omega\}\mid r] = \frac{r!}{\prod\limits_{\omega} u_\omega!}\prod_{\omega}\left(\frac{P_\omega}{1-Q}\right)^{u_\omega}. \tag{30}$$

(30), like (29), does not mention the uncaught individuals and does not truly describe any experiment. However, densities of this type are important as theoretical devices, as we can demonstrate by applying standard maximum likelihood theory to (30).

Maximizing $L = \log p[\{u_\omega\}\mid r]$ with respect to $\{p_i\}$, we obtain the equations

$$\frac{a_i}{p_i} = \frac{r}{1-Q} \quad (i = 1, 2, \ldots, s). \tag{31}$$

Let $\theta = (1-Q)^{-1}$. Then $\tilde{\theta} = (1-\tilde{Q})^{-1}$ is the maximum likelihood estimate of θ. (31) implies that $\tilde{\theta}$ satisfies the equation $\prod\limits_i (r\theta - a_i) = (r\theta)^{s-1}(r\theta - r)$, that is $r\tilde{\theta}$ satisfies (1). Therefore $r\tilde{\theta} = \tilde{n}$.

We can obtain the asymptotic mean square error of $\tilde{\theta}$ for (30) as $r\to\infty$ by first finding the information matrix, $\mathbf{V}^{-1}$ say, whose (i,j) element is $-E\left[\dfrac{\partial^2 L}{\partial p_i \partial p_j}\right]$. We find that

$$\mathbf{V}^{-1} = r\frac{Q}{(1-Q)^2}\mathbf{DWD},$$

where $\mathbf{D}$ is the diagonal matrix whose (i,i) element is q_i^{-1} and $\mathbf{W}$ is the matrix whose (i,i) element is $w_i - 1$, where $w_i = q_i(1-Q)/(p_iQ)$, and whose every other element is -1.

Now, as $r\to\infty$,

$$E[(\tilde{\theta}-\theta)^2\mid r] \sim \mathbf{d}'\mathbf{V}\mathbf{d},$$

where $\mathbf{d}$ is the vector whose ith element is the derivative $\partial\tilde{\theta}/\partial\tilde{p}_i$ evaluated at $\{\tilde{p}_i\} = \{p_i\}$. On differentiating $\tilde{\theta} = [1-\prod\limits_i (1-\tilde{p}_i)]^{-1}$, it is found that

$$\mathbf{d} = -\frac{Q}{(1-Q)^2}\mathbf{D1},$$

where $\mathbf{1}$ is the vector all of whose elements are 1. Hence

$$E[(\tilde{\theta}-\theta)^2 \mid r] \sim \frac{Q}{r(1-Q)^2}\mathbf{1}'\mathbf{D}\mathbf{D}^{-1}\mathbf{W}^{-1}\mathbf{D}^{-1}\mathbf{D}\mathbf{1} = \frac{Q}{r(1-Q)^2}\mathbf{1}'\mathbf{W}^{-1}\mathbf{1}.$$

To find $\mathbf{1}'\mathbf{W}^{-1}\mathbf{1}$, let $\mathbf{W}^{-1}\mathbf{1} = \mathbf{x}$. Then $\mathbf{W}\mathbf{x} = \mathbf{1}$ and the ith element of $\mathbf{x}$ is seen to be $\dfrac{w_i^{-1}}{1-\sum_i w_i^{-1}}$. Therefore

$$E[(\tilde{\theta}-\theta)^2 \mid r] \sim \frac{Q}{r(1-Q)^2}\mathbf{1}'\mathbf{x} = \frac{Q}{r(1-Q)^2}\,\frac{\sum_i w_i^{-1}}{1-\sum_i w_i^{-1}}.$$

Substituting for w_i

$$E[(\tilde{\theta}-\theta)^2 \mid r] \sim \frac{Q}{r(1-Q)^2}\left[\sum_i \frac{p_i}{q_i}\right]\left[\frac{1}{Q}+s-1-\sum_i \frac{1}{q_i}\right]^{-1} = \frac{\Theta}{r}, \quad \text{say.}$$

By means of the δ-technique we can further say that

$$\left.\begin{aligned} E[(\tilde{\theta}-\theta)^2 \mid r] &= \frac{\Theta}{r}+\frac{\lambda_2}{r^2}+o\left(\frac{1}{r^2}\right) \\ \text{and} \qquad E[(\tilde{\theta}-\theta) \mid r] &= \frac{\nu_1}{r}+\frac{\nu_2}{r^2}+o\left(\frac{1}{r^2}\right). \end{aligned}\right\} \tag{32}$$

Using (32) we can re-derive formula (18) for the asymptotic value of $E[(\tilde{n}-n)^2]$. For, since $\theta = (1-Q)^{-1}$ and $\tilde{n} = r\tilde{\theta}$,

$$E[(\tilde{n}-n)^2 \mid r] = r^2E[(\tilde{\theta}-\theta)^2 \mid r]+\theta^2(r-n(1-Q))^2+2r\theta(r-n(1-Q))\,E[(\tilde{\theta}-\theta) \mid r]. \tag{33}$$

In evaluating the expectation of the right-hand side of (33) over r, we consider the range $kn < r \leqslant n$, where $0 < k < 1-Q$ and kn is integral. This approximation permits the use of the asymptotic formulae (32) and produces errors which asymptotically are negligible. Two observations are sufficient to establish the latter claim. First, if P_r denotes the probability of r

$$\sum_{r=0}^{kn} P_r < P_{kn}\frac{(n-kn+1)(1-Q)}{(n+1)(1-Q)-kn}$$

(Feller, 1957, p. 140), and the last quantity is $O(n^{-\frac{1}{2}}c^n)$, where

$$c = \left(\frac{1-Q}{k}\right)^k\left(\frac{Q}{1-k}\right)^{1-k} < 1.$$

Secondly, $\tilde{\theta} < \phi(s)\,r$ provided always that we ignore the possibility that $\sum_i a_i = r$ which makes $\tilde{\theta}$ infinite. Substituting from (32) we find that

$$\begin{aligned} E[(\tilde{n}-n)^2] &= E[r\Theta]+\theta^2\operatorname{var}(r)+O(1) \\ &= n\left[\frac{1}{Q}+s-1-\sum_i \frac{1}{q_i}\right]^{-1}+O(1), \end{aligned}$$

which is (18).

The rederivation of (18) is not important in itself. However, when combined with the fact that $\tilde{\theta}$ is asymptotically efficient for θ, the above argument has an important consequence. Namely, that for the class of estimates $n^* = r\theta^*$ which satisfy the very reasonable conditions

$$\left.\begin{aligned} E[(\theta^*-\theta)^2 \mid r] &= \frac{\lambda_1}{r}+\frac{\lambda_2}{r^2}+o\left(\frac{1}{r^2}\right) \\ \text{and} \qquad E[(\theta^*-\theta) \mid r] &= \frac{\nu_1}{r}+\frac{\nu_2}{r^2}+o\left(\frac{1}{r^2}\right), \end{aligned}\right\} \tag{34}$$

$\tilde{n}$ has asymptotically minimum mean square error since Θ is the minimum attainable value of λ_1. (Conditions (34) are sufficient for this conclusion. They are quite possibly not necessary.) (34) implies that

$$\left.\begin{aligned} E[(n^*-n)^2] &= \alpha_1 n+\alpha_2+o(1) \\ \text{and}\qquad E[n^*-n] &= \beta_1+\frac{\beta_2}{n}+o\left(\frac{1}{n}\right), \end{aligned}\right\} \tag{35}$$

which are even more reasonable. Unfortunately, (35) does not imply (34), though it is difficult to imagine estimates which satisfy (35) but not (34).

In conclusion: among a wide class of estimates of n, those derived from (A) or (A_1) by the method of maximum likelihood are asymptotically best in that they have minimum mean square error.

I wish to acknowledge my considerable indebtedness to the referee for his invaluable comments on two previous drafts of this paper.

REFERENCES

BAILEY, N. T. J. (1951). On estimating the size of mobile populations from capture-recapture data. *Biometrika*, **38**, 293–306.

CHAPMAN, D. G. (1951). Some properties of the hypergeometric distribution with applications to zoological sample censuses. *Univ. Calif. Publ. Statist.* **1**, 131–60.

CHAPMAN, D. G. (1952). Inverse, multiple and sequential sample censuses. *Biometrics*, **8**, 286–306.

CHAPMAN, D. G. (1954). The estimation of biological populations. *Ann. Math. Statist.* **25**, 1–15.

CHAPMAN, D. G. (1955). Population estimation based on change of composition caused by selective removal. *Biometrika*, **42**, 279–90.

CHAPMAN, D. G. & JUNGE, C. O. (1956). The estimation of the size of a stratified animal population. *Ann. Math. Statist.* **27**, 375–89.

CRAIG, C. C. (1953). Use of marked specimens in estimating populations. *Biometrika*, **40**, 170–6.

FELLER, W. (1957). *An Introduction to Probability Theory and its Application*, 2nd ed. New York: John Wiley and Sons.

GOODMAN, L. A. (1953). Sequential sampling tagging for population size problems. *Ann. Math. Statist.* **24**, 56–69.

GULLAND, J. A. (1955). On the estimation of population parameters from marked members. *Biometrika*, **42**, 269–70.

HAMMERSLEY, J. M. (1953). Capture-recapture analysis. *Biometrika*, **40**, 265–78.

JORDAN, C. (1947). *Calculus of Finite Differences*, 2nd ed. New York: Chelsea Publishing Company.

LESLIE, P. H. (1952). The estimation of population parameters obtained by means of the capture-recapture method. II. The estimation of total numbers. *Biometrika*, **39**, 363–88.

LESLIE, P. H. & CHITTY, D. (1951). The estimation of population parameters obtained by means of the capture-recapture method. I. The maximum likelihood equations for estimating the death-rate. *Biometrika*, **38**, 269–92.

MORAN, P. A. P. (1952). The estimation of death-rates from capture-mark-recapture sampling. *Biometrika*, **39**, 181–8.

RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research.* New York: John Wiley and Sons.

[*From* Biometrika, *Vol.* 46. *Parts* 3 *and* 4, *December* 1959.]
[*All rights reserved*] PRINTED IN GREAT BRITAIN

THE MULTIPLE-RECAPTURE CENSUS

II. ESTIMATION WHEN THERE IS IMMIGRATION OR DEATH

BY J. N. DARROCH
Statistical Laboratory, University of Manchester

1. INTRODUCTION

1·1. In a previous paper (1958), which will be referred to as (I), we discussed the multiple-recapture census when the population is closed both to augmentation from outside and departure from inside. These restrictions are now removed.

Let the experimenter take s samples, as in (I). Also, let $1-\phi_k$ be the conditional probability that an individual dies (or permanently emigrates) between the kth and $(k+1)$th samples given that it is alive at the time of the kth, $k = 1, 2, \ldots, s-1$. Let n_1 be the size of the population at the time of the first sample and let $n_k - n_{k-1}$ new individuals immigrate (or be born) into the population between the $(k-1)$th and kth samples and be alive at the time of the kth, $k = 2, 3, \ldots, s$. Nothing is assumed about the $n_k - n_{k-1}$ and they are treated as parameters of the model. To treat them as random variables would entail assumptions about the manner in which they vary and would complicate rather than simplify the probability densities.

In § 2 we shall take $\phi_k = 1$ in which case n_i is the population size at the time of the ith sample, $i = 1, 2, \ldots, s$. In § 3, $n_i = n$ and the size of the population at the time of the ith sample $(i > 1)$ is a random variable with expected value $n\phi_1 \ldots \phi_{i-1}$. The general case when there is both immigration and death will be considered in § 4.

1·2. The main aims of this paper, as of (I), are to provide exact, fully stochastic models for the observed frequencies of individuals, to show how simply these frequencies naturally group themselves, and to obtain estimates of the unknown parameters. When there is immigration only or death only, the estimates are shown to be asymptotically efficient and their variances are found. In addition, a method of performing tests on the values of the parameters is given. When both immigration and death are operating, on the other hand, the complexity of the probability density prevents us from going further than obtaining the estimates and merely indicating how their variances can be found.

Both in (I) and in the present work we have been unable to obtain satisfactory tests of the underlying assumptions of the models, notably that tagged and untagged animals are captured with equal probabilities. (The above-mentioned tests on the values of the parameters assume the truth of the models.) We hope to fill this gap at a later date.

1·3. We gave a brief review of the literature and a list of references in (I) but, unfortunately, omitted to refer to a paper by Leslie, Chitty & Chitty (1953), the third of three papers by these authors. These three papers contain several ingenious mathematical approaches to multiple-recapture problems, together with a very full discussion of field data on populations of voles, and are uniquely valuable in the way that they dove-tail theory with practice. One of the points emphasized by the authors is this: although a basic feature of the multiple-recapture census is that caught animals are returned to the population alive and unhurt, in practice a few are either accidentally killed or have to be removed from the population.

This eventuality has not been taken account of in the following pages but we note here that it is easy to do so. Each of the usual observed classes 'caught in the $i, \ldots, l$ samples but not otherwise' is coupled with an extra one for those animals which are caught in the same samples but killed or removed at the lth. Also, an additional parameter is introduced to represent the probability that a captured animal is killed or removed at any sample. The estimates of the other parameters change slightly but, apart from this, there are no complications.

2. IMMIGRATION BUT NO DEATH

2·1. In this section, suffices i, j take all values from 1 to s while suffices k, l take all values from 2 to s.

Let p_i $(= 1-q_i)$ denote the probability that any member of the population is caught at the ith sample and let a_i be the size of the ith sample. Let $a_{<k}$ be the number of individuals caught before the kth sample and $a_{<k.k}$ the number which are caught before and at the kth sample. Further, let w denote any non-empty subset of the integers $1, 2, \ldots, s$ and u_w the number caught in every one of the w-samples but not otherwise. Let

$$r = \sum_w u_w = \sum_i u_i + \sum_{i<j} u_{ij} + \ldots + u_{12\ldots s},$$

the total number of individuals caught in the whole experiment.

The notation thus far is the same as in (I) and suffices for most of § 2. However, for the purpose of deriving the density $p[\{u_w\}]$, it is convenient to have a further notation. Namely, let a_w be the number of individuals caught in every one of the w-samples, regardless of whether or not they are caught otherwise.

We begin by finding $p[\{u_w\}]$ for $s = 3$, using a chainwise argument. The probability of catching a_1 individuals in the first sample is

$$\binom{n_1}{a_1} p_1^{a_1} q_1^{n_1-a_1}.$$

Given a_1, the probability of catching a_2 in the second, of which a_{12} are common to the first and $a_2 - a_{12}$ are new ones, is

$$\binom{a_1}{a_{12}} \binom{n_2-a_1}{a_2-a_{12}} p_2^{a_2} q_2^{n_2-a_2}.$$

Given a_1, a_2, a_{12}, the probability of catching a_3 in the third, of which a_{123} are common to the first and second, $a_{13}-a_{123}$ to the first only, $a_{23}-a_{123}$ to the second only and of which $a_3 - a_{13} - a_{23} + a_{123}$ are new ones is

$$\binom{a_{12}}{a_{123}} \binom{a_1-a_{12}}{a_{13}-a_{123}} \binom{a_2-a_{12}}{a_{23}-a_{123}} \binom{n_3-a_1-a_2+a_{12}}{a_3-a_{13}-a_{23}+a_{123}} p_3^{a_3} q_3^{n_3-a_3}.$$

Multiplying these three expressions together, cancelling and renaming the terms involved, we obtain

$$p[\{u_w\}] = \frac{n_1!}{(n_1-a_{<2})!} \frac{(n_2-a_{<2})!}{(n_2-a_{<3})!} \frac{(n_3-a_{<3})!}{(n_3-r)!} \frac{1}{\prod_w u_w!} \prod_{i=1}^{3} p_i^{a_i} q_i^{n_i-a_i}. \tag{1}$$

Note that, if $n_i = n$, (1) contracts to

$$\frac{n!}{(n-r)! \prod_w u_w!} \prod_i p_i^{a_i} q_i^{n-a_i},$$

which is the model forming the basis of (I).

(1) may also be written

$$p[\{u_w\}] = \frac{n_1!}{(n_1-a_{<2})!\,u_1!\,u_{12}!\,u_{13}!\,u_{123}!}\,q_1^{n_1-a_{<2}}\,(p_1q_2q_3)^{u_1}\,(p_1p_2q_3)^{u_{12}}\,(p_1q_2p_3)^{u_{13}}\,(p_1p_2p_3)^{u_{123}}$$
$$\times\frac{(n_2-a_{<2})!}{(n_2-a_{<3})!\,u_2!\,u_{23}!}\,q_2^{n_2-a_{<3}}\,(p_2q_3)^{u_2}\,(p_2p_3)^{u_{23}}$$
$$\times\frac{(n_3-a_{<3})!}{(n_3-r)!\,u_3!}\,q_3^{n_3-r}p_3^{u_3}. \tag{1'}$$

Another way of exhibiting the distribution of the u_w is by means of their generating function $\psi[\{t_w\}] = E[\prod_w t_w^{u_w}]$. It is soon shown that

$$\psi[\{t_w\}] = (q_1q_2q_3+p_1q_2q_3t_1+p_1p_2q_3t_{12}+p_1q_2p_3t_{13}+p_1p_2p_3t_{123}$$
$$+q_1p_2q_3t_2+q_1p_2p_3t_{23}+q_1q_2p_3t_3)^{n_1}$$
$$\times(q_2q_3+p_2q_3t_2+p_2p_3t_{23}+q_2p_3t_3)^{n_2-n_1}$$
$$\times(q_3+p_3t_3)^{n_3-n_2}. \tag{2}$$

2·2. Let us define $a_{<1} = 0$ and $a_{<s+1} = r$. Then for general s, (1) becomes

$$p[\{u_w\}] = \frac{1}{\prod_w u_w!}\prod_i \frac{(n_i-a_{<i})!}{(n_i-a_{<i+1})!}\,p_i^{a_i}q_i^{n_i-a_i}. \tag{3}$$

Maximizing with respect to p_i by equating $(\partial/\partial p_i)\log p[\{u_w\}]$ to zero,

$$\tilde{n}_i\tilde{p}_i = a_i. \tag{4}$$

Maximizing with respect to n_i by equating $\Delta_{n_i}\log p\,[\{u_w\}]$ to zero,

$$\frac{\tilde{n}_i-a_{<i+1}}{\tilde{n}_i-a_{<i}} = \tilde{q}_i. \tag{5}$$

The equation for $i = 1$ in (4) is the same as that in (5) and there is consequently no information on n_1 and p_1 separately. Otherwise, combining (4) and (5) and remembering that $a_{<k+1} = a_{<k}+a_k-a_{<k.k}$, we obtain

$$\tilde{p}_k = \frac{a_{<k.k}}{a_{<k}} \tag{6}$$

and

$$\tilde{n}_k = \frac{a_{<k}a_k}{a_{<k.k}}$$

provided $a_{<k.k} > 0$.

2·3. Having found maximum likelihood estimates from the density (3) we cannot apply maximum likelihood large sample theory to them directly, for reasons given in (I) § 3·2. The principal obstacle is the presence of the n_i in the factorial terms of (3). To get over this difficulty, we use a suitably modified version of the argument of (I) § 5·7. That is, we apply maximum likelihood large sample theory to a density which is conditional on some of the sample variables and afterwards take expectations over these variables. In this case, the most suitable conditional density is $p[\{u_w\}\mid\{a_{<i+1}\}]$.

The conditional distribution of $a_{<i+1}-a_{<i}$ given $a_{<i}$ is $B[n_i-a_{<i}, p_i]$ (including $i = 1$ when the distribution of $a_{<2}$ is $B[n_1, p_1]$) and therefore

$$p[\{a_{<i+1}\}] = \prod_i\binom{n_i-a_{<i}}{a_{<i+1}-a_{<i}}p_i^{a_{<i+1}-a_{<i}}q_i^{n_i-a_{<i+1}}.$$

Hence

$$p[\{u_w\}\mid\{a_{<i+1}\}] = \frac{1}{\prod_w u_w!}\prod_i(a_{<i+1}-a_{<i})!\prod_k p_k^{a_{<k.k}}q_k^{a_{<k}-a_{<k.k}}. \tag{7}$$

Reverting for the moment to the parent density, we see that

$$\tilde{n}_k = (a_{<k+1} - a_{<k})/\tilde{p}_k + a_{<k}, \tag{8}$$

where $\tilde{p}_k$ is given by (6). This equation will prove useful for two reasons. First, $a_{<k}$ and $a_{<k+1}$ are held constant in the conditional density (7), and secondly, $\tilde{p}_k$ is the maximum likelihood estimate of p_k for the conditional as well as the parent density. In view of (8), it turns out to be preferable to change from p_k to $\theta_k = p_k^{-1}$. (8) then becomes

$$\tilde{n}_k = (a_{<k+1} - a_{<k})\,\tilde{\theta}_k + a_{<k}. \tag{9}$$

Let $L = L[\{\theta_k\}] = \log p[\{u_w\} \mid \{a_{<i+1}\}]$. Then

$$\frac{\partial L}{\partial \theta_k} = -\frac{a_{<k}}{\theta_k} + \frac{a_{<k} - a_{<k.k}}{\theta_k - 1},$$

which, when equated to zero, gives $\tilde{\theta}_k = a_{<k}/a_{<k.k}$ provided $a_{<k.k} > 0$. Further

$$-E\left[\frac{\partial^2 L}{\partial \theta_k^2} \mid \{a_{<i+1}\}\right] = \frac{a_{<k}}{\theta_k^2(\theta_k - 1)},$$

$$-E\left[\frac{\partial^2 L}{\partial \theta_k \partial \theta_l} \mid \{a_{<i+1}\}\right] = 0 \quad (k \neq l),$$

yielding the asymptotic formulae

$$E[(\tilde{\theta}_k - \theta_k)^2 \mid \{a_{<i+1}\}] = \frac{\theta_k^2(\theta_k - 1)}{a_{<k}},$$

$$E[(\tilde{\theta}_k - \theta_k)(\tilde{\theta}_l - \theta_l) \mid \{a_{<i+1}\}] = 0.$$

When $a_{<k.k} = 0$, the Taylor expansion of $\partial L/\partial \theta_k$ at $\tilde{\theta}_k$ about its value at θ_k is meaningless since there is no finite solution for $\tilde{\theta}_k$. However, the resulting invalidity of the last formulae is easily avoided by changing from $\tilde{\theta}_k$ to

$$\theta_k' = \frac{a_{<k} + 1}{a_{<k.k} + 1},$$

since θ_k' is always finite and has an asymptotic variance differing only negligibly from that of $\tilde{\theta}_k$ over the finite part of the latter's range. (This can be shown using the δ-technique.) Moreover, θ_k' has a negligible bias. For, given $\{a_{<i+1}\}$, $a_{<k.k}$ is $B[a_{<k}, p_k]$. Therefore

$$E[\theta_k' \mid \{a_{<i+1}\}] = E[\theta_k' \mid a_{<k}] = \theta_k(1 - q_k^{a_{<k}+1})$$

and the difference between this expression and θ_k may be neglected.

Two further properties of the θ_k' should be mentioned here. First, they are independently distributed given $\{a_{<i+1}\}$. This follows from the independence of the variables $a_{<k.k}$, a property which is proved without difficulty. Secondly, using the δ-technique, we find that

$$E[(\theta_k' - \theta_k)^2 \mid \{a_{<i+1}\}] = E[(\theta_k' - \theta_k)^2 \mid a_{<k}]$$
$$= \frac{\theta_k^2(\theta_k - 1)}{a_{<k}} + O\left(\frac{1}{a_{<k}^2}\right),$$

as $a_{<k} \to \infty$.

Returning now to the estimation of the n_k and recalling (9), we replace $\tilde{n}_k$ by

$$n_k' = (a_{<k+1} - a_{<k})\,\theta_k' + a_{<k},$$

that is, by

$$n_k' = \frac{(a_{<k} + 1)(a_k + 1)}{a_{<k.k} + 1} - 1. \tag{10}$$

Now

$$n_k' - n_k = (a_{<k+1} - a_{<k})(\theta_k' - \theta_k) + \theta_k[(a_{<k+1} - a_{<k}) - (n_k - a_{<k})\,p_k]. \tag{11}$$

Therefore, taking expectations conditional on $\{a_{<i+1}\}$,

$$E[(n_k'-n_k)\,|\,\{a_{<i+1}\}] = -(a_{<k+1}-a_{<k})\,\theta_k q_k^{a_{<k+1}}+\theta_k[(a_{<k+1}-a_{<k})-(n_k-a_{<k})\,p_k].$$

Now taking expectations over $a_{<2}, \ldots, a_{<k-1}, a_{<k+1}, \ldots, a_{<s+1}$,

$$\begin{aligned} E[(n_k'-n_k)\,|\,a_{<k}] &= -(n_k-a_{<k})\,p_k\theta_k q_k^{a_{<k+1}} \\ &= -(n_k-a_{<k})\,q_k^{a_{<k+1}}, \end{aligned}$$

since $a_{<k+1}-a_{<k}=B[n_k-a_{<k}, p_k]$. We shall neglect this bias.

Squaring (11) and taking expectations,

$$\begin{aligned} E[(n_k'-n_k)^2\,|\,a_{<k}] &= [(n_k-a_{<k})^2 p_k^2+(n_k-a_{<k})\,p_k q_k]\,E[(\theta_k'-\theta_k)^2\,|\,a_{<k}] \\ &\quad +2\theta_k(n_k-a_{<k})\,p_k q_k E[(\theta_k'-\theta_k)\,|\,a_{<k}] \\ &\quad +\theta_k^2(n_k-a_{<k})\,p_k q_k. \end{aligned} \tag{12}$$

In deriving an asymptotic formula for $E[(n_k'-n_k)^2]$, a convenient limit process to use is: $n_i \to \infty$ such that $n_i/n_j \to c_{ij}$, constant. To evaluate the expectation of the first term of the right-hand side of (12) over $a_{<k}$, consider separately the ranges $0 \leqslant a_{<k} \leqslant hn_1$ and $hn_1 < a_{<k} \leqslant n_{k-1}$, where $h < 1-q_1 \ldots q_{k-1}$ and hn_1 is integral. Now $a_{<k}$ is the sum of independent binomial variables $B[n_1, 1-q_1 \ldots q_{k-1}], \ldots, B[n_{k-1}-n_{k-2}, p_{k-1}]$ and therefore

$$P[a_{<k} \leqslant hn_1] \leqslant P[B[n_1, 1-q_1 \ldots q_{k-1}] \leqslant hn_1] = O\left(\frac{1}{\sqrt{n_1}}c^{n_1}\right),$$

where $$c = \left(\frac{1-q_1 \ldots q_{k-1}}{h}\right)^h \left(\frac{q_1 \ldots q_{k-1}}{1-h}\right)^{1-h} < 1.$$

Also θ_k' can never exceed $a_{<k}+1$. Therefore, for the range $0 \leqslant a_{<k} \leqslant hn_1$, the contribution of the first term of (12) is at most $O(n_k^2 n_1^2 n_1^{-\frac{1}{2}} c^{n_1}) = O(n_1^{\frac{7}{2}} c^{n_1}) = o(1)$. For the other part of the range, we may use the asymptotic formula for the variance of θ_k'. Since the second term on the right-hand side of (12) is $o(1)$, we may thus far say that

$$\begin{aligned} E[(n_k'-n_k)^2] = \underset{a_{<k}>hn_1}{E} \Bigg[&((n_k-a_{<k})^2 p_k^2+(n_k-a_{<k})\,p_k q_k) \\ &\left(\frac{\theta_k^2(\theta_k-1)}{a_{<k}}+O\left(\frac{1}{a_{<k}^2}\right)\right)\Bigg] + \underset{a_{<k}\geqslant 0}{E}[\theta_k^2(n_k-a_{<k})\,p_k q_k]+o(1). \end{aligned}$$

Let $$\alpha_{<k} = E[a_{<k}] = n_1(1-q_1 \ldots q_{k-1})+\ldots+(n_{k-1}-n_{k-2})\,p_{k-1}.$$

Then, expanding $1/a_{<k}$ about $1/\alpha_{<k}$ and taking expectations over $a_{<k}$, it will be found that

$$E[(n_k'-n_k)^2] = \frac{n_k(n_k-\alpha_{<k})\,q_k}{\alpha_{<k}p_k}+O(1). \tag{13}$$

Note that, when all $n_i = n$, (13) becomes

$$E[(n_k'-n)^2] = \frac{nq_1 \ldots q_k}{(1-q_1 \ldots q_{k-1})\,p_k}+O(1),$$

which was derived previously in (I).

Thus, the n_k' are almost unbiased estimates of the n_k with variances given by (13) and, as is easily shown, with negligible covariances. Further, n_k' is the asymptotically most efficient estimate of the class of estimates

$$n_k^* = (a_{<k+1}-a_{<k})\,\theta_k^*+a_{<k}$$

for which $$E[(\theta_k^*-\theta_k)^2\,|\,a_{<k}, a_{<k+1}] = \frac{\lambda_k}{a_{<k}}+O\left(\frac{1}{a_{<k}^2}\right)$$

and for which $E[(\theta_k^* - \theta_k) \mid a_{<k}, a_{<k+1}]$ is negligible. This is a simple consequence of the fact that, since θ_k' is asymptotically efficient for θ_k, $\theta_k^2(\theta_k - 1)$ is the minimum attainable value of λ_k. In (I) we pointed out that n_k' may be evaluated from 'similar' tagging and that, if $n_i = n$, the n_k' contain between them all of the information on n.

2·4. Suppose that catchability is constant throughout the experiment and that e_i units of effort are expended in catching the ith sample. Then

$$q_i = e^{-e_i/\gamma}, \tag{14}$$

where the unknown coefficient of e_i is written as a reciprocal parameter for much the same reason that, in § 2·3, we changed from p_k to $\theta_k = p_k^{-1}$. In previous effort models, it has been customary to make the expected catch size proportional to the effort. This is equivalent here to writing $p_i = e_i/\gamma$ which involves a relative error $O(e_i/\gamma)$, and the equivalent approximation for q_i is $q_i = 1$ which is obviously too severe. Therefore, accepting $q_i = 1 - e_i/\gamma$ as a satisfactory approximation to (14), that is, neglecting terms which are relatively $O(e_i^2/\gamma^2)$, the approximation of the same order for p_i is $p_i = e_i/\gamma(1 - e_i/2\gamma)$. When it is expedient to do so, we shall use these expressions. The fact that they lead to $p_i + q_i = 1 - e_i^2/2\gamma^2$ is not, of course, inconsistent with our decision to neglect terms $O(e_i^2/\gamma^2)$.

Again, consider the conditional density $p[\{u_w\} \mid \{a_{<j+1}\}]$ which gives rise to the likelihood

$$e^{L(\gamma)} = \prod_k (1 - e^{-e_k/\gamma})^{a_{<k.k}} (e^{-e_k/\gamma})^{a_{<k} - a_{<k.k}}.$$

Hence

$$\frac{\partial L}{\partial \gamma} = -\frac{1}{\gamma^2}\left[\sum_k \frac{a_{<k.k} e_k}{1 - e^{-e_k/\gamma}} - \sum_k a_{<k} e_k\right].$$

Equating this to zero, the equation for $\tilde{\gamma}$ is

$$\sum_k \frac{a_{<k.k} e_k}{(e_k/\tilde{\gamma})(1 - e_k/2\tilde{\gamma})} = \sum_k a_{<k} e_k$$

and replacing $(1 - e_k/2\tilde{\gamma})^{-1}$ by $1 + e_k/2\tilde{\gamma}$, we find that

$$\tilde{\gamma} \doteq \frac{\sum_k (a_{<k} - \frac{1}{2}a_{<k.k})\, e_k}{\sum_k a_{<k.k}}.$$

This formula breaks down if $\sum_k a_{<k.k} = 0$ (and so does the whole experiment since it has not yielded a single recapture). Therefore, instead of $\tilde{\gamma}$, consider

$$\gamma' = \frac{\sum_k (a_{<k} - \frac{1}{2}a_{<k.k})\, e_k + \frac{1}{2}\bar{e}}{\sum_k a_{<k.k} + 1},$$

where $\bar{e} = \sum_k e_k/(s-1)$. γ' always exists and we suspect that it is almost unbiased, for the following reason. If all $e_k = \bar{e}$ and $p_k = \bar{p}$ say,

$$\gamma' = \bar{e}\,\frac{\sum_k a_{<k} + 1}{\sum_k a_{<k.k} + 1} - \tfrac{1}{2}\bar{e}$$

and $\Sigma a_{<k.k} = \Sigma B[a_{<k}, \bar{p}] = B[\Sigma a_{<k}, \bar{p}]$ since the $a_{<k.k}$ are independent. Hence

$$E[\gamma' \mid \{a_{<j+1}\}] \doteq \bar{e}\,\frac{1}{\bar{p}}\,[1 - \bar{q}^{\Sigma a_{<k}+1}] - \tfrac{1}{2}\bar{e}.$$

Neglecting $\bar{q}^{\Sigma a_{<k+1}}$ and replacing $\bar{p}$ by $(\bar{e}/\gamma)(1-\bar{e}/2\gamma)$,

$$E[\gamma' \mid \{a_{<j+1}\}] = \gamma.$$

In the general case when the e_k are unequal, it still seems reasonable to assume that γ' is almost unbiased.

To obtain $E[(\gamma'-\gamma)^2 \mid \{a_{<j+1}\}]$ we may invert

$$\begin{aligned} -E\left[\frac{\partial^2 L}{\partial \gamma^2} \mid \{a_{<j+1}\}\right] &= \gamma^{-4} \sum_k a_{<k} e_k^2 q_k / p_k \\ &= \gamma^{-2} \sum_k a_{<k} p_k \end{aligned}$$

on making use of our approximations for p_k and q_k. The δ-technique shows that

$$E[(\gamma'-\gamma)^2 \mid \{a_{<j+1}\}] = \frac{\gamma^2}{\Sigma a_{<k} p_k} + O((\Sigma a_{<k} p_k)^{-2}).$$

From now on, the argument follows very much the same lines as in § 2·3. For the parent density

$$\tilde{n}_i = (a_{<i+1} - a_{<i})/\tilde{p}_i + a_{<i}$$

and $\tilde{\gamma}$ has the same value as for the conditional density. We replace it by γ' and $\tilde{n}_i$ by m_i' say, defined as

$$m_i' = (a_{<i+1} - a_{<i})/(e_i/\gamma')(1 - e_i/2\gamma') + a_{<i}.$$

Hence

$$m_i' - n_i = \frac{(a_{<i+1} - a_{<i})}{e_i}(\gamma'-\gamma) + \frac{1}{p_i}[(a_{<i+1} - a_{<i}) - (n_i - a_{<i}) p_i]. \tag{15}$$

Taking expectations conditional on $\{a_{<j+1}\}$ and then over all $a_{<j+1}$ except $a_{<i}$, we see that m_i' is an almost unbiased estimate of n_i. Further, squaring (15) and omitting the region: $0 \leqslant a_{<k} \leqslant h_k n_{k-1}$, where $h_k < 1 - q_1 \ldots q_{k-1}$ and $h_k n_{k-1}$ is integral, all k, when taking expectations over $\{a_{<j+1}\}$, we find that

$$E[(m_i' - n_i)^2] = \frac{(n_i - \alpha_{<i})^2 p_i^2}{e_i^2} \frac{\gamma^2}{\sum_k \alpha_{<k} p_k} + (n_i - \alpha_{<i}) q_i / p_i + O(1) \tag{16}$$

for the limit process: $n_i \to \infty$ such that n_i/n_j is constant, all i, j. Besides being almost unbiased m_i' is also most efficient of the class of estimates m_i^* which satisfy obvious conditions.

It is worth examining the gain in information that constant catchability and knowledge of the effort bring. The information on n_1 provided by m_1' represents a total gain as there is none when the effort is not known. Otherwise,

$$F_k' = \frac{\operatorname{var}(n_k')}{\operatorname{var}(m_k')} = \frac{1 + (\alpha_{<k+1} - \alpha_{<k})/\alpha_{<k} p_k}{1 + (\alpha_{<k+1} - \alpha_{<k})/\sum_l \alpha_{<l} p_l},$$

on replacing $(n_k - \alpha_{<k}) p_k$ by $\alpha_{<k+1} - \alpha_{<k}$ and e_k^2/γ^2 by p_k^2/q_k. In order to evaluate F_k' numerically, some assumptions must be made about the p_k and about the rate of immigration. First of all, suppose that $p_k = p$ for all k. Secondly, suppose that $\alpha_{<k+1} - \alpha_{<k}$ is constant for all k; that is, the expected number of new individuals caught at each sample is constant. Taken in combination with the first assumption, this implies that $n_k - n_{k-1} = n_1 p$ for all k; that is, the number of immigrants between consecutive samples is constant and equal to the expected size of the first sample. The formula for F_k' can be shown to be very insensitive

to the particular assumptions made about the rate of immigration, and we have merely chosen those which make the formula as simple as possible. It now becomes

$$F'_k = \frac{1+[(k-1)p]^{-1}}{1+[s(s-1)p/2]^{-1}}$$

and is tabulated below for $s = 5$ and three values of p. For $p \leqslant 0{\cdot}001$, the formula $F'_k = s(s-1)/2(k-1)$ provides a very good approximation to the previous one.

k p	2	3	4	5
0·001	9·91	4·96	3·31	2·49
0·01	9·18	4·64	3·12	2·36
0·1	5·50	3·00	2·17	1·75

When we come to estimate the number of immigrants between samples, knowledge of the effort is doubly advantageous and results in a very considerable gain in information. For, $\operatorname{var}(m'_k - m'_{k-1}) = \operatorname{var}(m'_k) + \operatorname{var}(m'_{k-1}) - 2\operatorname{cov}(m'_k, m'_{k-1})$ and not only are the two variances smaller than their counterparts $\operatorname{var}(n'_k)$ and $\operatorname{var}(n'_{k-1})$, but their covariance is large and positive whereas $\operatorname{cov}(n'_k, n'_{k-1})$ is negligible. In fact,

$$\operatorname{cov}(m'_i, m'_j) = \frac{(\alpha_{<i+1} - \alpha_{<i})(\alpha_{<j+1} - \alpha_{<j})\gamma^2}{e_i e_j \Sigma \alpha_{<k} p_k} + O(1) \quad (i \neq j), \tag{17}$$

and, making the same simplifications as we did for F'_k, we find that

$$G'_h = \frac{\operatorname{var}(n'_h - n'_{h-1})}{\operatorname{var}(m'_h - m'_{h-1})} = 1 + \frac{1}{2p}\left[\frac{1}{h-1} + \frac{1}{h-2}\right] \quad (h = 3, \ldots, s).$$

G'_h is tabulated below for $s = 5$ and three values of p.

h p	3	4	5
0·001	751	418	293
0·01	76·0	42·7	30·2
0·1	8·50	5·17	3·92

2·5. We mentioned in (I) that the capture–recapture model used by Hammersley (1953), when dealing with a large accumulation of data on the ringing of Alpine Swifts, contains a flaw. It is appropriate here to show how this comes about.

Hammersley effectively made two postulates: that the overall likelihood could be taken simply as the product of the individual likelihoods for the captured birds; and that for any such bird, its likelihood need only record its behaviour onwards from the sample in which it was first captured. For $s = 3$, the likelihood arrived at in this way is

$$(p_1q_2q_3)^{u_1}(p_1p_2q_3)^{u_{12}}(p_1q_2p_3)^{u_{13}}(p_1p_2p_3)^{u_{123}}(p_2q_3)^{u_2}(p_2p_3)^{u_{23}}p_3^{u_3}, \tag{18}$$

which may also be written
$$p_1^{a_1}p_2^{a_2}q_2^{a_{<3}-a_2}p_3^{a_3}q_3^{r-a_3},$$

or, for general s,
$$\prod_i p_i^{a_i}q_i^{a_{<i+1}-a_i}.$$

The maximum likelihood estimate of p_i is $a_i/a_{<i+1}$. Dividing this into a_i to obtain an estimate of n_i, we get $a_{<i+1}$. In other words, the population size at any time is estimated by the number of individuals captured up to that time, which is plainly unsatisfactory. The reason why this model is at fault is indicated by comparing (18) with (1′) and with (7) for $s = 3$. The latter can be written

$$p[\{u_w\} \mid \{a_{<i+1}\}] = \frac{a_{<2}!}{u_1!\,u_{12}!\,u_{13}!\,u_{123}!}(q_2q_3)^{u_1}(p_2q_3)^{u_{12}}(q_2p_3)^{u_{13}}(p_2p_3)^{u_{123}}\frac{(a_{<3}-a_{<2})!}{u_2!\,u_{23}!}q_3^{u_2}p_3^{u_{23}}. \tag{19}$$

A remark worth making here is that, even if (18) is replaced by a true density such as (19) and n_i is estimated by
$$\tilde{n}_i = a_i/\tilde{p}_i,$$
the sampling variance of $\tilde{n}_i$ is not only attributable to that of $\tilde{p}_i$ but also to that of a_i (and also to their covariance). This remark is of course implicit in §§ 2·3 and 2·4 where, for other reasons, we took
$$\tilde{n}_i = (a_{<i+1}-a_{<i})/\tilde{p}_i + a_{<i}.$$

3. DEATH BUT NO IMMIGRATION

3·1. In this section, suffices i, j take all values from 1 to s and suffices k, l all values from 1 to $s-1$. p_i is now the conditional probability that an individual is captured at the ith sample, given that it is then alive. Let $a_{>k}$ be defined as the number of individuals caught after the kth sample, with a corresponding definition for $a_{k.>k}$.

Let χ_k be the conditional probability of an individual not being caught after the kth sample given that it is alive at the time of the kth sample. Then, for $s = 3$,

$$\begin{aligned}\chi_1 &= 1-\phi_1+\phi_1q_2(1-\phi_2)+\phi_1q_2\phi_2q_3\\ &= 1-\phi_1p_2-\phi_1\phi_2q_2p_3,\\ \chi_2 &= 1-\phi_2+\phi_2q_3,\\ &= 1-\phi_2p_3.\end{aligned}$$

Let π_0 denote the probability that an individual is never caught and π_i the probability that it is caught for the last time at the ith sample. Then, for $s = 3$,

$$\pi_0 = q_1\chi_1, \quad \pi_1 = p_1\chi_1, \quad \pi_2 = \phi_1p_2\chi_2, \quad \pi_3 = \phi_1\phi_2p_3$$

and
$$\pi_0+\pi_1+\pi_2+\pi_3 = 1,$$

which has an obvious probability interpretation.

The probability density of the u_w is

$$p[\{u_w\}] = \frac{n!}{(n-r)!\prod_w u_w!}\pi_0^{n-r}\prod_w P_w^{u_w},$$

where, for $s = 3$,

$$\begin{aligned} P_1 &= \pi_1, & P_3 &= q_1q_2\pi_3, \\ & & P_{13} &= p_1q_2\pi_3, \\ P_2 &= q_1\pi_2, & P_{23} &= q_1p_2\pi_3, \\ P_{12} &= p_1\pi_2, & P_{123} &= p_1p_2\pi_3. \end{aligned}$$

Hence,

$$p[\{u_w\}] = \frac{n!}{(n-r)!\prod_w u_w!}\pi_0^{n-r}\pi_1^{r-a_{>1}}\pi_2^{a_{>1}-a_{>2}}\pi_3^{a_{>2}}p_1^{a_{1.>1}}p_2^{a_{2.>2}}q_1^{a_{>1}-a_{1.>1}}q_2^{a_{>2}-a_{2.>2}}.$$

3·2. Define $a_{>0} = r$ and $a_{>s} = 0$. Then, for general values of s,

$$p[\{u_w\}] = \frac{n!}{(n-r)!\prod_w u_w!}\pi_0^{n-r}\prod_i \pi_i^{a_{>i-1}-a_{>i}}\prod_k p_k^{a_{k.>k}}q_k^{a_{>k}-a_{k.>k}}. \tag{20}$$

Besides n, the original parameters were $\{\phi_k\}$ and $\{p_i\}$, a total of $2s-1$. We have now changed to π_0, $\{\pi_i\}$ and $\{p_k\}$ subject to the constraints

$$\pi_0+\sum_i \pi_i = 1, \quad \text{and} \quad \pi_0p_1-\pi_1q_1 = 0.$$

The effective number of parameters is therefore $2s-2$, a reduction of one. The reason for this is that, if $p[\{u_w\}]$ is written as a function of $\{\phi_k\}$ and $\{p_i\}$, ϕ_{s-1} and p_s only appear in the combination $\phi_{s-1}p_s$ and are therefore really only one parameter. In other words, they are non-identifiable.

Maximizing (20) with Lagrangian multipliers λ_1, λ_2 for the two constraints, we find that

$$\begin{aligned} \lambda_1 = 0, \quad \lambda_2 &= \tilde{n}, \\ \tilde{p}_k &= \frac{a_{k.>k}}{a_{>k}}, \\ \tilde{n}\tilde{\pi}_i &= a_{>i-1}-a_{>i}. \end{aligned} \tag{21}$$

Let us define $\Phi_1 = 1$ and otherwise

$$\Phi_i = \phi_1\phi_2\ldots\phi_{i-1},$$

the probability of survival up to the time of the ith sample. Then

$$\Phi_kp_k = \pi_k+p_k(\pi_{k+1}+\ldots+\pi_s),$$

with obvious probability interpretation. Therefore

$$n\Phi_k = n\pi_k/p_k+(n\pi_{k+1}+\ldots+n\pi_s).$$

Now let $\tilde{n}_k$ denote the estimate of $n\Phi_k$, which is the expected population size at the time of the kth sample. Then substituting the above estimates of $n\pi_i$,

$$\tilde{n}_k = (a_{>k-1}-a_{>k})/\tilde{p}_k+a_{>k} \tag{22}$$

and hence, substituting for $\tilde{p}_k$,

$$\tilde{n}_k = \frac{a_ka_{>k}}{a_{k.>k}}.$$

3·3. At this stage, the argument becomes so nearly parallel to that of §§ 2·3 and 2·4 that we can omit almost all of it. The starting point is the comparison of (21) and (22) with (6) and (8). Next, $n-r$ and $\{a_{>i-1}-a_{>i}\}$ are distributed multinomially with parameters n, π_0 and $\{\pi_i\}$. Therefore,

$$p[\{a_{>i-1}\}] = \frac{n!}{(n-r)!\prod\limits_i (a_{>i-1}-a_{>i})!}\,\pi_0^{n-r}\prod_i \pi_i^{a_{>i-1}-a_{>i}}.$$

Hence
$$p[\{u_w\}\,|\,\{a_{>i-1}\}] = \frac{1}{\prod\limits_w u_w!}\prod_i (a_{>i-1}-a_{>i})!\prod_k p_k^{a_{k.>k}}q_k^{a_{>k}-a_{k.>k}},$$

which should be compared with (7).

Let
$$n''_k = \frac{(a_k+1)(a_{>k}+1)}{a_{k.>k}+1}-1.$$

Then n''_k is almost unbiased for $n\Phi_k$ as n'_k was for n_k. There is, however, an interesting additional term in the expressions for $\operatorname{var}(n''_k)$ and $\operatorname{cov}(n''_k, n''_l)$ which was not present in the corresponding formulae for n'_k, n'_l. We find that

$$\operatorname{var}(n''_k) = \frac{(n\Phi_k-\alpha_{>k})\,n\Phi_k q_k}{\alpha_{>k}p_k}+n\Phi_k(1-\Phi_k)+O(1)$$

as $n\to\infty$, all other parameters remaining constant. The second term is the variance of the *actual* population size at the time of the kth sample. In the same way,

$$\operatorname{cov}(n''_k, n''_l) = n\Phi_l(1-\Phi_k)+o(1)\quad (k<l),$$

and the first term is the covariance of the actual population sizes at the times of the kth and lth samples.

In (I) we remarked that, if all $\phi_k = 1$, the n''_k contain between them all of the information on n but that their evaluation requires 'differentiated' tagging.

3·4. If $q_i = e^{-e_i/\gamma}$ and the e_i are known, it is still true that

$$\tilde{n}\tilde{\pi}_i = a_{>i-1}-a_{>i}.$$

Hence
$$\tilde{n}\tilde{\Phi}_i = (a_{>i-1}-a_{>i})/\tilde{p}_i+a_{>i}.$$

Let
$$\gamma'' = \frac{\sum\limits_k (a_{>k}-\tfrac{1}{2}a_{k.>k})\,e_k+\tfrac{1}{2}\bar{e}}{\sum\limits_k a_{k.>k}+1},$$

the almost unbiased estimate of γ and let m''_i be the corresponding estimate of $n\Phi_i$. The terms $n\Phi_i(1-\Phi_i)$ and $n\Phi_j(1-\Phi_i)$ appear in the formulae for $\operatorname{var}(m''_i)$, $\operatorname{cov}(m''_i, m''_j)$, $i<j$, in addition to those corresponding to (16) and (17). But, as far as $F''_k = \operatorname{var}(n''_k)/\operatorname{var}(m''_k)$ is concerned, the additional terms may be omitted because, although they are of the same order, $O(n)$, as the others, they are negligible by comparison. Therefore,

$$F''_k = \frac{1+(\alpha_{>k-1}-\alpha_{>k})/\alpha_{>k}p_k}{1+(\alpha_{>k-1}-\alpha_{>k})/\sum\limits_l \alpha_{>l}p_l}.$$

The simplifying assumptions are now that $p_k = p$ and $\alpha_{>k-1}-\alpha_{>k}$ is constant for all k. Taken in conjunction, these two assumptions imply that $n\Phi_k(1-\phi_k) = n\Phi_s p_s$, that is the expected number of deaths between consecutive samples is constant and equal to the expected size of the last sample. With these simplifications,

$$F''_k = F'_{s+1-k}\quad (k = 1, 2, \ldots, s-1).$$

Also,
$$G''_h = \frac{\operatorname{var}(n''_{h-1} - n''_h)}{\operatorname{var}(m''_{h-1} - m''_h)} = G'_{s+2-h} \quad (h = 2, \ldots, s-1).$$

The experimenter will probably be more interested in estimating the survival probability ϕ_{h-1} than $n\Phi_{h-1} - n\Phi_h$, the expected number of deaths. Both n''_h/n''_{h-1} and m''_h/m''_{h-1} are biased estimates of ϕ_{h-1} but the biases are $O(n^{-1})$ and can easily be evaluated, if desired, using the δ-technique. The δ-technique also shows that, neglecting terms $O(n^{-1})$,

$$\frac{\operatorname{var}(n''_h/n''_{h-1})}{\operatorname{var}(m''_h/m''_{h-1})} = \frac{\operatorname{var}(n''_h - \phi_{h-1} n''_{h-1})}{\operatorname{var}(m''_h - \phi_{h-1} m''_{h-1})}$$

and, provided ϕ_{h-1} is not too much less than one, the latter ratio is well represented by G''_h.

3·5. Throughout (I) and so far in this paper, we have been exclusively occupied with estimation and have said nothing about testing hypotheses. In this subsection, the likelihood-ratio method will be adapted for the purpose of testing hypotheses about the values of $\{\phi_k\}$ and $\{p_i\}$.

Consider any density of the form

$$p[\{u_w\}] = \frac{n!}{(n-r)! \prod_w u_w!} Q^{n-r} \prod_w P_w^{u_w},$$

where Q and the P_w are functions of parameters $\{\theta_t\}$ say. Let the log-likelihood of n and $\{\theta_t\}$ be $L_1 = \log p[\{u_w\}]$. Next, consider the conditional density

$$p[\{u_w\} \mid r] = \frac{r!}{\prod_w u_w!} \frac{\prod_w P_w^{u_w}}{(1-Q)^r},$$

and let the corresponding log-likelihood of $\{\theta_t\}$ be $L_2 = \log p[\{u_w\} \mid r]$.

Let $\tilde{L}_1$ and $\tilde{L}_2$ denote the maximum values of L_1 and L_2. We show first that, to a good approximation, $\tilde{L}_1$ differs from $\tilde{L}_2$ only by an additive constant. Equating $\partial L_1/\partial\theta_t$ to zero,

$$\frac{n-r}{Q}\frac{\partial Q}{\partial \theta_t} + \sum_w \frac{u_w}{P_w}\frac{\partial P_w}{\partial \theta_t} = 0. \tag{23}$$

Equating $\Delta_n L_1$ to zero,
$$\log n - \log(n-r) + \log Q = 0. \tag{24}$$

(24) gives $(n-r)/Q = r/(1-Q)$ and, substituting in (23), we have

$$\frac{r}{1-Q}\frac{\partial Q}{\partial \theta_t} + \sum_w \frac{u_w}{P_w}\frac{\partial P_w}{\partial \theta_t} = 0.$$

But this is seen to be the equation $\partial L_2/\partial\theta_t = 0$. Therefore, the maximum likelihood estimates $\tilde{Q}$ and $\tilde{P}_w$ of Q and P_w are the same for L_1 and L_2; and for L_1, $\tilde{n} = r/(1-\tilde{Q})$. To find $\tilde{L}_1$ and $\tilde{L}_2$, an approximation for $n!$ is required, and we shall assume that $n! = Ke^{-n}n^n$. (This approximation is equivalent to our practice of maximizing L_1 with respect to n by equating $\Delta_n L_1$ to zero rather than $\partial L_1/\partial n$. For, if the exact identity $\Delta_n \log n! = \log n$ is replaced by the approximate one $\partial \log n!/\partial n = \log n$ and the latter is integrated, it leads directly to $n! = Ke^{-n}n^n$. Using this approximation instead of Stirling's, $n! = \sqrt{(2\pi)}\, e^{-n}n^{n+\frac{1}{2}}$, results in an increase of $\frac{1}{2}$ in the maximum likelihood estimate of n, a difference which can be ignored as it is $O(n^{-1})$.) Hence,

$$\begin{aligned}\tilde{L}_1 &= \tilde{n}[\log \tilde{n} - 1] - (\tilde{n} - r)[\log(\tilde{n} - r) - 1] - \log \prod_w u_w! + (\tilde{n} - r)\log \tilde{Q} + \sum_w u_w \log \tilde{P}_w \\ &= r[\log r - 1] - \log \prod_w u_w! - r\log(1 - \tilde{Q}) + \sum_w u_w \log \tilde{P}_w \\ &= \tilde{L}_2 - \log K,\end{aligned}$$

which proves our assertion.

Let H_0 denote a hypothesis which reduces the number of degrees of freedom of $\{\theta_t\}$ by ν and H the alternative hypothesis allowing $\{\theta_t\}$ full freedom. Then, as L_2 has the properties required for maximum-likelihood large sample theory, we can say, using self-explanatory notation, that

$$x_2 = -2[\tilde{L}_{2,H_0} - \tilde{L}_{2,H}]$$

is approximately distributed as χ^2_ν. More precisely (Bartlett, 1953),

$$P[a \leqslant x_2 < b \mid r] = \int_a^b f(\chi^2_\nu)\, d\chi^2_\nu + O(r^{-1}) \quad \text{as} \quad r \to \infty.$$

Let $$x_1 = -2[\tilde{L}_{1,H_0} - \tilde{L}_{1,H}].$$

Then, since $\tilde{L}_{1,H} = \tilde{L}_{2,H} - K$ and, similarly, $\tilde{L}_{1,H_0} = \tilde{L}_{2,H_0} - K$,

$$x_1 = x_2.$$

Therefore, if P_r is the probability of r individuals,

$$P[a \leqslant x_1 < b] = \sum_{r \leqslant hn} P_r P[a \leqslant x_2 < b \mid r] + \sum_{r > hn} P_r P[a \leqslant x_2 < b \mid r],$$

where $0 < h < 1 - Q$ such that hn is integral. Now

$$\sum_{r \leqslant hn} P_r P[a \leqslant x_2 < b \mid r] \leqslant \sum_{r \leqslant hn} P_r = O(n^{-\frac{1}{2}} c^n) = o(n^{-1}),$$

where $$c = \left(\frac{1-Q}{h}\right)^h \left(\frac{Q}{1-h}\right)^{1-h} < 1.$$

Also, $$\begin{aligned}\sum_{r > hn} P_r P[a \leqslant x_2 < b \mid r] &= \sum_{r > hn} P_r \left[\int_a^b f(\chi^2_\nu)\, d\chi^2_\nu + O(r^{-1})\right] \\ &= [1 - o(n^{-1})] \int_a^b f(\chi^2_\nu)\, d\chi^2_\nu + O(n^{-1}).\end{aligned}$$

Hence, $$P[a \leqslant x_1 < b] = \int_a^b f(\chi^2_\nu)\, d\chi^2_\nu + O(n^{-1})$$

which is in the same form as the usual likelihood-ratio test.

There are two applications of this test for which the algebra has already been performed. In each case H denotes the hypothesis allowing complete freedom to the parameters $\{\phi_k\}$ and $\{p_i\}$ and it is therefore easy to find $\tilde{L}_{1,H}$. As the first example, consider H_0: $\phi_k = 1$ all k. $\tilde{L}_{1,H_0}$ can be found from (I) and $\nu = (2s-2) - s = s - 2$. (If the test dictates that H_0 be accepted it will of course not mean that we believe it to be exactly true since this is strictly impossible for an animal population. It will mean, rather, that the $\tilde{\phi}_k$ cannot be regarded as reliable estimates of the ϕ_k.) As the second example, consider H_0: $q_i = e^{-e_i/\gamma}$ where the e_i are known. This will test whether or not catchability is constant. $\tilde{L}_{1,H_0}$ can be found from § 3·4 and, again, $\nu = (2s-2)-s = s-2$.

Finally, a remark concerning hypothesis testing when there is immigration but no death. The above adaptation of the likelihood-ratio method is not applicable to the density (3). There is, however, a way round this difficulty. We can set up a formal model for immigration but no death which is the exact inverse of the model for death but no immigration. Namely, let the population be of size n at the time of the last sample and let $1-\psi_k$ be the conditional probability that any individual immigrated between the $(k-1)$th and kth samples given that it was in the population at the time of the kth, $k = 2, 3, \ldots, s$. This model yields just the same estimates as (3) and any likelihood-ratio test can be written down automatically from the corresponding one for death but no immigration. (For this model $p[\{u_w\}]$ is most quickly arrived at directly as a multinomial probability, but it can also be developed as a chain of conditional probabilities $P[S_1]\,P[S_2 \mid S_1] \ldots P[S_s \mid S_1, \ldots, S_{s-1}]$ just as easily as for any other model. The main snag about this description of immigration is that it cannot be combined with stochastic death. To include both immigration and death, densities (3) and (20) must be 'combined', and we do this in the next and final section.)

4. IMMIGRATION AND DEATH

4·1. For $s = 3$, the generating function $E[\prod_w t_w^{u_w}]$ of $p[\{u_w\}]$ is clearly (cf. (2))

$$\begin{aligned}(q_1\chi_1 + p_1\chi_1 t_1 + p_1\phi_1 p_2\chi_2 t_{12} + p_1\phi_1 q_2\phi_2 p_3 t_{13} + p_1\phi_1 p_2\phi_2 p_3 t_{123}\\ + q_1\phi_1 p_2\chi_2 t_2 + q_1\phi_1 p_2\phi_2 p_3 t_{23} + q_1\phi_1 q_2\phi_2 p_3 t_3)^{n_1}\\ \times (q_2\chi_2 + p_2\chi_2 t_2 + p_2\phi_2 p_3 t_{23} + q_2\phi_2 p_3 t_3)^{n_2-n_1}(q_3+p_3t_3)^{n_3-n_2},\end{aligned} \tag{25}$$

where χ_1 and χ_2 have been defined in § 3·1. The coefficient of $\prod_w t_w^{u_w}$ in (25) turns out to be a double sum of probabilities which only contracts to a single probability if $n_1 = n_2$ or $\phi_1 = 1$ and if $n_2 = n_3$ or $\phi_2 = 1$.

In general, $p[\{u_w\}]$ is an $(s-1)$-dimensional sum of probabilities which cannot be written as a single expression unless, in every interval between successive samples, there is either no immigration or no death.

4·2. The unwieldy form of $p[\{u_w\}]$ clearly rules out estimation by maximizing the likelihood, but there is an obvious pointer to an alternative method. In all cases when there is no immigration or no death, the method of maximum likelihood is equivalent to equating some of the class sizes to their expected values. We recall from (I) that, for a closed population, these are $\{a_i\}$ and r. It is easily shown that, for immigration but no death, they are $\{a_i\}$, $a_{<3}, \ldots a_{<s}$ and r, and, for death but no immigration, they are $\{a_i\}$, r and $a_{>1}, \ldots, a_{>s-2}$. Therefore, when there is immigration and death, it seems eminently reasonable to equate

$$\{a_i\}, \quad a_{<3}, \ldots, a_{<s}, \quad r, \quad a_{>1}, \ldots, a_{>s-2} \tag{26}$$

to their expected values. We shall indicate how to solve the resulting equations by supposing that there are four samples, this being the minimum number from which the general solution may be inferred.

For $s = 4$, the aforesaid equations can be arranged as

$$a_1 = n_1 p_1, \tag{27a}$$

$$a_2 = [n_1\phi_1 + n_2 - n_1]\,p_2, \tag{27b}$$

$$a_3 = [n_1\phi_1\phi_2 + (n_2 - n_1)\,\phi_2 + n_3 - n_2]\,p_3, \tag{27c}$$

$$a_4 = [n_1\phi_1\phi_2\phi_3 + (n_2 - n_1)\,\phi_2\phi_3 + (n_3 - n_2)\,\phi_3 + n_4 - n_3]\,p_4, \tag{27d}$$

$$r - a_{>1} = n_1 p_1 \chi_1, \tag{28a}$$

$$a_{>1} - a_{>2} = [n_1\phi_1 + n_2 - n_1]\,p_2\chi_2, \tag{28b}$$

$$a_{>2} - a_{>3} = [n_1\phi_1\phi_2 + (n_2 - n_1)\,\phi_2 + n_3 - n_2]\,p_3\chi_3, \tag{28c}$$

$$a_{<2} = n_1 p_1, \tag{29a}$$

$$a_{<3} - a_{<2} = [n_1\phi_1 q_1 + n_2 - n_1]\,p_2, \tag{29b}$$

$$a_{<4} - a_{<3} = [n_1\phi_1\phi_2 q_1 q_2 + (n_2 - n_1)\,\phi_2 q_2 + n_3 - n_2]\,p_3, \tag{29c}$$

where (27*a*) reappears, for convenience, as (29*a*).

Combining (27*a*, *b*, *c*) with (28*a*, *b*, *c*), we have $\chi_k = (a_{>k-1} - a_{>k})/a_k$ or $1 - \chi_k = a_{k.>k}/a_k$, $k = 1, 2, 3$. There is a recurrence relationship between successive χ_k which may be written

$$\phi_k q_{k+1}(1 - \chi_{k+1}) + \phi_k p_{k+1} = 1 - \chi_k \quad (k = 1, 2).$$

Hence

$$\phi_1 q_2 \frac{a_{2.>2}}{a_2} + \phi_1 p_2 = \frac{a_{1.>1}}{a_1}, \tag{30a}$$

$$\phi_2 q_3 \frac{a_{3.>3}}{a_3} + \phi_2 p_3 = \frac{a_{2.>2}}{a_2}. \tag{30b}$$

If we multiply (29*a*) by $\phi_1 p_2$, add it to (29*b*) and refer to (27*b*), we obtain

$$a_{<3} - a_{<2} + \phi_1 p_2 a_{<2} = a_2$$

or

$$\phi_1 p_2 a_{<2} = a_{<2.2}. \tag{31a}$$

Similarly, multiplying (29*a*) by $\phi_1\phi_2 p_3$, (29*b*) by $\phi_2 p_3$ and adding them to (29*c*),

$$\phi_2 p_3 (a_{<3} - a_{<2}) + \phi_1\phi_2 p_3 a_{<2} = a_{<3.3}. \tag{31b}$$

Equations (30) and (31) enable us to solve for p_2, p_3 and ϕ_1, ϕ_2. Hence, from (27*b*, *c*), we can estimate $n_1\phi_1 + n_2 - n_1$ and $n_1\phi_1\phi_2 + (n_2 - n_1)\,\phi_2 + n_3 - n_2$, the expected population sizes at the times of the second and third samples. Also $n_3 - n_2$, the number of immigrants between these two samples. We note that (27*d*) has not proved to be of any use.

In general, it is possible to solve for $p_2, \ldots, p_{s-1}$ and hence for the expected population sizes at the times of all but the first and last samples. Also for the survival rate in all but the last interval and the number of immigrants in all but the first and last intervals. The remaining parameters cannot be estimated and it is fairly obvious that this would hold true of any method of estimation.

Since the distribution of the variables (26) is simply a combination of independent multinomial distributions, it is theoretically straightforward to find formulae for the variances and covariances of all the estimates by using the δ-technique. However, it would be too tedious to compile these formulae for the general case and we shall not attempt it here.

4·3. If $q_i = e^{-e_i/\gamma}$ and the e_i are known, we can again appeal to moment equations for the purpose of estimation. In (I) we remarked that, for a closed population, maximizing the likelihood is equivalent to equating $\sum_i a_i e_i/p_i$ and r to their expected values. When there is immigration but no death the corresponding variables are $\sum_i a_i e_i/p_i$, $a_{<2}\ (= a_1)$, $a_{<3}, \ldots, a_{<s}$ and r, and when there is death but no immigration they are $\sum_i a_i e_i/p_i$, r and $a_{>1}, \ldots, a_{>s-2}$, $a_{>s-1}\ (= a_s)$. With both immigration and death we therefore equate $\Sigma a_i e_i/p_i$, $a_{<2}, \ldots, a_{<s}$, r, $a_{>1}, \ldots, a_{>s-1}$ to their expected values, thus providing $2s$ equations. These enable us to estimate all of the $2s$ unknowns γ, $\{n_i\}$ and $\{\phi_k\}$; except for $s = 2$ when the $a_1e_1/p_1 + a_2e_2/p_2$ equation is implied by the $a_{<2}$ and $a_{>1}$ equations and there are only three equations for the four unknowns.

Without actually performing the straightforward but tedious task of finding the variances of the estimates, we can be confident that knowledge of the effort provides very considerable gains in information.

I am indebted to the referee for his most helpful comments.

REFERENCES

BARTLETT, M. S. (1953). Approximate confidence intervals. II. More than one unknown parameter. *Biometrika*, **40**, 306–17.

DARROCH, J. N. (1958). The multiple-recapture census. I. Estimation of a closed population. *Biometrika*, **45**, 343–59.

HAMMERSLEY, J. M. (1953). Capture-recapture analysis. *Biometrika*, **40**, 265–78.

LESLIE, P. H., CHITTY, DENNIS & CHITTY, HELEN (1953). The estimation of population parameters from data obtained by means of the capture-recapture method. III. An example of the practical applications of the method. *Biometrika*, **40**, 137–69.

The Two-Sample Capture-Recapture Census
When Tagging and Sampling are Stratified

By J. N. Darroch
Statistical Laboratory, University of Manchester.

1. Introduction.

1.1 We start by recalling the capture-recapture argument used for the simplest type of census with only two samples and negligible death and emigration rates.

Let a animals be taken from a population, marked and put back into it. After allowing time for these a individuals to "mix" with the others, let a second sample be taken and suppose that it comprises b unmarked individuals and c marked ones. Then, if it is assumed that *every individual has the same probability p of being a member of the second sample*, p is estimated by $\tilde{p} = c/a$ and, if n is the number of unmarked individuals in the population at the time of the second sample, n is estimated by $b/\tilde{p} = ab/c$. We shall denote this estimate by $\tilde{n}_P$ and refer to it as the Petersen estimate, although this name is usually given to $\tilde{n}_P + a = a(b+c)/c$, the estimate of total population size.

The italicised assumption is both the essence and the weakness of this and all other capture-recapture arguments and is implicit in most work on the subject. (See Darroch (1958, 1959) for a short review of the literature and a discussion of the multiple-recapture census.) In practice it can be violated in many ways which may be summarised roughly as follows:

(i) Animals can differ in their inherent catchability. (If p is the probability of an animal being captured when it is exposed to ε units of catching effort, we may define its catchability α by $p = 1-e^{-\alpha\varepsilon}$.) If this difference is known to be correlated with a recognisable characteristic such as species, sex, age, size, the problem reduces to a separate Petersen estimation of each of the sub-populations in which the characteristic is reasonably homogeneous and in which p may therefore be assumed uniform. On the other hand, if there is no such characteristic, $\tilde{n}_p$ will be a biased estimate of n unless the average probability of the marked individuals, $\bar{p}_m$ say, is equal $\bar{p}_u$, the average probability of the unmarked individuals. (In this case there is even a very small advantage in that the error variance of $\tilde{n}_p$ is smaller than it would be if the probabilities were equal.) The danger is that, in catching the first sample, one obtains individuals with a higher average catchability, making $\bar{p}_m > \bar{p}_u$ and giving $\tilde{n}_p$ a negative bias.

(ii) The catchability of an animal can be altered once it has been captured and marked. Thus, the caught ones may become comparatively "trap-shy", making $\bar{p}_m < \bar{p}_u$, or "trap-happy" making $\bar{p}_m > \bar{p}_u$. Further, when we take mortality (and emigration) into account and let ϕ_m and ϕ_u be the probabilities of survival to the time of the second sample, c/a estimates $\phi_m \bar{p}_m$ and therefore, provided $\bar{p}_m = \bar{p}_u$, ab/c estimates n/ϕ_m. If $\phi_m = \phi_u$, this is roughly the population size at the time of the first sample, neglecting immigration. However, if the marked animals incur any physical disability, it may happen that $\phi_m < \phi_u$, making $\tilde{n}_p$ a positively biased estimate of n/ϕ_u.

(iii) The probability p can vary considerably over the region occupied by the population, partly because the animals are more catchable in one locality than another and also because the effort expended in catching them is not uniform over the region.

1.2. It is much easier to describe the ways in which p may vary than it is to try and make allowance for them. The only possible way of coping with (i) would be to take a sequence of samples and assume that the new values of α had a given frequency function $f(\alpha;\xi)$ say, ξ denoting unknown parameters which could be estimated from the recapture data. Again, to allow for (ii) one would need more than two samples and some assumption about the way in which an animal's value of α changes each time it is captured.

1.3. By contrast, it is possible to modify the two-sample experiment to allow for (iii) and the modifications are the subject of this paper. We shall suppose that, at the time of the second sample, the region occupied by the population is divided into t sub-regions or strata in each of which p can be assumed uniform. Also that, at the time of the first sample, the region can be divided into s strata where each has the property that when the marked animals are 'randomly' released within it they have the same probability distribution of moving to the t strata. The second-sample stratification necessitates the first-sample one since it turns out that, in order to estimate N, s must be greater than or equal to t. The experimenter is required to use a distinctive mark in each of the s strata in order that he can record the stratum of origin of the recaptured animals.

Chapman and Junge (1956) mainly considered the experiment with $s = t$ using probability models which, to contrast them with ours, may be termed non-parametric. That is, only average movement and catching behaviour were specified; equal probabilities and independence of behaviour were not assumed except when it came to finding error variances. We shall start by making the latter assumptions and then relax some of them in an exact, parametric, fashion. Most of the problems raised in this paper were first raised by Chapman and Junge and several of the answers obtained are the same as or similar to theirs, but we shall not attempt to make a detailed comparison as it would lengthen the present work too greatly.

The other paper on this problem is by Schaefer (1951), arising from a tagging experiment on a population which was stratified temporally instead of spatially. Chapman and Junge pointed out that Schaefer's estimate was not a consistent one but they did not analyse his data. We do so, in § 6.

Beverton and Holt (1957) considered a similar problem but treated it deterministically and assumed that the marked animals are released in only one stratum. This effectively makes $s=1$ and consequently $s < t$ but they were able to estimate the unknown parameters by assuming that catchability was constant and the catching effort known for each stratum.

2. Probability Model and Estimation.

2.1. Subscript i will refer to the s first-sample strata and subscript j to the t second-sample strata.

Of the a animals tagged, let a_i be released in the ith stratum and, of these, let c_{ij} be caught in the jth stratum. Let $c_{i.} = \sum_j c_{ij}$, $c_{.j} = \sum_i c_{ij}$. Then $\sum_i c_{i.} = \sum_j c_{.j} = c$. Of the b untagged animals which are caught, let b_j denote the number caught in the jth stratum.

Let θ'_{ij} denote the probability of moving to the jth stratum for each of the a_i individuals, and write $\sum_j \theta'_{ij} = \phi'_i$, the probability of their survival to the time of the second sample. For the most part, the ϕ'_i will be assumed equal, with common value ϕ', say, since this is usually a reasonable biological assumption but, in §[illegible].6 and §6 they will be estimated on the supposition that they differ in a certain restricted fashion.

Let n'_j denote the number of unmarked individuals in the jth stratum (including possible immigrants from outside the region) and let $\sum_j n'_j = n'$. The fundamental assumption will be that all individuals in the jth stratum, marked or unmarked, have the same probability, p'_j say, of being a member of the second sample.

2.2. Assuming that all animals move and are caught independently of each other, the probability density of the c_{ij} is

$$p[\{c_{ij}\}] = \frac{\prod_i a_i!}{\prod_i (a_i - c_{i.})! \prod_{i,j} c_{ij}!} \prod_i \left(1 - \sum_j \theta'_{ij} p'_j\right)^{a_i - c_{i.}} \prod_{i,j} \left(\theta'_{ij} p'_j\right)^{c_{ij}}. \qquad (1)$$

Notice that $\{\theta'_{ij}\}$, $\{p'_j\}$ are thus seen to be non-identifiable to the extent of a multiplicative constant for their likelihood is the same as that of $\{k\theta'_{ij}\}$, $\{k^{-1} p'_j\}$. To overcome this non-identifiability, we shall work with $\theta_{ij} = \theta'_{ij}/\phi'$, $p_j = \phi' p'_j$.

~~$p = \phi' p$~~ . In the same way, let $n_j = n'_j/\phi'$, $n = \sum_j n_j$. Thus, $\{\theta_{ij}\}$, $\{p_j\}$, $\{n_j\}$, n are the true probabilities and sizes when and only when $\phi' = 1$.

Let $L = \log p\left[\{c_{ij}\}\right]$ denote the log-likelihood of $\{\theta_{ij}\}$, $\{p_j\}$. The number of these parameters is st-s (for the θ_{ij} subject to the constraints $\sum_j \theta_{ij} = 1$, $i = 1,2,\ldots,s$) plus t (for the p_j), making a total of

$$N(\theta,p) = st + t - s.$$

Compare this with the number of parameters involved when (1) is expressed in terms of $\psi_{ij} = \theta_{ij}\,p_j$, namely

$$N(\psi) = st.$$

(The point of introducing the ψ system is that it reveals in a simple manner certain important features of the θ, p system.) We see that

$$N(\theta,p) \gtreqless N(\psi)$$

according as $s \lesseqgtr t$

and the three possibilities:- $s < t$, $s > t$, $s=t$ will be treated separately.

2.3. When $s < t$, $N(\theta,p) > N(\psi)$ and $\{\theta_{ij}\}$, $\{p_j\}$ are not identifiable: only $N(\psi)$ functions of them are. The maximum-likelihood equations in the ψ system are easily seen to take the simple form

$$\tilde{\psi}_{ij} = c_{ij}/a_i, \qquad (2)$$

and if we try to estimate $\{p_j\}$ using (2) and $\sum_j \psi_{ij}\tilde{p}_j^{-1} = 1$ (or, equivalently, maximising L with respect to $\{\theta_{ij}\}$, $\{p_j\}$) we only get s equations $\sum_j c_{ij}\tilde{p}_j^{-1} = a_i$ for t unknowns. Since $\tilde{p}_j$ cannot be determined, neither can $\tilde{n}_j = b_j/\tilde{p}_j$, $\tilde{n} = \sum_j \tilde{n}_j$.

2.4. When $s > t$, $N(\theta, p) < N(\psi)$ and the ψ_{ij} are mathematically dependent functions of $\{\theta_{ij}\}$, $\{p_j\}$. This means that equations (2) are invalid since they are obtained assuming the ψ_{ij} to be independent. Consequently, we must maximise L directly with respect to $\{\theta_{ij}\}$, $\{p_j\}$. $\partial L/\partial p_j = 0$ gives

$$\sum_i \frac{(a_{i\cdot}-c_{i\cdot})\theta_{ij}p_j}{1-\sum_j \theta_{ij}p_j} = c_{\cdot j}. \tag{3}$$

Introducing s Lagrangian multipliers $\{\lambda_i\}$ corresponding to the constraints $\sum_j \theta_{ij} = 1$, $\partial L/\partial \theta_{ij} = 0$ gives

$$\frac{(a_{i\cdot}-c_{i\cdot})\theta_{ij}p_j}{1-\sum_j \theta_{ij}p_j} = c_{ij} + \lambda_i \theta_{ij} \tag{4}$$

It is not possible to solve (3) and (4) explicitly, except in the trivial case $t = 1$. They must therefore be solved numerically and the information matrix inverted to find error variances and covariances. This is not an attractive proposition and is made even less so by two further considerations. The first is that the solution to (3) and (4) is not, in general, unique though there is (see Chanda (1954)) only one consistent solution $\{\tilde{\theta}_{ij}\}$, $\{\tilde{p}_j\}$ say. Secondly, it may not be easy to recognise this solution since $\tilde{\theta}_{ij}$ and $\tilde{p}_j$ are not bound to lie in the interval $[0,1]$ although, provided the sample sizes are large as they must be for the success of any capture-recapture experiment, they will invariably do so simply because they are consistent.

Fortunately, it is possible to obtain much simpler estimation equations than the above when $s > t$ by relaxing the assumption of equal ϕ'_i. We do this in § 2.6.

2.5. When $s = t$, $N(\theta, p) = N(\psi)$ and the two parametric systems are interchangeable provided only that the transformation $\psi_{ij} = \theta_{ij} p_j$ is one-one. In matrix notation, it is

$$\underset{\sim}{\Psi} = \underset{\sim}{\Theta}\,\underset{\sim}{D}_p \qquad (5)$$

where $\underset{\sim}{\Psi} = (\psi_{ij})$, $\underset{\sim}{\Theta} = (\theta_{ij})$ and $\underset{\sim}{D}_p = (\delta_{ij} p_j)$, the diagonal matrix whose elements are those of the vector $\underset{\sim}{p} = (p_j)$. Now $\underset{\sim}{\Theta}\,\underset{\sim}{1} = \underset{\sim}{1}$, where $\underset{\sim}{1}$ is the vector of s 1's, and therefore $\underset{\sim}{\Psi}\,\underset{\sim}{D}_p^{-1}\underset{\sim}{1} = \underset{\sim}{1}$. Therefore, provided that $\underset{\sim}{\Psi}$ is non-singular, that is provided $\underset{\sim}{\Theta}$ is non-singular,

$$\underset{\sim}{D}_p^{-1}\underset{\sim}{1} = \underset{\sim}{\Psi}^{-1}\underset{\sim}{1}$$

or

$$\underset{\sim}{\rho} = \underset{\sim}{\Psi}^{-1}\underset{\sim}{1} \qquad (6)$$

where $\underset{\sim}{\rho} = (\rho_j) = (p_j^{-1})$.

(5) and (6) show that the transformation is one-one if and only if $\underset{\sim}{\Theta}$ is non-singular.

The maximum-likelihood equations for $\{\tilde{\psi}_{ij}\}$ are

$$\tilde{\psi}_{ij} = c_{ij} / a_i \qquad (7)$$

as in §2.3. They are also the method of moments equations $E[c_{ij}] = c_{ij}$ which, in effect, Chapman and Junge used. $\tilde{\psi}_{ij}$ is obviously a consistent estimate of ψ_{ij} as $a_i \to \infty$. Translating into the θ, p system, we deduce that

$$\tilde{\theta}_{ij}\,\tilde{p}_j = c_{ij} / a_i \qquad (8)$$

are the maximum likelihood equations and give consistent estimates of $\{\theta_{ij}\}$, $\{p_j\}$ provided that $\tilde{\underset{\sim}{\Theta}}$ and $\underset{\sim}{\Theta}$ are non-singular. From (7) it follows that $\tilde{\underset{\sim}{\Psi}}$ (and therefore also $\tilde{\underset{\sim}{\Theta}}$) is non-singular if and only if $\underset{\sim}{C} = (c_{ij})$ is.

Assuming $\underset{\sim}{C}$ to be non-singular, let us find $\tilde{p}_j$.

From (8), we have

$$\tilde{\underset{\sim}{\Theta}} = \underset{\sim}{D}_a^{-1}\underset{\sim}{C}\,\underset{\sim}{D}_p^{-1}$$

where $\underset{\sim}{D}_a = (\delta_{ij} a_i)$. Inverting

$$\tilde{\underset{\sim}{D}}_p\,\tilde{\underset{\sim}{\Theta}}^{-1} = \underset{\sim}{C}^{-1}\underset{\sim}{D}_a.$$

Now $\hat{\underset{\sim}{\Theta}}^{-1}\underset{\sim}{1} = \underset{\sim}{1}$ since $\hat{\underset{\sim}{\Theta}}\underset{\sim}{1} = \underset{\sim}{1}$, therefore

$$\tilde{\underset{\sim}{\beta}} = \underset{\sim}{C}^{-1}\underset{\sim}{D}_a\underset{\sim}{1} = \underset{\sim}{C}^{-1}\underset{\sim}{a} \tag{9}$$

As $n_i \to \infty$, $i = 1,2,\dots,s$, $\underset{\sim}{D}_a^{-1}\underset{\sim}{C}$ converges in probability to $\underset{\sim}{\Theta}\underset{\sim}{D}_\beta$. Therefore, if $\underset{\sim}{\Theta}$ is non-singular, $\tilde{\beta}$ converges in probability to $\underset{\sim}{D}_\beta^{-1}\underset{\sim}{\Theta}^{-1}\underset{\sim}{1} = \underset{\sim}{D}_\beta^{-1}\underset{\sim}{1} = \beta$, as already stated. On the other hand, if $\underset{\sim}{\Theta}$ is singular, the elements of $\underset{\sim}{C}^{-1}\underset{\sim}{D}_a$ tend to infinity (some of them, at least) and $\tilde{\underset{\sim}{\beta}}$ diverges.

To see what happens when $\underset{\sim}{C}$ is singular, we must look at the likelihood equations in the θ, p system. These are (3) and (4) and, summing (4) over i, we deduce from them that

$$\sum_i \lambda_i \tilde{\theta}_{ij} = 0, \quad j = 1,2,\dots,n.$$

If $\underset{\sim}{C}$ is non-singular, we know that $\tilde{\underset{\sim}{\Theta}}$ is also and hence $\lambda_i = 0$ all i. Putting $\lambda_i = 0$ in (4) we get $\tilde{\theta}_{ij}\tilde{p}_j = c_{ij}/a_i$, which is (8) again. However, if $\underset{\sim}{C}$ is singular, not all of the λ_i are zero and there is no explicit solution of this kind except when $s = t = 2$.

It is obvious from the foregoing that the experimenter must try and avoid stratifications which make $\underset{\sim}{\Theta}$ singular; and the further Θ is from singularity, the more reliable are $\{\tilde{\theta}_{ij}\}$, $\{\tilde{p}_j\}$, a fact which will be confirmed by the variance formulae in § 3. There are a few trivial cases of singularity which may be foreseeable. For example, if $\theta_{ij} = 0$ all i, the jth stratum is effectively non-existent. If $\theta_{ij} = \theta_{lj}$ all j, the ith and lth strata can be combined. If $\theta_{ij}/\theta_{ik} = \omega$, independent of i, the jth and kth strata can be combined into a Jth say with $\theta_{iJ} = \theta_{ij} + \theta_{ik}$, $p_J = (\omega p_j + p_k)/(\omega + 1)$. The more general instances of singularity are those which are less easily avoided.

If $\underset{\sim}{\Theta}$ is non-singular and the sample sizes are large, a singular $\underset{\sim}{C}$ will be extremely improbable but, if by some freak chance, it does occur, the best procedure will probably be to group two of the first-sample strata and two of the second-sample strata in such a way as to eliminate the singularity. If this is done then $s = t = 2$, the result will be an unstratified analysis and it is of interest that the maximum likelihood solution with λ_1 and λ_2 non-zero throws out strong hints that this is the correct procedure. For the latter solution is

$$\tilde{\theta}_{11} = \tilde{\theta}_{21} = [(a_1 c_{21} - a_2 c_{11}) + (c_{11} c_{22} - c_{12} c_{21})]/(a_1 c_{2\cdot} - a_2 c_{1\cdot}),$$
$$\tilde{\theta}_{12} = \tilde{\theta}_{22} = 1 - \tilde{\theta}_{11}, \quad \tilde{p}_1 = c_{\cdot 1}/a\tilde{\theta}_{11}, \quad \tilde{p}_2 = c_{\cdot 2}/a\tilde{\theta}_{22}.$$

(This solution exists when $\underset{\sim}{C}$ is non-singular but is not consistent as may be seen either by inspection or by applying Chanda's result.) When $c_{11} c_{22} - c_{12} c_{21} = 0$, besides $\tilde{\theta}_{11} = \tilde{\theta}_{21}$ and $\tilde{\theta}_{12} = \tilde{\theta}_{22}$ which indicates a combination of the first-sample strata, it turns out that $\tilde{p}_1 = \tilde{p}_2 = c/a$ which indicates a combination of the second-sample strata.

Finally, let us use (9) to estimate $\underset{\sim}{n} = (n_j)$. Since $\tilde{\underset{\sim}{n}} = (b_j/\tilde{p}_j) = \underset{\sim}{D}_b \tilde{\beta}$,

$$\hat{\underset{\sim}{n}} = \underset{\sim}{D}_b \underset{\sim}{C}^{-1} \underset{\sim}{a}, \tag{10}$$

first found by Chapman and Junge. Also, since $\tilde{n} = \underset{\sim}{1}' \tilde{\underset{\sim}{n}}$,

$$\hat{n} = \underset{\sim}{b}' \underset{\sim}{C}^{-1} \underset{\sim}{a}. \tag{11}$$

2.6. In § 2.4 we were unable to use the simple equations $\tilde{\psi}_{ij} = c_{ij}/a_i$ because $N(\theta, p) < N(\psi)$. We now contrive to make $N(\theta, p) = N(\psi)$ by allowing the ϕ'_i to differ.

Let $\bar{\phi}' = (\sum_i \phi'_i)/s$ and let us work with the parameters

$$\theta_{ij} = \theta'_{ij}/\bar{\phi}', \quad \phi_i = \sum_j \theta_{ij}, \quad p_j = \bar{\phi}' p'_j.$$

This gives a total of $st + t - 1$ parameters, the subtraction of one being due to $\sum_{ij} \theta_{ij} = s$. Instead of imposing $s-1$

constraints $\phi_1 = \phi_2 = \dots = \phi_s$ as we effectively did in § 2.4, let us instead only impose t-1. Then $N(\theta, p) = st$ and (with certain provisos) the estimation equations are $\tilde{\theta}_{ij}\tilde{p}_j = c_{ij}/a_i$. The t-1 (independent) constraints will usually take the form $\phi_i - \phi_l = 0$ or possibly $\phi_i + \phi_k - 2\phi_l = 0$ but we write them generally as

$$\sum_i u_{ki}\phi_i = 0, \quad k = 1,2,\dots,t-1.$$

Also, $$\sum_i u_{ti}\phi = 1$$

if we define $u_{ti} = 1/s$. The matrix $\underset{\sim}{U} = (u_{ki})$, k=1,2,...,t, i=1,2,...,s is of full rank t because the first t-1 rows are independent among themselves and, provided they make sense as constraints, the last row is independent of them. For, if not, the first t-1 rows could be combined to give $\sum_i \phi_i = 0$ which is nonsensical.

Let $\underset{\sim}{v}'$ denote the 1xt vector (0,0,...,0,1). Then

$$\underset{\sim}{U}\,\underset{\sim}{\Theta}\,\underset{\sim}{1} = \underset{\sim}{v}$$

replacing $\underset{\sim}{\Theta}\,\underset{\sim}{1} = \underset{\sim}{1}$ of § 2.5. Adapting the argument of § 2.5 we find that, provided $\underset{\sim}{U}\,\underset{\sim}{\Theta}$ and $\underset{\sim}{U}\,\underset{\sim}{D}_a^{-1}\underset{\sim}{C}$ are non-singular or, in other words, provided $\underset{\sim}{\Theta}$ and $\underset{\sim}{C}$ are of full rank t, consistent estimates of $\underset{\sim}{\rho} = (1/p_j)$ and $\underset{\sim}{\Theta}$ are given by

$$\tilde{\underset{\sim}{\rho}} = (\underset{\sim}{U}\,\underset{\sim}{D}_a^{-1}C)^{-1}\underset{\sim}{v}, \tag{9'}$$
$$\tilde{\underset{\sim}{\Theta}} = \underset{\sim}{D}_a^{-1}\underset{\sim}{C}\,\underset{\sim}{D}_{\tilde{\rho}}.$$

In the trivial case t=1, the $\phi_i = \theta_{i1}$ are estimated without constraints giving

$$\tilde{p}_1 = (\sum_i c_{i1}/a_i)/s, \quad \tilde{\phi}_i = c_{i1}/a_i\tilde{p}_1.$$

Finally, $\underset{\sim}{n} = (n_j) = (n_j'/\bar{\phi}')$ is estimated by

$$\tilde{\underset{\sim}{n}} = \underset{\sim}{D}_b(\underset{\sim}{U}\,\underset{\sim}{D}_a^{-1}\underset{\sim}{C}')^{-1}\underset{\sim}{v} \tag{10'}$$

and $n = n'/\bar{\phi}'$ by

$$\tilde{n} = \underset{\sim}{b}'(\underset{\sim}{U}\,\underset{\sim}{D}_a^{-1}\underset{\sim}{C}')^{-1}\underset{\sim}{v} \tag{11'}$$

3. First and Second Moments of the Estimates.

3.1. We first find the approximate bias of $\tilde{\underset{\sim}{\beta}}$ for the limit process: $\{\theta_{ij}\}$, $\{p_j\}$ constant and all $a_i \to \infty$ in such a way that a_i/a is fixed. Both $\underset{\sim}{\Theta}$ and $\underset{\sim}{C}$ will be assumed to be of rank t, and we shall neglect the probability that C is of rank less than t.

Let $\underset{\sim}{\Gamma} = E[\underset{\sim}{C}] = \underset{\sim}{D}_a \underset{\sim}{\Theta} \underset{\sim}{D}_\beta$

and $\underset{\sim}{C} - \underset{\sim}{\Gamma} = \underset{\sim}{Z} = (z_{ij})$.

When s = t, $\hat{\underset{\sim}{\beta}} = \underset{\sim}{C}^{-1} \underset{\sim}{a}$ and

$$\underset{\sim}{C}^{-1} = [\underset{\sim}{I} - (\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z}) + (\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z})^2 - \cdots + (-1)^{m-1}(\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z})^{m-1}]\underset{\sim}{\Gamma}^{-1} + (-1)^m(\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z})^m\underset{\sim}{C}^{-1}.$$

Therefore, since $\underset{\sim}{\Gamma}^{-1}\underset{\sim}{a} = \underset{\sim}{\beta}$, we have

$$E[\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}] = E[-(\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z}) + (\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z})^2 - \cdots + (-1)^{m-1}(\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z})^{m-1}]\underset{\sim}{\beta} + \underset{\sim}{r}$$

where $\underset{\sim}{r} = (r_i) = (-1)^m E[(\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z})^m \underset{\sim}{C}^{-1}\underset{\sim}{a}]$, the remainder vector.

Let $\underset{\sim}{F} = (f_{ij}) = (\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z})^m$ and let C_{ij} denote the co-factor of c_{ij} in $\underset{\sim}{C}$. Then, since $|C|$, the determinant of $\underset{\sim}{C}$ is assumed numerically greater than or equal to one,

$$|r_j| = |E[\sum_{k,l} f_{jk} C_{lk} a_l / |C|]| \leq \sum_{k,l} |E[f_{jk} C_{lk}]| a_l$$
$$\leq \sum_{k,l} (E[f_{jk}^2] E[C_{lk}^2])^{\frac{1}{2}} a_l$$

Now $E[f_{jk}^2]$ is the sum of products of terms of $\underset{\sim}{\Gamma}^{-1}$ of total power 2m and moments of $\underset{\sim}{Z}$ of order 2m. The former are $O(a^{-2m})$ and the latter, since they are multinomial moments, are $O(a^m)$. Also, C_{lk} is numerically less than $(\prod_i c_i)/c_l$ which is $O(a^{s-1})$. Therefore $|r_j| \leq O(a^{-m+\frac{1}{2}m+s-1+1})$, $= O(a^{-\frac{1}{2}m+s}) \to 0$ for m > 2s. It is therefore permissable to write $E[\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}]$ as the infinite series

$$E[-(\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z}) + (\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z})^2 - (\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z})^3 + \cdots]\underset{\sim}{\beta}$$

The first term is zero since $E[\underset{\sim}{Z}] = \underset{\sim}{0}$, the (2m)th is $O(a^{-m})$ and the (2m+1)th is $O(a^{-m-1})$, $m \geq 1$.

Therefore the bias of $\tilde{\rho}_j = 1/\hat{p}_j$ is $O(a^{-1})$ and this is asymptotically negligible compared with its root mean square which we shall show in § 3.2 to be $O(a^{-\frac{1}{2}})$. The same holds true, when $s > t$, of $= (\underset{\sim}{U}\hat{D}^{-1}_{\underset{\sim}{a}}C')^{-1}\hat{\underset{\sim}{v}}$.

Next, let us augment the above limit process by supposing that all $n_j \to \infty$ in such a way that n_j/n and s/n remain fixed. Let $\underset{\sim}{\beta} = E[\underset{\sim}{b}]$. As $\hat{\underset{\sim}{b}}$ is independent of $\underset{\sim}{C}$,

$$E[\hat{\underset{\sim}{n}} - \underset{\sim}{n}] = D_{\underset{\sim}{\beta}} E[\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}], \quad E[\hat{n} - n] = \underset{\sim}{\beta}' E[\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}]$$

which are both $O(1)$ and negligible compared with the root mean squares which are $O(n^{\frac{1}{2}})$.

3.2. Because the bias of $\hat{\tilde{\underset{\sim}{\beta}}}$ is negligible, its variance-covariance matrix is $E[(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})'] = \underset{\sim}{\Sigma}$ say, and, arguing as in § 3.1, this matrix can be expanded as a doubly infinite series whose terms are $O(a^{-m})$, $m \geq 1$. Suppose $s = t$. Then the leading term, of size $O(a^{-1})$, is

$$E[\underset{\sim}{\Gamma}^{-1}\underset{\sim}{Z}\beta\beta'\underset{\sim}{Z}'\underset{\sim}{\Gamma}'^{-1}]. \tag{12}$$

The (i,l) element of $E[\underset{\sim}{Z}\beta\beta'\underset{\sim}{Z}']$ is

$$E[\textstyle\sum_{jk} z_{ij}\beta_j\beta_k z_{lk}]. \tag{13}$$

Now
$$E[z_{ij} z_{lk}] = \delta_{il} a_i [\delta_{jk}\theta_{ij}p_j - \theta_{ij}p_j\theta_{ik}p_k] \tag{14}$$

since c_{ij} and c_{lk} are independent when $i \neq l$ and, when $i = l$, have a multinomial distribution. Substituting (14) in (13) we get

$$\delta_{il} a_i \left[\sum_j \theta_{ij}/p_j - \left(\sum_j \theta_{ij}\right)\left(\sum_k \theta_{ik}\right)\right] = \delta_{il} a_i \left[\sum_j \theta_{ij}/p_j - 1\right]$$

and this is the (i,l) element of $\underset{\sim}{D}_{\underset{\sim}{a}}\underset{\sim}{D}_{\underset{\sim}{\mu}}$ where $\underset{\sim}{\mu} = (\mu_i)$ and

$$\mu_i = \sum_j \theta_{ij}/p_j - 1. \tag{15}$$

Substituting in (13), we can say that

$$\begin{aligned}\underset{\sim}{\Sigma} &\sim \underset{\sim}{\Gamma}^{-1}\underset{\sim}{D}_{a}\underset{\sim}{D}_{\mu}\underset{\sim}{\Gamma}'^{-1} \\ &= \underset{\sim}{D}_{\beta}\underset{\sim}{\Theta}^{-1}\underset{\sim}{D}_{\mu}\underset{\sim}{D}_{a}^{-1}\underset{\sim}{\Theta}'^{-1}\underset{\sim}{D}_{\beta}\,. \end{aligned} \qquad (16)$$

Notice that, when $\underset{\sim}{\Theta}$ is near to singularity and the elements of $\underset{\sim}{\Theta}^{-1}$ are large, $\hat{\underset{\sim}{\beta}}$ will be an inefficient estimate of $\underset{\sim}{\beta}$.

When $s > t$, (16) is changed to

$$\underset{\sim}{\Sigma} \sim \underset{\sim}{D}_{\beta}(\underset{\sim}{U}\underset{\sim}{\Theta})^{-1}\underset{\sim}{U}\underset{\sim}{D}_{\mu}\underset{\sim}{D}_{a}^{-1}\underset{\sim}{U}'(\underset{\sim}{U}\underset{\sim}{\Theta})'^{-1}\underset{\sim}{D}_{\beta} \qquad (16')$$

where

$$\mu_i = \sum_j \theta_{ij}/p_j - \phi_i^2\,. \qquad (15')$$

3.3. Let $\underset{\sim}{y}$ denote the deviation of $\underset{\sim}{b}$ from its expected value $\underset{\sim}{\beta} = \underset{\sim}{D}_{\beta}\underset{\sim}{n}$. Then

$$\hat{\underset{\sim}{n}} - \underset{\sim}{n} = \underset{\sim}{D}_{b}\tilde{\underset{\sim}{\beta}} - \underset{\sim}{D}_{\beta}\underset{\sim}{\beta} = \underset{\sim}{D}_{\beta}(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) + \underset{\sim}{D}_{y}\tilde{\underset{\sim}{\beta}}\,.$$

Therefore

$$E[(\hat{\underset{\sim}{n}} - \underset{\sim}{n})(\hat{\underset{\sim}{n}} - \underset{\sim}{n})'] = \underset{\sim}{D}_{\beta}\underset{\sim}{\Sigma}\underset{\sim}{D}_{\beta} + E[\underset{\sim}{D}_{y}\tilde{\underset{\sim}{\beta}}\tilde{\underset{\sim}{\beta}}'\underset{\sim}{D}_{y}]\,. \qquad (17)$$

The leading term in (17) is $O(n)$ ($=O(s)$) and the others are $O(n^{-m})$, $m > 0$. Retaining only the former, the second term in (17) may be replaced by $E[\underset{\sim}{D}_{y}\underset{\sim}{\beta}\underset{\sim}{\beta}'\underset{\sim}{D}_{y}] = \underset{\sim}{D}_{n}(\underset{\sim}{D}_{\beta} - \underset{\sim}{I})$ since $E[y_j y_k] = \delta_{jk} n_j p_j (1-p_j)$. Substituting (16) when $s = t$,

$$E[(\hat{\underset{\sim}{n}} - \underset{\sim}{n})(\hat{\underset{\sim}{n}} - \underset{\sim}{n})'] \sim \underset{\sim}{D}_{n}\underset{\sim}{\Theta}^{-1}\underset{\sim}{D}_{\mu}\underset{\sim}{D}_{a}^{-1}\underset{\sim}{\Theta}'^{-1}\underset{\sim}{D}_{n} + \underset{\sim}{D}_{n}(\underset{\sim}{D}_{\beta} - \underset{\sim}{I})\,. \qquad (18)$$

Since $(\hat{n}-n)^2 = \underset{\sim}{1}'(\hat{\underset{\sim}{n}} - \underset{\sim}{n})(\hat{\underset{\sim}{n}} - \underset{\sim}{n})'\underset{\sim}{1}$,

$$\begin{aligned} E[(\hat{n}-n)^2] &\sim \underset{\sim}{n}'\underset{\sim}{\Theta}^{-1}\underset{\sim}{D}_{\mu}\underset{\sim}{D}_{a}^{-1}\underset{\sim}{\Theta}'^{-1}\underset{\sim}{n} + \underset{\sim}{n}'(\underset{\sim}{\beta} - \underset{\sim}{1}) \\ &= \sum_i \eta_i^2 \mu_i / a_i + \sum_j n_j(\beta_j - 1) \end{aligned} \qquad (19)$$

where

$$\underset{\sim}{\eta}' = (\eta_i) = \underset{\sim}{n}'\underset{\sim}{\Theta}^{-1}. \qquad (20)$$

If $\underset{\sim}{\Theta}$ is near to singularity the decreased efficiency of $\hat{\underset{\sim}{\beta}}$ impairs that of $\hat{\underset{\sim}{n}} = (\hat{n}_j)$ as is to be expected, but the efficiency of $\tilde{n} = \sum_j \tilde{n}_j$ is almost unimpaired. We have not so far needed to make any assumption about the movement of the unmarked individuals between the two samples but, to explain the last remark, suppose that they have the same movement pattern as the unmarked individuals. Suppose, also, that there is no immigration and let n_i denote the number of individuals in the ith stratum. There are two ways of formally equating the movement patterns. One way is to think of n'_j as a random variable with expectation $\sum_i n_i\theta'_{ij}$, but this involves some tiresome and unimportant complications. Instead, we shall continue to think of n'_j as a parameter by the device of defining $\{\theta'_{ij}\}$ by the movement of the unmarked animals. Then $\sum_i n_i\theta'_{ij} = n'_j$ or $\sum_i n_i\theta_{ij} = n_j$

$$\underset{\sim}{n}'\underset{\sim}{\Theta} = \underset{\sim}{n}'. \tag{21}$$

Thus, if the movement patterns are the same, $\underset{\sim}{\eta} = \underset{\sim}{n}$. The only point we wish to establish from these considerations is that, even if the movement patterns are different, the η_i are likely to be positive and, since $\sum_i \eta_i = n$, the fact that $\underset{\sim}{\Theta}$ is near to singularity has no effect on (19). On the other hand, if $\underset{\sim}{n}$ and $\underset{\sim}{\Theta}$ are "incompatible" in the sense that some of the η_i are negative, either because of very different movement patterns or because of immigrants increasing the numbers of individuals in the t strata disproportionately, this does increase $E[(\tilde{n}-n)^2]$ because the coefficient of η_i^2 in (19) is positive. Of course, these conclusions only apply to the leading terms of size $O(n)$ in $E[(\tilde{n}-n)^2]$. They do not apply to the terms of smaller order and the latter might have to be taken into account if $\underset{\sim}{\Theta}$ was very near to singularity.

When $s > t$,

(18')

To replace three lines after (20') on page 16.

From (20') we have $\eta' \textcircled{H} = \underset{\sim}{n}'$ and, if the movements of the marked and unmarked animals are the same, $\underset{\sim}{m}' \textcircled{H} = \underset{\sim}{n}'$. However, besides $\underset{\sim}{m}$, there is now an infinity of vectors ξ such that $\xi' \textcircled{H} = \underset{\sim}{n}'$ and we cannot infer that $\eta = \underset{\sim}{m}$. Therefore, the above remarks for s=t cannot be extended for s > t.

and $$E[(\tilde{n}-n)^2] \sim \sum_i \eta_i^2 \mu_i / a_i + \sum_j n_j(\rho_j - 1) \qquad (19')$$

where $$\eta = \underset{\sim}{n}'(\underset{\sim}{U}\Theta)^{-1}\underset{\sim}{U} \qquad (20')$$

If the marked and unmarked animals move in the same way, $\eta = \underset{\sim}{n}$ and all the above remarks for c = 0 may be adapted in an obvious way for c > 0.

Finally, $\tilde{n}$ is a consistent estimate of n in the sense that, for the limit process we are using, $\tilde{n}/n$ converges in probability to one.

4. 'Contagious' Movement And Catching.

4.1. In writing $p[\{c_{ij}\}]$ as a product of multinomial densities (see (1)) it was assumed that the a_i individuals released in the ith stratum move and are caught or not caught (i) independently of those released in any other stratum and (ii) independently of each other. (i) is a very reasonable assumption but (ii) is less likely to hold true in practice and in this section we examine the consequences of relaxing it.

Let us refer to the a_i individuals as $I_1, I_2, \ldots I_{a_i}$. In considering how, for instance, the "experimental histories" of I_α and I_β may be dependent we must distinguish between the two stages: movement to one of the t strata and, once there, being caught or not caught. To assist this distinction, let a_{ij} denote the number of individuals released in the ith stratum who are alive in the jth at the time of the second sample. Then $a_i - \sum_j a_{ij}$ is the number who die or emigrate between the two samples. We shall first find the distribution of the a_{ij} and then the conditional distribution of c_{ij} given a_{ij} and, in doing this, it is convenient to revert to the original parameters $\{\theta^0_{ij}\}$, $\{p^0_j\}$ which are now interpreted as common marginal probabilities.

We note that, whatever form the dependence takes,

$$E[c_{ij}] = E[E[c_{ij} \mid a_{ij}]] = E[a_{ij} p^0_j] = a_i \theta^0_{ij} p^0_j \quad \text{as in } \S 2$$

and therefore the $\tilde{\beta}_i$ are still consistent provided their variances still tend to zero as $a_i \to \infty$. It is these variances in which we shall be mainly interested, and since they principally depend on var(c_{ij}), cov(c_{ij}, c_{ik}) we must find how the latter are changed by dependence from $a_i\theta'_{ij}p'_j(1-\theta'_{ij}p'_j)$ and $-a_i\theta'_{ij}p'_j\theta'_{ik}p'_k$ respectively.

4.2. To construct $p[\{a_{ij}\}]$ it is necessary to specify how such probabilities as $P[I_\alpha \to j | I_\beta \to k, I_\gamma \to l]$ (using a self-explanatory notation) differ from θ'_{ij}. The assumption that there is no difference is really justified only if $I_1, I_2, \ldots, I_{a_i}$ are each released at a random point in the stratum and this would be difficult to achieve in practice. Two possibilities which are more likely are that they are either released close together in a randomly chosen sub-area or over a carefully spaced grid of points. In the former case we should expect that, for instance

$$P[I_\alpha \to j | I_\beta \to j, I_\gamma \to j] \geq P[I_\alpha \to j | I_\beta \to j] \geq \theta'_{ij},$$

and in the latter case a reversal of these inequalities. In other words, if the animals start close together, they are more likely than otherwise to move to the same stratum whereas, if they start with maximum possible distances between them, they are more likely than otherwise to move to different strata. The point of release may be thought of as determining the initial movement of an animal to the extent that the latter depends not only on such things as topography, wind or current, temperature, the local food distribution, but also, especially if the species is gregarious, on the location and movement of the nearest group of unmarked animals. (It should be noted that we are fortunately not required to try and describe the very complex interdependence of the movements of the unmarked animals nor the dependence of the marked on the unmarked, only the interdependence of the marked.)

Let us call the above two types of dependence "positive" and "negative" respectively. Then, as a first approximation, we may describe them in terms of the statistical concept of contagion by a simple generalization of Polya's urn model. Since death is a possible contingency it is reasonable to assume that it is contagious also and it can be thought of formally as the (t+1)th stratum, letting $a_{i\,t+1} = a_i - \sum_j a_{ij}$ and $\theta'_{i\,t+1} = 1 - \beta'_i$.

Consider an urn containing f_i balls of which f_{ih} are marked "h" and where $f_{ih}/f_i = \theta'_{ih}$, h = 1,2,....,t+1. Thus the proportions of balls in the urn are the marginal probabilities. If there is positive contagion, the conditional probability $P[I_r \to h_r \mid I_1 \to h_1, \dots, I_{r-1} \to h_{r-1}]$ is the proportion marked "h_r" after adding g_i marked "h_1", g_i marked "h_2",....,g_i marked "h_{r-1}". In this way we obtain -

$$P[I_r \to h_r, r = 1,2,\dots,a_i] = \prod_{h=1}^{t+1} \frac{(f_{ih}/g_i + a_{ih} - 1)!}{(f_{ih}/g_i - 1)!} \Big/ \frac{(f_i/g_i + a_i - 1)!}{(f_i/g_i - 1)!}$$

(If unequal numbers $g_{ih_1}, g_{ih_2}, \dots, g_{ih_{r-1}}$ of balls are added, the resulting density is very difficult to handle and, moreover, depends on the ordering of the individuals, a feature we do not desire.) Hence

$$p[\{a_{ih}\}] = \prod_{h=1}^{t+1} \binom{f_{ih}/g_i + a_{ih} - 1}{a_{ih}} \Big/ \binom{f_i/g_i + a_i - 1}{a_i}. \qquad (22)$$

It is readily deduced from (22) that var(a_{ih}) = $K_i a_i \theta'_{ih}(1-\theta'_{ih})$, cov($a_{ih}, a_{ik}$) = $-K_i a_i \theta'_{ih} \theta'_{ik}$, (23)
where $K_i = (f_i/g_i + a_i)/(f_i/g_i + 1)$. Thus $1 \le K \le a_i$, the extreme values corresponding respectively to no contagion and identical behaviour of the a_i individuals.

If there is negative contagion, balls are subtracted instead of added and provided $f_i > g_i a_i$, an obviously necessary condition, (23) still holds with $K_i = (f_i/g_i - a_i)/(f_i/g_i - 1)$. Thus $K \le 1$ and, in practice, would be substantially greater than zero.

4.3. If the catching in the jth stratum is not uniform but is concentrated in one or more sub-areas and if, having a common stratum of origin, the a_{ij} individuals are not uniformly distributed in the jth stratum, there will be a certain amount of positive dependence in their being caught or not caught. Otherwise, it is fair to assume independence. (We shall ignore the possible negative dependence of the catching of individuals from different strata of origin.) A density analogous to (22) can be constructed for $p[c_{ij} | a_{ij}]$ and

$$E[(c_{ij} - a_{ij}p_j')^2 | a_{ij}] = L_{ij}\, a_{ij}\, p_j'(1-p_j') \qquad (24)$$

where $1 \leq L_{ij} \leq a_{ij}$.

4.4. In the following we can generalise from specifically contagious dependence as described by urn models to any form of dependence which alters the variances and covariances in the simple manner given by (23) and (24), where $K_i \gtrless 1$ and $L_{ij} \geq 1$ (but probably not much greater than one).
Combining (23) and (24),

$$\begin{aligned} var(c_{ij}) &= L_{ij}\, a_i\, \theta_{ij}' p_j' (1-p_j') + K_i a_i \theta_{ij}' (1-\theta_{ij}') p_j'^2 \\ cov(c_{ij}, c_{ik}) &= - K_i a_i \theta_{ij}' p_j' \theta_{ik}' p_k' \end{aligned}$$

Suppose s = t. Then using these new variances and covariances in place of (14), we easily find that it is still true that

$$\underset{\sim}{\Sigma} \sim \underset{\sim}{D}_p \underset{\sim}{\Theta}^{-1} \underset{\sim}{D}_\mu \underset{\sim}{D}_a^{-1} \underset{\sim}{\Theta}'^{-1} \underset{\sim}{D}_p$$

except that, instead of (15),

$$\mu_i = \sum_j L_{ij}\theta_{ij}/p_j - \left(\sum_j L_{ij}\theta_{ij}\right)/\phi' + K_i\left(1/\phi - 1\right).$$

We can now note the interesting fact that the factor K_i has barely any effect on the efficiency of the $\tilde{p}_j$ and what little effect there is can be shown to be due to dependent mortality (or emigration) and not to dependent movement. Dependent catching, on the other hand, increases the largest terms in μ_i in the ratios $L_{ij} : 1$ and therefore also increases $\underset{\sim}{\Sigma}$ approximately in the ratio $\bar{L}:1$ where $\bar{L}$ is a suitably defined average of the L_{ij}.

When $s > t$, instead of (15'),

$$\mu_i = \sum_j L_{ij}\theta_{ij}/p_j - \sum_j L_{ij}\theta_{ij}/\bar{\phi}' + K_i\left(\phi_i/\bar{\phi}' - \phi_i^2\right)$$

and similar conclusions obtain.

We have, of course, only dealt with the leading terms in the expansion of $\underset{\sim}{\Sigma}$ but the others may still be neglected unless either form of dependence is very considerable.

5. Validity of the Peterson Estimate.

5.1. It was taken for granted in §§ 2,3,4 that the experimenter knows how the population is stratified and is able to conduct the experiment in the requisite manner. Suppose, however, that he does not know or that he is unable to mark the animals in s distinct ways and record t separate catches. He will then use the Peterson estimate $\tilde{n}_p = ab/c$ (or its unbiased version $(a+1)b/(c+1)$).

In § 5.2 we examine the legitimacy of doing this and then, in § 5.3, the appropriateness of the corresponding formula for $\mathrm{var}(\tilde{n}_p)$.

The Peterson estimate not only has simplicity to recommend it because even if a stratified experiment is performed, producing an estimate $\tilde{n}$, $\tilde{n}_p$ is preferable if it is valid as it is generally more accurate than $\tilde{n}$. In § 5.4, therefore, we consider tests for its validity. These tests are also necessary to complement the estimation theory of §§ 2,3 and reveal important facts about the movement and catching of the marked individuals. The extent to which $\mathrm{var}(\tilde{n}_p)$ is less than $\mathrm{var}(\tilde{n})$ will be discussed in § 5.5.

Except where otherwise stated, there will be no restriction on the relative values of s and t nor on the relative values of the ϕ'_i. We shall work with the parameters

$\theta_{ij} = \theta'_{ij}/\bar{\phi}'$, $\phi_i = \sum_j \theta_{ij}$, $p_j = \bar{\phi}' p'_j$, $n_j = n'_j/\bar{\phi}'$, $n = \sum_j n_j$,
where $\bar{\phi}' = (\sum_{i,j} \theta'_{ij})/s$.

5.2. Let $\beta = E[b] = \sum_j n_j p_j$ and $\gamma = E[c] = \sum_{i,j} a_i \theta_{ij} p_j$ and let

$$n_p = a\beta/\gamma.$$

For the limit process: $a_i \to \infty$ and $n_j \to \infty$ in such a way that a_i/a, n_j/n and a/n are all constant, n_p/n remains constant and $\tilde{n}_p/n$ converges in probability to n_p/n.

Therefore, $\tilde{n}_p$ estimates n if $n_p = n$, that is if

$$n \sum_{i,j} a_i \theta_{ij} p_j = a \sum_j n_j p_j \,. \tag{25}$$

(25) has an infinity of "accidental" solutions but we shall be concerned only with those having a simple physical interpretation. One of them is

$$\sum_i a_i \theta_{ij} = n_j / n \tag{26}$$

that is, the expected number of marked animals in the jth stratum is proportional to the number of unmarked. A special case of (26) is

$$\theta_{ij} = n_j / n \,.$$

Next, consider the condition that

$$p_j = p \quad \text{say} \tag{27}$$

This makes $n_p = n / \sum_i \frac{a_i}{a} \phi_i = n' / \sum_i \frac{a_i}{a} \phi_i'$ which is equal to $n = n' / \sum_i (\phi_i' / s)$ if the ϕ_i' or the a_i are equal and differs very little if they are not. Moreover, even this small difference is really irrelevant since, as we cannot estimate n' itself without the additional knowledge that all $\phi_i' = 1$, $n' / \sum_i \frac{a_i}{a} \phi_i'$ is just as good a substitute as $n' / \sum_i (\phi_i' / s)$. (In § 2.6 we could have chosen to estimate the former instead of the latter by writing $u_{ti} = a_i / a$ instead of $u_{ti} = 1/s$.)

A further condition for the validity of $\tilde{n}_p$ can be obtained by making a minor assumption about the relative values of $\underset{\sim}{n}$ and $\underset{\sim}{\Theta}$. It is that there exists $\underset{\sim}{\xi} = (\xi_i)$ such that

$$n_j = \sum_i \xi_i \theta_{ij} \,. \tag{28}$$

$\underset{\sim}{\xi}$ always exists if the movements of the marked and unmarked animals are the same since it may then be taken equal to $\underset{\sim}{m}$, the vector of first-sample strata sizes. Otherwise, if $s \geqslant t$, $\underset{\sim}{\xi}$ exists provided $\underset{\sim}{\Theta}$ is of full rank t. (When s = t, $\tilde{\underset{\sim}{\xi}}' = \eta' = \underset{\sim}{n}' \underset{\sim}{\Theta}^{-1}$ and when s > t there is an infinity of such $\underset{\sim}{\xi}$.

some of which at least are in the form $\underset{\sim}{\eta}' = \underset{\sim}{n}'(\underset{\sim}{U}\underset{\sim}{\Theta})^{-}\underset{\sim}{U}$ where $\underset{\sim}{U}$ is any matrix of the type described in § 2.6). If s < t, $\underset{\sim}{\xi}$ does not necessarily exist. Anyhow, assuming that there is a ξ satisfying (28), (25) becomes

$$n\sum_{i,j} a_i \theta_{ij} p_j = a \sum_{i,j} \xi_i \theta_{ij} p_j . \tag{29}$$

Consider

$$\sum_j \theta_{ij} p_j = \bar{p} \text{ say.} \tag{30}$$

This makes $n_p = \sum_i \xi_i = n'/\sum_i \left(\frac{\xi_i}{\sum \xi_i} \phi'_i \right)$ which equals n if the ϕ'_i or ξ_i are equal and, when it differs, is as good a substitute for n' as $n = n'/(\sum_i \phi'_i/s)$. Note that, while (27) specifies equal probabilities (of being caught in the second sample) at the time of the second sample, (30) specifies equal probabilities at the time of the first sample. If $\sum_j \theta_{ij} = 1$, (27) is a special case of (30). Another special case of (30) is

$$\theta_{ij} = \theta_j \text{ say.}$$

Lastly, when the movement patterns are the same (26) becomes $\sum_i a_i \theta_{ij} = (a/n) \sum_i m_i \theta_{ij}$ and a particular solution of this equation is

$$a_i/m_i = a/n ,$$

the first-sample counterpart of $p_j = p$.

5.3. Having listed the various conditions under which $\hat{n}_p$ is valid, namely (26), (27) and (30), let us suppose that the experimenter assumes as a matter of faith that one of them holds true and conducts an unstratified experiment. We now check that the formula for $var(\hat{n}_p)$ that he is obliged to use is an appropriate one.

It is an easy matter to show that

$$E[(\hat{n}_p - n_p)^2] \sim (\beta^2 a^2/\gamma^4) var(c) + (a^2/\gamma^2) var(b) \tag{31}$$

by retaining only leading terms as in §3.2, §3.3. Now
$\mathrm{var}(c) = \sum_i \mathrm{var}(c_{i.}) = \sum_i a_i(\sum_j \theta_{ij} p_j)(1 - \sum_j \theta_{ij} p_j)$

$$\leq (\sum_{i,j} a_i \theta_{ij} p_j)(1 - \sum_{i,j} a_i \theta_{ij} p_j / a) = a\,\gamma/a(1-\gamma/a) = \text{"var}(c)\text{" say.}$$

$\mathrm{var}(c)$ = "var(c)" only when $\sum_j \theta_{ij} p_j = \bar{p}$ say, that is when c is a true binomial variable. The experimenter has no other course but to use "var(c)" in (31) and we see that, in doing so, he overestimates var(c) if anything and therefore errs on the right side; but the difference is extremely small. Similarly, "var(b)" $= n\,\beta/n(1-\beta/n)$ overestimates $\mathrm{var}(b) = \sum_j n_j p_j(1-p_j)$ slightly unless $p = p$ say. Inserting "var(c)" and "var(b)" in (31),

$$\text{"}E[(\hat{n}_p - n_p)^2]\text{"} \sim (\beta^2 a^2/\gamma^3)(1-\gamma/a) + (a^2\beta/\gamma^2)(1-\beta/n) \qquad (32)$$

which is estimated by replacing β by b, γ by c and n by $\tilde{n}_p$.

If there is movement dependence among the marked animals, it has a small but negligible effect on the second of the four terms in (32) and, in view of the fact that the first term is by far the largest, we can certainly ignore this effect. Catch dependence does increase the largest term however, as in §4.4. For, if a_ω, c_ω and p_ω denote the number of marked animals present, the number caught and the marginal probability of being caught in a given subregion and, if $E[(c_\omega - a_\omega p_\omega)^2 \mid a_\omega] = L\, a_\omega p_\omega (1-p_\omega)$, the first term in (32) is multiplied by L.

5.4. Reverting to the stratified experiment, we turn from estimation to testing hypotheses. In most cases, the likelihood-ratio method is an obvious one to use and a great variety of hypotheses which might be of interest in particular experiments can be tested in this way. However, our attention will be confined to those tests which bear on the validity of $\tilde{n}_p$ and, moreover, which involve only simple functions of $\{a_i\}$, $\{b_j\}$, $\{c_{ij}\}$. Consider

H_1 : $\sum_j \theta_{ij} p_j = \bar{p}$, say; H_2: $p_j = p$ say; H_3: $\sum_i a_i \theta_{ij} = an_j/n$;

H_4: $\theta_{ij} = \theta_j$ say; H_5: $\theta_{ij} = n_j/n$.

Let H denote the general hypothesis which puts no restrictions on $\{\theta_{ij}\}$, $\{p_j\}$ except, of course, $\sum_{ij} \theta_{ij} = s$.

On H, the $\psi_{ij} = \theta_{ij} p_j$ are independent and, substituting $\tilde{\psi}_{ij} = c_{ij}/a_i$, the maximum value of the log-likelihood of $\{\theta_{ij}\}$, $\{p_j\}$ is

$$\tilde{L} = \sum_{ij} c_{ij} \log c_{ij} + \sum_i (a_i - c_{i.}) \log(a_i - c_{i.}) - \sum_i a_i \log a_i .$$

On H_1, the log-likelihood is

$$L_1 = \sum_{ij} c_{ij} \log \chi_{ij} + c \log \bar{p} + (a-c) \log(1-\bar{p})$$

where $\chi_{ij} = \theta_{ij} p_j / \bar{p}$. Maximising subject to $\sum_j \chi_{ij} = 1$,

$$\tilde{L}_1 = \sum_{ij} c_{ij} \log c_{ij} - \sum_i c_{i.} \log c_{i.} + c \log c + (a-c) \log(a-c) - a \log a .$$

The number of independent, identifiable parameters in L is $N(\psi) = st$ and the number in L_1 is $st-s+1$. To test H_1 against H we use the fact that, on H_1, $2(\tilde{L}-\tilde{L}_1) = \chi^2_{s-1}$ approximately. Now

$$2(\tilde{L}-\tilde{L}_1) = 2\sum_i c_{i.} \log(ac_{i.}/a_i c) + 2\sum_i (a_i - c_{i.}) \log(a(a_i - c_{i.})/(a-c)a_i)$$

and this expression is asymptotically equivalent to

$$\sum_i \frac{(c_{i.} - a_i c/a)^2}{a_i c/a} + \sum_i \frac{(a_i - c_{i.} - a_i(a-c)/a)^2}{a_i(a-c)/a} .$$

Therefore, the test is asymptotically equivalent to a goodness of fit test for proportionality in the last two columns of (33).

$$\begin{array}{ccc|c|c}
c_{11} & \cdots\cdots & c_{1t} & a_1 - c_{1.} & c_{1.} \\
\cdot & & \cdot & \cdot & \cdot \\
\cdot & & \cdot & \cdot & \cdot \\
\cdot & & \cdot & \cdot & \cdot \\
c_{s1} & \cdots\cdots & c_{st} & a_s - c_{s.} & c_{s.} \\
\hline
c_{.1} & \cdots\cdots & c_{.t} & & \\
\hline
b_1 & \cdots\cdots & b_t & &
\end{array} \tag{33}$$

This is hardly surprising when it is remembered that, on H_1, $E[c_{i.}] = a_i \bar{p}$.

There is no very satisfactory test of H_2 and, to see why, three possibilities have to be considered. First, suppose that nothing is assumed about the relative values of the ϕ_i. The $\psi_{ij} = \theta_{ij} p$ are then mathematically independent and L_2 is not identifiably distinct from L. Therefore H_2 cannot be tested. Secondly, if q constraints are imposed on the ϕ_i where $0 < q < s-1$, the minimisation of L_2 is awkward and, while a likelihood ratio test based on χ^2_q can be constructed, it does not take a simple form. Lastly, if it is assumed that $\phi_i = \sum_j \theta_{ij} = 1$, all i,

$$L_2 = \sum_{ij} c_{ij} \log \theta_{ij} + c\log p + (a-c)\log(1-p),$$

and this is not identifiably distinct from L_1. Therefore, in attempting to test the hypothesis: p_j independent of j, $\sum_j \theta_{ij} = 1$, one is really testing the more general hypothesis: $\sum_j \theta_{ij} p_j$ independent of i.

As H_3 involves the n_j which do not appear in $p\left[\{c_{ij}\}\right]$, it cannot be tested by the likelihood-ratio method. But $E\{c_{\cdot j}\} = (\sum_i a_i \theta_{ij})p_j$ and $E[b_j] = n_j p_j$ and a χ^2_{t-1} test of proportionality in the last two rows of (33) tests the hypothesis $\sum_i a_i \theta_{ij} = \nu n_j$, or $\sum_i a_i \theta_{ij} = (\sum_i a_i \phi_i) n_j/n$. This is not quite the same as H_3 and, if true, implies that $n_p = n/(\sum_i \frac{a_i}{a} \phi_i)$ rather than n but, as we pointed out in § 5.2, this sort of difference is immaterial.

H_4 is a particular case of H_1 and would not therefore be tested unless H_1 was accepted first. On H_4, the log-likelihood is $L_4 = \sum_j c_{\cdot j} \log \psi_j + (a-c)\log(1-\sum_j \psi_j)$, where $\psi_j = \theta_j p_j$. L_4 is maximised when $\psi_j = c_{\cdot j}/a$ giving $\tilde{L}_4 = \sum_j c_{\cdot j} \log c_{\cdot j} + (a-c)\log(a-c) - a\log a$. On H_4, $2(\tilde{L}-\tilde{L}_4) = \chi^2_{st-t}$ approximately and this is equivalent to a χ^2 test on the contingency table formed by the first s rows and t+1 columns of (33). The $\chi^2_{(s-1)(t-1)}$ test on the first s rows and t columns, that is on the c_{ij} alone, is easily seen to test H_4 against H_1 or, viewed differently, it tests the hypothesis: $\theta_{ij} = \phi_i \theta_j$ against H.

H_5 cannot be tested as L_5 is not identifiably distinct from L_4.

5.5. The theoretical discussion of the stratified experiment is concluded by a comparison of the variances of $\tilde{n}$ and $\tilde{n}_p$. Much simplicity is gained and nothing essential is lost by assuming in what follows that all $\phi_i = 1$. One advantage of this assumption is that it makes $n_p = n$ exactly whenever $\tilde{n}_p$ is valid. Writing $n_p = n$, the formulae to be compared are

$$E\left[(\tilde{n}_p-n)^2\right] \sim (\beta^2 a^2/\gamma^4)\ \mathrm{var}(c) + (a^2/\gamma^2)\ \mathrm{var}(b)$$

$$= P_1 + P_2 \quad \text{say, and}$$

$$E\left[(\tilde{n}-n)^2\right] \sim \sum_i \eta_i^2 \mu_i/a_i + \sum_j n_j(1/p_j - 1)$$

$$= S_1 + S_2 \quad \text{say, where}$$

$\underset{\sim}{\eta}' = \underset{\sim}{n}'\underset{\sim}{\Theta}^{-1}$ or $\underset{\sim}{n}'(\underset{\sim}{u}\underset{\sim}{\Theta})^{-1}\underset{\sim}{u}$ according as $s = t$ or $s > t$, and

$\mu_i = \sum_j \theta_{ij}/p_j - 1$.

It was observed in §5.3. that $P_2 = (a^2/\gamma^2)\sum_j n_j p_j(1-p_j) \leq (a^2/\gamma^2)\,\beta\,(1-\beta/n)$ with equality only if $p_j = p$. Therefore, since $a\beta/\gamma = n$, $P_2 \leq n^2/\beta - n$. In much the same way, $S_2 \geq n^2/\beta - n$ with equality only if $p_j = p$. Therefore $S_2 \geq P_2$.

We also observed in §5.3. that $P_1 \leq (\beta^2 a^2/\gamma^4)\,\gamma(1-\gamma/a) = n^2/\gamma - n^2/a$ with equality only if $\sum_j \theta_{ij} p_j = \bar{p}$. To compare S_1 with P_1, it is necessary to consider separately the conditions under which n is valid and these boil down to either H_1 or H_2.

On H_1, $\sum_j \theta_{ij} p_j = \bar{p}$ or

$$\sum_j \theta_{ij}(p_j - \bar{p}) = 0. \qquad (34)$$

Now, for $\tilde{n}$ to be consistent it is necessary that $\underset{\sim}{\Theta}$ is of full rank t and, this being so, (34) implies $p_j = \bar{p}$. Thus, it is sufficient to consider H_2. On H_2, $\mu_i = 1/p - 1 = a/\gamma - 1$. Therefore $S_1 = (a/\gamma - 1)\sum_i \eta_i^2/a_i = (a/\gamma - 1)\left[n^2/a + \sum_{i<\iota}(a_i\eta_\iota - a_\iota\eta_i)^2/a_i a_\iota a\right]$ ~~/a a a~~ $\leq (a/\gamma - 1)n^2/a = n^2/\gamma - n^2/a \geq P_1$.

On H_3, $\sum_i a_i \theta_{ij} = an_j/n$.

Also, $\sum_i \eta_i \theta_{ij} = n_j$. (35)

Therefore $S_1 = \sum_i (\eta_i/a_i)^2 a_i \mu_i = (n^2/a^2) \sum_i a_i \mu_i$ $+ 2(n/a) \sum_i (\eta_i/a_i - n/a)^2 a_i \mu_i + \sum_i (\eta_i/a_i - n/a)^2 a_i \mu_i$. The first of these terms is $(n^2/a^2) \left[\sum_i a_i \theta_{ij}/p_j - a\right]$ and, using H_3, equals $(n/a)\left[\sum_j n_j/p_j - n\right] = (n/a)S_2 \geqslant n^3/a\beta - n^2/a = n^2/\gamma - n^2/a$. Using H_3 and (35), the second term is zero. The third term is greater than or equal to zero. (It equals zero when $s = t$ since H_3 and (35) imply that $a_i = (a/n)\eta_i$. When $s > t$, however, $\eta_i/a_i \neq n/a$ in general.) Therefore, $S_1 \geqslant n^2/\gamma - n^2/a \geqslant P$.

Thus, in all cases when $\tilde{n}_p$ is valid,

$$E\left[(\tilde{n}-n)^2\right] \sim S_1 + S_2 \geqslant P_1 + P_2 \sim E\left[(\tilde{n}_p - n)^2\right]$$

with equality only if $p_j = p$ and $\eta_i/a_i = n/a$. The actual difference $S_1 + S_2 - P_1 - P_2$ is easily found from the above and, in practice, can be quite substantial.

6. Analysis of Sockeye Salmon Data.

6.1. In an experiment reported by Schaeffer (1951), both stratifications were with respect to time instead of place. The population comprised all adult sockeye salmon who passed a certain point of a river during a period of $s = 8$ weeks on their way up-stream to their spawning grounds. The fish were sampled and tagged according to the week in which they passed this point. Provided they succeed in reaching the spawning grounds, most adult salmon die after spawning. In this case, the deaths took place over a period of $t = 9$ weeks, and, during each of these weeks, a number of dead fish were recovered, presumably very soon after death. As Schaeffer's paper is not easily obtainable for reference, the data of his experiment are reproduced in table 1.

Table 1.

Table 1. $\{c_{ij}\}$ with $\{a_i\}$, $\{b_j\}$;

$s = 8,\ t = 9$

		Week of recovery (j)									Totals	a_i
		1	2	3	4	5	6	7	8	9		
Week of tagging (i)	1	1		2							3	15
	2	1	3	7							11	59
	3	1	11	33	24	5	1		1		76	410
	4		5	29	79	52	3	2	7	3	180	695
	5			11	67	77	2	16	7	3	183	775
	6				14	25	3	10	6	2	60	335
	7							1	5		6	59
	8							1			1	5
Totals		3	19	82	184	159	9	30	26	8	520	2,351
	b_j	16	113	718	2,664	3,317	635	1,217	904	368	9,952	

Table 2. $\{c_{ij}\}$ with $\{a_i\}$, $\{b_j\}$;

$s = 4,\ t = 4.$

	c_{i1}	c_{i2}	c_{i3}	c_{i4}	$c_{i.}$	$a_i - c_{i.}$	a_i	$c_{i.}/a_i$
c_{1j}	59	24	5	2	90	394	484	.186
c_{2j}	34	79	52	15	180	515	695	.259
c_{3j}	11	67	77	28	183	590	773	.237
c_{4j}	0	14	25	28	67	332	399	.168
$c_{.j}$	104	184	159	73				
b_j	847	2,664	3,317	3,124				

The frequencies in some of the outer weeks are too small to be used in what is essentially a large-sample theory and we have therefore reduced both s and t to four by grouping the first three and last three weeks of tagging into single strata and the first three and last four weeks of recovery into single strata. The new values of $\{a_i\}$, $\{b_j\}$, $\{c_{ij}\}$ are given in table 2.

Table 2.

$\{n'_j\}$ and $\{p'_j\}$ have the usual interpretations but θ'_{ij} now signifies the probability of dying in the jth stratum for a fish tagged in the ith and ϕ'_i the probability of dying on the spawning grounds during the nine weeks period. $1-\phi'_i$ therefore represents the probability of dying before reaching the spawning grounds or of surviving until after this period. (A small percentage of salmon do manage to reach the sea alive and return to spawn again. See Jones (1959).)

It is soon apparent that the Peterson estimate is not valid. For, in testing H, which specifies proportionality of the vectors $(c_{i.})$, $(a_i - c_{i.})$, we obtain $\chi^2 = 16.91$ and the .1% value is 16.27. The vectors $(c_{.j})$, (b_j) are so obviously not proportional that there is no need to apply a χ^2 test to see that H_1 is unacceptable.

6.2. Applying the theory of § 2.5, it is found, on evaluating $\underset{\sim}{\tilde{p}} = \underset{\sim}{C}^{-1}\,\underset{\sim}{a}$, that $\tilde{p}_1 = .1318$, $\tilde{p}_2 = 1.9461$, $\tilde{p}_3 = .1947$, $\tilde{p}_4 = .1063$. The unsatisfactory value of $\tilde{p}_2$ may be just a symptom of the general inaccuracy of capture-recapture estimation or it may indicate that the model is incorrect in assuming that the ϕ'_i are equal, this being necessary for the consistency of $\tilde{p}_1$. Both of these explanations are probably correct but while nothing can be done about the first, we can act on the second. The $c_{i\cdot}/n_i$ indicate where the possible differences in the ϕ'_i lie, the middle two being appreciably larger than the outer two. Let us therefore estimate subject only to the two constraints: $\phi'_1 = \phi'_4$, $\phi'_2 = \phi'_3$. This necessitates a reduction of t from four to three and, consequently, a grouping of two of the second-sample strata. It is permissible to group the jth and kth strata if (i) $(\theta_{ij}p_j + \theta_{ik}p_k)/(\theta_{ij} + \theta_{ik})$ is independent of i, in particular if (ii) $p_j = p_k$ or (iii) θ_{ij}/θ_{ik} is independent of i. (ii) cannot be tested but (iii) can as it implies proportionality of the jth and kth columns of (c_{ij}). In this case, the columns which are nearest to being proportional are the third and fourth and we therefore group them. (The hypothesis of proportionality is rejected at the .1% level but, even so, if p_3 is not too different from p_4, (i) will hold approximately true.)

6.3. Applying the theory of §2.6 to

$$\underset{\sim}{U} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ .25 & .25 & .25 & .25 \end{pmatrix}, \quad \underset{\sim}{C} = \begin{pmatrix} 59 & 24 & 7 \\ 34 & 79 & 67 \\ 11 & 14 & 105 \\ 0 & 14 & 53 \end{pmatrix}$$

$\underset{\sim}{a}' = (484, 695, 773, 399)$, $\underset{\sim}{b}' = (847, 2{,}664, 6{,}441)$

and evaluating $\tilde{\underset{\sim}{\rho}} = (\underset{\sim}{U}\underset{\sim}{D}_a^{-1}\underset{\sim}{C})^{-1}\underset{\sim}{v}$, we obtain

$$\tilde{\underset{\sim}{\rho}}' = (6.021,\ 1.607,\ 6.397) \text{ or } \tilde{\underset{\sim}{p}}' = (.1661,\ .6223,\ .1563).$$

$\hat{p}_1$ now lies in $[0,1]$ but is still curiously high. The other estimates are

$$\tilde{\underset{\sim}{\Theta}} = \underset{\sim}{D}_a^{-1}\underset{\sim}{C}\underset{\sim}{D}_{\tilde{\rho}} = \begin{pmatrix} .7339 & .0797 & .0925 \\ .2945 & .1827 & .6167 \\ .0857 & .1393 & .8689 \\ .0000 & .0564 & .8497 \end{pmatrix}$$

and, summing the rows,

$$\tilde{\phi}_1 = \tilde{\phi}_4 = .9061, \quad \tilde{\phi}_2 = \tilde{\phi}_3 = 1.0939.$$

Also $\quad \underset{\sim}{\tilde{n}}' = \tilde{\rho}' \underset{\sim}{D}_b = (5{,}099,\ 4{,}282,\ 41{,}204)$

and $\quad \tilde{\tilde{n}} = 50{,}585.$

The estimated variance-covariance matrix of $\tilde{\rho}$ is

$$\underset{\sim}{\tilde{\Sigma}} = \begin{pmatrix} 9.96 & -14.84 & 6.31 \\ -14.84 & 23.58 & -10.32 \\ 6.31 & -10.32 & 4.78 \end{pmatrix}.$$

Note the very high variance of $\tilde{\rho}_1$. It is unlikely that there was any catch dependence in this experiment so that we need make no mental reservations about $\underset{\sim}{\tilde{\Sigma}}$ underestimating $\underset{\sim}{\Sigma}$.

Next,

$$E\left[(\underset{\sim}{\tilde{n}}-\underset{\sim}{n})(\underset{\sim}{\tilde{n}}-\underset{\sim}{n})'\right] = 10^6 \begin{pmatrix} 7.168 & -33.474 & 34.441 \\ -33.474 & 167.347 & -177.113 \\ 34.441 & -177.113 & 198.694 \end{pmatrix}$$

and $$E\left[(\tilde{n}-n)^2\right] = 20.916 \times 10^6.$$

It is note-worthy that var($\tilde{n}_2$) and var($\tilde{n}_3$) are each considerably larger than var($\tilde{n}$).

Although $\tilde{n}_p$ is invalid, it is worth evaluating it and its variance to compare with $\tilde{n}$ and var($\tilde{n}$). Using the unbiased version $(a+1)b/(c+1)$,

$$\tilde{n} = 44{,}927,$$

and $$E\left[(\tilde{n}_p-n_p)^2\right] = 3.181 \times 10^6.$$

The latter is a good deal smaller than var($\tilde{n}$) as might be expected since the $\tilde{p}_j$ differ and, even more so, the η_i/a_i. For $\tilde{\eta}' = \underset{\wedge}{b}'(\underset{\wedge}{U}\underset{\wedge}{D}_a^{-1}\underset{\sim}{\Theta})^{-1}\underset{\sim}{U} = (9{,}896,\ -20{,}728,\ 46{,}020,\ 15{,}397).$

6.4. Finally, we consider how far we were justified in §6.2 in inferring that the ϕ differ. It can be shown that

$$\text{var}(\tilde{\phi}_1) \sim \mu_1/a_1 + \underset{\sim}{\theta}_1'(\underset{\sim}{U}\,\underset{\sim}{\Theta})^{-1}\underset{\sim}{U}\underset{\sim}{D}_\mu \underset{\sim}{D}_a^{-1}\underset{\sim}{U}'(\underset{\sim}{U}\,\underset{\sim}{\Theta})'^{-1}\underset{\sim}{\theta}_1$$
$$-2\mu_1/a_1\underset{\sim}{\theta}_1'(\underset{\sim}{U}\,\underset{\sim}{\Theta})^{-1}\underset{\sim}{U}\,\underset{\sim}{\delta}_1 \qquad (36)$$

where $\underset{\sim}{\theta}_1' = (\theta_{1j})$ and $\underset{\sim}{\delta}_1 = (\delta_{i1})$. Using this formula, we find that the estimated standard error of $\tilde{\phi}_1$ is .1509, and therefore the difference between $\tilde{\phi}_1 = .9061$ and one is not significant. However, this non-significance may probably be attributed mainly to the insensitivity of the estimation. The variances of $\tilde{\phi}_2, \tilde{\phi}_3, \tilde{\phi}_4$ can be found by formulae similar to (56) and a useful check on the computation is provided by verifying that they are all the same.

I wish to thank Professor M.S.Bartlett for some very helpful comments.

REFERENCES.

BEVERTON, R.J.H. & HOLT, S.J. (1957) On the dynamics of exploited fish populations. London: *Her Majesty's Stationery Office.*

CHANDA, K.C. (1954) Note on the consistency and maxima of the roots of likelihood equations. *Biometrika*, *41*, 56-61.

CHAPMAN, D.G. & JUNGE, C.O. (1956) The estimation of the size of a stratified animal population. *Ann.Math.Statist.*, *27*, 375-89.

DARROCH, J.N. (1958) The multiple-recapture census. I. Estimation of a closed population. *Biometrika*, *45*, 343-59.

\- (1959) The multiple-recapture census. II. Estimation when there is immigration or death. *Biometrika*, *46*, 336-351.

JONES, J.W. (1959) The salmon. London: Collins.

SCHAEFER, M.B. (1951) Estimation of the size of animal populations by marking experiments. *U.S.Fish and Wildlife Service Fisheries Bulletin*, *69*, 191-203.