Biometrika Trust

The Two-Sample Capture-Recapture Census when Tagging and Sampling are Stratified
Author(s): J. N. Darroch
Reviewed work(s):
Source: *Biometrika*, Vol. 48, No. 3/4 (Dec., 1961), pp. 241-260
Published by: Biometrika Trust
Stable URL: http://www.jstor.org/stable/2332748
Accessed: 14/05/2012 13:46

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Biometrika Trust is collaborating with JSTOR to digitize, preserve and extend access to *Biometrika*.

http://www.jstor.org

Biometrika (1961), **48**, 3 *and* 4, *p.* 241
Printed in Great Britain

The two-sample capture-recapture census when tagging and sampling are stratified

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

1. Introduction and summary

1·1. We start by recalling the capture-recapture argument used for the simplest type of experiment 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 Peterson estimate, although this name is usually given to

$$\tilde{n}_P + a = a(b+c)/c,$$

the estimate of total population size.

The italicized 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 summarized as follows:

(i) Animals can differ in their inherent catchability.

(ii) The catchability of an animal may change after being captured and marked.

(iii) The probability p can vary geographically 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. In order to cope with (i) and (ii) one would need more than two samples but, by contrast, it is possible to adapt the two-sample experiment to allow for (iii) and the adaptations 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 subregions 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 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.

This subject has been treated in three previous papers known to the author. Most of the problems raised here were first raised by Chapman & Junge (1956) and many of the answers obtained are the same as or similar to theirs, but we shall not attempt a detailed comparison as it would lengthen the present work too greatly. The main difference between their treatment and ours lies in the models used. They estimated the unknown parameters from sets of equations relating them to the expected values of the observed frequencies, and only when

it came to finding error variances did they assume equal probabilities and independence of behaviour for the different animals. We shall start by making the latter assumptions, estimate by the method of maximum likelihood, and then later relax the independence assumptions in an exact parametric fashion.

The earliest of the three papers is by Schaeffer (1951), arising from a tagging experiment which was stratified temporally instead of spatially. Chapman & Junge pointed out that Schaeffer's estimate was not a consistent one but they did not analyse his data. We do so, in § 6.

Beverton & Holt (1957) considered a similar problem but treated it deterministically and assumed that the marked animals are released in only one stratum.

1·3. *Summary.* The underlying assumptions are that all marked animals released in a given first-sample stratum have the same probability distribution of movement to the second-sample strata, and that all animals, marked and unmarked, in a given second-sample stratum have the same probability of capture. These assumptions are sufficient when $s \geqslant t$ but when $s < t$ it is necessary to assume that the unmarked animals move with the same probabilities as the marked animals. As one of the main aims of this paper is to keep the assumptions to a minimum, most of the attention is given to the case $s \geqslant t$ and we look at the case $s < t$ only in § 2·5.

The unknown parameters are the movement probabilities, the catching probabilities, the numbers of unmarked animals present in the different second-sample strata and, in particular, their total. It is usually assumed in two-sample censuses that every marked animal has probability zero of dying or emigrating in the interval between the two samples or, in other words, that the 'survival probability' is one. As this assumption is often not justifiable we have framed the theory in order that it may be avoided if desired. The price paid for dropping this assumption is that, instead of estimating the true (movement and catching) probabilities and true sizes, we have to be content with estimating them after they have been scaled either up or down by the survival probability. However, these scaled parameters do, for the most part, have a useful physical meaning. See §§ 2·1, 2.

Nearly all estimates are found by the method of maximum likelihood and are therefore asymptotically efficient. For $s = t$ the estimates of sizes are given by (8) and (9) and their variances and covariances by (18) and (19). For $s > t$, it is possible to allow the survival probabilities to depend to a certain extent on the stratum of origin. See § 2·4, (8′), (9′), (18′) and (19′).

In § 4 it is shown that if the (hitherto implicit) assumption of *independent* movement and catching of the marked animals is dropped, the estimates are still consistent and the formulae for their variances undergo only minor changes.

In § 5 we examine the validity of using the simple Peterson estimate since it would be unrealistic to assume that the experimenter always knows about the underlying stratification and is able to conduct his experiment to fit in with it. Our conclusions there support most of the previous literature on this question which has been mainly of a heuristic nature.

Finally, in § 6, most of the theory is illustrated on the data of Schaeffer's paper.

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 $\sum_i c_{i.} = \sum_j c_{.j} = c$. 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 survival probability. (The reason for using asterisks will appear in § 2·2.) 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 § 2·4 and § 6·3, they will be estimated on the supposition that they differ in a certain restricted fashion.

Turning to the untagged individuals, we shall assume nothing about their movement (except in § 2·5) and therefore they need not be brought into the picture until the time of the second sample. Accordingly let n^*_j denote the number in the jth stratum at that time and let $n^* = \sum_j n^*_j$. Of the n^*_j let b_j be caught in the second sample.

The fundamental assumption will be that all live 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 independent behaviour as regards movement and catching, we have the following two probability densities.

$$p[\{c_{ij}\}|\{a_i\}] = \frac{\prod_i a_i!}{\prod_i (a_i - c_{i.})! \prod_{ij} c_{ij}!} \prod_i (1 - \sum_j \theta^*_{ij} p^*_j)^{a_i - c_{i.}} . \prod_{ij} (\theta^*_{ij} p^*_j)^{c_{ij}}, \tag{1*}$$

$$p[\{b_j\}] = \prod_j \binom{n^*_j}{b_j} (p^*_j)^{b_j} (1 - p^*_j)^{n^*_j - b_j}. \tag{2*}$$

Thus the likelihood of $\{\theta^*_{ij}\}$, $\{p^*_j\}$, $\{n^*_j\}$ is obtained by multiplying (1*) and (2*), giving $e^{L^*} = e^{L^*_1} . e^{L^*_2}$ say. Differentiating L^* with respect to p^*_j gives

$$\frac{\partial L^*}{\partial p^*_j} = \frac{\partial L^*_1}{\partial p^*_j} + \frac{b_j}{p^*_j} - \frac{n^*_j - b_j}{1 - p^*_j},$$

and maximizing with respect to n^*_j by equating $\Delta_{n^*_j} L^*$ to zero gives

$$\frac{b_j}{p^*_j} = \frac{n^*_j - b_j}{1 - p^*_j}. \tag{3*}$$

Therefore, at the maximum-likelihood value of p^*_j, $\partial L^*/\partial p^*_j = \partial L^*_1/\partial p^*_j$. Also, since θ^*_{ij} is present only in L^*_1, $\partial L^*/\partial \theta^*_{ij} = \partial L^*_1/\partial \theta^*_{ij}$. In other words, maximizing L^* with respect to the unknown parameters is equivalent to maximizing L^*_1 and then (by (3*)) estimating n^*_j by the estimate of b_j/p^*_j. This allows us to make the simplification of speaking of (1*) as *the* likelihood and excluding (2*) from the explicit maximization process.

The next point to be made is that $\{\theta^*_{ij}\}$, $\{p^*_j\}$ must always be non-identifiable to the extent of a multiplicative constant, for their likelihood is the same as that of $\{k\theta^*_{ij}\}$, $\{k^{-1}p^*_i\}$. In particular it is not possible to estimate the survival probability ϕ^*. To tie down this non-identifiability, we shall work with the parameters $\theta_{ij} = \theta^*_{ij}/\phi^*$ and $p_j = \phi^* p^*_j$. These parameters *are* identifiable (provided $s \geqslant t$) and have fairly obvious probability interpretations. Thus, θ_{ij} is the probability of any one of the a_i individuals being in stratum j at the time of the second sample given that it is alive at that time; while p_j is the probability of any animal surviving and, if it is in the jth stratum, of being caught there. Correspondingly, it is not possible to estimate the n^*_j (unless it is known *a priori* that $\phi^* = 1$) and, instead, we estimate $n_j = n^*_j/\phi^*$, $n = \sum_j n_j$. n_j is the number of unmarked individuals in the jth stratum divided

by the probability of survival for the marked individuals. Therefore, assuming that there is no immigration during the experiment and that marked and unmarked individuals have the same probability of survival, n_j is roughly the number of unmarked individuals at the time of the first sample who finish up in the jth stratum and n is roughly the total unmarked population size at the time of the first sample.

To summarize the foregoing, our plan will be to maximize the likelihood of $\{\theta_{ij}\}$, $\{p_j\}$, namely

$$e^L \propto \prod_i (1-\sum_j \theta_{ij}p_j)^{a_i-c_i} . \prod_{ij} (\theta_{ij}p_j)^{c_{ij}}, \tag{1}$$

obtaining estimates $\{\tilde{\theta}_{ij}\}$, $\{\tilde{p}_j\}$ and then to estimate n_j by $\tilde{n}_j = b_j/\tilde{p}_j$ and n by $\tilde{n} = \sum_j \tilde{n}_j$.

The number of parameters in (1) 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 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 cases: $s = t$, $s > t$, $s < t$ will be treated separately.

2·3. 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

$$\mathbf{\Psi} = \mathbf{\Theta}\mathbf{D}_{\mathbf{p}}, \tag{2}$$

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

$$\mathbf{D}_{\mathbf{p}}^{-1}\mathbf{1} = \mathbf{\Psi}^{-1}\mathbf{1},$$

or

$$\boldsymbol{\rho} = \mathbf{\Psi}^{-1}\mathbf{1}, \tag{3}$$

where $\boldsymbol{\rho} = (\rho_j) = (p_j^{-1})$. (2) and (3) show that the transformation is one-one if and only if $\mathbf{\Theta}$ is non-singular.

Maximizing the likelihood with respect to $\{\psi_{ij}\}$ the estimates are

$$\tilde{\psi}_{ij} = c_{ij}/a_i. \tag{4}$$

$\tilde{\psi}_{ij}$ is obviously a consistent estimate of ψ_{ij} as $a_i \to \infty$. Translating into the θ, p system, we deduce that

$$\tilde{\theta}_{ij}\tilde{p}_j = c_{ij}/a_i$$

are the maximum-likelihood equations and give consistent estimates of $\{\theta_{ij}\}$, $\{p_j\}$ provided that $\tilde{\mathbf{\Theta}}$ and $\mathbf{\Theta}$ are non-singular. From (4) it follows that $\tilde{\mathbf{\Psi}}$ (and therefore also $\tilde{\mathbf{\Theta}}$) is non-singular if and only if $\mathbf{C} = (c_{ij})$ is. If $\mathbf{C}$ is non-singular then, from (3) and (4),

$$\tilde{\boldsymbol{\rho}} = \mathbf{C}^{-1}\mathbf{D}_{\mathbf{a}}\mathbf{1} = \mathbf{C}^{-1}\mathbf{a}, \tag{5}$$

and since $\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C}$ converges in probability to $\mathbf{\Theta}\mathbf{D}_{\mathbf{p}}$ as the $a_i \to \infty$, $\tilde{\boldsymbol{\rho}}$ converges to

$$\mathbf{D}_{\mathbf{p}}^{-1}\mathbf{\Theta}^{-1}\mathbf{1} = \mathbf{D}_{\mathbf{p}}^{-1}\mathbf{1} = \boldsymbol{\rho}$$

provided $\boldsymbol{\Theta}$ is non-singular (as already stated). On the other hand, if $\boldsymbol{\Theta}$ is singular, $\tilde{\boldsymbol{\rho}}$ diverges. It is easily seen that the elements of $\tilde{\boldsymbol{\rho}} = \mathbf{C}^{-1}\mathbf{a}$ are not necessarily positive and greater than one. In other words, it is quite possible to obtain a $\tilde{p}_j$ which does not lie in the interval [0, 1]. The first diagnosis of this anomaly that springs to mind is that the estimation procedure should have somehow incorporated the restriction that all p_j (and all θ_{ij}) lie in [0, 1]. Usually when estimating probabilities there is no need to do this explicitly as their estimates turn out to be proportions and, indeed, as far as the ψ_{ij} are concerned, this is what happens in the present case. It is only in transferring from $\tilde{\psi}_{ij} = c_{ij}/a_i$ to $\tilde{\theta}_{ij}$ and $\tilde{p}_j$ that the anomaly occurs. As this transfer is made by imposing the constraints $\sum_j \tilde{\theta}_{ij} = 1$ $(i = 1, 2, \ldots, s)$ it is these that are at the root of the trouble and there are, broadly speaking, two possible diagnoses. Namely, (i) the true ϕ_i^* are not equal or (ii) the true ϕ_i^* are equal but sampling errors are overwhelmingly large. (i) can be allowed for by reducing the number of second-sample strata and applying the theory of § 2·4. This is further discussed in the case of the numerical example in §§ 6·2, 3. (ii) means that the $\tilde{p}_j$ and the $\tilde{n}_j = b_j/\tilde{p}_j$ are virtually useless as estimates of p_j and n_j and this would no doubt be confirmed by their having large variances. However, this does not necessarily mean that the variance of $\tilde{n} = \sum_j \tilde{n}_j$ is also large as the covariances of the $\tilde{n}_j$ may well be large and negative. Again, see the discussion of the numerical example in § 6·3.

When $\mathbf{C}$ is singular the likelihood equations in the ψ system are invalid and we must look at those in the θ, p system. $\partial L/\partial p_j = 0$ gives

$$\sum_i \frac{(a_i - c_{i.})\,\theta_{ij} p_j}{1 - \sum_j \theta_{ij} p_j} = c_{.j}. \tag{6}$$

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

$$\frac{(a_i - c_{i.})\,\theta_{ij} p_j}{1 - \sum_j \theta_{ij} p_j} = c_{ij} + \lambda_i \theta_{ij}. \tag{7}$$

These equations have no easy solution except when $s = t = 2$. (They do, of course, have an easy solution when $\mathbf{C}$ is non-singular. For (6) and (7) imply that $\sum_i \lambda_i \tilde{\theta}_{ij} = 0$ all j, which, if $\tilde{\boldsymbol{\Theta}}$ is non-singular, implies that $\lambda_i = 0$ all i and hence that $\tilde{\theta}_{ij}\tilde{p}_j = c_{ij}/a_i$.)

It is obvious from the foregoing that the experimenter must try and avoid stratifications which make $\boldsymbol{\Theta}$ singular; and the further $\boldsymbol{\Theta}$ is from singularity the more reliable are the estimates $\{\tilde{\theta}_{ij}\}$, $\{\tilde{p}_j\}$, a fact which will be confirmed by the variance formulae in § 3. There are a few trivial instances of singularity which may be foreseeable. For example, if $\theta_{ij} = 0$ all i, the jth stratum is effectively non-existent. If $\theta_{ij} = \theta_{lj}$ for 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 $\theta_{iJ} = \theta_{ij} + \theta_{ik}$, $p_J = (\omega p_j + p_k)/(\omega + 1)$. But it will sometimes be less easy to avoid the more general instances of singularity.

If $\boldsymbol{\Theta}$ is non-singular and the sample sizes are large, a singular $\mathbf{C}$ will be extremely improbable. But, if it does occur, then, because of the difficulty of solving the likelihood equations and in view of the fact that the variance formulae of § 3 are derived conditional on $\mathbf{C}$ being non-singular, the best procedure will 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. (The

first-sample grouping may be omitted, making $s > t$ and involving us in the theory of § 2·4.) If both groupings are made when $s = t = 2$ the result is an unstratified analysis and it is interesting 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

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

(This solution, which always exists, is obviously not consistent.) 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 (5) to estimate $\mathbf{n} = (n_j)$. Since $\tilde{\mathbf{n}} = (b_j/\tilde{p}_j) = \mathbf{D}_{\mathbf{b}}\tilde{\boldsymbol{\rho}}$,

$$\tilde{\mathbf{n}} = \mathbf{D}_{\mathbf{b}}\mathbf{C}^{-1}\mathbf{a}, \tag{8}$$

first found by Chapman & Junge under similar assumptions. Also, since $\tilde{n} = \mathbf{1}'\tilde{\mathbf{n}}$,

$$\tilde{n} = \mathbf{b}'\mathbf{C}^{-1}\mathbf{a}. \tag{9}$$

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 the simple estimation equations $\tilde{\psi}_{ij} = c_{ij}/a_i$ are invalid and we have to solve (6) and (7). These equations have no explicit solution (except in the trivial case $t = 1$) and, for any given data, they would therefore have to be solved iteratively and the information matrix inverted to find error variances and covariances. If s and t are not small this is not an attractive proposition but, fortunately, it is easily avoided as follows.

We continue 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 parameters

$$\theta_{ij} = \theta_{ij}^*/\bar{\phi}^*, \quad \phi_i = \sum_j \theta_{ij}, \quad p_j = \bar{\phi}^* p_j^*.$$

These parameters correspond closely to the θ_{ij} and p_j of § 2·2 but do not, unfortunately, have simple probability interpretations. This gives a total of $st - 1 + t$ parameters, the subtraction of one being due to $\sum_{ij} \theta_{ij} = s$. Instead of imposing $s - 1$ constraints: $\phi_1 = \phi_2 = \ldots = \phi_s$, as we effectively did in § 2·3, 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 $\phi_i - \phi_k = 0$ or possibly $\phi_i + \phi_k - 2\phi_1 = 0$ but we can write them generally as

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

Also

$$\sum_i u_{ti}\phi_i = 1$$

if we define $u_{ti} = 1/s$. The matrix $\mathbf{U} = (u_{ki})\,(k = 1, 2, \ldots, t)$, $(i = 1, 2, \ldots, s)$ 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 $\mathbf{v}'$ denote the $1 \times t$ vector $(0, 0, \ldots, 0, 1)$. Then

$$\mathbf{U}\boldsymbol{\Theta}\mathbf{1} = \mathbf{v}$$

replacing $\boldsymbol{\Theta}\mathbf{1} = \mathbf{1}$ of § 2·3. Adapting the argument of § 2·3 we find that, provided $\mathbf{U}\boldsymbol{\Theta}$ and $\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C}$ are non-singular or, in other words, provided $\boldsymbol{\Theta}$ and $\mathbf{C}$ are of full rank t, consistent estimates of $\boldsymbol{\rho} = (p_j^{-1})$ and $\boldsymbol{\Theta}$ are given by

$$\tilde{\boldsymbol{\rho}} = (\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C})^{-1}\mathbf{v}, \tag{5'}$$

$$\tilde{\boldsymbol{\Theta}} = \mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C}\mathbf{D}_{\boldsymbol{\rho}}.$$

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, $\mathbf{n} = (n_j) = (n_j^*/\bar{\phi}^*)$ is estimated by

$$\tilde{\mathbf{n}} = \mathbf{D}_{\mathbf{b}}(\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C})^{-1}\mathbf{v} \tag{8'}$$

and $n = n^*/\bar{\phi}^*$ by

$$\tilde{n} = \mathbf{b}'(\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C})^{-1}\mathbf{v}. \tag{9'}$$

If $\bar{\phi}^*$ is also the 'average' probability of survival of the unmarked individuals, then n_j and n have the same interpretations as in § 2·3. It is of course not necessary to work with $\bar{\phi}^* = (\sum_i \bar{\phi}_i^*)/s$ and any linear combination $\sum_i \nu_i\phi_i^*$, where $\sum_i \nu_i = 1$, would serve instead, and would only involve changing u_{ti} from $1/s$ to ν_i.

2·5. When $s < t$, $N(\theta, p) > N(\psi)$ and $\{\theta_{ij}\}, \{p_j\}$ are not identifiable; only $N(\psi)$ functions of them are. If we try to estimate $\{p_j\}$ from the equations $\tilde{\psi}_{ij} = c_{ij}/a_i$ (or, equivalently, from (6) and (7)) 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 $\tilde{n}_j = b_j/\tilde{p}_j$.

At this point it is appropriate to consider the effect of assuming that marked and unmarked animals have the same movement pattern since, with this assumption, n can be estimated even though the n_j cannot. The unmarked individuals are now brought into the picture at the time of the first sample by letting m_i denote the number in the ith stratum. The density $p[\{b_j\}]$ becomes a convolution of s multinomial densities, the ith having parameters $m_i, \psi_{i1}, \psi_{i2}, \ldots, \psi_{it}$, and the likelihood equations are intractable. The best we can do, therefore, is to consider the moment equations

$$\sum_i m_i\psi_{ij} = b_j,$$

$$a_i\psi_{ij} = c_{ij}.$$

From these we get t equations

$$\sum_i m_i c_{ij}/a_i = b_j \tag{10}$$

for s unknown parameters $\{m_i\}$. When $s = t$ the solution of (10) is

$$\tilde{\mathbf{m}}' = \mathbf{b}'\mathbf{C}^{-1}\mathbf{D}_{\mathbf{a}},$$

giving

$$\sum_i \tilde{m}_i = \mathbf{b}'\mathbf{C}^{-1}\mathbf{a} \tag{11}$$

which is equivalent to (9). When $s < t$ there are more than enough equations and two courses are open: either equations (10) may be replaced by s linear combinations of them, or the number of parameters can be increased by relaxing the equality of some of the movement probabilities or by introducing some immigration parameters. We shall not go into details.

Finally we note that in deriving (10), we used ψ_{ij} merely as the probability of being caught in the jth stratum for an individual starting in the ith stratum and did not have to

express it as the product of a movement probability and a catching probability. Consequently, if it is expressed in this form, it can be taken as $\theta_{ij}p_{ij}$ with no restriction on the relative values of the $\phi_i(= \sum_j \theta_{ij})$. Thus (as pointed out by Chapman & Junge), in return for imposing the assumption that marked and unmarked animals have the same movement pattern, we can allow the survival probabilities and the probabilities of capture to depend on the stratum of origin. However, we consider that this involves losing more on the roundabouts than is gained on the swings and we shall therefore keep to our original assumptions for the remainder of the paper, and merely note that much of the ensuing theory for $s \geqslant t$ could be adapted under the different assumptions for $s \leqslant t$.

3. FIRST AND SECOND MOMENTS OF THE ESTIMATES

3·1. We first find the approximate bias of $\tilde{\boldsymbol{\rho}}$ 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 $\boldsymbol{\Theta}$ and $\mathbf{C}$ will be assumed to be of rank t, and we shall neglect the probability that $\mathbf{C}$ is of rank less than t.

Let
$$\boldsymbol{\Gamma} = E[\mathbf{C}] = \mathbf{D_a}\boldsymbol{\Theta}\mathbf{D_p}$$
and
$$\mathbf{C} - \boldsymbol{\Gamma} = \mathbf{Z} = (z_{ij}).$$

When $s = t$, $\tilde{\boldsymbol{\rho}} = \mathbf{C}^{-1}\mathbf{a}$ and
$$\mathbf{C}^{-1} = [\mathbf{I} - \boldsymbol{\Gamma}^{-1}\mathbf{Z} + (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^2 - \ldots + (-1)^{m-1}(\boldsymbol{\Gamma}^{-1}\mathbf{Z})^{m-1}]\boldsymbol{\Gamma}^{-1} + (-1)^m(\boldsymbol{\Gamma}^{-1}\mathbf{Z})^m\mathbf{C}^{-1}.$$
Therefore, since $\boldsymbol{\Gamma}^{-1}\mathbf{a} = \boldsymbol{\rho}$, we have
$$E[\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho}] = E[-(\boldsymbol{\Gamma}^{-1}\mathbf{Z}) + (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^2 - \ldots + (-1)^{m-1}(\boldsymbol{\Gamma}^{-1}\mathbf{Z})^{m-1}]\boldsymbol{\rho} + \mathbf{r},$$
where $\mathbf{r} = (r_j) = (-1)^m E[(\boldsymbol{\Gamma}^{-1}\mathbf{Z})^m\mathbf{C}^{-1}\mathbf{a}]$, the remainder vector. Let $\mathbf{F} = (f_{ij}) = (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^m$ and let C_{ij} denote the co-factor of c_{ij} in $\mathbf{C}$. Then, since $|C|$, the determinant of $\mathbf{C}$, is assumed numerically greater than or equal to one,
$$|r_j| = |E[\sum_{k,l} f_{jk}C_{lk}a_l/|C|]| \leqslant \sum_{k,l}|E[f_{jk}C_{lk}]|\,a_l \leqslant \sum_{k,l}(E[f_{jk}^2]E[C_{lk}^2])^{\frac{1}{2}}a_l.$$
Now $E[f_{jk}^2]$ is the sum of products of terms of $\boldsymbol{\Gamma}^{-1}$ of total power $2m$ and moments of $\mathbf{Z}$ 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
$$|r_j| \leqslant O(a^{-m+\frac{1}{2}m+s-1+1}) = O(a^{-\frac{1}{2}m+s}) \to 0$$
for $m > 2s$. It is therefore permissible to write $E[\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho}]$ as the infinite series
$$E[-(\boldsymbol{\Gamma}^{-1}\mathbf{Z}) + (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^2 - (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^3 + \ldots]\boldsymbol{\rho}.$$
The first term is zero since $E[\mathbf{Z}] = \mathbf{O}$, the $(2m)$th is $O(a^{-m})$ and the $(2m+1)$th is $O(a^{-m-1})$, $m \geqslant 1$.

Therefore the bias of $\tilde{\rho}_j = 1/\tilde{p}_j$ 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 $\tilde{\boldsymbol{\rho}} = (\mathbf{U}\mathbf{D}_\mathbf{a}^{-1}\mathbf{C})^{-1}\mathbf{v}$.

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 a/n remain fixed. Let $\boldsymbol{\beta} = E[\mathbf{b}]$. As $\mathbf{b}$ is independent of $\mathbf{C}$,
$$E[\hat{\mathbf{n}} - \mathbf{n}] = \mathbf{D}_{\boldsymbol{\beta}}E[\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}], \quad E[\tilde{n} - n] = \boldsymbol{\beta}'E[\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho}]$$
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 $\tilde{\boldsymbol{\rho}}$ is negligible, its variance-covariance matrix is

$$E[(\tilde{\boldsymbol{\rho}}-\boldsymbol{\rho})(\tilde{\boldsymbol{\rho}}-\boldsymbol{\rho})'] = \boldsymbol{\Sigma}$$

say, and, arguing as in § 3·1, this matrix can be expanded as a doubly infinite series whose terms are $O(a^{-m})$, $m \geqslant 1$. Suppose $s = t$. Then the leading term, of size $O(a^{-1})$, is

$$E[\boldsymbol{\Gamma}^{-1}\mathbf{Z}\boldsymbol{\rho}\boldsymbol{\rho}'\mathbf{Z}'\boldsymbol{\Gamma}'^{-1}]. \tag{12}$$

The (i, l) element of $E[\mathbf{Z}\boldsymbol{\rho}\boldsymbol{\rho}'\mathbf{Z}']$ is

$$E[\sum_{jk} z_{ij}\rho_j\rho_k z_{lk}]. \tag{13}$$

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

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

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

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

Substituting in (13), we can say that

$$\boldsymbol{\Sigma} \sim \boldsymbol{\Gamma}^{-1}\mathbf{D}_{\mathbf{a}}\mathbf{D}_{\boldsymbol{\mu}}\boldsymbol{\Gamma}'^{-1} = \mathbf{D}_{\boldsymbol{\rho}}\boldsymbol{\Theta}^{-1}\mathbf{D}_{\boldsymbol{\mu}}\mathbf{D}_{\mathbf{a}}^{-1}\boldsymbol{\Theta}'^{-1}\mathbf{D}_{\boldsymbol{\rho}}. \tag{16}$$

Notice that, when $\boldsymbol{\Theta}$ is near to singularity and the elements of $\boldsymbol{\Theta}^{-1}$ are large, $\tilde{\boldsymbol{\rho}}$ will be an inaccurate estimate of $\boldsymbol{\rho}$.

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

$$\boldsymbol{\Sigma} \sim \mathbf{D}_{\boldsymbol{\rho}}(\mathbf{U}\boldsymbol{\Theta})^{-1}\mathbf{U}\mathbf{D}_{\boldsymbol{\mu}}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{U}'(\mathbf{U}\boldsymbol{\Theta})'^{-1}\mathbf{D}_{\boldsymbol{\rho}}, \tag{16'}$$

where

$$\mu_i = \sum_j \theta_{ij}/p_j - \phi_i^2. \tag{15'}$$

3·3. Let $\mathbf{y}$ denote the deviation of $\mathbf{b}$ from its expected value $\boldsymbol{\beta} = \mathbf{D}_{\mathbf{p}}\mathbf{n}$. Then

$$\tilde{\mathbf{n}} - \mathbf{n} = \mathbf{D}_{\mathbf{b}}\tilde{\boldsymbol{\rho}} - \mathbf{D}_{\boldsymbol{\beta}}\boldsymbol{\rho} = \mathbf{D}_{\boldsymbol{\beta}}(\tilde{\boldsymbol{\rho}}-\boldsymbol{\rho}) + \mathbf{D}_{\mathbf{y}}\tilde{\boldsymbol{\rho}}.$$

Therefore

$$E[(\tilde{\mathbf{n}}-\mathbf{n})(\tilde{\mathbf{n}}-\mathbf{n})'] = \mathbf{D}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{\beta}} + E[\mathbf{D}_{\mathbf{y}}\tilde{\boldsymbol{\rho}}\tilde{\boldsymbol{\rho}}'\mathbf{D}_{\mathbf{y}}]. \tag{17}$$

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

$$E[(\tilde{\mathbf{n}}-\mathbf{n})(\tilde{\mathbf{n}}-\mathbf{n})'] \sim \mathbf{D}_{\mathbf{n}}\boldsymbol{\Theta}^{-1}\mathbf{D}_{\boldsymbol{\mu}}\mathbf{D}_{\mathbf{a}}^{-1}\boldsymbol{\Theta}'^{-1}\mathbf{D}_{\mathbf{n}} + \mathbf{D}_{\mathbf{n}}(\mathbf{D}_{\boldsymbol{\rho}} - \mathbf{I}\phi^{*-1}). \tag{18}$$

Since $(\tilde{n}-n)^2 = \mathbf{1}'(\tilde{\mathbf{n}}-\mathbf{n})(\tilde{\mathbf{n}}-\mathbf{n})'\mathbf{1}$,

$$\begin{aligned} E[(\tilde{n}-n)^2] &\sim \mathbf{n}'\boldsymbol{\Theta}^{-1}\mathbf{D}_{\boldsymbol{\mu}}\mathbf{D}_{\mathbf{a}}^{-1}\boldsymbol{\Theta}'^{-1}\mathbf{n} + \mathbf{n}'(\boldsymbol{\rho} - \mathbf{1}\phi^{*-1}) \\ &= \sum_i \eta_i^2\mu_i/a_i + \sum_j n_j(\rho_j - \phi^{*-1}), \end{aligned} \tag{19}$$

where

$$\boldsymbol{\eta}' = (\eta_i) = \mathbf{n}'\boldsymbol{\Theta}^{-1}. \tag{20}$$

If $\boldsymbol{\Theta}$ is near to singularity the low accuracy of $\tilde{\boldsymbol{\rho}}$ impairs that of $\tilde{\mathbf{n}} = (\tilde{n}_j)$ as is to be expected, but the accuracy of $\tilde{n} = \sum_j \tilde{n}_j$ is almost unimpaired. We have so far specifically avoided 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 m_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 $\sum_i m_i \theta_{ij}^*$, 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 $\sum_i m_i \theta_{ij}^* = n_j^*$ or $\sum_i m_i \theta_{ij} = n_j$ or

$$\mathbf{m}'\boldsymbol{\Theta} = \mathbf{n}'. \tag{21}$$

Thus, if the movement patterns are the same, $\boldsymbol{\eta} = \mathbf{m}$. 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 $\boldsymbol{\Theta}$ is near to singularity has no effect on (19). On the other hand, if $\mathbf{n}$ and $\boldsymbol{\Theta}$ 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 $\boldsymbol{\Theta}$ was very near to singularity.

When $s > t$,

$$E[(\tilde{n}-n)(\tilde{\mathbf{n}}-\mathbf{n})'] \sim \mathbf{D}_{\mathbf{n}}(\mathbf{U}\boldsymbol{\Theta})^{-1}\mathbf{U}\mathbf{D}_{\boldsymbol{\mu}}\mathbf{U}'(\mathbf{U}\boldsymbol{\Theta})'^{-1}\mathbf{D}_{\mathbf{n}} + \mathbf{D}_{\mathbf{n}}(\mathbf{D}_{\boldsymbol{\rho}} - \mathbf{I}\bar{\phi}^{*-1}) \tag{18'}$$

and

$$E[(\tilde{n}-n)^2] \sim \sum_i \eta_i^2 \mu_i / a_i + \sum_j n_j(\rho_j - \bar{\phi}^{*-1}), \tag{19'}$$

where

$$\boldsymbol{\eta}' = \mathbf{n}'(\mathbf{U}\boldsymbol{\Theta})^{-1}\mathbf{U}. \tag{20'}$$

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

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 $p[\{c_{ij}\}|\{a_i\}]$ 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 $\{\theta^*_{ij}\}$, $\{p^*_j\}$ which are now interpreted as common *marginal* probabilities, that is the probabilities for any single individual disregarding the histories of the other individuals.

We note that, whatever form the dependence takes,

$$E[c_{ij}] = E[E[c_{ij}|a_{ij}]] = E[a_{ij}p^*_j] = a_i\theta^*_{ij}p^*_j$$

as in § 2 and therefore the $\tilde{\rho}_j$ 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 released either close together in a randomly chosen subarea or over a carefully spaced grid of points. In the former case we should expect that, for instance

$$P[I_\alpha \to j | I_\beta \to j, I_\gamma \to j] \geqslant P[I_\alpha \to j | I_\beta \to j] \geqslant \theta^*_{ij},$$

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 $\theta^*_{i,t+1} = 1-\phi^*_i$.

Consider an urn containing f_i balls of which f_{ih} are marked 'h' and where $f_{ih}/f_i = \theta^*_{ih}$ $(h = 1, 2, \ldots, t+1)$. Thus the proportions of balls in the urn are the marginal probabilities. If there is positive contagion, the conditional probability $P[I_r \to h_r \,|I_1 \to h_1, \ldots, I_{r-1} \to h_{r-1}]$ 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

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

(If unequal numbers $g_{ih_1}, g_{ih_2}, \ldots, g_{ih_{r-1}}$ 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

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

It is readily deduced from (22) that

$$\operatorname{var}(a_{ih}) = K_i a_i \theta^*_{ih}(1-\theta^*_{ih}), \quad \operatorname{cov}(a_{ih}, a_{ik}) = -K_i a_i \theta^*_{ih}\theta^*_{ik}, \tag{23}$$

where $K_i = (f_i/g_i + a_i)/(f_i/g_i + 1)$. Thus $1 \leqslant K \leqslant a_i$, 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 $K_i \leqslant 1$ 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 subareas 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

$$E[(c_{ij} - a_{ij}p^*_j)^2|a_{ij}] = L_{ij}a_{ij}p^*_j(1-p^*_j), \tag{24}$$

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

4·4. In the following we can generalize 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 $L_{ij} \geqslant 1$ (but probably not much greater than one). Combining (23) and (24),

$$\operatorname{var}(c_{ij}) = L_{ij}a_i\theta^*_{ij}p^*_j(1-p^*_j) + K_i a_i \theta^*_{ij}(1-\theta^*_{ij})p^{*2}_j,$$
$$\operatorname{cov}(c_{ij}, c_{ik}) = -K_i a_i \theta^*_{ij}p^*_j\theta^*_{ik}p^*_k.$$

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

$$\mathbf{\Sigma} \sim \mathbf{D}_\rho \mathbf{\Theta}^{-1}\mathbf{D}_\mu \mathbf{D}_{\mathbf{a}}^{-1}\mathbf{\Theta}'^{-1}\mathbf{D}_\rho$$

except that, instead of (15),

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

We can now note the interesting fact that the factor K_i has barely any effect on the efficiency of the $\tilde{\rho}_j$ 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 $\mathbf{\Sigma}$ 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′),

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

and similar conclusions obtain.

We have, of course, only dealt with the leading terms in the expansion of $\mathbf{\Sigma}$ 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 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 generally has a smaller variance than $\tilde{n}$ (see § 5·5). In § 5·4, therefore, we consider tests for its validity. These tests are also necessary to complement the estimation theory of §§ 2, 3 and they reveal important facts about the movement and catching of the marked individuals.

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

$$\theta_{ij} = \theta_{ij}^*/\bar{\phi}^*, \quad \phi_i = \sum_j \theta_{ij}, \quad p_j = \bar{\phi}^* p_j^*, \quad n_j = n_j^*/\bar{\phi}^*, \quad n = \sum_j n_j,$$

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

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

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 $n_P = n/\sum_i (a_i/a)\,\phi_i = n^*/\sum_i (a_i/a)\,\phi_i^*$ 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.

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

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

$\boldsymbol{\xi}$ always exists if the movements of the marked and unmarked animals are the same, since it may then be taken equal to $\mathbf{m}$, the vector of first-sample strata sizes. Otherwise, if $s \geqslant t$, $\boldsymbol{\xi}$ exists provided $\boldsymbol{\Theta}$ is of full rank t. (When $s = t$, $\boldsymbol{\xi}' = \boldsymbol{\eta}' = \mathbf{n}'\boldsymbol{\Theta}^{-1}$ and when $s > t$ there is an infinity of such $\boldsymbol{\xi}$, some of which at least are in the form $\boldsymbol{\eta}' = \mathbf{n}'(\mathbf{U}\boldsymbol{\Theta})^{-1}\,\mathbf{U}$ where $\mathbf{U}$ is any

matrix of the type described in § 2·6.) If $s < t$, $\boldsymbol{\xi}$ does not necessarily exist. Anyhow, assuming that there is a $\boldsymbol{\xi}$ 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

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

This makes

$$n_P = \sum_i \xi_i = n^* / \sum_i (\xi_i / \sum_j \xi_j)\, \phi_i^*$$

which equals n if the ϕ_i^* or ξ_i are equal and, otherwise, will not usually differ very much from n. If $\sum_j \theta_{ij} = 1$, (27) is a special case of (30). Another special case of (30) is

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

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 $\tilde{n}_P$ 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 $(\tilde{n}_P)$ that he is obliged to use is an appropriate one.

It is an easy matter to show that

$$E[(\tilde{n}_P - n_P)^2] \sim (\beta^2 a^2/\gamma^4) \operatorname{var}(c) + (a^2/\gamma^2) \operatorname{var}(b) \tag{31}$$

by retaining only leading terms as in §§ 3·2, 3·3. Now

$$\begin{aligned} \operatorname{var}(c) &= \sum_i \operatorname{var}(c_{i.}) = \sum_i a_i (\sum_j \theta_{ij} p_j)(1 - \sum_j \theta_{ij} p_j) \\ &\leqslant (\sum_{i,j} a_i \theta_{ij} p_j)(1 - \sum_{i,j} a_i \theta_{ij} p_j / a) = a(\gamma/a)(1 - \gamma/a) = \text{`var}(c)\text{'} \quad \text{say.} \end{aligned}$$

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)' $= n(\beta/n)(1 - \beta/n\bar{\phi}^*)$ overestimates

$$\operatorname{var}(b) = \sum_j n_j p_j (1 - p_j/\bar{\phi}^*)$$

slightly unless $p_j = p$ say. Inserting 'var (c)' and 'var (b)' in (31),

$$\text{`}E[(\tilde{n}_P - n_P)^2]\text{'} \sim (\beta^2 a^2/\gamma^3)(1 - \gamma/a) + (a^2\beta/\gamma^2)(1 - \beta/n\bar{\phi}^*) \tag{32}$$

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 sub-region and, if $E[(c_\omega - a_\omega p_\omega)^2 | a_\omega] = L a_\omega p_\omega (1 - p_\omega)$, 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

$$H_1\colon \sum_j \theta_{ij} p_j = \bar{p}, \text{ say}; \quad H_2\colon p_j = p \text{ say}; \quad H_3\colon \sum_i a_i \theta_{ij} = a n_j/n;$$

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

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

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

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

On H_1, the log-likelihood is

$$L_1 = \sum_{i,j} c_{ij} \log \kappa_{ij} + c \log p + (a-c) \log (1-\bar{p}),$$

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

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

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

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

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 χ^2 goodness-of-fit test for proportionality in the last two columns of (33).

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

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 maximization of L_2 is awkward and, while a likelihood ratio test based on χ_q^2 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,

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

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

As H_3 involves the n_j it cannot be tested by the likelihood-ratio method without involving the density $p[\{b_j\}]$ as well as $p[\{c_{ij}\}|\{a_i\}]$. We shall not do this but, instead, simply note that it is equivalent to observe that $E[c_{.j}] = (\sum_i a_i \theta_{ij}) p_j$ and $E[b_j] = n_j p_j$ and that a χ_{t-1}^2 test of proportionality in the last two rows of (33) tests the hypothesis

$$\sum_i a_i \theta_{ij} = \nu n_j, \quad \text{or} \quad \sum_i a_i \theta_{ij} = (\sum_i a_i \phi_i) n_j / n.$$

This is not quite the same as H_3 and, if true, implies that $n_P = n/\{\sum_i (a_i/a) \phi_i\}$ 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 $L_4 = \sum_j c_{.j} \log \psi_j + (a-c) \log (1 - \sum_j \psi_j)$, where $\psi_j = \theta_j p_j$. L_4 is maximized when $\psi_j = c_{.j}/a$ giving $\tilde{L}_4 = \sum_j c_{.j} \log c_{.j} + (a-c) \log (a-c) - a \log a$. On H_4, $2(\tilde{L} - \tilde{L}_4) = \chi_{st-t}^2$ 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 $\chi_{(s-1)(t-1)}^2$ 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

$$E[(\tilde{n}_P - n)^2] \sim (\beta^2 a^2/\gamma^4) \operatorname{var}(c) + (a^2/\gamma^2) \operatorname{var}(b) = P_1 + P_2 \text{ say},$$

and

$$E[(\tilde{n} - n)^2] \sim \sum_i \eta_i^2 \mu_i / a_i + \sum_j n_j (1/p_j - \bar{\phi}^{*-1}) = S_1 + S_2 \text{ say},$$

where $\boldsymbol{\eta}' = \mathbf{n}'\boldsymbol{\Theta}$ or $\mathbf{n}'(\mathbf{U}\boldsymbol{\Theta})^{-1}\mathbf{U}$ according as $s = t$ or $s > t$, and $\mu_i = \sum_j \theta_{ij}/p_j - 1$.

It was observed in § 5·3 that $P_2 = (a^2/\gamma^2) \sum_j n_j p_j (1 - p_j/\bar{\phi}^*) \leqslant (a^2/\gamma^2) \beta (1 - \beta/n\bar{\phi}^*)$ with equality only if $p_j = p$. Therefore, since $a\beta/\gamma = n$, $P_2 \leqslant n^2/\beta - n/\bar{\phi}^*$. In much the same way, $S_2 \geqslant n^2/\beta - n/\bar{\phi}^*$ with equality only if $p_j = p$. Therefore $S_2 \geqslant P_2$.

We also observed in § 5·3 that $P_1 \leqslant (\beta^2 a^2/\gamma^4) \gamma (1 - \gamma/a) = n^2/\gamma - n^2/a$ 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 $\tilde{n}_P$ is valid and these boil down to either H_1 or H_3.

On H_1,
$$\sum_j \theta_{ij} p_j = \bar{p} \quad \text{or} \quad \sum_j \theta_{ij}(p_j - \bar{p}) = 0. \tag{34}$$

Now, for $\tilde{n}$ to be consistent it is necessary that $\boldsymbol{\Theta}$ 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

$$S_1 = (a/\gamma - 1) \sum_i \eta_i^2/a_i = (a/\gamma - 1)\,[n^2/a + \sum_{i<l} (a_i\eta_l - a_l\eta_i)^2/a_i a_l a]$$
$$\geqslant (a/\gamma - 1)\, n^2/a = n^2/\gamma - n^2/a \geqslant P_1.$$

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

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

Therefore

$$S_1 = \sum_i (\eta_i/a_i)^2 a_i \mu_i = (n^2/a^2) \sum_i a_i \mu_i + 2(n/a) \sum_i (\eta_i/a_i - n/a)\, a_i \mu_i + \sum_i (\eta_i/a_i - n/a)^2 a_i \mu_i.$$

The first of these terms is $(n^2/a^2)\,[\sum_i a_i \theta_{ij}/p_j - a]$ and, using H_3, equals

$$(n/a)\,[\sum_j n_j/p_j - n] = (n/a)\, S_2 \geqslant n^3/a\beta - n^2/a = n^2/\gamma - n^2/a.$$

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, $S_1 \geqslant n^2/\gamma - n^2/a \geqslant P_1$.

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

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

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. 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 3 and last 3 weeks of tagging into single strata and the first 3 and last 4 weeks of recovery into single strata. The new values of $\{a_i\}$, $\{b_j\}$, $\{c_{ij}\}$ are given in 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 9 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_1 which specified proportionality of the vectors $(c_{i.})$, $(a_i - c_{i.})$, we obtain $\chi_3^2 = 16{\cdot}91$ and the 0·1 % 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_3 is unacceptable.

Table 1. *Schaeffer's data.* $\{c_{ij}\}$ *with* $\{a_i\}$, $\{b_j\}$; $s = 8$, $t = 9$

Week of tagging (i)	Week of recovery (j) 1	2	3	4	5	6	7	8	9	Totals	a_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	773
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	2351
b_j	16	113	718	2664	3317	635	1217	904	368	9952	

Table 2. *Schaeffer's data.* $\{c_{ij}\}$ *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	0·186
c_{2j}	34	79	52	15	180	515	695	0·259
c_{3j}	11	67	77	28	183	590	773	0·237
c_{4j}	0	14	25	28	67	332	399	0·168
$c_{.j}$	104	184	159	73				
b_j	847	2664	3317	3124				

6·2. Applying the theory of § 2·5, it is found, on evaluating $\tilde{\boldsymbol{\rho}} = \mathbf{C}^{-1}\mathbf{a}$, that

$$\tilde{p}_1 = 0{\cdot}1318, \quad \tilde{p}_2 = 1{\cdot}9461, \quad \tilde{p}_3 = 0{\cdot}1947, \quad \tilde{p}_4 = 0{\cdot}1063.$$

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 $\tilde{p}_2$. Both of these explanations are probably correct but while nothing can be done about the first, we can act on the second. The $c_{i.}/a_i$ 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 0·1 % 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 § 2·4 to

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

$$\mathbf{a}' = (484, 695, 773, 399), \quad \mathbf{b}' = (847, 2664, 6441)$$

and evaluating $\tilde{\boldsymbol{\rho}} = (\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C})^{-1}\mathbf{v}$, we obtain

$$\tilde{\boldsymbol{\rho}}' = (6{\cdot}021,\ 1{\cdot}607,\ 6{\cdot}397) \quad \text{or} \quad \tilde{p}' = (0{\cdot}1661,\ 0{\cdot}6223,\ 0{\cdot}1563).$$

$\tilde{p}_2$ now lies in [0,1] but is still curiously high. The other estimates are

$$\tilde{\boldsymbol{\Theta}} = \mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C}\mathbf{D}_{\boldsymbol{\rho}} = \begin{pmatrix} 0{\cdot}7339 & 0{\cdot}0797 & 0{\cdot}0925 \\ 0{\cdot}2945 & 0{\cdot}1827 & 0{\cdot}6167 \\ 0{\cdot}0857 & 0{\cdot}1393 & 0{\cdot}8689 \\ 0{\cdot}0000 & 0{\cdot}0564 & 0{\cdot}8497 \end{pmatrix}$$

and, summing the rows,

$$\tilde{\phi}_1 = \tilde{\phi}_4 = 0{\cdot}9061, \quad \tilde{\phi}_2 = \tilde{\phi}_3 = 1{\cdot}0939.$$

Also $$\tilde{\mathbf{n}}' = \tilde{\boldsymbol{\rho}}'\mathbf{D}_{\mathbf{b}} = (5{,}099,\ 4{,}282,\ 41{,}204),$$

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

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

$$\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} 9{\cdot}96 & -14{\cdot}84 & 6{\cdot}31 \\ -14{\cdot}84 & 23{\cdot}58 & -10{\cdot}32 \\ 6{\cdot}31 & -10{\cdot}32 & 4{\cdot}78 \end{pmatrix}.$$

Note the very high variance of $\tilde{p}_2$. It is unlikely that there was any catch dependence in this experiment so that we need make no mental reservations about $\tilde{\boldsymbol{\Sigma}}$ underestimating $\boldsymbol{\Sigma}$. Next,

$$E[(\tilde{\mathbf{n}} - \mathbf{n})(\tilde{\mathbf{n}} - \mathbf{n})'] = 10^6 \begin{pmatrix} 7{\cdot}168 & -33{\cdot}474 & 34{\cdot}441 \\ -33{\cdot}474 & 167{\cdot}347 & -177{\cdot}113 \\ 34{\cdot}441 & -177{\cdot}113 & 198{\cdot}694 \end{pmatrix}$$

and $$E[(\tilde{n} - n)^2] = 20{\cdot}916 \times 10^6.$$

It is noteworthy 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 $(a+1)\,b/(c+1)$,

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

and $$E[(\tilde{n}_P - n_P)^2] = 3{\cdot}181 \times 10^6.$$

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 $\tilde{\eta}_i/a_i$. For

$$\tilde{\boldsymbol{\eta}}' = \mathbf{b}'(\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C})^{-1}\,\mathbf{U} = (9896,\ -20{,}728,\ 46{,}020,\ 15{,}397).$$

6·4. Finally, we consider how far we were justified in § 6·2 in inferring that the ϕ_i differ. It can be shown that

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

where $\boldsymbol{\theta}_1' = (\theta_{1j})$ and $\boldsymbol{\delta}_1 = (\delta_{i1})$. Using this formula, we find that the estimated standard error of $\hat{\phi}_1$ is 0·1509, and therefore the difference between $\hat{\phi}_1 = 0{\cdot}9061$ and 1 is not significant. However, this non-significance may probably be attributed mainly to the insensitivity of the estimation. The variances of $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ can be found by formulae similar to (36) and a useful check on the computation is provided by verifying that they are all the same.

I wish to thank Professors M. S. Bartlett and D. V. Lindley for their very helpful comments.

REFERENCE

BEVERTON, R. J. H. & HOLT, S. J. (1957). *On the Dynamics of Exploited Fish Populations.* London: Her Majesty's Stationery Office.

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.

DARROCH, J. N. (1959). The multiple-recapture census. II. Estimation when there is immigration or death. *Biometrika*, **46**, 336–51.

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

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