THE ESTIMATION OF THE SIZE OF A STRATIFIED ANIMAL POPULATION

By D. G. Chapman[1] and C. O. Junge, Jr.

University of Washington and Washington State Department of Fisheries

0. Summary. The estimation of the size of an animal population is considered for a situation where the population is stratified and only partial mixing takes place between strata. A consistent estimate is found and its variance determined. It is shown that estimates previously given or frequently used in this situation are not necessarily consistent and, in fact, may be meaningless. Conditions for their consistency are determined. Some further statistical problems in estimating the interstrata migration are discussed.

1. Introduction. The use of tagging procedures in estimating the size of animal problems is now well known; also the problems of sampling such populations so that the procedure conforms to the mathematical models used in the analysis of the resulting data have been stressed, particularly by DeLury [4].

We recall that the simplest procedure involves marking or tagging some members of the population and subsequently taking a sample of the population, the sampling being random with respect to the marked animals. This procedure and various extensions of it have been carried out on many populations, particularly on a small scale, and it has been customary to assume that where the sampling occurred without replacement, the random number of tag recoveries would follow the hypergeometric distribution or its various approximations (binomial, Poisson, normal). This will certainly be true if each member of the population is equally "catchable" and the capture of one member does not affect the chances of capture of others.

If the experimenter has a large and widespread population under study, it is no longer safe to make these assumptions. Thus, populations of fish in the ocean are subject to widely different fishing intensities in different areas. Consequently, if the sample is obtained from such a fishery it is hardly to be expected that it will be random, unless there is a complete mixing of the population throughout the differentially fished areas in the time that elapses between tagging and sampling.

However, it is known that some animal and many fish populations are made up of several groups ["tribes" is a word suggested by European fisheries biologists; see [1] for example] differentiated by their location. Some degree of mixing occurs between these adjacent groups or tribes, either continuously or at intervals. Hence, during the time interval between tagging and sampling, while

---

Received May 20, 1955, revised December 30, 1955.

[1] The research of this author has been supported in part by O.N.R. and part of it was carried out while he was at the Readership in the Design and Analysis of Scientific Experiment, Oxford University, on a Guggenheim Fellowship.

the tagged animals are mixing within their groups or tribes, there may also occur between-groups mixing. The nature and amount of this is almost certainly unknown, and, in fact, the experimenter will often wish to analyze the tag returns to cast some light on just this aspect of the population dynamics. Moreover, in general, it is not possible to state that an animal captured in strata j had previously belonged to strata i, though it is usually possible to identify the tags so that this information is available for recaptured tagged animals.

In fact, it is to be expected that in almost all extensive populations of marine animals this is the situation the biologist will face, i.e., a population with segregation by area but some mixing between areas and a population where the different subgroups are usually indistinguishable, at least on superficial examination.

Estimation of the population size in a situation of partial mixing of this type seems to have been first considered by Schaefer [8]; he dealt with a migrating salmon population. An example involving stratification by area in connection with the Pribilof fur seals has been noted in [6]. Of the many other possible examples we note only two: the important halibut population (see [9]) and a smaller flatfish population which has been the subject of a recent intensive study [7].

In this paper formulae are given for estimating the size of such a mixing population in those cases where tags are put out in all strata. Estimates can be given of the population migration between strata as well as of the total population size. Asymptotic variances of these estimates are obtained. We also determine under what assumptions the Petersen estimate made disregarding stratification and the estimate proposed by Schaefer in [8] are valid in the sense of being consistent. Examples are given to show further that these estimates may give meaningless results.

2. Notation and assumptions.

N_{ij} = number of individuals that are in stratum i at the time of tagging and in stratum j at time of sampling,

t_{ij} = number of tagged individuals in stratum i at tagging time and in stratum j at sampling time,

n_{ij} = number of sampled individuals in stratum i at tagging time and in stratum j at sampling time.

s_{ij} = number of tagged individuals, tagged in stratum i and subsequently recovered in a sample from stratum j $(i, j = 1, 2, \cdots r)$.

Sums over any subscript will be denoted by replacing the subscript by a dot. Thus,

$$t_{i.} = \sum_{j=1}^{r} t_{ij} = \text{number of tags put out in stratum } i,$$

$$n_{.j} = \sum_{i=1}^{r} n_{ij} = \text{number sampled in stratum } j.$$

The N_{ij}, $N_{.j}$, and $N_{..}$ (the total population size) are regarded as unknown parameters, the $t_{i.}$, $n_{.j}$, as known parameters, the t_{ij} and n_{ij} are unobservable random variables, while the s_{ij} are the observed random variables. It is assumed that all the parameters are positive.

Let S be the $r \times r$ matrix having s_{ij} in the ith row and jth column, $i = 1, \cdots, r, j = 1, \cdots, r$; let $|S|$ be the determinant of S, and let

$$\mathbf{n}' = (n_{.1}, n_{.2}, n_{.3}, \cdots, n_{.r})$$

$$\mathbf{t}' = (t_{1.}, t_{2.}, t_{.3}, \cdots, t_{r.}),$$

$$\left(\frac{\hat{\mathbf{N}}}{\mathbf{n}}\right)' = \left(\frac{\hat{N}_{.1}}{n_{.1}} \frac{\hat{N}_{.2}}{n_{.2}}, \cdots, \frac{\hat{N}_{.r}}{n_{.r}}\right).$$

Our model may be thought of as consisting of r urns, where urn i contains $N_{i.}$ marbles of which $t_{i.}$ have been marked, a different mark being used in each urn. After stirring the marked marbles in each urn, an unknown number of marbles (possibly zero) are placed in each of the other $(r - 1)$ urns. After this process, $N_{.j}$ is the number of marbles in the jth urn and N_{ij} is the number of these which were originally in the ith urn. After again stirring the marbles in each urn, a sample of $n_{.j}$ marbles is taken from the jth urn; of these, s_{ij} are observed to have been marked originally in the ith urn. From the known values $t_{i.}$, $n_{.j}$, and s_{ij}, it is desired to estimate the total number of marbles in the r urns, i.e.,

$$N_{..} = \sum_{i=1}^{r} N_{i.} = \sum_{j=1}^{r} N_{.j},$$

It appears to be useful to set out in detail the assumptions that need to be made for the determination of any estimate and, in particular, to distinguish those which relate to the experimenter's actions and those which relate to nature. The minimum possible assumption that could be made seems to be

$$\text{I.} \qquad E(s_{ij} \mid n_{ij}, t_{ij}) = \frac{n_{ij} t_{ij}}{N_{ij}} \qquad \text{for all } i, j.$$

This expected value would occur if a random sample is taken within the ijth substratum.

However, a model constructed on assumption I appears to be inadequate to yield an estimate of $N_{..}$, for it involves $3r^2$ unknowns (n_{ij}, t_{ij}, N_{ij}'s) and there are only r^2 observable random variables (s_{ij}) plus $2r$ side conditions ($\sum_i n_{ij} = n_{.j}$, $\sum_j t_{ij} = t_{i.}$) to determine these. The information is inadequate, except in case $r = 1$, so that it is necessary to make further assumptions to set up some structure relating the various substrata. In this respect it is sufficient that

$$\text{II.} \quad E(n_{ij}) = n_{.j}\left(\frac{N_{ij}}{N_{.j}}\right) \qquad \text{for all } i, j \text{ with the distribution of } t_{ij} \text{ arbitrary.}$$

Assumption II would be satisfied if the $n_{.j}$ marbles are taken from the $N_{.j}$ in the jth urn by a random sampling procedure.

It is seen that I and II together imply

$$\text{III. } E(s_{ij} \mid t_{ij}) = n_{.j} \frac{t_{ij}}{N_{.j}} \qquad \text{for all } i, j.$$

Assumptions I and III also imply II, but II and III do not imply I. For example, consider II and III holding together with $E(s_{ij} \mid n_{ij}, t_{ij}) = \delta_{ij} n_{ij} + \epsilon_{ij} n_{ij}^2$. Then it is trivial to determine δ_{ij}, ϵ_{ij} for each $E(n_{ij}^2)$ so that this assumption, together with II and III, is consistent. Consequently, it follows that III alone does not imply I and II and is therefore a weaker assumption. However, in an actual field situation it is likely that III will be satisfied only if I and II are.

It can be seen that assumption III can be satisfied even though no mixing of the marbles took place before redistribution; in other words, the validity of the procedure does not depend on any assumption on the behaviour of the animals after tagging within their respective strata or on the effect of tagging on the migration pattern, provided that a random sampling procedure can be used.

If stratification is disregarded, then the usual estimate of $N_{..}$ is

$$\hat{N}_0 = \frac{n_{..} t_{..}}{s_{..}}, \tag{1}$$

although for small $s_{..}$,

$$\hat{N}_1 = \frac{(n_{..} + 1)(t_{..} + 1)}{s_{..} + 1} \tag{2}$$

is preferable (see [3]).

The estimate proposed by Schaefer in the notation given here is

$$\hat{N}_2 = \sum_{i=1}^{r} \sum_{j=1}^{r} \left(\frac{n_{.j} t_{i.} s_{ij}}{s_{i.} s_{.j}} \right), \tag{3}$$

though in the derivation of this estimate Schaefer found it necessary to add some further assumptions.

An estimate based only on assumptions I and II, or III, is derived by observing that we can write

$$E\left(\sum_{j=1}^{r} s_{ij} \frac{N_{.j}}{n_{.j}} \right) = t_{i.}, \qquad i = 1, 2, \cdots, r. \tag{4}$$

The set of equations

$$\sum_{j=1}^{r} s_{ij} \frac{\hat{N}_{.j}}{n_{.j}} = t_{i.}, \qquad i = 1, 2, \cdots, r, \tag{5}$$

form a set of r equations in r unknowns which has a unique solution provided that $|S| \neq 0$. The estimate of $N_{..}$ is then simply the sum of the $\hat{N}_{.j}$. The solution of (5) and the estimate of $N_{..}$ are most simply expressed in matrix notation:

$$\left(\frac{\hat{\mathbf{N}}}{\mathbf{n}}\right)' = S^{-1}\mathbf{t}, \tag{6}$$

$$\hat{N}_3 = \mathbf{n}'S^{-1}\mathbf{t}. \tag{7}$$

3. Consistency of the estimates. The distribution of the s_{ij}, depending as it does on the random variables t_{ij}, is complex. We can make additional assumptions as to the behaviour of the t_{ij} and will do so later. If nothing is assumed about the t_{ij}, then it appears that it is possible to study only the consistency properties of the proposed estimates.

The property of consistency of estimates based on samples from a finite population has been variously defined. Following one such usage, an estimate $\hat{N}$ of $N_{..}$ would be called consistent if $\hat{N} = N_{..}$ whenever all $n_{.j} = N_{.j}$, i.e., whenever the sample taken without replacement exhausts the population. This usage makes the definition particular not only to the finiteness of the sample, but also to the method of sampling. Moreover, it is certainly satisfied in this problem if whenever $s_{ij} = t_{ij}$ for all i, j, $\hat{N} = N_{..}$, and it is easy to construct estimates that satisfy this condition and are otherwise meaningless. Finally, from a practical point of view, in the study of populations that number several hundred thousand or several millions, it is unreasonable to think of a sample equalling or nearly equalling the population size.

Yet at the same time it is possible that the samples may be very large—e.g., in the study noted in [7] it was of the order of 100,000 or roughly one-fifth of the population. With samples of this size, it is to be expected that if the samples are random the law of large numbers should be applicable.

Hence, we consider the case $n_{.j}, N_{.j} \to \infty$, $n_{.j}/N_{.j} \to \lambda_j$, and

$$\frac{s_{ij}}{t_{ij}} \to_p \frac{E(s_{ij} \mid t_{ij})}{t_{ij}} = \frac{n_{.j}}{N_{.j}} = \lambda_j,$$

and say that $\hat{N}$ is a consistent estimator of $N_{..}$ if under these conditions $\hat{N}/N_{..} \to_p 1$.

This assumption on the asymptotic behavior of the s_{ij} is certainly fulfilled if given t_{ij} the conditional distribution of the s_{ij} is multihypergeometric (or multinormal). In terms of the sampling procedure, it may be said that this situation is to be reasonably expected if the subarea is sufficiently small so that the sampling is uniform over it and each member of the population in it has an equal chance of capture. Or, again, it is a reasonable assumption if the subarea is so small that the $N_{.j}$ members of the population may be expected to mix freely and completely, regardless of the sampling uniformity.

The degree of within-area mixing is a pervasive problem in population estimation—this and the effect of tagging on the subsequent behavior of the tagged animal. Ultimately, these questions can only be answered by using independent methods of population estimation. Still, it will always be important to analyze the data from any experiment to determine whether there is internal evidence to support the assumptions necessary for the population study. Some suggestions in this direction were made in [3] for a different tag-sample procedure, which might be adapted to this situation. Some further considerations along these lines are elaborated on in Section 5.

Returning, then, to the question of consistency, under the conditions specified,

$$\frac{\hat{N}_0}{N_{..}} \text{ tends in probability to } \frac{n_{..}t_{..}}{\sum_j \frac{t_{.j}n_{.j}N_{..}}{N_{.j}}} = \frac{\lambda}{\sum_j a_j \lambda_j},$$

where $a_j = t_{.j}/t_{..}$, $\lambda = n_{..}/N_{..}$.

For arbitrary t_{ij}, and hence also $t_{.j}$, this ratio equals 1 if and only if $\lambda_j = \lambda$ or $n_{.j}/N_{.j}$ is the same for all strata, i.e., if the sampling is proportional to the population size in all strata.

Now consider $\hat{N}_2/N_{..}$. This ratio converges in probability to

$$\sum_i \sum_j \frac{t_{ij}t_{i.}n_{.j}}{t_{.j}\sum_\alpha t_{i\alpha}\left(\frac{n_{.\alpha}}{N_{.\alpha}}\right)N_{..}},$$

which equals 1 for arbitrary t_{ij}, provided $n_{.j}/N_{.j} = \lambda_j$ is constant, i.e., proportional sampling is required in all strata.

Turning to $\hat{N}_3$, it is seen that, substituting $E(s_{ij})$ for s_{ij} and $N_{.j}$ for $\hat{N}_{.j}$, equations (5) are satisfied. Hence, if $|S| \neq 0$, the uniqueness of the solution of this set of linear equations ensures that $\hat{N}_3$ is a consistent estimate of $N_{..}$.

It is a trivial exercise in arithmetic to construct examples to show that the estimate $\hat{N}_0$, may be absurd for all values of the random variables when $n_{.j}/N_{.j}$ varies with j. One such is given by the case where $r = 2$ $t_{i.} = t_{2.} = n_{.1} = n_{.2}$ and all animals migrate from stratum 1 to stratum 2 during the experiment except those tagged in stratum 1. Then it is seen that

$$\hat{N}_0 = \frac{n_{..}(2t_{i.})}{n_{.1} + s_{22}} = \frac{(2t_{i.})^2}{t_{i.} + s_{22}} \leqq 2t_{..}$$

for all observed s_{22}.

This is a pathological example, but perhaps it is not so unreal as might appear. Migration to stratum 2 may be normal behavior, but it is quite possible that tagging may produce abnormal behaviour such as failure to follow a migration instinct. In any case, it seems undesirable to use an estimate that may be misleading even for larger and larger samples, unless there is good reason for supposing that proportionate sampling has occurred.

In the example above, if the estimate $\hat{N}_2$ were used, it would be satisfactory; it may be checked that $\hat{N}_2$ is consistent if $t_{ij} = \delta_{ij}t_{i.}$, where δ_{ij} is the Kronecker delta. While general examples might be given to illustrate the pathological behaviour of $\hat{N}_2$ (though these, as well as a general study of the sampling properties of the estimate, are made more difficult by the presence of the random s_{ij} in both numerator and denominator), we content ourselves with a simple numerical example.

Let $r = 2\ t_{1.} = t_{2.} = n_{.1} = n_{.2} = 1000$; $N_{1.} = 100{,}000$, $N_{2.} = 2000$; suppose now that all tagged animals in area 1 migrate to area 2 but no others do. On the other hand, the area 2 animals distribute themselves in the two areas equally. Then $N_{11} = 99{,}000$, $N_{12} = 1000$, $N_{21} = 1000$, $N_{22} = 1000$. If s_{21} is zero, $\hat{N}_2$ is undefined; with probability >0.99 however, $s_{21} > 0$, in which case $\hat{N}_2$ is (again with probability >0.99) less than 10,000, whereas $N_{..} = 102{,}000$.

4. An alternative model. It has been noted that from some points of view the operations of tagging and sampling in these experiments are dual operations. It might then be thought reasonable to make assumption I plus assumption II′

$$E(t_{ij}) = t_{i.} \frac{N_{ij}}{N_{i.}} \qquad \text{for all } i, j$$

with the distribution of n_{ij} arbitrary.

I and II′ together imply a parallel to III, viz., III′:

$$E(s_{ij} \mid n_{ij}) = t_{i.} \frac{N_{ij}}{N_{i.}}.$$

Assumption III requires that, on the average, in the sample of size $n_{.j}$ from the population $N_{.j}$, the various tagged groups are proportionately represented. The dual assumption III′ makes the same requirement, but treats the tagged group as the sample and the subsequent recovery as the property of being marked. This appears to be a less reasonable practical assumption in that it requires predicting the future behaviour of the animals marked.

With respect to the urn model, this assumes that the $t_{i.}$-marked marbles are completely stirred before any marbles are transferred to the other urns, so that in choosing the marbles to be transferred there is no preference for marked or unmarked marbles. Although this does not require stirring before sampling the $N_{.j}$ marbles, it does still require that there be no preference for marked or unmarked marbles in this final sample.

It is of interest to note that whereas no effect due to tagging must usually be assumed, I and II put no restriction on possible differential migration between tagged and untagged fish. Assumptions I and II′, however, do require that the migration pattern into the different recovery strata be the same for tagged and untagged fish.

From III′ a set of equations to yield estimates of the $N_{i.}$ can be written down. They are

$$(8) \qquad \sum_{i=1}^{r} s_{ij} \frac{\hat{N}_{i.}}{t_{i.}} = n_{.j} \qquad (j = 1, 2, \cdots, r)$$

and

$$\hat{N}_3 = \sum_{i=1}^{r} \hat{N}_{i.} = \mathbf{t}'(S')^{-1}\mathbf{n} = \mathbf{n}'S^{-1}\mathbf{t}, \tag{9}$$

the same estimate as derived in the first model. If we further assume that as $t_{i.}$, $N_i \to \infty$, $t_{i.}/N_i \to \mu_i$ (s_{ij}/n_{ij}) tends in probability to $E(s_{ij} \mid n_{ij})/n_{ij} = t_{i.}/N_{i.} = \mu_i$, then $\hat{N}_3$ is a consistent estimator of $N_{..}$.

However $\hat{N}_0$, $\hat{N}_2$ will not necessarily be consistent estimators of $N_{..}$ unless $t_{i.}/N_i = \mu$ for all $i = 1, 2, \cdots, r$. There may be special values of the random n_{ij} (or in the earlier model, of the random t_{ij}) for which $\hat{N}_0$, $\hat{N}_2$ are consistent. Thus, in this model N_2 is consistent if $n_{ij} = \delta_{ij} n_{.j}$, which would, in general, be true only if there were no intermixing (i.e., $N_{ij} = 0$ if $i \neq j$). In this case $N_{..}$ is the sum of r separate subpopulations N_{ii} $(i = 1, 2, \cdots, r)$ and the estimate $\hat{N}_2$ reduces to the sum of r estimates of the form of $\hat{N}_0$. $\hat{N}_3$ also reduces to the same form in this special situation.

It may be thought that $\hat{N}_0$ would be a consistent estimator, if both $n_{.j}$ and $t_{i.} \to \infty$ as $N_{i.}$, $N_{.j} \to \infty$ with $t_{i.}/N_{i.} \to \mu_i$, $n_{.j}/N_{.j} \to \lambda_j$ and if assumptions I, II, and II′ hold. N_0 will then be consistent if all μ_i are equal, or all λ_j are equal. Consider, however, the case where neither of these is true. Assumptions I, II, and II′ imply

$$E(s_{ij}) = t_{i.} n_{.j} \frac{N_{ij}}{N_{i.} N_{.j}}, \tag{10}$$

and consistency of $\hat{N}_0$ under these circumstances will hold in general only if $N_{ij}/N_{i.}N_{.j} =$ constant for all i, j. This condition, which means that there is random or "independent" mixing between the various strata, will be considered in Section 8. It has been noted that equations (5) or (8) may be solved only if S is nonsingular. In many cases the nature of the situation dictates that S be nonsingular. If stratification is with respect to time of migration and the time periods are defined so that an animal marked in any period i cannot be recovered in a period where $j < i$, then $s_{ij} = 0$ for all $j < i$. Hence, S is nonsingular provided that no $s_{ii} = 0$; it will thus certainly converge in probability to a nonsingular matrix.

In cases of a real migration or mixing, usually N_{ii} will be much larger than N_{ij} $(j \neq i)$, so that again in large samples $|S|$ will be not zero. The fur seal study [6] already referred to is an example.

In general, S' will converge in probability to

$$\begin{bmatrix} \frac{n_{.1}}{N_{.1}} t_{11}, & \frac{n_{.1}}{N_{.1}} t_{21}, & \cdots, & \frac{n_{.1}}{N_{.1}} t_{r1} \\ \frac{n_{.2}}{N_{.2}} t_{12}, & \frac{n_{.2}}{N_{.2}} t_{22}, & \cdots, & \frac{n_{.2}}{N_{.2}} t_{r2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{n_{.r}}{N_{.r}} t_{1r}, & \frac{n_{.r}}{N_{.r}} t_{2r}, & \cdots, & \frac{n_{.r}}{r_{.r}} t_{rr} \end{bmatrix},$$

which is nonsingular provided $T = (t_{ij})$ is. The model and limiting conditions are those associated with assumptions I and II. If, in addition, II′ is made and $t_{i.}/N_{i.} \to \mu_i$ as $t_{i.}, N_{i.} \to \infty$, then S will converge in probability to a matrix which is nonsingular if the matrix (N_{ij}) is nonsingular.

Whether it is possible to construct a test of the hypothesis that (N_{ij}) is nonsingular remains an open question. One case of singularity which has a simple biological interpretation is that associated with the condition

$$N_{ij} \propto N_{i.}N_{.j}, \tag{11}$$

i.e., the random mixing just referred to.

5. Variance of $\hat{N}_3$. To derive the formula for the asymptotic variance of $\hat{N}_3$ it is necessary to make additional assumptions on the distributions of the random variables involved. To this end we make the following assumption IV: The distribution of the t_{ij} and the conditional distribution of the s_{ij} given t_{ij} are multinomial with expectation given by II′ and III.

It might be more reasonable from a practical point of view to assume that these distributions are multihypergeometric. In this case we here neglect the finite sampling corrections.

It is now elementary to derive the variance of each s_{ij} by working with conditional expectations. In fact,

$$\sigma^2(s_{ij}) = \frac{t_{i.}n_{.j}N_{ij}}{N_{i.}N_{.j}}\left[1 + \frac{1}{N_{.j}}(n_{.j} - 1)\left(1 - \frac{N_{ij}}{N_{i.}}\right) - t_{i.}\frac{N_{ij}}{N_{i.}N_{.j}}\right]. \tag{12}$$

Also,

$$\sigma(s_{aj}s_{bj}) = -\frac{n_{.j}t_{a.}t_{b.}N_{aj}N_{bj}}{N_{a.}N_{b.}N_{.j}^2}, \qquad a \neq b, \tag{13}$$

and

$$\sigma(s_{ai}s_{bj}) = 0, \qquad i \neq j. \tag{14}$$

From theorems on matrix differentiation,

$$-\frac{\partial N_3}{\partial s_{ab}} = \mathbf{n}'S^{-1}I_{ab}S^{-1}\mathbf{t}, \tag{15}$$

where I_{ab} is the matrix with 1 in the abth place and 0 everywhere else. This reduces to

$$-\frac{\partial \hat{N}_3}{\partial s_{ab}} = \left(\sum_j n_{.j}\frac{S_{aj}}{|S|}\right)\left(\sum_i \frac{S_{ib}}{|S|}t_{i.}\right), \tag{16}$$

$|S|$ being the determinant of S, S_{ij} the signed cofactor of s_{ij}.

Under the assumptions,

$$E(s_{ij}) = \frac{t_{i.}n_{.j}N_{ij}}{N_{i.}N_{.j}} \tag{17}$$

and substituting $E(s_{ab})$ for s_{ab}, it is readily seen that

$$(18) \qquad -\left(\frac{\partial \hat{N}_3}{\partial s_{ab}}\right) = \frac{N_{a.}}{t_{a.}} \cdot \frac{N_{.b}}{n_{.b}}.$$

Now, as usual, terms of second and higher order in the Taylor expansion of N_3 may be neglected in the limit, i.e., as $t_{i.}, n_{.j}, N_{i.}, N_{.j} \rightarrow \infty$. Then the asymptotic variance of $\hat{N}_3$ is easily calculated to be

$$(19) \qquad \sum_i \sum_j \frac{N_{ij} N_{i.} N_{.j}}{t_{i.} n_{.j}} \Bigl[1 + \frac{1}{N_{.j}} (n_{.j} - 1) \Bigl(1 - \frac{N_{ij}}{N_{i.}} - t_{i.} \frac{N_{ij}}{N_{i.} N_{.j}}\Bigr] - \sum_j \sum_a \sum_b \frac{N_{aj} N b_j}{n_{.j}}.$$

We write (19) in this form, in which it will be used, although from a mathematical point of view under the conditions stated, (19) must be infinite. The requirements of mathematical rigor can be met by considering the asymptotic variance of $\hat{N}_3/N_{..}$. The second term in this variance will be much smaller than the first. In fact, for most applications the approximation

$$(20) \qquad \sigma^2(\hat{N}_3) \doteq \sum_{i=1}^{r} \sum_{j=1}^{r} \frac{N_{ij} N_{i.} N_{.j}}{t_{i.} n_{.j}}$$

will be sufficient.

Since the estimates $\hat{N}_0$ and $\hat{N}_2$ may not be consistent if proportionate sampling (or tagging) does not occur, these are not in competition with $\hat{N}_3$ in general. If proportionate sampling does occur, then the several samples may be regarded as a single random sample from the whole population. In this case $\hat{N}_0$ is the maximum likelihood estimate and hence optimum from an asymptotic point of view.

Assumption IV is much stronger than any of the earlier assumptions and opens up the possibility of obtaining maximum likelihood or minimum χ^2 estimates. However, the modified minimum χ^2 estimates obtained by the use of Lagrange multipliers would require the solution of $r^2 + 2r$ linear equations. Even for r as small as 3 this is hardly feasible for general usage in the absence of special computing facilities. Whether these procedures could be simplified or other estimates with optimum properties found, is unsolved.

With Assumption IV it is seen that for the large sample case with $n_{.j}/N_{.j} \rightarrow 0$, $t_{i.}/N_{i.} \rightarrow 0$, but $n_{.j} t_{i.}/N_{.j}$, $n_{.j} t_{i.}/N_{i.}$ remaining finite, the estimate $\hat{N}_3$ is a maximum likelihood estimate. Even in this situation it is not to be expected that an unbiased estimate of $N_{..}$ exists, since such an estimate would have to be a nonlinear function of the s_{ij}. That this is in fact so may be proven similarly to the situation in the case of a single sample from a single strata as was done in [2], by an appeal to the theorem of Barankin [1] on necessary and sufficient conditions for the existence of unbiased estimates with finite variance.

When Assumption IV is made, further questions are opened up, viz., what practical sampling procedures may be expected to yield sample observations which satisfy this requirement. It was indicated that uniformity of the sampling over the subareas or uniformity of behavior within the subareas could be sufficient to ensure the conditions that $\hat{N}_3$ be consistent. Assumption IV, however, requires even more, namely, that the migration pattern of the animals be not affected by tagging and that the migrations taken by different members be independent of one another. Consequently, it is difficult to visualize a realistic situation where there is a priori information that IV is in fact reasonable. The experimenter will wish therefore to make what tests are possible on the data to determine if it is consistent with this assumption. One such method of analysis is to divide each of the r groups of tags of size $t_{i.}$ randomly into q_i subgroups t_{ik} $(k = 1, 2, \cdots, q_i)$. Then if Assumption IV holds,

$$E(s_{ijk}) = t_{i.k} n_{.j} \frac{N_{ij}}{N_{i.} N_{.j}} = \lambda_{ij} t_{i.k} n_{.j}$$

and for each i, j a χ^2 test for homogeneity may be made. Acceptance of this hypothesis does not confirm the validity of Assumption IV, but it does lend support to the analysis based on it.

On the contrary, if the hypothesis of homogeneity is rejected by the χ^2 test for one or several subgroups, the experimenter will do well to assume that formula (19) does not give the correct variance of $\hat{N}_3$. To derive another formula would require making other assumptions as to the distributions involved or as to the lower moments of the several random variables. Since there seems to be no reasonable alternative information in this direction, we suggest instead a method to *estimate* the variance of $\hat{N}_3$. For suppose all q_i defined above are equal, say, to q (i.e., each group of $t_{i.}$ tags is divided randomly into q subgroups $t_{i.1}, t_{i.2}, \cdots, t_{i.q}$). Then utilizing formula (7), we can construct q estimates $\hat{N}^{(1)}, \hat{N}^{(2)} \cdots, \hat{N}^{(q)}$ of $N_{..}$ and hence determine an estimate of the variance of $\bar{N} = 1/q \sum_{i=1}^{q} \hat{N}^{(i)}$ with $q - 1$ d.f.

The choice of the number of subgroups, either for this purpose or for the test of homogeneity outlined above, will be strictly limited in practice; for if q is chosen too large, the subgroup $t_{i.k}$ will be quite small and $E(s_{ij}) \leqq t_{i.k}(n_{.j}/N_{.j})$ also. There will be many zeros in the matrix S and the estimates $\hat{N}^{(k)}$ will be extremely variable. With some knowledge of the extraneous variance and of the degree of migration, it would be possible to set down working rules for the choice of q. In many large-scale field experiments, the biologist will do well to proceed in this manner until there is adequate assurance that more can be assumed about the distribution of the t_{ij}.

6. Additional strata; mortality. It is traditional that even though mortality may occur in a population between the time of tagging and of sampling, then if the chance of survival is the same among the tagged and untagged animals, the Petersen estimate (i.e., $\hat{N}_0$), based on a single random sample from a population, is essentially unaffected. In fact, in one case the distribution of the random number of tag recoveries remains unchanged. For let

N, N' = population size at tagging and at sampling, respectively,
t, t' = number of tags at similar times,
s = number of tag recoveries,
n = sample size.

If n is large relative to t and hence t', then the usual hypergeometric distribution for s is adequately replaced by a binomial distribution, i.e., $s \mid t'$ is $B(t', n/N')$. Further, if t' is $B(t, N'/N)$, then s is $B(t, n/N' \cdot N'/N)$, i.e., $B(t, n/N)$, and hence s has the same distribution as if there were no mortality. Clearly, it will be desirable to establish whether $\hat{N}_3$ enjoys a similar "robustness."

Consider first a somewhat more general situation where there are $k + 1$ strata, in the last of which no tags are placed and no sampling is done. We can still write $E[s_{ij}(N_{.j}/n_{.j})] = t_{ij}$, but it no longer follows that $\sum_{j=1}^{k} E[s_{ij}(N_{.j}/n_{.j})] = t_{i.}$ since $\sum_{j-1}^{k+1} t_{ij} = t_{i.}$. Since there are $(k + 1)^2$ essential parameters and only k^2 observations (no additional information is contributed by $n_{.j} - \sum_{i=1}^{k} t_{ij}$ in each sample), clearly $N_{..}$ cannot be estimated. Also, certainly, if we do not trust the t_{ij} to reflect the behaviour of the N_{ij}, the $N_{.j}$ or $\sum_{j=1}^{k} N_{.j}$ cannot be estimated. It is also easy to give numerical examples in the case $r = 2$ to show that even if t_{ij} is assumed to have expectation $t_{i.}N_{ij}/N_{i.}$, the distribution of the s_{ij} may be the same for quite different values of the several parameters. Clearly, identification is not possible.

If, however, sampling takes place in all strata, there will be $(k + 1)^2$ observations and the solutions of equations (5) will lead to consistent estimates, provided $|S| \neq 0$. The case where mortality occurs is related to this situation. All strata are assumed to be tagged and sampled. Those dying form the $(k + 1)$st strata so that $N_{k+1.} = 0$. Hence, if assumptions I and II′ are regarded as valid, then the equations (8) lead to consistent estimates of $N_{i.}$ for $\sum_{i=1}^{k+1} n_{ij} = \sum_{i=1}^{k} n_{ij} = n_{.j}$. Hence, $N_{..}$ is also estimable.

Furthermore, if an analysis, similar to that at the beginning of this section for the simple model, is carried out, i.e., consider t'_{ij} the random survivors of t_{ij} with probability of survival N'_{ij}/N_{ij}, then the modified variance of s_{ij} can be obtained. The distribution is not identical with the original but the leading (and dominant) term in the asymptotic variance formula is unchanged. This of course assumes a binomial model for survival as well as the original assumptions and limiting conditions used to calculate (19). If any of these fail to hold here, replication must be resorted to for variance estimates.

7. Variable number of strata. In some situations the number of strata will change between the times of tagging and sampling. This may occur either where the distribution is by area or by time.

Suppose there are m strata at the time of tagging or marking and r strata at the time of sampling or recovery. Consider $m > r$, with assumption III holding. The equations (8) yield m equations in r unknowns. The simplest device is the combination of some of the tagging periods or areas to form a system that has a unique solution. Of course, using assumption IV, an optimum asymptotic solution could be found by determining the modified minimum χ^2 estimates. If assumption III′ holds, there are r equations in m unknowns, from (11), and no solution is possible.

In case $m < r$, these conclusions are reversed; i.e., no estimation possible with assumption III, estimation possible with assumption III′. In general, it is reasonable to expect that either assumptions III or III′ can be made so that estimation of $N_{..}$ is possible.

8. Estimates of migration. In studies of migration, the N_{ij} will be of interest. If assumptions I, II, and II′ are made, then $N_{i.}$ and $N_{.j}$ $(i, j = 1, 2, \cdots, r)$ are both estimable, estimates being determined as solutions of equations (8) and (5), respectively. Also, an estimate of N_{ij} is

$$\hat{N}_{ij} = \frac{s_{ij}\hat{N}_{i.}\hat{N}_{.j}}{t_{i.}n_{.j}}. \tag{21}$$

Since (17) now holds, the consistency of $\hat{N}_{i.}$, $\hat{N}_{.j}$ proven earlier, implies the consistency of $\hat{N}_{ij}$. It is necessary to consider the dual limiting conditions that $n_{i.}$, $n_{.j}$, $N_{i.}$, $N_{.j}$ all tend to infinity.

If it is permissible to make assumption IV and further, for simplicity, that $t_{i.}/N_{i.} \to 0$, $n_{.j}/N_{.j} \to 0$, then by methods similar to those used in Section 6, the asymptotic variances of the $\hat{N}_{i.}$, $\hat{N}_{.j}$, and $\hat{N}_{ij}$ may be derived.

Let

$$\eta = \begin{bmatrix} N_{11}N_{12} & \cdots & N_{1r} \\ N_{21}N_{22} & \cdots & N_{2r} \\ \cdots\cdots & \cdots & \cdots\cdots \\ N_{r1}N_{r2} & \cdots & N_{rr} \end{bmatrix}$$

and η_{aj} be the (signed) cofactor of N_{aj} in η. Then

$$\text{A.V.}\left(\frac{\hat{N}_{.j}}{N_{.j}}\right) = \frac{1}{|\eta|^2}\sum_a\sum_b \frac{\eta_{aj}^2 N_{ab}N_{a.}N_{.b}}{t_{a.}n_{.b}} \tag{22}$$

$$\text{A.V.}\left(\frac{\hat{N}_{ij}}{N_{ij}}\right) = \frac{1}{|\eta|^2}\sum_a\sum_b \left[\frac{[\eta_{aj} + \eta_{ib}]^2 N_{ab}N_{a.}N_{.b}}{t_{a.}n_{.b}}\right] + \frac{N_{i.}N_{.j}}{t_{i.}n_{.j}}\left[\frac{1}{N_{ij}} + \frac{4\eta_{ij}}{|n|}\right]. \tag{23}$$

It was noted in an earlier section that the situation

$$N_{ij} \propto N_{i.}N_{.j}$$

is of importance both from the biological aspect and from its effect on the estimation problem.

Under the restrictions just noted it is easy to construct a test of this hypothesis. For the s_{ij} will be asymptotically normal and having asymptotically zero covariances, and it follows that they are asymptotically independent.

Thus, under the restrictions that $n_{.j}$ and $t_{i.}$ be small relative to $N_{..}$, an approximate test of the hypothesis of complete mixing, i.e., (11), is based on the statistic

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{r} \frac{\left(s_{ij} - \frac{t_{i.}n_{.j}}{\hat{N}}\right)^2}{\frac{t_{i.}n_{.j}}{\hat{N}}}, \tag{24}$$

where $\hat{N} = n_{..}t_{..}/s_{..}$.

If the $n_{.j}$ are considerably larger than the $t_{i.}$ and are not small relative to $N_{..}$, this test should be used with caution, for the type 1 error may be much larger than the nominal significance level. This is partly due to the fact that N_0 is not exactly the modified minimum χ^2 estimate of $N_{..}$. The inflation of χ^2 in (14), caused by the underestimate of $\sigma^2(s_{ij})$, is more serious. The exact variance of the s_{ij} contains terms involving $N_{.j}$, $N_{i.}$, N_{ij} which cannot be estimated by the modified minimum χ^2 method. Hence, no asymptotically efficient estimates of these parameters exist under the hypothesis. An approximate correction may be obtained by estimating the $N_{.j}$ from equations (5), and substituting these estimates in (13).

Similar χ^2 tests may be employed to test the hypothesis that sampling (or tagging) is proportionate in the different strata, i.e.,

$$H: \frac{n_{.j}}{N_{.j}} = \lambda_j = \lambda \qquad \text{for } j = 1, 2, \cdots, r.$$

For, under H, $E[s_{ij} \mid t_{ij}] = \lambda t_{ij}$ so that $E(s_{i.}) = \lambda t_{i.}$. Also, using assumption IV (actually we may dispense with the assumption that the t_{ij} have expectations given by II′), for t_{ij} small relative to $N_{.j}$ $\sigma^2(s_{i.}) = t_i(\lambda - \lambda^2)$. Under H we may estimate λ by

$$\hat{\lambda} = \frac{\sum_j n_{.j}}{\hat{N}_0} = \frac{n_{..}}{\hat{N}_0}$$

so that the χ^2 statistic (with $r - 1$ d.f.) is

$$\sum_{i=1}^{r} \frac{\left[s_{i.} - t_{i.}\frac{n_{..}}{\hat{N}}\right]^2}{t_{i.}\left[\frac{n_{..}}{\hat{N}_0} - \left(\frac{n_{..}}{\hat{N}_0}\right)^2\right]}. \tag{25}$$

The test for proportionate tagging is obtained by interchanging n and t in this formula.

REFERENCES

[1] Olav Aasen, "Estimation of the stock strength of the Norwegian winter herring." Mimeographed paper presented to Special Scientific meeting of the International Council for the Exploration of the Sea, Paris, 1954.

[2] E. W. BARANKIN, "Locally best unbiased estimates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 477–501.

[3] DOUGLAS G. CHAPMAN, "Some properties of the hypergeometric distribution with applications to zoological sample censuses," *Univ. California Publ. Stat.*, 1 (1951), pp. 131–160.

[4] D. B. DELURY, "On the assumptions underlying estimates of mobile populations," *Statistics and Mathematics in Biology*, Iowa (1954), pp. 287–293.

[5] PAUL S. DWYER AND M. S. MACPHAIL, "Symbolic matrix derivatives," *Ann. Math. Stat.*, Vol. 19, (1948), pp. 517–534.

[6] KARL W. KENYON, VICTOR B. SCHEFFER, AND DOUGLAS G. CHAPMAN, "A population study of the Alaska fur-seal herd," *U. S. Fish and Wildlife Service Special Scientific Report—Wildlife*, No. 12 (1954), pp. 1–77.

[7] A. T. PRUTER AND R. VAN CLEVE, "A preliminary study of the population of English Sole, *Parophrys vetulus*, in Holmes Harbor, Washington," *Fisheries Research Papers*, Washington Dept. of Fisheries, 1954, pp. 3–18.

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

[9] WILLIAM F. THOMPSON AND WILLIAM C. HERRINGTON, "Life history of the Pacific halibut. (1) Marking experiments," *Report of the International Fisheries Commission*, No. 2 (1930), pp. 1–137.