
Improvements in the Interpretation of Recapture Data

Author(s): Norman T. J. Bailey

Source: *Journal of Animal Ecology*, Vol. 21, No. 1 (May, 1952), pp. 120-127

Published by: British Ecological Society

Stable URL: <https://www.jstor.org/stable/1913>

Accessed: 03-06-2024 02:40 +00:00

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.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

British Ecological Society is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Animal Ecology*

IMPROVEMENTS IN THE INTERPRETATION OF RECAPTURE DATA

By NORMAN T. J. BAILEY

Department of Human Ecology, University of Cambridge

1. INTRODUCTION

It is usually possible to estimate the size of human populations and communities of sessile organisms such as plants by using methods which involve fixed sampling units. On the other hand, with mobile individuals such as animals different procedures are required. One of the most widely known techniques is the use of recapture data. Records showing the proportions of marked animals recaptured seem to have been first employed by C. G. J. Petersen as early as 1889 in the study of the growth and migration of fish. Subsequently (1894 and later) he derived estimates of mortality from such data. However, Petersen's method of estimating relative abundance from length-frequency graphs did not provide values of the *total* population size. So far as I am aware, the application of recapture data for such a purpose was first made by Lincoln (1930). He estimated the total number of duck in North America, and the formula he used is sometimes called the 'Lincoln Index'. The principle is simple. A number of animals are captured, marked and released. When the marked animals are thought to have mingled thoroughly with the unmarked, a further sample is captured and the proportion of marked individuals noted. Then an estimate of the total population size is given by dividing the number of marked animals released by the proportion found to be marked in the second group of captures. Jackson (1933) independently used the same method for estimating the density of tsetse flies. In a subsequent extension (Jackson, 1937, 1939) he used his 'negative' and 'positive' methods to allow for the complication introduced by birth- and death-rates. Later work by Jackson (1940) discusses the effect of migration. Dowdeswell, Fisher & Ford (1940, 1949) have applied recapture methods to Lepidoptera populations, catching, marking and releasing moths on several different days. This kind of data can be set out in triangular arrays—the well-known 'trellis diagrams'. Fisher & Ford (1947), in their discussion of trellis diagrams, showed how to estimate daily numbers, taking into account a death-rate obtained from the average time interval between the release and recapture of marked specimens. The published papers referred to above have been largely concerned with *estimates* of population size, and of birth- and death-rates; there has been little discussion of the *precision* of these estimates. Jackson (1937, 1939) did give some attention to the latter problem but his method of fitting curves to series of 'standardized' recapture rates is not entirely satisfactory owing to absence of proper weighting factors. Although Schnabel (1938) considered the maximum-likelihood estimation of the total number of fish in a lake using data for repeated sampling, she gave no discussion of the relevant precisions. Schumacher & Eschmeyer (1943) have used a somewhat different method of analysis, but this has the disadvantage of not leading to efficient estimates, although their standard errors can be calculated. More recently,

Gilmour, Waterhouse & McIntyre (1946) have made an ingenious study of the population density of the sheep blowfly with a combination of recapture data and random walk theory. The review paper by Ricker (1945), which discusses some applications of recapture methods to fishery problems, is perhaps also worth mentioning. A fairly elaborate mathematical treatment of the estimation of death-rates alone from recapture material has been given by Leslie & Chitty (1951). In another recent paper (Bailey 1951) I examined both the bias and precision of the usual estimate of population size for simple recapture, and suggested some modifications to improve estimates from small samples. Maximum-likelihood treatments of Jackson's 'positive' and 'negative' methods, involving birth- and death-rates as well as population sizes, were also given. There was in addition some discussion of trellis diagrams, and particular attention was drawn to the simplification introduced by using the 'triple-catch' method, which involves the simple trellis diagram obtained where catching is confined to three occasions only.

The purpose of the present discussion is to give an elementary account of the results obtained in my 1951 paper, to which the reader may refer for the detailed mathematical analysis and for a description of some of the more complicated applications. A few new results are also given, with an indication of the method of derivation. These comprise the special cases of Jackson's 'positive' and 'negative' methods when catching occurs on only three occasions. It will be assumed that the reader is familiar with the simple ideas involved in considering estimates of unknown quantities, together with their bias and variance (or the latter's square root, the standard error). Thus if an estimate, of population size, for example, is biased, this means that in the repeated use of such an estimate it will be on average too high or too low, according to whether the bias is positive or negative. It is desirable where possible to employ estimates which are unbiased or nearly so. The standard error is a measure of the accuracy of the estimate to which it is attached. In large samples at any rate, when the estimate obtained by maximum likelihood will be approximately normally distributed, there will be roughly one chance in three that the true population size lies outside a range of one standard error on each side of the estimate obtained; or about one chance in twenty that the true value is more than two standard errors away.

The formulae given require no more than arithmetical substitution for the various algebraic symbols. In the general case of Jackson's 'positive' method maximum-likelihood estimation necessitates the use of scores and information functions, for an account of which reference should be made to my 1951 paper.

It is further assumed in the present elementary treatment that animals are selected at random from the population; and that marked animals have the same birth- and death-rates, and are equally likely to be caught, as the unmarked. Whether these assumptions are justified in any particular case is for the investigator concerned to decide by the light of experience. The more complicated situation occurring when there are important differences of behaviour between marked and unmarked animals requires a more complex analysis beyond the scope of the present paper.

2. SIMPLE RECAPTURE

(a) *Direct sampling*

In simple recapture we catch, mark and then release a animals from a total population of size x . When these marked animals have freely mixed with the unmarked we catch a further sample of size n , of which r are found to be marked. The problem is to find a satisfactory estimate of x , together with its precision in terms of a , r and n . We assume that there is no emergence of new unmarked animals and that any death-rate which operates affects marked and unmarked animals in the same way (although the latter may hold for large numbers it is unlikely to be true for small ones). In the present subsection we are considering *direct* sampling, that is the standard errors are calculated with respect to repeated sampling for which the number of animals caught on the second occasion, n , remains constant. In practice this condition may be very closely fulfilled, even if not exactly. For example, we may decide beforehand to catch 50 animals on the second occasion, but in the event it may be easy to capture the first 45 but prohibitively difficult to obtain all the remaining 5 in the time available. The alternative procedure of *indirect* sampling will be considered in the following subsection.

Now the ordinary maximum-likelihood estimate of x is

$$\hat{x} = an/r, \quad (1)$$

with variance

$$\text{var } \hat{x} = a^2 n(n-r)/r^3. \quad (2)$$

These formulae are quite satisfactory for samples which are large enough for the average value of r to be fairly large. For if we write

$$m = Er = an/x, \quad \text{where } E \text{ represents the expectation or average,}$$

then, omitting samples for which $r = 0$, it can be shown the estimate \hat{x} has a positive bias of order m^{-1} . Thus suppose $x = 1000$, and $n = a = 100$. Then $m = 10$, and although we catch as many as 100 animals the average relative bias is as high as 10 %.

A better estimate to use is one obtained by a slight modification of \hat{x} :

$$\check{x} = a(n+1)/(r+1). \quad (3)$$

The bias of \check{x} is of order e^{-m} , which is quite small even for moderate m . In the example given above, the relative bias is less than 2.5×10^{-5} . In large samples, of course, \check{x} approximates to the maximum-likelihood estimate \hat{x} , given in (1). An exact value for the variance of \check{x} in terms of x , a and n is not available, but a nearly unbiased estimate of $\text{var } \check{x}$ can be obtained and is quite satisfactory in practice. If we put

$$V = a^2(n+1)(n-r)/(r+1)^2(r+2), \quad (4)$$

then we can use V as an estimate of $\text{var } \check{x}$. The relative bias is of order $m^2 e^{-m}$, which in the example above is only 2.5×10^{-3} .

In some ecological problems it may be convenient, or even more appropriate, to use the reciprocal of population size, $y = x^{-1}$. The maximum-likelihood estimate of y is unbiased and gives no trouble. We have

$$\left. \begin{aligned} \hat{y} &= r/an, \\ \text{var } \hat{y} &= y(1-ay)/an. \end{aligned} \right\} \quad (5)$$

(b) *Inverse sampling*

As an alternative to *direct* sampling we may consider *inverse* sampling. With this method, instead of aiming to catch a predetermined total number of animals on the second occasion, we continue catching until we have a predetermined number of marked animals. This procedure is not always possible in practice, but when it is it very conveniently leads to an unbiased estimate of x , and an exact value of the variance, both of which are valid for any size of sample, however small. We have the estimate

$$x' = \frac{n(a+1)}{m} - 1, \quad (6)$$

with variance $\text{var } x' = (a-m+1)(x+1)(x-a)/m(a+2).$ (7)

3. JACKSON'S 'NEGATIVE' METHOD

The principle of the 'negative' method (Jackson 1937) is that we catch, mark and release groups of animals on several successive occasions, making no record of the numbers of *recaptures* until the last occasion. Such a method is suitable when the preliminary marking of successive batches can be carried out by relatively unskilled workers. Careful classification of animals according to the number of marks they bear is then required only on the final occasion.

Suppose that on the j th day previous to the last day a_j freshly marked animals are released. Then if there is a constant death-rate γ , this means that $a_j e^{-\gamma j}$ will survive to stand a chance of recapture on the final day. Let us catch n animals altogether on this occasion, of which r_j were first marked on the j th day previously, while r_0 are unmarked. A few animals may have been marked more than once. Although we shall lose a small amount of information it is convenient to take notice of the earliest mark only. Let x be the size of the population on the last day. In addition to estimating x , we have to estimate γ as well, or, what is perhaps a little simpler, ν , where

$$\nu = e^{-\gamma} \quad \text{and} \quad \gamma = -\log_e \nu. \quad (8)$$

Let us write

$$A = \sum_{j=1} j r_j, \quad (9)$$

and

$$c_j = a_j \left(j - \frac{A}{n - r_0} \right). \quad (10)$$

We can now obtain maximum-likelihood estimates of x and ν as follows. First we solve the equation

$$\sum_{j=1} c_j \nu^j = 0, \quad (11)$$

to give the maximum-likelihood estimate $\hat{\nu}$. This is most easily done in practice by the usual sort of iterative procedure, for which one of the standard text-books may be consulted (e.g. Whittaker & Robinson 1944). If we now write

$$\left. \begin{aligned} F(\nu) &= \sum_{j=1} a_j \nu^j, \\ F'(\nu) &= - \sum_{j=1} j a_j \nu^j, \\ F''(\nu) &= \sum_{j=1} j^2 a_j \nu^j, \end{aligned} \right\} \quad (12)$$

then the maximum-likelihood estimate of x is

$$\hat{x} = nF(\hat{\nu})/(n - r_0). \quad (13)$$

The corresponding variances are

$$\begin{aligned} \text{var } \hat{x} &= (x^2/n) \{[xF''/(FF'' - F'^2)] - 1\}, \\ \text{and} \quad \text{var } \hat{\nu} &= \nu^2 x F/n (FF'' - F'^2). \end{aligned} \quad (14)$$

We can find, if desired, the corresponding value of $\hat{\gamma}$ from Equation (8), the variance of $\hat{\gamma}$ being given by

$$\text{var } \hat{\gamma} = (d\gamma/d\nu)^2 \text{var } \hat{\nu} = xF/n (FF'' - F'^2). \quad (15)$$

It should be noticed that the foregoing treatment is unaffected by the birth or emergence of new, unmarked individuals. This analysis of the general case of Jackson's negative method may be a little too complicated for those who are not very mathematically inclined, but for a worked example reference may be made to my 1951 paper.

If the release of marked animals is confined to only two or three occasions before the last, then the analysis can be somewhat simplified as the solution of (11) becomes very much easier. It is perhaps worth giving the results for one special case, namely, when animals are released on two occasions only, $2p$ and p days before the final day. Suppose that the numbers of marked animals released on these days are a_{2p} and a_p , and that the numbers caught on the last day are r_{2p} and r_p , respectively. Then, applying the previous results we eventually obtain the estimates

$$\begin{aligned} \hat{\nu} &= (a_p r_{2p}/a_{2p} r_p)^{1/p}, \\ \text{and} \quad \hat{x} &= na_p^2 r_{2p}/a_{2p} r_p^2, \end{aligned} \quad (16)$$

together with the large sample variances

$$\begin{aligned} \text{var } \hat{\nu} &= \frac{\nu^2}{p^2} \left(\frac{1}{r_{2p}} + \frac{1}{r_p} \right), \\ \text{and} \quad \text{var } \hat{x} &= x^2 \left(\frac{4}{r_p} + \frac{1}{r_{2p}} - \frac{1}{n} \right). \end{aligned} \quad (17)$$

These maximum-likelihood estimates are of course valid for large samples. For only moderate values of n we can, as in § 2, use an improved modified estimate of x :

$$\check{x} = (n+1) a_p^2 r_{2p}/a_{2p} (r_p+1)(r_p+2), \quad (18)$$

which is approximately unbiased. As before, no exact population value of the variance of \check{x} is available, but we can use the quantity V , where

$$V = \hat{x}^2 - \frac{a_p^4 (n+1)(n+2) r_{2p} (r_{2p}-1)}{a_{2p}^2 (r_p+1)(r_p+2)(r_p+3)(r_p+4)}, \quad (19)$$

which is an approximately unbiased estimate of $\text{var } \check{x}$.

4. JACKSON'S 'POSITIVE' METHOD

The principle of the 'positive' method (Jackson 1937) is, in contrast to the 'negative' method, to release a large number of marked animals on a single occasion. Catching then takes place on a number of later occasions, the number of marked animals being recorded each time. The majority of animals captured on these latter occasions should be released again, otherwise we should have to take special account of the diminution

in the number of marked animals available for recapture. A death-rate which applies to marked and unmarked animals equally will not affect our analysis, and may be ignored, but allowance must be made for the existence of a birth-rate β . Suppose that the population is initially of size x , of which a animals are marked. After j days there will be $xe^{\beta j}$ animals altogether, of which a will be marked. On the j th day let us catch n_j animals, r_j of which bear marks.

Now the maximum-likelihood treatment of the general case of the 'positive' method is more difficult than that of the negative method, as it does not lead to such easily solved equations as (11) and (13). We have, in fact, to adopt the method of using maximum-likelihood scores and information functions. An account of this procedure is beyond the scope of the present treatment, but those who are prepared to undertake such an analysis should consult my 1951 paper (p. 300).

However, in a manner somewhat analogous to that occurring in the 'negative' method, simplified formulae are available when catching is confined to two subsequent occasions only, p and $2p$ days after the first, respectively. It can be shown that the maximum-likelihood estimates of x and $\rho = e^{\beta}$ are

$$\left. \begin{aligned} \hat{\rho} &= e^{\hat{\beta}} = (r_p n_{2p} / r_{2p} n_p)^{1/p}, \\ \hat{x} &= a n_p^2 r_{2p} / n_{2p} r_p^2, \end{aligned} \right\} \quad (20)$$

and together with variances

$$\left. \begin{aligned} \text{var } \hat{\rho} &= \frac{\rho^2}{p} \left(\frac{1}{r_p} + \frac{1}{r_{2p}} \right), \\ \text{var } \hat{x} &= x^2 \left(\frac{4}{r_p} + \frac{1}{r_{2p}} - \frac{4}{n_p} - \frac{1}{n_{2p}} \right). \end{aligned} \right\} \quad (21)$$

As with the 'negative' method we can use the modified estimate of \hat{x}

$$\check{x} = a(n_p + 1)(n_p + 2)r_{2p}/n_{2p}(r_p + 1)(r_p + 2), \quad (22)$$

which is approximately unbiased even for only moderate values of n_p and n_{2p} . An approximately unbiased estimate of the variance of \check{x} is

$$V = \check{x}^2 - \frac{a^2(n_p + 1)(n_p + 2)(n_p + 3)(n_p + 4)r_{2p}(r_{2p} - 1)}{n_{2p}(n_{2p} - 1)(r_p + 1)(r_p + 2)(r_p + 3)(r_p + 4)}. \quad (23)$$

5. TRELLIS DIAGRAMS

Jackson's 'negative' and 'positive' methods are really only special cases of a more general situation in which the majority of animals captured on a particular day are released again, while all captured specimens are examined on every occasion for evidence of previous marking. As before, it is convenient to take notice of the earliest mark only, when there is more than one. The complete data showing the numbers of captures, releases and recaptures can be exhibited as a triangular array or 'trellis' diagram. Fisher & Ford (1947) have used a method of analysing such arrays, which allows one to take into account a death-rate which is estimated from the average time interval between release and recapture. Daily estimates of population size can then be calculated, though it is unfortunately difficult to know what precision to attach to the results thus obtained.

The triple-catch method

The general treatment of a large trellis diagram presents considerable difficulty, but if we confine our attention to catching on 3 days only it is possible to obtain fairly simple estimates of population size, and birth- and death-rates. Suppose that on the first day we release s_1 marked animals. On the second day we catch n_2 animals, of which n_{21} were marked on the first day and n_{20} are unmarked; s_2 freshly marked animals are released. On the third and final day n_3 animals are caught, of which n_{31} were marked on the first day, n_{32} were first marked on the second day and n_{30} are unmarked. Let the population size on the second day be x . Suppose that the time intervals between the first and second days, and between the second and third days, are respectively t_1 and t_2 . Birth- and death-rates are β and γ , with $\rho = e^\beta$ and $\nu = e^{-\gamma}$, as before. Application of the method of maximum-likelihood leads to the estimates

$$\left. \begin{aligned} \hat{\rho} &= (n_{21} n_3 / n_2 n_{31})^{1/t_2}, \\ \hat{\nu} &= (s_2 n_{31} / s_1 n_{32})^{1/t_1}, \\ \hat{x} &= s_2 n_2 n_{31} / n_{21} n_{32}. \end{aligned} \right\} \quad (24)$$

and

The corresponding large-sample variances are

$$\left. \begin{aligned} \text{var } \hat{\rho} &= \frac{\rho^2}{t_2^2} \left(\frac{1}{n_{21}} + \frac{1}{n_{31}} - \frac{1}{n_2} - \frac{1}{n_3} \right), \\ \text{var } \hat{\nu} &= \frac{\nu^2}{t_1^2} \left(\frac{1}{n_{32}} + \frac{1}{n_{31}} \right), \\ \text{var } \hat{x} &= x^2 \left(\frac{1}{n_{21}} + \frac{1}{n_{32}} + \frac{1}{n_{31}} - \frac{1}{n_2} \right). \end{aligned} \right\} \quad (25)$$

and

For data in which n_{21} , n_{32} and n_{31} are only moderate there is some advantage in using the approximately unbiased estimate

$$\check{x} = \frac{s_2(n_2 + 1) n_{31}}{(n_{21} + 1)(n_{32} + 1)}, \quad (26)$$

for which we may use the approximately unbiased estimate of variance

$$V = \check{x}^2 - \frac{s_2^2(n_2 + 1)(n_2 + 2) n_{31}(n_{31} - 1)}{(n_{21} + 1)(n_{21} + 2)(n_{32} + 1)(n_{32} + 2)}. \quad (27)$$

REFERENCES

- Bailey, Norman T. J. (1951). On estimating the size of mobile populations from recapture data. *Biometrika*, **38**, 293.
- Dowdeswell, W. H., Fisher, R. A. & Ford, E. B. (1940). The quantitative study of populations in the Lepidoptera. 1. *Polyommatus icarus* Rott. *Ann. Eugen., Lond.*, **10**, 123-36.
- Dowdeswell, W. H., Fisher, R. A. & Ford, E. B. (1949). The quantitative study of populations in the Lepidoptera. 2. *Maniola jurtina* L. *Heredity*, **3**, 67-84.
- Fisher, R. A. & Ford, E. B. (1947). The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity*, **1**, 143-74.
- Gilmour, D., Waterhouse, D. F. & McIntyre, G. A. (1946). An account of experiments undertaken to determine the natural population density of the sheep blowfly *Lucilia cuprina* Wied. *Bull. Aust. Coun. Sci. Ind. Res.* no. 195.
- Jackson, C. H. N. (1933). On the true density of tsetse flies. *J. Anim. Ecol.* **2**, 204-9.

- Jackson, C. H. N. (1937).** Some new methods in the study of *Glossina morsitans*. *Proc. Zool. Soc. Lond.* 1936, (4), 811-96.
- Jackson, C. H. N. (1939).** The analysis of an animal population. *J. Anim. Ecol.* 8, 238-46.
- Jackson, C. H. N. (1940).** The analysis of a tsetse-fly population. *Ann. Eugen., Lond.*, 10, 332-69.
- Leslie, P. H. & Chitty, Denis (1951).** The estimation of population parameters from data obtained by means of the capture-recapture method. I. The maximum likelihood equation for estimating the death-rate. *Biometrika*, 38, 269-92.
- Lincoln, F. C. (1930).** Calculating waterfowl abundance on the basis of banding returns. *Circ. U.S. Dep. Agric.* no. 118.
- Petersen, C. G. J. (1889).** *Fisk. Beretn., Kbh.*, 1888-9.
- Petersen, C. G. J. (1894).** On the biology of our flat-fishes and on the decrease of our flat-fish fisheries. *Rep. Danish Biol. Sta.* no. IV (1893-4), 146 pp.
- Ricker, W. E. (1945).** Some applications of statistical methods to fishery problems. *Biom. Bull.* 1, 73-9.
- Schnabel, Z. E. (1938).** The estimation of the total fish population of a lake. *Amer. Math. Mon.* 45, 348.
- Schumacher, F. X. & Eschmeyer, R. W. (1943).** The estimate of fish population in lakes and ponds. *Tenn. Acad. Sci.* 18, 228.
- Whittaker, E. T. & Robinson, G. (1944).** *The Calculus of Observations*, 4th ed. London.