

05 May 1981

Kenneth H. Pollock
Department of Statistics
North Carolina State University
P. O. Box 5457
Raleigh, NC 27650
919-737-2535

A CAPTURE-RECAPTURE SAMPLING DESIGN ROBUST TO UNEQUAL CATCHABILITY

KENNETH H. POLLOCK, Department of Statistics, North Carolina State University,
Box 5457, Raleigh, NC 27650

The Jolly-Seber capture-recapture model (Jolly 1965, Seber 1965) is now becoming widely used to estimate population parameters in open populations. A major problem with using this model in practice is that the assumption of equal catchability of animals is rarely true. Unequal catchability can cause serious bias to population size estimates while survival rate estimators are less affected (Cormack 1972).

In recent years there has been a lot of research on capture-recapture models for closed populations which allow unequal catchability (Otis et al. 1978). A major practical problem with these models is that often the biologist is interested in long-term studies where a closed population model is inappropriate.

Here a design is suggested which allows application of both closed and open population models to the data. It provides "robust" estimators of the population parameters when unequal catchability is present. The robustness properties of the estimators are investigated briefly using simulation and an example is given for illustration.

INTRODUCTION

This author believes that the design of capture-recapture experiments

deserves a lot more attention from statisticians and biologists. Study design should be oriented around satisfaction of as many model assumptions as practically possible so that a simple and reasonably efficient model can be used for estimation.

At the basis of many capture-recapture sampling models is the assumption that all animals are equally likely to be caught in each sample (The Equal Catchability Assumption). This assumption is most unlikely to be realistic in wildlife populations and two general types of alternatives exist:

(i) Heterogeneity: The probability of capture in any sample is a property of the animal and may vary over the population. That is, animals may vary in capture probabilities according to age, sex, social status and many other factors.

(ii) Trap Response: The probability of capture in any sample depends on the animal's prior history of capture. That is, animals may become "trap-shy" or "trap happy" depending on the type of trapping method used. Either one or both of these two types of alternatives may be acting in a particular animal population.

The traditional capture-recapture model used by biologists for closed populations (populations closed to additions or deletions) in short-term studies is the Schnabel Model (Schnabel 1938) which requires The Equal Catchability Assumption. In recent years there has been a lot of research on models for closed populations which allow heterogeneity and or trap response of the capture probabilities. Otis et al. (1978) published an important monograph on these models which allows their routine use by biologists.

The capture-recapture model becoming used by biologists for open

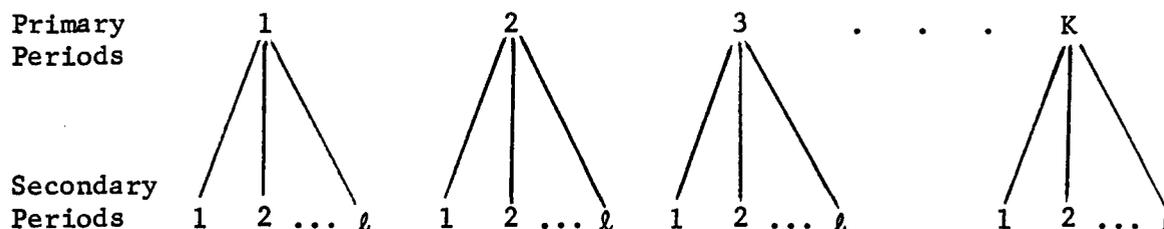
populations in long-term studies is the Jolly-Seber Model (Jolly 1965, Seber 1965). This model requires The Equal Catchability Assumption and while some weakening of the assumption is possible the complexity of open population models is likely to preclude general models which allow heterogeneity and or trap response.

During the preparation of a review of capture-recapture methods recently (Pollock 1981) I realized that statisticians have drawn a very sharp distinction between closed and open population models which is perhaps rather artificial. To quote "In practice a series of short-term studies may be carried out. One approach to analysis would be to analyze each short-term study using the closed population models ... which allow unequal catchability. Then all the sampling periods in each short-term study could be pooled and survival estimators between these short-term studies could be estimated using the Jolly-Seber Model. This approach allows population size estimation under models allowing unequal catchability while survival estimation, which is not so affected by unequal catchability, is under the Jolly-Seber Model."

In this article I discuss the design described above which is obviously motivated by the desire to find a design for long-term studies which is robust to heterogeneity and or trap response. There is a brief examination of its robustness properties using simulation and an example is given in some detail to illustrate the methodology for biologists.

THE DESIGNDescription

Consider the following representation of a capture-recapture sampling experiment:



where we have K primary sampling periods (for example years) and within each one of these we have l secondary sampling periods which are very close to each other in time (for example l consecutive days of trapping).

The biologist is interested in the population size for each of the primary sampling periods (N_1, N_2, \dots, N_K) assuming that the population is constant over the secondary sampling periods within a primary sampling period. He will also be interested in survival and birth rates between the primary sampling periods $(\phi_1, \phi_2, \dots, \phi_{K-1})$ and $(B_1, B_2, \dots, B_{K-1})$.

Estimation of Population Parameters

Assuming that the population is approximately closed over the secondary sampling periods within a primary period then two estimation procedures are possible.

(i) Jolly-Seber

Under this procedure all the secondary sampling periods within a primary sampling period would be "pooled". Here by "pooled" I mean that we are just interested if an animal is captured at least once or uncaptured in the primary sampling period. Then the Jolly-Seber estimates of population sizes (N_1, \dots, N_K) , survival rates $(\phi_1, \dots, \phi_{K-1})$ and birth rates (B_1, \dots, B_{K-1})

would be calculated using the pooled data.

A brief description of these estimators will now be given. For more details see Seber (1973:196) or Cormack (1973). Suppose to begin with that M_i , the number of marked animals in the population just before the i th sample, is known for all values $i=2, \dots, K$ (there are no animals marked at the time of the first sample so $M_1=0$).

An intuitive estimator of N_i , the population size at time i , is the Petersen estimator

$$\hat{N}_i = \frac{n_i M_i}{m_i} \quad (1)$$

which is based on equating the ratios of marked to total animals captured in the i th sample (m_i/n_i) to the ratio in the population (M_i/N_i).

An intuitive estimator of the survival rate from sample i to sample $(i+1)$ is

$$\hat{\phi}_i = \frac{M_{i+1}}{(M_i - m_i + R_i)} \quad (2)$$

where R_i is the number of the n_i animals captured in the i th sample which are released.

An intuitive estimator of the recruitment from time i to $(i+1)$ is

$$\hat{B}_i = \hat{N}_{i+1} - \hat{\phi}_i (\hat{N}_i - n_i + R_i) \quad (3)$$

which is simply the estimated difference between the population size at time $(i+1)$ and the expected number of survivors from i to $(i+1)$.

To complete this outline we need an estimator of the number of marked animals at each sample time (M_i) which are obviously unknown parameters in an open population. This can be obtained by equating the two ratios

$$\frac{Z_i}{M_i - m_i} \approx \frac{r_i}{R_i}, \quad (4)$$

which are the future recovery rates of two distinct groups of marked animals:

- (i) $M_i - m_i$ are the marked animals not seen at i and
- (ii) R_i are the marked animals seen at i and released for possible for possible recapture.

Note that Z_i and r_i are the members of $(M_i - m_i)$ and R_i , which are recaptured again at least once. The estimator of M_i is thus given by

$$\hat{M}_i = m_i + \frac{R_i Z_i}{r_i}. \quad (5)$$

It should be emphasized that \hat{M}_i is only defined for $i = 2, \dots, K - 1$ and thus \hat{N}_i in (1) is only defined for $i = 2, \dots, K - 1$; $\hat{\phi}_i$ in (2) for $i = 1, \dots, K - 2$; \hat{B}_i in (3) for $i = 2, \dots, K - 2$. For approximate variance formulae see Seber (1973; 202).

Notice that heterogeneity and or trap response will have a large effect on the population size estimators (1) because the sample ratio (m_i/n_i) will no longer accurately reflect the population ratio (M_i/N_i) . Notice that the marked population estimators (5) will not be so effected by unequal catchability because the two ratios in (4) will both tend to be influenced similarly. It follows that the survival rate estimators (2) which are simply ratios of the M_i 's will also be less effected by unequal catchability than the population size estimators. Cormack (1972) suggested this using an intuitive argument and Carothers (1973) documented it using a simulation for populations where heterogeneity of capture probabilities occurs.

(ii) Modified Procedure

Here we devise a procedure which attempts to minimize the influence of unequal catchability on our estimators by exploiting our two levels of sampling. Survival rate estimators which are not so influenced will be estimated exactly as under the Jolly-Seber procedure. Recall that these survival estimators will only be available for $i = 1, 2, \dots, K - 2$.

Here it is suggested that population size estimators for each primary sampling period (N_1, \dots, N_K) be obtained using closed population models allowing for unequal catchability and based only on the captures and recaptures within a primary sampling period. The easiest way for a biologist to obtain these estimates is to use the program CAPTURE developed by Otis et al. (1978). It considers a range of different models allowing for heterogeneity and or trap response and also gives an objective method of choosing the appropriate model to use. Notice that an additional advantage of this approach is that estimators are available for all primary periods ($i = 1, \dots, K$) whereas under the Jolly-Seber approach estimators are only available for $i = 2, \dots, K - 1$.

Finally the birth rate estimators can be obtained from (3) as before but now the population size estimators used are those described in the preceding paragraph. Notice that it is possible to estimate B_1, \dots, B_{K-2} whereas under Jolly-Seber it is only possible to estimate B_2, \dots, B_{K-2} . If we had an estimator of ϕ_{K-1} then B_{K-1} would also be estimable. In some experiments it may be reasonable to estimate ϕ_{K-1} by $\hat{\phi}_{K-2}$ or perhaps by an average of $\hat{\phi}_1, \dots, \hat{\phi}_{K-2}$.

The approximate variances of the $\hat{\phi}_i$ and \hat{N}_i estimators are available (Seber 1973, Otis et al. (1978)). The variances of the \hat{B}_i given here can

thus be obtained as

$$\begin{aligned} \text{var}(\hat{B}_i) &= \text{var}(\hat{N}_{i+1}) + \phi_i^2 \text{var}(\hat{N}_i) + N_i^2 \text{var}(\hat{\phi}_i) \\ &+ \text{var}(\hat{\phi}_i) \text{var}(\hat{N}_i) . \end{aligned} \tag{6}$$

To obtain (6) we assume no losses on capture and independence of the estimators $\hat{\phi}_i$ and \hat{N}_i .

Simulation

Here a small simulation study is presented to illustrate the robustness properties of the modified procedure discussed previously. Let us consider a population which is sampled for 3 years and within each year there are 5 sampling periods close enough together so that closure can be assumed. Let us assume that the population is subject to heterogeneity alone (Model M_h of Otis et al. 1978) and that each animal has the same capture probability for all 15 (3 x 5) sampling periods. The seven trials we consider are all taken from Otis et al. (1978) Table N.4.a and have differing degrees of heterogeneity as can be seen from Table 1.

(TABLE 1 TO APPEAR HERE)

For each trial we also consider two different birth and death processes acting on the population:

(i) There are no births or deaths so that the population stays constant from year to year.

(ii) There is a survival rate of 0.5 for all animals in the population and births exactly matches deaths to keep the population constant in size from year to year.

Using the simulation procedure of POPAN-2 (Arnason and Baniuk 1978)

approximate means and standard errors were obtained for \hat{N}_2 , the Jolly-Seber estimator given by (1). These values were each based on 99 simulation runs.

If we consider the 5 sampling periods within year 2 the appropriate estimate of N_2 is the jackknife estimator (\hat{N}_h) and Otis et al. (1978) have derived approximate means and standard errors for our trials using simulation.

In Table 2 comparison of the two estimators described above is carried out. When the population has no births and deaths then the Jolly-Seber estimates have smaller standard errors than the jackknife estimators but in the more realistic case when survival is 0.5 and births match deaths the standard errors of both estimators are very similar. Under both types of birth and death process the Jolly-Seber estimators typically have a much larger bias than the jackknife estimators.

(TABLE 2 TO APPEAR HERE)

AN EXAMPLE

Manley Fuller (unpublished thesis, North Carolina State University) carried out a capture-recapture study on a population of American Alligator (Alligator mississippiensis) in the vicinity of Lake Ellis Simon, North Carolina between 1976 and 1979. The study was not designed exactly as I suggest here but it does illustrate that the suggested methodology has advantages.

In Table 3, the data on between year captures and recaptures is given. This data was then used to obtain the Jolly-Seber estimates of population size in Year 2 (1977) and Year 3 (1978) and survival estimates.

(TABLE 3 TO APPEAR HERE)

In 1978 there were enough within year captures and recaptures to apply the closed population models of Otis et al. (1978). The model selection procedure in CAPTURE chose Model M_h , the heterogeneity model and hence the jackknife estimator \hat{N}_h was computed.

In Table 4 the Jolly-Seber population estimates for 1978 and 1977 are given with their standard errors together with the jackknife estimator for 1978 and its standard error. The jackknife estimator is double the Jolly-Seber estimate which is consistent with the hypothesis that there is strong heterogeneity of capture probabilities for this population. Fuller also suggests that the jackknife estimate is more consistent with other independent estimates of the population size based on other sampling techniques.

(TABLE 4 TO APPEAR HERE)

DISCUSSION

The author believes that the example and simulation study suggest that the design and modified estimation procedure given in this paper should have practical importance. Heterogeneity and or trap response of capture probabilities is extremely common for wildlife populations.

The major practical difficulty in using this design is obviously its large size. Otis et al. (1978) suggest that a closed population model requires 5 to 10 sampling periods with average capture probabilities of at least 0.1 per period for reasonable results. Thus the smallest practical design would be 3 primary periods each containing 5 secondary sampling periods. If, however, this is the size design necessary for robust estimation of population parameters it is important for the biologist to

be aware of it. In the past there have been too many small inadequate studies carried out because statisticians have not given biologists good design guidelines.

ACKNOWLEDGEMENTS -- The author wishes to thank Manley Fuller for allowing him to use his alligator data.

LITERATURE CITED

- Arnason, A. N. and L. Baniuk. 1978. POPAN-2, A data maintenance and analysis system for recapture data. Charles Babbage Research Centre, St. Pierre, Manitoba, Canada.
- Carothers, A. D. 1973. The effects of unequal catchability on Jolly-Seber estimates. *Biometrics* 29:79-100.
- Cormack, R. M. 1972. The logic of capture-recapture estimates. *Biometrics* 28:337-343.
- _____. 1973. Commonsense estimates from capture-recapture studies. Pages 225-254 in M. S. Bartlett and R. W. Hiorns, eds. *The mathematical theory of the dynamics of biological populations*. Academic Press, London.
- Jolly, G. M. 1965. Explicit estimates from capture-recapture data with both death and immigration--stochastic model. *Biometrika* 52:225-247.
- Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. *Statistical inference for capture data from closed populations*. Wildl. Monogr. 62, 135 pp.
- Pollock, K. H. 1981. Capture-recapture models: A review of current methods, assumptions and experimental design. Pages 000-000 in C. J. Ralph and J. M. Scott, eds. *Estimating the numbers of terrestrial birds*. Stud. Avian Biol. 6.
- Schnabel, Z. E. 1938. The estimation of the total fish population of a lake. *Amer. Math. Mon.* 45:348-352.
- Seber, G. A. F. 1965. A note on the multiple recapture census. *Biometrika* 52:249-259.
- _____. 1973. *The estimation of animal abundance and related parameters*. Griffin, London.

Received

Accepted

Table 1. Description of trials of model M_h .

Trial	N	$P_j, j=1, 2, \dots, N$
1	400	$P_j=0.05, j=1,200; P_j=0.15, j=201,300; P_j=0.50, j=301,400.$
2	400	$P_j=0.01, j=1,100; P_j=0.05, j=101,200; P_j=0.10, j=201,300; P_j=0.20, j=301,400.$
3	400	$P_j=0.10, j=1,100; P_j=0.20, j=101,200; P_j=0.25, j=201,300; P_j=0.30, j=301,400.$
4	400	$P_j=0.01, j=1,50; P_j=0.15, j=51,200; P_j=0.25, j=201,300; P_j=0.30, j=301,400.$
5	400	$P_j=0.20, j=1,100; P_j=0.30, j=101,200; P_j=0.40, j=201,300; P_j=0.50, j=301,400.$
8	100	$P_j=0.05, j=1,40; P_j=0.10, j=41,80; P_j=0.30, j=81,100.$
12	100	$P_j=0.10, j=1,40; P_j=0.20, j=41,80; P_j=0.30, j=81,100.$

Table 2. Comparison of estimators (standard errors).

Trial	N	\hat{N}_h	$\hat{N}_2(\phi=1)$	$\hat{N}_2(\phi=0.5)$
1	400	331.06(23.99)	250.99(13.22)	244.38(22.96)
2	400	298.06(24.52)	257.08(26.23)	258.14(52.74)
3	400	461.22(32.64)	367.06(12.44)	369.04(35.79)
4	400	417.08(30.91)	338.07(12.07)	338.27(29.19)
5	400	443.95(19.71)	386.89(4.94)	387.83(17.81)
8	100	84.66(14.32)	70.17(10.49)	72.93(30.74)
12	100	107.57(16.80)	89.98(7.83)	84.13(16.54)

Table 3. Alligator data capture histories for between year captures.

1976	1977	1978	1979
$n_1^a = 20$	$n_2 = 51$	$n_3 = 41$	$n_4 = 25$
$R_1^b = 17$	$R_2 = 49$	$R_3 = 37$	$R_4 = 25$
	$X_{11} = 9$	$X_{111}^c = 4$	$X_{1111} = 1$
	$X_{01} = 42$	$X_{101} = 1$	$X_{1011} = 1$
		$X_{011} = 18$	$X_{0111} = 6$
		$X_{001} = 18$	$X_{0011} = 5$
			$X_{1001} = 0$
			$X_{1101} = 2$
			$X_{0101} = 4$
			$X_{0001} = 5$

^a $n_1 = 20$ is the number of alligators captured in year 1.

^b $R_1 = 17$ is the number of alligators released in year 1.

^c $X_{111} = 4$ is the number of alligators first captured in year 1 which are recaptured year 2 and then again in year 3.

Pollock

Table 4. Comparison of estimators for the alligator data.

	<u>Estimate</u>	<u>Standard Error</u>	<u>95% Confidence Interval</u>
Jolly-Seber 1977	56	6.7	43-69
Jolly-Seber 1978	69	12.6	44-94
Jackknife 1978	140	28.5	84-197