Webinar

Session 3: Introduction to Modern Methods for Analyzing Capture-Recapture Data: Open Populations 1

Bryan F.J. Manly
Western EcoSystems Technology Inc.
Cheyenne, Wyoming
bmanly@west-inc.com

Summary

! The Jolly-Seber model

! The Cormack-Jolly-Seber model

! The Manly-Parr method

! Some example calculations

! The Lebreton *et al*. approach using covariates to model survival and capture probabilities

Introduction

- So far we have considered **closed population** situations where the population size does not change during the study.

- With **open population** models the processes of birth, death, and migration are allowed, and therefore the population size can change during the study.

- Studies of open populations often cover extended time periods, and the population changes that occur are of great interest to ecologists and managers.

- A popular model of this open model class is the Jolly-Seber (JS) model which requires that the number of uniquely marked and unmarked animals be recorded on each trapping occasion so that a complete **capture history** of each captured animal is available.

- Allows estimation of the parameters pertaining to population sizes, survival rates, recruitment numbers, and capture probabilities.

- Is not possible to separate survival from emigration or recruitment from immigration without additional information.

- Early proposals for analyzing data from open populations were proposed by Jackson (1939,1940,1944,1948), Fisher and Ford (1947), Leslie and Chitty, (1951) and Leslie *et al*. (1953).

- Entomological research was the motivation for much of the early work.

- The early models have been superseded by the **Jolly-Seber** model that is based on a properly defined probability model.

- Before the Jolly-Seber model was developed there was a crucial paper was by Cormack (1964). He derived one component of the likelihood used by Jolly and Seber in their more general model.

! To recognize Cormack's contribution to development of these models the term **Cormack-Jolly-Seber (CJS)** model is often used when referring to the marked animal component of the likelihood function, which allows the estimation of survival and capture probabilities.

The Jolly-Seber Model

The following notation is used, assuming that samples of animals are taken at k different times (periods):

p_j = the probability of capture in period j,

φ_j = the probability of survival from period j to period j+1,

M_j = the marked population size just before period j,

U_j = the unmarked population size just before period j,

N_j = $M_j + U_j$, the population size in period j, and

B_j = the birth numbers in the interval from period j to period j+1.

Important statistics obtained from the data and needed for the estimation of the population parameters are:

m_j = the number of animals captured at sampling occasion j that are marked,

u_j = the number of animals captured at sampling occasion j that are unmarked,

n_j = $m_j + u_j$, the total number of animals captured at sampling occasion j,

R_j = the total number of animals captured at sampling occasion j that are released,

r_j = the number of members of the R_j captured again later, and

z_j = the number of members of the marked population not captured at sampling occasion j (M_j - m_j) that are captured again later.

Parameter Estimation

- In practice estimates of population parameters will probably be obtained using one of the available computer programs (e.g., JOLLY, POPAN, or MARK) but they can be calculated in a spreadsheet.

- The estimation equations are intuitively sensible but here this will just be indicated briefly. See the Handbook for more details. The intuitive explanation of estimates can help users to understand the structure of the model and see why the various assumptions made are so crucial.

! The proportion of marked animals not seen at time j that are recaptured after time j should be approximately equal to the proportion of the animals that are released at time j that are seen again, so that

$$z_j/(M_j - m_j) \approx r_j/R_j.$$

Solving this equation for M_i gives the estimated number of **marked animals in the population just before the jth sampling period**, which is the observed number plus an estimate of the number of marked animals not seen,

$$\hat{M}_j = m_j + (R_j\, z_j) / r_j.$$

- The **survival rate** estimator is obtained from the ratio of marked animals present just before time j+1 to those present just before time j,

$$\hat{\varphi}_j = \hat{M}_{j+1}/(\hat{M}_j + R_j - m_j),$$

which allows the number of marked animals released at time j (R_j) to differ from the number captured (m_j).

- The survival estimator does not distinguish between losses due to **death** and **permanent emigration** without more information. It is therefore often called the apparent survival.

! The **capture probability** is estimated as the ratio of marked animals caught at time j to the number present in the population at time j,

$$\hat{p}_j = m_j / \hat{M}_j.$$

! The **population size** for period j can be determined by equating the sample and population ratios of marked to total animals, i.e.

$$m_j / n_j \approx M_j / N_j,$$

leading to

$$\hat{N}_j = (\hat{M}_j n_j) / m_j,$$

or

$$\hat{N}_j = \hat{M}_j / \hat{p}_j.$$

- Finally, the number of **births** between period j and period j+1 can be estimated as the difference between the estimated number in the population at time j+1 minus the number expected to have survived from those alive at time j, i.e.

$$\hat{B}_j = \hat{N}_{j+1} - \varphi_j (\hat{N}_j - n_j + R_j).$$

- See the Handbook for bias corrected estimators and variance equations.

Assumptions

Many assumptions need to be made for the Jolly-Seber model to be valid and for the estimators to be approximately unbiased. In particular:

1. Every animal alive in the population at sample time j has the same probability p_j of being captured in that sample.

2. Every marked animal alive in the population at sample time j has the same probability φ_j of survival until the next sampling occasion. This applies to all animals, marked and unmarked, for estimating the survival of all animals in the population.

3. Marked animals do not lose their marks and marks are not overlooked.

4. Sampling periods are short and effectively instantaneous so that the population does not change while sampling is taking place.

5. All emigration from the population is permanent, so that it is similar in effect to death.

Properties of JS Estimators

! Many studies have been made of the properties of the Jolly-Seber estimators. See the Handbook for more information about these studies.

The JS Likelihood Components

- The estimates of the JS model parameters are justified by maximum likelihood.

- Jolly (1965) and Seber (1965) used different likelihoods but came up with the same estimators.

The Seber likelihood can be viewed as three conditionally independent components with the overall likelihood the product of these, $L_1 \times L_2 \times L_3$, where:

L_1, is a product binomial likelihood that relates the unmarked population and sample sizes at each time to the capture probabilities.

L_2, just contains parameters for probabilities of being lost on capture.

L_3, is the component that contains all the recapture information conditional on the numbers of marked animals released at different times, and the parameters for capture and survival probabilities.

! L_3 is the likelihood that was originally derived by Cormack (1964). One way to view estimation in the full JS model is that capture and survival probabilities are estimated from L_3 only. This is then what is called the **Cormack-Jolly-Seber** model.

! Once capture probabilities are estimated the size of the population at the time of sample j can be estimated from

$$n_j \approx N_j \, p_j,$$

leading to

$$\hat{N}_j = n_j / \hat{p}_j,$$

which turns out to be the same as the JS estimator.

Restrictions and Generalizations of the Jolly-Seber Model

! Before and after the JS model was introduced various special cases and generalizations were developed and included in the computer program JOLLY.

! The **deaths only** model is useful if there is no immigration and recruitment is not occurring.

! The **births only** model is useful if there is no mortality or emigration.

! Versions of the model with constant survival and/or capture probabilities are also available in program JOLLY.

! Some generalizations of the JS model that allow for temporary effects on survival and capture rates have also been developed and are available in programs JOLLY or MARK.

Manly and Parr Method

- Manly and Parr (1968) noticed that it is possible to estimate p_j in a manner that is robust to heterogeneity of survival rates due to age or other causes.

- They defined C_j as the class of marked animals known to be alive at time j because they were captured both before and after j, and they defined c_j as the members of that class C_j that are captured in sample j. Then an estimate of p_j is

$$\hat{p}_j = c_j / C_j.$$

- Once p_j is estimated there is the obvious estimator of the population size

$$\hat{N}_j = n_j / \hat{c}_j.$$

- Survival rates and birth numbers can be estimated, and also some variances (see the Handbook).

- Modifications to the equations to reduce bias are available (see Handbook).

- A more recent development is the application of the idea behind the Manly and Parr method to animals of known ages (Manly *et al*., 2003; Handbook page 272).

Goodness of Fit and Model Selection

! There are two aspects to assessing whether the best model for an open capture-recapture model has been chosen. There is the overall **goodness of fit** of a model and the decision about which of a series of related models is the **best model**. See the Handbook for more information about these issues.

Example

! Data from a capture-recapture study on a population of the **American alligator** (*Alligator mississippiensis*) at Lake Ellis Simon, North Carolina, between 1976 and 1979 (Fuller, 1981).

! There was an error in the data as originally published, with the pattern 1001 occurring once.

! Jolly-Seber estimates are in the spreadsheet Alligatr.xls. This spreadsheet shows how you can estimate population parameters with simple calculations.

Capture history matrix for the alligator data.

1976	1977	1978	1979	Alligators
1	1	1	1	1
1	0	1	1	1
0	1	1	1	6**
0	0	1	1	5
1	0	0	1	1
1	1	0	1	2
0	1	0	1	4
0	0	0	1	5
1	1	1	0	3
0	1	1	0	12
0	0	1	0	9
0	0	2*	0	4
1	1	0	0	3
0	1	0	0	18
0	2*	0	0	2
1	0	0	0	6
2*	0	0	0	3

**Reflects 6 identical lines in the capture history.

*These animals were lost on capture and therefore not released. There were no alligators with capture history (1010).

Flexible Modeling Procedures

- An approach to modeling mark-recapture data that has been advocated by Lebreton *et al*. (1992) represents a generalization of the work of Cormack, Jolly, and Seber.

- Models are proposed for **survival** and **capture** probabilities.

- The likelihood function for a set of data is then constructed by multiplying together the probabilities of observing the recapture patterns under the model for all animals captured at least once, given the time of first capture.

- This function is then maximized with respect to the unknown parameters in the model.

Example

! Consider a study of an animal population that lasts seven years, and suppose that the captures and recaptures of an animal are recorded as

0110100,

where (as before) 1 indicates a capture and 0 indicates no capture.

! For such an animal, the **probability of the recapture pattern** is assumed to be

$\varphi_2 p_3 \varphi_3 (1 - p_4) \varphi_4 p_5 [(1 - \varphi_5) + \varphi_5 (1 - p_6)(1 - \varphi_6) + \varphi_5 (1 - p_6) \varphi_6 (1 - p_7)]$,

where φ_i is the probability of surviving from year i to year i+1, p_i is the probability of being captured in year i. and the three terms that are added within the square brackets are (i) the probability of dying in the year following the last sighting, (ii) the probability of surviving to year six, not being captured in that year, and dying in the following year, and (iii) the probability of surviving to year seven, but not being captured in either year six or year seven.

- For a given model, probabilities can be obtained for all other recapture patterns in a similar manner and the construction of the likelihood function is therefore in principle fairly straightforward.

- An important feature of this approach for modelling is the focus on the **estimation of survival and capture probabilities** instead of population sizes.

- One justification for this is that mark-recapture estimation works better for estimating these parameters than it does for the estimation of population sizes.

- Another justification is that it is survival probabilities that are important for population dynamics and a population size is just the outcome of survival and reproduction rates.

! In some cases it really will be estimates of population size that are needed. These can, however, always be obtained by using the estimation equation

$$\hat{N}_j = n_j / \hat{p}_j,$$

! This estimator is valid providing that the probability of capture is the same for marked and unmarked animal - an assumption that is **not required** for the estimation of survival probabilities.

Models for Capture and Survival Probabilities

! Because the capture and survival probabilities must be within the range 0 to 1 it is sensible to build this constraint into the model. One way that this can be done involves replacing φ_i and p_i with the logistic functions

$$p_j = \exp(u_j)/\{1 + \exp(u_j)\}.$$

and

$$\varphi_j = \exp(v_j)/\{1 + \exp(v_j)\}.$$

! Remember the same idea with the Huggins models for closed populations.

! This has the further advantage of making it easy to incorporate into the model the effects of **covariates** in the manner that is used with logistic regression.

! Example: if x_j is a measure of the **severity of the weather** from year j to year j + 1 and y_j is a measure of the **effort put into capturing animals** in year i then it may be considered appropriate to model the probabilities of capture and survival by

$$p_j = \exp(ß_0 + ß_1 y_j)/\{1 + \exp(ß_0 + ß_1 y_j)\}.$$

and

$$\varphi_j = \exp(\alpha_0 + \alpha_1 x_j)/\{1 + \exp(\alpha_0 + \alpha_1 x_j)\}$$

- There is a good deal of **flexibility** in the modeling process.

- Several covariates can be used, capture and survival probabilities can be allowed to vary with time, the age of animals, which group an animal is in, etc.

- In all cases the likelihood is maximized with respect to the α and ß parameters.

- Once these parameters have been estimated the logistic functions can be used to determine estimated survival and capture probabilities for any animal.

- A problem with using logistic regression functions with modelling survival probabilities is that there is no simple way to take into account changes in the time involved.

- E.g., if there are two years between some samples and only one year between others then this cannot be allowed for in a simple way by introducing a time between samples parameter into the model.

- A better function to use in this respect is the **proportional hazards function**

$$\varphi_j = \exp\{-\exp(-u_j)t_j\}$$

where t_j is the time between when samples j and j + 1 are taken.

- So far this approach to modelling survival does not seem to have been used.

- Care is needed in setting up models for which parameters vary with time. For example, the Jolly-Seber model with k samples has 2k - 2 parameters φ_1, φ_2, ..., φ_{k-1} and p_2, p_3, ..., p_k. However, it is only possible to estimate the product $\varphi_{k-1}p_k$, rather than separate values for φ_{k-1} and p_k.

- This type of complication occurs with other models as well. It is therefore important to know **which parameters can be estimated from the available data** and adopt an appropriate set of parameters for the logistic functions.

- The **variable metric algorithm** for maximizing the likelihood function that is currently favored is quite capable of determining a maximum even in cases where the model being fitted contains more parameters than it is really possible to estimate separately. An inappropriate parameterization may therefore not be obvious just from the results of the fitting process.

Possible Candidate Models

The type of approach that is used for analysing data involves defining a set of candidate models from which one will be chosen. For example if the sex is recorded for each individual that is marked then the following six models might be entertained for **capture probabilities**:

(a) sex*time: the capture probability varies with the sample time, and also differs for males and females;

(b) sex+time: the capture probability varies with the sample time and the constant term in the logistic model also varies for males and females;

(c) sex: the capture probability is constant over time but is different for males and females;

(d) time: the capture probability varies with time but is the same for males and females;

(e) trend: there is a trend in capture probabilities; and

(f) constant: capture probabilities are constant.

These models can all be expressed through **logistic function equations**.

- The six models for **survival probabilities** would probably be the same as those used for **capture probabilities** and by considering all different combinations of the six capture probability models and the six survival probability models there are then 6x6 = 36 possible models to be considered for the data.

- These 36 models may or may not be suitable for a particular mark-recapture study, but indicate the flexibility of the modeling approach.

Summary

- The Jolly-Seber model
- The Cormack-Jolly-Seber model
- The Manly-Parr method
- Some example calculations
- The Lebreton *et al*. approach using covariates to model survival and capture probabilities