Cormack-Jolly-Seber Models

Estimating Apparent Survival from Mark-Resight Data & Open-Population Models
Ch. 17 of *WNC*, especially sections 17.1 & 17.2

For these models, animals are captured on *k* occasions (say years) and given a unique mark during a relatively short tagging period (say week) each year. Time periods could also be weeks, months, or multiple year intervals. After occasion 1, both marked and unmarked animals are caught; tag numbers of the marked animals are recorded and unmarked animals are marked. Animals are released back into the population; accidental deaths (losses on capture) are allowed.

Mark	Mark & Resight	Mark & Resight	Mark & Resight	Mark & Resight	Mark & Resight	Resight

$$1 \xrightarrow{\phi_1} \underset{p_2}{2} \xrightarrow{\phi_2} \underset{p_3}{3} \xrightarrow{\phi_3} \underset{p_4}{4} \xrightarrow{\phi_4} \underset{p_5}{5} \xrightarrow{\phi_5} \underset{p_6}{6} \xrightarrow{\phi_6} \underset{p_7}{7}$$

The Jolly-Seber model (after Jolly 1965 and Seber 1965) is fairly general and serves as a starting point for open C-R modeling (see White et al (1982, chapter 8). This model allows year-specific estimates of apparent survival (φ), capture probability (p), population size (N), and the number of new individuals entering the population (B). While population size can be estimated, it is often very difficult to avoid substantial bias in the estimation of this parameter set because of individual heterogeneity and other issues. We will discuss the actual Jolly-Seber model later in the course but begin our work with discussion of the Cormack-Jolly-Seber model, which is a restricted model that allows only year-specific estimates of φ and p. Although less general, it has proven to be the more useful model for several reasons.

One important biological issue is that **only apparent survival can be estimated with CJS modeling**; that is 1– φ represents both animals that died and animals that merely left the population (emigration). In general, φ < S. This can be a significant matter and often misunderstood.

In open C-R studies, all sampling is done by researchers on a relatively small area. That is, animals are marked on a relatively small area each year for at least 3 occasions, e.g., years. Recaptures are only made for those animals that come back to the same (relatively very small) area where capturing is being conducted. Thus, an animal that comes back to the same general area may not be recaptured if it is a mile or two away from the capture site. This issue can be very problematic with migratory birds or fish that move considerable distances during the course of a year. Still, there are often cases where there is biological interest in φ and the fact that some animals merely "left" is not problematic to the interpretation of the data. **Note**: data can be collected by *recapturing or resighting* animals.

The approach we'll use conditions on the initial capture of an animal and models its subsequent capture history as functions of parameters associated with sampling (p) and real population change (φ). To get started and in keeping with section 17.1 of *WNC*, we'll first consider single-age models and the standard Cormack-Jolly-Seber model (CJS) in which φ and p are time-specific.

Data Structure for study with 3 occasions (LLL format = 1 if seen alive on occasion, 0 otherwise):

Capture History	Number of animals with history
100	89
110	41
101	16
111	19
010	75
011	37
001	82

Note: for 3 occasions, there are 7 observable histories (000 is not seen).

Model Structure:

Known Constants:

R_i - The number of animals released in year *i*. These releases are typically made up of releases of newly captured and marked animals AND re-releases of re-captured animals that were marked on earlier occasions.

Parameters:

φ_i is the probability that a marked animal in the study population at sampling period *i* survives until period *i+*1 and remains in the population (does not permanently emigrate). Thus, apparent survival in year *i* relates to the interval between resighting (or capture) periods *i* and *i+1*.

p_i is the probability that a marked animal in the study population at sampling period *i* is captured or observed during period *i.*

χ_i is the probability that an animal alive and in the study population at sampling period *i* is **not** caught or observed again at any sampling period **after** period *i*. For a study with *T* sampling periods, $\chi_T = 1$, and values for periods with $i<T$ can be obtained recursively as: $\chi_i = (1-\phi_i) + \phi_i(1-p_{i+1})\chi_{i+1}$

Assumptions (we will later learn how to change some of these):

1. Every marked animal present in the population at sampling period *i* has the same probability p_i of being captured or resighted.
2. Every marked animal present in the population at sampling period *i* has the same probability φ_i of survival until sampling period *i+1*.
3. Marks are neither lost nor overlooked and are recorded correctly.
4. Sampling periods are instantaneous (in reality they are very short periods) and recaptured animals are released immediately.
5. All emigration from the sampled area is permanent.
6. The fate of each animal with respect to capture and survival probability is independent of the fate of any other animal.

Model Structure Illustrations:

$$\mathbf{1} \xrightarrow{\phi_1} \underset{p_2}{\mathbf{2}} \xrightarrow{\phi_2} \underset{p_3}{\mathbf{3}}$$

Pr(111 | release at period 1) = $\phi_1 \cdot p_2 \cdot \phi_2 \cdot p_3$

Pr(110 | release at period 1) = $\phi_1 \cdot p_2 \cdot x_2$, where $x_2 = (1-\phi_2) + \phi_2 \cdot (1-p_3)$

Pr(101 | release at period 1) = $\phi_1 \cdot (1-p_2) \cdot \phi_2 \cdot p_3$

Pr(100 | release at period 1) = χ_i, where $\chi_1 = (1-\phi_1) + \phi_1 \cdot (1-p_2)(1-\phi_2 \cdot p_3)$

…

Pr(011 | release at period 2) = $\phi_2 \cdot p_3$

Competing models:

When individual covariates are not considered, the models that can be considered for 1 group of animals within an identifiable class of animals are:

1. $\phi(t), p(t)$
2. $\phi(t), p(.)$
3. $\phi(.), p(t)$
4. $\phi(.), p(.)$

Intuition on Estimation:

Capture History	Number of animals with history
100	176
110	11
101	239
111	74

$$\mathbf{1} \xrightarrow{\phi_1} \underset{p_2}{\mathbf{2}} \xrightarrow{\phi_2} \underset{p_3}{\mathbf{3}}$$

Model $\phi(t), p(t)$:

$\hat{p}_2 = 0.236421$ and $\hat{\phi}_1 = 0.719054$

How do the histories and their frequencies provide information on these 2 parameters? First, consider the histories '101' and '111'. Animals with these histories are known to have survived from year 1 to year 2 (and also to have survived to year 3). There are 313 such animals (239 with '101' history plus 74 with '111' history) and 74 of those 313 were *detected* in year 2. And, 74/313=0.236421.

Thus, because we have information from the third occasion, we can separately estimate the survival and recapture rates φ_1 and p_2 respectively. We can see how the probability statements show this:

$$\frac{N_{111}}{N_{101}} = \frac{\phi_1 p_2 \phi_2 p_3}{\phi_1 (1-p_2) \phi_2 p_3} = \frac{p_2}{1-p_2}$$

What about $\hat{\phi}_1$? Well, $\phi_1 p_2$ is the probability that an animal survives from year 1 to year 2 AND is detected in year 2 (it's referred to as a *return rate* in some literature and isn't all that useful by itself). But, given an estimate of p_2, we could then obtain an estimate of ϕ_1. Of the 500 individuals released on occasion 1, how many *returned* on occasion 2 (survived and captured)? To get this number add 11 and 74 (11 '110's and 74 '111's) to obtain 85. Thus, $\phi_1 p_2$ = (74+11)/500=0.1700. And, 0.1700/0.236421 = 0.719054.

Quoting from Chapter 4 of *C&W (4.4 – counting parameters): "*But, it is important to note that we can't separately estimate all the parameters. Consider for instance φ_2 and p_3. Can we separate them? No! In fact, the product of these two parameters is completely analogous to a return rate between occasions 2 and 3. If we wanted to separate these 2 parameters, we'd need a fourth occasion, and so on. Thus, in such a model where both survival and recapture rate are time-dependent, the terminal parameters are not individually identifiable - all we can do is estimate the product of the 2. Lebreton et al. (1992) refer to this product term as β_3. Thus, we can re-write our table, and the probability statements, as:"

encounter history	*probability*
111	$\phi_1 p_2 \beta_3$
110	$\phi_1 p_2 (1 - \beta_3)$
101	$\phi_1 (1 - p_2) \beta_3$
100	$1 - \phi_1 p_2 - \phi_1 (1 - p_2) \beta_3$

It turns out that we can estimate φ_1, p_2, *and* (φ_2 p_3) in the $\phi(t), p(t)$ model. If we constrain parameters (as we do in the other 3 models), then we can avoid that terminal product. For example, the $\phi(t), p(.)$ model allows us to estimate φ_1 *and* φ_2 because we now assume that *p* is constant across occasions.

Our actual modeling is done using maximum likelihood estimation and a multinomial distribution where each of the encounter histories is a possible outcome.

Multinomial Distribution:

$$P(X = \{x_1, x_2, ..., x_c\}) = \frac{N!}{(x_1!)(x_2!)...(x_c!)} p_1^{x_1} p_2^{x_2} ... p_c^{x_c}$$

Example from page 35 of *WNC:4*

$$f(x_1, x_2 \mid N = 5, p_1 = .6, p_2 = .38, p_3 = (1 - .6 - .38)) = \frac{5!}{(x_1!)(x_2!)(x_3!)} (0.6)^{x_1} (0.38)^{x_2} (0.02)^{5 - x_1 - x_2}$$

```
dmultinom(x=c(5,0,0),prob=c(.6,.38,.02)) # = 0.07776 = probability of 5 completed passes,
                                          # 0 incomplete passes, and 0 intercepted passes

dmultinom(x=c(4,1,0),prob=c(.6,.38,.02)) # = 0.24624
```

Maximum likelihood for the CJS model using the encounter histories

From page 421 of *WNC*:

$$\Pr(\{x_\omega\} \mid R_1) = \frac{R_1!}{\prod_\omega x_\omega!}(\phi_1 p_2 \phi_2 p_3)^{x_{111}}[\phi_1 p_2 (1-\phi_2 p_3)]^{x_{110}}$$
$$\times[\phi_1(1-p_2)\phi_2 p_3]^{x_{101}}(x_1)^{x_{100}}$$

- As for known-fate modeling, maximum likelihood finds that combination of the parameters that maximizes the likelihood of observing the set of encounter histories according to the frequencies at which they were observed in the study.
- MLE's of the parameters, their variances, and their covariances are available of the parameters, their variances, and their covariances are available.
- Different models can be evaluated that constrain the parameters in various ways.
- And, those models can include sub-models that allow the various parameters to be functions of covariates.
- AICc can be used to evaluate the competing models of interest for the problem at hand.

Animals that aren't captured on the 1st occasion are useful and the likelihood can be adjusted to handle them just fine. And, animals can be removed if desired without problem as discussed in Ch. 17 of *WNC*.

The following shows the EH for an animal first uniquely marked on occasion 3.

$$0 \rightarrow 0 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 0 \rightarrow 1$$
$$\phi_3 \quad \phi_4 \quad \phi_5 \quad \phi_6$$
$$p_4 \quad p_5 \quad p_6 \quad p_7$$

This animal provides no information about ϕ_1 or ϕ_2 or about p_2 or p_3. The probability of observing this encounter history is: $\phi_3 p_4 \ \phi_4 p_5 \ \phi_5(1-p_6) \ \phi_6 p_7$.

Input format for MARK (live recaptures or LLLL… format):

For Program MARK, the EH for each animal can be entered on 1 row, e.g.,

1101 1;
1110 1;
1001 1;
etc.

OR, you can use summary formatting to indicate EH's for 112 animals. This is obviously much more efficient than writing out 112 lines; 12 repeats of the 1st line, 20 repeats of the 2nd line, etc..

1111 12;
1011 20;
1000 80;

You can also indicate multiple groups this way.

1111 12 15;
1011 20 32;
1000 80 101;

And, … you can include individual covariates.

So, this type of modeling is quite useful

IF … the biology of the animal makes ϕ a biologically meaningful parameter for the study in question.

Background reading:

- Ch. 17 – *WNC*
- Ch. 3 & 4 – *CW*
- Lebreton, J. D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modeling Survival and Testing Biological Hypotheses Using Marked Animals - A Unified Approach with Case-Studies. Ecological Monographs 62:67-118.
    - *not required for class but excellent background for those using these models in their work.*