Lab 7: Parameter estimation: mark-recapture analysis!

NRES 470/670

April 4, 2017

## Open versus closed populations

**Closed population** â€“ a population in which there is no recruitment (birth or immigration) or losses (death or emigration) during the period of study. Geographic and demographic closure.

**Open population** â€“ a population that changes in size and composition due to births, deaths, and movements.

## Overview of capture-recapture models:

### Closed-population models

* *two samples* â€“ Lincoln-Petersen model
* *several samples* (k>2) â€“ Schnabel (Schumacher-Eschmeyer) model and models in program CAPTURE (which can be run through Program MARK.

#### LINCOLN-PETERSEN MODEL (L-P) (2-sample closed-population model)

Basic underlying concept:

A sample of animals is caught, marked, and released. Later a sample of animals is Captured, of which animals are *Recaptures* that were previously Marked.

If capture probability (*p*) is independent of marking status, then the proportion of marked animals in the second sample should be equivalent to the proportion of marked animals in the total population (*N*) so that

where is the total catchable population size. Solving for yields the estimator:

If sample size is small, the above estimator is biased. For example, what happens if the number of recaptures is zero?

A modified version with less bias was originally developed by Chapman (1951) and is commonly called the modified Petersen estimate in fisheries:

This formula is a **statistic** estimating total population size on the basis of a **sample**.

We still need to make inference about the **parameter**, , on the basis of the statistic, .

To do this, we need to understand the **sampling variability**. That is, if we collected a different sample from the *same population*, we might get a very different answer for . If we took 100 or 1000 different samples, we might get 100 or 1000 different estimates for !! The variation among these estimates is the **sampling variability**.

For the Lincoln-Peterson estimator, the sampling variability (variance) can be estimated as:

The standard deviation of the sampling variance is just the square root of the sampling variance:

Finally, the 95% confidence interval for is approximately 2 standard deviations from the value of .

#### Fundamental Assumptions of Lincoln-Petersen estimator:

* The population is closed (geographically and demographically).
* All animals are equally likely to be captured in each sample.
* Capture and marking do not affect catchability.
* Each sample is random.
* Marks are not lost between sampling occasions.
* All marks are recorded correctly and reported on recovery in the second sample.

**Q** What is the estimate of per-capita capture probability in the Peterson estimator??

#### SCHNABEL ESTIMATOR (k-sample closed-pop'n model)

Schnabel 1938, Schumacher and Eschmeyer 1943

This method extends the Lincoln-Peterson method to a series of samples in which there are 2, 3, 4,..., k samples. Individuals caught at each sample are first examined for marks, then marked and released.

Only a single type of mark need be used because we just need to distinguish 2 types of individuals: marked, caught in one or more prior samples; and unmarked, never caught before. For each sample t, the following is determined:

= Total number of individuals caught in sample

= Number of individuals already marked (Recaptures) when caught in sample

= Number of marked animals in the population just before the th sample is taken.

Schnabel treated the multiple samples as a series of Lincoln-Peterson (L-P) samples and obtained a population estimate as a weighted average of the L-P estimates which is an approximation to the maximum likelihood estimate of N:

What about the sampling variance? It can be computed as:

Note that this formula gives you the sampling variance on the *inverse* of N! How do we compute a confidence interval around ??

One way is to compute a confidence interval on the inverse of N, then take the inverse of the lower and upper bounds (limits) of this confidence interval!

#### Assumptions of the Schnabel method

Same as Lincoln-Petersen estimator but assumptions apply to all sampling periods. In other words, every individual in the population is assumed to have the same capture probability for a given sampling occasion (although capture probabilities can vary among sampling periods).

##### Advantage of k-sample methods

The major advantage of multiple sampling is that it is possible to evaluate the data for violations of assumptions, such as unequal capture probabilities.

## Exercise 1: Working with closed populations!

Let's use the following example from the ["gentle introduction to program mark"](http://www.phidot.org/software/mark/docs/book/pdf/chap14.pdf). This example represents a closed population that has been surveyed 6 times.

Load up the CSV file [here](simple_closed.csv). The first few lines should look something like this:

## sample1 sample2 sample3 sample4 sample5 sample6  
## 1 0 0 0 1 1 1  
## 2 0 0 0 1 1 1  
## 3 0 0 0 1 1 1  
## 4 0 0 0 1 1 1  
## 5 0 0 0 1 1 1  
## 6 0 0 0 1 1 1

1a. First, let's imagine that we only have samples 1 and 2! Using this data, compute the L-P estimate of abundance using Eq. 2. Also compute the confidence interval for your abundance estimate. Show your work! I recommend using EXCEL for this!

1b. Now, use the Schnabel method to estimate abundance (and confidence interval). Show your work! Again, I recommend using EXCEL for this!

### Open-population models in Program MARK

(note: geographic closure is still a critical assumption)

*Cormack-Jolly-Seber* model is the most basic model in Program MARK.

See the [parameter estimation lecture](LECTURE15.html) for more information about open-population mark-recapture analysis!

#### TUTORIAL: working with open populations!

For this exercise, we will use the classic European Dipper data!

These data should look like this! Here it is in R (just the first 15 lines)!

library(mra)  
  
data(dipper.histories)  
head(dipper.histories,15)

## h1 h2 h3 h4 h5 h6 h7  
## 1 1 1 1 1 1 1 0  
## 2 1 1 1 1 1 0 0  
## 3 1 1 1 1 0 0 0  
## 4 1 1 1 1 0 0 0  
## 5 1 1 0 1 1 1 0  
## 6 1 1 0 0 0 0 0  
## 7 1 1 0 0 0 0 0  
## 8 1 1 0 0 0 0 0  
## 9 1 1 0 0 0 0 0  
## 10 1 1 0 0 0 0 0  
## 11 1 1 0 0 0 0 0  
## 12 1 0 1 0 0 0 0  
## 13 1 0 1 0 0 0 0  
## 14 1 0 0 0 0 0 0  
## 15 1 0 0 0 0 0 0

Program MARK wants a particular type of input file (.INP). For the dipper data, it should look something like [this](ed_males.inp)

1. Open Program MARK! You can download the software [here](http://www.phidot.org/software/mark/downloads/index.html)
2. Mark automatically creates an output file when you open a data file so move the dipper data file to your personal folder to prevent existing examples from being overwritten.
3. Double-click the Mark icon to open Program Mark. Click the spreadsheet icon in the upper left corner to open a menu for Specifications for Mark Analysis. This menu allows you to specify the kind of analysis you will conduct (Select Data Type). Today we will start with a data set that includes live recaptures only so be sure this Data Type is selected (Cormack-Jolly-Seber model).
4. Look to the right and you will see a button: Click to Select File. Click this button and browse to find the "ed\_males.inp"" file. Double click this file to open this file in Program Mark. Now click the view this file button, which will allow you to see the data file. You will see some file information at the top enclosed in //, which tells Mark this information is not part of the data. Below, you will see encounter histories followed by a space pairs of 1s and 0s, also separated by a space, all of this followed by a ;. The encounter history indicates the occasions when each individual was encountered (actually observed), indicated by a 1, or not ed, indicated by a 0. The column to the right indicate how many individuals in the population exhibit this particular capture history. The ; at the end indicates the end of the record. Note that in this encounter history each individual has its own record (the value in the final colummn is always 1). However, it is possible to specify only the unique observed encounter histories and indicate the number of individuals with each history.
5. We now have to provide Mark some information about the data. You should provide a title for the data to keep you results organized. Below the data file selection area you will find some buttons and counters to provide additional information. Encounter occasions needs information about the number of possible times an individual was encountered. Count the number of columns in the dipper encounter history (there are 7) and enter this number for encounter occasions. Once you have completed these tasks click OK; Mark has the basic information it needs to perform an analysis.
6. A window will open entitled â€œapparent survival parameter (phi) males of live recapturesâ€. Before we discuss this window we need to open an additional window. Click on the PIM button on the top toolbar, then click on â€œopen parameter index matrixâ€. Click select all then OK. Click on the Window button on the top toolbar then click on Tile. You should see 2 similar appearing windows all with the upper diagonal of a matrix. Look more closely and youâ€™ll see that the window for male survival has numbers ranging from 1 to 6 as columns go from left to right. The encounter probability matrices have numbers 7 to 12. These numbers specify the model structure by defining the number of survival and capture probabilities we wish to estimate! The model you have specified allows survival and encounter probabilities to vary annually.
7. Another useful way to visualize the parameters you wish to estimate is the "Parameter Index Chart". Click on the PIM button on the top toolbar, then click on â€œopen parameter index chartâ€. Here you see all parameters in one window- six different survival parameters and six different encounter probability parameters.

**Q**: Why are there only 6 survival parameters, when there are seven surveys??

**Q**: Why are there only 6 capture probabilities, when there are seven surveys??

1. To run this model click on the small button with the green arrow (third from left). A new window will open asking for the title for the analysis and the model name. Use 'dippertest' or another descriptive name for the analysis. Identify the model as: "{phi(t),p(t)}", which indicates that survival and encounter probabilities can each vary across time, independently. This model is among the most general we can run for this data set.
2. Click OK, and a new window will ask you if you want to use the identity matrix because no design matrix was specified. Click yes (or OK) weâ€™ll learn more about the design matrix later. A new black window with scrolling text will open indicating that MARK is doing calculations (the numerical methods to maximize the likelihood for the data and specified model).
3. When Mark is finished a new window will open asking you if you want to append the model to the database. Click yes and a new table (The Results Browser) will open. The model is identified on the left based on the notation you provided, AIC, AIC weight, number of parameters and deviances are all reported. For now you can consider AIC as a ranking of the quality of the models from best (low AIC) to worst (high AIC). "Deviance" is a measure of how well the model fits the data.
4. Re-open the PIMs for survival and capture probability. Use the minus button to reduce the numbers in survival windows to 1 for both males and females and 2 for the both the windows for encounter probabilities (for the latter reduce all matrix entries to 1 then use the plus button to increase them to 2). Use the green arrow to run this model and follow the same procedure as for the earlier model to run this model. Identify the model as {phi(.),p(.)}, which indicates that both parameters are constant across both groups and time. This is the simplest model we can run for these data. Again, use the identity matrix and append the results to the Results Browser. The â€œdotâ€ model performs better (lower AIC) and has fewer parameters so it is the best of the two models run so far!
5. **Examine Parameter Estimates**: To examine parameter estimates click on the model, then move the cursor to the top tool bar and click on 'Retrieve'. Then click on current model. To see the parameter estimates for the retrieved model return the curser to the Results Browser and click the fourth icon from the left (the third minipage from the left). A text file will open with a list of parameters and their estimates ("view estimates of real parameters in notepad window"). For the 'dot' model you will only see one survival estimate and one encounter probability because you specified that both parameters would be constant across time.

Now retrieve the {phi(g*t),p(g*t)} model and examine parameter estimates for this model. You will see 6 survival estimates and 6 estimates for detection probability. These are indexed using the numbers you provided in the PIMs. Notice that the 6th estimates for both phi and p have standard error that are either very large or zero. *These are the estimates for the last survival and encounter probability for each group, which cannot be estimated*; only their product can be estimated.

## Exercise 2- CJS models in Program MARK

2a. In the CJS model with time-constant survival and capture probability, write out the *likelihood* of the following capture history:

1 1 1 1

2b. In the CJS model with time-constant survival and capture probability, write out the *likelihood* of the following capture history:

1 0 0 1

2c. In the CJS model with time-constant survival and capture probability, write out the *likelihood* of the following capture history:

1 0 0 0

2d. [building off the demonstration in Program MARK]. Build and run the following model: capture probability varies by year, but survival is constant across time. Is this model better than the current top model? What are the parameter estimates for this model? What are the *confidence intervals* for these parameters?

2e. Build and run the following model: capture probability is constant across time, but survival exhibits temporal variability. Is this model better than the current top model? What are the parameter estimates for this model? What are the *confidence intervals* for these parameters?

2f. Use your results to estimate *environmental stochasticity* in survival for European dippers. Explain how you obtained your answer, and show your work!

## Checklist for Lab 7 completion

* Please bundle all your responses into a single Word document and submit *using WebCampus*!
* Where appropriate, Excel files should be included as part of your lab submission. See details below...

***Due Mar. 31 at 11 am.***

* Word document with short answers, and and Excel document
  + **Exercise 1**
    - *Short answer (1a.)*
    - *Short answer (1b.)*
  + **Exercise 2**
    - *Short answer (2a.)*
    - *Short answer (2b.)*
    - *Short answer (2c.)*
    - *Short answer (2d.)*
    - *Short answer (2e.)*
    - *Short answer (2f.)*