Survival of European dippers

It would help if you have use MARK before, eg. for a "closed captures" analysis to estimate population size. However, sufficient detail is given here for you to carry out the analysis.

For this lab you will need the data input file, "Dippers.inp", and you will need the MARK package installed on your computer.

A. The dipper data

The white-throated or European dipper, *Cinclus cinclus*, is a small bird which lives alongside streams in much of Europe, foraging underwater for invertebrates and small fish.

The data come from trapping dippers at the same sites for seven consecutive years. All were adults (at least 1 yr old) when first trapped, and were marked with numbered leg bands. We want to use the information on recaptures to estimate the probability of survival.

Open the file "Dippers.inp" in Notepad or similar text editor.

The first line is just a comment to confirm what data are in the file; MARK ignores anything enclosed with /* … */ characters.

The file contains capture histories for individual birds: a series of seven 1s and 0s indicating whether the bird was caught (1) or not caught (0) each year for seven years. So the first bird (1111110) was caught every year for six years but not caught in the seventh year; we don't know if the bird was still alive in year 7 and just wasn't trapped, or if it had died.

The capture history is followed by a "1", indicating that it refers to a single individual and each line of data finishes with a semi-colon (;). Birds often have identical capture histories, eg. in rows 3 to 6; we could replace these four rows with a single row: "1100000 4;" to indicate that four birds have this capture history.

Close the Notepad file.

B. Setting up the project and entering data in MARK

Start MARK and select 'File > New' from the pull-down menus. In the "Enter Specifications for MARK Analysis" window, under "Select Data Type", make sure that the first item, "Recaptures only" is selected.

Type a title for the set of data (eg "Dippers"). Then click on 'Click to select file', browse to 'Dippers.inp' and click 'Open'.

MARK enters the name of the Results file as 'Dippers.dbf', but you can change that if you wish (e.g., if you are doing different sets of analyses with the same input file).

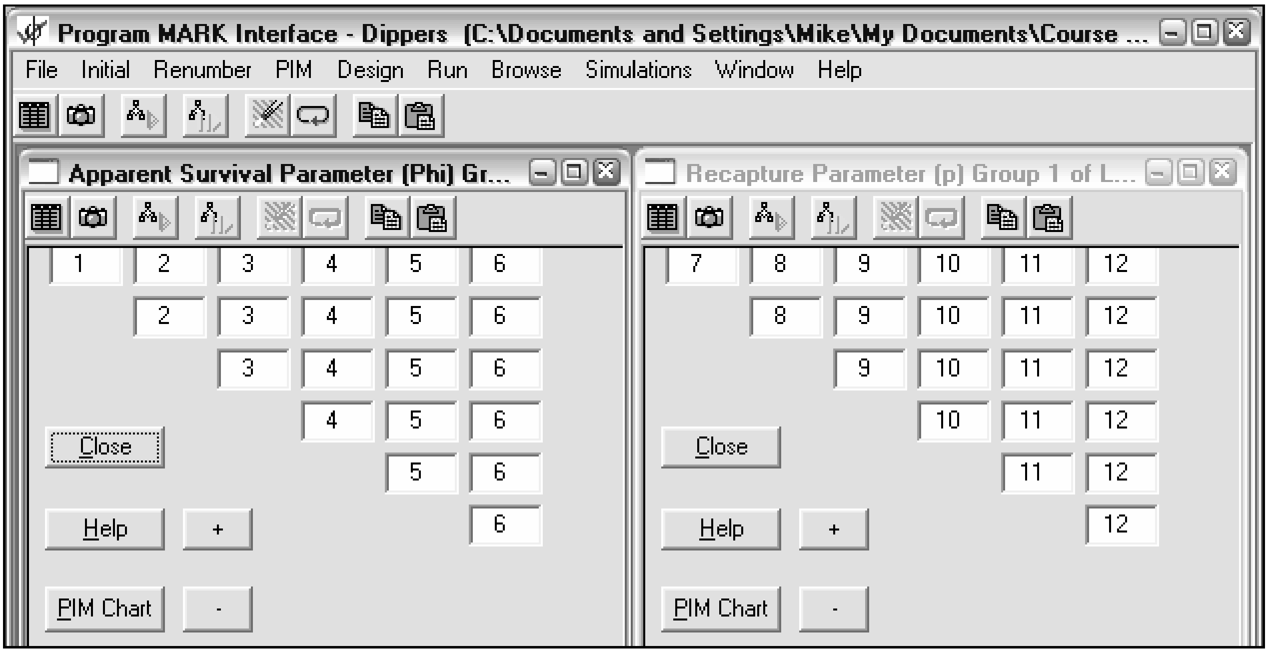
Change 'Encounter occasions' to 7, and leave 'Attribute groups' as 1. Click on OK.

A dialogue box pops up to inform you that the output file 'DIPPERS.DBF' was created. Click OK.

C. Running a first analysis in MARK

When you've entered all the specifications, a PIM (Parameter Index Matrix) window opens. There are two of these, so open both.

From the pull-down menus, select 'PIM > Open Parameter Index Matrix'. In the box that appears click 'Select All' then 'OK'. Select 'Windows > Tile' to see them side-by-side (see screen shot below).



The left window in the screen shot has six columns showing indices for six different survival parameters, corresponding to the probability of survival from year 1 to year 2, year 2 to year 3, and so on.

The top row corresponds to the birds first captured in year 1, referred to as the first *cohort*. The second row is for the second cohort, first captured in year 2; for these, we cannot estimate survival from year 1 to year 2, so this row has only five parameters. Each cohort has a row in the PIM, with fewer parameters for the later cohorts.

The right window has six columns with recapture parameters corresponding to years 2 to 7. We are not interested in the initial capture probability, and there are no recaptures for year 1, so there are only six columns in the matrix. As before, we have one row for each cohort.

The default model has a different survival probability for each interval between years (parameters 1 to 6) and a different recapture probability for each year after the first (parameters 7 to 12), but it doesn't matter when the bird was first capture, ie. they are the same for all cohorts.

Actually, it's impossible to estimate all 12 parameters: specifically, we cannot separate out the last survival parameter from the last recapture parameter. So let's begin with the minimum number of parameters, the same survival probability and the same recapture probability for all years.

Change all the Survival Parameter indices to "1": you could type "1" in each box, but a quicker way is to click on any of the boxes and select 'Initial > Constant' from the pull-down menus. Then click anywhere in the Recapture Parameters window and select 'Initial > Constant' again: all the indices change to "7", which is okay, it just has to be different to the Survival Parameter.

Now we're ready to run this simple model, which we'll call the "phi(.) p(.)" model. The usual symbol for survival is the Greek letter ϕ ("phi") and the dot inside the brackets indicates that a fixed value is used for all years.

Select 'Run > Current Model' from the pull-down menus, or click the "Run…" button – the third button on the tool bar..

In the 'Setup Numerical Estimation Run' window, give the model the name "phi(.) p(.)", leave all the other default values as they are and click 'OK to Run'.

Click 'Yes' when MARK asks if it should use an identity design matrix, and 'Yes' again when it asks if it should add the output to the database.

The Results Browser appears, which has nothing interesting yet, as we have only run one model, but you might note that No. Parameters is 2, as we intended. Let's look at the detailed output.

Highlight the phi(.) p(.) model in the Results Browser and click on the second button on the toolbar.

The first part summarizes the input, including the model name and setup details. There's a warning that "At least a pair of the encounter histories are duplicates"; we put each dipper on a separate line, so that is okay. Then there's some technical information and, near the bottom, the *real* stuff. MARK deals in probabilities which vary from 0 to 1, and these cannot be modelled directly by a linear function, so MARK uses a 'linear estimator' which is linked to the real values of ϕ and *p*; the default link in MARK is the SIN link function (PRESENCE uses the logit link by default). The SIN Link Function Parameters are then converted to real values, ie. probabilities of survival and recapture.

Real Function Parameters of {phi(.) p(.)}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:Phi	0.5658226	0.0355404	0.4953193	0.6337593
2:p	0.9231757	0.0363182	0.8149669	0.9704014

Check the output for signs that MARK has failed to calculate the results properly. Things to look for are:

- Warnings; some of these may not be serious, but do take a look at each one. Here we have a warning about duplicate capture histories, but that is not a problem.
- A standard error = 0
- A probability estimate close to 0 or close to 1, or with a huge confidence interval, from near 0 to near 1.
- A note at the bottom of the output file of any *singular values* (more on those below).

There are no signs of any problems in the output for the phi(.) p(.) model.

The estimated recapture probability is high at 0.92, but is clearly less than 1, so has to be taken into account.

The estimated survival probability is 0.57 with 95% confidence interval from 0.50 to 0.63, ie. the probability that an adult male dipper survives from one year to the next is a little better than 50-50.

Now let's see what happens with more complex (and maybe more realistic) models.

D. Comparing models

1. phi(year) p(.)

Survival is likely to vary from year to year, in particular reflecting the severity of the winter: if streams are largely frozen over, the dippers have a problem getting food. Let's try a model with a different survival probability for each year, but the same capture probability: this is model phi(year) p(.).

Use Windows > Tile to show the two PIM windows again. (If they have been closed, use 'PIM > Open Parameter Index Matrices' to get them back.)

Notice that MARK did some tidying up before running the previous model and renumbered the Recapture Parameter to "2".

Right-click in the Survival Parameter window and select Time from the context menu. Parameters 1 to 6 now correspond to the (different) survival probabilities for the six time intervals between surveys.

In the Recapture Parameter window, click on the "+" button until 7 appears in all the boxes: parameter 7 corresponds to the (same) recapture probability for all surveys.

Run the analysis as before, naming the model "phi(year) p(.)".

The new model appears in the browser below phi(.) p(.) as it has a higher value for AICc. The difference (DeltaAICc) is 7.5, which means that the phi(.) p(.) model is better and the five extra parameters in the phi(year) p(.) model don't improve the fit sufficiently to be worthwhile.

Look at the detailed output for the new model.

> As before, check the output for signs of problems.
>
> There are no signs of any problems with the phi(year) p(.) model.

The survival probabilities are different from year to year, but all the estimates fall within the 95% confidence interval for phi in the phi(.) p(.) model (viz. 0.49 to 0.63). From the hypothesis testing point of view, all 6 values are consistent with the simpler model, and we have no evidence of variation from year to year.

2. phi(.) p(year)

It's possible that recapture probability may vary from year to year. Even if the same techniques are used by the same people, the behaviour of the birds may vary depending on factors such as the weather.

Go back to the two PIM windows again. This time make the Survival Parameter constant (ie. the same index in all the boxes), and make the Recapture Parameter different for each column (select Time from the context menu). Make sure that the same parameter index doesn't appear in both windows. Run the analysis as before, naming the model "phi(.) p(year)".

> As before, check the detailed output for signs of problems.
>
> Notice that the 2nd and last values for *p* are exactly 1, with a confidence interval from 1 to 1. Are these values plausible? Are they useful?

Could the recapture probability in year 3 really be 1? Look at the original data in "Dippers.inp". The first 32 rows are birds caught and marked before year 3. Some of these were caught again in year 3, some were not. If a marked bird was not caught in year 3 but was caught later, we know it was still alive in year 3 and recapture probability is less than 1. Of the first 32 birds, 20 were not recaptured in year 3, and *none of these 20 was ever seen later*. So, yes, it is possible that all 20 were dead by year 3 and recapture probability really is 1! The same logic applies to *p* for the final year too.

In theory you should be able to estimate separate recapture probabilities for each year except the last, but sometimes you get a quirky data set like this one which doesn't allow you to do it properly, so always check the detailed output.

But, from a practical point of view, it's hardly plausible that *p* really was 1! Fortunately this model turns out to be rather poor (DeltaAICc = 8.13) so we don't need to worry about it too much.

3. phi(year) p(year)

Just for completeness, we'll run the full model, phi(year) p(year), though we anticipate problems estimating two of the parameters for *p*.

Make both the Survival Parameter and the Recapture Parameter different for each column (select Time from the context menu). Make sure that the same parameter index doesn't appear in both windows: it should look like the screen shot on page 2.

This model appears at the bottom of the list in the Results Browser, with DeltaAICc = 13.89. Notice also that the number of parameters is 11, although when we set up the model we specified 12 different parameters. Check the detailed output:

As before, check the detailed output for signs of problems. Notice that:

- the 2nd value for *p* is exactly 1, as before.
- the last Phi value and the last *p* value both have confidence intervals from 0 to 1; moreover the point estimates are the same.
- At the bottom of the file is the note "Beta number 12 is a singular value" (you may have a different number instead of 12).

The note about a "singular value" tells us that a particular parameter cannot be estimated separately: in this case it is the recapture parameter for the final year, which cannot be separated from survival probability. Recapture could be high and survival low, or vice versa, hence the confidence intervals from 0 to 1. MARK estimates phi x p for the last year, hence only 11 parameters have actually been estimated, not 12. The reported the values for the final phi and p are just the square root of phi x p.

The unfortunate consequence is that the value shown for the last Phi, 0.764, is not really a proper estimate of phi. For example, although it is higher than all the other values, it does *not* mean that survival is higher during the last interval than the others.

Now let's look at the summary of the models in the Results Browser:

Results Browser: Live Recaptures (CJS)

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance	-2Log(L)
{phi(.) p(.)}	322.5527	0.0000	0.96003	1.0000	2	41.8147	318.4938
{phi(year) p(.)}	330.0567	7.5040	0.02253	0.0235	7	38.8147	315.4939
{phi(.) p(year)}	330.6794	8.1267	0.01650	0.0172	7	39.4374	316.1166
{phi(year) p(year)}	336.4343	13.8816	0.00093	0.0010	11	36.4013	313.0805

The best model in the set we tested is clearly the simplest model, phi(.) p(.); models where survival probability or recapture probability vary from year to year have very little support from the data. Accordingly, our estimate of annual survival probability for adult dippers is 0.57 (95% confidence interval 0.50 to 0.63).

Conclusion

This has been a first run through the analysis of the dipper data in MARK. We have left out a couple of important topics, such as overdispersion, which you would need to consider is before writing up a paper or a thesis. These are covered in much more detail in the online book by Evan Cooch and Gary White (2009), which you should work through if you are going to use MARK for serious analysis.

Exercise

One hypothesis that the researchers were interested in testing concerned the effect of floods during the breeding season on survival probabilities. Floods occurred during the 2^{nd} and 3^{rd} intervals between surveys. Set up a model which has different survival probabilities for ‘flood’ years and ‘no flood’ years, ie. phi(flood) p(.). How does this model compare with the others? Is there evidence that survival is different when there is a flood during the breeding season?

Literature cited and further reading

Amstrup, S. C., T. L. McDonald, and B. F. J. Manly (eds). 2005. Handbook of capture-recapture analysis. Princeton University Press, Princeton, NJ.

Cooch, E., and G. White 2009. Program MARK: a gentle introduction. Available in .pdf format for free download at http://www.phidot.org/software/mark/docs/book/