MA888 Ecological Statistics

`Eagle`: Ring-recovery models and score test model selection

R.S.McCrea and B.J.T.Morgan

Spring Term 2012

1 Introduction

The package Eagle was written by Ted Catchpole. It is written in MATLAB, however, it is conversational, and so you do not need to know anything about MATLAB in order to run the package. The package will be used to fit models using maximum likelihood, and later we shall use it to consider score and likelihood ratio tests for comparing different, nested models.

There are 4 real data sets available: `mallard_dat.m`, `gulls_dat.m`, `teal_dat.m`, `heron_med.m` and further simulated data sets available for experimentation: `wc_dat.m`, `wc2_dat.m` and `test_dat.m`.

You will be given the option of generating two possible output files. If you have used the data set `data1.m`, then the two files will be called `data1.res`, and `data1.mat`. The latter is needed if at a future `Eagle` session you want to continue the current analysis. The former is simply a diary of everything you have done in the current session. These output files will appear in your own file-space on the computer.

The output of current interest is the parameter estimates, as well as the estimated standard errors. Also of interest is the table of fitted values, and the table showing the residuals.

2 Gulls Example

1. Download the folder `Eagle.zip` from the MA888 module on moodle. Extract the zipped file.

2. Start MATLAB by accessing it through central software on the student computers.

3. We now need to change the directory to the location of your newly downloaded Eagle folder. This can be done at the top of the right hand window of Matlab 2011a.

4. To start `Eagle`, type `eagle` in the command window next to the >> sign.

5. Select `1` for a new session.

6. When instructed, change the default directory to the directory you have just navigated to (this can be done by copying the contents of the directory box and pasting into the command window).

7. Select the `gulls_dat` data set and accept the default storage `gulls_dat.res` for the results.

8. Recall the Catchpole-Freeman-Morgan model notation is $x/y/z$ where:

   - $x = c$ if first year survival probabilities are constant;
   - $x = t$ if they vary with time;

   - $y = c$ if adult survival probabilities are constant;
   - $y = a$ if they are fully age dependent;
   - $y = ar$ if they are age-dependent up to year r;

   - $z = c$ if recovery probabilities are constant;
   - $z = t$ if they vary with time;

9. Fit the constant model $c/c/c$.

10. The first year survival probability estimate is 0.529 with associated standard error 0.0223.

11. What do you observe about the size of the recovery probability?

12. The goodness-of-fit statistic is given in the output. Is this an appropriate model for the data?

13. Now suppose we wish to undertake model selection. In order to perform a score test you need to have fitted the null model to the data. Given we have already fitted the $c/c/c$ model, try performing score tests of the alternative models:

    (a) $t/c/c$

    (b) $c/a2/c$

    (c) $c/c/t$

14. Which has resulted in the most significant test statistic?

15. Remembering to fit the null model first, perform a score test of the $t/c/c$ model versus the $t/c/t$ model.

16. Perform a score test of the $t/c/c$ model versus the $t/a2/c$ model.

17. Out of all of the models you have considered, which is the most appropriate for the gull data set?

18. Another approach which can be used involves using likelihood-ratio tests instead of score tests. Fit the $c/c/c$ model again. You will notice that the output shows: `The maximized loglikelihood is -83.9410`.

19. Now fit model $t/c/c$. By looking at the difference in the maximized log likelihood values, calculate the likelihood-ratio statistic[1]. Is it close in value to the score statistic for testing these two models?

20. Look at the `AIC` values for the models, filling in the following table. Based on AIC, which is the best model for this data set?

Model	np	AIC
$c/c/c$		
$t/c/c$		
$c/a2/c$		
$c/c/t$		

[1]Recall that the likelihood ratio statistic is given by

$$2 \log \frac{L_2}{L_1}$$

MA888: Ecological Statistics Assessment 2

Ring recovery models and score test model selection: Eagle

Please hand in your assessment to the SMSAS general office
by 12.00 *on* **Wednesday 21st March 2012**.

Question 1: Fitting models to the cormorant data

(a) Select the `corms_dat` data set and type `corms` to store your output in the file `corms.res`.

(b) Fit the constant model $c/c/c$.

(c) What do you notice when you compare the first year and adult survival probabilities?

(d) Look at the correlation structure of the parameters of the model. Are the parameters highly correlated?

(e) The AIC and chi-squared goodness-of-fit statistics are provided in the program output. Does the model fit the data? Which of these values would enable you to compare relative fit of models?

(f) Make a note of the AIC value for the constant model.

(g) What do you observe about the size of the recovery probability? Is the probability of a dead bird being reported high?

(h) Look at the matrix of fitted values and the residuals. Is there any pattern in the residuals?

Question 2: Score test model selection

(a) Now suppose we wish to undertake model selection. In order to perform a score test you need to have fitted the null model to the data. Given we have already fitted the $c/c/c$ model, try performing score tests of the alternative models:

(i) $t/c/c$

(ii) $c/a2/c$

(iii) $c/c/t$

(b) Which has resulted in the most significant test statistic? Which model (if any) do you select as the model under the new null hypothesis? What is your new set of alternative models?

(c) Remembering to fit the null model first, continue the score test model selection procedure.

(d) If we have selected a 5% significance level stopping rule, what is the model we select using this model selection procedure?

(e) Out of all of the models you have considered, which is the most appropriate for the cormorant data set? Does your selected model fit the data? How does the AIC value of your selected model compare with the AIC of the constant model?

Question 3: Use of Winter NAO Covariate

(a) The implemented score test model selection procedure should have identified the model $t/c/t$ as the best of the models proposed for this data set.

(b) We now want to see if the temporal variation in survival can be explained by the winter NAO index covariate.

(c) We now want to fit a model $v(NAO)/c/t$ to the data. What is the AIC value of this model? How does this compare with the AIC of the model $t/c/t$?

Question 4: Comparison with Likelihood-ratio tests

(a) Another approach which can be used involves using likelihood-ratio tests instead of score tests. Fit the $c/c/c$ model again. You will notice that the output shows: `The maximized loglikelihood is -256.0080`.

(b) Now fit model $c/c/t$. By looking at the difference in the maximized log likelihood values, calculate the likelihood-ratio statistic. Is it close in value to the score statistic for testing these two models?

(c) Now fit model $t/c/t$. Perform a likelihood ratio test of $c/c/t$ against $t/c/t$. Is it close in value to the score statistic for testing these two models?

(d) Now fit model $t/a2/t$. Perform a likelihood ratio test of $t/c/t$ against $t/a2/t$. Is it close in value to the score statistic for testing these two models?