

# MA888 Ecological Statistics

## **EstimateN**: Estimating Population Size

M.S.Ridout

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

Spring Term 2012

## 1 Introduction

**EstimateN** was written by M.S.Ridout and is software for fitting capture heterogeneity models to estimate the size of a closed population. The models are described in two papers:

Morgan, B.J.T. and Ridout, M.S. (2008) A new mixture model for capture heterogeneity. *Applied Statistics*, **57** (4), 433-446.

Morgan, B.J.T. and Ridout, M.S. (2008) Estimating N: a robust approach to capture heterogeneity. In *Modeling demographic processes in marked populations*. Thomson, D.L., Cooch, E.G., Conroy, M.J. (Eds.). Springer Series: Environmental and Ecological Statistics. *To appear*.

The software allows four different models to be fitted:

**bin** - no heterogeneity, captures follow a binomial distribution;

**betabin** - captures follow a beta-binomial distribution;

**twobin** - captures follow a mixture of two binomial distributions;

**binbbin** - captures follow a mixture of a binomial and a beta-binomial.

A full user guide (on which this worksheet is based) can be found at <http://www.kent.ac.uk/ims/personal/msr/estimateN.html>. We use data on Cormorants to demonstrate the fitting of these four models.

## 2 Cormorants

During 1994 between period March and October biologists visited a colony Vorskø in the Horesns Fjord in Denmark. Resightings of ringed individuals were made. The data contain

information on when each individual bird was seen. From this, we are able to compile a Schnabel census for each month of data. The data are given in the text document `cormdata` which can be found in the `EstimateN` directory and a printed copy is attached to the back of this worksheet.

From these data, we wish to estimate the size of the population. The population size we are estimating will be the number of ringed individuals within the resighting region of the colony, and therefore will not be an estimate of the total population of the colony. The data is further divided into successful breeders and non-successful breeders.

### 3 Accessing the software

The `EstimateN.r` function can be found on the MA888 module on moodle. Download the zipped file `EstimateN.zip` and extract the file to a folder on your workspace. Open R and change directory (File-Change dir...) to the folder where you extracted your file. Commands to be typed into R are marked by a `>` sign.

### 4 The function `fitallmodels`

The specification of this function is

```
fitallmodels<-function(freq,kmax)
```

where `freq` is an array such that `freq[k]` is the number of animals seen on `k` occasions and `kmax` is the total number of sampling occasions.

We demonstrate the `fitallmodels` function on the March cormorant data:

1. Load the `EstimateN` functions:

```
> source("estimateN.r")
```

2. Enter the cormorant data (a text file containing the monthly cormorant data is in the `EstimateN` folder):

```
> MarSB = c(13,24,14,6,11,10,6,5,5,4,2,5,3,3,2,1,0,2,1,0,1)
```

3. Enter the number of sampling occasions. Typically `freq` will be of length `kmax`, but zeros at the end of the array may be omitted.

```
> kMar = 29
```

4. Call the `fitallmodels` function:

```
> fitallmodels(MarSB,kMarSB)
```

5. Study the output of the four models. How does the number of parameters differ between the models?
6. What do you note about the output from the binomial model?
7. Given the output provided, how might you decide which is the best model for the data?

## 5 The function `fitonemodel`

The specification of this function is rather more involved because it provides several options. However, most options have default settings and the function remains quite easy to use. The full specification is:

```
fitonemodel<-function(freq,kmax,model="bin",initparam=NULL,
  nrandstart=100,plotprofile=FALSE,logn0values=NULL,
  confpercent=95,nrandprof=10)
```

Unless you are an advanced user of the software, it is best to leave `initparam` unset, and in this case the function works from a number of random starting points, specified by the argument `nrandstart`. The model is fitted from each possible starting value, using the Nelder-Mead simplex algorithm, and the best fit overall is reported.

It is also possible to plot the profile-log-likelihood for  $n_0$  (the number of unseen animals) by setting `plotprofile=TRUE`. Argument `confpercent` determines the confidence level of the interval provided for  $N$ . Finally, the argument `nrandprof` relates to the number of starting points for the profile likelihood optimisations.

For our purposes there is little need to define all of these arguments.

### 5.1 Return to the Cormorant Data

1. From the previous model fitting you will have noticed that the estimate of the number of unseen individuals is zero for the binomial model. Fit a two binomial mixture model to the same March data using 25 random starting points for the optimisation:

```
> fitonemodel(MarSB,kMar,model="twobin",nrandstart=25)
```

2. Now fit the same model, but increase to 100 random starting points for the optimisation. Has this made a difference?
3. Using appropriate argument values, fit the beta-binomial and the binomial and beta-binomial mixture model using the `fitonemodel` command.
4. Suppose we now want to obtain a profile plot of  $n_0$  under the two binomial mixture model. To achieve this type:

```
> fitonemodel(MarSB,kMar,model="twobin",nrandstart=25,  
  plotprofile=TRUE,nrandprof=10)
```

Note that this will take longer to run, as constructing the profile likelihood requires multiple optimisations.

5. By entering appropriate arguments, construct profile plots under the beta-binomial model.

# MA888: Ecological Statistics Assessment 1

## Estimating population size: EstimateN

*Please hand in your assessment to the SMSAS general office  
by 12.00 on Friday 9th March 2012.*

### Question 1: Cormorants

Now that you have been shown how to use the `fitallmodels` and `fitonemodel` commands, we would like you to analyse the monthly data for both the successful and unsuccessful breeders (SB and UB respectively).

The number of visits conducted per month (from March to October) are [29, 30, 31, 30, 31, 28, 28, 19] and the recapture data are summarised by the Schnabel census by month:

- $\text{MarSB} = c(13, 24, 14, 6, 11, 10, 6, 5, 5, 4, 2, 5, 3, 3, 2, 1, 0, 2, 1, 0, 1)$
- $\text{AprSB} = c(13, 14, 10, 8, 11, 7, 7, 12, 7, 9, 6, 10, 7, 2, 0, 3, 1, 0, 0, 0, 1)$
- $\text{MaySB} = c(17, 22, 15, 12, 8, 7, 7, 10, 9, 2, 1, 4, 3, 1, 1, 0, 1, 1)$
- $\text{JunSB} = c(28, 21, 12, 12, 8, 6, 6, 6, 8, 7, 7, 6, 5, 4, 2, 3, 0, 1, 0, 0, 1)$
- $\text{JulSB} = c(13, 11, 12, 9, 1, 8, 6, 3, 6, 6, 8, 3, 1, 0, 1, 2)$
- $\text{AugSB} = c(6, 5, 3, 4, 5, 3, 2, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1)$
- $\text{SepSB} = c(1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1)$
- $\text{OctSB} = c(1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1)$
- $\text{MarUB} = c(5, 4, 5, 1, 1, 0, 2, 2, 3, 3, 1, 0, 0, 0, 1, 0, 1, 0, 1)$
- $\text{AprUB} = c(1, 4, 5, 6, 4, 5, 5, 4, 4, 2, 2, 1, 1, 2, 2, 1)$
- $\text{MayUB} = c(4, 12, 6, 6, 1, 1, 5, 3, 2, 3, 0, 0, 0, 1, 0, 2)$
- $\text{JunUB} = c(7, 8, 3, 2, 2, 1, 0, 0, 0, 0, 2)$
- $\text{JulUB} = c(2, 1, 1, 1, 1, 2, 1, 0, 1)$
- $\text{AugUB} = c(2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1)$

- (a) Construct a table of population size estimates by month for successful and unsuccessful breeders. Remember to include confidence intervals with your parameter estimates.

- (b) Discuss which model is most appropriate for each of the data sets using the minimised negative log-likelihoods of the four models.
- (c) For each of the data sets does the the model fit the data?
- (d) What can you conclude about the proportions of successful and unsuccessful breeders within this colony?
- (e) Briefly discuss the time-variation exhibited by the parameter estimates.

## Question 2: Other data sets

For each of the following data sets, fit the four models to the data sets and examine the differences in the estimates of total population size. Also, decide which is the most appropriate model for each of the data sets by (i) examining the observed frequencies and fitted values from the model and (ii) comparing the minimised negative log-likelihood values.

`housemice = c(2, 64, 40, 31, 16, 13, 5, 1, 0, 1), k = 10;`

`pocketmice = c(16, 15, 6, 5, 5, 5, 3), k = 7;`

`woodmice = c(71, 59, 41, 39, 20, 26, 19, 12, 9, 5, 8, 4, 9, 2, 1, 3, 3, 3), k = 21;`

`voles1 = c(29, 15, 16, 27), k = 5;`

`voles2 = c(18, 15, 8, 6, 6), k = 5;`