FISH 621 Laboratory #5: Bayesian Mark-Recapture

Curry Cunningham 2022

Instructions

The purpose of this lab is to:

- Explore the Jolly-Seber mark-recapture model for open populations.
- Develop familiarity with implementing simple Bayesian analyses within the Stan platform.

If you have a question during the lab, please un-mute yourself and ask, or type it into the chat box. There is a high likelihood that someone else has the same question. It is more fun if we all learn together in our distance-learning world.

I have posted the lecture slides to the ***Canvas site***, so you can reference this material as you work through the lab.

This and all other labs will be graded based on your attendance and participation.

Lab Contents

621_Lab 5_Bayes.pdf	(this file)
621_Lab 5_Bayes.R	R script with exercises.
Jolly Seber.csv	Example of Jolly-Seber analysis of Large-mouth bass at Par Pond from Hightower and Gilbert (1984)
Bristol Bay Spawner-Recruit Data.csv	Spawner-recruit data for Bristol Bay Alaska sockeye salmon

Exercise 1: Jolly-Seber

We will explore implementation of the Jolly-Seber model for a dataset of large-mouth bass at Par Pond from Hightower and Gilbert (1984). The sampling periods $i\ in\ 1{:}s = 6$ are subsequent weeks.The available data include:

- The number of bass captured during each sampling period: n_i
- The total number of sampled individuals returned to the population during each sampling event, with tags: R_i
- The matrix describing the number of individuals from each sample release that are captured during subsequent sampling events: m_{hi}. Recall in this type of experiment we mark individuals in

such a way that upon recapture we can tell in which sampling period they were originally marked (e.g. using differently colored tags during each sampling event).

- The rows h reference the release period (release cohorts)
- The columns i are the recovery periods
- The elements of the m_{hi} matrix are the number of fish from release period h that are recaptured during period i

Please open the spreadsheet called ***Jolly Seber.xlsx***. Green cells are data, and differently colored cells represent different derived parameters of the Jolly-Seber model. Please use the following steps to generate estimates:

1. Calculate m_i the total number of marked individuals observed during each sampling event (Cells B11:G11) for periods $i = 1:6$, as: $m_i = \sum_{h=1}^{i-1} m_{hi}$.
2. Calculate r_h, the number of R_i releases that are later recaptured (Cells I5:I9), for release groups $h = 1:5$, as: $r_h = \sum_{i=h+1}^{s} m_{hi}$.
3. Calculate the c_{hi} matrix (Cells B15:G19) where:
    a. The first row of c_{hi} is equal to the first row of m_{hi}, or $c_{1i} = m_{1i}$
    b. Remaining rows for each recapture period i are the sum of captures of all earlier release groups h, or $c_{hi} = c_{h-1,i} + m_{hi}$
4. Calculate the number of individuals for each release group h that before period i, that were not captured in period i, and are captured after period i, or z_{h+1}, based on the c_{hi} matrix.
    a. Where z_{h+1} is calculated by summing each row of the c_{hi} matrix, as: $z_{h+1} = \sum_{i=h+2}^{s} c_{hi}$
5. Next, we will calculate summary statistics in ***Cells B24:O29***
6. Begin by copying summary values calculated from your m_{hi} and c_{hi} matrices into the appropriate columns of the ***Summary Statistics*** table.
    a. Copy the total marks observed in each sampling period m_i from Cells B11:G11, into Cells D24:D29.
    b. Copy the total number of releases that are later recaptured r_h from Cells I5:I9, into Cells E24:E28.
    c. Copy z_{h+1} from Cells I15:I18 into Cells F25:F28.
7. Calculate the mark fraction in each sampling event $\rho_i = m_i/n_i$, in Cells G24:G29.
8. Calculate the ***unbiased*** estimate for the total number of marks $M_i^* = \frac{(R_i+1)}{(r_i+1)}(z_i) + m_i$ for $i = 2:5$, in Cells H25:H28.
9. Calculate the ***unbiased*** estimate for the total number of unmarked individuals in the population $U_i^* = \frac{M_i^*(n_i+2)}{m_i+1} - M_i^*$ for $i = 2:5$, in Cells I25:I28.
10. Calculate $M_i^* - m_i$ for $i = 2:5$, in Cells J25:J28.
11. Calculate $M_i^* - m_i + R_i$ for $i = 2:5$, in Cells K25:K28.
12. Calculate $(1/r_i) - (1/R_i)$ for $i = 1:5$, in Cells L24:L28.
13. Calculate the ***potentially biased*** estimate of the total number of marked individuals in the population $\widehat{M}_i = \frac{R_i z_i}{r_i} + m_i$ for $i = 2:5$, in Cells M25:M28.
14. Calculate $\widehat{M}_i - m_i$ for $i = 2:5$, in Cells N25:N28.
15. Finally, calculate $\widehat{M}_i - m_i + R_i$ for $i = 2:5$, in cells O25:O28.
16. Now we have all of the pieces to calculate estimates for survival. Under the Population Estimates section please calculate

   a. For survival during the first sampling period use the approximation: $\phi^*_{i=1} = M^*_{i+1}/R_i$ or $\phi^*_2 = M^*_2/R_1$in Cell F33.
   b. For subsequent sampling periods $i = 2:(s-2) = 2:4$, calculate survival as $\phi^*_i = \frac{M^*_{i+1}}{\widehat{M}_i \ -m_i+R_i}$

17. Next, we can calculate our total population size estimates N^*_i
    a. For periods $i = 2:5$ calculate $M^*_i + U^*_i = \frac{M^*_i(n_i+2)}{m_i+1}$, in Cells B34:B37.
    b. Set cells in the N^*_i column (B34:B37) for periods $i = 2:5$, equal to the $M^*_i + U^*_i$ column (Cells C34:C37).
    c. To estimate total abundance at the start ($N^*_{i=1}$ in Cell C33) of period $i = 1$, we will use $N^*_{i=2}$ and the estimated survival rate ϕ^*_i, as: $N^*_{i=1} = \frac{N^*_{i=2}}{\phi^*_{i=1}}$
    d. ***Congratulations! You have estimated abundance at the start of each sampling period!***
18. Now that we have point estimates for abundance, we will focus on uncertainty in our estimates.
    a. Remember an estimate isn't all that useful if we don't have a sense of its precision.
19. First calculate the uncaptured number of individuals in the population at each time point $N^*_i - n_i$ in Cells D33:D37.
20. We will first calculate the standard error in our estimate of N^*_i in Cells E33:E37, as
    a. $SE(N^*_i) = \sqrt{N^*_i(N^*_i - n_i)\left(\left(\frac{M^*_i - m_i + R_i}{M^*_i}\right)\left(\frac{1}{r_i} - \frac{1}{R_i}\right) + \left(\frac{1-\rho_i}{m_i}\right)\right)}$ For periods $i = 2:5$ in Cells E34:E37.
21. Calculate the coefficient of variation for the abundance estimates N^*_i as: $CV(N^*_i) = SE(N^*_i)/N^*_i$.
22. Next, we will calculate our estimate of recruitment B^*_i for periods $i = 2:4$ in Cells I35:I37, using the formula $B^*_i = N^*_{i+1} - \phi^*_i(N^*_i - n_i + R_i)$.

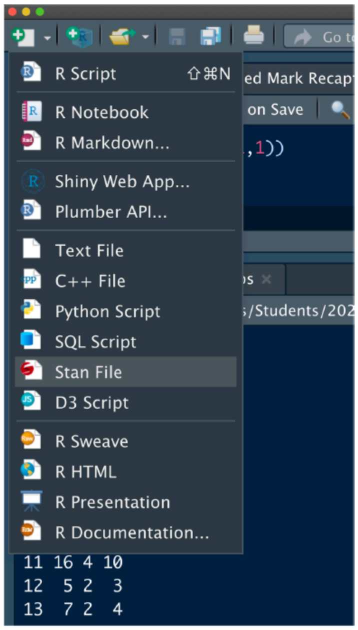
Exercise 3: Bayesian Linear Regression

To familiarize ourselves with how we simulate data with R, and define and fit a Bayesian model with Stan we will start with a simple linear regression: $y_i = \alpha + \beta x_i + e_i$, where our observation errors are normally-distributed: $e_i \sim Normal(0, \sigma^2)$ with standard deviation σ.

Please follow through the R script as we simulate our data, and then open the ***lin_reg.stan*** script to see how we encode the Bayesian regression model using our Stan syntax.

The .stan Script

- data
  - Define data inputs
    - Including dimensionality
- parameters
  - Define “free” or estimated (true) parameters
    - Names, dimensions, ect.
- transformed parameters
  - Define derived parameters
    - Quantities that depend on your estimated (true) parameters
  - Do calculations
  - ***This is where the meat of your code is likely to exist!***
- model
  - Define priors for estimated parameters
  - Define likelihoods for the data
    - Probability of the data, given the model
- generated quantities
  - Calculations you want to do based on your
    - Estimated or derived parameters and data



```
11 16  4 10
12  5  2  3
13  7  2  4
```

Stan Data Types

- Primitive types
  - real
    - Continuous values: 1.4, 0.9, -99.1, 100, ect.
  - int
    - Integer values: 1,2,3, ect.
- Vector and matrix types
  - Matrix-based types
    - `vector`, `matrix`, and `row_vector`
  - Examples
    - `vector[3] myVect` – vector of length 3, named “***myVect***”
    - `matrix[3,3] myMat` – matrix with 3 rows, 3 columns named “***myMat***”

I don’t use these often

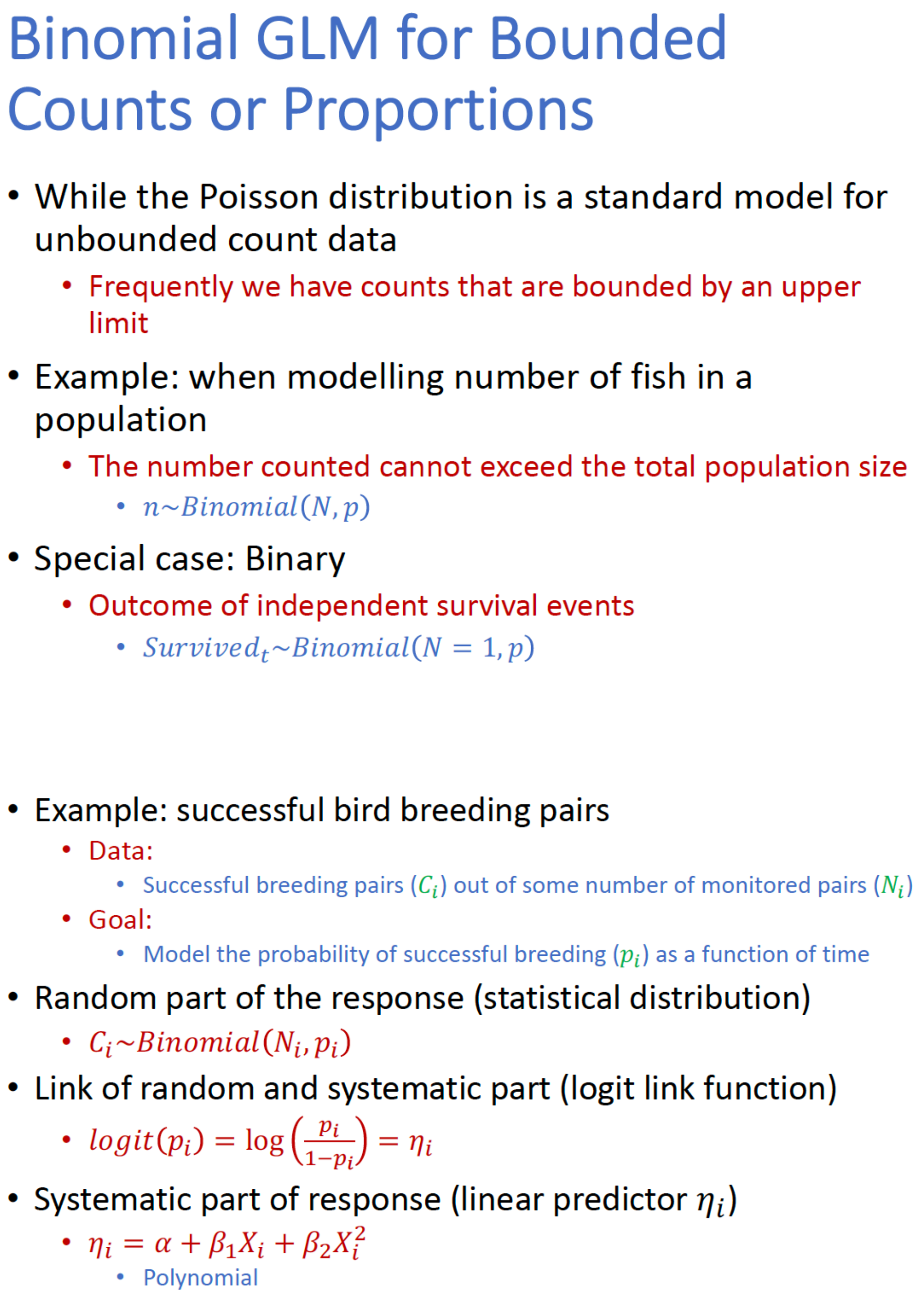
- Array types
  - Any data type can be made into an array type
  - `array[10] real x;`
    - One-dimensional array of size 10 containing real values
  - `array[6,7] matrix[3,3] m;`
    - Declares “m” to be a two-dimensional array of size 6 x 7
    - Containing values that are ***each*** 3 x 3 matrices
- Alternative declarations
  - `real x[10];`
  - `matrix[3,3] m[6,7];`
- A vector of vectors
  - `vector[N] pred[S];`
    - A vector of length S where each element is a vector of length N
    - Accessed like a matrix or 2d-array: `pred[s,n]`

Exercise 4: Poisson Regression

R script

Poisson Regression in Stan

- Estimate counts of peregrine falcons over n years
  - Linear predictor is a cubic polynomial
  - Random part of the response (statistical distribution)
    - $C_i \sim Poisson(\lambda_i)$
  - Link function of random and systematic part (log link)
    - $\log(\lambda_i) = \eta_i$
  - The systematic part of the response (linear predictor of η_i)
    - $\eta_i = \alpha + \beta_1 year_i + \beta_2 year_i^2 + \beta_3 year_i^3$



Binomial GLM for Bounded Counts or Proportions

- While the Poisson distribution is a standard model for unbounded count data
  - Frequently we have counts that are bounded by an upper limit
- Example: when modelling number of fish in a population
  - The number counted cannot exceed the total population size
    - $n \sim Binomial(N, p)$
- Special case: Binary
  - Outcome of independent survival events
    - $Survived_t \sim Binomial(N = 1, p)$

- Example: successful bird breeding pairs
  - Data:
    - Successful breeding pairs (C_i) out of some number of monitored pairs (N_i)
  - Goal:
    - Model the probability of successful breeding (p_i) as a function of time
- Random part of the response (statistical distribution)
  - $C_i \sim Binomial(N_i, p_i)$
- Link of random and systematic part (logit link function)
  - $logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \eta_i$
- Systematic part of response (linear predictor η_i)
  - $\eta_i = \alpha + \beta_1 X_i + \beta_2 X_i^2$
    - Polynomial