

Mark-Recapture The Bayesian Way

Where We Are

- We know lots about the ideas behind Bayesian analysis
 - estimating parameters
 - marginalisation
 - hierarchical models
 - missing data
 - prediction
- But, what do we do with them?

Mark-Recapture

- Originally used to estimate the sizes of populations
- Based on releasing and recapturing individuals
- Had a problem that animals might die between sampling times
 - needed to correct the estimates for this
- Recent developments have been on estimating survival

Lincoln-Peterson Model

- The Simplest Example
- 'Catch individuals at two times'
 - e.g. running around with butterfly nets
- Want to estimate the total population size
- Assumptions
 - Catchability of individuals constant
 - but catch effort can vary
 - closed population
 - no migration, births or deaths

The Symbols

- First sample Time
 - N_{mark} : number caught and marked
 - N_{unc} : number not caught
- Second Sample Time
 - N_{rec} : number recaptured
 - number captured with marks
 - N_{cap} : number captured without marks
- N_{tot} : total population size
 - what we are interested in
 - $N_{tot} = N_{mark} + N_{unc}$

The Sums

- The proportion of marked individuals caught at the second sample period, p , is

$$p = \frac{N_{rec}}{N_{cap} + N_{rec}}$$

- This is an estimate of the probability of one individual having being caught at the first time.

- Hence

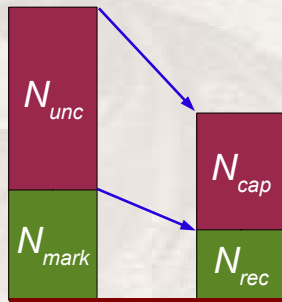
$$p = \frac{N_{mark}}{N_{mark} + N_{unc}}$$

The Estimate in Pictures

- A simple estimate of N_{tot} can be calculated as

$$\frac{N_{rec}}{N_{cap} + N_{rec}} = p = \frac{N_{mark}}{N_{tot}}$$

$$N_{tot} = \frac{N_{mark}(N_{cap} + N_{rec})}{N_{rec}}$$



The Estimate

- The estimate is a simple point estimate
- But how do we estimate the uncertainty?
- We need to develop a parametric model
 - i.e. using probability distributions
- This will give us a likelihood
- Estimate parameters
 - either by maximum likelihood or
 - Bayesian inference

Some Useful Theory

- We try and capture butterflies in a field
- Run around for time T (e.g. one hour)
- At each point in time during our sampling time we have a constant probability (q) of capturing a butterfly
- We count the number of butterflies caught
- The number will follow a Poisson distribution
 - mean $\lambda = qT$

Another Poisson

- Now imagine we only want to capture marked butterflies
- Proportion marked = p
- Probability of capture per unit time = pq
- Number caught is another Poisson distribution
 - mean: pqT

Binomial?

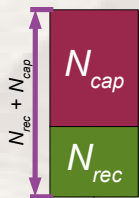
- The total number of butterflies captured follows a Poisson distribution
- The number of mark butterflies captured follows a Poisson distribution
- Obviously the marked number cannot be greater than the total number
- What is the distribution of marked butterflies, given the total number of butterflies
 - Binomial!

Useful Theory

- If $X \sim \text{Po}(\lambda_1)$ and $Y \sim \text{Po}(\lambda_2)$ then
- $X | X+Y \sim \text{Bin}(X+Y, \lambda_1/(\lambda_1+\lambda_2))$
- Useful because it unites logistic regression and log-linear models
 - chi-squared tests are based on Poisson distributions
- Useful here as a way of estimating the sample size of a Binomial distribution

MRR Sampling Theory

- From the second sample, the distribution of the number of marked individuals captured given the total number is



- $N_{rec} | N_{rec} + N_{cap} \sim \text{Bin}(N_{rec} + N_{cap}, p)$
- or $N_{rec} \sim \text{Po}(\lambda_{rec})$ and $N_{cap} \sim \text{Po}(\lambda_{cap})$
- where $\lambda_{rec} = p(\lambda_{rec} + \lambda_{cap})$

MRR Sampling Theory

- From the first sample

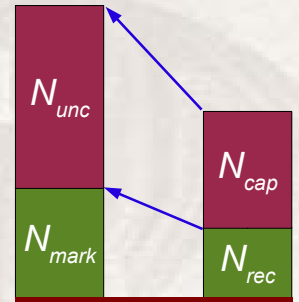
- $N_{mark} | N_{tot} \sim \text{Bin}(N_{tot}, p)$

- or $N_{mark} \sim \text{Po}(\lambda_{mark})$ where $\lambda_{rec} = p(\lambda_{mark} + \lambda_{unc})$

- So, $\frac{\lambda_{rec}}{\lambda_{cap} + \lambda_{rec}} = p = \frac{\lambda_{mark}}{\lambda_{mark} + \lambda_{unc}}$

- Hence

$$\lambda_{unc} = \lambda_{mark} \left(\frac{\lambda_{cap} + \lambda_{rec}}{\lambda_{rec}} - 1 \right)$$



The Estimate

- $N_{tot} = N_{unc} + N_{mark}$
- We know N_{mark} , and we now know that
- $N_{unc} \sim \text{Po}(\lambda_{unc})$ where $\lambda_{unc} = \lambda_{mark} \left(\frac{\lambda_{cap} + \lambda_{rec}}{\lambda_{rec}} - 1 \right)$
- We estimate λ_{mark} , λ_{rec} and λ_{cap} from the data
- Calculate λ_{unc} and then simulate N_{unc}

Bayesian Implementation

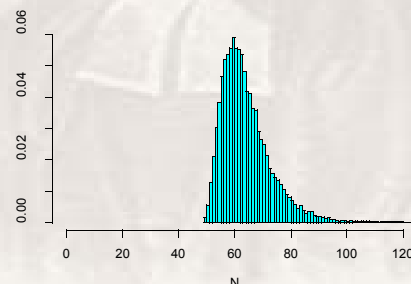
- The likelihoods for the observations are independent Poisson distributions
- We have to put priors on the means
 - exponential priors
 - conjugate, so posteriors also exponential
- The joint posterior for λ_{mark} , λ_{rec} and λ_{cap} , and hence the posterior for λ_{unc} can be calculated
- Use MCMC (Gibbs sampler)
- Then simulate N_{unc}

An Example

- Data from Fox & Pople (1984)
 - *Mus musculus*
 - part of a larger data set
- 31 individuals marked
- Second sample: 18 mice captured, 13 marked
- Uninformative priors: exponential, mean 100
- Run the MCMC

The Results

- Posterior mode: 59.1
- 95% confidence interval: 50.0 – 80.2



A worse alternative

- We could also write the model as Binomial distributions:
- $N_{mark} | N_{tot} \sim \text{Bin}(N_{tot}, p)$ and
- $N_{rec} | N_{rec} + N_{cap} \sim \text{Bin}(N_{rec} + N_{cap}, p)$
- Then, knowing N_{mark} , N_{rec} and N_{cap} we can estimate N_{tot}
 - but, more difficult to set a natural prior

A General Approach

- We can use log linear models more generally
- Can write as hierarchical models
- Then we estimate the parameters we are interested in
- The number of individuals is Poisson distributed
 - mean depends on several parameters

An Example

- The model is from Cormack, 1989
- Samples at several times
- Allow for arrivals and deaths
- Jolly-Seber model, re-written
- Have several possible catch histories
- e.g. 1=caught, 0=not caught: 010011101
- For n trapping times, $2^n - 1$ observable histories

The Modelling

- For each possible history, the number of individuals with that history will follow a Poisson distribution
 - making usual independence assumptions
- We can then model the expected numbers of individuals with each history.
- More interested in estimating the parameters
 - survival
 - birth rates/immigration

Parameters

- N – population size
- p_i – P(observed at time i | alive)
- φ_i – P(survive from i to $i+1$)
- χ_i – P(not seen after i)
- ψ_i – birth rate
 - ψ_i^{-1} = P(individual marked at $i+1$ was in pop. at i)
- ω_i – death rate
 - ω_i^{-1} = P(individual alive at i is alive at $i+1$)

One rate

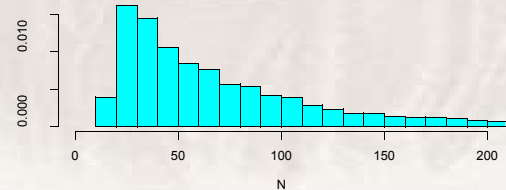
- As an example, look at the number observed at all 3 sampling times
- $\lambda_{111} = N p_1 \varphi_1 p_2 (1 - \chi_2)$
 - sampled at time 1 (probability p_1)
 - survives from time 1 to 2 (probability φ_1)
 - sampled at time 2 (probability p_2)
 - seen after time 2 (probability $1 - \chi_2$)
 - Multiply by the population size (N)

The Model

- $\lambda_{111} = Np_1\phi_1p_2(1-\chi_2)$
- $\lambda_{011} = N(1-p_1)\phi_1\psi_1p_2(1-\chi_2)$
- $\lambda_{101} = Np_1\phi_1(1-p_2)(1-\chi_2)$
- $\lambda_{001} = N(1-p_1)\phi_1(1-p_2)(1-\chi_2)$
- $\lambda_{110} = Np_1\phi_1p_2\chi_2$
- $\lambda_{010} = N(1-p_1)\phi_1\psi_1p_2\chi_2$
- $\lambda_{100} = Np_1\phi_1\omega_1(1-p_2)\chi_2$

Some Results

- Same data as before, different part of data set
- Posterior mode: 30.3
- 95% confidence interval: 5.7 – 198.4
 - wider than before
 - model more flexible



Bird Ringing

- Ornithology: birders ring birds at birth
- If a dead bird is captured, the ring is returned
- Some populations also surveyed
 - e.g. colonial birds
- Have two observations
 - neither perfect
- Birds are born in different years
- Interested in survival

Typical Data

Year of ringing	Number ringed	Year of recovery																						
		71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93
1970	1963	8	3	2	0	0	0	1	2	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1971	2163		1	1	2	2	1	3	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
1972	3092			7	2	2	5	1	5	2	1	3	1	1	1	2	0	0	0	0	0	0	0	0
1973	3412				15	1	1	1	2	3	2	0	1	1	1	0	0	0	0	0	0	0	0	0
1974	3945					12	1	7	4	1	1	1	1	0	1	2	1	1	0	0	0	0	0	0
1975	2561						12	3	0	3	1	1	1	2	0	3	0	0	0	0	0	0	0	0
1976	3304							11	3	5	1	3	4	2	3	0	1	0	1	1	0	0	0	0
1977	3478								13	6	0	5	4	2	1	2	4	0	0	0	1	0	0	0
1978	3165									18	5	2	2	0	5	3	0	0	0	1	0	0	0	0
1979	3351										7	4	7	1	4	3	3	0	1	0	0	1	0	1
1980	3487											12	8	2	2	4	5	3	0	1	2	0	0	1
1981	3939												26	2	3	3	3	4	1	0	0	0	0	0
1982	2881													11	0	6	3	0	1	2	0	0	1	0
1983	4119														19	7	7	4	0	0	2	1	0	0
1984	4006															13	3	2	0	0	0	0	1	0
1985	4867																23	2	5	2	0	3	3	2
1986	4769																	13	4	3	3	4	2	2
1987	5027																		14	2	1	2	3	2
1988	4804																			17	4	4	3	0
1989	3632																				11	4	2	4
1990	4170																					12	3	3
1991	4311																						9	4
1992	3180																							18

Complications

- Birds of different age may have different chances of survival
- Chances of survival may vary between years
- Chance of sighting or capture may also vary

An Example

- King & Brooks (2002)
- Data from Catchpole *et al.* 1998
- Scottish shags (which have a *Latin name*)
- 8637 ringed over 10 years of study
 - mainly as pulli (chicks)
- Immature shags spend 2 years at sea before returning to breeding site

The Parameters

- We need 3 types of term:
- For each an animal of age c :
- ϕ_t^c – Pr(animal alive at t survives to $t+1$)
- λ_t^c – Pr(animal that dies in $[t,t+1]$ has its band returned)
- p_{t+1}^c – Pr(animal alive at t is re-sighted then)
- λ_t^c and p_{t+1}^c are nuisance parameters

The Model

- We can write down the probability of any capture history from the parameters
- Don't need all histories, only these statistics:
 - The number of animals of age c :
 - z_t^c – observed after t , but not at t
 - d_t^c – recovered dead between t and $t+1$
 - w_t^c – observed alive at t
 - v_t^c – observed alive at t but not after

The Likelihood

- The likelihood can be written as:

$$l(x|\theta) = \prod_{c \in C} \left\{ \prod_{t=1}^{T-1} (\phi_t^c)^{w_t^c + z_t^c} (p_{t+1}^c)^{w_t^c} (1 - p_{t+1}^c)^{z_t^c} ((1 - \phi_t^c) \lambda_t^c)^{d_t^c} \prod_{t=1}^{T-1} (\chi_t^c)^{v_t^c} \right\}$$

- where

$$1 - \chi_t^c = (1 - \phi_t^c) \lambda_t^c + \phi_t^c (1 - (1 - p_{t+1}^c)) \chi_{t+1}^c$$

- A product of several terms
- Each term: parameter ^{statistic}
- Give the parameters flat priors

Models Selection

- Three parameters:
 - survival, observation, ring returned
- The model above has separate parameters for each age class and year
- But this is not efficient
 - we need a lot of data to estimate the parameters
- Can a model with less parameters be found?
 - still needs to fit the data well

Model Comparison

- We have several models, M_i ($i=1 \dots m$)
 - each has parameters $\theta(M_i)$
 - may be different number of parameters in each model
- We can calculate their posterior probabilities:

$$P(M_i|X) = \int P(M_i|\theta(M_i), X) P(\theta(M_i)|X) d\theta(M_i)$$
- Integrate over the uncertainty in the parameters
- How do we do this in practice?

Reverse Jump MCMC

- We can usually fit each model separately by MCMC
- We could fit each model, and then calculate the probabilities for each one
- But can be inefficient if there are a lot of models
- Instead, we can do it in one (long) run
- Added advantage: we can average over the different models

MCMC

- Reverse step MCMC uses the Metropolis-Hastings algorithm
 - starting from the current set of parameter values (θ^t), we propose a new value (θ^p) from a distribution, $g(\theta^p|\theta^t)$
 - accept with probability

$$r = \frac{P(\theta^p|X) g(\theta^t|\theta^p)}{P(\theta^t|X) g(\theta^p|\theta^t)}$$

- Repeat lots of times

rjMCMC

- M-H will not work for moves different models
 - if the dimensions are different, we have to correct the acceptance ratio for this
- We may also have to propose new parameters for the new model
- Propose extra parameters (u) from some distribution $q(u)$
- We alter the acceptance ratio to:

$$r = \frac{P(\theta^p|X) g(\theta^t|\theta^p)}{P(\theta^t|X) g(\theta^p|\theta^t) q(u)} \left| \frac{\partial \theta^p}{\partial (\theta^t, u)} \right|$$

rjMCMC: The Output

- The output is a list of parameter estimates
 - as with normal MCMC
- One column will be an indicator of the model
- Posterior probability for a model is the proportion of times it appears in the output
- The parameter estimates can be given conditional on a model
 - or averaged over models if they are common to them

The Shags

- Three parameters:
 - survival, observation, ring returned
- Four age classes:
 - Pulli (0-12months), 2nd year, 3rd year, Adult
 - can combine age classes, with some restrictions
- For each age class, parameters could be different for each year
- Total of 477 144 models

Shags: Results

- Different models:

Survival	Posterior model probability
{P;1}, {2,3}, {1;1}	0.366
{P;1}, {2,3}, {1}	0.219
{P;1}, {2}, {3;1}	0.142
{P;1}, {2}, {3}, {1;1}	0.140
{P;1}, {2}, {3}, {1}	0.092
<hr/>	
Recapture	
{P}, {2}, {3;1}, {1;1}	0.996
<hr/>	
Recovery	
{P}, {2,3}, {1}	0.597
{P,3}, {2}, {1}	0.182
{P}, {2}, {3,1}	0.070
{P}, {2}, {3}, {1}	0.053

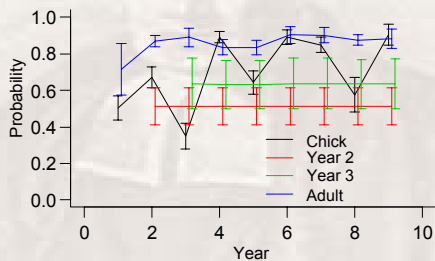
Shags: Survival

- The models:

Survival	Posterior model probability
{P;1}, {2,3}, {1;1}	0.366
{P;1}, {2,3}, {1}	0.219
{P;1}, {2}, {3;1}	0.142
{P;1}, {2}, {3}, {1;1}	0.140
{P;1}, {2}, {3}, {1}	0.092

- Two most likely: pulli, adolescents and adults have different survival (probability=0.585)
- Chick survival varies with year
- Adult survival may vary with year

Parameter Estimates



- More variation in chick survival

Comments

- Does it make sense to choose models in this way?
- Is there “one true model”?
- “All models are wrong, but some are useful”
 - G.E.P. Box
- In reality, survival of all age classes probably does vary between years
 - but in some age classes the variation is less

Why do model selection?

- One view of models is that they are simplified descriptions of reality
- A good model is one which is
 - realistic,
 - simple, and
 - fits the data
- Using rjMCMC concentrates on the fit of the data
 - but complex models do get penalised

Model Selection

- If “all models are wrong”, then what is the point of giving them a probability?
 - are we concentrating on “just a number”?
- Other approaches: less formal
- Use a criterion that balances model fit and complexity
 - DIC
- Use our own ideas about what is a reasonable model of the data

Another Criticism

- The analysis gives two choices
 - is/is not yearly variation
- But if there is always variation, then why do this?
 - with no data, the prior will make these equally likely
 - with a lot of data, we will learn that there are differences
- An alternative?

Hierarchical Models

- Rather than say there is/is not variation, treat it as a random effect
- Summarise the amount of variation with the variance
- Also helps prediction – we can now predict next year's survival
- Compare the variation between years for different age classes by comparing the variances

References

- Catchpole et al. (1998) *Biometrics* **54**: 292-304
- Cormack (1989) *Biometrics* **45**: 395-413
- Fox & Pople (1984) *Austr. J. Ecol.* **9**: 323-334
- King & Brooks (2002) *Biometrics* **58**: 841-851