

ECOLOGICAL STATISTICS: Analysis of Capture-Recapture Data

Rachel McCrea and Byron Morgan

National Centre for Statistical Ecology,
University of Kent,
Canterbury.

Outline

1. Introduction
2. Estimating abundance
3. Survival estimation
4. Model selection
5. Complex models
6. Parameter redundancy
7. State-space modelling
8. Bayesian analysis

Outline: Computer practicals

1. Introduction
2. Estimating abundance: **EstimateN**
3. Survival estimation: **Mark** and **RMark**
4. Model selection: **Eagle**
5. Complex models
6. Parameter redundancy: **Maple**
7. State-space modelling: **Kalm**
8. Bayesian analysis: **WinBUGs**

SECTION 1

INTRODUCTION

Introduction

- ❑ We are seeing changes in wild animal populations worldwide.
- ❑ In Britain there are notable declines in sparrows (*Passer domesticus*), lapwing (*Vanellus vanellus*), spotted fly-catcher (*Muscicapa striata*), etc.
- ❑ Effects of global warming and changes in farming practice



Motivation

- Importance of anthropogenic change: 6th mass extinction.
- 2010: The year of biodiversity
- <http://www.eol.org>





Large literature

- ❑ **Papers in journals such as:** Biometrics, JABES, JRSS, J.An. Ecology. Ecology, Oikos, J. Wildlife Management, Can. J. Fish. Aq. Sci., Meth. Ecol. Evol., Epidemiology
- ❑ **Books:** Seber (1982), Williams et al (2002), King et al (2009), McCrea and Morgan (2011)
- ❑ **Web-sites:** www.bto.org
- ❑ **Computer packages:** DISTANCE, MARK, E-SURGE, Presence.

Statistical Ecology; capture-recapture

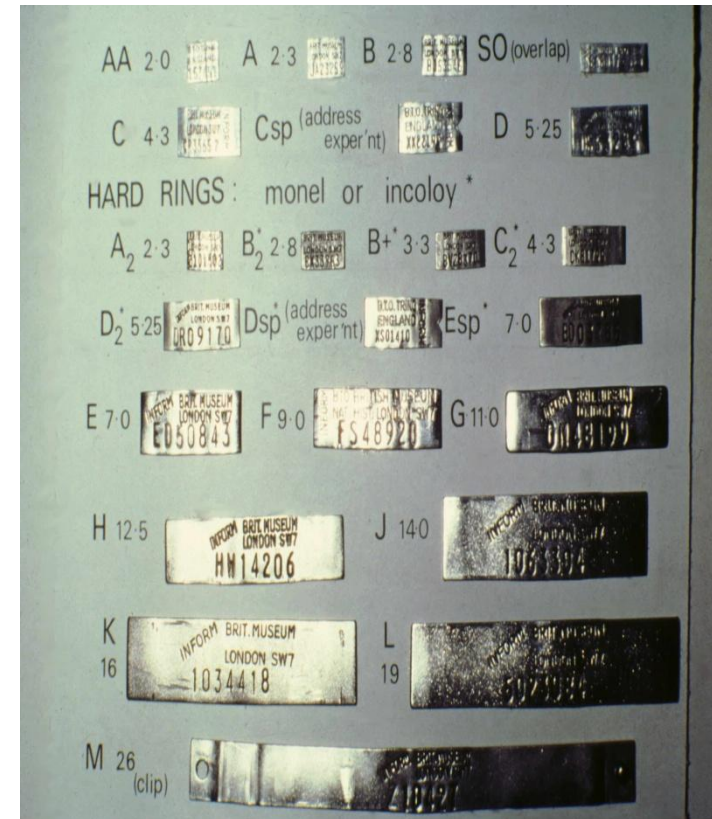
- ❑ Estimation of population sizes
- ❑ Estimation of important demographic rates (cf human demography):
 - Survival
 - Productivity
 - Movement
- ❑ Dealing with complexity
- ❑ Use of **capture-recapture** methods; started with H.C. Mortensen, 1899.

Recent Kent projects

- Jose multi-species survival; synchrony; IoM
- Guru Occupancy; tigers, Sumatra
- Beth Benthic organisms; diversity; trawling
- Rachel Model selection; goodness-of-fit; multi-state models
- Diana Parameter redundancy
- Lauren Spatio-temporal capture-recapture; cheetahs; missing time-varying individual covariates.
- Martin Meercat survival; closed population sizes
- Teresa Over-wintering wildfowl
- David Demography of the lizard orchid
- Dan local weather; use of lasso for variable selection
- Achaz Ibex; winter snow vs. reproductive senescence
- Eleni Stop-over models; demographic uses
- Takis Integrated population modelling; batch marking
- Ben Parameter redundancy

Marking

- We obtain information on survival from studying previously marked animals.
- These may be observed again alive or **dead**.
- It is assumed that marking does not affect behaviour (See *J. An. Ec.* 2005).



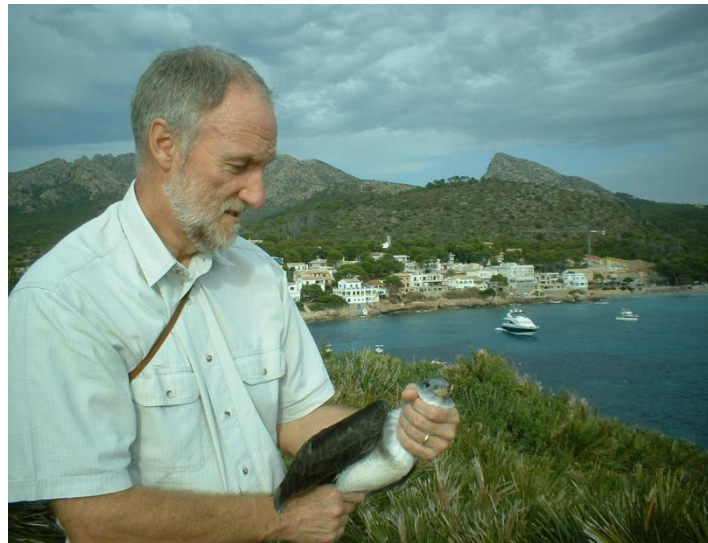
Shag rings, Isle of May: José



Identification/recapture/resighting of Cormorant, *Phalacrocorax carbo sinensis*, and great crested newt, *Titurus cristatus*



Further examples



Practicalities



Try again



Deer collars on Rum



Alternatives

- ▣ Radio tracking
- ▣ DNA
- ▣ Note possibility of errors
- ▣ Problems with ring loss

Fecundity

The key demographic processes involve *survival*, *reproduction* and *movement*. We shall be mainly concerned with the estimation of *survival*.

Estimation of fecundity can be challenging, but is typically easier than the estimation of survival. For example, **Cory's shearwaters** lay just one egg.



The Classical Approach

- ❑ For a given data set we consider a **family of alternative probability models**.
- ❑ Each model may be fitted by means of optimising some criterion, for example by **maximising a likelihood** over the parameter space. Many parameters may be involved.
- ❑ Models may be compared in terms of maximised likelihoods, or information criteria, such as the **AIC** or **BIC**.
- ❑ Usually a **single best model** is selected to represent the data, in terms of the model parameter estimates and estimates of their precision.

Classical continued

- ❑ Models may be **averaged**, using model weights derived from AICs. Cf. Bayesian approach.
- ❑ Model averaging, in general, should not be uncritical, as differences resulting from different models should be understood (McAllister).
- ❑ **Computing** is usually done using packages such as MARK or E-SURGE. It is necessary to interface with packages in order to make model specifications. For example one can use **parameter-index matrices**, PIMs.
- ❑ Diagnostics include looking at residuals of various kinds, and constructing tests of **goodness-of-fit**.

The Bayesian approach

- ▣ Brooks Catchpole and Morgan, (2000)
Bayesian Animal Survival Estimation,
Statistical Science, **15**, 357-376.
- ▣ King, Morgan, Gimenez and Brooks (2009)
Bayesian Analysis for Population Ecology.
Chapman & Hall/CRC, London.

SECTION 2

ESTIMATING ABUNDANCE

Estimating Abundance

- ▣ Standard estimators
- ▣ The importance of heterogeneity
- ▣ Illustrations from different areas

How many individuals? From skinks..



..to taxis: need for individual
identification



Medical and social applications

- How many French people are there (Laplace)?
- How many illegal firearms?
- How many drugs users? Eg., in a Bangkok study of drug users, 7062 individuals had hospital encounters, with $f_1=2955$, $f_2=1186$, $f_3=803$, ...
- How many traffic offences?

How many animals?

The **Lincoln-Petersen** estimate is given by the solution in N to the equation below, where a first sample of n_1 animals are captured and marked, and then a later sample resulted in n_2 animals, of which m_2 were marked.

$$\frac{n_1}{N} = \frac{m_2}{n_2}$$

The Schnabel census: data

Repetition of the sampling scheme results in data of the form below

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-------|-----|----|----|----|----|---|---|---|---|----|
| hares | 25 | 22 | 13 | 5 | 1 | 2 | 0 | 0 | 0 | 0 |
| taxis | 142 | 81 | 49 | 7 | 3 | 1 | 0 | 0 | 0 | 0 |
| voles | 29 | 15 | 5 | 16 | 27 | 0 | 0 | 0 | 0 | 0 |
| voles | 18 | 15 | 8 | 6 | 5 | 0 | 0 | 0 | 0 | 0 |
| chips | 14 | 13 | 18 | 12 | 7 | 5 | 1 | 1 | 0 | 1 |

The Horvitz-Thompson estimate

Let p_i denote the probability that the i^{th} animal is caught.

The Horvitz-Thompson estimate of population size is:

$$\hat{N} = \sum \{(1 - (1 - p_i)^t\}^{-1},$$

where the sum is taken over all caught animals, and t is the number of samples.

Horvitz-Thompson

This leads to:

$$\hat{N} = \sum_{j=1}^t f_j \{1 - (1 - j/t)^t\}^{-1},$$

where f_j is the number of animals caught j times.

Likelihood for the Schnabel census

The general form of the likelihood is given by

$$L(N, \theta) \propto \frac{N!}{(N - D)!} \prod_{j=0}^k p_j^{f_j},$$

where N is the population size, θ denotes the model parameters, p_j now denotes the probability an animal is caught j times, f_j is the number of animals caught j times, k is the number of sampling occasions and D is the number of distinct animals caught.

Violation of assumptions

- ❑ This approach is harder than it seems because, amongst other things, animals differ in their re-capture rates.
- ❑ Nevertheless, this remains an area of active research.

Models: the bias due to unmodelled heterogeneity can be severe!

- Example: Ayre (1962) estimated an anthill population to be 109 when there were known to be 3,000 ants in it.

Ayre, L.G. 1962. Problems in using the Lincoln index for estimating the size of ant colonies (*Hymenopter formicidae*). *J.N.Y. Ent. Soc.* **70**: 159-166.

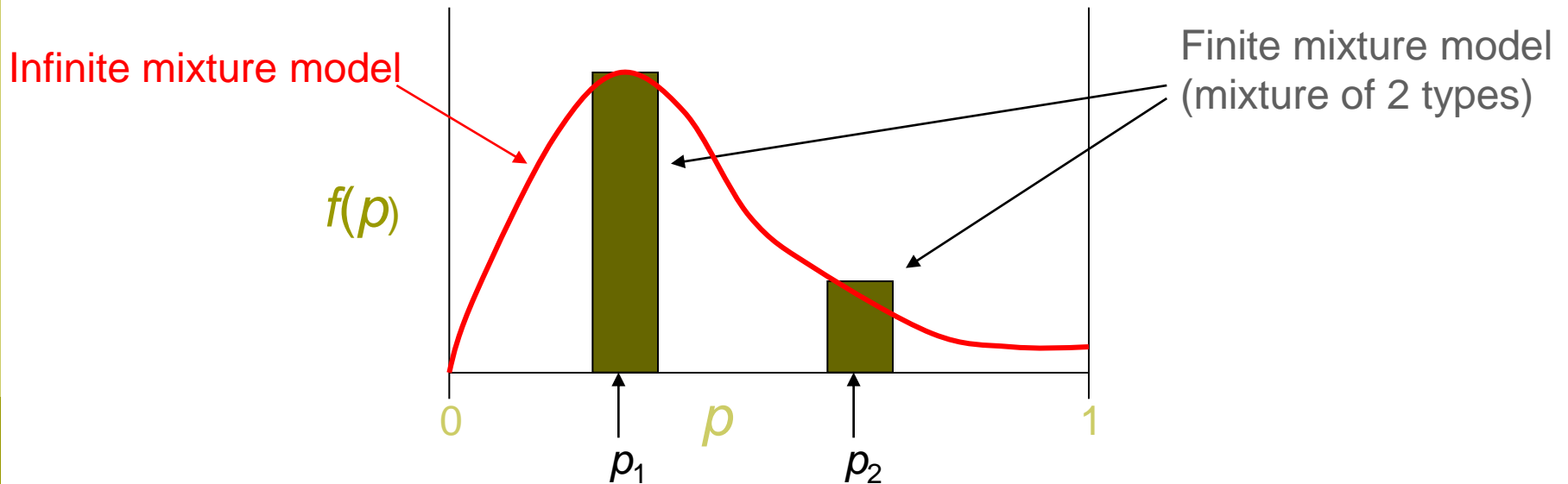
Two recent approaches to unobservable heterogeneity

- Finite Mixture models¹: approximate the distribution of p using a mixture of a few p s (conceptually from a few sub-populations).
- Continuous (or infinite mixture) models²: Model the distribution of p using some flexible continuous distribution (e.g. Beta)

¹ Pledger, S. 2000. Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* **56**: 434-442.

² Dorazio, R.M. and Royle, J.A. 2003. Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **59**: 352-364.

Alternative mixture models



Types of mark-recapture model

- M_0 : No heterogeneity
- M_t : Capture probabilities vary with time (occasion)
- M_b : Capture probabilities depend on behaviour (trap-shy or trap-happy)
- M_h : Capture probabilities are heterogeneous (vary between individuals)
- And combinations: M_{tb} , M_{th} , M_{bh} , M_{tbh}

New mixture model

We proposed a mixture of **binomial and beta-binomial** distributions for estimating the size of closed populations.

The new model includes discrete (two populations) and infinite mixtures, used for the recapture probability by Pledger, and Dorazio and Royle, as special cases

Modelling heterogeneity of recapture

$$p_j \propto \sum \gamma_a \pi_a^j (1 - \pi_a)^{k-j}$$

$$p_j \propto \Pi_r(\mu + r\theta) \Pi_r(1 - \mu + r\theta) / \Pi_r(1 + r\theta)$$

$$\equiv \pi_j^{\text{Be}}$$

$$p_j \propto \{\gamma \pi^j (1 - \pi)^{k-j} + (1 - \gamma) \pi_j^{\text{Be}}\}$$

Alternatives

Nonparametric (E and Q), and the logistic-normal-binomial model:

$$p_j \propto \phi^j (1 - \phi)^{k-j},$$

where

$$\phi = \int_{-\infty}^{\infty} \frac{1}{\{1 + \exp(-x)\}} f(x) dx,$$

and $f(x)$ denotes the $N(\mu, \sigma^2)$ density.

Illustration of different precision for a single data set, where it is thought $N=102$.

| Model | N | 95% Profile interval |
|----------------|-------|----------------------|
| Beta-bin | 106.0 | 81.2-313.2 |
| 2 Bins | 72.8 | 72.0-76.4 |
| Bin + Beta-bin | 77.7 | 72.5-155.8 |
| E | 84.5 | 75.6-103.8 |
| Q | 80.7 | 77.6-84.4 |

Change to beta parameters

$$f(p) \propto p^{\alpha-1} (1-p)^{\beta-1}.$$

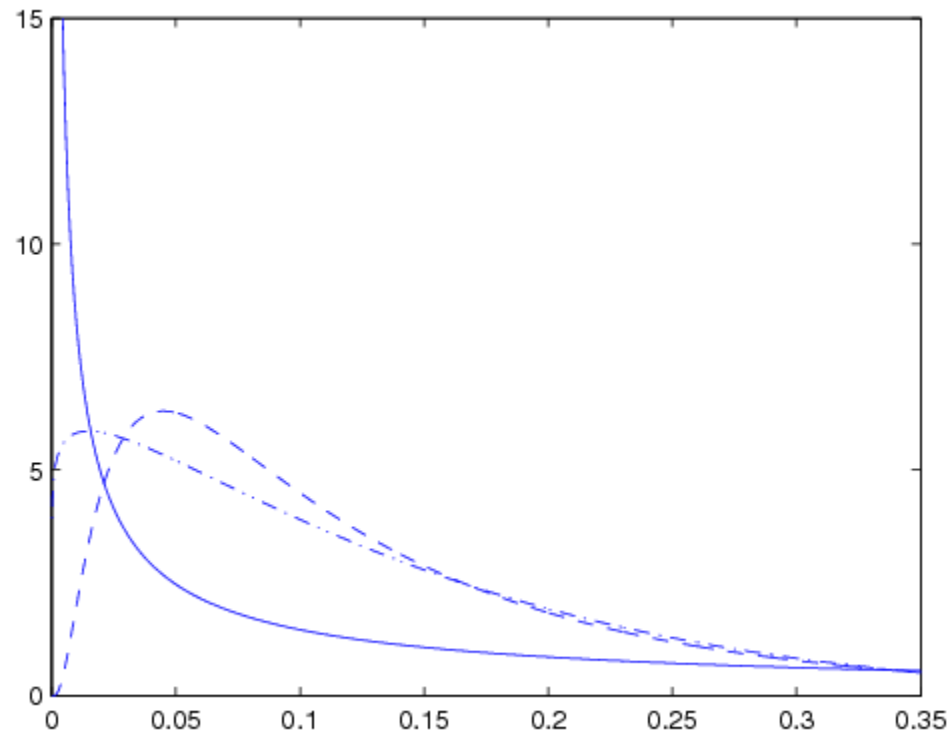
- For the beta-binomial model, we have $\alpha=0.40, \beta=3.41$.
- For the new mixture we have $\alpha=1.30, \beta=5.95$.

Modelling the parameter p near zero

Thus the new model allows flexible modelling of the probability of recapture, p , near zero, **which is important for estimating N .**

Shown in the next graph is how the logistic-normal-binomial, beta and the beta component of the new mixture can differ in modelling p near zero, for a particular data set.

Different capture probabilities near zero.



Results: illustrative comparisons of model fits: -max log-lik., values.

| Data | Binomial | Beta-binomial | 2 Bins | Bin + Beta-bin |
|------------|--------------|---------------|--------------|----------------|
| House mice | 44.43 | 43.55 | 39.88 | 39.58 |
| Skinks | 86.71 | 22.35 | 23.04 | 18.57 |
| Wood mice | 357.26 | 47.45 | 87.73 | 45.34 |
| Taxicabs | 16.95 | 16.44 | 16.34 | 16.34 |
| Squirrels | 39.38 | 19.24 | 17.82 | 17.76 |

Illustrative coverage, 95% profile CI

| True model | Fitted model | Coverage |
|------------|----------------|----------|
| Beta-bin | Binomial | 0.003 |
| Beta-bin | Beta-bin | 0.953 |
| Beta-bin | 2 Bins | 0.750 |
| Beta-bin | Bin + Beta-bin | 0.933 |
| 2 Bins | Binomial | 0.168 |
| 2 Bins | Beta-bin | 0.288 |
| 2 Bins | 2 Bins | 0.938 |
| 2 Bins | Bin + Beta-bin | 0.955 |

Further work

- ❑ Mix a binomial with a logistic-normal-binomial, in order to include also the cases of time and behavioural response.
- ❑ Extend mixture to modelling **survival** and **site-occupancy** models.
- ❑ Include covariates.

SECTION 3

SURVIVAL ESTIMATION

Survival estimation

- ▣ 3.1: Analysing recovery data
- ▣ 3.2: Capture-recapture data
- ▣ 3.3: Integrated recaptures and recoveries

Analysing Recovery Data

- ❑ Cohort of individuals marked and re-released into the population
 - When individuals die, they may be recovered dead; or their rings/marks may be recovered.
 - Form encounter histories for each individual:
 - 1: time individual marked
 - 2: time individual recovered dead
- ❑ 100020
- ❑ 010000

Recovery Data

- R_i : number of marked individuals released at occasion t_i
- d_{ij} : number of individuals released at occasion t_i and recovered dead in the time interval (t_{j-1}, t_j)
 - Consider a 4 encounter occasion study, then individuals released at time t_1 can be summarised as:

| | | | | |
|-------|-----------|-----------|-----------|-------------------------------------|
| R_1 | $d_{1,2}$ | $d_{1,3}$ | $d_{1,4}$ | $R_1 - d_{1,2} - d_{1,3} - d_{1,4}$ |
|-------|-----------|-----------|-----------|-------------------------------------|

Multinomial Distribution

| | | | | |
|----------------|-----------|-----------|-----------|-------------------------------------|
| $\mathbf{R_1}$ | $d_{1,2}$ | $d_{1,3}$ | $d_{1,4}$ | $R_1 - d_{1,2} - d_{1,3} - d_{1,4}$ |
|----------------|-----------|-----------|-----------|-------------------------------------|

□ Defining:

- ϕ_i : probability an individual alive at occasion t_i survives until occasion t_{i+1}
- λ_i : probability an individual who dies between occasion t_i and t_{i+1} is recovered (or has its ring/tag) and reported

| | | | | |
|----------------|-----------------------|-----------------------------|-----------------------------------|---------------------------|
| $\mathbf{R_1}$ | $(1-\phi_1)\lambda_1$ | $\phi_1(1-\phi_2)\lambda_2$ | $\phi_1\phi_2(1-\phi_3)\lambda_3$ | $1 - \sum_j \Pr(d_{i,j})$ |
|----------------|-----------------------|-----------------------------|-----------------------------------|---------------------------|

Product Multinomial

- Consider a general m-array for a 4 occasion recovery study

| Releases | Recovery Occasion | | | Never Recovered |
|----------|-------------------|-----------|-----------|-----------------|
| R_1 | $d_{1,2}$ | $d_{1,3}$ | $d_{1,4}$ | $d_{1,\infty}$ |
| R_2 | | $d_{2,3}$ | $d_{2,4}$ | $d_{2,\infty}$ |
| R_3 | | | $d_{3,4}$ | $d_{3,\infty}$ |

where $d_{i,\infty} = R_i - \sum_j d_{i,j}$

Product Multinomial

- Then the corresponding probabilities are

| Releases | Recovery Occasion | | | Never Recovered |
|----------|-------------------------|-------------------------------|-------------------------------------|---------------------------|
| R_1 | $(1 - \phi_1)\lambda_1$ | $\phi_1(1 - \phi_2)\lambda_2$ | $\phi_1\phi_2(1 - \phi_3)\lambda_3$ | $1 - \sum_j \Pr(d_{1,j})$ |
| R_2 | | $(1 - \phi_2)\lambda_2$ | $\phi_2(1 - \phi_3)\lambda_3$ | $1 - \sum_j \Pr(d_{2,j})$ |
| R_3 | | | $(1 - \phi_3)\lambda_3$ | $1 - \sum_j \Pr(d_{3,j})$ |

- Assuming independence between release cohorts, the log-likelihood function is then defined by

$$\log(L) = \text{constant} + \sum_i \sum_j d_{i,j} \log(\Pr(d_{i,j}))$$

Example: Mallards, *Anas platyrhynchos*

| Year of ringing | No. ringed | Year of recovery (-1962) | | | | | | | | |
|--------------------|---------------|--------------------------|----|----|----|----|----|-----|----|----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1963 | 231 | 10 | 13 | 6 | 1 | 1 | 3 | 1 | 2 | 0 |
| 1964 | 649 | | 58 | 21 | 16 | 15 | 13 | 6 | 1 | 1 |
| 1965 | 885 | | | 54 | 39 | 23 | 18 | 11 | 10 | 6 |
| 1966 | 590 | | | | 44 | 21 | 22 | 9 | 9 | 3 |
| 1967 | 943 | | | | | 55 | 39 | 23 | 11 | 12 |
| 1968 | 1077 | | | | | | 66 | 46 | 29 | 18 |
| 1969 | 1250 | | | | | | | 101 | 59 | 30 |
| 1970 | 938 | | | | | | | | 97 | 22 |
| 1971 | 312 | | | | | | | | | 21 |

MLEs: Mallard model $\phi(\cdot), \lambda(t)$

| Parameter | MLE | SE |
|---------------|-------|--------|
| $\phi(\cdot)$ | 0.638 | 0.0153 |
| λ_1 | 0.118 | 0.0369 |
| λ_2 | 0.247 | 0.0290 |
| λ_3 | 0.162 | 0.0180 |
| λ_4 | 0.189 | 0.0186 |
| λ_5 | 0.167 | 0.0156 |
| λ_6 | 0.193 | 0.0152 |
| λ_7 | 0.200 | 0.0143 |
| λ_8 | 0.226 | 0.0151 |
| λ_9 | 0.156 | 0.0144 |

Freeman-Morgan Model

- For bird populations, first-year survival is often dramatically different from adult survival
- If individuals marked as nestlings, then it is possible to adapt the model to incorporate age-dependent parameters

Freeman-Morgan Model

- Introduce model notation $x/y/z$:
 - x : first-year-survival probability
 - y : adult-survival probability
 - z : recovery probability
- Possible parameter dependencies
 - c : constant
 - t : time-dependence
 - a_k : age-dependence up to age k
 - v : covariate-dependence (see later section)

Model c/c/t

- $\phi(1)$: first-year survival probability
- $\phi(a)$: adult survival probability

| Releases | Recovery Occasion | | | Never Recovered |
|----------|--------------------------|---------------------------------|--|---------------------------|
| R_1 | $(1 - \phi(1))\lambda_1$ | $\phi(1)(1 - \phi(a))\lambda_2$ | $\phi(1)\phi(a)(1 - \phi(a))\lambda_3$ | $1 - \sum_j \Pr(d_{1,j})$ |
| R_2 | | $(1 - \phi(1))\lambda_2$ | $\phi(1)(1 - \phi(a))\lambda_3$ | $1 - \sum_j \Pr(d_{2,j})$ |
| R_3 | | | $(1 - \phi(1))\lambda_3$ | $1 - \sum_j \Pr(d_{3,j})$ |

More general age-dependence

$$Pr(d_{ij}(a)) = \begin{cases} 0 & i > j \\ (1 - \phi_j(a))\lambda_j(a) & i = j \\ \prod_{k=1}^{j-1} \phi_k(a + k - i) \\ \quad \times (1 - \phi_j(a + j - i))\lambda_j(a + j - i) & i < j \end{cases}$$

Example: Mallards (marked as young)

| Year of ringing | No. ringed | Year of recovery (-1962) | | | | | | | | |
|--------------------|---------------|--------------------------|-----|----|-----|-----|-----|-----|----|----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1963 | 962 | 82 | 35 | 18 | 16 | 6 | 8 | 5 | 3 | 1 |
| 1964 | 702 | | 103 | 21 | 13 | 11 | 8 | 6 | 6 | 0 |
| 1965 | 1132 | | | 82 | 36 | 26 | 24 | 15 | 18 | 4 |
| 1966 | 1201 | | | | 153 | 39 | 22 | 21 | 16 | 8 |
| 1967 | 1199 | | | | | 109 | 38 | 31 | 15 | 1 |
| 1968 | 1155 | | | | | | 113 | 64 | 29 | 22 |
| 1969 | 1131 | | | | | | | 124 | 45 | 22 |
| 1970 | 906 | | | | | | | | 95 | 25 |
| 1971 | 353 | | | | | | | | | 28 |

MLEs: Mallards t/c/t model

| Parameter | MLE | SE | Parameter | MLE | SE |
|-------------|-------|--------|-------------|-------|--------|
| $\phi_1(1)$ | 0.441 | 0.0487 | λ_1 | 0.154 | 0.0201 |
| $\phi_2(1)$ | 0.481 | 0.0545 | λ_2 | 0.280 | 0.0327 |
| $\phi_3(1)$ | 0.585 | 0.0461 | λ_3 | 0.181 | 0.0217 |
| $\phi_4(1)$ | 0.459 | 0.0419 | λ_4 | 0.223 | 0.0189 |
| $\phi_5(1)$ | 0.425 | 0.0424 | λ_5 | 0.173 | 0.0148 |
| $\phi_6(1)$ | 0.604 | 0.0439 | λ_6 | 0.239 | 0.0222 |
| $\phi_7(1)$ | 0.589 | 0.0450 | λ_7 | 0.271 | 0.0228 |
| $\phi_8(1)$ | 0.560 | 0.0549 | λ_8 | 0.236 | 0.0207 |
| $\phi_9(1)$ | 0.293 | 0.1348 | λ_9 | 0.152 | 0.0172 |
| $\phi(2+)$ | 0.692 | 0.0198 | | | |

Extensions

- ❑ Modelling ring-recovery data when the number of ringed individuals is not known (Brown, 2010)
- ❑ Modelling adult-survival heterogeneity (Besbeas et al, 2009)
- ❑ Mixture models for modelling age-dependent survival (McCrea et al, 2010)

References

- Besbeas et al (2009) in *Modelling Demographic processes in marked populations*. pp 513-539
- Brown (2010) PhD Thesis, University of Kent
- Brownie et al (1985) *Statistical inference from band recovery data – A handbook*.
- Freeman and Morgan (1992) *Biometrics*, **48**: 217-235.
- McCrea et al, (2010) Submitted.

Capture-recapture data

- Suppose now, that instead of recovering dead individuals, attempts are made instead to recapture live individuals
- 1: initial capture and subsequent live recaptures
- Examples:
 - 101101
 - 100000

Recapture M-array

| Number Released | Recaptured t_2 | Recaptured t_3 | Recaptured t_4 | Never Recaptured |
|-----------------|------------------|------------------|------------------|----------------------------------|
| R_1 | m_{12} | m_{13} | m_{14} | $R_1 - m_{12} - m_{13} - m_{14}$ |
| R_2 | | m_{23} | m_{24} | $R_2 - m_{23} - m_{24}$ |
| R_3 | | | m_{34} | $R_3 - m_{34}$ |

- R_i : Number of individuals released in year t_i
- m_{ij} : Number of individuals released in year t_i and next recaptured in year t_j

Encounter Histories to M-arrays

| Number Released | Recaptured t_2 | Recaptured t_3 | Recaptured t_4 | Never Recaptured |
|-----------------|------------------|------------------|------------------|----------------------------------|
| R_1 | m_{12} | m_{13} | m_{14} | $R_1 - m_{12} - m_{13} - m_{14}$ |
| R_2 | | m_{23} | m_{24} | $R_2 - m_{23} - m_{24}$ |
| R_3 | | | m_{34} | $R_3 - m_{34}$ |

- **1011**
- R_i : newly marked individuals AND re-released individuals

Singlesite Probabilities

| Number Released | Recaptured t_2 | Recaptured t_3 | Recaptured t_4 | Never Recaptured |
|-----------------|------------------|---------------------------|--|---------------------------------|
| R_1 | $\phi_1 p_2$ | $\phi_1(1-p_2)\phi_2 p_3$ | $\phi_1(1-p_2)\phi_2(1-p_3)\phi_3 p_4$ | $1-\Sigma \text{probabilities}$ |
| R_2 | | $\phi_2 p_3$ | $\phi_2(1-p_3)\phi_3 p_4$ | $1-\Sigma \text{probabilities}$ |
| R_3 | | | $\phi_3 p_4$ | $1-\Sigma \text{probabilities}$ |

- ϕ_i : probability an animal alive at time t_i survives to time t_{i+1}
- p_{i+1} : probability an animal alive at time t_{i+1} is recaptured at time t_{i+1}
- Cormack-Jolly-Seber Model: $\phi(t), p(t)$

Example: Great Crested Newts

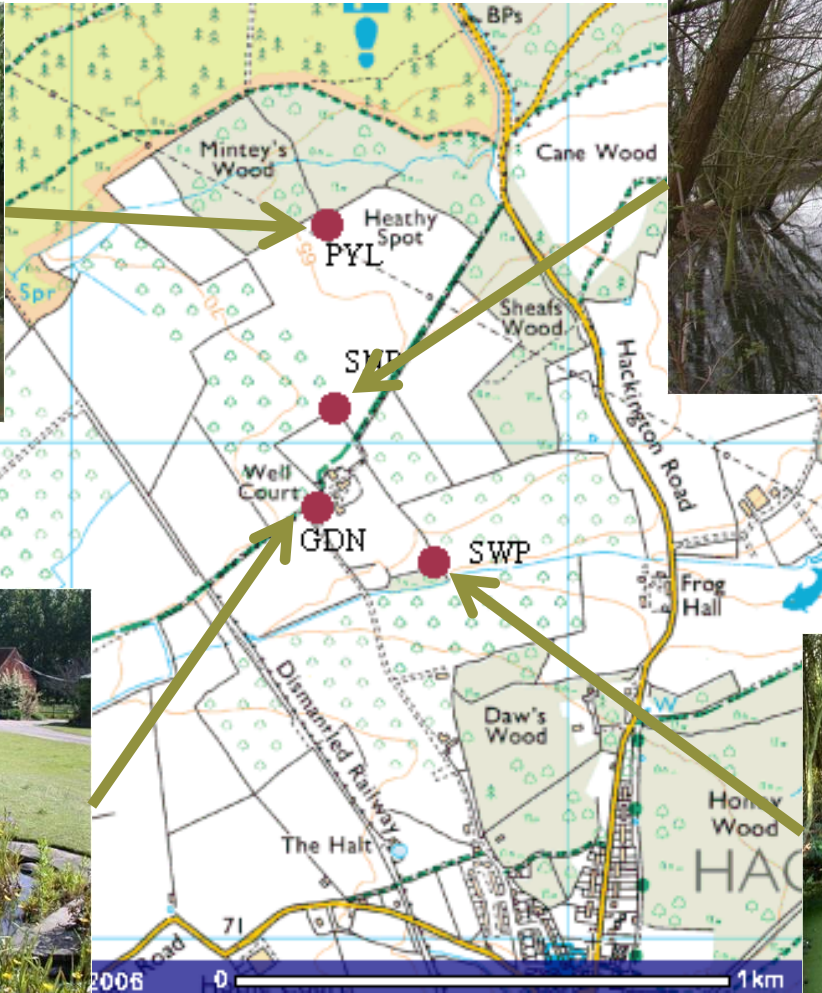


- Mark-recapture data collected between 1995 and 2006
- Population of Great crested newts, *Tristurus cristatus*, close to the University of Kent campus.
- Meta-population studied over four groups of ponds
- Best model: $\phi(t), p(\text{pond})$

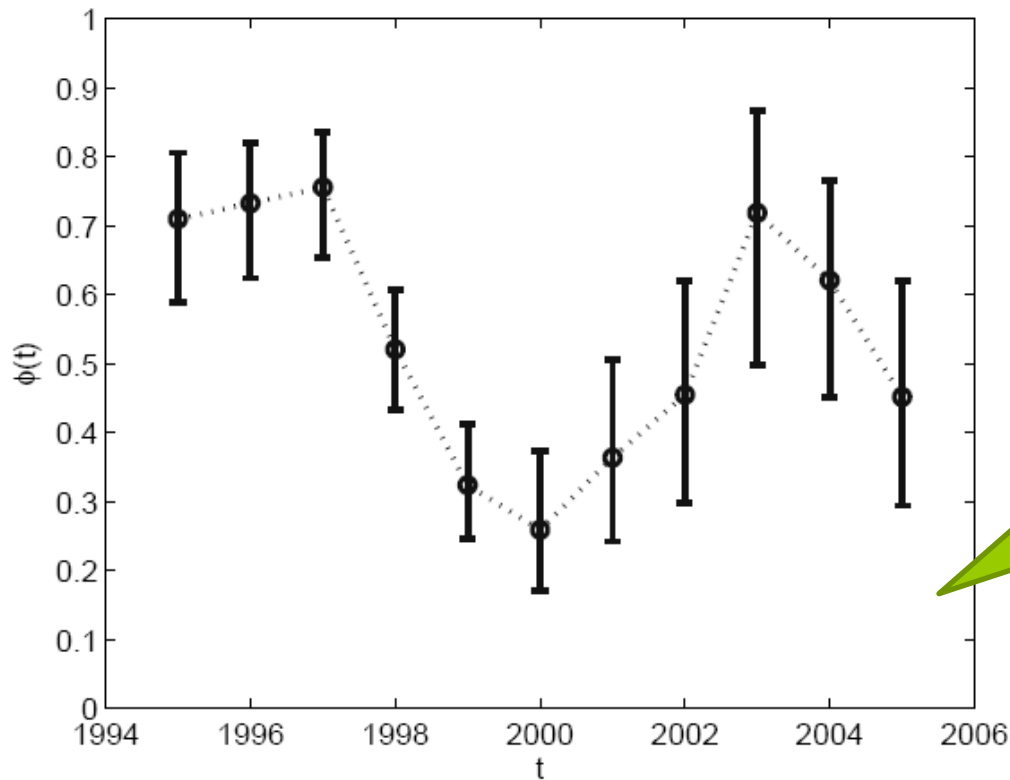




Wellcourt Study Site



Great Crested Newts: $\phi(t), p(\text{pond})$



What is causing this temporal variation?

Griffiths et al (2010)

- Garden Pond: $P = 0.697$ (0.629-0.757)
- Swimming Pool: $p = 0.845$ (0.726-0.919)
- Snake Pond: $p = 0.294$ (0.225-0.373)
- Pylon Pond: $p = 0.233$ (0.100-0.455)

Extensions

- ❑ General age and time-dependent models
- ❑ Individual heterogeneity (random effect) models (Gimenez and Choquet, 2010)
- ❑ Incorporation of behavioural traits, e.g. trap-response

References

- ❑ Cormack (1964) *Biometrika*, **51**: 429-438
- ❑ Gimenez and Choquet (2010) *Ecology*, **91**: 951-957
- ❑ Griffiths et al (2010) *Biological Conservation*, **143**: 485-491
- ❑ Jolly (1965) *Biometrika*, **52**:225-247
- ❑ Seber (1965) *Biometrika*, **52**:249-259

Joint recapture and recovery models

- What if both recapture and recovery data are collected on the same individuals?

Sufficient Statistics

- ❑ Catchpole, Freeman, Morgan and Harris (1998) Integrated Recovery/Recapture Data Analysis. *Biometrics*. **54**, 33-46
- ❑ General – cohort/age/time-dependence
- ❑ Assumes no emigration
- ❑ Alternative modelling approach in Barker (1997,1999).

CM Parameters (Time-dependent Model)

- ϕ_j : probability an animal alive at time t_j survives until time t_{j+1} ;
- λ_j : probability an animal which dies in the interval (t_j, t_{j+1}) , has its death reported;
- p_j : probability an animal alive at time t_j , is captured at t_j .

CM Likelihood Construction

1 1 2 0 0

1 0 0 2 0

1 1 1 1 1

1 1 0 0 0

- 1: Live recapture
- 2: Dead recovery

CM Likelihood Construction

$$1 \ 1 \ 2 \ 0 \ 0 \quad \phi_1 p_2 (1 - \phi_2) \lambda_2$$

$$1 \ 0 \ 0 \ 2 \ 0 \quad \phi_1 (1 - p_2) \phi_2 (1 - p_3) (1 - \phi_3) \lambda_3$$

$$1 \ 1 \ 1 \ 1 \ 1 \quad \phi_1 p_2 \phi_2 p_3 \phi_3 p_4 \phi_4 p_5 \chi_5$$

$$1 \ 1 \ 0 \ 0 \ 0 \quad \phi_1 p_2 \chi_2$$

- χ_j : probability an animal alive at time t_j is not seen again, alive or dead after t_j

CM Likelihood Construction

$$1 \ 1 \ 2 \ 0 \ 0 \quad \phi_1 p_2 (1 - \phi_2) \lambda_2$$

$$1 \ 0 \ 0 \ 2 \ 0 \quad \phi_1 (1 - p_2) \phi_2 (1 - p_3) (1 - \phi_3) \lambda_3$$

$$1 \ 1 \ 1 \ 1 \ 1 \quad \phi_1 p_2 \phi_2 p_3 \phi_3 p_4 \phi_4 p_5 X_5$$

$$1 \ 1 \ 0 \ 0 \ 0 \quad \phi_1 p_2 X_2$$

- Known survival.....

$$\alpha_j = \prod_{s=1}^{j-1} \phi_s$$

CM Likelihood Construction

$$1 \ 1 \ 2 \ 0 \ 0 \quad \phi_1 p_2 (1 - \phi_2) \lambda_2$$

$$1 \ 0 \ 0 \ 2 \ 0 \quad \phi_1 (1 - p_2) \phi_2 (1 - p_3) (1 - \phi_3) \lambda_3$$

$$1 \ 1 \ 1 \ 1 \ 1 \quad \phi_1 p_2 \phi_2 p_3 \phi_3 p_4 \phi_4 p_5 X_5$$

$$1 \ 1 \ 0 \ 0 \ 0 \quad \phi_1 p_2 X_2$$

- Disappearing individuals.....

$$1 - \chi_j = (1 - \phi_j) \lambda_j + \phi_j (1 - (1 - p_{j+1}) \chi_{j+1})$$

CM Likelihood Construction

$$1 \ 1 \ 2 \ 0 \ 0 \quad \phi_1 p_2 (1 - \phi_2) \lambda_2$$

$$1 \ 0 \ 0 \ 2 \ 0 \quad \phi_1 (1 - p_2) \phi_2 (1 - p_3) (1 - \phi_3) \lambda_3$$

$$1 \ 1 \ 1 \ 1 \ 1 \quad \phi_1 p_2 \phi_2 p_3 \phi_3 p_4 \phi_4 p_5 X_5$$

$$1 \ 1 \ 0 \ 0 \ 0 \quad \phi_1 p_2 X_2$$

- Recaptures.....

CM Likelihood Construction

$$1 \ 1 \ 2 \ 0 \ 0 \quad \phi_1 p_2 (1 - \phi_2) \lambda_2$$

$$1 \ 0 \ 0 \ 2 \ 0 \quad \phi_1 (1 - p_2) \phi_2 (1 - p_3) (1 - \phi_3) \lambda_3$$

$$1 \ 1 \ 1 \ 1 \ 1 \quad \phi_1 p_2 \phi_2 p_3 \phi_3 p_4 \phi_4 p_5 X_5$$

$$1 \ 1 \ 0 \ 0 \ 0 \quad \phi_1 p_2 X_2$$

- Death and recovery...

CM Likelihood Construction

$$1 \ 1 \ 2 \ 0 \ 0 \quad \phi_1 p_2 (1 - \phi_2) \lambda_2$$

$$1 \ 0 \ 0 \ 2 \ 0 \quad \phi_1 (1 - p_2) \phi_2 (1 - p_3) (1 - \phi_3) \lambda_3$$

$$1 \ 1 \ 1 \ 1 \ 1 \quad \phi_1 p_2 \phi_2 p_3 \phi_3 p_4 \phi_4 p_5 \chi_5$$

$$1 \ 1 \ 0 \ 0 \ 0 \quad \phi_1 p_2 \chi_2$$

- Non-captures.....

CM Sufficient Statistics

- ▣ $D(j)$: number of animals recovered dead in the interval (t_j, t_{j+1}) ;
- ▣ $V(j)$: number of animals captured or recaptured at t_j and not seen again during the study;
- ▣ $W(j)$: number of animals recaptured at t_{j+1} ;
- ▣ $Z(j)$: number of animals not recaptured at t_{j+1} but encountered later either dead or alive

CM Likelihood Construction

$$L \propto \prod_{j=1}^{T-1} [\alpha_j (1 - \phi_j) \lambda_j]^{D(j)} \prod_{j=1}^T [\alpha_j \chi_j]^{V(j)} \prod_{j=1}^{T-1} p_{j+1}^{W(j)} (1 - p_{j+1})^{Z(j)}$$

- Note that many animals are counted in both $V(j)$ and $W(j)$
- Sufficient statistics D , V , W and Z are therefore not independent multinomials
- Time-dependence structure given here, CM form completely general.

SECTION 4

MODEL ASSESSMENT

□ Model Selection

- Comparing models: which one is best?
- AIC/LRT tests
- Step-wise approaches using **score tests**

□ Goodness-of-fit

- **Absolute** goodness-of-fit tests
- **Diagnostic** assessment – departure from model assumptions

Model Selection using AIC

- ❑ Return to the Great crested newt example presented earlier
- ❑ Cormack-Jolly-Seber model:
 - Constant or time-dependent parameters
 - Pond specific parameters
 - Weather covariates?

AIC

Akaike's information criterion is defined by:

$$AIC = -2 \log L(\hat{\theta} \mid x) + 2K$$

where $L(\cdot \mid x)$ is the likelihood function given the observed data x , $\hat{\theta}$ are the MLEs of the components of parameter θ and K is the size of $\hat{\theta}$, which can be interpreted as the number of estimable parameters of the model.

AIC

- The smaller the AIC, the smaller the $-2\log L$, which will signify a comparatively better fitting model. Of course the number of parameters is also discriminated against within the information criterion, and the AIC can be used to rank models in a model set.
- Models not in the set remain out of consideration. AIC is useful in selecting the best model in the set; however, if all models are very poor, AIC will still select the one estimated to be the best, but even that relatively best model might be poor in an absolute sense.
- A full discussion and derivation of the AIC can be found in Burnham and Anderson (2002).

Relative AIC

- It is not the absolute size of the AIC value that is important, rather it is the relative values over the set of models considered.
- Particularly the differences between particular AIC values, Δ_i , that are important:

$$\Delta_i = \text{AIC}_i - \text{AIC}_{\min}$$

where AIC_i is the AIC of model i and AIC_{\min} is the AIC of the model with the smallest AIC.

AICc

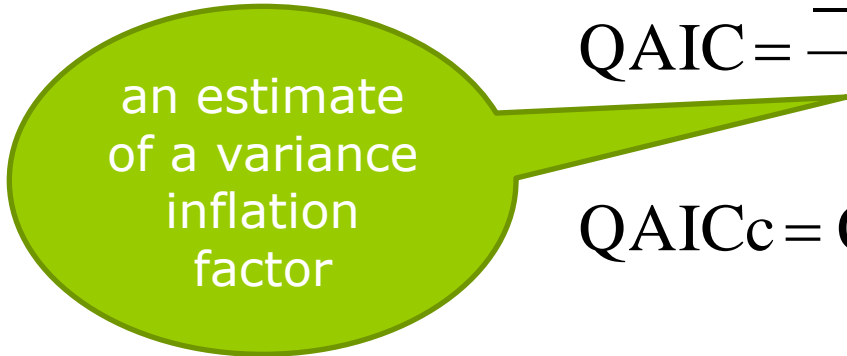
- If the sample size, n , is small an adapted AIC is recommended:

$$\text{AICc} = \text{AIC} + \frac{2K(K+1)}{n-K-1}$$

for $n > K+1$.

QAIC/QAICc

- If **overdispersion** exists, that is, *the sample variance exceeds the theoretical (model-based) variance*, then quasi-likelihood theory has been employed to modify the AIC (and AICc), which result respectively in the QAIC (and QAICc), defined by



an estimate
of a variance
inflation
factor

$$\text{QAIC} = \frac{-2 \log L}{\hat{c}} + 2K$$

$$\text{QAICc} = \text{QAIC} + \frac{2K(K+1)}{n-K-1}$$

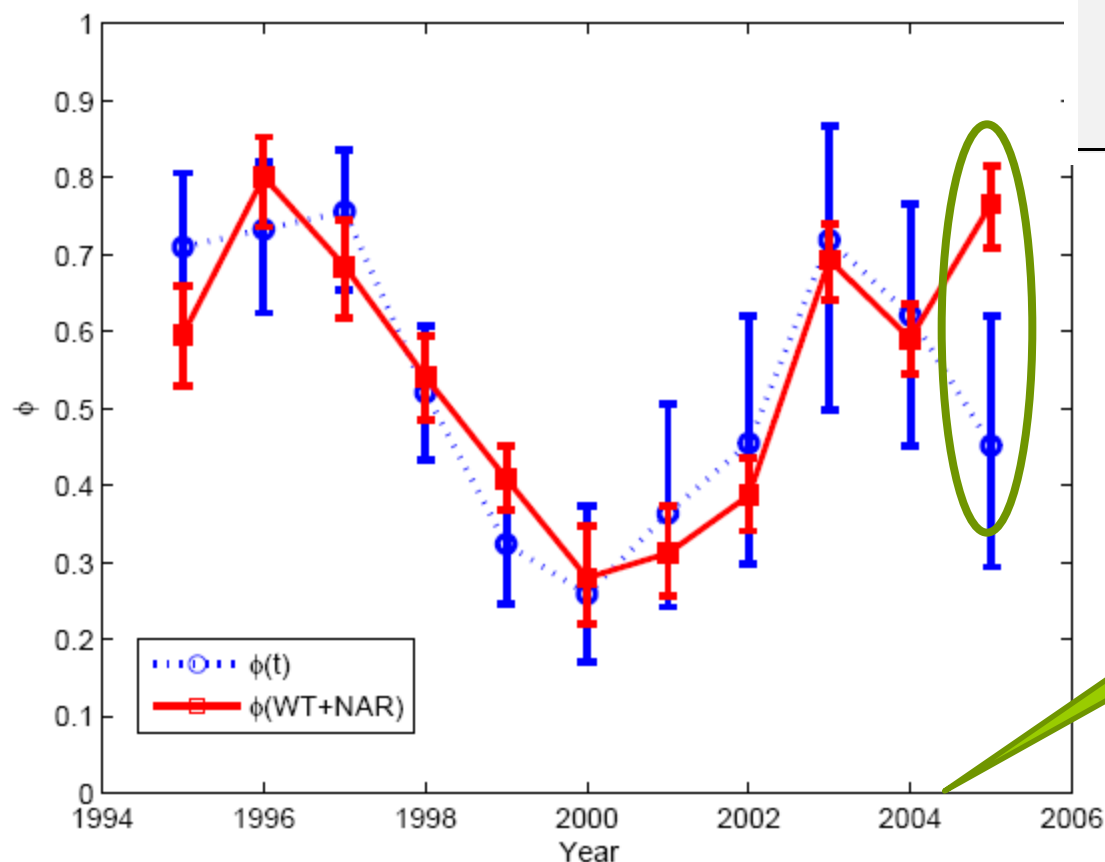
Great Crested Newt Model Selection

| Code | Model | QAICc | Δ QAICc | n |
|------|---------------------------|---------|----------------|----|
| 1.1 | $\phi(t), p(pond + sex)$ | 1941.62 | 0.00 | 16 |
| 1.2 | $\phi(t), p(pond)$ | 1941.86 | 0.24 | 15 |
| 1.3 | $\phi(t), p(pond * sex)$ | 1942.60 | 0.98 | 19 |
| 1.4 | $\phi(t), p(pond * time)$ | 1945.99 | 4.37 | 52 |

Great Crested Newt Model Selection

| Code | Model | QAICc | Δ QAICc | n |
|------|---------------------------|---------|----------------|----|
| 1.1 | $\phi(t), p(pond + sex)$ | 1941.62 | 0.00 | 16 |
| 1.2 | $\phi(t), p(pond)$ | 1941.86 | 0.24 | 15 |
| 1.3 | $\phi(t), p(pond * sex)$ | 1942.60 | 0.98 | 19 |
| 1.4 | $\phi(t), p(pond * time)$ | 1945.99 | 4.37 | 52 |

$\phi(t)$ versus $\phi(\text{WT+NAR})$



| Model | QAICc | Delta QAICc |
|--|---------|-------------|
| $\phi(\text{WT} + \text{NAR} + t2005), p(\text{site})$ | 1940.21 | 0.00 |
| $\phi(\text{time}), p(\text{site} + \text{sex})$ | 1941.62 | 1.41 |
| $\phi(\text{time}), p(\text{site})$ | 1941.86 | 1.65 |
| $\phi(\text{time}), p(\text{sex} + \text{site})$ | 1942.60 | 2.39 |
| $\phi(\text{WT} + \text{SR} + \text{NAR} + \text{WGF} + t2005), p(\text{site})$ | 1942.78 | 2.57 |
| $\phi(\text{time}), p(\text{site} + \text{time})$ | 1945.99 | 5.78 |
| $\phi(\text{WT} + \text{SR} + \text{NAR} + \text{WGF}), p(\text{site})$ | 1947.09 | 6.88 |
| $\phi(\text{WT} + \text{NAR} + \text{WGF}), p(\text{site})$ | 1947.26 | 7.06 |
| $\phi(\text{WT} + \text{SR} + \text{NAR} + \text{WGF}), p(\text{site} + \text{sex})$ | 1947.37 | 7.16 |
| $\phi(\text{WT} + \text{SR} + \text{NAR}), p(\text{site})$ | 1948.04 | 7.84 |
| $\phi(\text{WT} + \text{SR} + \text{NAR} + \text{WGF}), p(\text{site} + \text{sex})$ | 1948.48 | 8.27 |
| $\phi(\text{WT} + \text{NAR}), p(\text{site})$ | 1949.16 | 8.95 |
| $\phi(\text{sex} + \text{time}), p(\text{site})$ | 1950.53 | 10.33 |

Changes to ponds

Model selection using hypothesis tests

Consider testing the null hypothesis $H_0 : \theta = \theta_0$ versus the alternative $H_1 : \theta \neq \theta_0$. Denote the likelihood obtained from data x by $L(\theta | x)$, then the test statistic for a likelihood ratio test (LRT) is

$$\lambda = \frac{L(\theta_0 | x)}{L(\hat{\theta} | x)}$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ . If the null hypothesis is true, then subject to certain regularity conditions, asymptotically, $-2 \log \lambda$ has a χ_d^2 distribution, where d is the difference in the numbers of parameters between the two models.

Score tests for comparing models

The score test is due to Rao (1948). Suppose we want to test $H_0 : \theta = \theta_0$.

In a likelihood-ratio test we compare $\hat{\theta}$ with θ_0 through the corresponding likelihood values. In a score test we compare through **the values taken by the score statistic**, or gradient.

The advantage of this is that **we do not need to evaluate $\hat{\theta}$** . Note examples such as **Cochran-Armitage** and **log-rank tests**.

Score tests

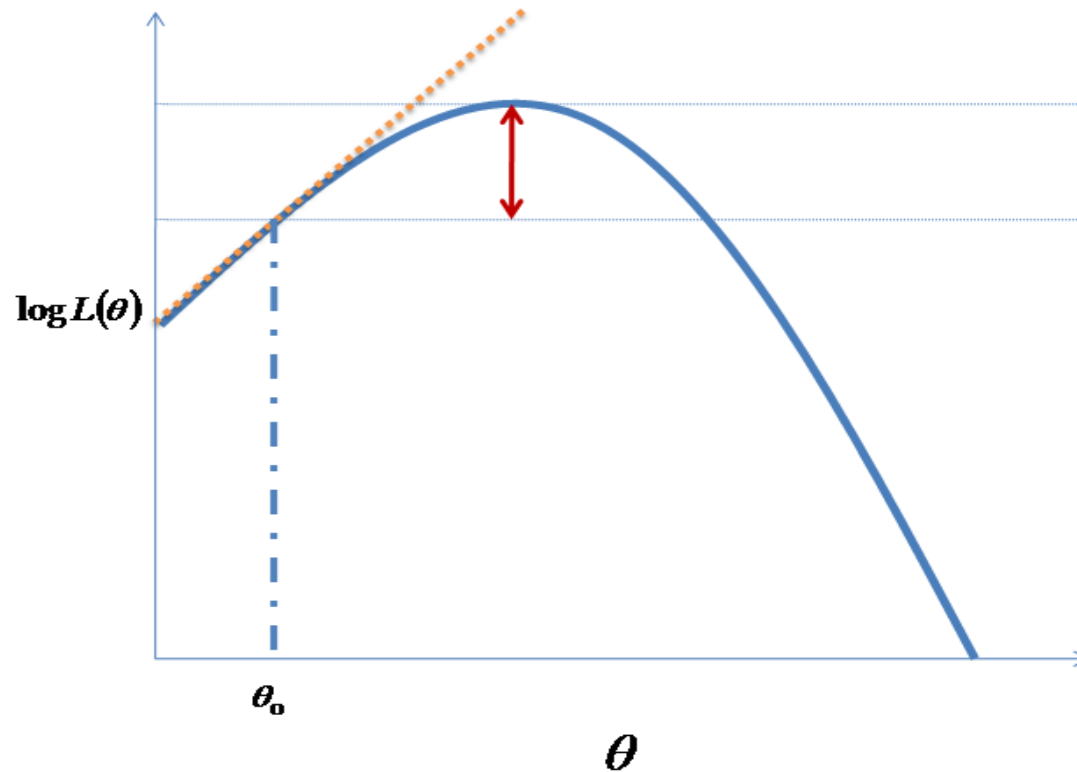
Suppose we wish to test the null hypothesis $H_0 : \varphi = \varphi_0$

The score test statistic is given by $z = U'J^{-1}U$ in which U is the scores vector defined by $U_i = \partial l / \partial \theta_i$ and J is the Fisher Information matrix, given by

$$J_{ij} = E\left(\partial^2 l / \partial \theta_i \partial \theta_j\right)$$

Under the null hypothesis, $z \sim \chi_d$.

Graphical illustration of score test



Comparing nested models

- When we compare nested models using a score test, **we only need to fit the simpler of the two models.**
- This is a clear advantage for when the data do not support the more complex model. Of course if the test is significant we need to fit the complex model, but we can do so with confidence.

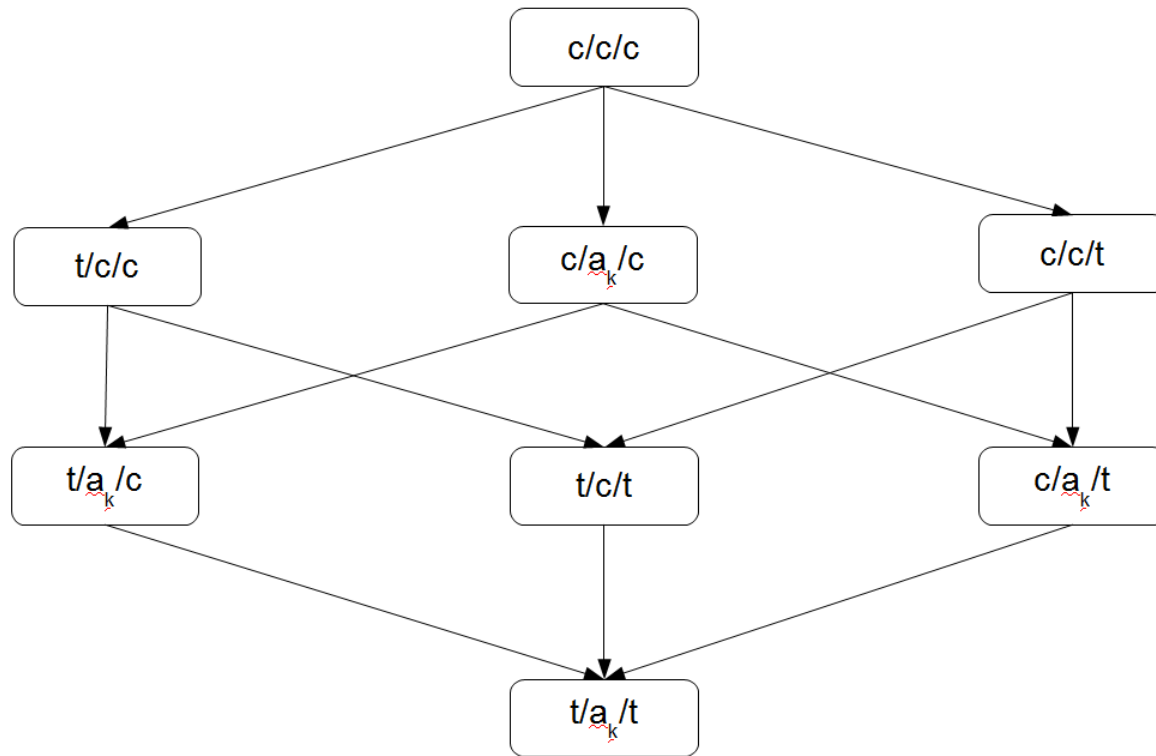
Pathways for ring-recovery model selection

Level 1

Level 2

Level 3

Level 4



Illustrative comparisons: importance of using the expected information matrix

| Score | | LR |
|--------------|------------|-----------|
| Obs | Exp | |
| 32.71 | 21.16 | 23.53 |
| 5.44 | 5.39 | 5.71 |
| 2.44 | 2.40 | 2.42 |
| 4.38 | 4.80 | 4.80 |
| 21.66 | 6.60 | 6.30 |
| 1.47 | 8.92 | 8.95 |

Example: Cormorant ring-recovery data

| Comparison | S | LR | df | P (score) | Model | AIC |
|----------------------------------|--------|--------|----|------------------------|--------------------------|--------|
| | | | | | <i>c/c/c</i> | 518.02 |
| <i>c/c/c : c/c/t</i> | 104.96 | 134.83 | 11 | 1.85×10^{-17} | <i>c/c/t</i> | 405.19 |
| <i>c/c/c : t/c/c</i> | 31.69 | 36.83 | 10 | 4.50×10^{-4} | <i>t/c/c</i> | 501.19 |
| <i>c/c/c : c/a₂/c</i> | 0.09 | 0.09 | 1 | 0.77 | <i>c/a₂/c</i> | 519.93 |
| <i>c/c/t : t/c/t</i> | 30.51 | 27.84 | 10 | 7.05×10^{-4} | <i>t/c/t</i> | 397.35 |
| <i>c/c/t : c/a₂/t</i> | 0.19 | 0.19 | 1 | 0.66 | <i>c/a₂/t</i> | 407.00 |
| <i>t/c/t : t/a₂/t</i> | 0.30 | 0.31 | 1 | 0.59 | <i>t/a₂/t</i> | 399.04 |

Example: Cormorant ring-recovery data

| Comparison | S | LR | df | P (score) | Model | AIC |
|-------------------|--------|--------|----|------------------------|-----------|--------|
| $c/c/c : c/c/t$ | 104.96 | 134.83 | 11 | 1.85×10^{-17} | $c/c/c$ | 518.02 |
| $c/c/c : t/c/c$ | 31.69 | 36.83 | 10 | 4.50×10^{-4} | $c/c/t$ | 405.19 |
| $c/c/c : c/a_2/c$ | 0.09 | 0.09 | 1 | 0.77 | $t/c/c$ | 501.19 |
| $c/c/t : t/c/t$ | 30.51 | 27.84 | 10 | 7.05×10^{-4} | $c/a_2/c$ | 519.93 |
| $c/c/t : c/a_2/t$ | 0.19 | 0.19 | 1 | 0.66 | $t/c/t$ | 397.35 |
| $t/c/t : t/a_2/t$ | 0.30 | 0.31 | 1 | 0.59 | $c/a_2/t$ | 407.00 |
| | | | | | $t/a_2/t$ | 399.04 |

Absolute Goodness-of-fit Tests

- ❑ Absolute goodness-of-fit measures the fit of the final selected model.
- ❑ Why do we need to assess this?
 - All of the models in the model set may not be appropriate for the data
 - Underlying violation of model assumptions, which has not been detected by diagnostic goodness-of-fit step.

Cormorant Model: t/c/t

[illegible][illegible]

Cormorant Model: t/c/t

Observed m-array

$\chi^2 = 57.27$
df = 53
p-value = 0.32

Expected m-array

Observed m-array

| | | | | | | | | | | | |
|-----|-----|-----|-----|-----|----|----|----|----|----|----|----|
| 117 | 10 | 11 | 9 | 2 | 4 | 6 | 2 | 1 | 4 | 4 | 0 |
| | 120 | 13 | 18 | 8 | 7 | 8 | 4 | 7 | 8 | 7 | 1 |
| | | 160 | 16 | 13 | 6 | 10 | 1 | 1 | 1 | 1 | 1 |
| | | | 267 | 16 | 10 | 10 | 1 | 1 | 1 | 1 | 1 |
| | | | | 188 | 11 | 11 | 1 | 1 | 1 | 1 | 1 |
| | | | | | 23 | 11 | 1 | 1 | 1 | 1 | 1 |
| | | | | | | 11 | 1 | 1 | 1 | 1 | 1 |
| | | | | | | | 11 | 1 | 1 | 1 | 1 |
| | | | | | | | | 11 | 1 | 1 | 1 |
| | | | | | | | | | 11 | 1 | 1 |
| | | | | | | | | | | 11 | 1 |
| | | | | | | | | | | | 11 |

Expected m-array

| | | | | | | | | | | | |
|-----|--------|--------|-------|--------|--------|--------|--------|------|-------|-------|------|
| 117 | 12.02 | 7.65 | 7.07 | 5.52 | | 4.38 | 3.19 | 2.51 | 2.36 | 3.27 | 0.18 |
| | 118.52 | 13 | 12 | 9.38 | 8.24 | 7.44 | 5.41 | 4.27 | 4 | 5.55 | 0.31 |
| | | 162.75 | 13.82 | 11.8 | 9.49 | 8.57 | 6.23 | 4.91 | 4.61 | 6.4 | 0.36 |
| | | | 275.6 | 17.86 | 15.69 | 14.17 | 10.31 | 8.13 | 7.63 | 10.58 | 0.59 |
| | | | | 184.14 | 14.5 | 13.09 | 9.53 | 7.51 | 7.05 | 9.78 | 0.54 |
| | | | | | 227.42 | 12.74 | 9.27 | 7.31 | 6.86 | 9.52 | 0.53 |
| | | | | | | 107.04 | 7.36 | 5.8 | 5.44 | 7.55 | 0.42 |
| | | | | | | | 135.32 | 6.08 | 5.71 | 7.92 | 0.44 |
| | | | | | | | | 63.8 | 3.59 | 4.98 | 0.28 |
| | | | | | | | | | 57.65 | 10.16 | 0.57 |
| | | | | | | | | | | 42.24 | 0.78 |

$\chi^2 = 57.27$
df = 53
p-value = 0.32

Example: Cormorant Ring-Recovery

| Model | AIC | X^2 | df | P (X^2) |
|-----------|--------|--------|----|------------------------|
| $c/c/c$ | 518.02 | 188.30 | 74 | 6.35×10^{-12} |
| $c/c/t$ | 405.19 | 85.71 | 63 | 0.03 |
| $t/c/c$ | 501.19 | 149.10 | 64 | 9.57×10^{-9} |
| $c/a_2/c$ | 519.93 | 188.40 | 73 | 3.79×10^{-12} |
| $t/c/t$ | 397.35 | 57.27 | 53 | 0.32 |
| $c/a_2/t$ | 407.00 | 85.78 | 62 | 0.02 |
| $t/a_2/t$ | 399.04 | 56.86 | 52 | 0.30 |

M-array for Goodness-of-fit Assessment

- ❑ We have seen that the ring-recovery m-array can be used as a set of sufficient statistics to assess absolute goodness-of-fit
- ❑ Similarly for single-site recapture data, the corresponding m-array can be used
- ❑ What about more general models?
 - Multi-state recapture m-arrays
 - Integrated data?

Sufficient statistics

- ▣ King and Brooks (2003) proposed a closed-form likelihood for age or time-dependent multi-state capture-recapture-recovery models
- ▣ Likelihood is product multinomial
- ▣ Sufficient statistics can be used to assess absolute goodness-of-fit
 - Pearson X^2 or Likelihood Ratio G^2 statistic
- ▣ Note: The Catchpole-Morgan sufficient statistics are **not multinomial**, therefore cannot be used in this way

Diagnostic Goodness-of-fit Tests

- ❑ Model Assumptions are necessary for all models presented here

- ❑ Cormack-Jolly-Seber Model:
 - Every marked animal present in the population at sampling time t_i has the same probability of being recaptured;
 - Every marked animal present in the population immediately following the sampling at time t_i has the same probability of survival until sampling time t_{i+1} ;
 - Marks are neither lost nor overlooked, and are recorded correctly;
 - Sampling periods are instantaneous and recaptured animals are released immediately;
 - All emigration from the sampled area is permanent.
 - The fate of each animal with respect to capture and survival is independent of the fate of any other animal.

Diagnostic tests

- ❑ Diagnostic goodness-of-fit tests seek to detect violations of model assumptions within the data set
- ❑ Carried out as a preliminary analysis prior to model fitting
- ❑ Should guide the model structure used for model fitting
- ❑ Diagnostic tests exist for single and multi-site mark-recapture data

Existing diagnostic tests

- ❑ **Test 1:** detects group effects
- ❑ **Test 2:** detects differences in future encounters between individuals encountered and not at a given occasion
- ❑ **Test 3:** detects differences in future encounters between `new' (newly marked) and `old' (recaptured) individuals

Partitioning

- Test 2 and Test 3 are **contingency table homogeneity tests** based upon partitioned sections of the m-array and generalised m-array
- Possible to partition the tests into biologically interpretable components:
 - Test C, a subcomponent of Test 2 tests for immediate **trap-dependence**
 - Test T, a subcomponent of Test 3 tests for **transient individuals**

Software to perform GoF tests

- ❑ Program Release: Run within Program Mark
 - Tests for single-site capture-recapture data
- ❑ Software U-Care
 - Test for single-site capture-recapture data
 - Tests for multi-site capture-recapture data

Cormorant Goodness-of-fit Tests

| Test | Component Test | χ^2 | <i>df</i> | P-Value |
|--------------|----------------|----------|-----------|---------|
| 2 | | | | |
| | 2.CT | 15.61 | 2 | < 0.001 |
| | 2.CL | 0.00 | 1 | 1.000 |
| 3 | | | | |
| | 3.SR | 20.38 | 3 | < 0.001 |
| | 3.SM | 0.99 | 2 | 0.610 |
| Total | | 36.98 | 8 | <0.001 |

Trap
dependence
test

Transience
test

Cormorant Goodness-of-fit Tests

| Test | Component Test | χ^2 | df | P-Value |
|--------------|----------------|--------------|----------|------------------|
| 2 | | | | |
| | 2.CT | 15.61 | 2 | < 0.001 |
| | 2.CL | 0.00 | 1 | 1.000 |
| 3 | | | | |
| | 3.SR | 20.38 | 3 | < 0.001 |
| | 3.SM | 0.99 | 2 | 0.610 |
| Total | | 36.98 | 8 | <0.001 |

$$\hat{c} = \frac{36.98}{8}$$

Should adapt model to account for trap-dependence and transience, rather than just using QAIC

Extensions

- ▣ Adaptation of goodness-of-fit tests for joint recapture and recovery data and recovery data alone
 - McCrea, Morgan and Pradel (2011) In prep.

Diagnostic vs Absolute

- ❑ Diagnostic tests do not involve fitting any models:
 - Examine properties of the data
 - Defined for time-dependent models only

- ❑ Absolute goodness-of-fit
 - Fit your best model and then compare observed and expected values
 - Completely general: cohort, age, state and time-dependence
 - Potentially allow you to compare the fit of models including heterogeneous capture/survival

References

- Burnham and Anderson (2002) *Model selection and multimodel inference: A practical information theoretic approach*.
- Buse, *The American Statistician*, **36**, 1982.
- Catchpole and Morgan, *Biometrics*, **52**, 1996.
- Catchpole, Morgan, Freeman and Peach, *Bird Study* 46 Supplement, 1999.
- McCrea and Morgan, *Biometrics*, In press.
- Morgan, Palmer and Ridout, *The American Statistician*, **61**, 2007.
- Rao, *Proc. Camb. Phil. Soc.*, **44**, 1948.

References

- Burnham (1993) In *The study of bird population dynamics using marked individuals*.
- Choquet et al. (2009) *Ecography*, **32**: 1071-1074
- McCrea et al. (2010) *Journal of Ornithology*. In press.
- Pollock et al. (1985) *Biometrics*, **41**: 399-410
- Pradel et al. (2003) *Biometrics*, **59**: 43-53
- Pradel et al. (2005) *ABC*, **28**: 189-204.

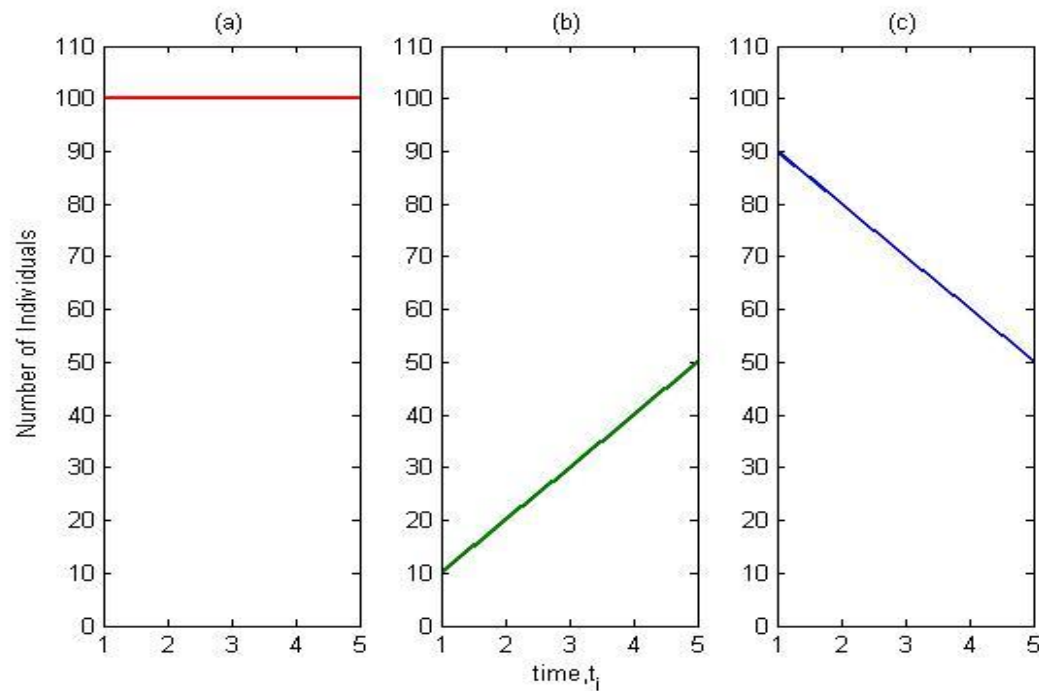
SECTION 5

COMPLEX MODELS

Complex models

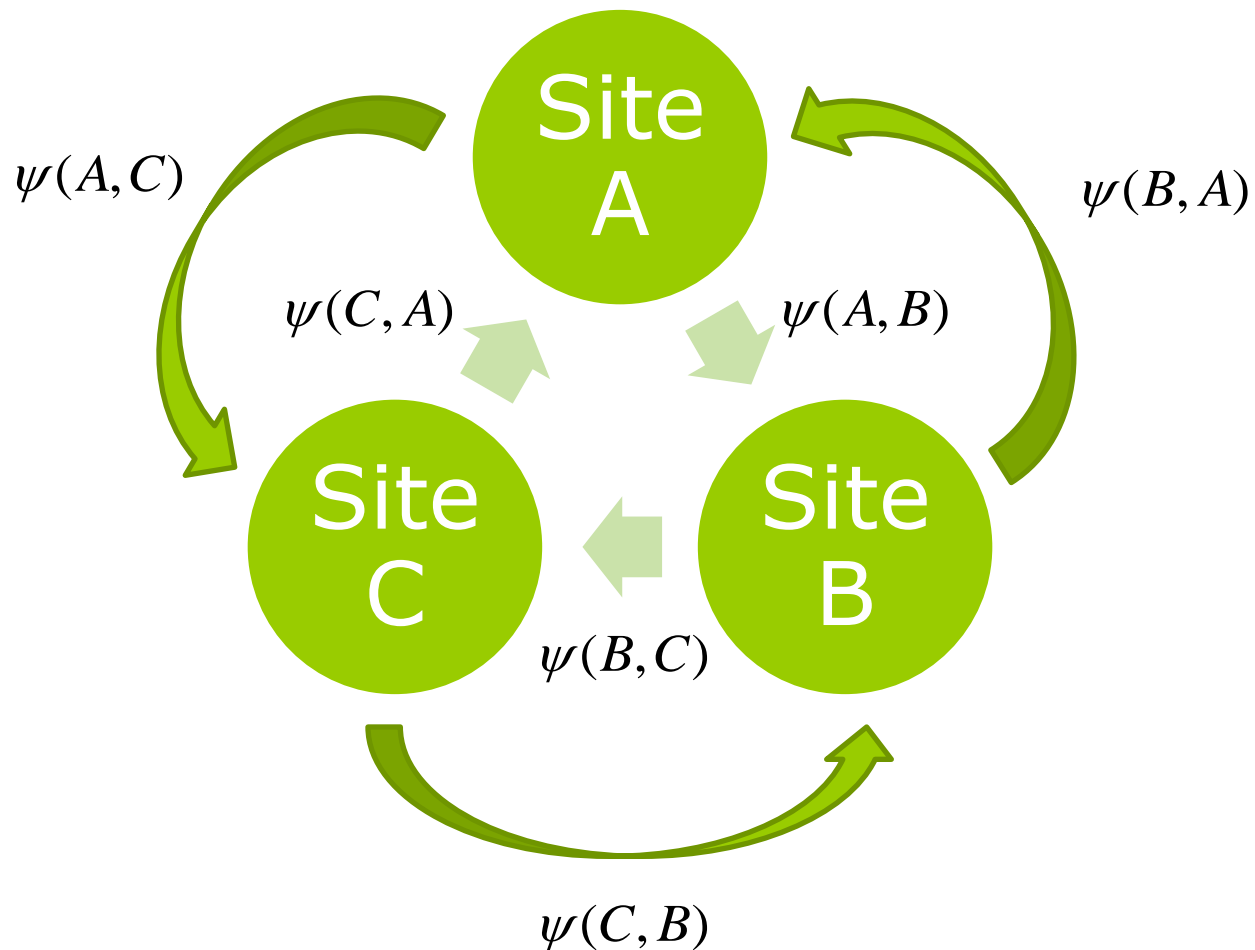
- ▣ 5.1 Multi-state models
- ▣ 5.2 Describing survival through covariates

Multi-state models

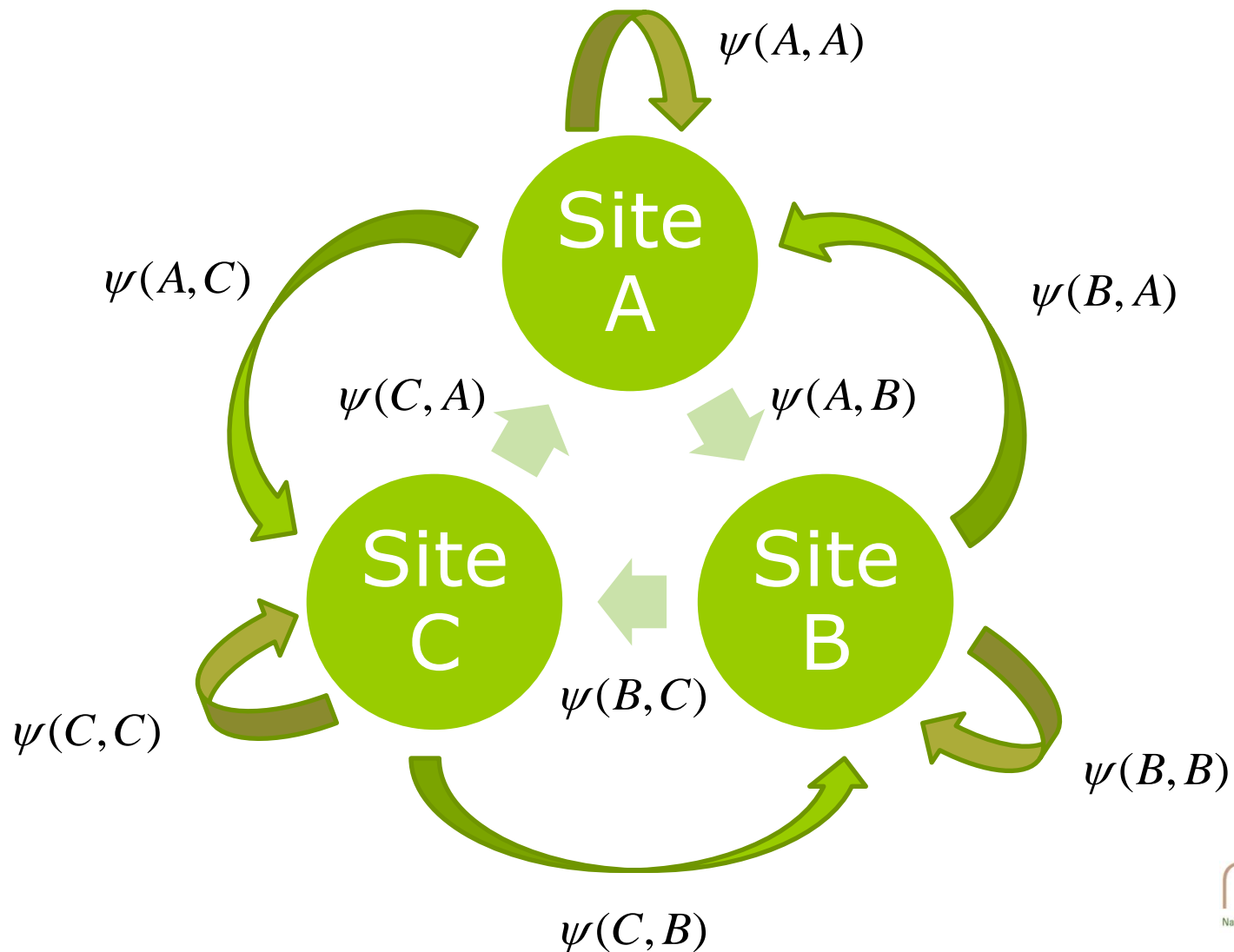


- (a) Total Population
- (b) Population at site 1
- (c) Population at site 2

Arnason-Schwarz Model



Account for Movement

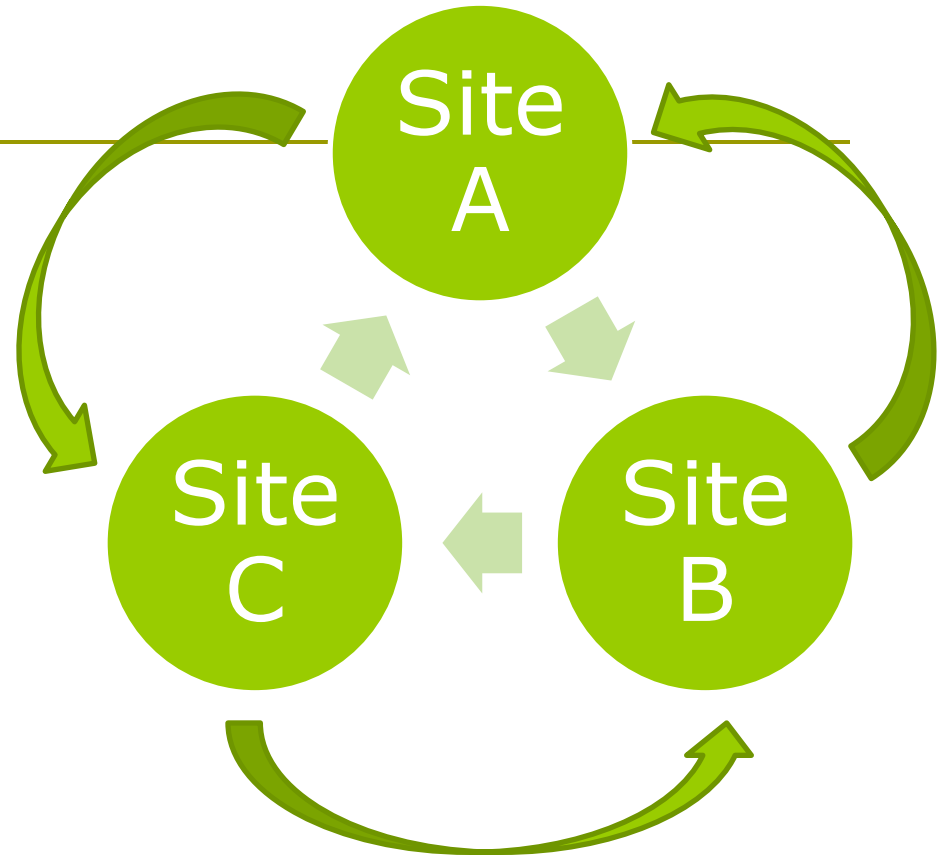


Movement Ctd.

$$\begin{bmatrix} \psi(A, A) & \psi(A, B) & \psi(A, C) \\ \psi(B, A) & \psi(B, B) & \psi(B, C) \\ \psi(C, A) & \psi(C, B) & \psi(C, C) \end{bmatrix}$$

□ “Closed” Population

- Transitions sum to 1: $\psi(A, A) = 1 - \psi(A, B) - \psi(A, C)$



Multisite Encounter Histories

- 1 0 1 1 0 0
- 1 1 0 1 0 0
- 1 1 0 0 0 0
- 1 0 1 0 0 0
- . . .

- A 0 B B 0 0
- B A 0 B 0 0
- A A 0 0 0 0
- B 0 B 0 0 0
- . . .

Multisite M-Array

| Number Released | Recaptured t_2 | | Recaptured t_3 | | Recaptured t_4 | | Never Recaptured |
|-----------------|------------------|---------------|------------------|---------------|------------------|---------------|------------------|
| | A | B | A | B | A | B | |
| $R_1(A)$ | $m_{12}(A,A)$ | $m_{12}(A,B)$ | $m_{13}(A,A)$ | $m_{13}(A,B)$ | $m_{14}(A,A)$ | $m_{14}(A,B)$ | $R_1(A)-\Sigma$ |
| $R_1(B)$ | $m_{12}(B,A)$ | $m_{12}(B,B)$ | $m_{13}(B,A)$ | $m_{13}(B,B)$ | $m_{14}(B,A)$ | $m_{14}(B,B)$ | $R_1(B)-\Sigma$ |
| $R_2(A)$ | | | $m_{23}(A,A)$ | $m_{23}(A,B)$ | $m_{24}(A,A)$ | $m_{24}(A,B)$ | $R_2(A)-\Sigma$ |
| $R_2(B)$ | | | $m_{23}(B,A)$ | $m_{23}(B,B)$ | $m_{24}(B,A)$ | $m_{24}(B,B)$ | $R_2(B)-\Sigma$ |
| $R_3(A)$ | | | | | $m_{34}(A,A)$ | $m_{34}(A,B)$ | $R_3(A)-\Sigma$ |
| $R_3(B)$ | | | | | $m_{34}(B,A)$ | $m_{34}(B,B)$ | $R_3(B)-\Sigma$ |

- $R_i(r)$: Number of individuals released at time t_i in site r
- $m_{ij}(r,s)$: Number of individuals released at time t_i in site r and next recaptured at time t_j in site s .

Encounter History to M-Array

| Number Released | Recaptured t_2 | | Recaptured t_3 | | Recaptured t_4 | | Never Recaptured |
|-----------------|------------------|---------------|------------------|---------------|------------------|---------------|-------------------|
| | A | B | A | B | A | B | |
| $R_1(A)$ | $m_{12}(A,A)$ | $m_{12}(A,B)$ | $m_{13}(A,A)$ | $m_{13}(A,B)$ | $m_{14}(A,A)$ | $m_{14}(A,B)$ | $R_1(A) - \Sigma$ |
| $R_1(B)$ | $m_{12}(B,A)$ | $m_{12}(B,B)$ | $m_{13}(B,A)$ | $m_{13}(B,B)$ | $m_{14}(B,A)$ | $m_{14}(B,B)$ | $R_1(B) - \Sigma$ |
| $R_2(A)$ | | | $m_{23}(A,A)$ | $m_{23}(A,B)$ | $m_{24}(A,A)$ | $m_{24}(A,B)$ | $R_2(A) - \Sigma$ |
| $R_2(B)$ | | | $m_{23}(B,A)$ | $m_{23}(B,B)$ | $m_{24}(B,A)$ | $m_{24}(B,B)$ | $R_2(B) - \Sigma$ |
| $R_3(A)$ | | | | | $m_{34}(A,A)$ | $m_{34}(A,B)$ | $R_3(A) - \Sigma$ |
| $R_3(B)$ | | | | | $m_{34}(B,A)$ | $m_{34}(B,B)$ | $R_3(B) - \Sigma$ |

■ A0BB

Matrix Multisite M-array

| Number Released | Recaptured t2 | Recaptured t3 | Recaptured t4 |
|-----------------|---------------|---------------|---------------|
| R_1 | M_{12} | M_{13} | M_{14} |
| R_2 | | M_{23} | M_{24} |
| R_3 | | | M_{34} |

$$\mathbf{R}_i = \begin{bmatrix} R_i(A) \\ R_i(B) \end{bmatrix}$$

$$\mathbf{M}_{ij} = \begin{bmatrix} m_{ij}(A, A) & m_{ij}(A, B) \\ m_{ij}(B, A) & m_{ij}(B, B) \end{bmatrix}$$

Parameters

- $\phi_i(r)$: probability of an animal alive at time t_i in site r , survives until time t_{i+1}
- $p_{i+1}(s)$: probability of an animal alive in site s at time t_{i+1} being recaptured
- $\psi_i(r,s)$: probability of an animal alive in site r at time t_i moving to site s by time t_{i+1}

Multisite Probabilities

| Number Released | Recaptured t_2 | Recaptured t_3 | Recaptured t_4 |
|-----------------|---------------------|---------------------------------------|---|
| R_1 | $\Phi_1 \Psi_1 P_2$ | $\Phi_1 \Psi_1 Q_2 \Phi_2 \Psi_2 P_3$ | $\Phi_1 \Psi_1 Q_2 \Phi_2 \Psi_2 Q_3 \Phi_3 \Psi_3 P_4$ |
| R_2 | | $\Phi_2 \Psi_2 P_3$ | $\Phi_2 \Psi_2 Q_3 \Phi_3 \Psi_3 P_4$ |
| R_3 | | | $\Phi_3 \Psi_3 P_4$ |

| Number Released | Recaptured t_2 | Recaptured t_3 | Recaptured t_4 |
|-----------------|------------------|------------------|------------------|
| R_1 | M_{12} | M_{13} | M_{14} |
| R_2 | | M_{23} | M_{24} |
| R_3 | | | M_{34} |

Multisite Probabilities

| Number Released | Recaptured t_2 | Recaptured t_3 | Recaptured t_4 |
|-----------------|---------------------|---------------------------------------|---|
| R_1 | $\Phi_1 \Psi_1 P_2$ | $\Phi_1 \Psi_1 Q_2 \Phi_2 \Psi_2 P_3$ | $\Phi_1 \Psi_1 Q_2 \Phi_2 \Psi_2 Q_3 \Phi_3 \Psi_3 P_4$ |
| R_2 | | $\Phi_2 \Psi_2 P_3$ | $\Phi_2 \Psi_2 Q_3 \Phi_3 \Psi_3 P_4$ |
| R_3 | | | $\Phi_3 \Psi_3 P_4$ |

$$\Phi_i = \begin{bmatrix} \phi_i(A) & 0 \\ 0 & \phi_i(B) \end{bmatrix}$$

Multisite Probabilities

| Number Released | Recaptured t_2 | Recaptured t_3 | Recaptured t_4 |
|-----------------|---------------------|---------------------------------------|---|
| R_1 | $\Phi_1 \Psi_1 P_2$ | $\Phi_1 \Psi_1 Q_2 \Phi_2 \Psi_2 P_3$ | $\Phi_1 \Psi_1 Q_2 \Phi_2 \Psi_2 Q_3 \Phi_3 \Psi_3 P_4$ |
| R_2 | | $\Phi_2 \Psi_2 P_3$ | $\Phi_2 \Psi_2 Q_3 \Phi_3 \Psi_3 P_4$ |
| R_3 | | | $\Phi_3 \Psi_3 P_4$ |

$$\Psi_i = \begin{bmatrix} \psi_i(A,A) & \psi_i(A,B) \\ \psi_i(B,A) & \psi_i(B,B) \end{bmatrix}$$

Multisite Probabilities

| Number Released | Recaptured t_2 | Recaptured t_3 | Recaptured t_4 |
|-----------------|---------------------|---------------------------------------|---|
| R_1 | $\Phi_1 \Psi_1 P_2$ | $\Phi_1 \Psi_1 Q_2 \Phi_2 \Psi_2 P_3$ | $\Phi_1 \Psi_1 Q_2 \Phi_2 \Psi_2 Q_3 \Phi_3 \Psi_3 P_4$ |
| R_2 | | $\Phi_2 \Psi_2 P_3$ | $\Phi_2 \Psi_2 Q_3 \Phi_3 \Psi_3 P_4$ |
| R_3 | | | $\Phi_3 \Psi_3 P_4$ |

$$P_i = \begin{bmatrix} p_i(A) & 0 \\ 0 & p_i(B) \end{bmatrix}$$

$$Q_i = \begin{bmatrix} 1-p_i(A) & 0 \\ 0 & 1-p_i(B) \end{bmatrix}$$

Simulated Data Example

- One Cohort – no new marking during the study

- 100 individuals in total – equally spread between two sites

| | | | | | | | | | | |
|----|----|----|---|---|---|---|---|---|---|---|
| 50 | 12 | 4 | 5 | 2 | 4 | 0 | 2 | 0 | 0 | 0 |
| 50 | 2 | 13 | 3 | 6 | 3 | 1 | 0 | 1 | 0 | 0 |
| 14 | | | 5 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 17 | | | 3 | 2 | 1 | 2 | 0 | 0 | 0 | 1 |
| 16 | | | | | 2 | 1 | 2 | 1 | 0 | 2 |
| 11 | | | | | 0 | 0 | 3 | 1 | 1 | 0 |
| 10 | | | | | | | 3 | 0 | 2 | 1 |
| 5 | | | | | | | 1 | 0 | 0 | 0 |
| 11 | | | | | | | | | 6 | 0 |
| 3 | | | | | | | | | 0 | 0 |

Estimates from Fitted Model

Model fitted

- Site-dependent survival
 - $\phi(A)$
 - $\phi(B)$
- Site-dependent movement
 - $\psi(A,B)$
 - $\psi(B,A)$
- Constant capture probabilities
 - p
- *Note:* No time-dependence
- 5 parameters in total

Parameter estimates

| Parameter | True Value | MLE | SE |
|-------------|------------|-------|--------|
| $\phi(A)$ | 0.8 | 0.791 | 0.0490 |
| $\phi(B)$ | 0.8 | 0.751 | 0.0568 |
| $\psi(A,B)$ | 0.2 | 0.192 | 0.0506 |
| $\psi(B,A)$ | 0.3 | 0.295 | 0.0671 |
| p | 0.4 | 0.407 | 0.0458 |

Estimates from Fitted Model

Model fitted

- Site-dependent survival
 - $\phi(A)$
 - $\phi(B)$
- Site-dependent movement
 - $\psi(A,B)$
 - $\psi(B,A)$
- Constant capture probabilities
 - p
- *Note:* No time-dependence
- 5 parameters in total

Parameter estimates

| Parameter | True Value | MLE | SE |
|-------------|------------|-------|--------|
| $\phi(A)$ | 0.8 | 0.791 | 0.0490 |
| $\phi(B)$ | 0.8 | 0.751 | 0.0568 |
| $\psi(A,B)$ | 0.2 | 0.192 | 0.0506 |
| $\psi(B,A)$ | 0.3 | 0.295 | 0.0671 |
| p | 0.4 | 0.407 | 0.0458 |

MultiSTATE

- Does not have to be a movement from one site to another that is being modelled.
- Perhaps wish to study whether or not an individual is classified as a “breeder” or “non-breeder” in a particular year.
- To do this, simply label your states:
 - A=Breeder
 - B=Non-Breeder
- Biological importance to measure these transitions.

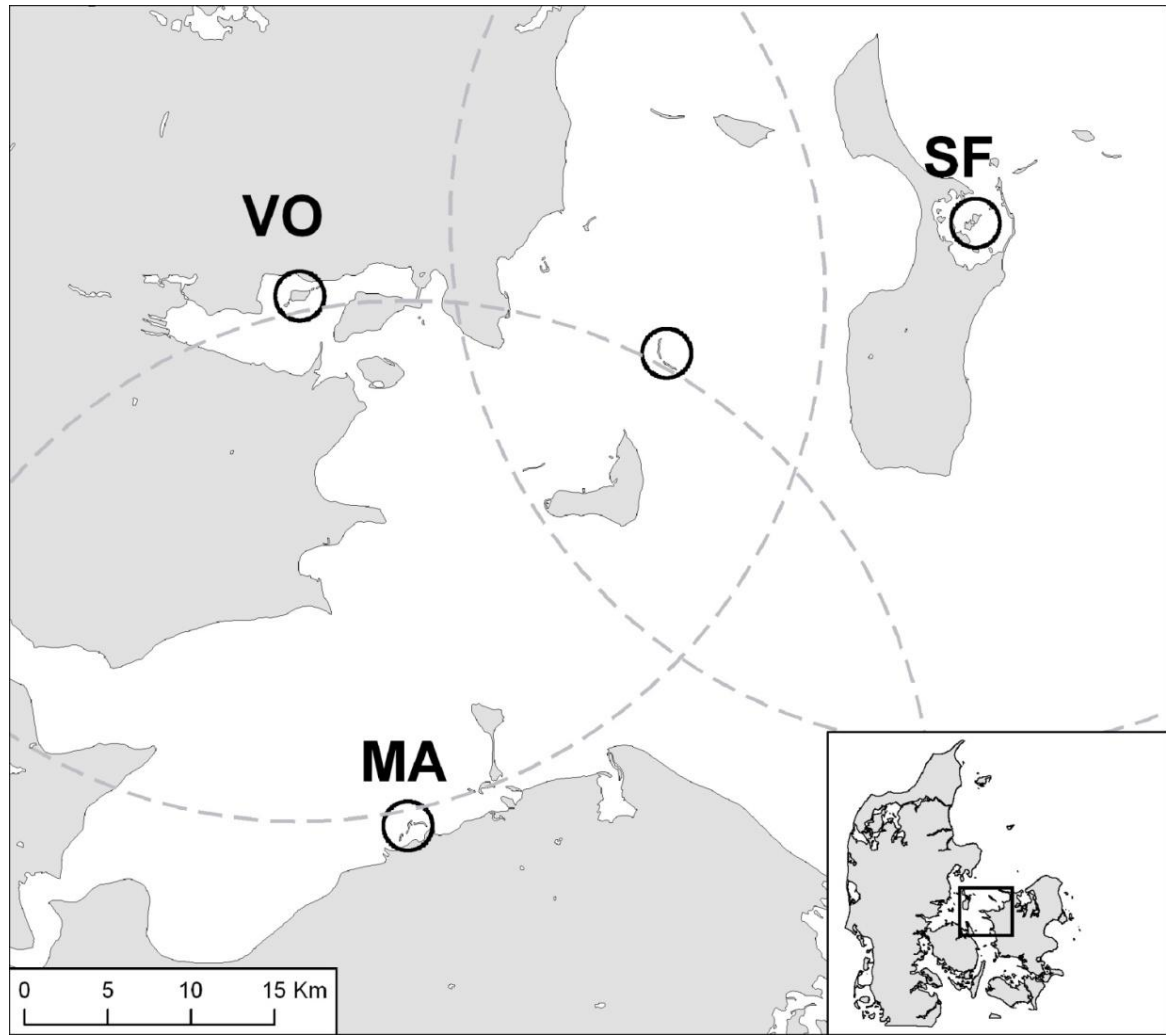
Example: Great Cormorants



Phalacrocorax carbo sinensis

Three neighbouring colonies:

- Vorsø (VO)
- Mågeøerne (MA)
- Stavns Fjord (SF)



Recapture Data



- **Recapture data collected between 1989 and 1994**
- **Initial ringing was carried out on non-breeding individuals**
- **Birds were not recaptured until they had become breeding individuals**
- **Must model transition between non-breeding and breeding states as well as geographical movements between the colonies**

What transitions must be considered?

- We have a total of six states:
 - Non-breeders & Breeders at each of the sites
- Background biology:
 - Once an individual has attained a breeding status, it remains in the breeding state for the remainder of its life



Transitions



- ❑ **Natal Dispersal:** movement between the colonies whilst a non-breeder
- ❑ **Recruitment:** transition from non-breeding to breeding state
- ❑ **Breeding Dispersal:** movement between the colonies whilst a breeder

Diagram showing transitions of interest

Reminder of Transitions

Natal Dispersal:
movement between the colonies whilst a non-breeder

Recruitment:
transition from non-breeding to breeding state

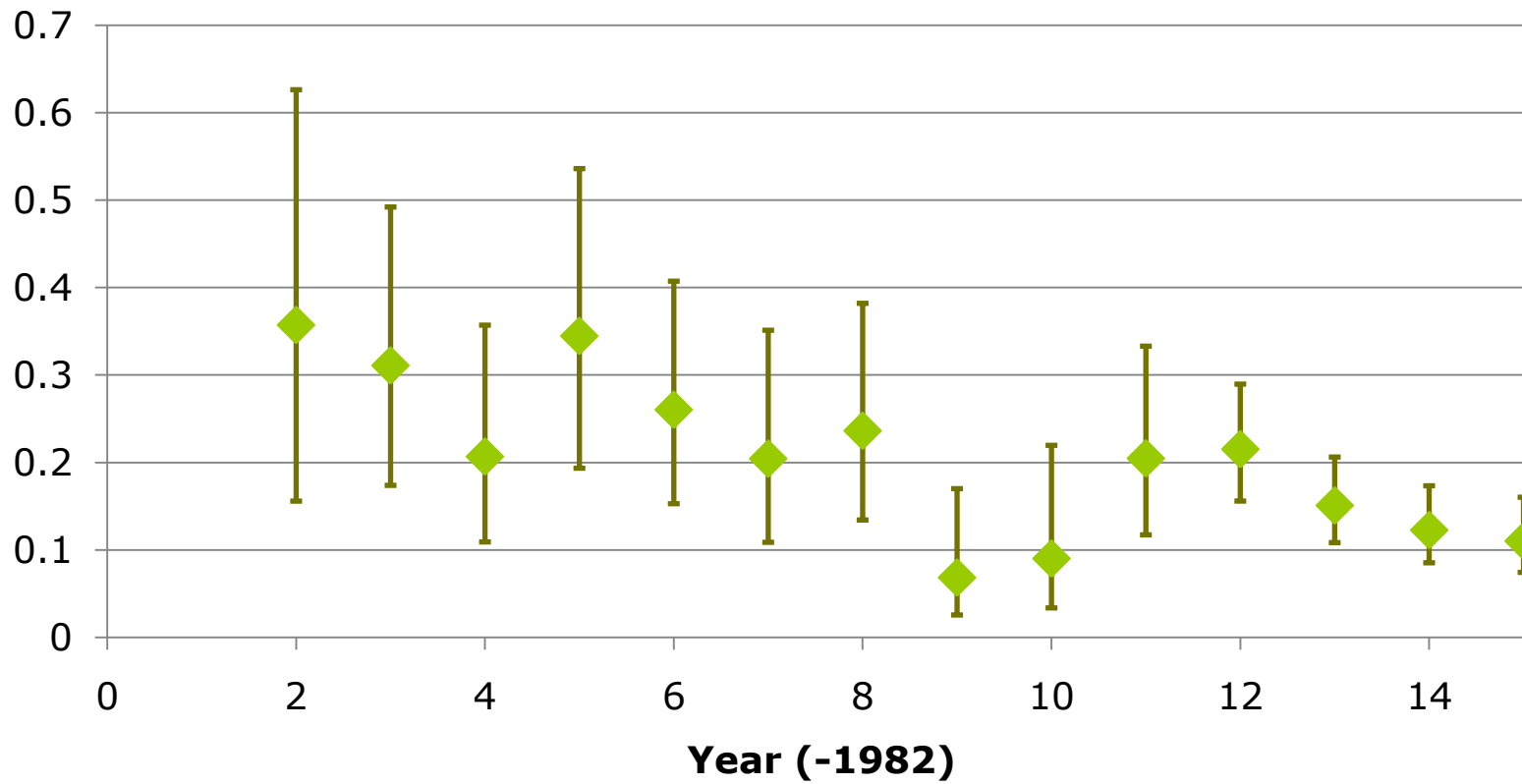
Breeding Dispersal:
movement between the colonies whilst a breeder

FROM

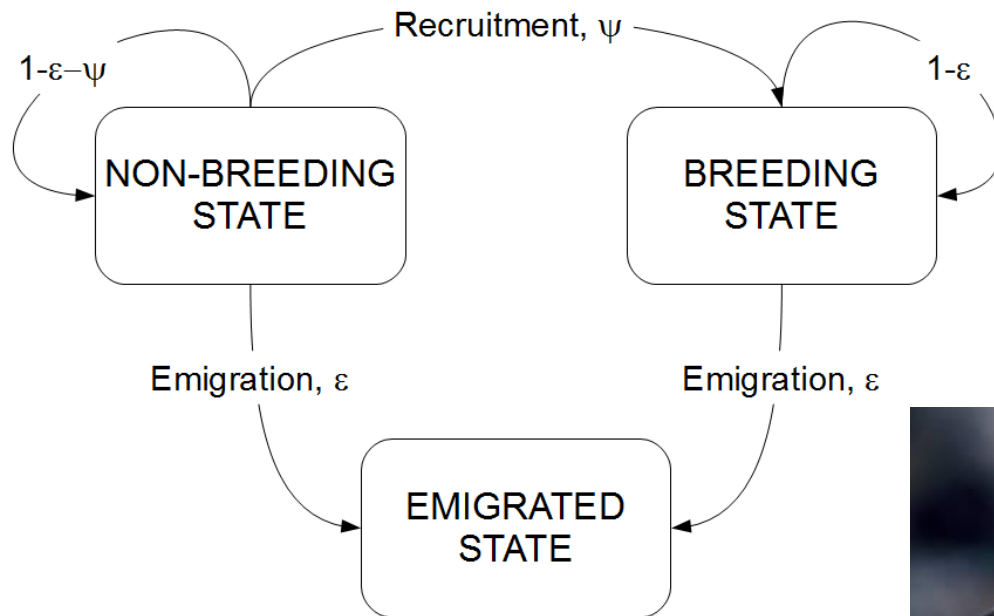
TO

| | | NB in VO | NB in MA | NB in SF | B in VO | B in MA | B in SF |
|----------|--|-----------------|----------|----------|---------|--------------------|---------|
| NB in VO | | Natal Dispersal | | | | | |
| NB in MA | | | | | | | |
| NB in SF | | | | | | | |
| B in VO | | | | | | Breeding Dispersal | |
| B in MA | | | | | | | |
| B in SF | | | | | | | |

Recruitment Probability



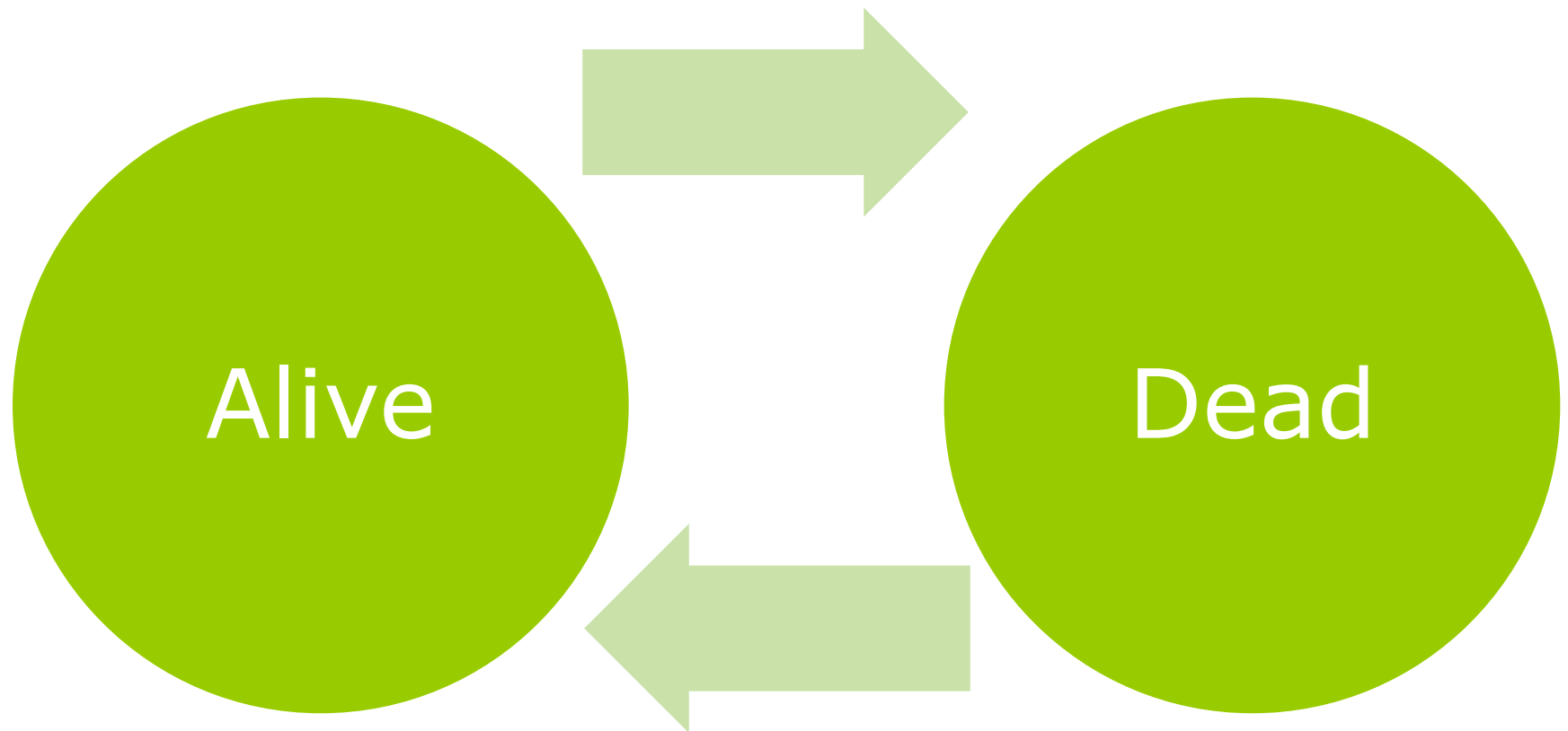
Emigration: Adapted Cormorant model



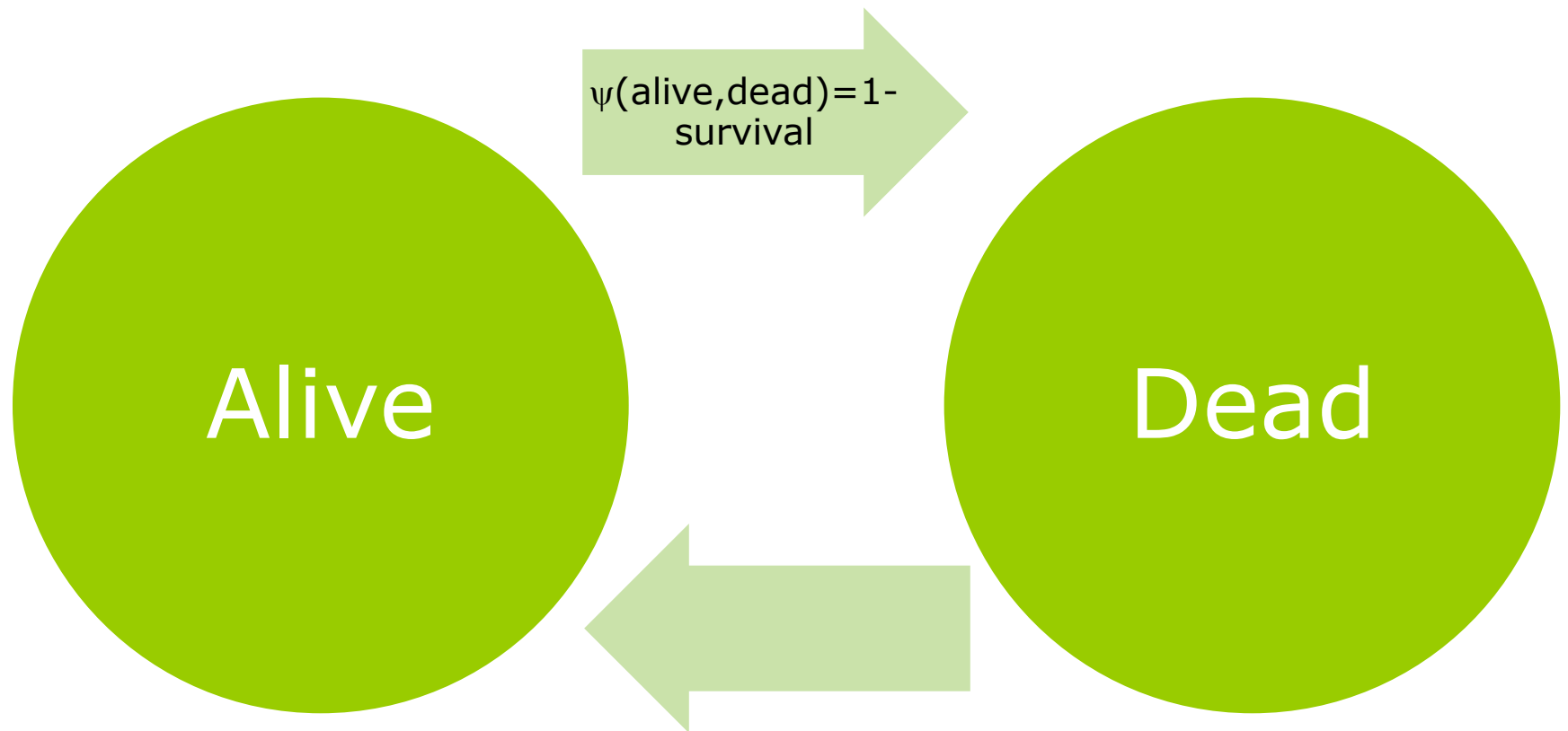
A Unified Framework

- ▣ Lebreton et al (1999) proposed that multistate models have the potential to be a unified framework for mark-recapture-recovery models.
- ▣ Clearly the single-site Cormack-Jolly-Seber model is a special case of the multisite Arnason-Schwarz model... however, how do you model **joint multisite recapture and recovery** data..?

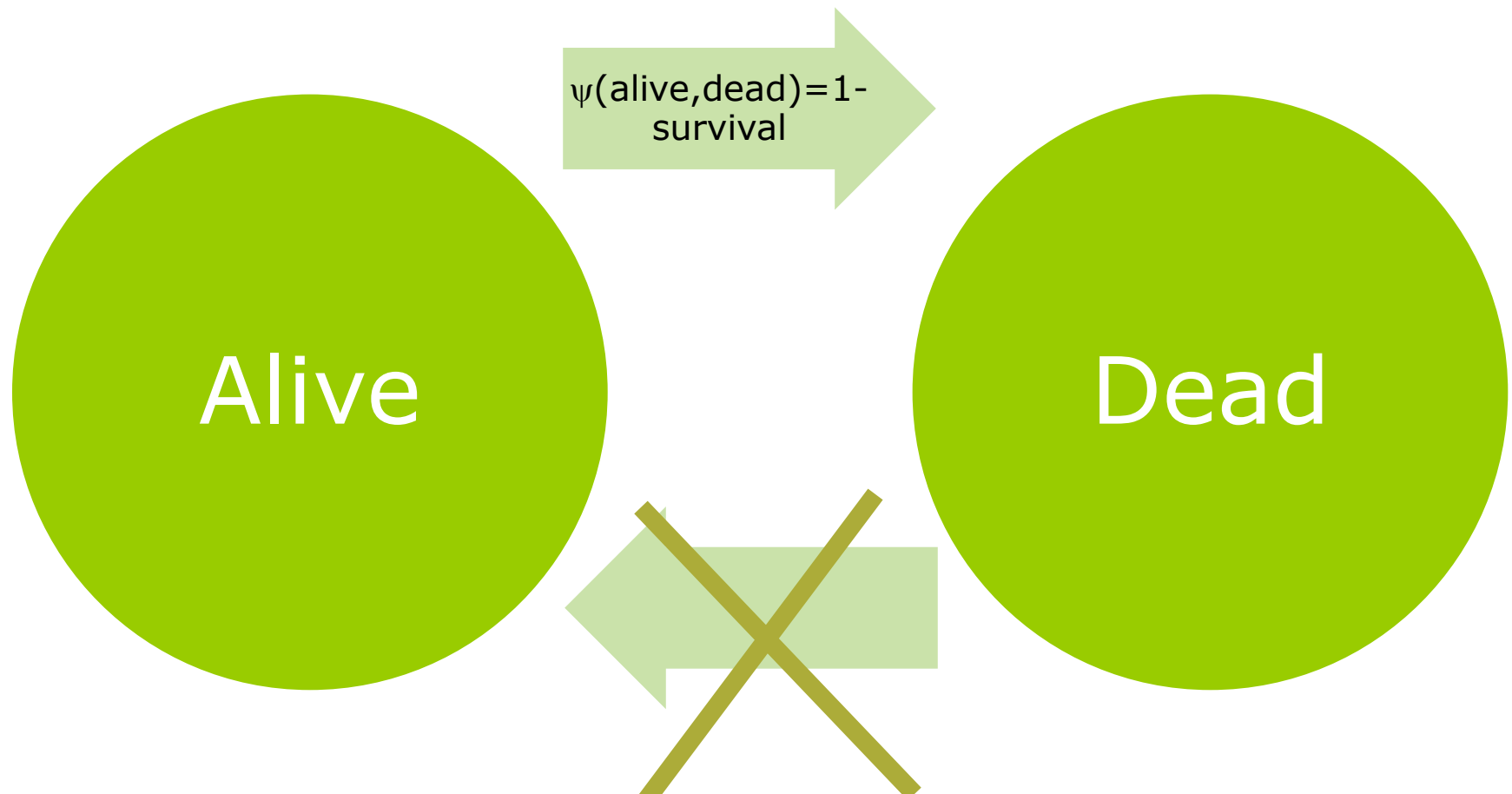
Multistate models: Modelling Joint Recaptures and Recoveries



Multistate models: Modelling Joint Recaptures and Recoveries

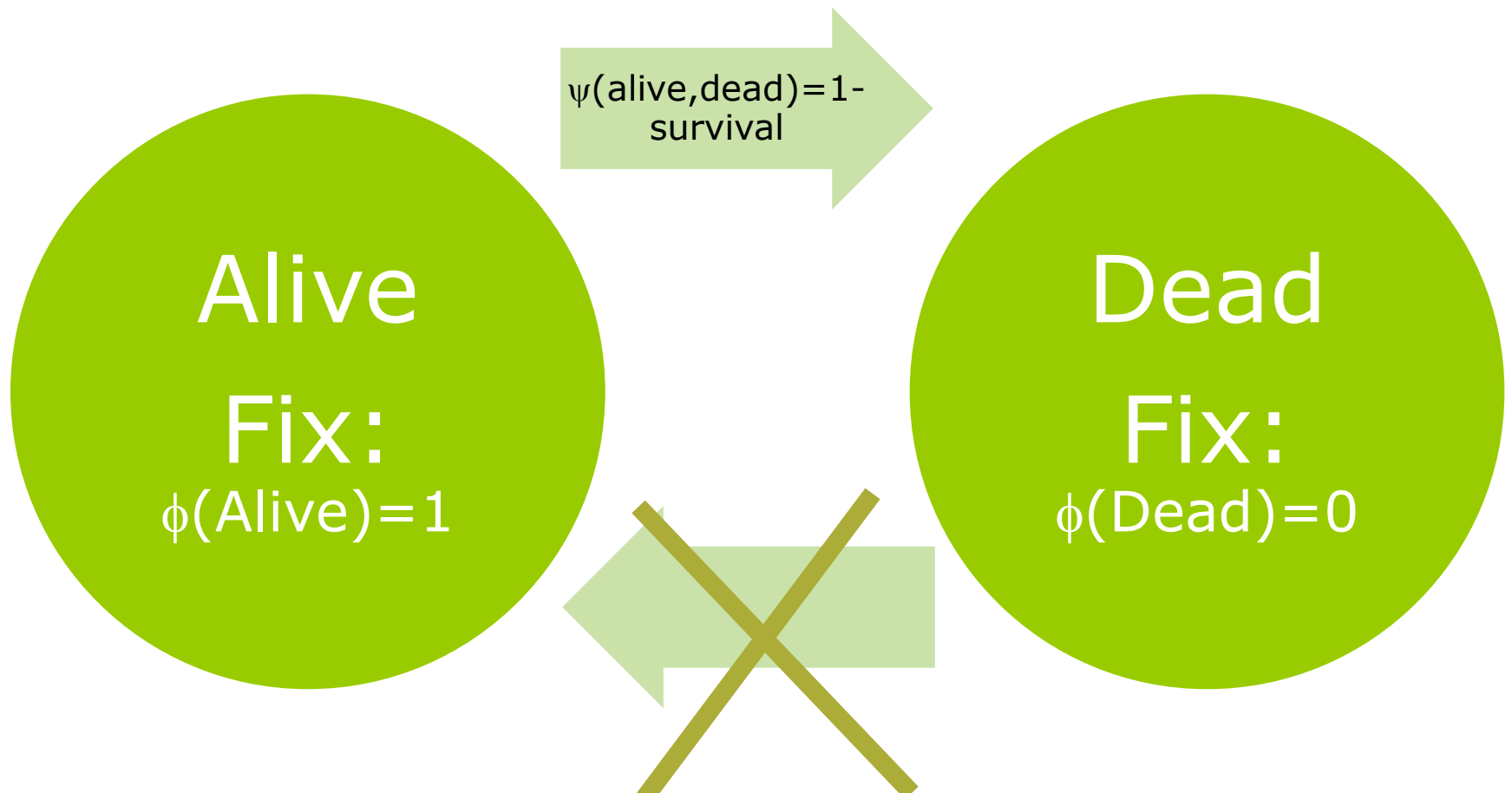


Multistate models: Modelling Joint Recaptures and Recoveries



Fix transition $\psi(\text{dead}, \text{alive}) = 0$, i.e. Once an animal has died it remains dead

Multistate models: Modelling Joint Recaptures and Recoveries



Fix transition $\psi(\text{dead,alive})=0$, i.e. Once an animal has died it remains dead

Multistate models: Modelling Joint Recaptures and Recoveries

$p(\text{Alive}) = \text{recapture probability}$

$p(\text{Dead}) = \text{recovery probability}$

$\psi(\text{alive}, \text{dead}) = 1 - \text{survival}$

Alive

Fix:

$\phi(\text{Alive}) = 1$

Dead

Fix:

$\phi(\text{Dead}) = 0$

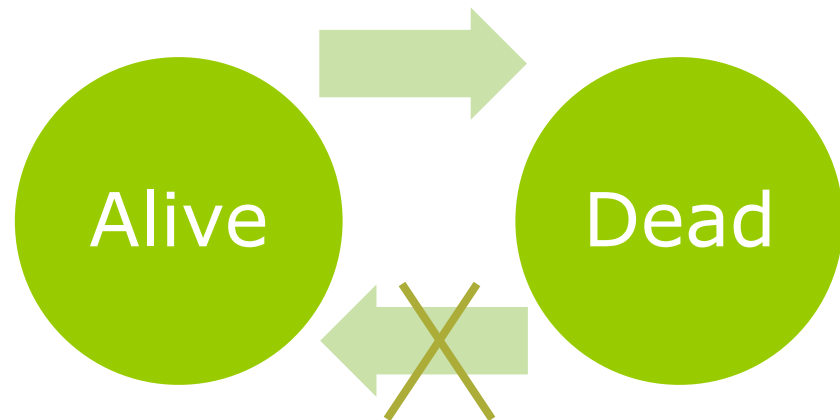
Fix transition $\psi(\text{dead}, \text{alive}) = 0$, i.e. Once an animal has died it remains dead

Multistate models for joint recaptures and recoveries

Example Simulated Data:

| | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|
| 1 | 0 | 0 | 2 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 1 | 2 | 0 |
| 1 | 0 | 0 | 0 | 0 | 2 | 0 |
| 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... |

1: Live Recapture
2: Dead Recovery



Multistate models for joint recaptures and recoveries

Example Simulated Data:

| | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|
| 1 | 0 | 0 | 2 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 1 | 2 | 0 |
| 1 | 0 | 0 | 0 | 0 | 2 | 0 |
| 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... |

1: Live Recapture
2: Dead Recovery

- Simulated 200 individual encounter histories
- Parameter values:
 - Survival probability = 0.6
 - Recapture probability = 0.4
 - Recovery probability = 0.3
- Use software `M-Surge` to fit a multistate model to the data

Multistate models for joint recaptures and recoveries

Example Simulated Data:

| | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|
| 1 | 0 | 0 | 2 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 1 | 2 | 0 |
| 1 | 0 | 0 | 0 | 0 | 2 | 0 |
| 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... |

1: Live Recapture
2: Dead Recovery

Output from fitting the constant model

- $\psi(\text{alive}, \text{dead}) = 0.393 (0.0278)$
- $p(\text{alive}) = 0.449 (0.0387)$
- $p(\text{dead}) = 0.305 (0.0343)$

Recall the following parameters fixed:

- $\psi(\text{dead}, \text{alive}) = 0$
- $\phi(\text{alive}) = 1$
- $\phi(\text{dead}) = 0$

Sufficient Statistics for multi-state integrated recapture and recovery data

- ❑ Similar to the single site likelihood formation – directly models live and dead encounters
- ❑ Each individual encounter history can be divided into three component parts:
 - Last observation and beyond
 - Consecutive live recaptures
 - Recovery of dead animals

Sufficient Statistics

- $\mathbf{v}_j^c(\mathbf{r})$: the number of animals from cohort c that are recaptured for the last time in region r at time t_j
- $\mathbf{n}_{(k,j)}^c(\mathbf{r}, \mathbf{s})$: the number of animals from cohort c that are observed in location r at time t_k , and next observed alive at location s at time t_{j+1}
- $\mathbf{d}_{(k,j)}^c(\mathbf{r})$: the number of animals from cohort c recovered dead between times t_j and t_{j+1} that were last observed alive at time t_k in location r .

Probabilities associated with the Statistics

- $\chi_j^c(\mathbf{r})$: the probability an animal from cohort c seen at time t_j in region r is not seen again in the study
- $\mathbf{O}_{(k,j)}^c(\mathbf{r}, \mathbf{s})$: the probability an animal from cohort c observed in location r at time t_k is unobserved until time t_{j+1} and is recaptured in location s at this time
- $\mathbf{D}_{(k,j)}^c(\mathbf{r})$: the probability an animal is recovered dead in the interval (t_j, t_{j+1}) given that it was last seen at time t_k in location r .

Likelihood Function

- Combining the sufficient statistics and derived probabilities....

$$L \propto \prod_{c=1}^C \prod_{r=1}^R \left[\prod_{j=1}^K \{\chi_j^c(r)\}^{v_j^c(r)} \prod_{k=1}^{K-1} \prod_{j=k}^{K-1} \{D_{(k,j)}^c(r)\}^{d_{(k,j)}^c(r)} \prod_{k=1}^{K-1} \prod_{j=k}^{K-1} \prod_{s=1}^R \{O_{(k,j)}^c(r,s)\}^{n_{(k,j)}^c(r,s)} \right]$$

Results in the same MLEs for the simulated data...

Multistate analysis

- ▣ $\psi(\text{alive}, \text{dead}) = 0.393$
(0.0278)
- ▣ $p(\text{alive}) = 0.449$
(0.0387)
- ▣ $p(\text{dead}) = 0.305$
(0.0343)

Sufficient statistic analysis

- ▣ $\phi = 0.607$ (0.0278)
- ▣ $p = 0.449$ (0.0387)
- ▣ $\lambda = 0.305$ (0.0343)

Extensions

- ❑ Models with additional memory (Hestbeck et al, 1993; Brownie et al, 1993)
- ❑ Models with state uncertainty
 - Multievent models (Pradel, 2004)
 - Partial observation models (King and McCrea, 2010)
- ❑ Robust design models
 - Combine open and closed capture-recapture models (Pollock, 1982)
- ❑ Models which do not condition on first capture
 - Stopover models (Pledger et al, 2009)

References

- Arnason (1972,1973) *Res. Popul. Ecology*
- Barker (1997) *Biometrics*, **53**: 666-677
- Barker (1999) *Bird study*, **46**: S82-S91
- Brownie et al (1993) *Biometrics*, **49**: 1173-1187
- Catchpole et al (1998) *Biometrics*, **54**: 33-46
- Hestbeck et al (1991) *Ecology*, **72**:523-533
- King and Brooks (2003) *Biometrika*, **90**: 435-444
- King and McCrea (2010) *submitted*
- Lebreton (1999) *Bird study*, **46**: S39-S46
- McCrea et al (2010) *JABES*, **15**: 539-561
- Pledger et al. (2009) in *Modelling Demographic processes in marked populations*.
- Pollock (1982) *Journal of Wildlife Management*. **46**:757-760.
- Pradel (2004) *Biometrics*, **61**: 442-447
- Schwarz (1993) *Biometrics*, **59**: 291-318

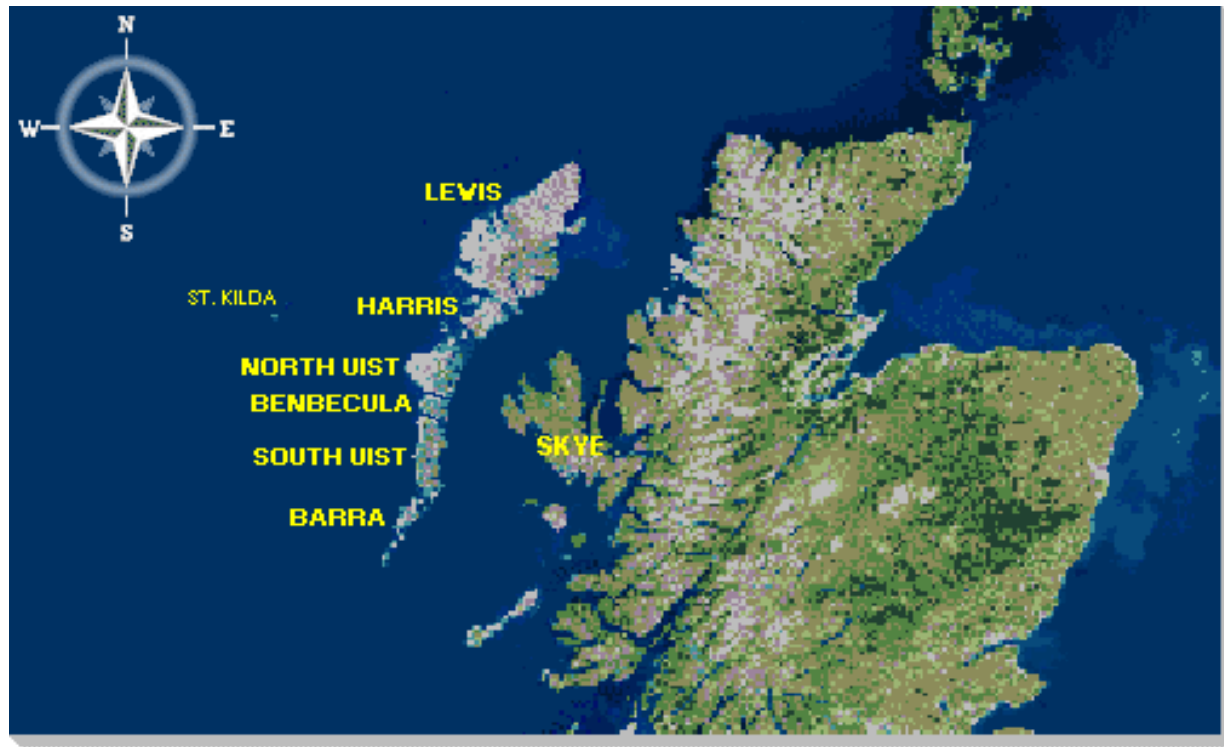
Covariates

- **Hérons**: North and Morgan, *Biometrics*, 1979.
- **Sheep**
- Distinction between **environmental** and **individual** covariates

Soay sheep, *Ovis aries*

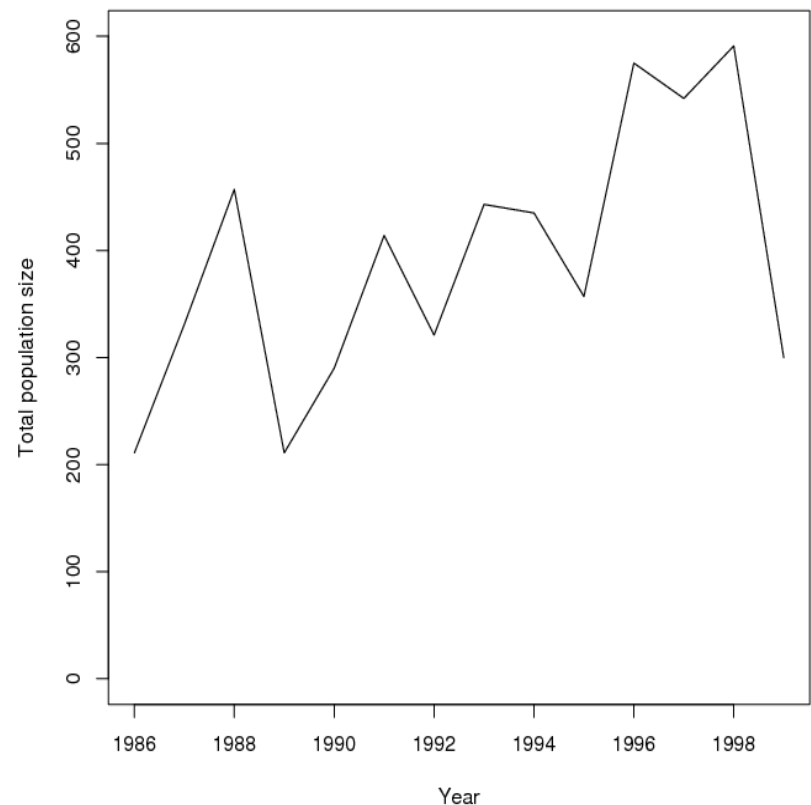


St. Kilda



Behaviour of the total population size over time

The total population of Soay sheep, *Ovis aries* on St. Kilda is well known for reflecting large **population crashes**, when there is high mortality due to a combination of factors such as severe winter and high population density.



Classical model for female sheep

An early model, resulting from likelihood-ratio tests between a small number of selected models is:

$$\phi_1(P+M+h), \phi_2(M), \phi_{3:7}(M), \phi_{8+}(P+M)$$

where P denotes population size, M denotes March rain and h denotes horn type. Regressions are **logistic**, and those on M are parallel.

Note the use of **age-classes**.

More detailed models result from **Bayesian inference**.

Individual covariates

- ❑ Age (use of age-classes)
- ❑ Sex (separate analyses: here only females)
- ❑ Climate
- ❑ Food
- ❑ Density
- ❑ Breeding history
- ❑ Size
- ❑ Health

Illustration of individual covariates



The likelihood for life-history data

For general life-history data, the likelihood has the form below:

$$L \propto \prod_{c=1}^C \left[\prod_{j=c}^{k-1} \{\alpha_{c,j}(1 - \phi_{c,j})\lambda_{c,j}\}^{d_{c,j}} \prod_{j=c}^k \{\alpha_{c,j}\chi_{c,j}\}^{v_{c,j}} \prod_{j=c}^{k-1} p_{c,j+1}^{w_{c,j}} q_{c,j+1}^{z_{c,j}} \right]$$

where c is the time an individual was first captured, k is the time the animal was last known to be alive, d indicates the known death of the animal (1 or 0), w is an indicator variable for being seen occasion, and χ is a probability that an animal is never seen after an occasion, when it was known to be alive.

Time-varying individual covariates

Important individual covariates often vary with time, and contain **missing values**.

One approach is to use models to **impute** missing values. Done by Bonner and Schwarz (*Biometrics*, 2006).

Another is to construct the likelihood **conditionally**.

Dealing with missing values

Consider now the case history: (1 0 1 1 0 0).

The **traditional** likelihood contribution is:

$$\phi_1(1-p_1)\phi_2 p_2 \phi_3 p_3 \chi_4$$

Forming the likelihood in stages, we write:

$$\pi_1(1,0)\pi_2(0,1) \pi_3(1,1) \pi_4(1,0) \pi_5(0,0)$$

and **to accommodate missing values**, we may write:

$$\pi_1(1,0) \pi_3(1,1) \pi_4(1,0),$$

in contrast to the traditional approach, in which this case history is **omitted: a complete case analysis**.

Age classes

- For Soay sheep we identify 4 **age-classes** for survival. These correspond to animals of years of life:
- 1, 2, 3-7, >7
- We use these age-classes in the data summary table that follows.

Trinomial data for Soay sheep

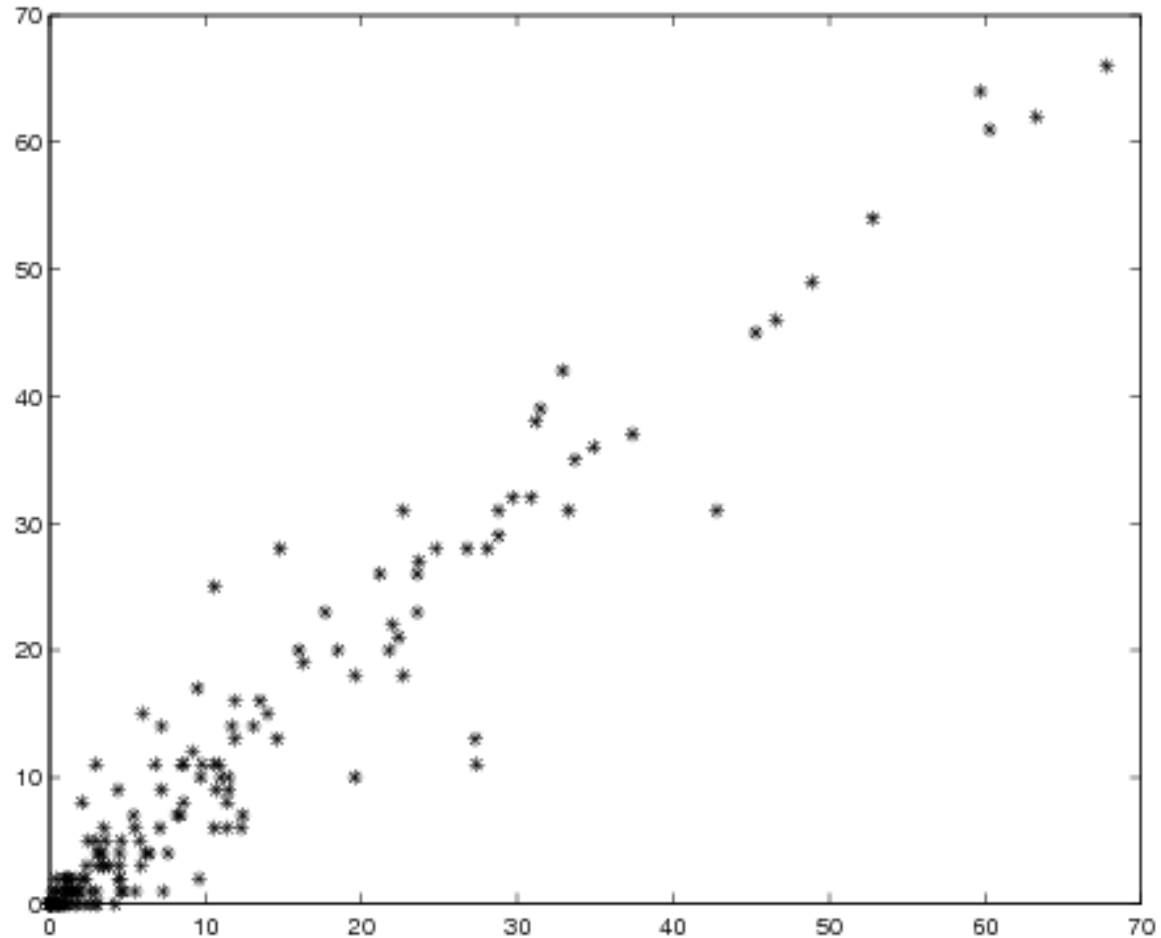
| Year | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
|------|----|----|----|---|----|---|---|----|----|---|----|----|
| 87 | 9 | 21 | 1 | 0 | 2 | 0 | 0 | 11 | 0 | 1 | 3 | 1 |
| 88 | 7 | 32 | 11 | 0 | 13 | 1 | 2 | 28 | 0 | 2 | 14 | 1 |
| 89 | 11 | 7 | 36 | 5 | 6 | 8 | 1 | 18 | 11 | 1 | 6 | 15 |
| 90 | 2 | 16 | 3 | 0 | 0 | 0 | 1 | 7 | 1 | 1 | 11 | 0 |
| 91 | 0 | 20 | 6 | 0 | 8 | 2 | 2 | 13 | 0 | 0 | 31 | 2 |
| 92 | 1 | 10 | 42 | 0 | 9 | 5 | 0 | 23 | 3 | 0 | 31 | 16 |
| 93 | 3 | 28 | 5 | 0 | 6 | 1 | 0 | 29 | 0 | 3 | 28 | 4 |
| 94 | 9 | 26 | 46 | 1 | 15 | 3 | 4 | 31 | 0 | 5 | 22 | 7 |
| 95 | 17 | 13 | 38 | 0 | 14 | 2 | 0 | 37 | 2 | 0 | 18 | 14 |
| 96 | 4 | 35 | 4 | 0 | 4 | 0 | 0 | 49 | 0 | 0 | 19 | 4 |
| 97 | 6 | 20 | 45 | 2 | 27 | 1 | 0 | 61 | 1 | 0 | 26 | 4 |
| 98 | 10 | 28 | 31 | 0 | 20 | 0 | 1 | 64 | 0 | 0 | 32 | 10 |
| 99 | 12 | 10 | 66 | 0 | 8 | 9 | 2 | 54 | 11 | 5 | 11 | 25 |
| 00 | 3 | 39 | 6 | 0 | 11 | 1 | 1 | 62 | 2 | 1 | 23 | 1 |

Comparison: current weight vs birth weight

Comparison of estimates of
weight coefficients

| | Trinomial with W | Trinomial with B |
|----------|---------------------|---------------------|
| Lamb | 0.88(0.16) | 0.91(0.17) |
| Yearling | 0.28(0.09) | -0.05(0.29) |
| Adult | 0.20(0.07) | 0.60(0.25) |

Goodness-of-fit



Classical model for female sheep

A classical model, resulting from likelihood-ratio tests between a small number of selected models is:

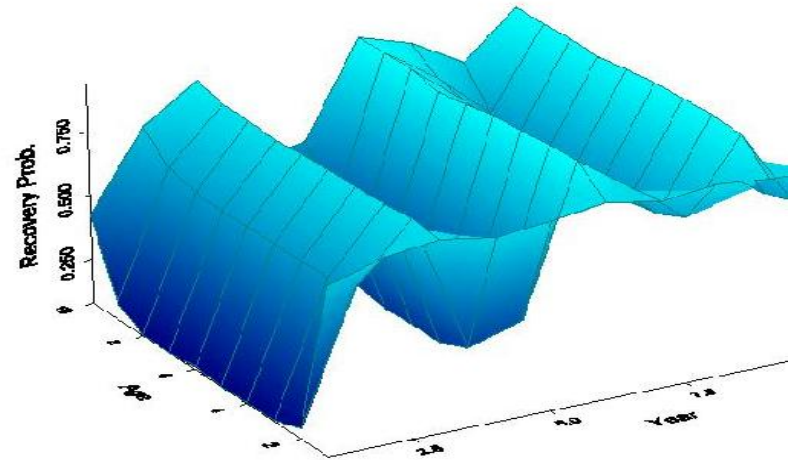
$$\phi_1(P+M+h), \phi_2(M), \phi_{3:7}(M), \phi_{8+}(P+M)$$

where P denotes population size, M denotes March rain and h denotes horn type.

Regressions are **logistic**, and those on M are **parallel**.

Soay sheep (*Ovis aries*)

Search for age-classes with constant survival- **Bayesian approach** obtains finer/better categorisation, and more detailed description of survival in terms of covariates.



Age-dependence (females): Bayes

| Age-structure | Posterior probability |
|---|-----------------------|
| $\phi_1; \phi_{2:4}; \phi_{5:7}; \phi_{8+}$ | 0.384 |
| $\phi_1; \phi_{2:7}; \phi_{8+}$ | 0.286 |
| $\phi_1; \phi_{2:4}; \phi_{5:7}; \phi_7; \phi_{8+}$ | 0.093 |
| $\phi_1; \phi_{2:3}; \phi_{4:7}; \phi_{8+}$ | 0.075 |

- Note that with probability 1, lambs have a distinct survival rate.
- Often, the models with most posterior support are close neighbours of each other.

The Bayes solution is richer

| Cov | 1 | 2 | 3 | 4 | 8 | 9 | >9 |
|-----|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| c | 0.160 | 0.101 | 0.101 | 0.105 | 0.768 | 0.841 | 0.841 |
| h | 0.369 | 0.039 | 0.039 | 0.043 | 0.055 | 0.014 | 0.014 |
| b | 1.000 | 1.000 | 1.000 | 0.952 | 0.225 | 0.162 | 0.162 |
| N | 0.666 | 0.403 | 0.403 | 0.427 | 0.919 | 0.919 | 0.919 |
| P | 0.935 | 0.203 | 0.203 | 0.153 | 0.355 | 0.419 | 0.419 |
| M | 0.703 | 0.464 | 0.464 | 0.413 | 0.186 | 0.167 | 0.167 |
| T | 0.070 | 0.195 | 0.195 | 0.194 | 0.735 | 0.671 | 0.671 |
| A | 0.094 | 0.215 | 0.215 | 0.215 | 0.154 | 0.190 | 0.169 |
| RE | 0.255 | 0.942 | 0.942 | 0.907 | 0.023 | 0.012 | 0.012 |

References

SECTION 6

PARAMETER REDUNDANCY

Determining the parameter redundancy of non-linear models

Collaborators:
Ted Catchpole,
Diana Cole

Outline

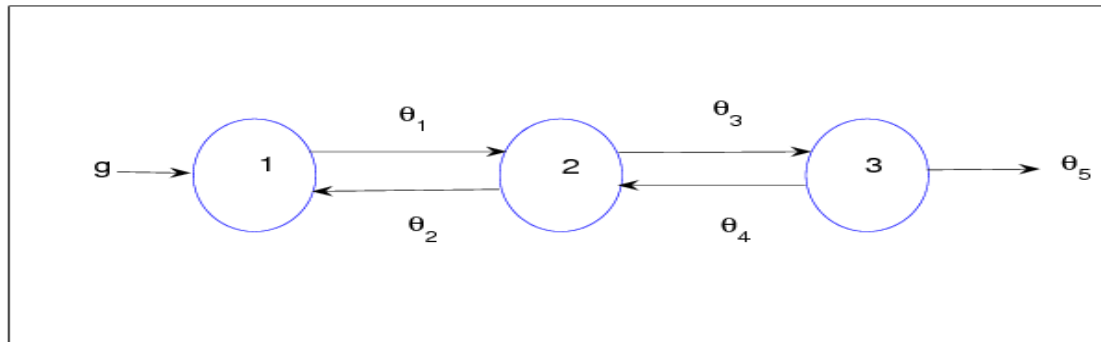
- Introduction and motivation
- Definitions; general rules
- Use of symbolic algebra; expansion theorems

- The PLUR decomposition; exhaustive summaries; reparameterisation; identifiability; weak identifiability
- Future work

Complex models and their parameters

- Compartment models
- Ecology
- Econometrics
- Hidden Markov models

Compartment models



Econometrics

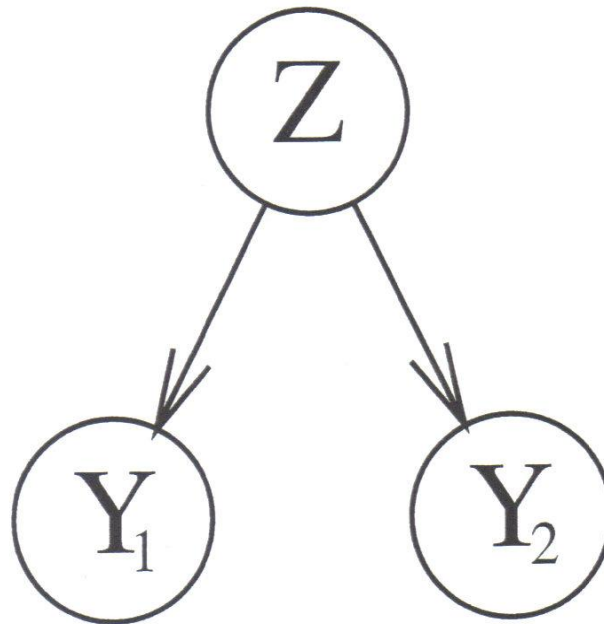
- ▣ Identifiability of the **simultaneous equation model**:

$$By_t + \Gamma x_t = u_t,$$

where y_t and u_t are vectors of random variables, x_t is a vector of non-random exogenous variables, B and Γ are matrices of parameters, and u_t has a normal distribution, with dispersion matrix Σ .

- ▣ The parameter space is $[B, \Gamma, \Sigma]$, some of which may be constrained.

A simple naïve Bayesian network



A naïve Bayesian network with a binary root node and two binary observable nodes.

Ecology

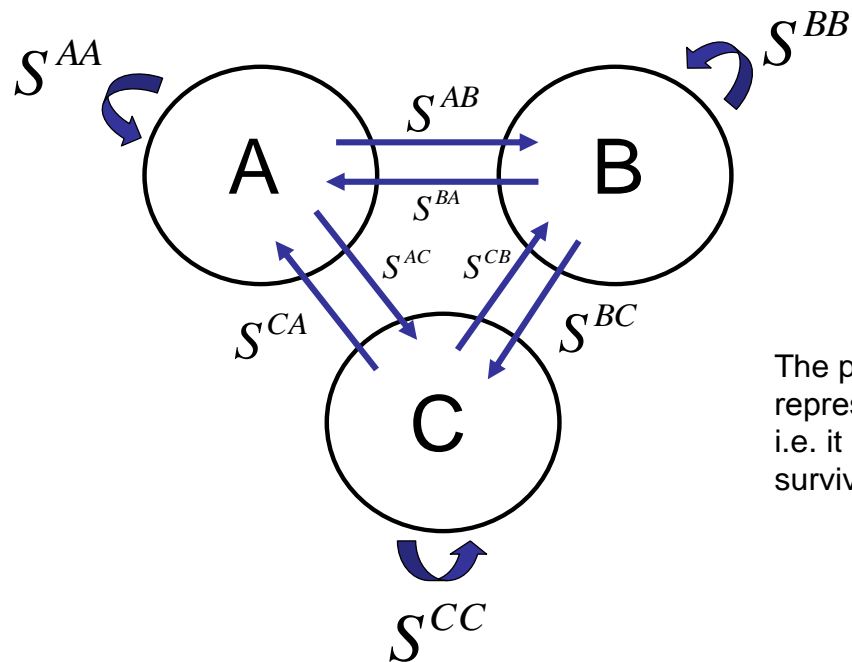
- ❑ Estimation of the annual survival probabilities of wild animals.
- ❑ Collect data on previously marked animals.
- ❑ These are either found **dead or alive**.
- ❑ Form probability models.
- ❑ Fit to data using maximum likelihood, or Bayesian methods.

Complexity

- ❑ Models may be complicated, incorporating age, cohort and time components.
- ❑ Models may be simplified by the use of **covariates**.
- ❑ Modern focus on **multi-site data** can produce models with many parameters.
- ❑ It is often unclear how many parameters can be estimated.

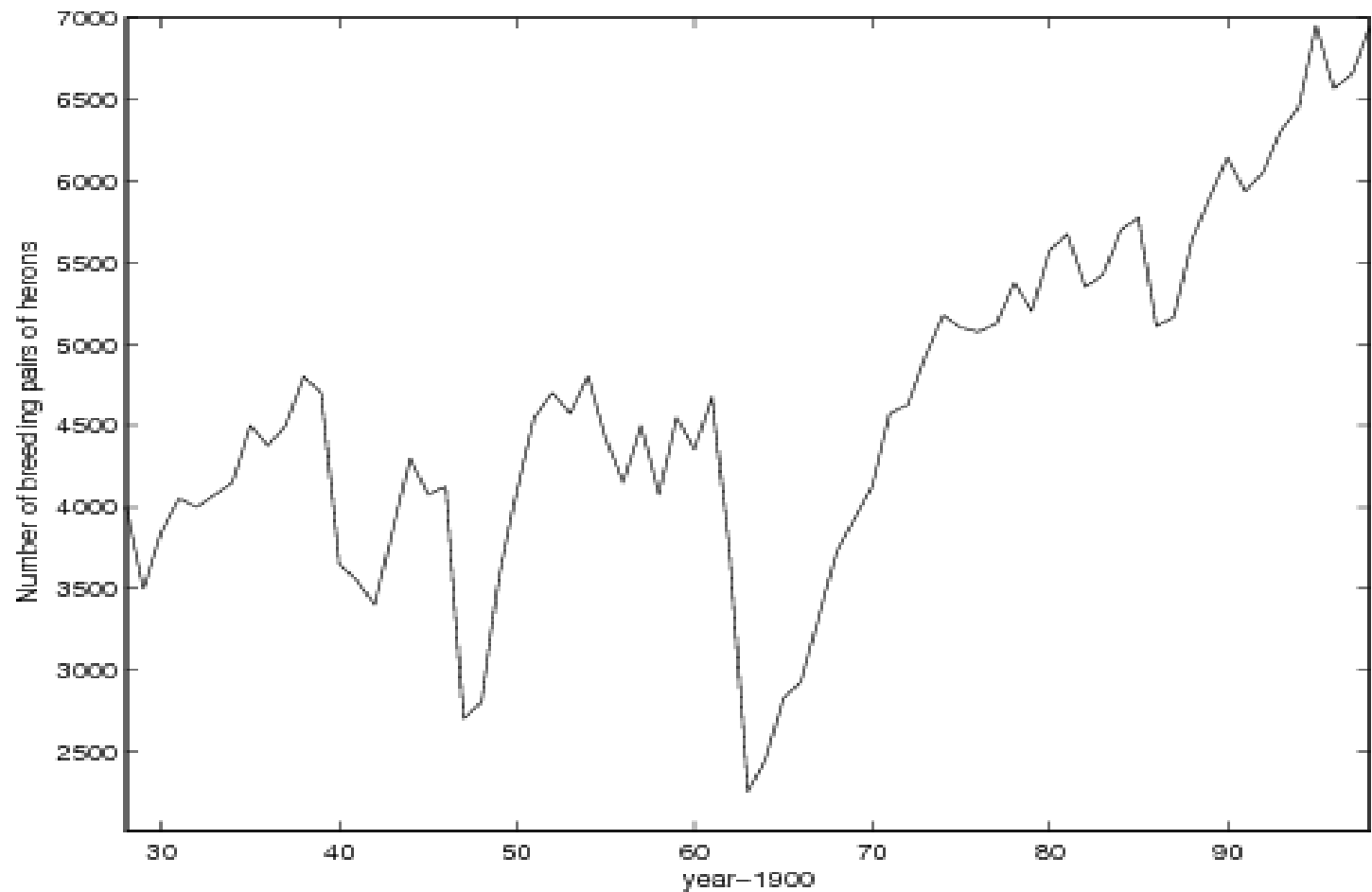
An example of a multi-site system

Multisite Systems

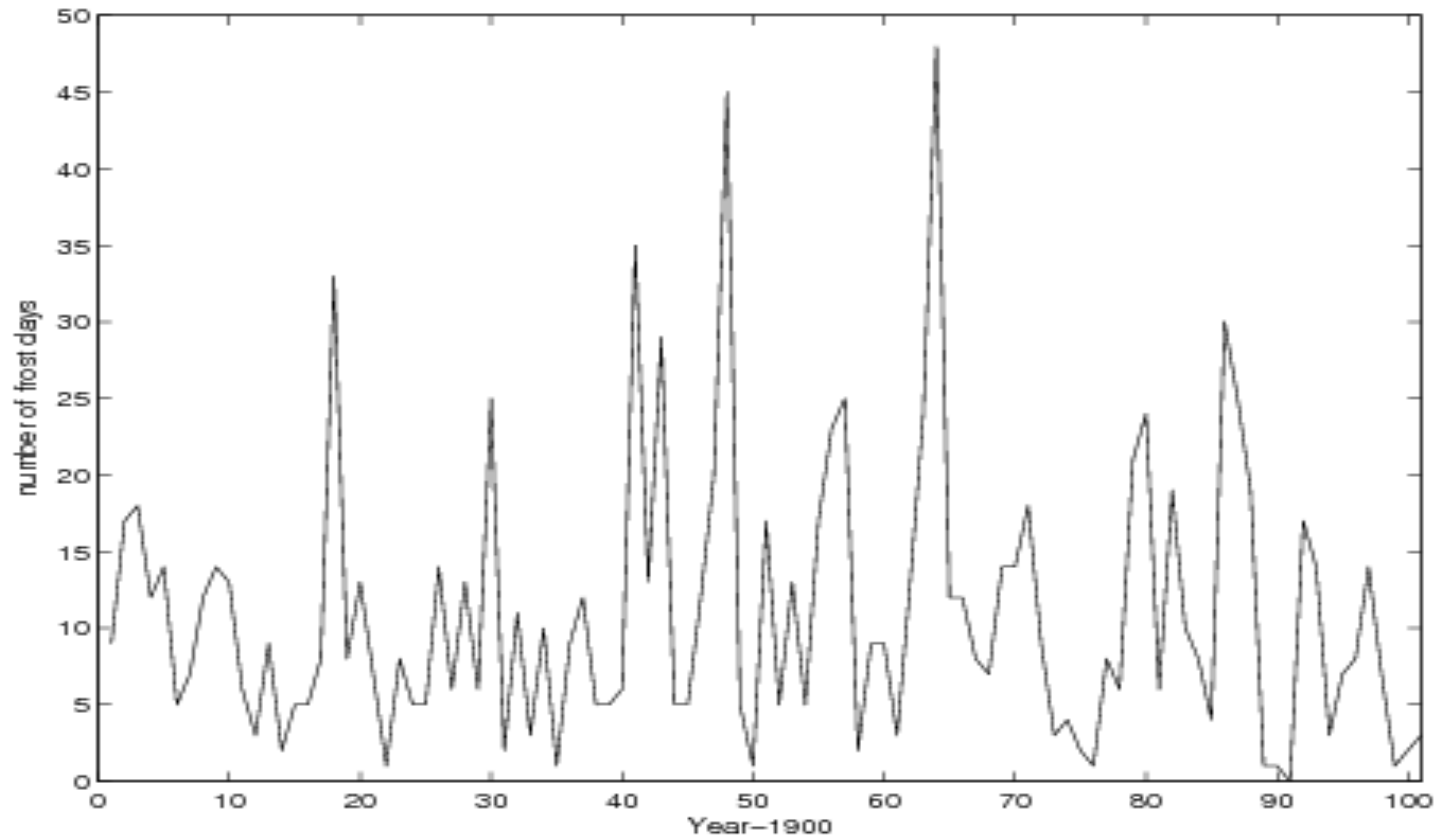


The parameter S represents the “transition”, i.e. it represents both survival and movement

The British heron census, *Ardea cinerea*



Climatic covariates: number of frost-days in **Central** England.



The Cormack-Jolly-Seber (CJS) model (1965)

Consider a simple case in which all animals are adults, sharing a common probability of annual survival, ϕ . If p denotes the probability of recapture then the **multinomial** probabilities corresponding to any cohort, of known size, of marked birds have the form:

$$\phi p, \quad \phi^2 p(1-p), \quad \phi^3 p(1-p)^2, \quad \dots$$

Parameters may be time-dependent – appropriate for adult animals.

Illustration of CJS recapture probabilities: a 3-year study

| | | |
|--------------|----------------------------|---|
| $\phi_1 p_2$ | $\phi_1 \phi_2 (1-p_2)p_3$ | $\phi_1 \phi_2 \phi_3(1-p_2)(1-p_3)p_4$ |
| | $\phi_2 p_3$ | $\phi_2 \phi_3(1-p_3)p_4$ |
| | | $\phi_3 p_4$ |

CJS recapture probabilities: what we can estimate

| | | |
|--------------|----------------------------|--|
| $\phi_1 p_2$ | $\phi_1 \phi_2 (1-p_2)p_3$ | $\phi_1 \phi_2 \phi_3 (1-p_2)(1-p_3)p_4$ |
| | $\phi_2 p_3$ | $\phi_2 \phi_3 (1-p_3)p_4$ |
| | | $\phi_3 p_4$ |

Parameter redundancy

- This model has deficiency of one: we can only estimate the product, $\phi_3 p_4$. All the other parameters can be estimated.

Parameter redundancy

- This model has deficiency of one: we can only estimate the product, $\phi_3 p_4$. All the other parameters can be estimated.
- What if we only have two years of ringing?

Illustration of CJS recapture probabilities: a 3-year study + 2 cohorts

| | | |
|--------------|----------------------------|--|
| $\phi_1 p_2$ | $\phi_1 \phi_2 (1-p_2)p_3$ | $\phi_1 \phi_2 \phi_3 (1-p_2)(1-p_3)p_4$ |
| | $\phi_2 p_3$ | $\phi_2 \phi_3 (1-p_3)p_4$ |

References

- ❑ Bellman and Åström, (1970), *Mathematical Biosciences*.
- ❑ Goodman, (1974), *Biometrika*.
- ❑ Rothenberg, (197), *Econometrica*.
- ❑ Walter, (1982), *Identifiability of state space models*.

Section

- Introduction and motivation
- **Definitions; general rules**
- Use of symbolic algebra; expansion theorems
- The PLUR decomposition; exhaustive summaries; reparameterisation; identifiability; weak identifiability
- Future work

Parameter redundancy and identifiability

- ❑ A model is **identifiable** if no two values of the parameters give the same probability distribution for the data.
- ❑ A model is **locally identifiable** if there is a distance $\delta > 0$, such that any two parameter values that give the same distribution must be separated by at least δ .
- ❑ A **parameter redundant** model has parameters that cannot be estimated.
- ❑ A parameter redundant model is not locally identifiable.
- ❑ Full rank models are **essentially** or **conditionally** full rank.
- ❑ An essentially full rank model is locally identifiable.
- ❑ Are essentially full rank models identifiable?

General rules

- ❑ In some cases it is possible to establish general rules for models of particular structures.
- ❑ This avoids having to use Maple (see later).
- ❑ A particular illustration of this occurs with **age-dependent** recovery models

Model notation for recovery models

Ring-recovery models are described as, for example:

C/A/C, T/A/C, T/A/T, C/C/T.

In this notation, each model is specified by 3 letters, which designate, in order,

1. The way we model **first-year survival**: C or T;
2. The way we model **adult survival**: C, A or T; and A can have categories.
3. The way we model the **recovery probability**: C, A or T.

Steps: age-dependence also in λ .

□ Consider, for example, the model denoted by C/A(2,2,3)/A(2,1,1,4). What can we estimate here?

□ Here we have the parameters:

$\phi_1, \phi_2, \phi_2, \phi_3, \phi_3, \phi_4, \phi_4, \phi_4$

$\lambda_1, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_4, \lambda_4, \lambda_4$

Steps: age-dependence also in λ .

- Consider, for example, the model denoted by C/A(2,2,3)/A(2,1,1,4). What can we estimate here?
- Here we have a **single step**, as shown:

$$\begin{array}{c} \phi_1, \phi_2, \phi_2 \mid \phi_3, \phi_3, \phi_4, \phi_4, \phi_4 \\ \lambda_1, \lambda_1, \lambda_2 \mid \lambda_3, \lambda_4, \lambda_4, \lambda_4, \lambda_4 \end{array}$$

Theorem 1

- Suppose the first step occurs at age n , and let m be the number of parameters used in the first n years.
- If $m = n+1$, the model is parameter redundant.
- If $1 < m < n+1$, then the step does not cause parameter redundancy.
Furthermore, to test for parameter redundancy, the parameters occurring in the first n years can be discarded, and the count started anew in year $n+1$.

Theorem 2

- ❑ In the age-dependent model T/A/A
- ❑ The step at age 1 year does **not** cause parameter-redundancy
- ❑ To determine any possible redundancy caused by a subsequent step, the age and parameter counts begin again after age 1 year, as in Theorem 1.

Section

- Introduction and motivation
- Definitions; general rules
- Use of symbolic algebra and expansion theorems
- The PLUR decomposition; exhaustive summaries; reparameterisation; identifiability; weak identifiability
- Future work

How to test for parameter redundancy in general

- ❑ Form an appropriate **derivative matrix**, D .
- ❑ Use **Maple** to determine the **symbolic row rank** of D . Use this to determine if the model is parameter redundant or full rank.
- ❑ We can also determine which **parameter combinations** can be estimated, if the model is parameter redundant.

The method

The approach was for **exponential family** models. It is performed using a symbolic algebra package such as Maple.

1. Calculate $\mathbf{D} = \left[\frac{\partial \mu_j}{\partial \theta_i} \right]$ (μ is the mean, θ are parameters).

2. The number of estimable parameters = $\text{rank}(\mathbf{D})$.

3. Solve $\alpha^T \mathbf{D} = 0$. The location of the zeros in α indicates which are the estimable parameters.

4. Solve $\sum_{i=1}^p \alpha_{ij} \frac{\partial f}{\partial \theta_i} = 0$ to find the full set of estimable

parameters; (j is the index for >1 solution to $\alpha^T \mathbf{D} = 0$).

Example 1: Cormack-Jolly-Seber Model

Little Penguins, *Eudyptula minor*, capture recapture data (1994 to 1997)

$$\mathbf{N} = \begin{bmatrix} 30 & 58 & 37 \\ 0 & 20 & 37 \\ 0 & 0 & 18 \end{bmatrix}$$

ϕ_i – probability a penguin survives from occasion i to $i+1$

p_i – probability a penguin is recaptured on occasion i

The set of parameters is: $\theta = [\phi_1, \phi_2, \phi_3, p_2, p_3, p_4]$



$$\mathbf{P} = \begin{bmatrix} \phi_1 p_2 & \phi_1 \bar{p}_2 \phi_2 p_3 & \phi_1 \bar{p}_2 \phi_2 \bar{p}_3 \phi_3 p_4 \\ 0 & \phi_2 p_3 & \phi_2 \bar{p}_3 \phi_3 p_4 \\ 0 & 0 & \phi_3 p_4 \end{bmatrix}$$

$$\bar{p}_2 = 1 - p_2 \text{ etc}$$

Forming the derivative matrix (take logs first)

$$\mathbf{D} = \frac{\partial \ln(\mathbf{P})}{\partial \theta} = \begin{bmatrix} \phi_1^{-1} & \phi_1^{-1} & \phi_1^{-1} & 0 & 0 & 0 \\ 0 & \phi_2^{-1} & \phi_2^{-1} & \phi_2^{-1} & \phi_2^{-1} & 0 \\ 0 & 0 & \phi_3^{-1} & 0 & \phi_3^{-1} & \phi_3^{-1} \\ p_2^{-1} & -\overline{p_2}^{-1} & -\overline{p_2}^{-1} & 0 & 0 & 0 \\ 0 & p_3^{-1} & -\overline{p_3}^{-1} & p_3^{-1} & -\overline{p_3}^{-1} & 0 \\ 0 & 0 & p_4^{-1} & 0 & p_4^{-1} & p_4^{-1} \end{bmatrix}$$

$\text{rank}(\mathbf{D}) = 5 < 6$, so the model is parameter redundant.

In order to see which of the original parameters we can estimate:

Set $\alpha^T \mathbf{D} = 0 \Rightarrow \alpha^T = [0, 0, -\phi_3 / p_4, 0, 0, 1]$

Solving PDE, we find that the estimable parameters are: $\phi_1, \phi_2, p_2, p_3, \phi_3 p_4$

Expansion theorems

These give conditions that ensure that results which hold for a particular configuration also hold for **larger** configurations.

For instance, the CJS model **always** has deficiency one.

Example – Cormack-Jolly-Seber Model with covariates

We now set

$$\phi_i = 1/\{1+\exp(a + bx_i)\}$$

For example, x_i could be the mean annual banding weight, or the SOI.

$$\theta = [a, b, p_2, p_3, p_4],$$

and we find that **the model is now full rank.**



References

- ❑ Bekker et al (1994) *Identification, equivalent models and computer algebra*.
- ❑ Catchpole, Freeman and Morgan, (1996), *JRSS B*.
- ❑ Catchpole and Morgan, (1997), *Biometrika*
- ❑ Catchpole, Morgan and Freeman, (1998), *Biometrika*.
- ❑ Catchpole and Morgan, (2001), *Biometrika*.
- ❑ Cole and Morgan, (2009), Parameter redundancy with covariates. submitted.

Section

- Introduction and motivation
- Definitions; general rules
- Use of symbolic algebra; expansion theorems
- The PLUR decomposition; exhaustive summaries; reparameterisation; identifiability; weak identifiability
- Future work

Use of the PLUR decomposition

If parameter redundant:

Solve $\alpha^T \mathbf{D} = 0$. Zeros in α indicate estimable parameters.

Solve $\sum_{i=1}^p \alpha_{ij} \frac{\partial f}{\partial \theta_i} = 0$ to find full set of estimable parameters.

If full rank:

Determine whether **essential** ($\forall \theta$) or **conditionally** ($\exists \theta$) full rank using the PLUR decomposition.

D = PLUR. If $\det(\mathbf{U}) = 0$, model is parameter redundant.
If $\det(\mathbf{U})$ is close to 0 model is **near parameter redundant**.

Components of the PLUR decomposition, available in Maple

We have **$D=PLUR$** .

P is a permutation matrix,

L is lower triangular, with 1s on the diagonal,

U is upper triangular,

R is reduced echelon form.

Use of the PLUR decomposition: penguin covariates

We have **D=PLUR**.

We find that

$$\text{Det}(\mathbf{U}) = \frac{-(x_1 - x_2)(1 - p_2)p_3p_4 \exp(a + bx_1)\exp(a + bx_2)}{\{1 + \exp(a + bx_1)\}^4 \{1 + \exp(a + bx_2)\}^4 \{1 + \exp(a + bx_3)\}^2}$$

Hence the model is full rank only if $x_1 \neq x_2$, irrespective of x_3 .

Example 2: Near-singular model

Consider the model with parameter set,

$$\theta = [\phi_{1,1}, \phi_{1,2}, \phi_{1,3}, \phi_a, \lambda_1, \lambda_a]$$

This is a T/C/A(1,1) recovery model. It is full rank, but from the **PLUR** decomposition, we find, irrespective of $\phi_{1,3}$

$$\text{Det}(U) = \frac{(\phi_{1,2} - \phi_{1,1})}{(1 - \phi_{1,1})\lambda_1\phi_{1,1}\phi_a(1 - \phi_{1,2})\lambda_a\phi_{1,2}(1 - \phi_{1,3})}.$$

A complete theory?

- ❑ Maple (eg) can provide all the answers:
 - Form the derivative matrix
 - Find its rank
 - Solve the Lagrange equations (parameter redundant)
 - Form the PLUR decomposition (full rank)
- ❑ However it can **run out of memory** if used **routinely** for complex problems.
- ❑ Numerical procedures may then be used.
- ❑ The symbolic approach may still follow from identifying **exhaustive summaries** and **reduced form exhaustive summaries**

CJS model: removal of a cell

| | | |
|--------------|----------------------------|---|
| $\phi_1 p_2$ | $\phi_1 \phi_2 (1-p_2)p_3$ | $\phi_1 \phi_2 \phi_3(1-p_2)(1-p_3)p_4$ |
| | $\phi_2 p_3$ | $\phi_2 \phi_3(1-p_3)p_4$ |
| | | $\phi_3 p_4$ |

The set of **cells** here has a redundancy.

Examples of exhaustive summaries

- We see from the last slide that we may simplify the formation of the derivative matrix by identifying a *sufficient* set of means.
- In fact **we have already been doing that**, by not using the last cell in each multinomial, from each row.
- Note the effect of **missing data**. These may affect rank.

Recapture of Dippers, *Cinclus cinclus*



The table shows capture-recapture data for European Dippers in 1981-1986.

| | | | | | | | |
|------|----|----|----|----|----|----|----|
| 1981 | 22 | 11 | 2 | 0 | 0 | 0 | 0 |
| 1982 | 60 | | 24 | 1 | 0 | 0 | 0 |
| 1983 | 78 | | | 34 | 2 | 0 | 0 |
| 1984 | 80 | | | | 45 | 1 | 2 |
| 1985 | 88 | | | | | 51 | 0 |
| 1986 | 98 | | | | | | 52 |

Example 3 – Tag Returns Fisheries Model

Jiang et al (2007): Striped Bass,
Morone saxatilis.



$$\theta = [F, M_1, M_2, M_3, C_1, C_2, C_3, \lambda],$$

F – instantaneous fishing mortality rate

M_a – instantaneous natural mortality rate, at age a

C_a – selectivity coefficient for age a ($a > 3$ $C_a = 1$)

λ – reporting probability

P_{ijk} – probability fish tagged at age k , released year i
harvested and returned year j

$$P_{ijk} = \left[\prod_{v=i}^{j-1} \exp \{ - (F C_{k+v-i} + M_{k+v-i}) \} \right] \left[1 - \exp \{ - (F C_{k+j-i} + M_{k+j-i}) \} \right] \frac{F C_{k+j-i} \lambda}{F C_{k+j-i} + M_{k+j-i}}$$

Reparameterisation

A reparameterisation with entries of the form $F_y C_a \lambda_{y,a}$ and $F_y C_a + M_{y,a}$ could be used, or to simplify **D** further an alternative reparameterisation has s_1 to s_{12} equal to

$$s_{4(y-1)+a} = \frac{F_y C_a \lambda_y^Y \lambda_a^A}{F_y C_a + M_y^Y M_a^A}$$

and s_{13} to s_{24} equal to

$$s_{12+4(y-1)+a} = \exp(-F_y C_a - M_y^Y M_a^A)$$

for $a = 1, \dots, 4$ and $y = 1, \dots, 3$, with $C_4 = 1$, $\lambda_4^A = \lambda_3^A$, $M_4^A = M_3^A$, $\lambda_1^Y = 1$ and $M_1^Y = 1$.

Reparameterisation to produce a structurally simpler Q

The probability matrix written in terms of s_i is given below. Note that we now have 24 parameters, compared to the original 16.

$$Q = \begin{bmatrix} s_1(1 - s_{13}) & s_6s_{13}(1 - s_{18}) & s_{11}s_{13}s_{18}(1 - s_{23}) \\ 0 & s_5(1 - s_{17}) & s_{10}s_{17}(1 - s_{22}) \\ 0 & 0 & s_9(1 - s_{21}) \\ s_2(1 - s_{14}) & s_7s_{14}(1 - s_{19}) & s_{12}s_{14}s_{19}(1 - s_{24}) \\ 0 & s_6(1 - s_{18}) & s_{11}s_{18}(1 - s_{23}) \\ 0 & 0 & s_{10}(1 - s_{22}) \\ s_3(1 - s_{15}) & s_8s_{15}(1 - s_{20}) & s_{12}s_{15}s_{20}(1 - s_{24}) \\ 0 & s_7(1 - s_{19}) & s_{12}s_{19}(1 - s_{24}) \\ 0 & 0 & s_{11}(1 - s_{23}) \end{bmatrix}.$$

Conclusions from the fisheries example

- ❑ In this example Maple **lacks memory**.
- ❑ In this example we move from 16 to 24 parameters.
- ❑ We find a deficiency of 8 in the new parameter space. Thus the model is full rank.
- ❑ Note that Jiang et al., used numerical analysis and found a deficiency of 9, as the model is **near-singular** (revealed by a PLUR decomposition in the new parameterisation).

Example 4: a multi-state example

$$\log L = \sum_{r=1}^{N-1} \sum_{c=r+1}^N \sum_{i=1}^4 \sum_{j=1}^4 m_{i,j}^{(r,c)} \log \psi_{i,j}^{(r,c)},$$

$$\psi^{(r,c)} = (\Pi_{r+1} \Phi_r)' \quad \text{for } c = r + 1$$

$$\psi^{(r,c)} = \{\Pi_c \Phi_{c-1} (I - \Pi_{c-1}) \Phi_{c-2} \dots (I - \Pi_{r+1}) \Phi_r\}' \quad \text{for } c > r + 1.$$

Reparameterisation example

- ❑ Form a set of sufficient means from the **non-zero elements** of the Ψ .
- ❑ Form an exhaustive summary from the **non-zero elements** of the component matrices.
- ❑ The result is 14 parameter combinations.
- ❑ Resulting derivative matrix has rank 12.
- ❑ Determine the **reduced form exhaustive summary** from solving the Lagrange partial differential equations.
- ❑ Use extension theorem.
- ❑ Previously (Hunter and Caswell, 2008) a numerical approach has been used.

Section

- Introduction and motivation
- Definitions; general rules
- Use of symbolic algebra; expansion theorems
- The PLUR decomposition; exhaustive summaries; reparameterisation; identifiability; **weak identifiability**
- Future work

Weak identifiability: the Bayesian context

- A parameter θ is said to be **weakly identifiable** when $\pi(\theta|Y) \approx p(\theta)$.
- This is the counterpart to near-redundancy.
- For each parameter in a model, Garrett and Zeger(2000) considered the overlap of prior and posterior.
- Form $\tau = \int \min(p(\theta), \pi(\theta|Y))d\theta$.
- Garrett and Zeger suggest ad-hoc threshold of $\tau = 0.35$. This works well for ecological applications.

Recapture of Dippers, *Cinclus cinclus*

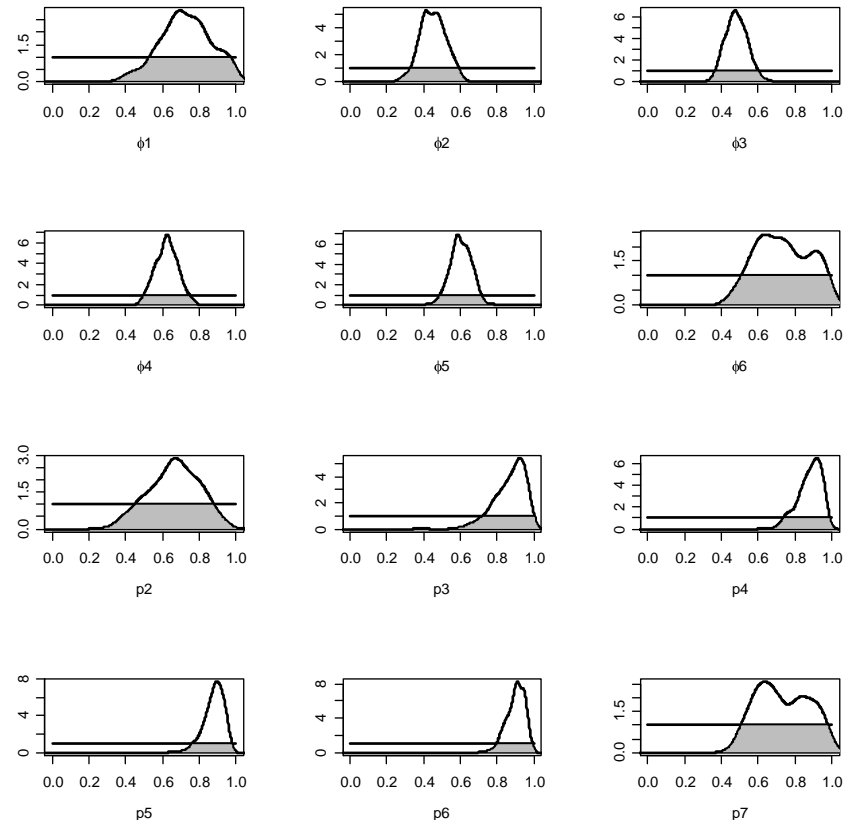


The table shows capture-recapture data for European Dippers in 1981-1986.

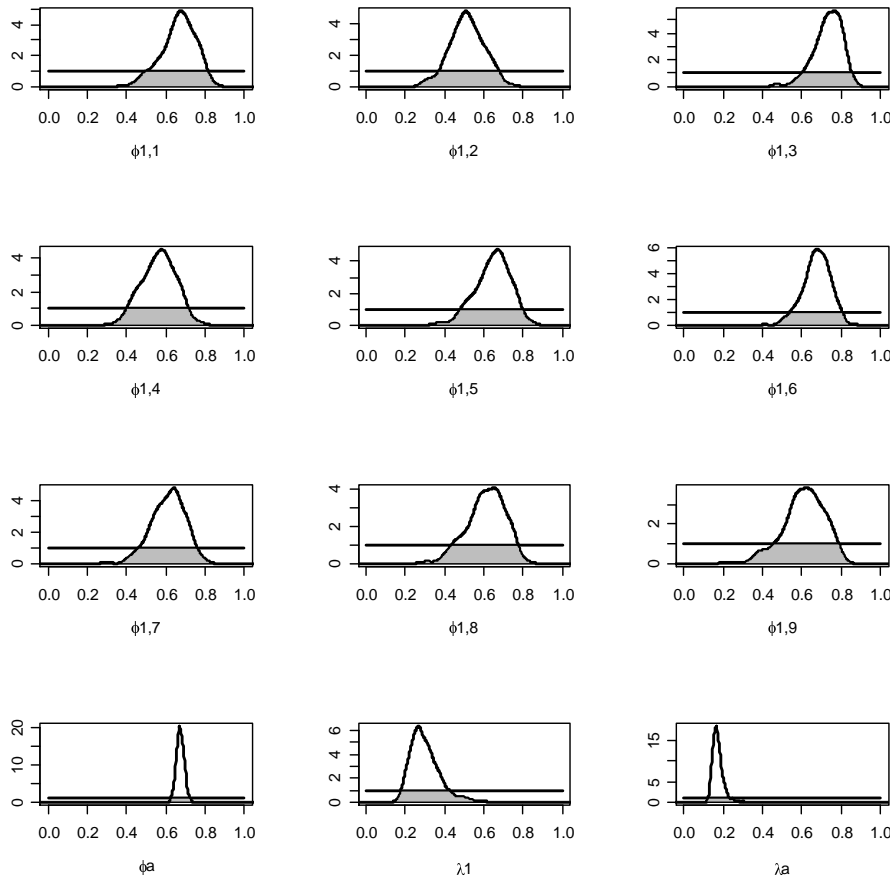
| | | | | | | | |
|------|----|----|----|----|----|----|----|
| 1981 | 22 | 11 | 2 | 0 | 0 | 0 | 0 |
| 1982 | 60 | | 24 | 1 | 0 | 0 | 0 |
| 1983 | 78 | | | 34 | 2 | 0 | 0 |
| 1984 | 80 | | | | 45 | 1 | 2 |
| 1985 | 88 | | | | | 51 | 0 |
| 1986 | 98 | | | | | | 52 |

A Bayesian perspective: the CJS model

- In population ecology we may devise models with parameters that cannot be estimated from the data.
- Symbolic algebra can be used to examine whether a model is parameter-redundant.
- In a Bayesian context, it is interesting to consider the overlap between priors, $p(\theta)$ and posteriors $\pi(\theta|x)$.

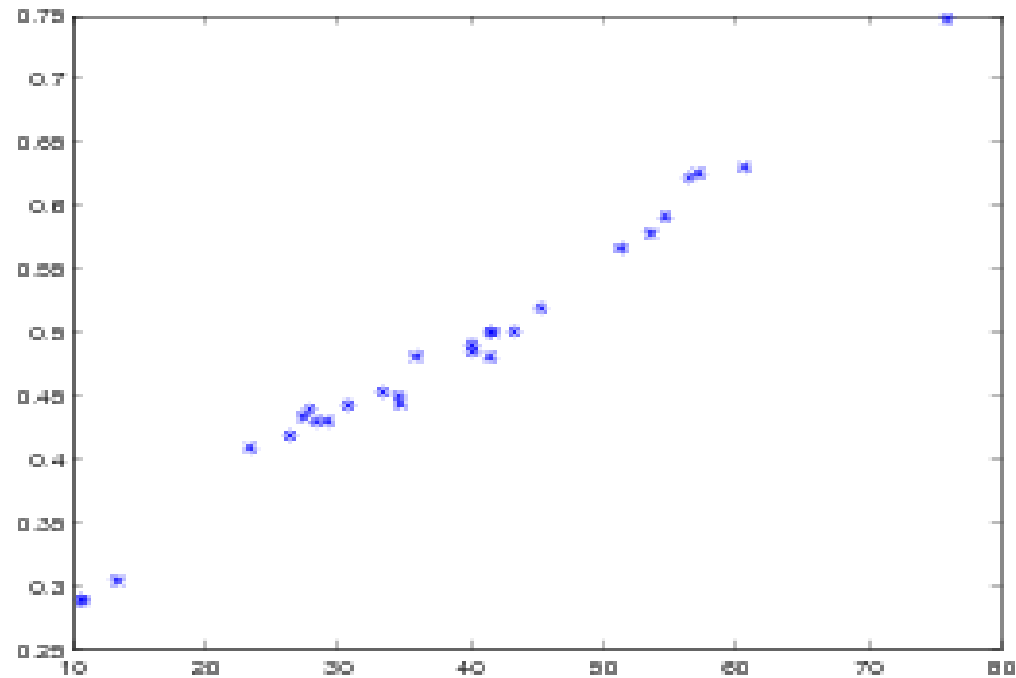


Male mallard, *Anas platyrhynchos*



Model: T/C/A_(1,1)
 $\phi_{1,i}$, ϕ_a , λ_1 , λ_a here
 only two
 parameters, ϕ_a and
 λ_a are strongly
 identified. The
 model is near-
 redundant.

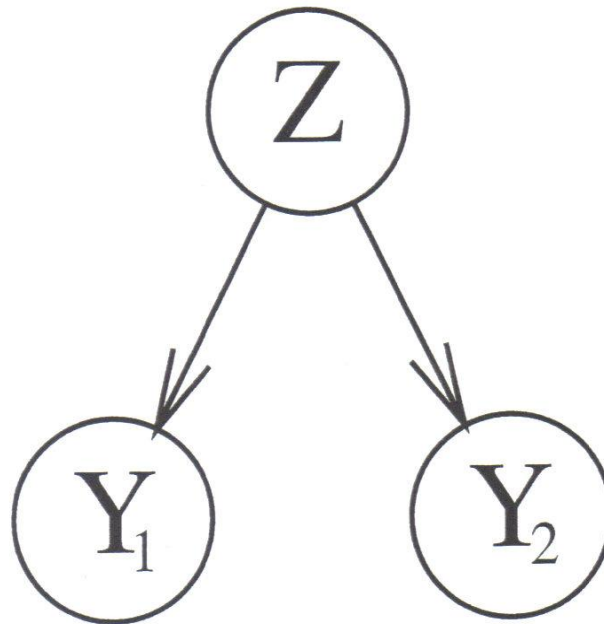
Relationship of overlap to interquartile range: simpler to calculate



Other areas

- ❑ Econometrics (Rothenburg)
- ❑ Compartment modelling (Walter)
- ❑ Contingency tables (Goodman)
- ❑ Naïve Bayesian Networks (Whiley)

A simple naïve Bayesian network



A naïve Bayesian network with a binary root node and two binary observable nodes.

Naïve Bayesian Networks in general

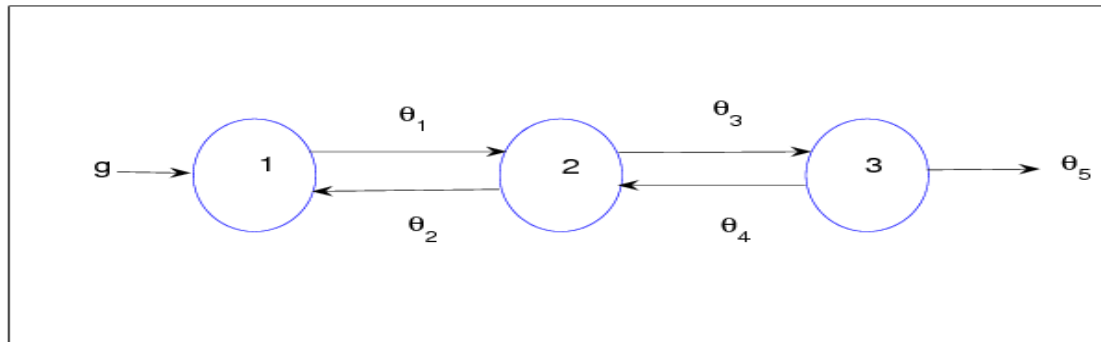
- We have n observable nodes, Y_1, \dots, Y_n , and a single observable node Z .
- All nodes are binary.
- $2n+1$ parameters: $p, \theta_{1|1}, \dots, \theta_{n|1}, \theta_{1|0}, \dots, \theta_{n|0}$.

$$P(y) = p \prod_{i=1}^n \theta_{i|1}^{y_i} (1 - \theta_{i|1})^{1-y_i} + (1-p) \prod_{i=1}^n \theta_{i|0}^{y_i} (1 - \theta_{i|0})^{1-y_i}$$

Naïve Bayesian network ctd

- In this example we can use a reparameterisation to show that
- For $n > 2$ the model is full rank
- We can use the PLUR decomposition to determine parameter redundant sub-models: for example, when $n=3$,
$$\text{Det}(U) = -p^3(1-p)^3(\theta_{1|1}-\theta_{1|0})^2(\theta_{2|1}-\theta_{2|0})^2(\theta_{3|1}-\theta_{3|0})^2.$$
- Previously conclusions followed a particular analysis.

Compartment models



Simple compartment model

$$\frac{dx_1}{dt} = -(\theta_1 + \theta_2)x_1 + \theta_3x_2 + u$$

$$\frac{dx_2}{dt} = \theta_2x_1 - (\theta_3 + \theta_4)x_2$$

$$y = x_1$$

Transfer function

$$Q(s) = \frac{s + \theta_3 + \theta_4}{s^2 + s(\theta_1 + \theta_2 + \theta_3 + \theta_4) + \theta_1\theta_3 + \theta_1\theta_4 + \theta_2\theta_4}$$

Estimable parameters are:

$$(\theta_1 + \theta_2), -\theta_2\theta_3, (\theta_3 + \theta_4).$$

Conclusion: new work

- ❑ Use of covariates
- ❑ Use of exhaustive summaries
- ❑ Use of PLUR decomposition
- ❑ Conclusions for complex examples using the **same approach**
- ❑ Overlap of priors and posterior

Acknowledgement

The work of Diana Cole was supported by the EPSRC grant for the NCSE.

References

- ❑ Cole, D.J., Morgan, B.J.T. and Titterton, D. M. (2010) Determining the parametric structure of models, *Mathematical Biosciences*.
- ❑ Cole, D.J. and Morgan, B.J.T. (2010) Parameter redundancy with covariates, *Biometrika*.
- ❑ Garrett and Zeger, (2000), *Biometrics*.
- ❑ Gimenez et al., (2003), *Biometrical J.*
- ❑ Gimenez, Morgan and Brooks, (2008), *J. Env. and Ecol. Stats*.
- ❑ Hunter and Caswell, (2008), *J. Env. and Ecol Stats*.
- ❑ Jiang et al., (2007), *JABES*.
- ❑ Whiley, (1999), Glasgow PhD thesis.

SECTION 7

STATE-SPACE MODELLING

References

Information about wildlife systems

- ❑ Diverse sources of data often exist:
 - Demographic studies
 - Population surveys (censuses)
- ❑ Trends in demographic processes (survival, productivity) and abundance are often investigated separately.
- ❑ But abundance and demography are related
 - Combined estimation is appealing

Questions

- ❑ How do we form models and likelihoods for the different sources of data?
- ❑ How do we combine likelihoods?
- ❑ Do we have to form exact likelihoods?
- ❑ Why not take estimates from some likelihoods and plug them into other likelihoods?

Fecundity

- The key demographic processes involve *survival*, *reproduction* and *movement*.
- Estimation of fecundity can be challenging, but is typically easier than the estimation of survival. For example, Cory's shearwaters (*Calonectis diomedea*) lay just one egg.

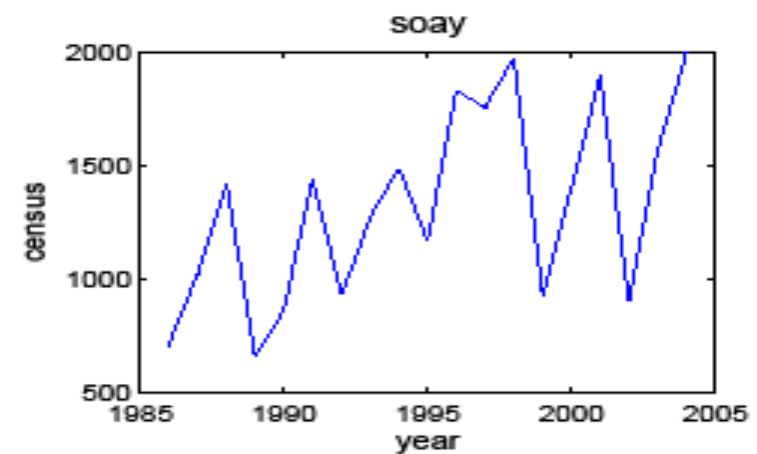
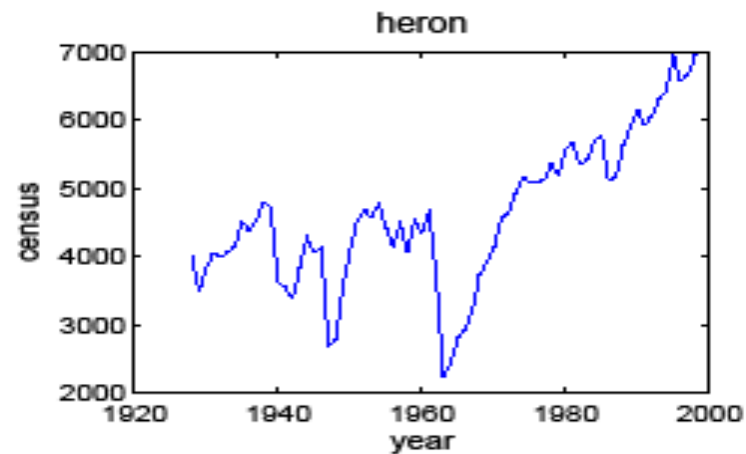
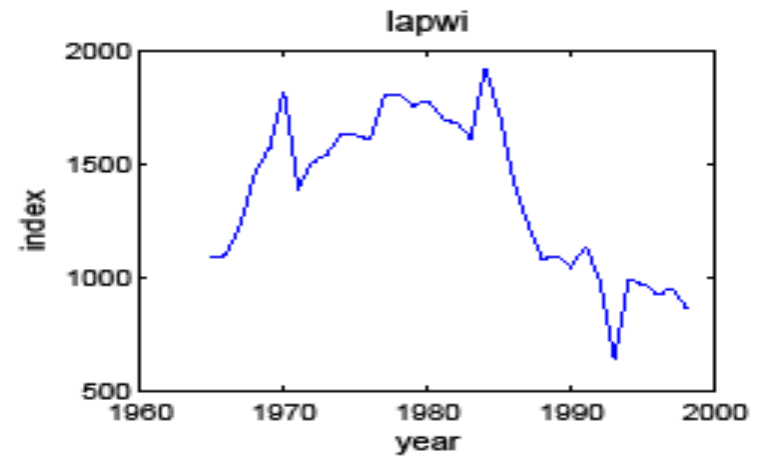
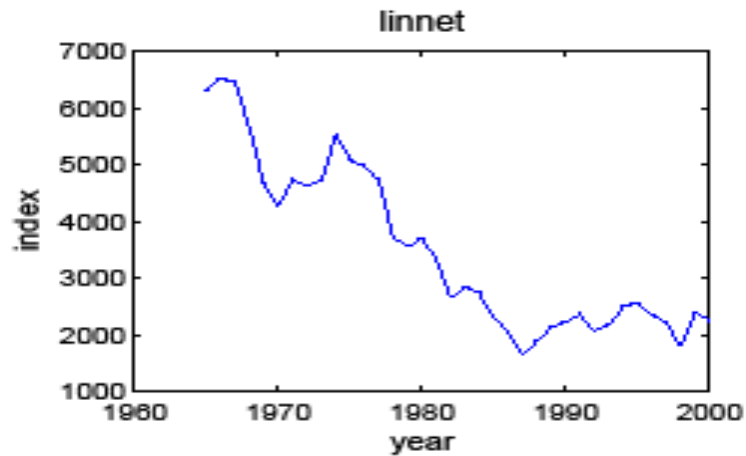


Population surveys

- We obtain information on abundance from population surveys
- Typically the surveys are annual and the end-result is a single number summarising abundance:
 - number of breeders at time t
 - total number of animals
- The numbers are either estimates or indices of the actual numbers
- Populations exhibit a range of dynamics
 - Compare dynamics of linnets, lapwings, herons and Soay sheep



Dynamics



Modelling survey data I

- There are, potentially, several aims in modelling survey data:
 - To estimate true population size;
 - To estimate biologically relevant parameters;
 - To predict future population sizes.
- We shall provide an all-purpose method
- The first task is to build a model for all variables, observable or not

Modelling survey data II

- Abundance and demography are related:
 - Abundance at time t is a function of survival, fecundity and abundance at time $t-1$.
- A state-space model framework is often used to make this idea explicit.

State-space models I

- A state-space model consists of 2 stochastic parts:
 1. A dynamic model describing the evolution of a 'state' from one year to the next:

$$n_{t+1} = T_t n_t + \varepsilon_t$$

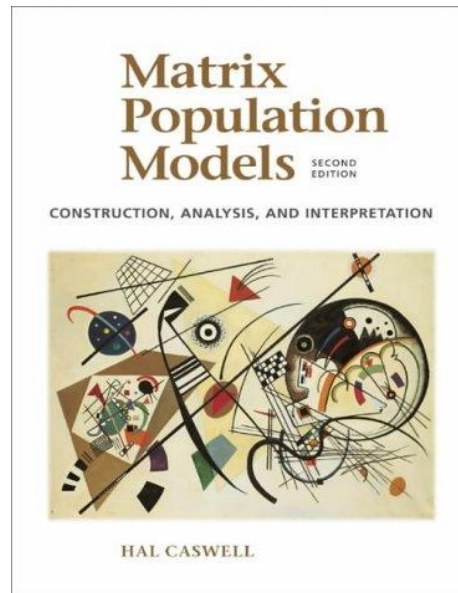
2. A model linking the observations to the 'state':

$$y_t = Z_t n_t + \eta_t$$

- The terms ε_t and η_t are errors incorporating:
 - ε : process variability
 - η : measurement uncertainty

State-space models II

- State-space models can be recognised as stochastic versions of



- But they are much more general (e.g. non-linear etc)

General definition

- A state-space model consists of two stochastic processes:
 - An unobserved state process, n_t , that is a function of past values, n_{t-1}
 - State process density: $g(n_t | n_{t-1}; \theta)$.
 - A known observation process, y_t , that is a function of the current state
 - Observation process density: $f(y_t | n_t; \theta)$.
- θ : system parameters
 - Could be time and space-dependent

Inference for state-space models

- Inference problems for state-space models fall into three broad categories:

1. Filtering: $p(n_t | y_{1:t})$
2. Smoothing: $p(n_t | y_{1:T})$
3. Prediction: $p(n_{t+m} | y_{1:t}), m > 0.$

Inference for state-space models

- Inference problems for state-space models fall into three broad categories:
 1. Filtering: $p(n_t | y_{1:t})$
 2. Smoothing: $p(n_t | y_{1:T})$
 3. Prediction: $p(n_{t+m} | y_{1:t}), m > 0.$
- Filtering is the key problem, since it forms the basis of everything else
- In general, it involves evaluating high-dimensional integrals, which can be problematic.
- But see.....Modelling Population Dynamics Workshop (Andrew Thomas)

Linear Gaussian model

- A special case is the linear Gaussian model:

$$\begin{aligned}n_{t+1} &= T_t n_t + \varepsilon_t & \varepsilon_t &\sim N(0, Q_t) \\ y_t &= Z_t n_t + \eta_t & \eta_t &\sim N(0, H_t) \\ & & n_1 &\sim N(a_1, P_1)\end{aligned}$$

n_t : numbers at age t

T_t : Leslie matrix

Y_t : observed abundances

θ includes survival, fecundity etc

- In this case, the Kalman algorithms can be used to estimate n_t and θ
 - No need for numerical or stochastic integration
- Details in Besbeas et al (2002).

The Kalman filter (KF)

- The following equations constitute the KF (Durbin and Koopman, 2001):

$$v_t = y_t - Z_t a_t$$

$$F_t = Z_t P_t Z_t' + H_t$$

$$M_t = P_t Z_t'$$

$$t = 1, \dots, T$$

$$a_{t|t} = a_t + M_t F_t^{-1} v_t$$

$$P_{t|t} = P_t - M_t F_t^{-1} M_t'$$

$$a_{t+1} = T_t a_{t|t}$$

$$P_{t+1} = T_t P_{t|t} T_t' + Q_t$$

with $a_1 = E(n_1)$ and $P_1 = \text{Cov}(n_1)$.

- The algorithm can be derived and expressed in several ways.

Related algorithms

- ❑ Missing values, prediction, smoothing follow from basic output:
 - Missing values/prediction
 - ❑ Use of a_{t+1}
 - ❑ Natural extension to m-steps
 - Smoothed estimates
 - ❑ Result from $a_{t|t}$ and another set of explicit recursions

The likelihood function

- ▣ The likelihood function also follows from basic output:

$$\log L_{kf}(\theta) = \text{constant} - \frac{1}{2} (\log |F_t| + v_t' F_t^{-1} v_t)$$

- ▣ The parameters θ include demographic parameters ϕ , f from the matrix population model.

Combining likelihoods

- ▣ Assuming independence, we integrate the demographic and survey information on θ into a joint likelihood:

$$L_j(\phi, \lambda, p, f, \sigma) = L_{mr}(\phi, \lambda, p) \times L_{kf}(\phi, f, \sigma)$$

**Mark
recapture
recovery
data**

**Census
data**

Integrated Population Modelling
Combining Sources of Information

$$L_{mrr}(\phi, \lambda, p)$$

$$L_{kf}(\phi, f, \sigma)$$

$$\begin{aligned} L_j(\phi, \lambda, p, f, \sigma) \\ = L_{mrr}(\phi, \lambda, p) \times L_{kf}(\phi, f, \sigma) \end{aligned}$$

Advantages of integration

1. Simultaneous description of all the data
2. Generally more precise parameter estimators
3. Reduction in the correlation of estimators arising from census data alone
4. Coherent estimation of parameters not estimable from separate analyses

Example I: Lapwings (Besbeas et al, 2002)



- The Lapwing in Britain has been declining for several years.
- Amber list of species of conservation concern in Britain.
- Regarded as an indicator species for farmland birds.

Lapwing ring-recovery data

- Data on Lapwings ringed as chicks during the years 1963-1997

| Year Ringed | Number Ringed | Year Recovered | | | | | |
|-------------|---------------|----------------|------|------|-----|------|------|
| | | 1964 | 1965 | 1966 | ... | 1997 | 1998 |
| 1963 | 1147 | 14 | 4 | 1 | ... | 0 | 0 |
| 1964 | 1285 | | 20 | 3 | ... | 0 | 0 |
| 1965 | 1106 | | | 10 | ... | 0 | 0 |
| ... | ... | | | | ... | ... | ... |
| 1996 | 4488 | | | | | 7 | 0 |
| 1997 | 4339 | | | | | | 5 |

Lapwing survey data

- Use an **index** based on the Common Birds Census.
- Annual territory counts are made at a number of survey sites, and from these an index is estimated from a generalized linear model.
- Lapwing index is available from 1965 to 1998.

| Year | Population Size |
|------|-----------------|
| 1965 | 1092 |
| 1966 | 1100 |
| 1967 | 1234 |
| ... | ... |
| 1997 | 953 |
| 1998 | 866 |



SSM: Observation model

- ▣ Let $N_{1,t}$ denote the true number of one-year old female birds at time t , and $N_{a,t}$ denote the number of adult female birds at time t .
- ▣ We assume that

$$y_t \sim N(N_{a,t}, \sigma^2)$$

where σ^2 is essentially an estimation error.

- ▣ We then impose a process model on the underlying population sizes

SSM: Process model

- A natural model will be to assume that

$$N_{a,t} \sim \text{Bin}(N_{1,t-1} + N_{a,t-1}, \phi_{a,t-1})$$

and

$$N_{1,t} \sim \text{Po}(N_{a,t-1} f_{t-1} \phi_{1,t-1})$$

- Here f_t denotes the productivity rate (number of females per female)
- These can be easily and accurately approximated by Normal distributions, even for small population sizes.

Normal SSM

$$\begin{pmatrix} N_1 \\ N_a \end{pmatrix}_{t+1} = \begin{pmatrix} 0 & f\phi_1 \\ \phi_a & \phi_a \end{pmatrix} \begin{pmatrix} N_1 \\ N_a \end{pmatrix}_t + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_a \end{pmatrix}_t$$

$$y_t = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} N_1 \\ N_a \end{pmatrix}_t + \eta_t$$

with

$$\text{Var}(\varepsilon_{1,t}) = N_{a,t-1} f\phi_1$$

$$\text{Var}(\varepsilon_{a,t}) = (N_{1,t-1} + N_{a,t-1})\phi_a(1 - \phi_a)$$

$$\text{Var}(\eta_t) = \sigma^2$$

Parameter Modelling

- We allow each of our parameters potentially to vary with time as follows:

- Survival:

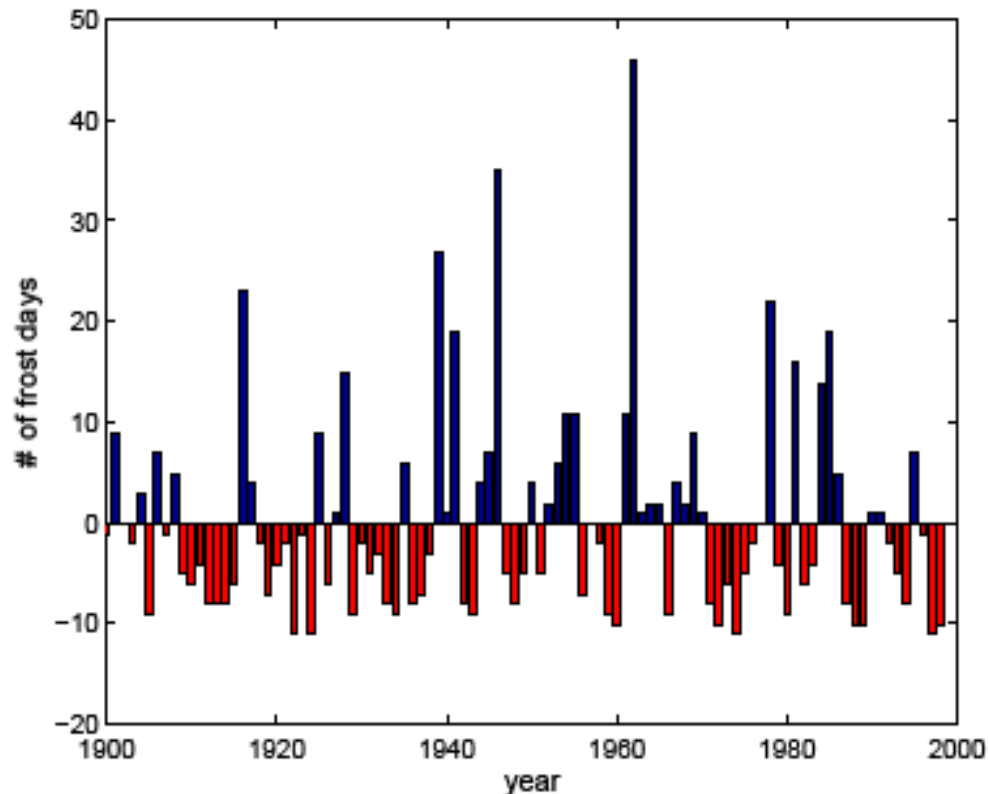
- $\text{logit } \Phi_{1,t} = \alpha_1 + \beta_1 \text{fdays}_t$

- $\text{logit } \Phi_{a,t} = \alpha_1 + \beta_1 \text{fdays}_t$

- Productivity & recovery:

- $\log f_t = \alpha_\rho + \beta_\rho t$

- $\text{logit } \lambda_t = \alpha_\lambda + \beta_\lambda t$



Likelihoods and Integrated Model

- Component likelihoods:

- Mrr: $L_{mrr}(\phi_1, \phi_a, \lambda)$

- Survey: $L_{kf}(\phi_1, \phi_a, f, \sigma)$

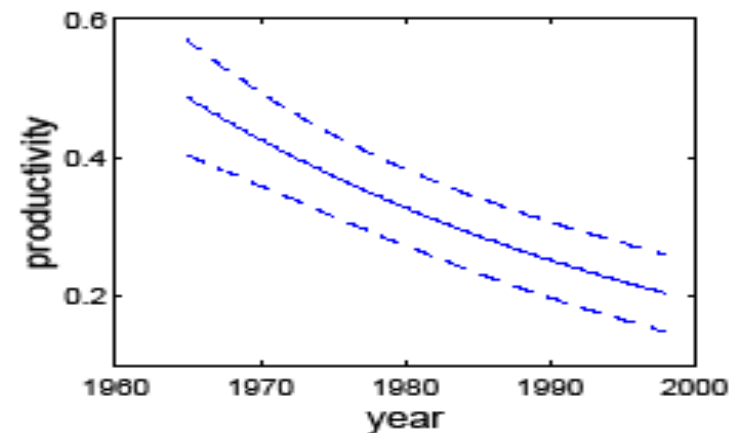
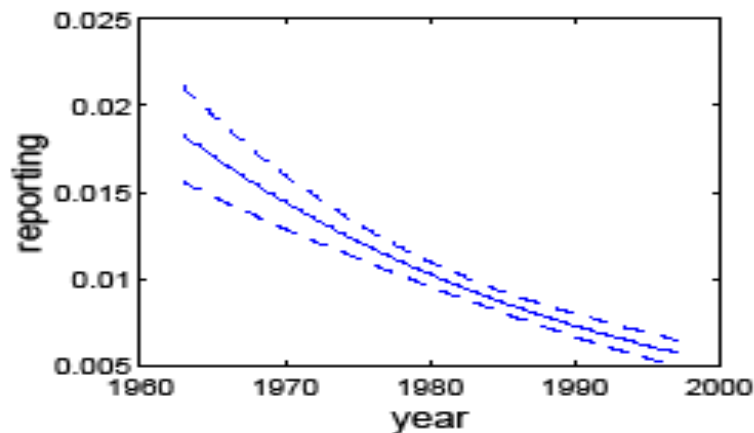
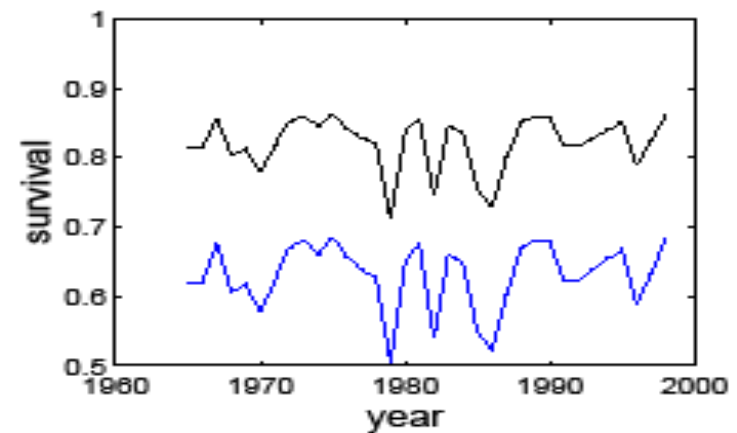
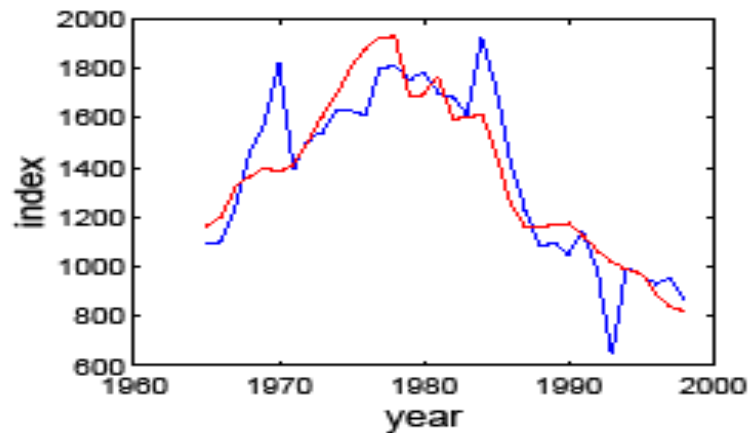
- Joint Likelihood:

- Assuming mrr and census data are independent

$$L_j(\phi_1, \phi_a, \lambda, f, \sigma) = L_{mrr}(\phi_1, \phi_a, \lambda) \times L_{kf}(\phi_1, \phi_a, f, \sigma)$$

Lapwing results

■ Model $\Phi_1(\text{fdays}), \Phi_a(\text{fdays})/\lambda(\text{year})/f(\text{year})$



Parameter estimates

| Parameter | Estimate | | SE | |
|---------------------|----------|--------|-------|-------|
| | Mrr | Joint | Mrr | Joint |
| Φ_1 intercept | 0.516 | 0.523 | 0.067 | 0.067 |
| Φ_1 slope | -0.024 | -0.023 | 0.007 | 0.007 |
| Φ_a intercept | 1.501 | 1.521 | 0.068 | 0.069 |
| Φ_a slope | -0.036 | -0.028 | 0.005 | 0.004 |
| λ intercept | -4.567 | -4.563 | 0.035 | 0.035 |
| λ slope | -0.573 | -0.584 | 0.064 | 0.063 |
| f intercept | | -1.151 | | 0.088 |
| f slope | | -0.432 | | 0.074 |
| σ | | 159.47 | | 21.87 |

- Minor differences between mrr and joint estimates/SEs
- The original Poisson/binomial model provides very similar results – but is more difficult to fit
- Integrated modelling allows estimation of productivity

Assumptions made

- ❑ Normal errors
- ❑ Linear state-space model
- ❑ Known Kalman filter starting values
- ❑ Dispersion properly ascribed
- ❑ Independence between different data sets
- ❑ Observation variance cannot depend on the current state(s)

How to initiate the filter

- ▣ Using prior information (when available)

Otherwise

- ▣ Unconditional initialisation
- ▣ Diffuse initialisation
 - Approximate diffuse
 - Exact diffuse
- ▣ Maximum-likelihood
- ▣ Use of stable age-distribution

Using the stable age-distribution

- The initial mean vector a_1 is taken to be proportional to the stable age distribution, with a maximal positive CI, e.g.

$$a_1 \propto v, \quad P_1 = \text{diag}((a_1/1.96)^2)$$

- Practical issues with v :
 - Unknown parameters is Leslie matrix, T
 - Time-varying matrix model, T_t .
- Details in Besbeas and Morgan (2007a)

Pros and cons of the methods

- ❑ Unconditional initialisation: limited to stationary systems.
- ❑ Approximate diffuse initialisation: arbitrary decisions are required.
- ❑ Exact diffuse initialisation: difficult for complex models and multivariate observations.
- ❑ Maximum likelihood: Multiple and/or boundary optima.
- ❑ Use of stable age-distribution: simple, and uses the Leslie matrix. Best in simulation comparisons.

Approximate Integration

- Recall we maximise:

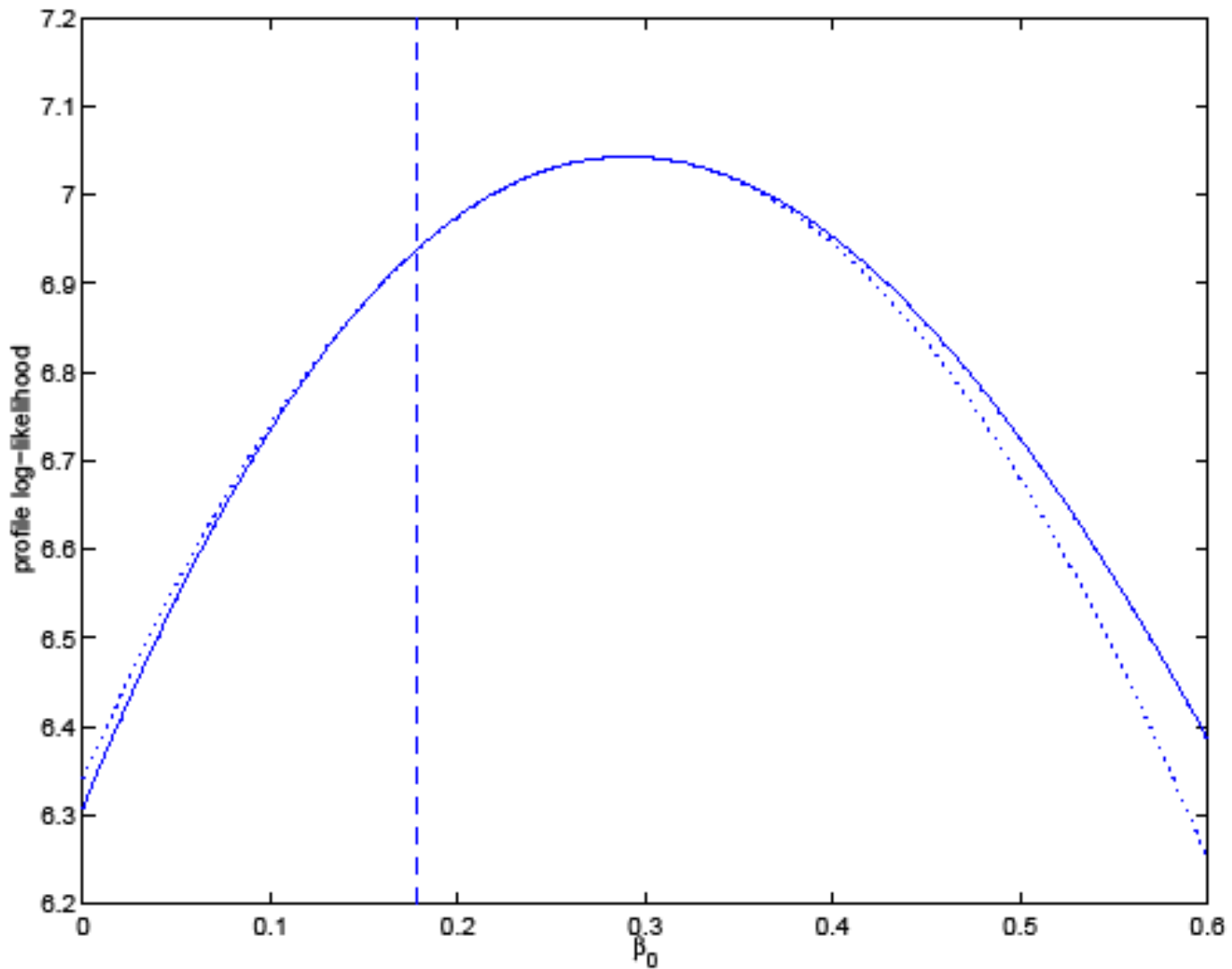
$$L_j(\phi, \lambda, f, \sigma) = L_{mrr}(\phi, \lambda) \times L_{kf}(\phi, f, \sigma)$$

- Calculation of L_j difficult when L_{mrr} is derived from specialised programs (e.g. MARK) or raw data not available.
- Simplify calculations using asymptotic distribution of MLEs:

$$\hat{\theta} \sim N(\theta, \Sigma) \quad \text{replacing } \Sigma \text{ by } \hat{\Sigma}.$$

- Equivalently,

$$\begin{aligned} \log L_{mrr}(\theta) &\approx \text{const} - 0.5(\hat{\theta} - \theta)' \hat{\Sigma}^{-1}(\hat{\theta} - \theta) \\ &\equiv \log \hat{L}_{mrr}(\theta; \hat{\theta}, \hat{\Sigma}) \end{aligned}$$



Comparison of $L_{mrr}(\theta)$ and $\hat{L}_{mrr}(\theta; \hat{\theta}, \hat{\Sigma})$

Approximate Joint Likelihood

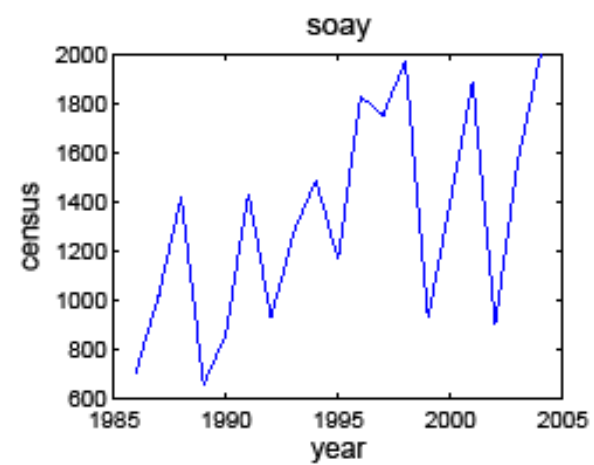
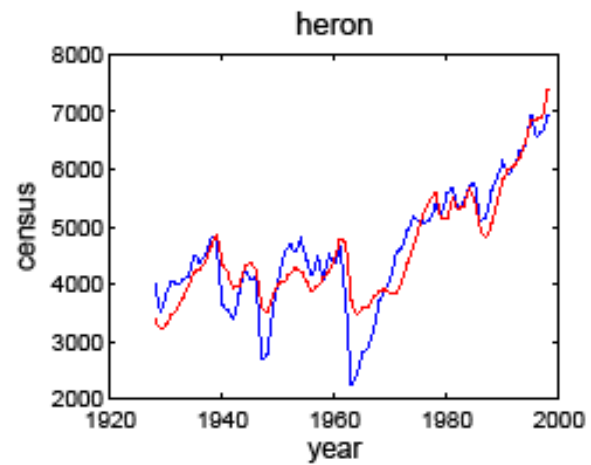
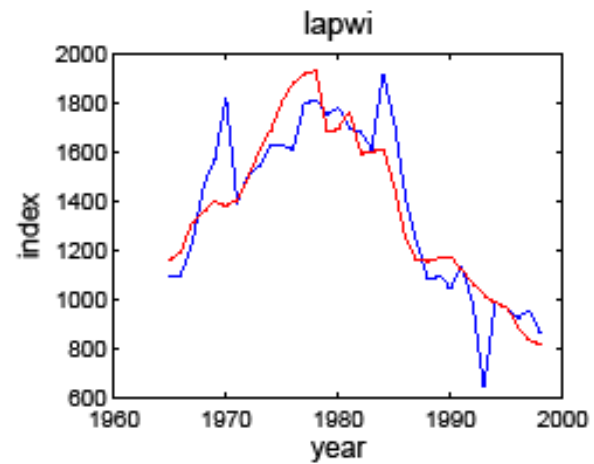
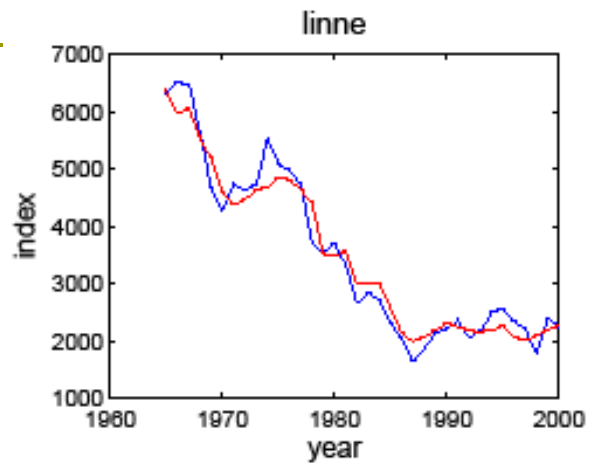
- Now maximise

$$\hat{L}_j(\phi, f, \sigma) = \hat{L}_{mrr}(\phi, \hat{\phi}, \hat{\Sigma}) \times L_{kf}(\phi, f, \sigma)$$

- Nuisance parameters in L_{mrr} (e.g. λ) omitted – details in Besbeas et al (2003)

Lapwing Parameter Estimates

| Parameter | Estimate | | Standard Error | |
|---------------------|----------|-------------|----------------|-------------|
| | Exact | Approximate | Exact | Approximate |
| ϕ_1 intercept | 0.5231 | 0.5226 | 0.0679 | 0.0678 |
| ϕ_1 slope | -0.0228 | -0.0227 | 0.0070 | 0.0070 |
| ϕ_a intercept | 1.5210 | 1.5191 | 0.0693 | 0.0686 |
| ϕ_a slope | -0.0279 | -0.0280 | 0.0045 | 0.0045 |
| λ intercept | -4.5632 | | 0.0352 | |
| λ slope | -0.5841 | | 0.0637 | |
| ρ intercept | -1.1513 | -1.1489 | 0.0886 | 0.0876 |
| ρ slope | -0.4323 | -0.4314 | 0.0743 | 0.0740 |
| σ | 159.47 | 159.61 | 22.06 | 21.87 |



Multi-site state-space models

- Basic structure of the state space model is the same:

$$n_{t+1} = T_t n_t + \varepsilon_t$$

- State vector, n_t will now include site-specific states

$$y_t = Z_t n_t + \eta_t$$

- Observations y_t will now be multivariate as site-specific observations will be made at each time point.

**Multi-site
mark
recapture
data**

**Multi-site
census
data**

Integrated Population Modelling
Combining Sources of Information

$$L_{mrr}(\phi, \psi, p)$$

$$L_{kf}(\phi, \psi, f, \sigma)$$

$$L_j(\phi, \psi, p, f, \sigma) \\ = L_{mrr}(\phi, \psi, p) \times L_{kf}(\phi, \psi, f, \sigma)$$

Advantages of MULTI-SITE integration

1. Simultaneous description of all the data
2. **Generally more precise parameter estimators**
3. Reduction in the correlation of estimators arising from census data alone
4. Coherent estimation of parameters not estimable from separate analyses

Simulation

Ringed animals are located in two distinct geographical sites. The animals are marked at an initial time point and then on subsequent occasions the sites are revisited and any re-sightings of marked individuals are recorded and recoveries of dead marked animals are made.

Suppose also that total census counts are made at each time period for each of the sites.

The simulation is run for 5 periods of recapture 15 census counts.

Simulation

- Parameters arising from mark-recapture-recovery data:
 - $\phi(r)$: annual survival probability at site r
 - $p(r)$: capture probability at site r
 - λ : recovery probability
 - $\psi(r,s)$: movement probability of an animal from site r to site s
- Standard Arnason-Schwarz Model

Simulation

- ▣ $N^J(x)$: Number of juvenile animals in site x
- ▣ $N^A(x)$: Number of adult animals in site x

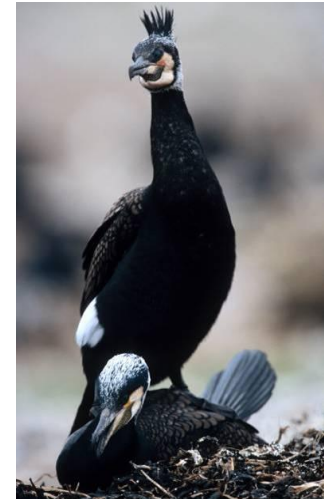
$$\begin{pmatrix} N^J(1) \\ N^J(2) \\ N^A(1) \\ N^A(2) \end{pmatrix}_t = \begin{pmatrix} 0 & 0 & f\phi(1)\psi(1,1) & f\phi(2)\psi(2,1) \\ 0 & 0 & f\phi(1)\psi(1,2) & f\phi(2)\psi(2,2) \\ \phi(1)\psi(1,1) & \phi(2)\psi(2,1) & \phi(1)\psi(1,1) & \phi(2)\psi(2,1) \\ \phi(1)\psi(1,2) & \phi(2)\psi(2,2) & \phi(1)\psi(1,2) & \phi(2)\psi(2,2) \end{pmatrix} \begin{pmatrix} N^J(1) \\ N^J(2) \\ N^A(1) \\ N^A(2) \end{pmatrix}_{t-1} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}_t$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}_t = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} N^J(1) \\ N^J(2) \\ N^A(1) \\ N^A(2) \end{pmatrix}_t + \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}_t$$

Simulation Results

| Parameter | Simulated Value | MRR Data Only | MRR and Census |
|-------------|-----------------|-----------------|-----------------|
| $\phi(1)$ | 0.85 | 0.832 (0.0999) | 0.807 (0.0798) |
| $\phi(2)$ | 0.41 | 0.316 (0.1064) | 0.325 (0.0970) |
| $p(1)$ | 0.00 | 0.157 (0.1291) | 0.184 (0.1207) |
| $p(2)$ | 0.41 | 0.301 (0.1889) | 0.263 (0.1669) |
| $\psi(1,2)$ | -2.20 | -2.238 (0.2101) | -2.140 (0.1153) |
| $\psi(2,1)$ | -1.39 | -1.321 (0.1893) | -1.354 (0.1691) |
| λ | 0.00 | -0.017 (0.1012) | -0.019 (0.1009) |
| f | 0.10 | - | 0.185 (0.0714) |
| σ | 4.61 | - | 4.448 (0.2089) |

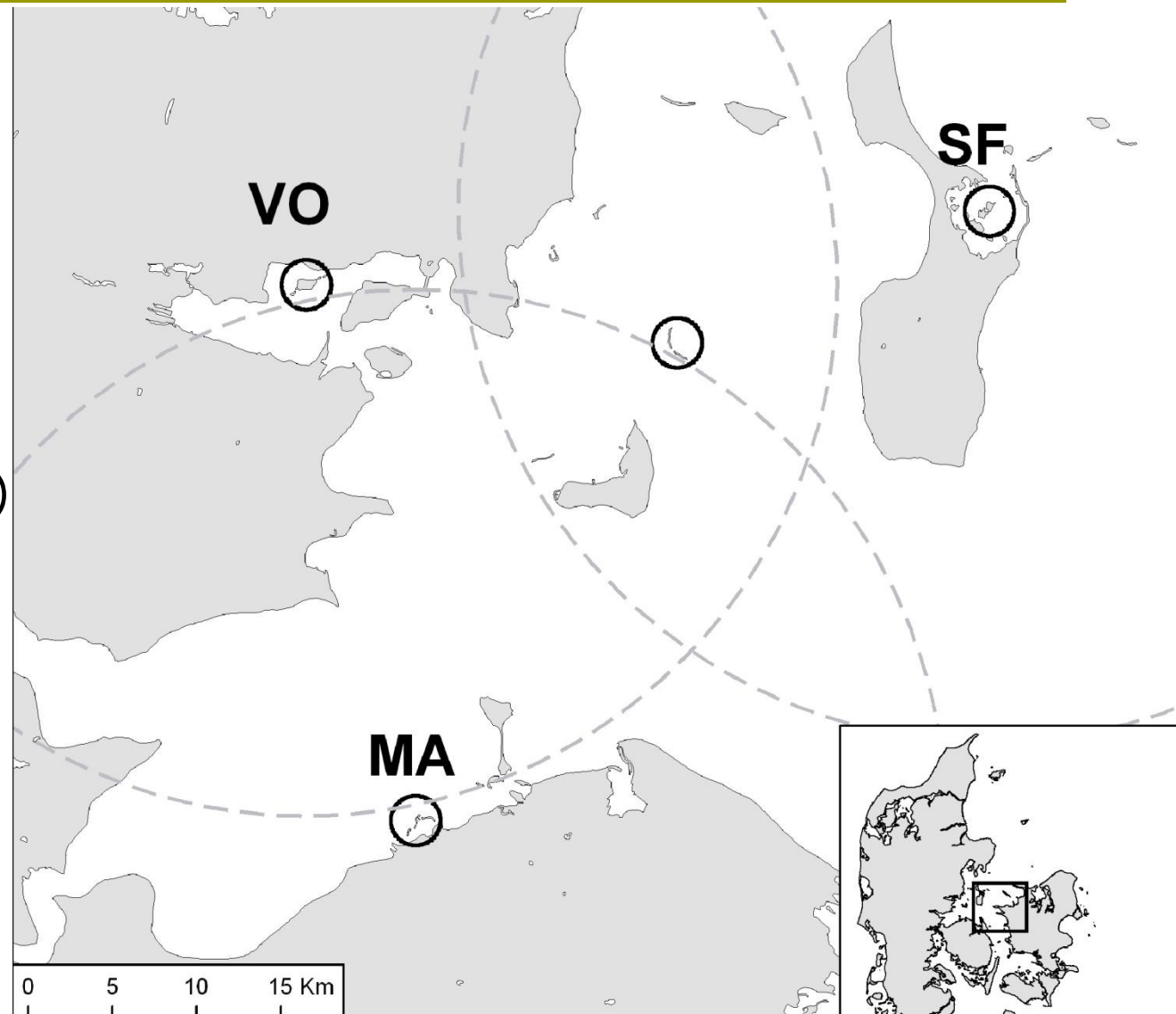
Example II: Cormorants (McCrea et al, 2010)



Great Cormorant, *Phalacrocorax carbo sinensis*

Three
neighbouring
colonies:

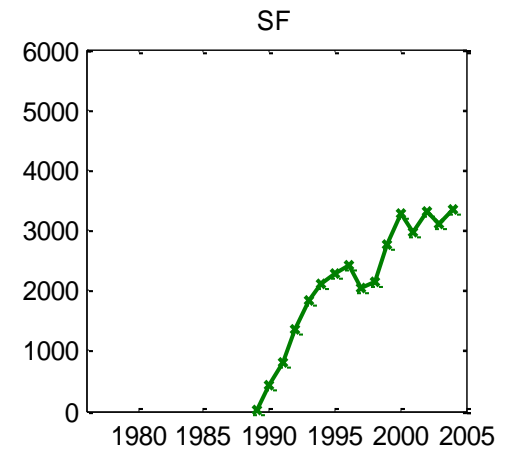
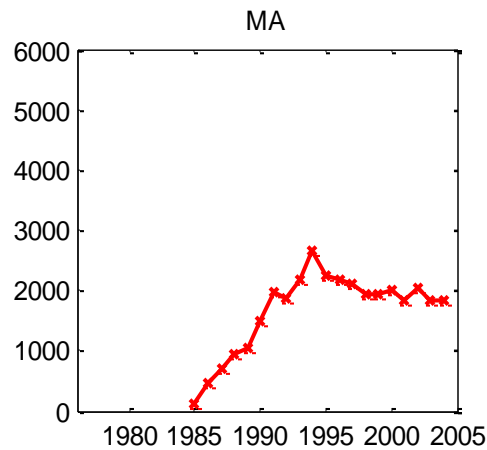
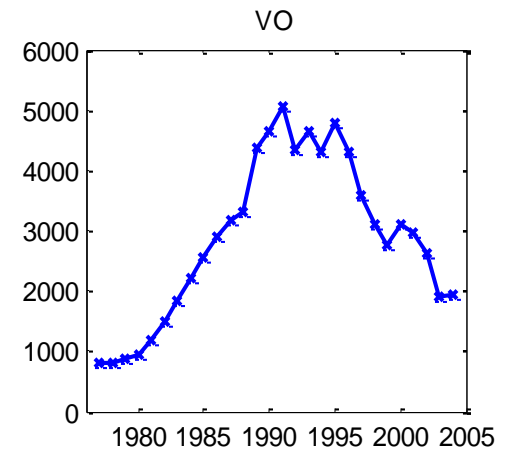
- Vorsø (VO)
- Mågeøerne (MA)
- Stavns Fjord (SF)



- **Recapture data collected between 1989 and 1994**
- **Initial ringing was carried out on non-breeding individuals**
- **Birds were not recaptured until they had become breeding individuals**
- **Must model transition between non-breeding and breeding states as well as geographical movements between the colonies**



Multi-site Cormorant census data



Transitions

| | | TO | | | | | | | | |
|------|-----------------|-----------------|-----------------|-----------------|-----------------|----|----|--------------------|----|----|
| | | vo ₁ | ma ₁ | sf ₁ | vo | ma | sf | VO | MA | SF |
| FROM | vo ₁ | | | | | | | | | |
| | ma ₁ | | | | Natal Dispersal | | | | | |
| | sf ₁ | | | | | | | | | |
| | vo | | | | | | | | | |
| | ma | | | | Non Maturation | | | Recruitment | | |
| | sf | | | | | | | | | |
| | VO | | | | | | | | | |
| | MA | | | | | | | Breeding Dispersal | | |
| | SF | | | | | | | | | |

State and Observation Equations

$$\begin{pmatrix} N_{VO} \\ N_{MA} \\ N_{SF} \\ B_{VO} \\ B_{MA} \\ B_{SF} \end{pmatrix}_t = \begin{pmatrix} \dots \\ \dots \\ \dots \end{pmatrix} \begin{pmatrix} N_{VO} \\ N_{MA} \\ N_{SF} \\ B_{VO} \\ B_{MA} \\ B_{SF} \end{pmatrix}_t + \begin{pmatrix} \varepsilon_{VO}^N \\ \varepsilon_{MA}^N \\ \varepsilon_{SF}^N \\ \varepsilon_{VO}^B \\ \varepsilon_{MA}^B \\ \varepsilon_{SF}^B \end{pmatrix}_t$$

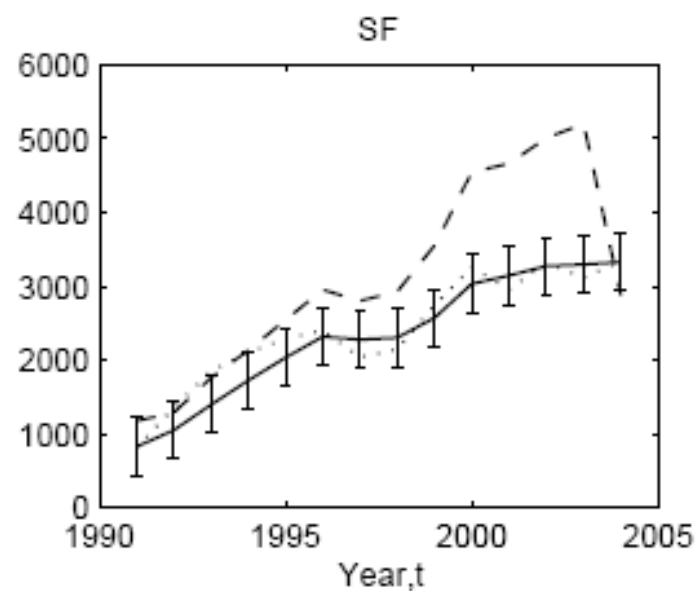
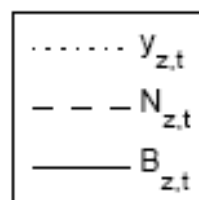
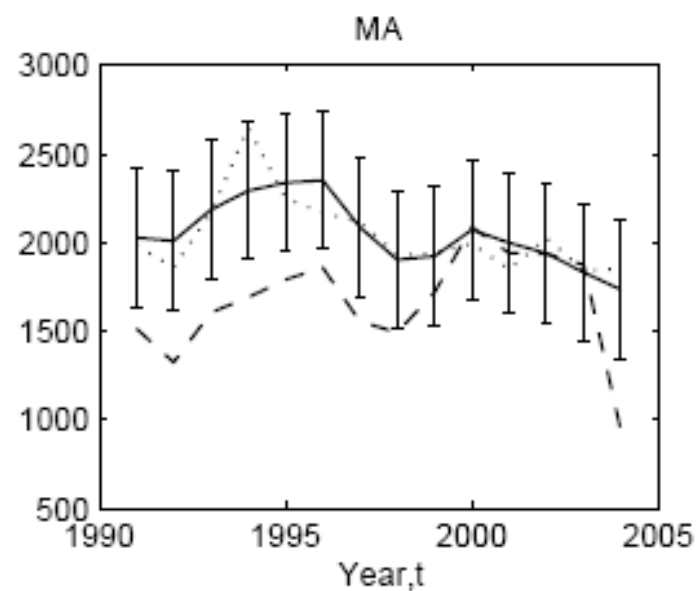
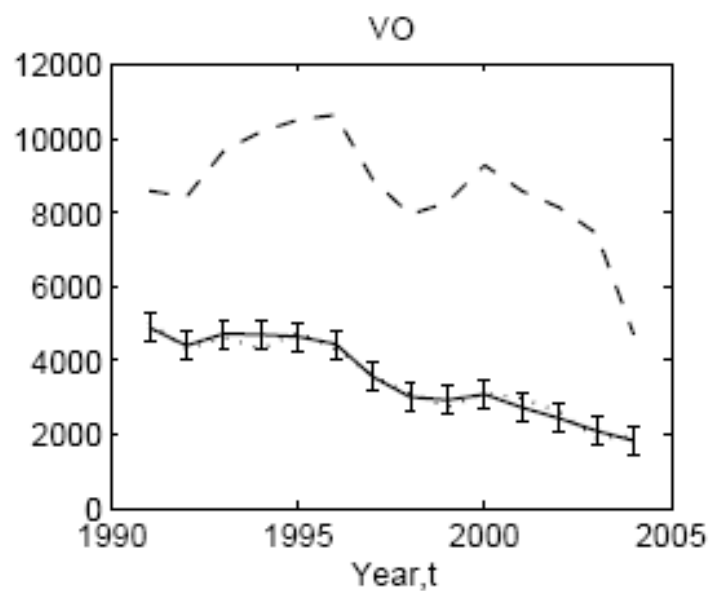
$$\begin{pmatrix} y_{VO} \\ y_{MA} \\ y_{SF} \end{pmatrix}_t = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} N_{VO} \\ N_{MA} \\ N_{SF} \\ B_{VO} \\ B_{MA} \\ B_{SF} \end{pmatrix}_t + \begin{pmatrix} \eta_{VO} \\ \eta_{MA} \\ \eta_{SF} \end{pmatrix}_t$$

Model Selection

| Model | n | AIC | Δ AIC |
|---|----|------|--------------|
| $\phi(\text{from} * \text{state} + \text{time}), B(\text{from} * \text{to}), N(\text{to}), R(\text{from} + T), f(\text{site}), \sigma(\cdot)$ | 70 | 13.2 | 0.0 |
| $\phi(\text{from} * \text{state} + \text{time}), B(\text{from}), N(\text{to}), R(\text{from} + T), f(\text{site}), \sigma(\cdot)$ | 67 | 14.4 | 1.2 |
| $\phi(\text{from} * \text{state} + \text{time}), B(\text{to}), N(\text{to}), R(\text{from} + T), f(\text{site}), \sigma(\cdot)$ | 67 | 16.4 | 3.2 |
| $\phi(\text{from} * \text{state} + \text{time}), B(\text{from}), N(\text{from} * \text{to}), R(\text{from} + T), f(\text{site}), \sigma(\cdot)$ | 70 | 16.9 | 3.8 |
| $\phi(\text{from} * \text{state} + \text{time}), B(\cdot), N(\text{to}), R(\text{from} + T), f(\text{site}), \sigma(\cdot)$ | 65 | 19.0 | 5.8 |
| $\phi(\text{from} * \text{state} + \text{time}), B(\cdot), N(\text{from} * \text{to}), R(\text{from} + T), f(\text{site}), \sigma(\cdot)$ | 68 | 21.5 | 8.4 |
| $\phi(\text{from} * \text{state} + \text{time}), B(\cdot), N(\text{to}), R(\text{from} + T), f(\cdot), \sigma(\cdot)$ | 63 | 27.7 | 14.5 |
| $\phi(\text{from} * \text{state} + \text{time}), B(\cdot), N(\text{from} * \text{to}), R(\text{from} + T), f(\cdot), \sigma(\cdot)$ | 66 | 29.8 | 16.6 |
| $\phi(\text{from} * \text{state} + \text{time}), B(\cdot), N(\text{to}), R(T), f(\cdot), \sigma(\cdot)$ | 61 | 51.8 | 38.6 |
| $\phi(\text{from} * \text{state}), B(\cdot), N(\text{to}), R(\text{from} + T), f(\cdot), \sigma(\cdot)$ | 49 | 54.7 | 41.6 |
| $\phi(\text{from} * \text{state}), B(\cdot), N(\text{from} * \text{to}), R(T), f(\cdot), \sigma(\cdot)$ | 50 | 70.0 | 56.8 |
| $\phi(\text{from} * \text{state}), B(\cdot), N(\text{to}), R(T), f(\text{site}), \sigma(\cdot)$ | 49 | 79.4 | 66.2 |
| $\phi(\text{from} * \text{state}), B(\text{from}), N(\text{to}), R(T), f(\cdot), \sigma(\cdot)$ | 49 | 80.9 | 67.8 |
| $\phi(\text{from} * \text{state}), B(\text{to}), N(\text{to}), R(T), f(\cdot), \sigma(\cdot)$ | 49 | 81.5 | 68.3 |
| $\phi(\text{from} * \text{state}), B(\cdot), N(\text{to}), R(T), f(\cdot), \sigma(\cdot)$ | 47 | 82.4 | 69.3 |
| $\phi(\text{from} * \text{state}), B(\cdot), N(\text{to}), R(T), f(\cdot), \sigma(\text{site})$ | 49 | 82.5 | 69.3 |
| $\phi(\text{from} * \text{state}), B(\cdot), N(\text{to} + \text{time}), R(T), f(\cdot), \sigma(\cdot)$ | 61 | 84.0 | 70.8 |
| $\phi(\text{from} * \text{state}), B(\text{time}), N(\text{to}), R(T), f(\cdot), \sigma(\cdot)$ | 61 | 85.5 | 72.4 |

Integrated Population Model

| Parameter | Mark-recapture MLE (SE) | Integrated MLE (SE) |
|-----------|-------------------------|---------------------|
| ϕ_1 | 0.61 (0.31) | 0.47 (0.26) |
| ϕ_2 | 1.20 (0.45) | 1.48 (0.24) |
| ϕ_3 | 0.93 (0.28) | 0.60 (0.15) |
| ϕ_4 | 1.14 (0.27) | 1.18 (0.12) |
| ϕ_5 | 1.22 (0.30) | 0.94 (0.16) |
| ϕ_6 | 0.68 (0.21) | 0.91 (0.16) |
| ϕ_7 | 1.37 (0.36) | 0.87 (0.16) |
| ϕ_8 | 0.30 (0.18) | 0.39 (0.12) |
| ϕ_9 | 0.54 (0.20) | 0.58 (0.15) |



References

- ❑ Besbeas et al (2002) *Biometrics*, **58**: 540-547.
- ❑ Besbeas et al (2003) *JRSSC*, **52**: 95-102.
- ❑ Besbeas and Morgan (2007) Submitted
- ❑ Besbeas and Morgan (2007a) Submitted
- ❑ Besbeas et al (2009) in *Modelling Demographic processes in marked populations*.
- ❑ Borysiewicz et al (2009) in *Modelling Demographic processes in marked populations*.
- ❑ Durbin and Koopman (2001) *Time series analysis by state-space methods*
- ❑ Gautier et al (2007) *Ecology*, **88**: 1420-1429
- ❑ Harvey (1989) *Forecasting, structural time series models and the Kalman filter*.
- ❑ King et al (2007) *JRSSC*, **57**: 609-632
- ❑ McCrea et al (2010) *JABES*, **15**: 539-561
- ❑ Tavecchia et al (2009) *American Naturalist*, **173**: 722-733

SECTION 8

BAYESIAN ANALYSIS

Bayesian Methods in Ecology

Collaborators:

Steve Brooks

Ruth King

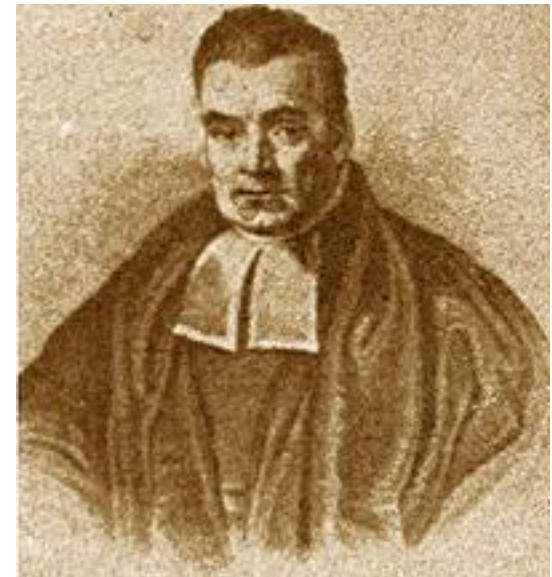
Olivier Gimenez

Bayesian methods in ecology: Outline

- History and Introduction
- Bayes vs Classical
- Selecting a prior
- Bayesian Integration
- Bayesian p-values
- Model selection

History

- Bayesian methods date to the original paper by the Rev. Thomas Bayes which was read to the Royal Society in 1763.
- Known then as *inverse probability* the approach dominated statistical thinking throughout the 19th Century.



What is the Bayesian Approach?

- ❑ The approach is based upon the idea that the experimenter begins with some prior beliefs about the system
- ❑ And then updates these beliefs on the basis of observed data
- ❑ This updating procedure is based upon what is known as Bayes' Theorem:

$$\pi(\theta \mid Data) \propto f(Data \mid \theta) p(\theta)$$

A Difference in Philosophy

- ❑ The Classical approach provides a quite distinct philosophy from the Bayesian one.
- ❑ Bayesians believe that the model parameters have a true but unknown *distribution* which is represented by π .
- ❑ The Classical statistician believes that the model parameters have a true but unknown *value*, which is estimated by the MLE.

Summarising the Bayesian's Beliefs

- ❑ The posterior distribution or the corresponding marginal distributions are the best summaries of the data.
- ❑ However, point estimates and uncertainty intervals are often more interpretable.
- ❑ It is the process of summarising the posterior that is the source of the computational complexity of the Bayesian approach.
- ❑ The classical approach also requires computation, but the maximisation of the likelihood function is usually relatively straightforward.

Point Estimates

- Though there are various ways to obtain point estimates for parameters of interest from the posterior, the most common is to take the posterior mean.
- The posterior mean is obtained via integration:

$$\mathbb{E}_{\pi}(\theta) = \int \theta \pi(\theta \mid \textit{Data}) d\theta.$$

Integration Problems

- There are two problems with this integration:
 1. For realistically complex problems the integration is not analytically tractable and numerical methods must be used.
 2. Bayes' theorem only gives us the posterior distribution up to a constant of proportionality.
[Recall: $\pi(\theta \mid \text{Data}) \propto f(\text{Data} \mid \theta) p(\theta)$]
Thus the normalisation constant must be determined before the mean can be computed

The Classical Interval

- The Classical interval is known as the *confidence interval*.
- Given data \mathbf{Y} , the $100(1-\alpha)\%$ confidence interval $[a(\mathbf{Y}), b(\mathbf{Y})]$ for θ then satisfies the statement:
$$\mathbb{P}([a(\mathbf{Y}), b(\mathbf{Y})] \text{ contains } \theta) = 1-\alpha.$$
- Since θ is assumed fixed, it is the interval (constructed from the data) that is uncertain - it's randomness derives from the stochastic nature of the data under theoretical replications of the data collection process.

The Bayesian Interval

- The Bayesian interval is known as the *credible interval*.
- Given data \mathbf{Y} , the $100(1-\alpha)\%$ credible interval $[a,b]$ for θ then satisfies the statement:
$$\mathbb{P}_{\pi}(\theta \text{ lies in } [a,b]) = 1-\alpha.$$
- Here, it is θ that is the random variable and the interval is fixed - the Bayesian conditions on the data observed rather than thinking about the long-term properties of the estimation procedure under theoretical replications of the data-collection experiment.

[Recall Classical interval: $\mathbb{P}([a(\mathbf{Y}), b(\mathbf{Y})] \text{ contains } \theta) = 1-\alpha.$]

The Benefits of Being Bayesian

□ For Example:

- It's no longer necessary to restrict attention to normal models.
- More complex (integrated, state-space) processes can be analysed with little additional effort.
- Unrealistic assumptions and simplifications can be avoided.
- Random effects and missing values are easily dealt with.
- Model choice and model averaging are relatively easy even with large model spaces.

Prior Elicitation and Specification

- Before we go on to discuss further the computational tools that have fuelled the “Bayesian Revolution” we pause to think a little more deeply about the prior elicitation and specification process.
- There are basically two situations that can be encountered:
 - There is no prior information
 - There is prior information that needs to be expressed in the form of a suitable probability distribution.

Informative Priors

- ❑ These priors aim to reflect information available to the analyst that is gained independently of the data being studied.
- ❑ A prior elicitation process is often used which involves choosing a suitable family of prior distributions for each model parameter and then attempting to find parameters for that prior that accurately reflects the information available.
- ❑ This is often an iterative process.

Prior Elicitation

- ❑ The analyst might begin by getting some idea of the range of plausible model parameter values and using this to get suitable prior parameters.
- ❑ Plots can be made and summary statistics calculated which can then be shown to the expert to see if they are consistent with their beliefs.
- ❑ Any mismatches can be used to alter the prior parameters and the process is repeated.

No Information

- ❑ In the absence of any prior information on one or more model parameters we wish to ensure that this lack of knowledge is properly reflected in the prior.
- ❑ Typically, this is done by choosing a prior distribution with a suitably wide variance.
- ❑ At first, it may seem that taking a flat prior that assigns equal probability to all possible parameter values is a sensible approach.

Flat Priors

- ❑ However, flat priors are rarely invariant to reparameterisations of the model.
- ❑ What may appear flat for parameters under one parameterisation may be far from flat under another.
- ❑ In ecological applications where many parameters of interest are probabilities (e.g., survival, recapture etc.) expressed as logistic regressions e.g.,

$$\text{logit}(\phi_t) = \alpha + \beta t.$$

- ❑ Do you want a prior that is flat on ϕ -space or (α, β) -space?

Flat Priors ctd...

- ❑ Another problem with flat priors is that for continuous variables, these are improper distributions unless bounds on the parameter space are imposed.
- ❑ Improper priors can lead to improper posteriors which, in turn, can mean that the posterior mean simply does not exist.
- ❑ On the other hand, the imposition of bounds which are too restrictive may mean an unrealistically restricted posterior is obtained.

Jeffreys' Prior

- ❑ One solution is to use the Jeffreys' prior.
- ❑ This prior attempts to minimise the amount of influence of the prior on the posterior and is invariant to reparameterisations of the model.
- ❑ It is based upon estimation of Fisher's Information matrix – the same matrix used to estimate Classical SE's.

Vague Priors

- In practice most people choose vague priors which are both proper and have large variance.
- An alternative is to adopt a hierarchical prior structure in which the prior parameters are assumed unknown and themselves given hyper-prior distributions.

For example, rather than saying

$$\alpha \sim N(0, 1000)$$

We might instead say

$$\alpha \sim N(0, \sigma^2), \quad \text{with } \sigma^2 \sim \Gamma(a, b).$$

Vague Priors ctd...

- ❑ This dilutes the influence on the posterior of any prior assumptions made and essentially creates random effects of the model parameters.
- ❑ We will explore the parallel between hierarchical priors and random effects later on.
- ❑ The influence of the prior can always be assessed via a sensitivity analysis and we will also return to this later.

A Simple Example

- ❑ Let's take a simple example to recap some of the main issues discussed so far.
- ❑ 120 deer were radio-tracked over Winter.

| | Number released | Alive | Dead | Other |
|-----------|-----------------|-------|------|-------|
| Treatment | 61 | 19 | 38 | 4 |
| Control | 59 | 21 | 38 | 0 |

Example ctd...

- ❑ So, $n=57$ deer were assigned to the “treatment group” of which $m=19$ survived the Winter,
- ❑ Of interest is the probability of over-Winter survival, call it q , for the general population within the treatment area.
- ❑ The obvious estimate is simply to take the ratio $m/n=19/57$.
- ❑ How would the Classical statistician justify this estimate?

The Model and MLE

- Our model is that we have a simple Binomial experiment (assuming independent and identically distributed draws from the population), so that

$$f(m|q) = \binom{n}{m} q^m (1 - q)^{n-m}.$$

- The classical approach is to maximise the corresponding likelihood with respect to q to obtain the entirely plausible MLE:

$$q = m/n = 19/57.$$

The Bayesian Approach

- The Bayesian starts off with a prior.
- Now, the one thing we know about q is that is a continuous random variable and that it lies between zero and one.
- Thus, a suitable prior distribution might be the Beta which is defined on this range.
- Suppose we assume *a priori* that

$$q \sim \beta(a, b)$$

- So that $p(q) \propto q^{a-1}(1-q)^{b-1}.$

The Posterior

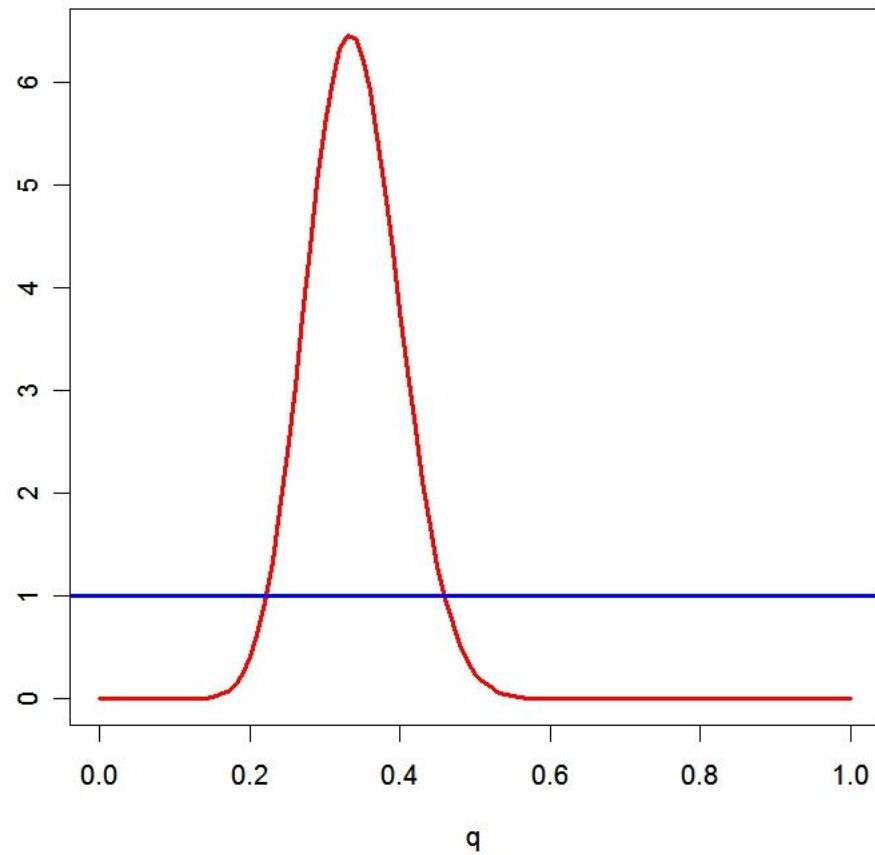
□ Then, we have that

$$\begin{aligned}\pi(q|m) &\propto \binom{n}{m} q^m (1-q)^{n-m} \times B(a, b) q^{a-1} (1-q)^{b-1} \\ &\propto q^{(a+m)-1} (1-q)^{(b+n-m)-1}\end{aligned}$$

i.e., $q \sim \beta(a+m, b+n-m)$ *a posteriori*.

□ If we take a Uniform prior i.e., $a=b=1$, then we get the following prior and posterior distributions.

Posterior Plot



Posterior Summaries

- Similarly, under the Uniform prior we get:
 - Posterior mean = 0.338
 - Posterior mode = 0.333
 - Posterior standard deviation = 0.061
-
- Recall that the MLE is 0.333 with a standard error of 0.062.

The Posterior ctd...

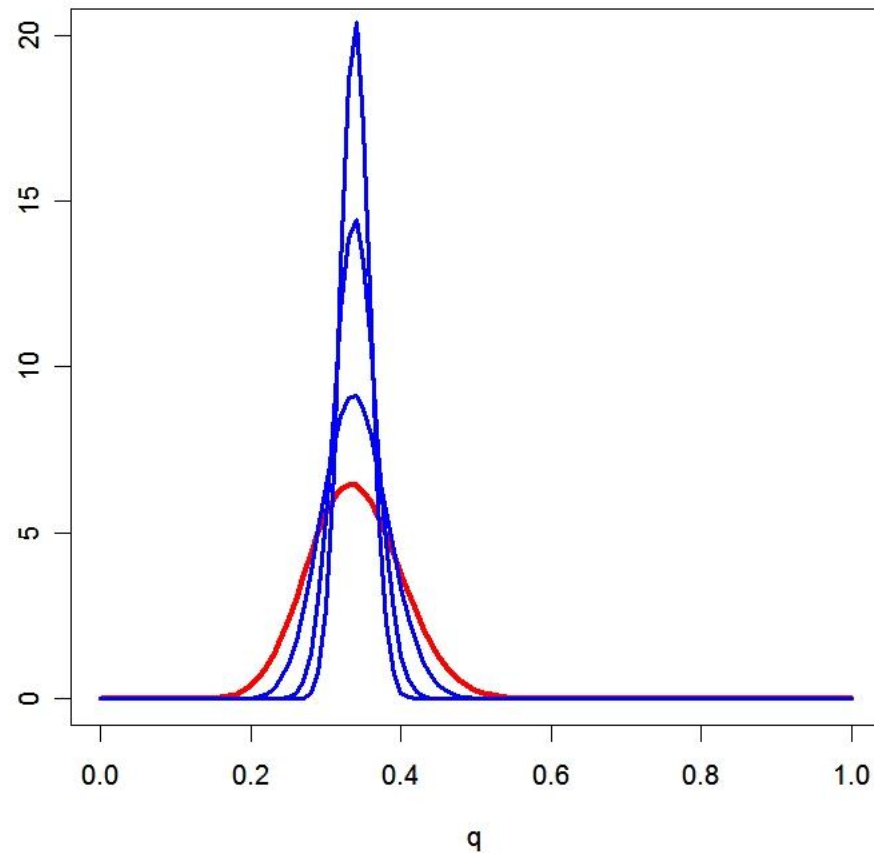
- For a general prior, the posterior has mean

$$\frac{a + m}{a + b + n} = (1 - w) \left(\frac{a}{a + b} \right) + w \frac{m}{n}$$

where $w = n/(a+b+n)$.

- Thus, the posterior mean is a weighted average of the prior mean and the MLE.
- Note that as n increases, the weight w tends to 1 and so the posterior mean converges to the MLE.

More Posterior Plots



The Posterior ctd...

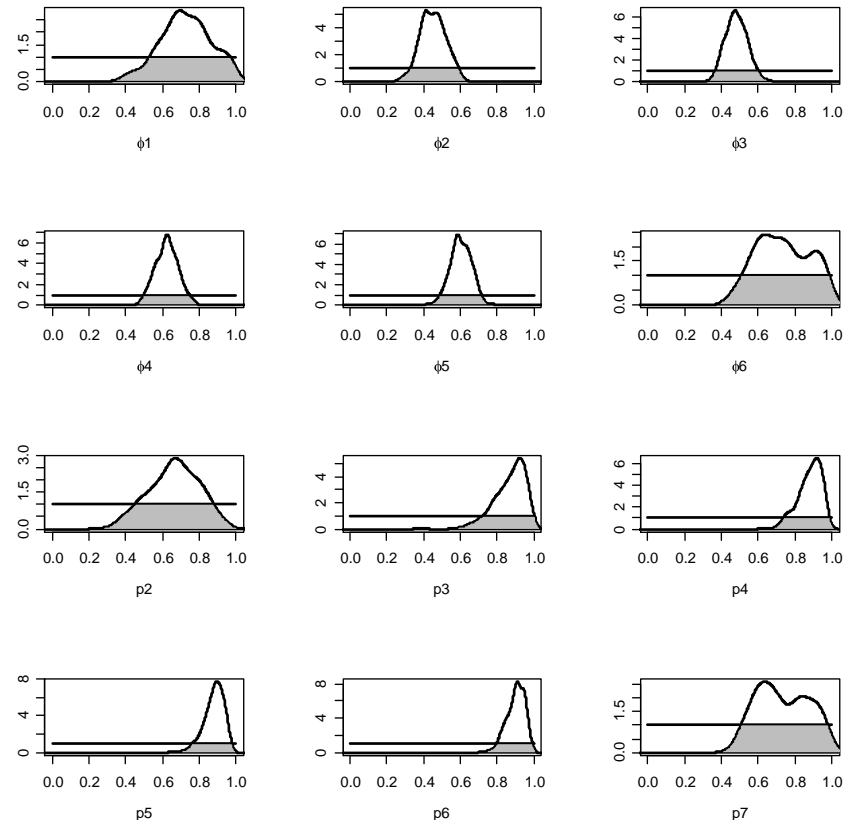
- ❑ The posterior mean converges to the MLE irrespective of the prior parameters a and b (though the speed at which convergence occurs will vary with the prior).
- ❑ Similarly, the posterior variance converges to zero as the sample size increases.
- ❑ Thus, the posterior itself converges to a point mass at the MLE.

A General Result

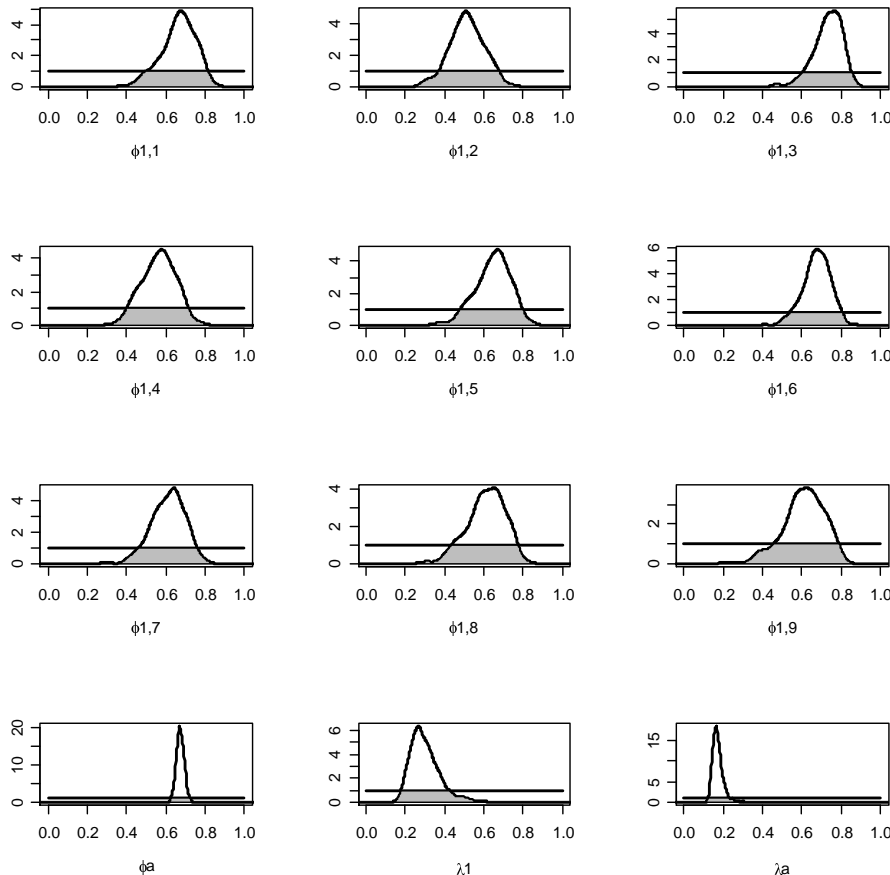
This is a general result, the Bayesian and Classical estimates will always agree if there is sufficient data, so long as the MLE is not explicitly ruled out by the prior.

Which parameters can we identify?

- In population ecology we may devise models with parameters that cannot be estimated from the data.
- Symbolic algebra can be used to examine whether a model is parameter-redundant.
- In a Bayesian context, it is interesting to consider the overlap between priors, $p(\theta)$ and posteriors $\pi(\theta|x)$.



Male mallard, *Anas platyrhynchos*



Model: $\phi_{1,i}$, ϕ_a , λ_1 , λ_a here only two parameters, ϕ_a and λ_a are strongly identified.

Bayesian Integration Made Easy

- Recall that Bayesian inference is based upon the estimation of posterior summaries such as the mean.
- And that these require integration of the posterior density e.g.

$$\mathbb{E}_{\pi}(\theta) = \int \theta \pi(\theta \mid \text{Data}) d\theta.$$

- Expectations of this sort can be estimated by drawing samples $\theta^1, \dots, \theta^T$ from the posterior distribution and then calculating the sample mean.
- This is known as *Monte Carlo Integration*.

From Integration to Sampling

- ❑ Thus, we replace the integration problem by a sampling problem.
- ❑ Two problems:
 - In general, $\pi(\theta|Data)$ represents a high-dimensional and complex distribution from which samples would usually be difficult to obtain.
 - In addition, large sample sizes are often required and so powerful computers are needed to generate these samples.

The 1990's

- At the beginning of the 1990's two things happened:
 - The computer revolution – which made powerful computers an affordable reality.
 - MCMC was introduced to the statistical literature – which provided a framework for generating samples from arbitrarily complex distributions.

Markov Chain Monte Carlo

- MCMC methods perform Monte Carlo Integration using a Markov chain to generate observations from π .
- A Markov chain is simply a stochastic sequence of numbers where each value in the sequence depends only upon the last.
- We might label the sequence $\theta^0, \theta^1, \theta^2, \dots$ etc.

MCMC ctd...

- The value for θ^0 is chosen from some arbitrary starting distribution.
- At time t , we *update* the state of the chain from θ^t to θ^{t+1} .
- This updating is performed in such a way that the probability distribution associated with the t^{th} observation gets closer and closer to $\pi(\theta \mid \text{Data})$ as t increases.
- We say that the distribution of the chain *converges* to π .
- We return to the concept of convergence later on.

MCMC Magic

- ❑ MCMC therefore allows us to construct a sequence of values whose distribution converges to the posterior distribution of interest.
- ❑ We can then use this sequence of values to obtain empirical (Monte Carlo) estimates of any posterior summaries of interest e.g., posterior means.
- ❑ The beauty of MCMC is that the updating procedure remains relatively simple no matter how complex the posterior distribution of interest.

Updating our Chain

- ❑ OK, so we can do the integration by sampling from the posterior.
- ❑ We can sample from the posterior by generating a Markov chain.
- ❑ We construct a Markov chain using an appropriate updating scheme.
- ❑ But how do we do these updates?
- ❑ There are several standard approaches...

Metropolis Hastings

- Most MCMC methods are special cases of the general Metropolis Hastings scheme.
- The MH scheme generates θ^{t+1} from θ^t in a two-step procedure.

- Step 1. At iteration t , sample $\phi \sim q(\phi \mid \theta^t)$.

ϕ is known as the *candidate point*

q is known as the *proposal distribution*

MH ctd...

- Step 2. With probability

$$\alpha(\theta^t, \phi) = \min \left(1, \frac{\pi(\phi|Data)q(\theta^t|\phi)}{\pi(\theta^t|Data)q(\phi|\theta^t)} \right)$$

set

$$\theta^{t+1} = \phi \quad (\textit{acceptance})$$

else, set

$$\theta^{t+1} = \theta^t \quad (\textit{rejection})$$

- Steps 1 and 2 are then repeated to obtain first θ^1 and then θ^2 etc.

“Flavours” of MH

- Within the MH scheme there are a number of special cases depending upon the (essentially arbitrary) choice of proposal distribution.
- One basic and common choice is to base the proposal around the current point e.g., $\phi \sim N(\theta^t, \sigma^2)$.
- In this case, the acceptance ratio simplifies so that

$$\alpha(\theta^t, \phi) = \min [1, \pi(\phi)/\pi(\theta^t)]$$

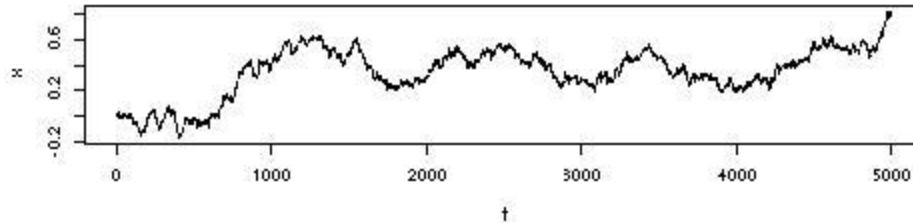
Tuning the RW Metropolis Algorithm

- ❑ This is known as the random walk Metropolis algorithm.
- ❑ In order to balance the size of the proposed moves with the chance of accepting them the proposal variance is often *tuned* to obtain an acceptance rate of 20-40%.
- ❑ Most MH algorithms require some initial *pilot tuning* in order to get satisfactory performance.

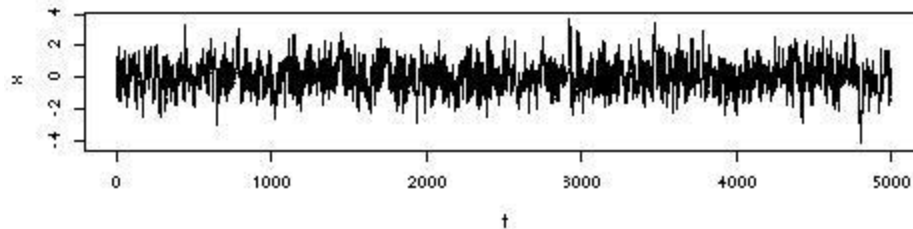
Why is the Size of the Move Important?

- ❑ If you propose extremely small jumps, you are bound to accept them, but you will take a long time to move around the state space.
- ❑ Small jumps induce correlations (known as *autocorrelations*) between successive observations.
- ❑ If you have strongly correlated observations, you require far bigger sample sizes and therefore longer MCMC runs.

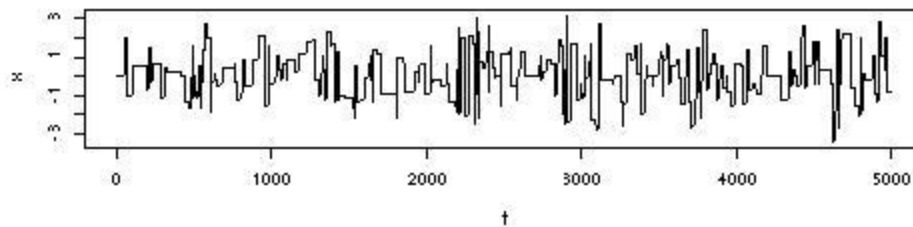
Good Trace/Bad Trace



Variance too small



Just right



Too large

Single-Update MH

- ❑ Typically, the posterior will be high-dimensional.
- ❑ In this case, we can either update all parameters simultaneously.
- ❑ Or, we can update each parameter in turn. This is known as *single-update Metropolis Hastings*.
- ❑ Sometimes it is convenient to do something in between and update parameters in *blocks*.

Gibbs Updates

- A special case of the single-update MH algorithm arises if we set the proposal distribution for any parameter to be the conditional posterior distribution of that parameter given current value of the others.
- In this case, the acceptance probability is always exactly 1.
- This is known as a *Gibbs-sampler* (or *simply Gibbs*) *update*.

Gibbs Updates ctd...

- ❑ One useful feature of the Gibbs update (beyond no longer having to perform the accept-reject step) is that it requires no pilot-tuning.
- ❑ Also the proposal is adaptive in that it depends upon the state of the other parameters and so changes from iteration to iteration.
- ❑ Gibbs updates tend to be highly efficient MCMC-wise but can be computationally more burdensome.

Hybrid Schemes

- ❑ In practice we often use a mixture of MH and Gibbs updates.
- ❑ For parameters with standard conditional distributions, we use Gibbs updates.
- ❑ And, for the rest, we might use random walk updates.
- ❑ Since Gibbs updates are so good, we would like to use them as often as possible...

Conjugacy

- ❑ When selecting our prior, it is often useful to select what is known as a *conjugate prior* family.
- ❑ For a particular parameter, the associated prior is conjugate if the corresponding posterior conditional is of the same form.
- ❑ E.g., a Normal prior leading to a Normal posterior conditional.
- ❑ Conjugate priors are often adopted so that Gibbs updates can be used.

Posterior Summaries

- Posterior means and variances etc. are easily obtained.
- Intervals usually come in the form of Highest Posterior Density Intervals which are credible intervals with the minimum width.
- E.g., for a $(1-\alpha)\%$ HPDI, order the observations and take $[\theta_{(i)}, \theta_{(j)}]$ where $i+j = \alpha T$ and T denotes the sample size.

Density Estimates

- Marginal density estimates are also often of interest.
- Basic histograms or kernel smoothed empirical estimates can be used.
- Rao-Blackwell estimates are also useful.

□ Here,

$$f(\theta_i) = \frac{1}{T} \sum_{t=1}^T \pi(\theta_i | \theta_{(i)}^t, Y)$$

- If the posterior conditional is standard, then the density estimate, f , is easily normalised.

Run Lengths

- There are two elements to be considered when determining the simulation length:
 - The time required for convergence.
 - The post-convergence sample size required for suitably small Monte Carlo errors.
- We begin with convergence assessment...

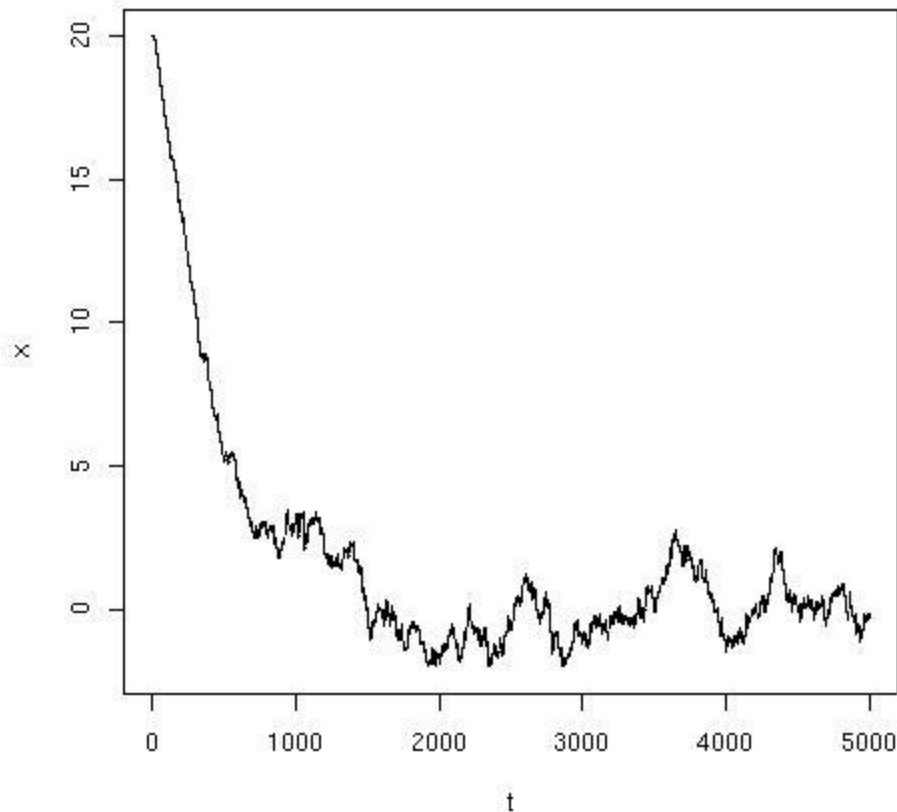
Convergence Assessment

- ❑ Here, we're looking to determine how long it takes for the Markov chain to converge to the target distribution.
- ❑ In practice, we discard observations from the start of the chain and just use observations from the chain once it has converged.
- ❑ The initial observations that we discard are referred to as the *burn-in*.

Trace Plots

- ❑ The simplest method to determine the length of the burn-in period is to look at trace plots.
- ❑ You can often see the individual parameters converging from their starting position to values based around a constant mean.
- ❑ This is a fairly efficient method, but is not robust.

Trace Plot Example



Converged after, say,
2000 iterations?

Multiple Replications

- ❑ Running several replications from different starting points provides additional reassurance.
- ❑ Basically, if you run the chain several times from different starting points and they all give you the same posterior estimates...
- ❑ Then, this suggests that no major modes have been missed in any one simulation and that each has probably converged.

Convergence Assessment

- ❑ This approach is formalised in the so-called BGR diagnostic. Available in WinBUGS.
- ❑ There are various implementations, all based upon the idea of using an analysis of variance to determine whether or not there are any differences in estimates from different replications.
- ❑ The simplest implementation is to compare the width of the empirical 80%CI obtained from all chains combined, with the corresponding mean within-chain interval width.
- ❑ Convergence is assumed when these are roughly equal, implying that all chains have roughly equal variability i.e., similar amounts of information.

Prior Sensitivity

- ❑ As discussed earlier it is important to assess the sensitivity of key results to the prior assumptions.
- ❑ In the presence of reasonably informative data, there should be little prior sensitivity.
- ❑ Prior sensitivity in itself is not a problem, but...
- ❑ Extreme prior sensitivity often points to problems such as parameter redundancy or overly restrictive prior assumptions.

Bayesian P-values

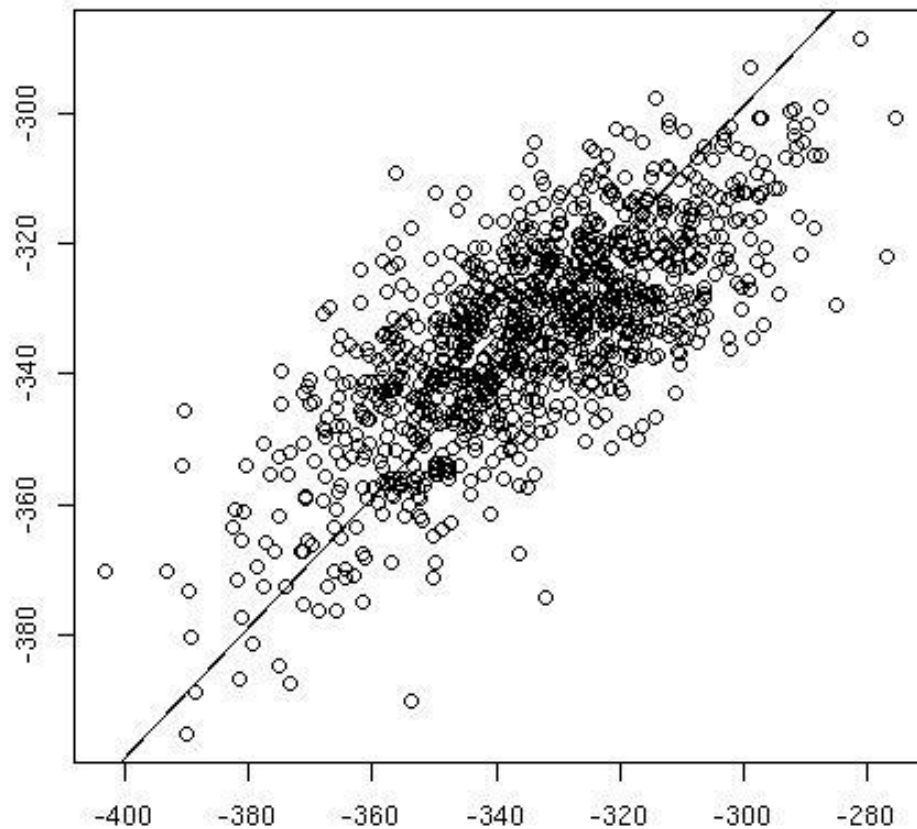
- Bayesian p-values involve generating a new data set \mathbf{Y}_t from the model at each MCMC iteration.
- A discrepancy statistic $D(\mathbf{Y}_t, \mathbf{e}_t)$ is then used to measure the distance between the generated data and the corresponding expected values, \mathbf{e}_t .
- The proportion of times that

$$D(\mathbf{Y}, \mathbf{e}_t) > D(\mathbf{Y}_t, \mathbf{e}_t)$$

is recorded and should be close to 0.5.

- Plots of $D(\mathbf{Y}, \mathbf{e}_t)$ vs $D(\mathbf{Y}_t, \mathbf{e}_t)$ may also be useful.

Bayesian P-values Plot



P-value here
is 0.513

Model Selection

- ❑ In the classical framework, likelihood ratio tests or information criterion (e.g. AIC) are often used.
- ❑ There is a “similar” Bayesian statistic – the DIC.
- ❑ This is programmed within WinBUGS – however its implementation is not suitable for hierarchical models (e.g. random effect models).
- ❑ In addition, the DIC is known to give falacious results in even simple problems.
- ❑ Within the general Bayesian framework, there is a more natural way of dealing with the issue of model discrimination.

Bayesian Approach

- We treat the model itself to be a unknown parameter to be estimated.
- Then, applying Bayes' Theorem we obtain the posterior distribution over both parameter and model space:

$$\pi(\theta_m, m \mid \text{data}) \propto L(\text{data} \mid \theta_m, m) p(\theta_m) p(m).$$

- Here θ_m denotes the parameters in model m .

Likelihood

Prior on parameters
in model m

Prior on model m

Posterior Model Probabilities

- The Bayesian approach then allows us to quantitatively discriminate between competing models via posterior model probabilities:

$$\begin{aligned}\pi(m \mid \text{data}) &= \int \pi(\theta_m, m \mid \text{data}) d\theta_m \\ &\propto p(m) \int L(\text{data} \mid \theta_m, m) p(\theta_m) d\theta_m\end{aligned}$$

- Note that we need to specify priors on both the parameters and now also on the models themselves.
- Thus we need to specify a prior probability for each model to be considered.

Bayes Factors

- These are often used to compare models and are simply defined to be the ratio of the posterior odds to prior odds. For two models, m_1 and m_2 , the Bayes factor of model 1 to model 2 is given by:

$$BF_{12} = \frac{\pi(m_1 | \text{data}) / p(m_1)}{\pi(m_2 | \text{data}) / p(m_2)}$$

- Kass and Raftery (1995) give the following “rule of thumb” guide:

| Bayes factor | Interpretation |
|--------------|----------------------|
| < 3 | Not worth mentioning |
| 3-20 | Positive evidence |
| 20-150 | Strong evidence |
| > 150 | Very strong evidence |

Model Averaging

- Model averaging obtains an estimate of a parameter, based on all plausible models, and takes into account both model and parameter uncertainty.
- For parameter θ , common to models m_1, \dots, m_k , the model averaged estimate is:

$$\begin{aligned}\pi(\theta \mid \text{data}) &= \sum_{i=1}^k \pi(\theta, m_i \mid \text{data}) \\ &= \sum_{i=1}^k \pi(\theta \mid \text{data}, m_i) \pi(m_i \mid \text{data}).\end{aligned}$$

- Thus each estimate of θ is weighted according to the corresponding posterior probability.

Monte Carlo estimates

- The posterior probability of model m is given by:

$$\pi(m \mid \text{data}) \propto p(m) \int L(\text{data} \mid \theta_m, m) p(\theta_m) d\theta_m$$

- However $E_p[f(x)] = \int f(x) p(x) dx$.

- So that,

$$\pi(m \mid \text{data}) \propto p(m) E_p[L(\text{data} \mid \theta_m, m)].$$

- Thus, we can estimate this expression by using a *Monte Carlo* estimate.

Monte Carlo estimates ctd.

- The simplest approach is to draw observations $\theta_1, \dots, \theta_N$ from the prior distribution $p(\theta)$, and estimate,

$$\begin{aligned} \hat{L}(\text{data} \mid m) &= N^{-1} \sum_{i=1}^N L(\text{data} \mid \theta_i, m) \\ &\rightarrow E_p[L(\text{data} \mid \theta_m, m)] \text{ as } N \rightarrow \infty \end{aligned}$$

- Thus, this estimate converges to the posterior model probability (up to proportionality).
- Posterior model probabilities can be estimated by calculating this for each model and renormalising (so that the probabilities sum to one).

Further Monte Carlo estimates

- However, this approach is often very inefficient, and can be very unstable.
- Alternative approaches essentially involve changing the distribution from which the θ are drawn (and hence the expression for the Monte Carlo estimate).
- For example, an alternative estimate would be to sample θ from the posterior distribution and use,

$$L(\text{data} \mid m) = [N^{-1} \sum_{i=1}^N L(\text{data} \mid \theta_i, m)^{-1}]^{-1}$$

General Comments

- ❑ The Monte Carlo estimates are the easiest to program and conceptualise.
- ❑ However, they are often very inefficient and do not always converge within a feasible number of iterations.
- ❑ In addition, each individual model needs to be considered in turn, which may be infeasible when there are a large number of models.
- ❑ We consider in more detail alternatives which use the same underlying principle as MCMC.

MCMC-based estimates

- We have a posterior distribution (over parameter and model space) defined up to proportionality:
$$\pi(\theta_m, m \mid \text{data}) \propto L(\text{data} \mid \theta_m, m) p(\theta_m \mid m) p(m)$$
- If we can sample from this posterior distribution then we are able to obtain posterior estimates of summary statistics of interest.
- In particular the posterior model probabilities can be estimated as the proportion of time that the chain is in each model.
- So, all we need to do is define how we construct such a Markov chain.

Reversible Jump MCMC

- The reversible jump MCMC algorithm allows us to construct a Markov chain with stationary distribution equal to the posterior distribution.
- It is simply an extension of the Metropolis-Hastings algorithm that allows moves between different dimensions.
- This algorithm is needed because the number of parameters, θ_m , in model m , may differ between models.
- Note that this algorithm needs only one Markov chain irrespective of the number of models!!

Markov chain

- Each iteration of the Markov chain essentially involves two steps:
 1. **Within model moves** – updating each parameter using standard MCMC moves (Gibbs sampler, Metropolis-Hastings)
 2. **Between model moves** – updating the model using a reversible jump type move.
- Then, standard MCMC-type processes apply, such as using an appropriate burn-in, obtaining summary statistics of interest etc.

Within model moves: MH algorithm

- Recall, the MH algorithm involves two steps:
 1. At each iteration k , with current parameter values θ_k , sample a candidate point ϕ from some proposal distribution q , i.e. $\phi \sim q(\phi \mid \theta_k)$.
 2. Accept the proposed value with probability, $\min(1, A)$, where,

$$A = \frac{\pi(\phi \mid \text{data}) q(\theta_k \mid \phi)}{\pi(\theta_k \mid \text{data}) q(\phi \mid \theta_k)}$$

and set $\theta_{k+1} = \phi$.

Else, if the move is rejected, set $\theta_{k+1} = \theta_k$.

Between model moves: RJ step

- ❑ This is simply a generalisation of the MH algorithm, which allows moves between different models.
- ❑ A single move involves two steps:
 1. Proposing to move to a different model;
 2. Accepting this proposed model with some probability.
- ❑ To illustrate the RJ algorithm, we consider a particular example relating to variable selection (e.g. covariate analysis).

Example: variable selection

- Suppose that the survival rate ϕ is logistically regressed on possible covariates x and y .
- However, it is unknown *a priori* which covariates (if any) affect the survival rates.
- There are two possible models:
 1. $\text{logit } \phi = \alpha + \beta x;$
 2. $\text{logit } \phi = \alpha + \beta x + \gamma y$where α and β (and γ for model 2) are parameters to be estimated.

RJ update

- Suppose that the Markov chain is currently in model 1 (logit $\phi = \alpha + \beta x$).
- **Step 1:** Proposing new model.
 - Set $\theta = (\alpha, \beta)$.
 - Suppose that at iteration k of the Markov chain we are in state $(\theta, m)_k$.
 - We propose to move to state (θ', m') , where
$$\theta' = (\alpha, \beta, \gamma).$$
 - Then, we simulate parameter value γ from some *proposal* distribution q .

RJ Update (cont)

□ Step 2: Accept/reject step:

- Accept the new model m' , and set $(\theta, m)_{k+1} = (\theta', m')$ with probability $\min(1, A)$, where,

$$A = \frac{\pi(\theta', m' | \text{data}) P(m | m')}{\pi(\theta, m | \text{data}) P(m' | m) q(\gamma)},$$

where $P(m | m')$ denotes the probability of proposing to move from model m to model m' .

- Else reject the move and set $(\theta, m)_{k+1} = (\theta, m)_k$.
- Then, in the reverse move, where we propose to move from (θ', m') to state (θ, m) , we remove parameter γ (or equivalently set it equal to zero).
- This move is then accepted with probability, $\min(1, A^{-1})$.

****General RJ updates****

- ❑ The previous RJ step can only be applied to nested models – more generally, moves are possible between any two models:
- ❑ **Step 1:** Proposing new model.
 - Suppose that at iteration k the Markov chain is in model m with parameters $(\theta, m)_k$.
 - We propose to move to state (θ', m') .
 - Then, we define a (bijective) function g , such that,

$$(\theta', \mathbf{u}') = g(\theta, \mathbf{u})$$

where \mathbf{u} and \mathbf{u}' are sets of random variables with density function $q(\mathbf{u})$ and $q'(\mathbf{u}')$, respectively.

General RJ updates (cont)

□ Step 2: Accept/reject step:

- Accept the new model m' , and set $(\theta, m)_{k+1} = (\theta', m')$ with probability $\min(1, A)$, where,

$$A = \frac{\pi(\theta', m' | \text{data}) P(m | m') q'(\mathbf{u}')}{\pi(\theta, m | \text{data}) P(m' | m) q(\mathbf{u})} \left| \frac{\delta(\theta', \mathbf{u}')}{\delta(\theta, \mathbf{u})} \right|$$

where $P(m | m')$ denotes the probability of proposing to move from model m to model m' .

- Else reject the move and set $(\theta, m)_{k+1} = (\theta, m)_k$.
- This final term is called the Jacobian.
- This formula is not as complicated as it appears!!

Case Study: Soay sheep

- Heterogeneity within a population can often be explained via different covariates.
- For illustration we shall consider in detail MRR (mark-recapture-recovery) data relating to Soay sheep.



Covariate values

Capture history

| ID | Capture History | | | | | | | | | | | | Sex | YOB | Coat | Horn | BW |
|-------|-----------------|---|---|---|---|---|---|---|---|---|---|---|-----|-----|------|------|-------|
| ny068 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 85 | 1 | 3 | 0.51 |
| ny168 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 85 | 1 | 2 | 0.69 |
| op110 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 85 | 1 | 3 | 0 |
| ny207 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 0 | 1 | 85 | 1 | 3 | 1.2 |
| ny211 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 85 | 1 | 2 | 2.57 |
| ny212 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 85 | 2 | 1 | 0 |
| op176 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 85 | 1 | 3 | 0 |
| np055 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 86 | 1 | 0 | -1.00 |
| np027 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 86 | 0 | 0 | -0.78 |
| np028 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 86 | 2 | 0 | -0.73 |
| np035 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 86 | 1 | 2 | -0.60 |

Missing values

Soay sheep

- ❑ We place vague priors on the parameters present in each model.
- ❑ Priors also need to be specified on the models.
- ❑ Placing an equal prior on each model places a high prior mass on models with a large number of age groups, since the number of models increases with the number of age groups.
- ❑ Thus, we specify we an equal prior probability on each marginal age structure and a flat prior over the covariate dependence given the age structure

Results – Soay sheep

- We use the RJ algorithm to explore and summarise the posterior distribution.
- The posterior distribution is very complex, and so we present results from the marginal distributions for the survival rates only.
- We initially consider the age-dependence structure before the covariate dependence.

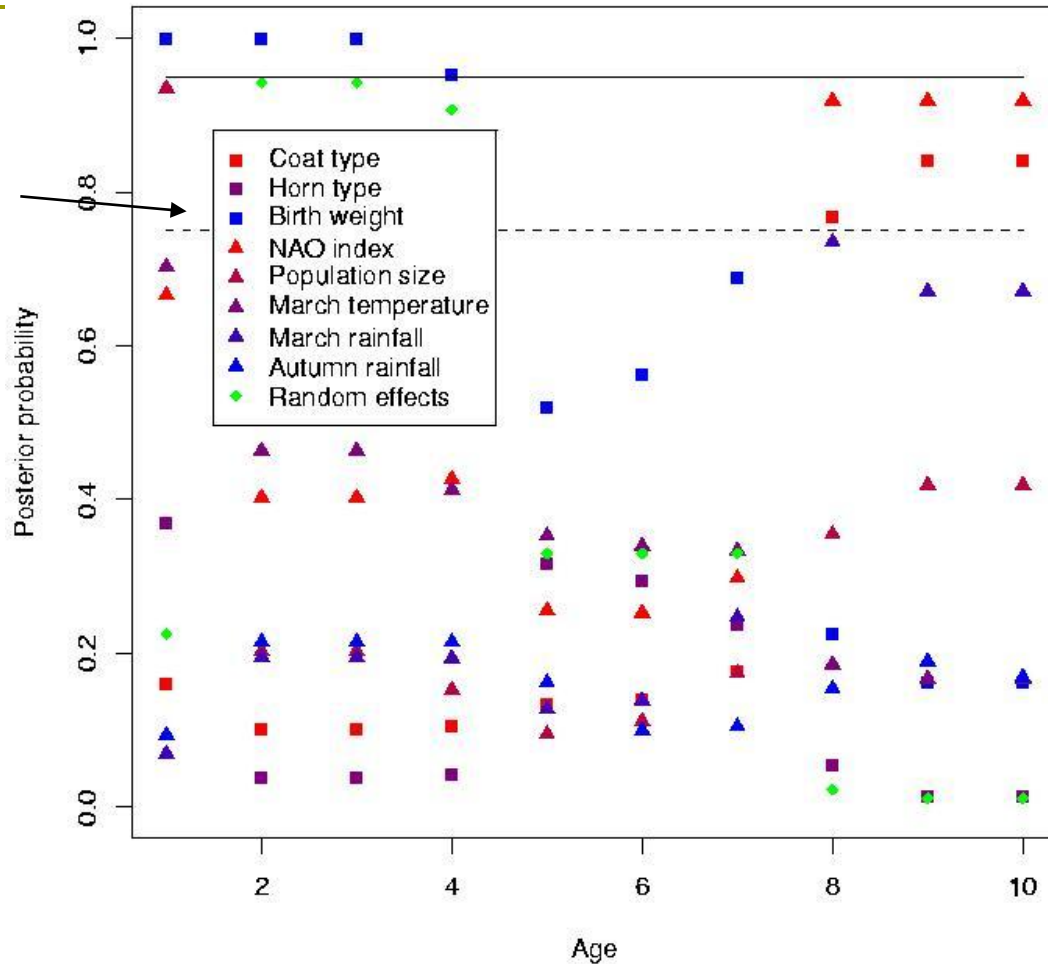
Age-dependence (females)

| Age-structure | Posterior probability |
|---|-----------------------|
| $\phi_1; \phi_{2:4}; \phi_{5:7}; \phi_{8+}$ | 0.384 |
| $\phi_1; \phi_{2:7}; \phi_{8+}$ | 0.286 |
| $\phi_1; \phi_{2:4}; \phi_{5:7}; \phi_7; \phi_{8+}$ | 0.093 |
| $\phi_1; \phi_{2:3}; \phi_{4:7}; \phi_{8+}$ | 0.075 |

- Note that with probability 1, lambs have a distinct survival rate.
- Often, the models with most posterior support are close neighbours of each other.

Covariate Dependence

BF = 20

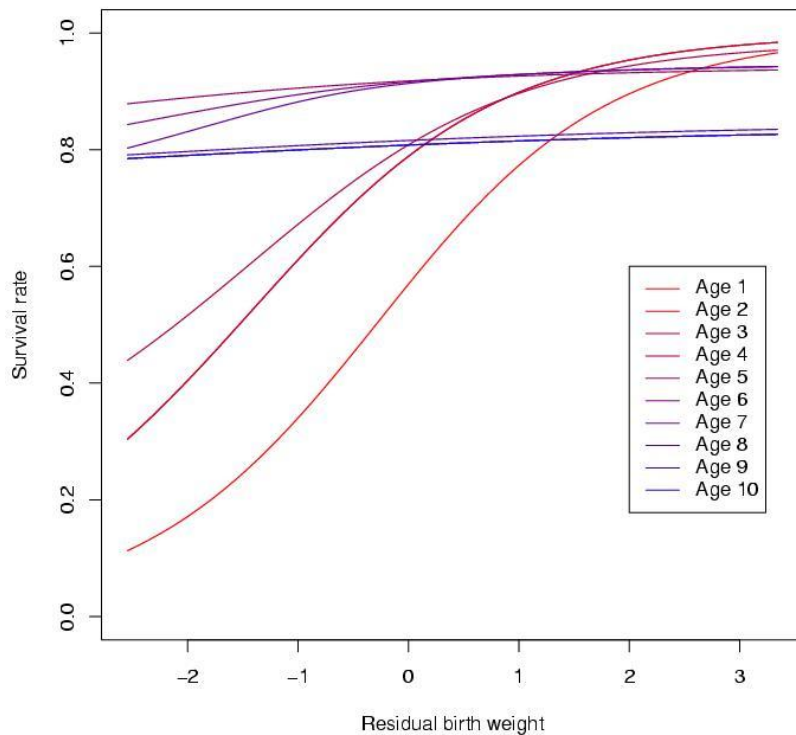


BF = 3

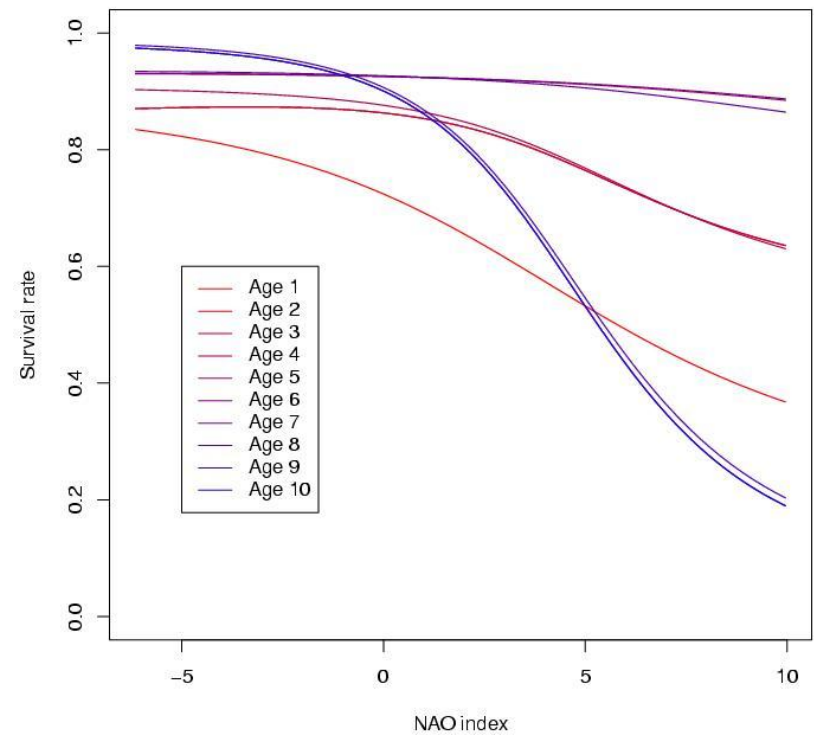
BF = Bayes factor

Influence of parameters

Birth weight



NAO Index



Age 1

Results – Shags (MRR data)

| Survival rate | Recapture rates | Recovery rate | Posterior probability |
|--|----------------------------|---------------------------------------|-----------------------|
| $\phi_1(t), \phi_{2,3}, \phi_A(t)$ | $p_1, p_2, p_3(t), p_A(t)$ | $\lambda_1, \lambda_{2,3}, \lambda_A$ | 0.309 |
| $\phi_1(t), \phi_{2,3}, \phi_A$ | $p_1, p_2, p_3(t), p_A(t)$ | $\lambda_1, \lambda_{2,3}, \lambda_A$ | 0.204 |
| $\phi_1(t), \phi_2, \phi_{3,A}$ | $p_1, p_2, p_3(t), p_A(t)$ | $\lambda_{1,3}, \lambda_2, \lambda_A$ | 0.126 |
| $\phi_1(t), \phi_2, \phi_3, \phi_A(t)$ | $p_1, p_2, p_3(t), p_A(t)$ | $\lambda_1, \lambda_{2,3}, \lambda_A$ | 0.047 |

Some variability but
neighbouring models

This model for the recapture rates has
posterior probability 0.996.

General Comments

- ❑ RJMCMC is the most widely used algorithm when there is model uncertainty.
- ❑ The acceptance probabilities for RJ moves are typically lower than for MH updates.
- ❑ Longer simulations are generally needed to explore both parameter and model space.
- ❑ Only a single Markov chain is necessary, irrespective of the number of possible models!!
- ❑ However, RJMCMC cannot (currently) be generally programmed in WinBUGS and specific codes generally need to be written.

The benefits of a Bayesian approach

- ❑ Random effects are easily incorporated
- ❑ Missing values are easily dealt with
- ❑ Hierarchical models are easy to construct
- ❑ Model averaging is easily done...
- ❑ State-space models can be constructed without the need for normal approximations
- ❑ Easy application to longitudinal models
- ❑ Easy application of complex spline regressions.

Are there dangers in Bayesian methods?

“The main danger is that they will be used automatically. In the application of Bayesian methods, you first need to assign some basic probabilities (and perhaps utilities) to give the problem structure; then the computer can derive other probabilities. One must *think* about the basic values and it is not usually satisfactory to use a normal density and non-informative priors. This distinction between the thinking you and the un-thinking, calculating personal computer is essential.” Dennis Lindley, *Significance*, 2004.

Possible reading

- ❑ Bayesian Model Averaging: a tutorial. (1999), Hoeting, J.A. et al. *Statistical Science*, **14**, 382-417.
- ❑ Model choice: A minimum posterior predictive loss approach. (1998) Gelfand, A.E. and Ghosh, S.K. *Biometrika*, **85**, 1-12.
- ❑ Bayesian model choice via Markov chain Monte Carlo methods. (1995) Carlin, B.P. and Chib, S. *JRSSB*, **57**, 473-484.
- ❑ Sensitivity of Bayes estimators to hyper-parameters with an application to maximum yield from fisheries. (2004) Millar, R.B. *Biometrics*, **60**, 536-542.

The end

