Workshop on Parameter Redundancy Part I

Byron Morgan
University of Kent

St. Andrews, June, 2009

ncse
National Centre for Statistical Ecology

Collaborators

- Ted Catchpole (Canberra and Kent)
- Diana Cole (Kent)

National Centre for Statistical Ecology

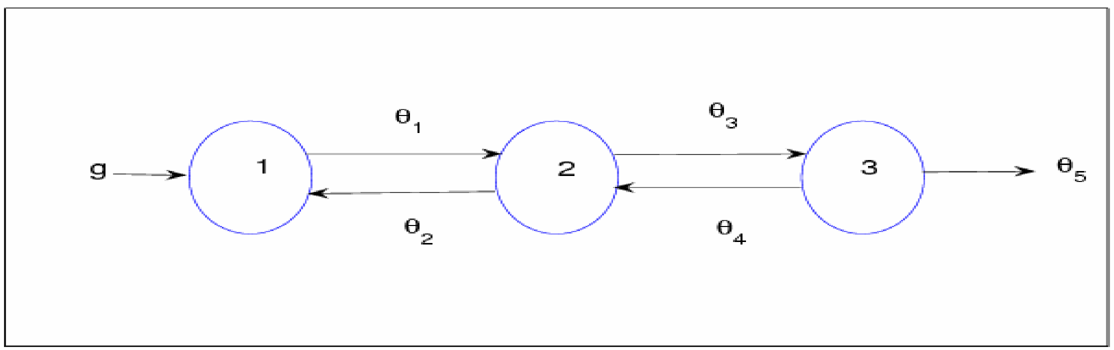
Outline

- Introduction and motivation
- Definitions
- General rules
- Use of symbolic algebra
- Extension theorems
- Near redundancy
- Weak identifiability

Complex models and their parameters

- Compartment models
- Ecology
- Econometrics
- Hidden Markov models

Compartment models



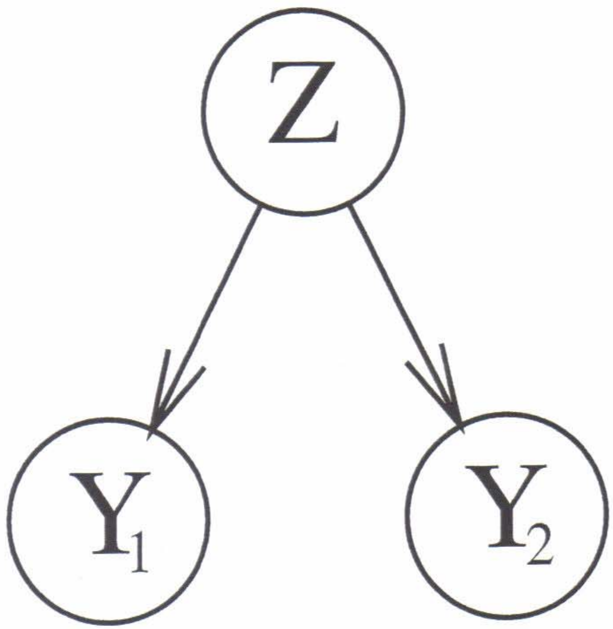
Econometrics

- 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 naive Bayesian network with a binary root node and two binary observable nodes.

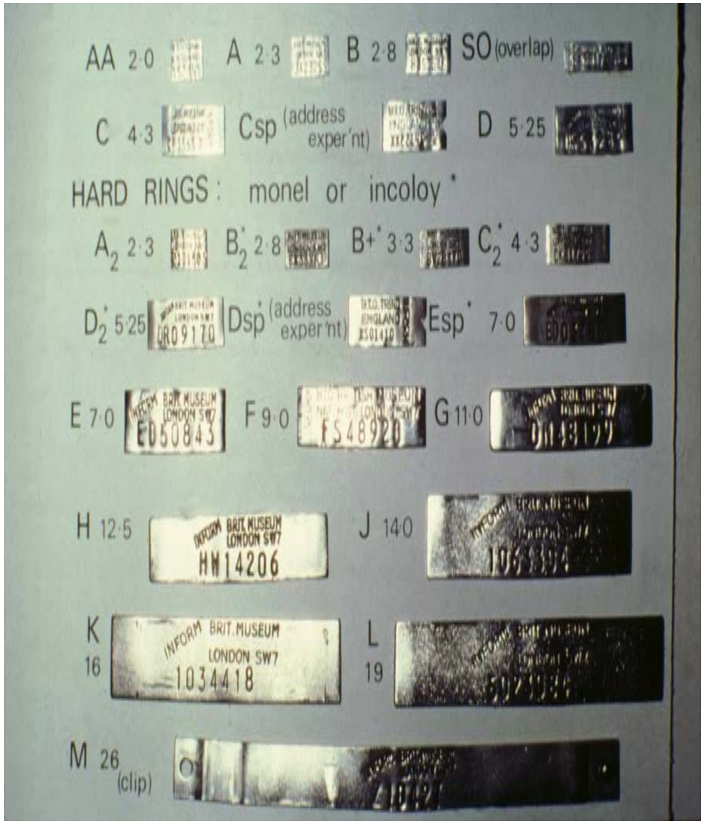
Ecology

- 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.

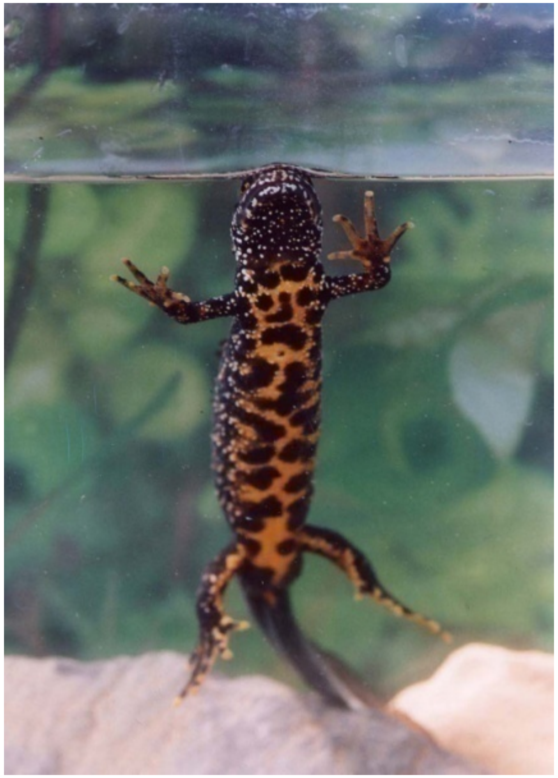
ncse
National Centre for Statistical Ecology

Models for survival: 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

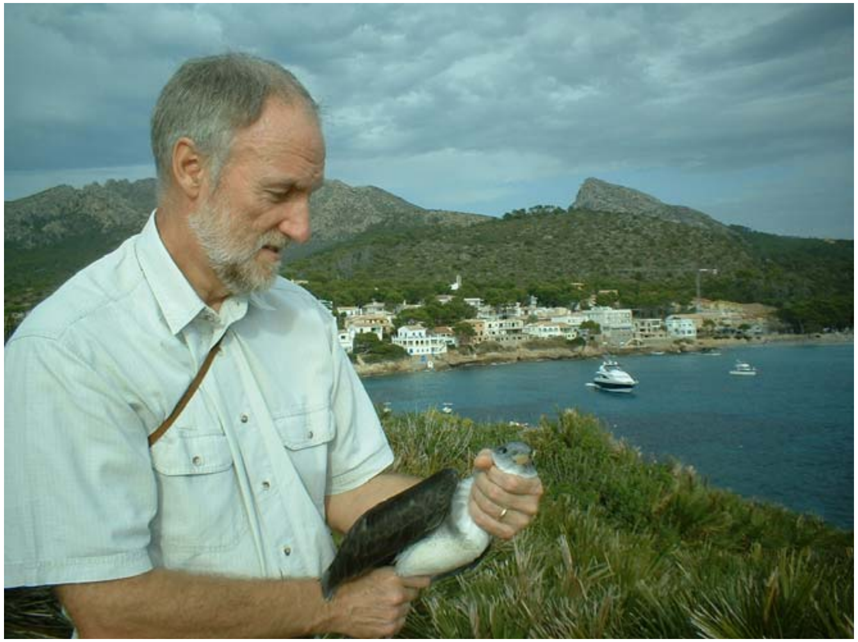


Identification of Cormorant, *Phalacrocorax carbo sinensis*, and great crested newt, *Titurus cristatus*



National Centre for Statistical Ecology

Recapture of Cory's shearwater, *Calonectris diomedea*



ncse
National Centre for Statistical Ecology

Recovery/recapture

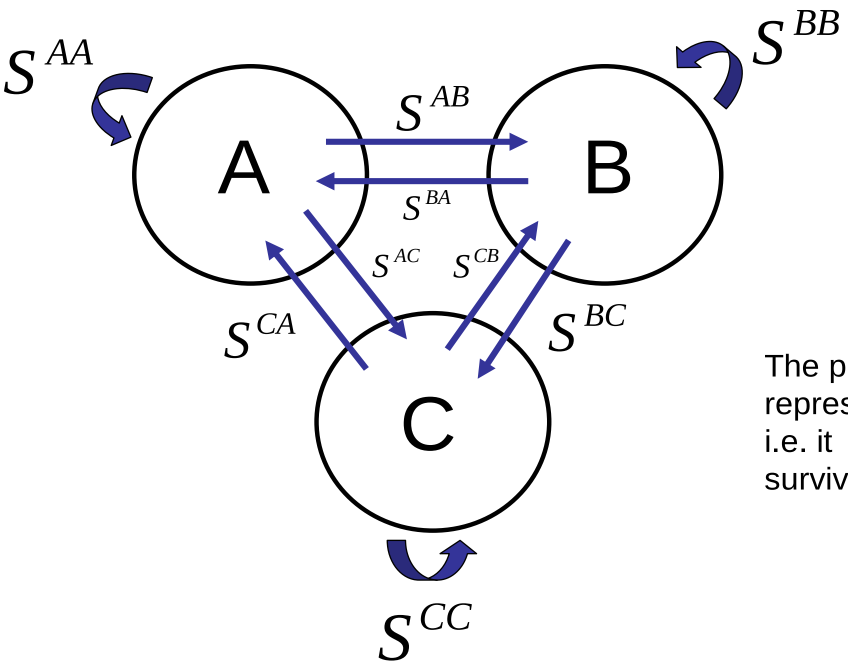
Estimation of survival/mortality from information on the recovery of dead marked animals and from observations on live marked animals.

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.

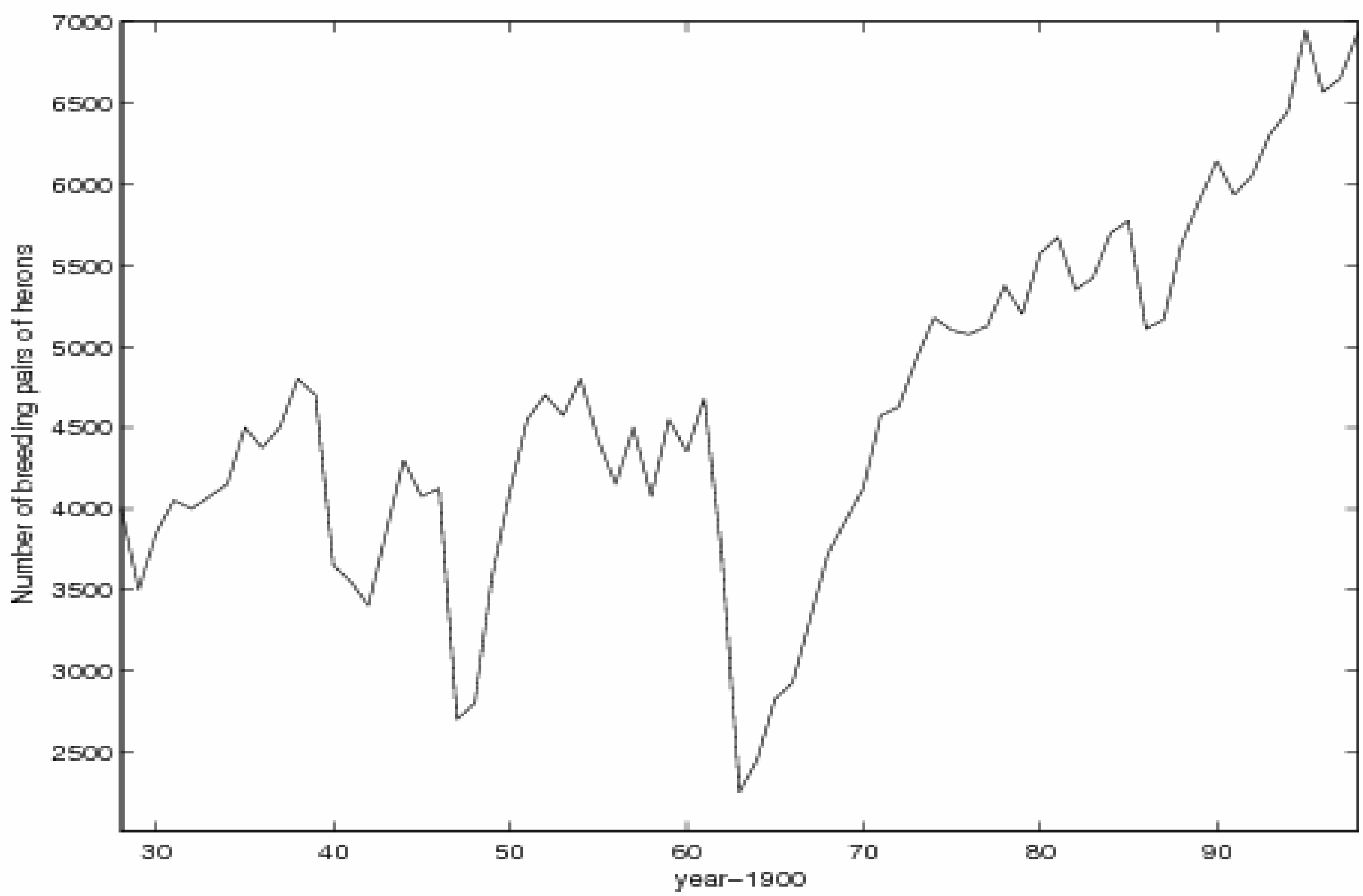
An example of a multi-site system

Multisite Systems

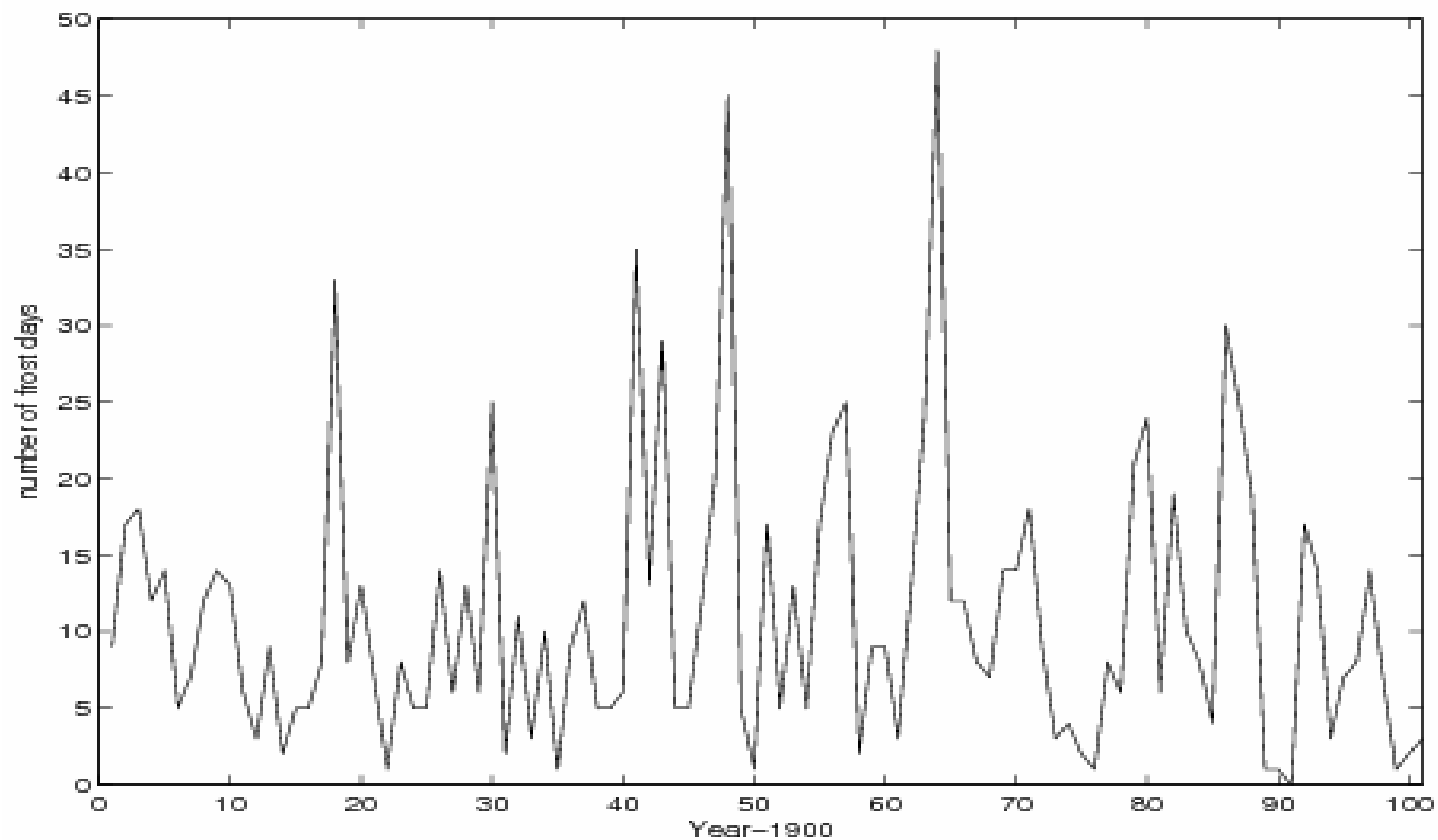


The parameter S represents the "transition", i.e. it represents both survival and movement

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:

ϕ p, ϕ^2 p(1-p), ϕ^3 p (1-p)^2 , ...

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

National Centre for Statistical Ecology

Illustration of CJS recapture probabilities: a 3-year study

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

ncse
National Centre for Statistical Ecology

CJS recapture probabilities: what we can estimate

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

National Centre for Statistical Ecology

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

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

National Centre for Statistical Ecology

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

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

ncse
National Centre for Statistical Ecology

References

- Bellman and Åström, 1970, *Mathematical Biosciences.*
- Goodman, 1974, *Biometrika*.
- Rothenberg, 1971, *Econometrica*.
- Walter, 1982, *Identifiability of state space models.*

National Centre for Statistical Ecology

Section

- Introduction and motivation
- Definitions
- General rules
- Use of symbolic algebra
- Expansion theorems
- Near redundancy
- Weak identifiability

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?

ncse
National Centre for Statistical Ecology

Section

- Introduction and motivation
- Definitions
- General rules
- Use of symbolic algebra
- Expansion theorems
- Near redundancy
- Weak identifiability

ncse
National Centre for Statistical Ecology

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.

We use this notation in Example 4 of the first Practical.

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:

$$\phi_1, \phi_2, \phi_2, \phi_3, \phi_3, \phi_4, \phi_4, \phi_4$$
$$\lambda_1, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_4, \lambda_4, \lambda_4$$

ncse
National Centre for Statistical Ecology

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:

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

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

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
- Expansion theorems
- Near redundancy
- Weak identifiability

National Centre for Statistical Ecology

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

2. The number of estimable parameters = rank(**D**).

3. Solve α^T**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 α^T**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: $\theta = [\phi_1, \phi_2, \phi_3, p_2, p_3, p_4\ ]$

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

$\overline{p}_2 = 1 - p_2$ etc

We can now use P.

Forming the derivative matrix (take logs first)

$$\mathbf{D} = \frac{\partial \ln(\mathbf{P})}{\partial \theta} = \begin{bmatrix} \phi_1^{-1} & \phi_1^{-1} & \phi_1^{-1} & 0 & 0 & 0 \\ 0 & \phi_2^{-1} & \phi_2^{-1} & \phi_2^{-1} & \phi_2^{-1} & 0 \\ 0 & 0 & \phi_3^{-1} & 0 & \phi_3^{-1} & \phi_3^{-1} \\ p_2^{-1} & -\bar{p}_2^{-1} & -\bar{p}_2^{-1} & 0 & 0 & 0 \\ 0 & p_3^{-1} & -\bar{p}_3^{-1} & p_3^{-1} & -\bar{p}_3^{-1} & 0 \\ 0 & 0 & p_4^{-1} & 0 & p_4^{-1} & p_4^{-1} \end{bmatrix}$$

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: ϕ_1, ϕ_2, p_2, p_3, $\phi_3 p_4$

National Centre for Statistical Ecology

Example – Cormack-Jolly-Seber Model with covariates

We now set

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

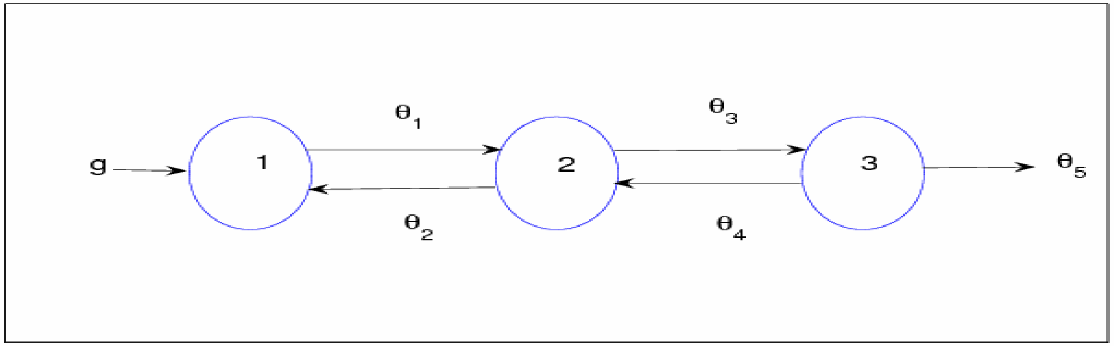
For example, x_i could be the mean annual banding weight, or the SOI.

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

and we find that the model is now full rank.

See Maple practical.

Compartment models



Simple compartment model

$$\frac{dx_1}{dt} = -(\theta_1 + \theta_2)x_1 + \theta_3 x_2 + u$$

$$\frac{dx_2}{dt} = \theta_2 x_1 - (\theta_3 + \theta_4)x_2$$

$$y = x_1$$

Transfer function

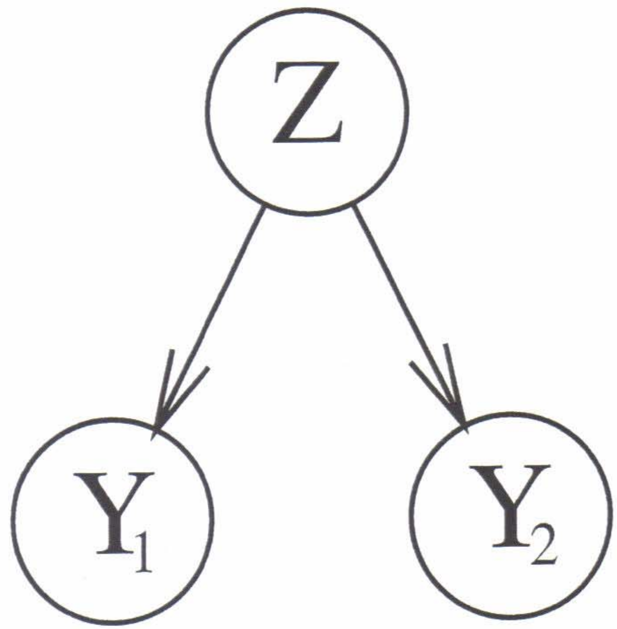
$$Q(s) = \frac{s + \theta_3 + \theta_4}{s^2 + s(\theta_1 + \theta_2 + \theta_3 + \theta_4) + \theta_1\theta_3 + \theta_1\theta_4 + \theta_2\theta_4}$$

Estimable parameters are:

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

ncse
National Centre for Statistical Ecology

A simple naïve Bayesian network



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

Naïve Bayesian Networks in general

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

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

ncse
National Centre for Statistical Ecology

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 $n=3$,

  $\mathrm{Det}(U)=-p^3(1-p)^3(\theta_{1|1}-\theta_{1|0})^2(\theta_{2|1}-\theta_{2|0})^2(\theta_{3|1}-\theta_{3|0})^2.$
- Previously conclusions followed a particular analysis.

Section

- Introduction and motivation
- Definitions
- General rules
- Use of symbolic algebra
- Extension theorems
- Near redundancy
- Weak identifiability

Extension theorems

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

For instance, the CJS model always has deficiency one, when the number of cohorts equals the number of years of recapture.

Extension Theorem

- Suppose a model has exhaustive summary κ_1 and parameters θ_1.

$$\mathbf{D}_1 = \left[\frac{\partial \kappa_{1,j}}{\partial \theta_{1,i}}\right]$$

- Now extend that model by adding extra exhaustive summary terms κ_2, and extra parameters θ_2. (Eg add more years of ringing/recovery) New model's exhaustive summary is $\kappa = [\kappa_1\ \kappa_2]^T$ and parameters are $\theta = [\theta_1\ \theta_2]^T$.

$$\mathbf{D} = \begin{bmatrix} \left[\frac{\partial \kappa_{1,j}}{\partial \theta_{1,i}}\right] & \left[\frac{\partial \kappa_{2,j}}{\partial \theta_{1,i}}\right] \\ 0 & \left[\frac{\partial \kappa_{2,j}}{\partial \theta_{2,i}}\right] \end{bmatrix} = \mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & \left[\frac{\partial \kappa_{2,j}}{\partial \theta_{1,i}}\right] \\ 0 & \mathbf{D}_2 \end{bmatrix}$$

- If $\mathbf{D}_1$ is full rank and $\mathbf{D}_2$ is full rank, the extended model will be full rank. The result can be further generalised by induction.
- Result is trivially always true, if you add zero or one extra parameters
- Method can also be used for parameter redundant models by first rewriting the model in terms of its estimable set of parameters.

ncse
National Centre for Statistical Ecology

Extension

- Example: Ring-recovery model $\theta = [\phi_{1,1} \quad \phi_{1,2} \quad \phi_a \quad \lambda_1 \quad \lambda_a]$ is a trivial example of extension theorem. Adding an extra year of ringing, adds one parameter. Adding an extra year of recovery adds no extra parameters. $\therefore$ by induction general model rank is always full rank.
- Example: Ring-recovery Model (T/A/T) 4 years ringing 5 years recovery (p = 13)

$$\mathbf{P} = \begin{bmatrix} (1-\phi_{1,1})\lambda_1 & \phi_{1,1}(1-\phi_2)\lambda_2 & \phi_{1,1}\phi_2(1-\phi_3)\lambda_3 & \phi_{1,1}\phi_2\phi_3(1-\phi_4)\lambda_4 & \phi_{1,1}\phi_2\phi_3\phi_4(1-\phi_5)\lambda_5 \\ 0 & (1-\phi_{1,2})\lambda_2 & \phi_{1,2}(1-\phi_2)\lambda_3 & \phi_{1,2}\phi_2(1-\phi_3)\lambda_4 & \phi_{1,2}\phi_2\phi_3(1-\phi_4)\lambda_5 \\ 0 & 0 & (1-\phi_{1,3})\lambda_3 & \phi_{1,3}(1-\phi_2)\lambda_4 & \phi_{1,3}\phi_2(1-\phi_3)\lambda_5 \\ 0 & 0 & 0 & (1-\phi_{1,4})\lambda_4 & \phi_{1,4}(1-\phi_2)\lambda_5 \end{bmatrix}$$

Rank($\mathbf{D}_1$) = 13, $\therefore$is full rank

- Add an extra year of recovery:

$$\mathbf{P} = \begin{bmatrix} (1-\phi_{1,1})\lambda_1 & \phi_{1,1}(1-\phi_2)\lambda_2 & \phi_{1,1}\phi_2(1-\phi_3)\lambda_3 & \phi_{1,1}\phi_2\phi_3(1-\phi_4)\lambda_4 & \phi_{1,1}\phi_2\phi_3\phi_4(1-\phi_5)\lambda_5 & \phi_{1,1}\phi_2\phi_3\phi_4\phi_5(1-\phi_6)\lambda_6 \\ 0 & (1-\phi_{1,2})\lambda_2 & \phi_{1,2}(1-\phi_2)\lambda_3 & \phi_{1,2}\phi_2(1-\phi_3)\lambda_4 & \phi_{1,2}\phi_2\phi_3(1-\phi_4)\lambda_5 & \phi_{1,2}\phi_2\phi_3\phi_4(1-\phi_5)\lambda_6 \\ 0 & 0 & (1-\phi_{1,3})\lambda_3 & \phi_{1,3}(1-\phi_2)\lambda_4 & \phi_{1,3}\phi_2(1-\phi_3)\lambda_5 & \phi_{1,3}\phi_2\phi_3(1-\phi_4)\lambda_6 \\ 0 & 0 & 0 & (1-\phi_{1,4})\lambda_4 & \phi_{1,4}(1-\phi_2)\lambda_5 & \phi_{1,4}\phi_2(1-\phi_3)\lambda_6 \end{bmatrix}$$

θ_2 = [ϕ_6 λ_6]. Rank($\mathbf{D}_2$) = 2, $\therefore$is full rank, $\therefore$ extended model is full rank.

- Add an extra year of ringing, adds one parameter $\therefore$ extended model is full rank. By induction model is T/A/T model is always full rank.

Section

- Introduction and motivation
- Definitions
- General rules
- Use of symbolic algebra
- Expansion theorems
- Data & near redundancy
- Weak identifiability

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
1086	98						52

ncse
National Centre for Statistical Ecology

Dealing with missing data & near-redundancy

- Missing data in any application can change the parameter redundancy status.
- This is easily dealt with by removing the probabilities associated with empty cells.
- Near-redundant models are full-rank, but for certain data can result in poor estimation. This may be due to similarity to a parameter-redundant sub-model. Check eigen-values of Hessian.

Section

- Introduction and motivation
- Definitions
- General rules
- Use of symbolic algebra
- Expansion theorems
- Near redundancy
- Weak identifiability

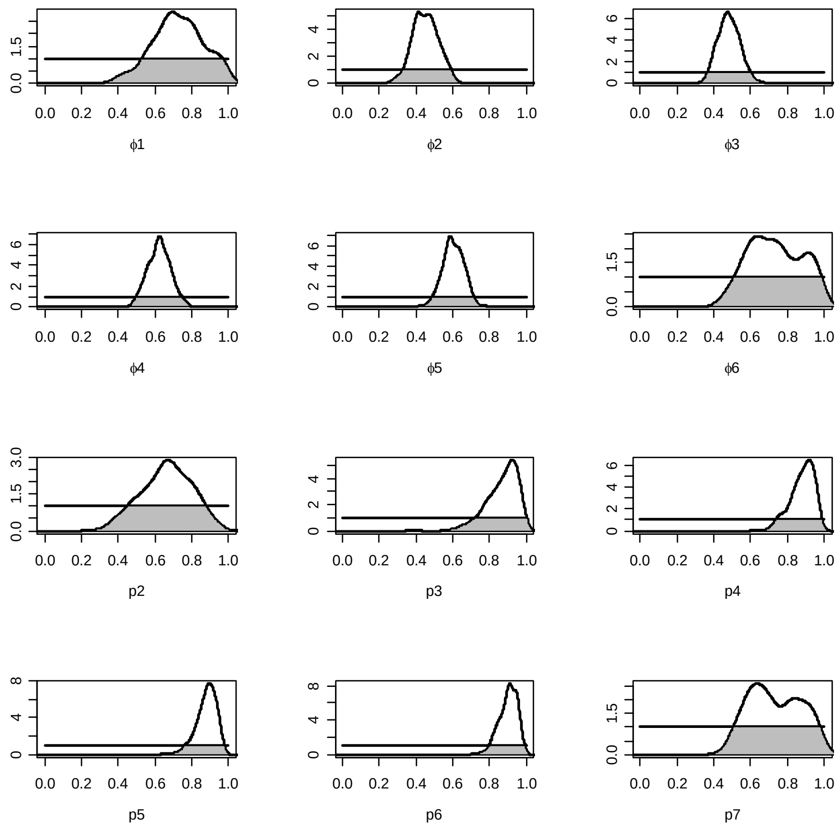
Weak identifiability: the Bayesian context

- 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.

ncse
National Centre for Statistical Ecology

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 $\pi(\theta|x)$.



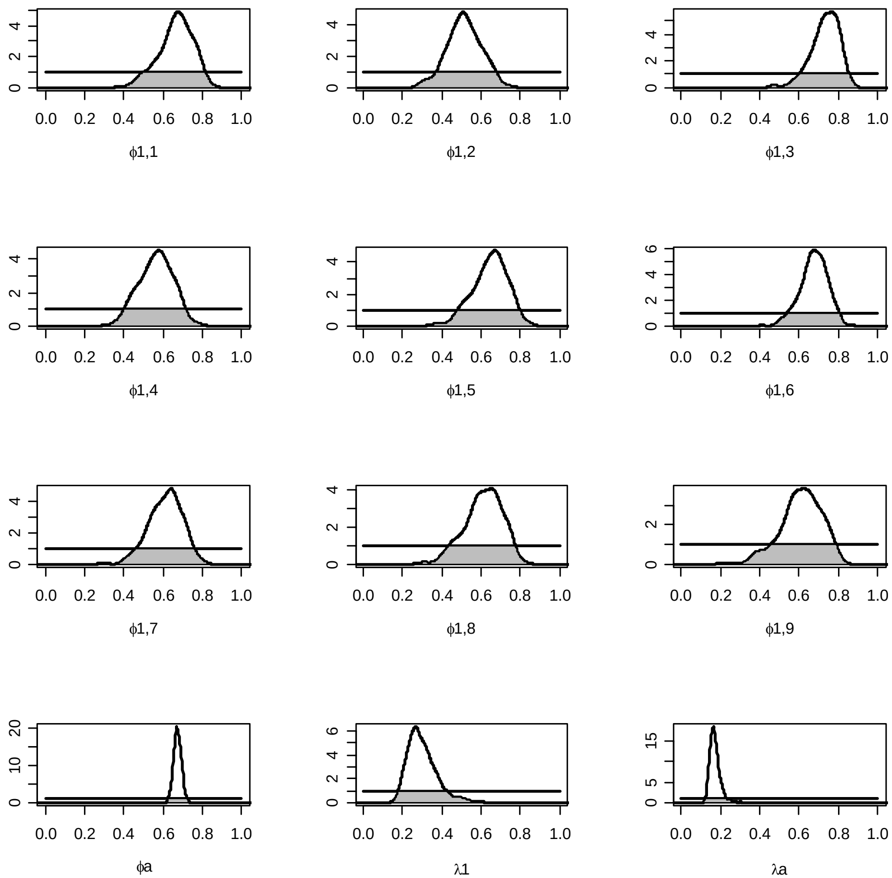
Recapture of Dippers, *Cinclus cinclus*

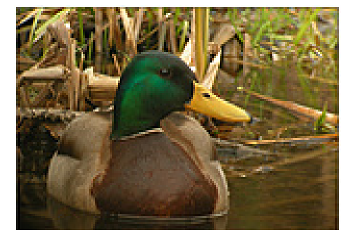
Note small cohort size in 1981.

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
1086	98						52

ncse
National Centre for Statistical Ecology

Male mallard, *Anas platyrhyncos*



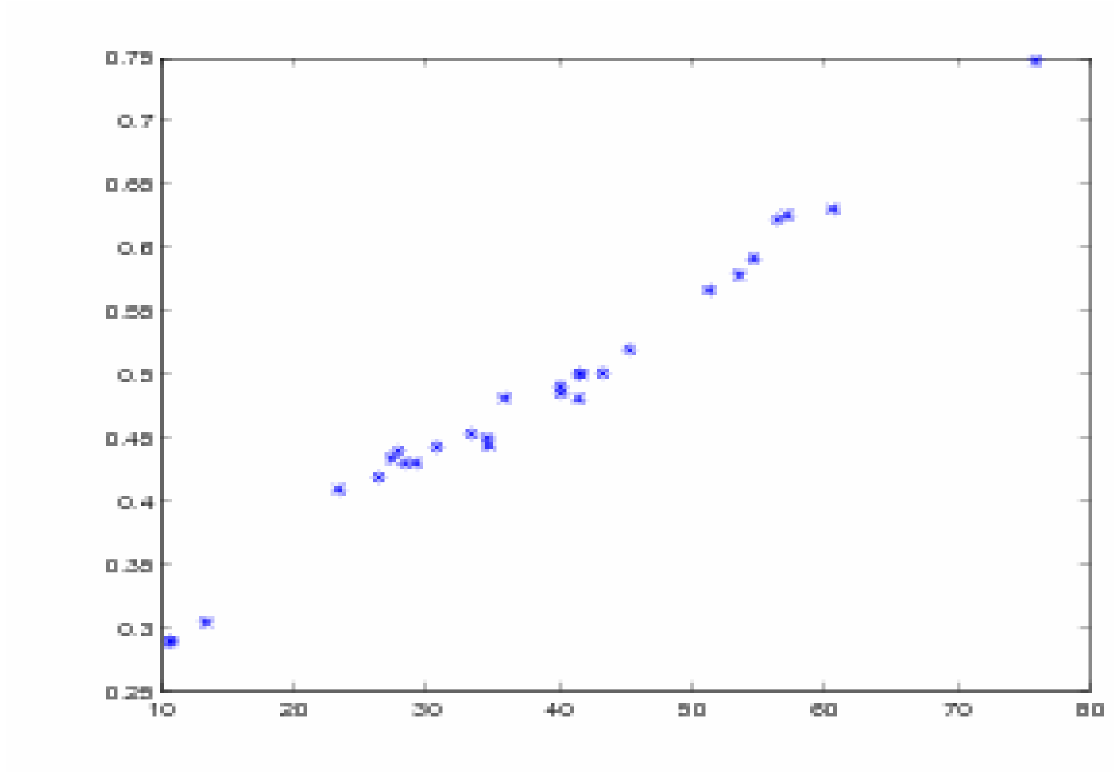


Model: T/C/A $_{(1,1)}$

$\phi_{1,i}$, ϕ_a, λ_1, λ_a here only two parameters, ϕ_a and λ_a are strongly identified. The model is near-redundant.

National Centre for Statistical Ecology

Relationship of overlap to interquartile range: simpler to calculate



National Centre for Statistical Ecology

References

- Bekker et al (1994) *Identification, equivalent models and computer algebra*.
- Catchpole, Freeman and Morgan, 1996, *JRSS B*.
- Catchpole, Kgosi and Morgan, 2001, *Biometrics.*
- 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.
- Garrett and Zeger, 2000, *Biometrics*.
- Gimenez, Morgan and Brooks, 2009. *J. Env. and Ecol. Stats*.
- Whiley, 1999, Glasgow PhD thesis.

ncse
National Centre for Statistical Ecology

Acknowledgement

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