Identifiability

Diana Cole, University of Kent, UK

Plan

Material available at:

https://www.dropbox.com/sh/hwlbfpnpxwzglvj/AABzePS9QN6tnsFBu9DrUvV6a?dl=0

Or

https://tinyurl.com/SSMident

- 50 mins - background on identifiability and method
  - Extra material provided but not covered marked with *
- 50 mins - practical split into 3 groups
  - R (Hessian and Likelihood Profile Method)
  - Winbugs (Data Cloning / Prior and Posterior Overlap)
  - Maple (Symbolic Method / Hybrid Method)
- 20 mins - discuss of results from each group

Outline

1. Introduction

1.1 Introductory Example – Lapwing Census

- Lapwing (*Vanellus vanellus*) 'census' data consists of a yearly index of abundance derived from counts of adult Lapwings.

$$\boldsymbol{y} = \begin{bmatrix} 1092 \\ 1100 \\ 1234 \\ \vdots \end{bmatrix}$$

- Let $x_{1,t}$ denote the number of 1 year old birds (unobserved) and $x_{2,t}$ number of adults (observed). Besbeas *et al* (2002) considered the following state-space model

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & \rho\phi_1 \\ \phi_a & \phi_a \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{a,t} \end{bmatrix}$$

$$y_t = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} + \eta_t$$

ϕ_1 juvenile survival probability; ϕ_a adult survival; ρ productivity; η_t and ϵ_i error processes.

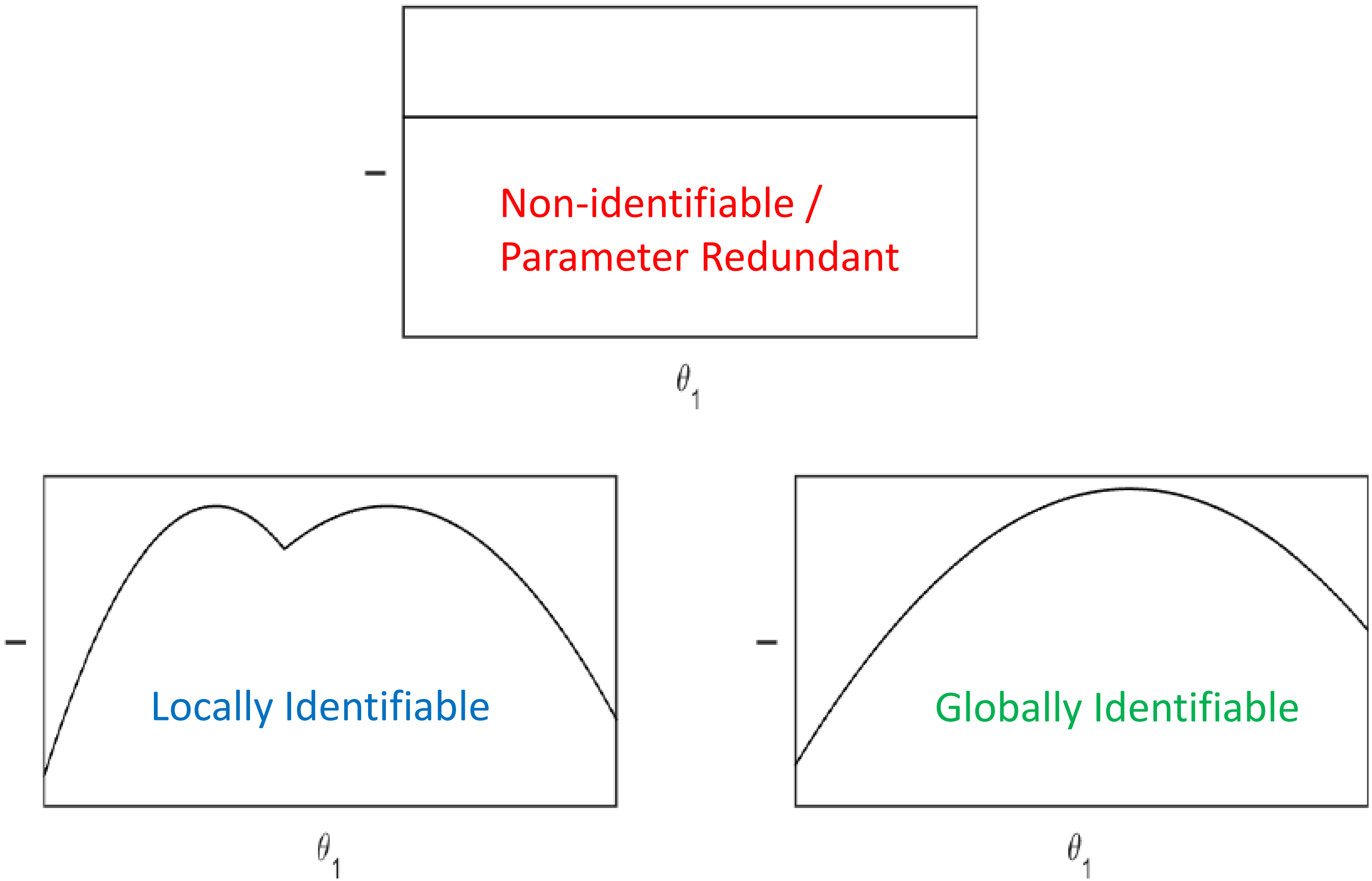
- The two parameters ρ and ϕ_1 only ever appear as a product. It will only ever be possible to estimate the product and never the two parameters separately.
- This is an example of parameter redundancy or non-identifiability.
- Reparameterised identifiable model: $\beta = \rho\phi_1$

1.2 Definitions

- Suppose we have a model $M(\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$. A model is globally (locally) identifiable if $M(\boldsymbol{\theta}_1) = M(\boldsymbol{\theta}_2)$ implies that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ (for a neighbourhood of θ).
- A model is parameter redundant if it can be reparameterised in terms of a smaller set of parameters. A parameter redundant model is non-identifiable.
- Intrinsic Parameter Redundancy / Structural Non-Identifiability – non-identifiability caused by the model.
- Extrinsic Parameter Redundancy / Non-estimability of Parameters – parameter redundancy caused by specific data (model identifiable).
- Near Redundancy – identifiable model that behaves like a parameter redundant model (typically occurs for specific data sets because there is a nested parameter redundant model).

1.2 Definitions

Log-Likelihood profiles:



1.3 Problems with Parameter Redundancy

- There will be a flat ridge in the likelihood of a parameter redundant model (Catchpole and Morgan, 1997), resulting in more than one set of maximum likelihood estimates. (A profile likelihood will be flat for a non-identifiable parameter.)
- Numerical methods to find the MLE will not pick up the flat ridge, although it could be picked up by trying multiple starting values and looking at profile log-likelihoods. If a parameter redundant model is fitted parameter estimates could be biased.
- The Fisher information matrix will be singular (Rothenberg, 1971) and therefore the standard errors will be undefined.
- However the exact Fisher information matrix is rarely known. Standard errors are typically approximated using a Hessian matrix obtained numerically. Can get explicit (but incorrect) estimates of standard errors in some cases.
- Model selection is based on identifiable models. For example AIC needs number of parameters we can estimate. If we know the number of estimable parameters can use AIC directly.
- Recommended you check the identifiability of a model before fitting (or at least as part of model fitting process).

1.3 Problems with Parameter Redundancy
Lapwing Example

Parameter	Starting Value	Estimate	SE	Starting Value	Estimate	SE
ϕ_1	0.5	0.51	45.42	0.2	0.27	NaN
ϕ_a	0.6	0.58	0.011	0.4	0.58	0.011
ρ	1.5	1.64	145.71	2	3.16	NaN
log-lik		-180.27			-180.27	

- Different starting values results in different parameter estimates for the two non-identifiable parameters (ϕ_1, ρ).
- Standard errors are either large or do not exist. Do not always obtain standard errors for identifiable parameters, theoretically SEs do not exist for any parameters in a parameter redundant model.

1.4 Bayesian Identifiability

- Lindley (1971) noted that identifiability does not present any difficulty when a Bayesian approach is used. (Their reasoning was that priors can be used to provide extra information on non-identifiable parameters.) But:
- Bayesian Identifiability (Dawid, 1979): If the parameters θ are partitioned as $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \theta_2]$, then the parameter θ_2 is not identifiable if $f(\boldsymbol{\theta}_2|\theta_1, x) = f(\boldsymbol{\theta}_2|\theta_1)$.
- Gelfand and Sahu (1999) show this is equivalent to non-identifiability of the likelihood (ignores the priors).
- If a classical model is non-identifiable then the equivalent Bayesian model will also be non-identifiable.
- Posterior identifiability: general definition of identifiability (slide 6), where the model, $M(\boldsymbol{\theta})$, is the posterior. Fits with Lindley (1971) statement.

1.4 Bayesian Identifiability

- Identifiability can be a problem in Bayesian models as well as classical models.
- Performing the MCMC analysis can be more difficult in non-identifiable models due to problems with poor mixing in MCMC samples and slow convergence (Carlin and Louis, 1996, Rannala, 2002, Gimenez et al., 2009).
- Whilst using informative priors can result in a posterior that is technically identifiable, the results will be heavily dependent on the choice of prior, even with large sample sizes (Neath and Samaniego, 1997). Therefore results for non-identifiable parameters may be misleading (Garrett and Zeger, 2000, Gimenez et al., 2009).

1.4 Bayesian Identifiability
Lapwing Example

- Posterior mean and variance:

Parameter	Prior	Mean	SD	Prior	Mean	SD	Prior	Mean	SD
ϕ_1	U(0,1)	0.38	0.21	Beta(1,3)	0.27	0.07	Beta(2,2)	0.42	0.06
ϕ_a	U(0,1)	0.63	0.01	U(0,1)	0.63	0.02	Beta(2,2)	0.63	0.02
ρ	U(0,10)	2.32	1.27	Gamma(6,3)	2.78	0.73	Gamma(50,30)	1.63	0.21

- Different priors give different results for non-identifiable parameters.
- Reparameterised identifiable model: $\beta = \rho\phi_1$:

Parameter	Prior	Mean	SD	Prior	Mean	SD	Prior	Mean	SD
ϕ_a	U(0,1)	0.63	0.02	U(0,1)	0.63	0.02	Beta(2,2)	0.61	0.01
β	U(0,10)	0.69	0.05	Gamma(6,3)	0.70	0.04	Gamma(50,30)	0.76	0.04

2. Numerical Methods

2.1 Hessian Method

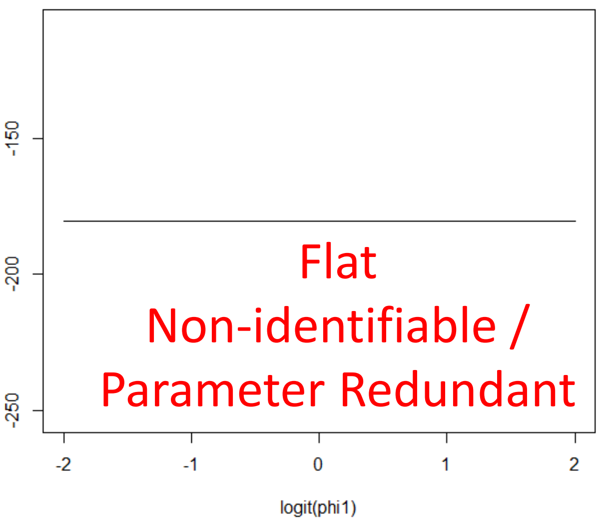
- In non-identifiable models the Fisher Information and Hessian matrix will be singular and the matrix rank will be less than the number of parameters. However when found numerically it is often not singular.
- In a non-identifiable model the matrix will have at least one eigenvalue of zero. In practice for a numerical approximation this will be very close to zero.
- If one or more of the eigenvalues are close to zero, the model is non-identifiable / parameter redundant (Viallefont, et al., 1998).
- The number of eigenvalues not close to zero is number of estimable parameters. (Standardise eigenvalues by taking modulus dividing through by the largest.)
- Suggested cut-off for standardised eigenvalues: $t = p\delta$ (p number of parameters, δ error in Hessian approximation).
- Misclassification occurs depending on cut-off used.
- In a near-redundant model the smallest eigenvalue will be close to zero as well (Catchpole et al, 2001). Results can change with data used.
- Lapwing Example:
  ϕ_1, ϕ_a, ρ standardised eigenvalues: 1, 0.0336, 0.000000038
  Non-identifiable, 2 estimable parameters
  β, ϕ_a ($\beta = \phi_1 \rho$) standardised eigenvalues: 1, 0.034
  Identifiable.

2.2 Likelihood Profile

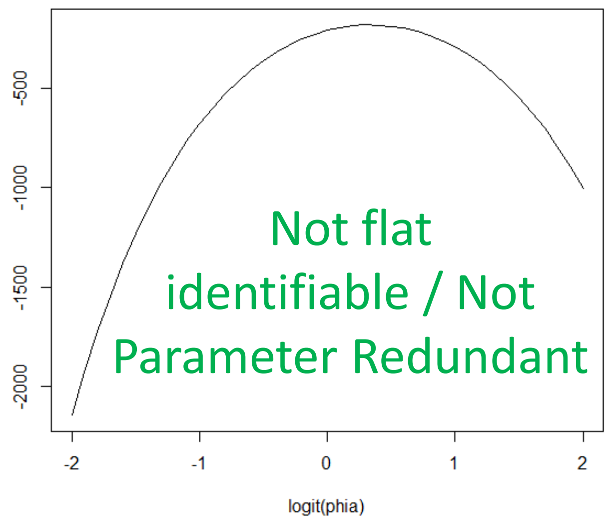
- Profile log-likelihood created by fixing a parameter at a range of values, and maximising with respect to the other parameters.
- If profile flat for parameter, parameter is non-identifiable.
- Near-redundant models will have a relatively flat profile.

- Lapwing Example:

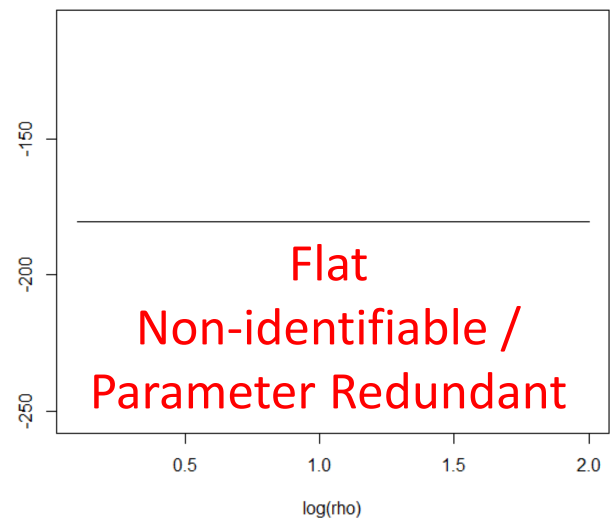
ϕ_1



ϕ_a



ρ



2.3 Data Cloning

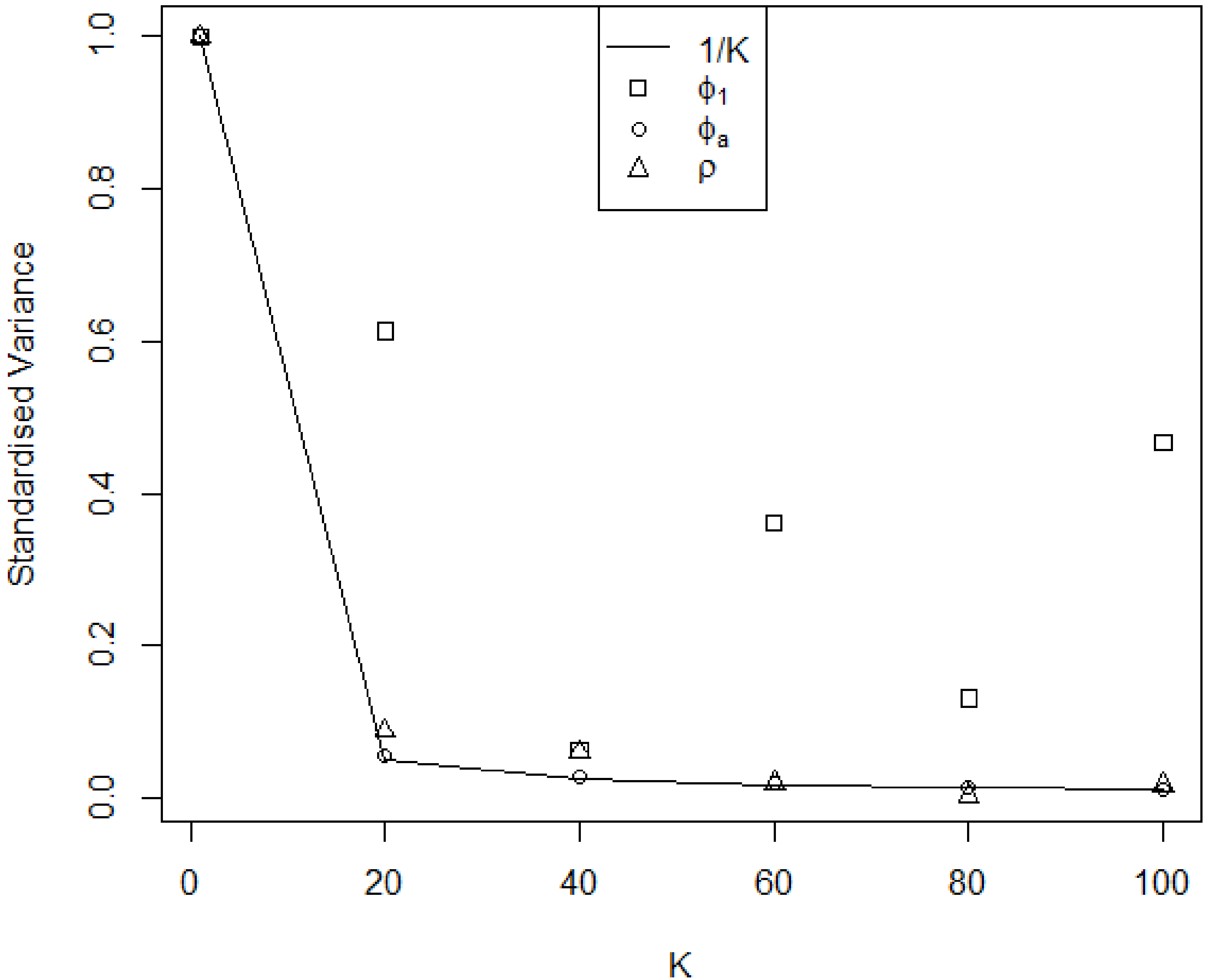
- Data cloning involves using Bayesian methodology with a likelihood based on *K* copies of the data (clones).
- The posterior mean of a parameter will tend to the MLE as *K* tends to infinity.
- The posterior variance of a parameter will tend to *K* times the asymptotic variance of the parameter.
- If a parameter is identifiable the posterior variance will tend to zero as *K* tends to infinity (Lele et al., 2010).
- If a parameter is not identifiable the posterior variance will tend to a fixed (non-zero) value.

2.3 Data Cloning

- Method:

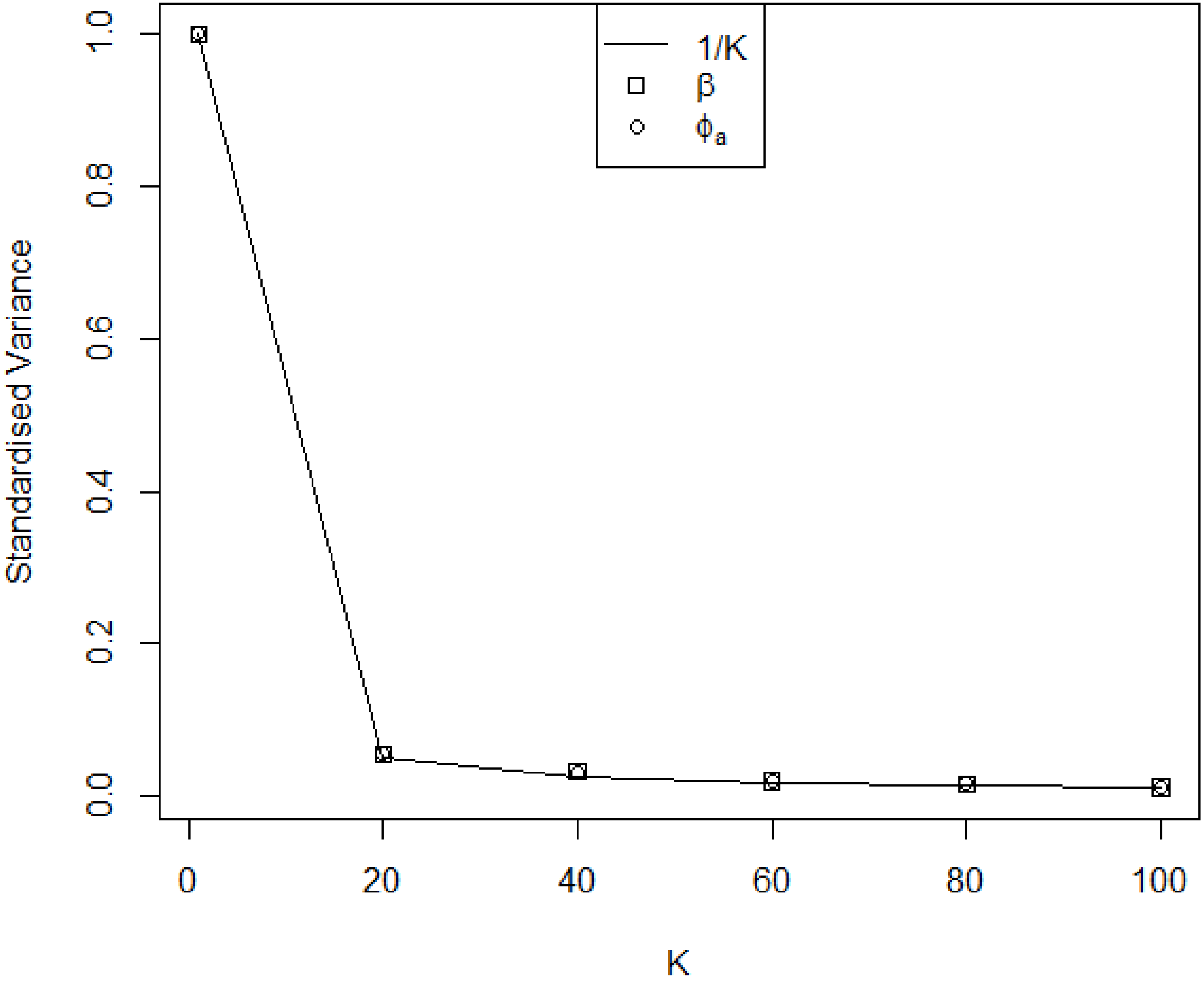
1. Perform Bayesian analysis with likelihood based on the original data; record the variance of the parameters.
2. Create a data set with K clones of the data set; perform Bayesian analysis; record variance; scale variance (÷ by variance from step 1).
3. Repeat step 2 with increasing values of K.
4. If the scaled variance is approximately equal to $1/K$ for a parameter then that parameter is identifiable. If the scaled variance is much larger than $1/K$ then that parameter is non-identifiable. If there is at least one non-identifiable parameter then the model with that set of data is non-identifiable / parameter redundant.

2.3 Data Cloning – Lapwing Example



2.3 Data Cloning – Lapwing Example

- Reparameterised identifiable model: $\beta = \rho\phi_1$:



* 2.3 Extra Data Cloning

- Different priors can be used to check estimability of parameters. If posterior mean is not sensitive to the choice or prior, the parameter is estimable (Campbell and Lele, 2014).
- Data Cloning results and the ANOVA test can be used to determine estimable parameter combinations (Campbell and Lele, 2014).
- Data cloning could be used in classical framework too. Couch and Evans (2017) suggest that for identifiable parameters

  $$\text{SE(original)} = \text{SE(cloned)} \times (\text{number of clones})^{0.5}$$

  However note that the actual standard error is undefined in a non-identifiable model.

* 2.4 Simulation

- Simulated data can be used to check for parameter redundancy by simulating a large data set from the model of interest and then fitting the same model to the data set.
- A large bias on at least one parameter indicates the model is probably parameter redundant, whereas a very small bias indicates the parameter is not parameter redundant (Gimenez et al., 2004, Kendall and Nichols, 2002).
- The threshold for the bias varies from example to example, and would depend on the sample size used.
- Depending on the threshold used the method can be inaccurate.

3. Symbolic Methods

3.1 Method

- Symbolic methods can be used to detect parameter redundancy / non-identifiability (see for example Catchpole and Morgan, 1997, Cole *et al*, 2010).
- An exhaustive summary, $\boldsymbol{\kappa}$, is a vector of parameter combinations that uniquely define the model. For example if log-likelihood is $l = \sum_{i=1}^{n} l_i$.
- The model has *p* parameters, $\boldsymbol{\theta}$.
- Form a derivative matrix $\boldsymbol{D} = \frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \kappa_1}{\partial \theta_1} & \frac{\partial \kappa_2}{\partial \theta_1} & \cdots & \frac{\partial \kappa_n}{\partial \theta_1} \\ \frac{\partial \kappa_1}{\partial \theta_2} & \frac{\partial \kappa_2}{\partial \theta_2} & & \frac{\partial \kappa_n}{\partial \theta_2} \\ \vdots & & \ddots & \vdots \\ \frac{\partial \kappa_1}{\partial \theta_p} & \frac{\partial \kappa_2}{\partial \theta_p} & \cdots & \frac{\partial \kappa_n}{\partial \theta_p} \end{bmatrix}$
- Then calculate the rank, *r*, of $\boldsymbol{D}$.
- When *r* = *p,* model is full rank and at least locally identifiable (can estimate all parameters).
- When *r* < *p,* model is parameter redundant / non-identifiable (cannot estimate all the parameters).
- To distinguish between local and global identifiability solve $\boldsymbol{\kappa}(\boldsymbol{\theta}) = \boldsymbol{k}$. Unique solution = globally identifiable.

3.1 Method: State-space models

- See Cole and McCrea (2016).
- Linear state-space model format:

$$\boldsymbol{y}_t = \boldsymbol{B}_t \boldsymbol{x}_t + \boldsymbol{\eta}_t \text{ and } \boldsymbol{x}_t = \boldsymbol{A}_t \boldsymbol{x}_{t-1} + \boldsymbol{\epsilon}_t \text{ with } \boldsymbol{x}_0 = \boldsymbol{c}_0$$

$\boldsymbol{y}_t$ observation process, $\boldsymbol{x}_t$ state equation, $\boldsymbol{B}_t$ measurement matrix, $\boldsymbol{A}_t$ transition matrix, $\boldsymbol{\eta}_t$ and $\boldsymbol{\epsilon}_t$ are error processes.

- Z-transform exhaustive summary:

  Coefficients of z in $\boldsymbol{y}_z = \boldsymbol{B}(z\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{c}_0$
- Only suitable for time-invariant linear models.
- Expansion exhaustive summary:

$$\boldsymbol{\kappa} = \begin{bmatrix} E(\boldsymbol{y}_1) \\ E(\boldsymbol{y}_2) \\ E(\boldsymbol{y}_3) \\ \vdots \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_1 \boldsymbol{A}_1 \boldsymbol{c}_0 \\ \boldsymbol{B}_2 \boldsymbol{A}_2 \boldsymbol{A}_1 \boldsymbol{c}_0 \\ \boldsymbol{B}_3 \boldsymbol{A}_3 \boldsymbol{A}_2 \boldsymbol{A}_1 \boldsymbol{c}_0 \\ \vdots \end{bmatrix}$$

- If $\boldsymbol{A}_t = \boldsymbol{A}$ is an $m \times m$ matrix then need to expand up to $\mathrm{E}(\boldsymbol{y}_{2m})$. Otherwise an extension theorem (Catchpole and Morgan, 1997, Cole *et al*, 2010) can be used.
- If the error processes involve parameters, then we can also expand the variance to extend the exhaustive summary.
- Method also extends to non-linear models.

* 3.1 Method: State-space models

- Expansion method with non-linear models (Cole and McCrea, 2016):

$$\mathbf{y}_t = h(\mathbf{x}_t, \boldsymbol{\theta}) + \boldsymbol{\eta}_t \text{ with } \mathbf{x}_t = g(\mathbf{x}_{t-1}, \boldsymbol{\theta}) + \boldsymbol{\epsilon}_{t-1}$$

$$\boldsymbol{\kappa} = \begin{bmatrix} h\{g(\mathbf{x}_0)\} \\ h[g\{g(\mathbf{x}_0)\}] \\ h(g[g\{g(\mathbf{x}_0)\}]) \\ \vdots \end{bmatrix} = \begin{bmatrix} h\{g(\mathbf{x}_0)\} \\ h\{g^2(\mathbf{x}_0)\} \\ h\{g^3(\mathbf{x}_0)\} \\ \vdots \end{bmatrix}$$

3.1 Method:
Lapwing Example – Z-transform Exhaustive Summary

$$\boldsymbol{y}_z = \boldsymbol{B}(z\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{c}_0 = \frac{-\phi_a\left(x_{0,1} + x_{0,2}\right)z - \rho\phi_1\phi_a x_{0,2}}{-z^2 + \phi_a z + \rho\phi_1\phi_a}$$

$$\boldsymbol{\kappa} = \begin{bmatrix} -\rho\phi_1\phi_a x_{0,2} \\ -\phi_a\left(x_{0,1} + x_{0,2}\right) \\ \rho\phi_1\phi_a \\ \phi_a \end{bmatrix}$$

$$\boldsymbol{\theta} = [\phi_1, \phi_a, \rho]$$

$$\boldsymbol{D} = \frac{\partial\boldsymbol{\kappa}}{\partial\boldsymbol{\theta}} = \begin{bmatrix} -\rho\phi_a x_{0,2} & 0 & \rho\phi_a & 0 \\ -\rho\phi_1 x_{0,2} & -x_{0,1} - x_{0,2} & \rho\phi_1 & 1 \\ -\phi_1\phi_a x_{0,2} & 0 & \phi_1\phi_a & 0 \end{bmatrix}$$

Rank($\boldsymbol{D}$) = r = 2, p = 3, r < p, therefore model is parameter redundant.

3.1 Method:
Lapwing Example – Expansion Exhaustive Summary

$$\boldsymbol{\kappa} = \begin{bmatrix} x_{0,1}\phi_a + x_{0,2}\phi_a \\ x_{0,2}\phi_1\phi_a\rho + \phi_a^2(x_{0,1} + x_{0,2}) \\ \phi_1\phi_a^2\rho(x_{0,1} + 2x_{0,2}) + \phi_a^3(x_{0,1} + x_{0,2}) \\ \phi_a\rho\phi_1\{x_{0,2}\phi_1\phi_a\rho + \phi_a(x_{0,1}\phi_a + x_{0,2}\phi_a)\} + \phi_a\{\phi_1\phi_a\rho\phi_a(x_{0,1}\phi_a + x_{0,2}\phi_a) + \cdots\} \end{bmatrix}$$

$$\boldsymbol{\theta} = [\phi_1, \phi_a, \rho]$$

$$\boldsymbol{D} = \frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} 0 & x_{0,2}\phi_1\rho & \\ x_{0,1} + x_{0,2} & x_{0,2}\phi_1\rho + 2\phi_a(x_{0,1} + x_{0,2}) & \ldots \\ 0 & x_{0,2}\phi_1\phi_a & \end{bmatrix}$$

Rank($\boldsymbol{D}$) = r = 2, p = 3, r < p, therefore model is parameter redundant.

3.1 General Method: Estimable Parameter Combinations

- In a parameter redundant model we can find individually identifiable parameters, and estimable parameter combinations (locally identifiable reparameterisation) – see Catchpole *et al* (1998) or Cole *et al* (2010).
- Consider a model with p parameters, rank r, deficiency $d = p - r > 0$.
- There will be d non-zero solutions to $\boldsymbol{\alpha}^T\mathbf{D} = 0$.
- Zeros in $\boldsymbol{\alpha}$s (in all d solutions) indicate estimable parameters.
- Solve PDEs to find full set of estimable pars:

$$\sum_{i=1}^{p} \alpha_{ij} \frac{\partial f}{\partial \theta_i} = 0, \qquad j = 1, \ldots d.$$

- Lapwing Example:

$\boldsymbol{\theta} = [\phi_1, \phi_a, \rho]$

$p = 3$, $r = 2$, $d = 1$

$\boldsymbol{\alpha} = \begin{bmatrix} -\frac{\phi_1}{\rho} & 0 & 1 \end{bmatrix}$ (ϕ_a identifiable)

PDE: $-\frac{\phi_1}{\rho}\frac{\partial f}{\partial \phi_1} + \frac{\partial f}{\partial \rho} = 0$

Solution: $\phi_a, \phi_1\rho$

3.1 General Method: Generalising Results

- Extension Theorem (Catchpole and Morgan, 1997, Cole *et al*, 2010) used to generalise the results – e.g. *T* years of data, or *S* states.
- Original exhaustive summary has terms $\boldsymbol{\kappa}$ and p parameters $\boldsymbol{\theta}$.
- Extend the exhaustive summary by adding extra terms $\boldsymbol{\kappa}_{ex}$ and extra p_{ex} parameters $\boldsymbol{\theta}_{ex}$.
- If $\boldsymbol{D} = \frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\theta}}$ has rank p and $\boldsymbol{D}_{ex} = \frac{\partial \boldsymbol{\kappa}_{ex}}{\partial \boldsymbol{\theta}_{ex}}$ has rank p_{ex} then the extended model has rank $p + p_{ex}$.
- If rank $= q < p$, first reparameterise in terms of q estimable parameters and then add extra q_{ex} (reparameterised) parameters. If $\boldsymbol{D}_{ex} = \frac{\partial \boldsymbol{\kappa}_{ex}}{\partial \boldsymbol{\beta}_{ex}}$ has rank q_{ex}, the extended model has rank $q + q_{ex}$.
- By extension Theorem can extend to any number of years (or states).
- If adding 0 extra parameters do not need to find $\boldsymbol{D}_{ex}$: $p_{ex} = 0$ (or $q_{ex} = 0$).
- If adding 1 extra parameter do not need to find $\boldsymbol{D}_{ex}$: $p_{ex} = 1$ (or $q_{ex} = 1$).

3.1 General Method:
Generalising Results – Lapwing Example

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & \rho\phi_{1,t} \\ \phi_{a,t} & \phi_a \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{a,t} \end{bmatrix}$$

$$y_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} + \eta_t$$

Expansion exhaustive summary $T = 2$ years:

$$\boldsymbol{\kappa} = \begin{bmatrix} \rho\phi_{1,1}x_{0,2} \\ \phi_{a,1}x_{0,1} + \phi_{a,1}x_{0,2} \\ \rho\phi_{1,2}\phi_a(x_{0,1} + x_{0,2}) \\ \rho\phi_{1,1}\phi_{a,2}x_{0,2} + \phi_{a,1}\phi_{a,2}x_{0,1} + \phi_{a,1}\phi_{a,2}x_{0,2} \end{bmatrix}$$

$$\theta = [\phi_{1,1} \quad \phi_{1,2} \quad \phi_{a,1} \quad \phi_{a,2} \quad \rho]$$

$\boldsymbol{D} = \frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\theta}}$ has rank $q = 4 < p = 5$

Estimable parameter combinations: $\rho\phi_{1,1}, \frac{\phi_{1,2}}{\phi_{1,1}}, \phi_{a,1}, \phi_{a,2}$

Equivalent to: $\beta_1 = \rho\phi_{1,1}, \beta_2 = \rho\phi_{1,2}, \phi_{a,1}, \phi_{a,2}$

3.1 Method:
Generalising Results – Lapwing Example

Reparameterise:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & \beta_t \\ \phi_{a,t} & \phi_a \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{a,t} \end{bmatrix}$$

$$y_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} + \eta_t$$

Expansion exhaustive summary $T = 2$ years:

$$\boldsymbol{\kappa} = \begin{bmatrix} \beta_1 x_{0,2} \\ \phi_{a,1} x_{0,1} + \phi_{a,1} x_{0,2} \\ \beta_2 \phi_a (x_{0,1} + x_{0,2}) \\ \beta_1 \phi_{a,2} x_{0,2} + \phi_{a,1} \phi_{a,2} x_{0,1} + \phi_{a,1} \phi_{a,2} x_{0,2} \end{bmatrix}$$

Adding an extra year adds extra terms:

$$\kappa_{ex} = \begin{bmatrix} \beta_{1,3} \phi_{a,2} (\beta_1 \phi_{a,2} x_{0,2} + \phi_{a,1} \phi_{a,2} x_{0,1} + \phi_{a,1} \phi_{a,2} x_{0,2}) \\ \beta_1 \phi_{a,2} \phi_{a,3} x_{0,2} + \beta_2 \phi_{a,1} \phi_{a,3} (x_{0,1} + x_{0,2}) + \phi_{a,1} \phi_{a,2} \phi_{a,3} (x_{0,1} + x_{0,2}) \end{bmatrix}$$

D_{ex} has rank $q_{ex} = 2$. Extended model has rank $q + q_{ex} = 4 + 2 = 6$, but 7 parameters, therefore parameter redundant / non-identifiable.

For T years rank is $2T$, but there are $2T + 1$ parameters. So will always be non-identifiable.

* 3.1 Method: Extended Symbolic Method

- In complex models, symbolic algebra packages run out of memory calculating rank (examples Jiang et al, 2007, Hunter and Caswell, 2009).
- How do you proceed:
  - Numerically – but only valid for specific value of parameters. But can't find combinations of parameters you can estimate. Not possible to generalise results.
  - Extended Symbolic Method.
  - Hybrid Symbolic-Numeric Method.
- Extended Symbolic Method: Simpler exhaustive summaries can be created using reparameterisation. First reparameterise model to simplify structure, then find the rank of the structurally simpler derivative matrix. Rank for reparameterised model is same as original parameterisation. (Method Cole et al, 2010, examples Cole and Morgan, 2010, Cole, 2012, Cole et al, 2014).

3.2 Hybrid Symbolic-Numeric Method

- Method: Choquet and Cole (2012), example application: Allen et al (2017).
- Find derivative matrix, **D**, symbolically (or using automatic differentiation).
- Find the rank of D at 5 random points, and the model rank is the maximum of all 5 ranks.
- Can get a numerical version of $\boldsymbol{\alpha}$. So can also show if any of the original parameters can be estimated, but not the estimable parameter combinations.
- Can be added to a software package. Has been added to M-Surge (Choquet et al, 2004) and E-Surge (Choquet et al, 2009).

3.2 Hybrid Symbolic-Numeric Method Example

$$FN := \begin{bmatrix} 2 & \left\{ \begin{bmatrix} -0.850857857468406 \\ 7.21089854494039\ 10^{-13} \\ 0.525395952005983 \end{bmatrix} \right\} \\ 2 & \left\{ \begin{bmatrix} -0.442598509561338 \\ 1.13036247029186\ 10^{-11} \\ 0.896719889003295 \end{bmatrix} \right\} \\ 2 & \left\{ \begin{bmatrix} -0.537993845269066 \\ 2.51465515077598\ 10^{-14} \\ 0.842948766208602 \end{bmatrix} \right\} \\ 2 & \left\{ \begin{bmatrix} -0.999997613511473 \\ 9.26082127650218\ 10^{-15} \\ 0.00218471310683500 \end{bmatrix} \right\} \\ 2 & \left\{ \begin{bmatrix} -0.694898369991373 \\ 6.09388928207721\ 10^{-12} \\ 0.719107958086498 \end{bmatrix} \right\} \end{bmatrix}$$

- Rank is 2. $p = 3$, $r < p$, therefore model is parameter redundant / non-identifiable.
- 2nd entry of Vector is almost 0. 2nd parameter is identifiable (ϕ_a).

* 3.2 Extend Hybrid Symbolic-Numeric Method

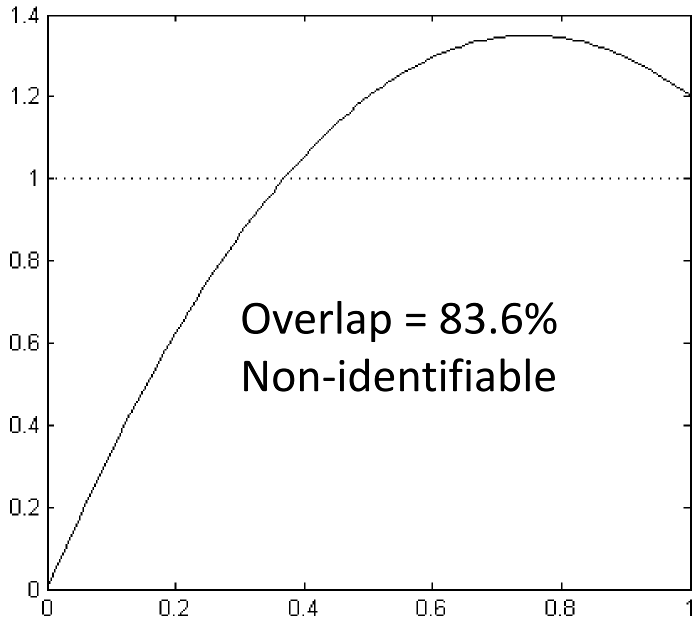
- Current research (Cole and Choquet).
- Extension theorem can be used with Hybrid method. Identical to symbolic extension theorem, but rank is calculated at 5 random points.
- Estimable parameter combinations can be found using information in the likelihood profile.
  - Eisenberg and Hayashi (2014) use subset profiling to identify estimable parameter combinations manually in compartment modelling.
  - Combine hybrid method (Choquet and Cole, 2012) with subset profiling (Eisenberg and Hayashi, 2014) and extend to automatically detect relationships between confounded parameters.

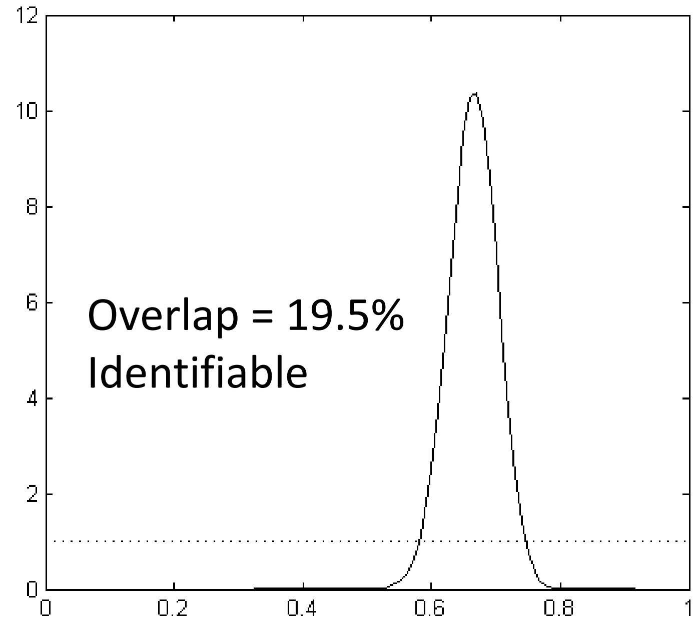
4. Bayesian Identifiability Methods

National Centre for Statistical Ecology

4.1 Overlap of Prior and Posterior
Weak Identifiability

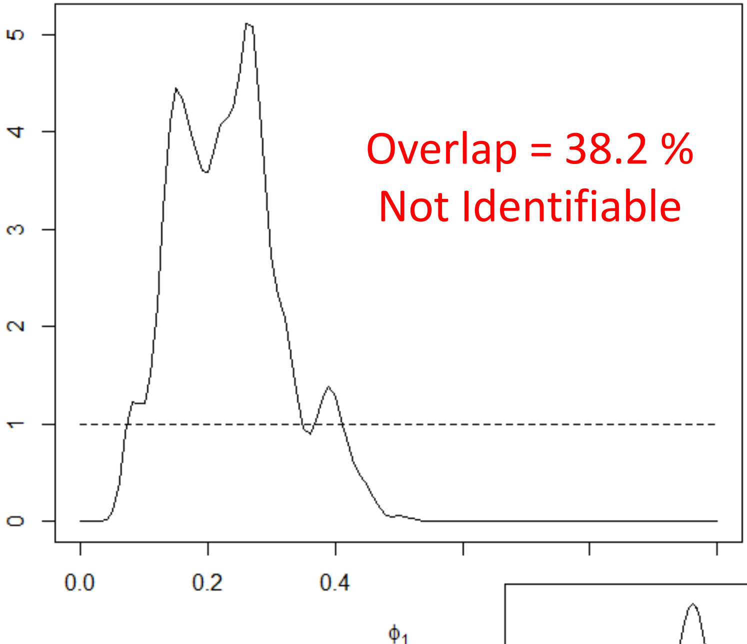
- Garrett and Zeger (2000) definition of weak identifiability is: if posterior ≈ prior.
- Check for weak identifiability by checking overlap between posterior and prior.
- Overlap of more than 35% = weakly identifiable.

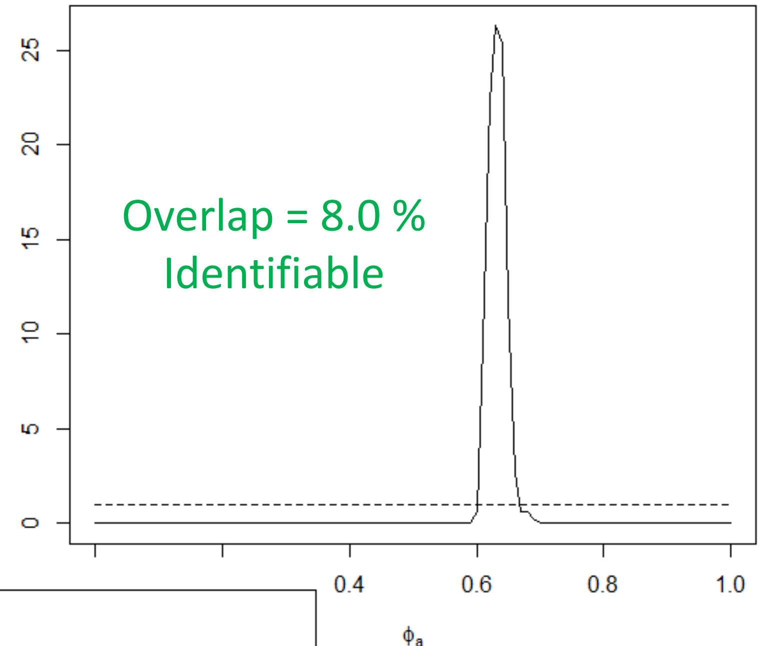


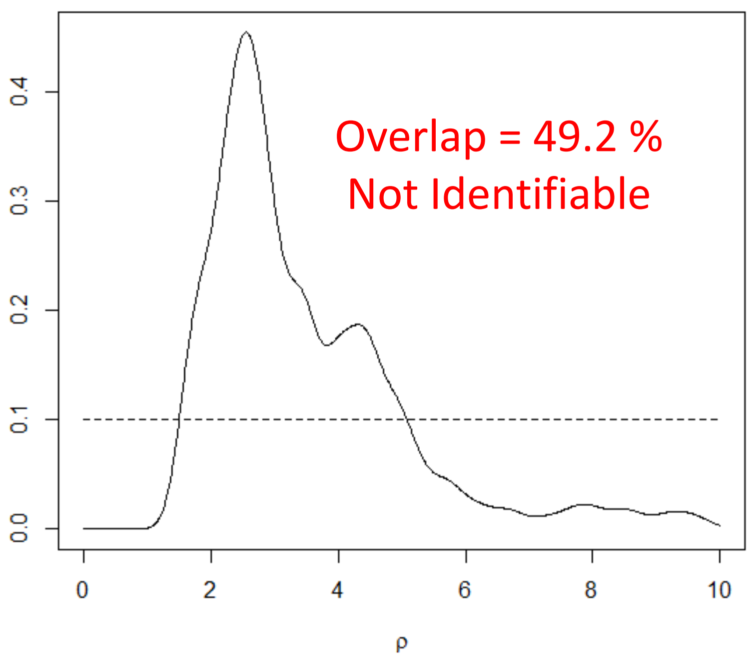


- Can lead to the wrong conclusion about identifiability, e.g. in an identifiable model with similar posterior and prior.

4.1 Overlap of Prior and Posterior
Lapwing Example







4.2 Symbolic Method

- Bayesian identifiability – exhaustive summary based on likelihood.
- Posterior identifiability – exhaustive summary with likelihood terms and prior terms.
- Can replace the likelihood exhaustive summary with any alternative exhaustive summary (e.g. Z-transform exhaustive summary).
- Lapwing example:
  - Bayesian identifiability: same result as classic model, non-identifiable.
  - Posterior identifiability: With uniform priors on all parameters, non-identifiable. With an informative prior on either ϕ_1 and/or ρ identifiable, e.g. beta prior on ϕ_1.

$$\boldsymbol{\kappa} = \begin{bmatrix} -\rho\phi_1\phi_a x_{0,2} \\ -\phi_a\left(x_{0,1} + x_{0,2}\right) \\ \rho\phi_1\phi_a \\ \phi_a \\ \phi_1^{a-1}(1-\phi_1)^{b-1} \end{bmatrix}$$

Rank of $\boldsymbol{D} = \frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\theta}}$ is 3. As there are 3 parameters, model is identifiable.

5. Discussion

5.1 Comparison of Methods

Method	Accurate	Identifiable Parameters	Estimable Parameter Combinations	General Results	Complex Models	Automatic
Hessian	×	×	×	×	✓	✓
Lik. Profile	×	✓	×	×	✓	×
Simulation	×	✓	x	x	Slow	x
Data Cloning	×	✓	×	×	Slow	×
Symbolic	✓	✓	✓	✓	×	×
Ext. Symbolic	✓	✓	✓	✓	✓	×
Hybrid	✓	✓	×	×	✓	✓
Ext. Hybrid	✓	✓	✓	✓	✓	Semi
Prior/Post overlap	×	✓	×	×	✓	×

5.2 Using non-identifiable models

- How can we use a non-identifiable model?
  - Don't use in current format – will result in biased results and misleading standard errors.
  - Reparameterise (use estimable parameter combinations – but not necessarily biologically meaningful).
  - Add constraints on parameters (results will depend on constraints, only recommended if constraints are biologically sensible).
  - Use informative priors (results will depend on priors, only recommended if have reliable prior information).
  - Combine with other data sets to create an integrated population model (Besbeas et al, 2002). Not all integrated models will be identifiable. Symbolic method involves combining exhaustive summaries (Cole and Morgan, 2016).

6. Practical

- Split into 3 groups
- Group 1: Use R
  - A. Hessian Method
  - B. Profile Method
- Group 2: Use Winbugs (via R)
  - A. Data Cloning
  - B. Overlap of Prior and Posterior
- Group 3: Use Maple
  - A. Symbolic Method
  - B. Hybrid Method
- 3 Examples
  - Eg1. Introductory Example (tutorial example)
  - Eg2. Immigration SSM + IPM
  - Eg3. Salmon SSM

6. Practical Results
Example 1: Introductory Example

	Hessian	Profile	Data Cloning	Overlap	Symbolic	Hybrid
Identifiable	No	No	No	No	No	No
Number of Estimable Parameters (AIC)	2	-	-	-	2	2
Identifiable parameters	-	ϕ_a	ρ, ϕ_a	ϕ_a	ϕ_a	ϕ_a
Estimable parameter combinations	-	-	-	-	$\phi_1 \rho$	-

6. Practical Results

Example 2: Immigration SMM + IPM (Abdai et al, 2010)

		Hessian	Profile	Data Cloning	Overlap	Symbolic	Hybrid
Census Alone	Identifiable						
	Number of Estimable Parameters (AIC)						
	Identifiable parameters						
	Estimable parameter combinations						
Census + CR	Identifiable						
	Number of Estimable Parameters (AIC)						
	Identifiable parameters						
	Estimable parameter combinations						

6. Practical Results
Example 3: Salmon SSM (Newman et al, 2014)

		Hessian	Profile	Data Cloning	Overlap	Symbolic	Hybrid
Constant α	Identifiable						
	Number of Estimable Parameters (AIC)						
	Identifiable parameters						
	Estimable parameter combinations						
Time dependent α	Identifiable						
	Number of Estimable Parameters (AIC)						
	Identifiable parameters						
	Estimable parameter combinations						

References

- Abadi, F., Gimenez, O**.,** Arlettaz, R. and Schaub, M. (2010). An assessment of integrated population models: bias, accuracy, and violation of the assumption of independence. *Ecology,* **91**, 7-14.
- Allen, S. D. and Satterthwaite, W. H. and Hankin, D. G. and Cole, D. J. and Mohr, M. S. (2017) Temporally varying natural mortality: Sensitivity of a virtual population analysis and an exploration of alternatives. *Fisheries Research*, **185**, 185-197.
- Besbeas, P., Freeman, S. N., Morgan, B. J. T. and Catchpole, E. A. (2002) Integrating Mark-Recapture-Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters. *Biometrics*, **58**, 540-547.
- Campbell, D. and Lele, S. (2014), An ANOVA test for parameter estimability using data cloning with application to statistical inference for dynamic systems, Computational Statistics and Data Analysis (70), 257-267.
- Catchpole, E. A. and Morgan, B. J. T. (1997) Detecting parameter redundancy. *Biometrika*, **84**, 187-196.
- Catchpole, E. A., Morgan, B. J. T and Freeman, S. N. (1998) Estimation in parameter redundant models. *Biometrika*, **85**, 462-468.
- Catchpole, E. A., P. M. Kgosi and B. J. T. Morgan (2001), On the near singularity of models for animal recovery data. *Biometrics* **57**, 720–726.
- Carlin, B. P. and T. A. Louis (1996) *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall/CRC.
- Choquet, R., Reboulet, A., Pradel, R., Gimenez, O. and Lebreton J.D. (2004) M-surge: new software specifically designed for multistate recapture models. *Animal Biodiversity and Conservation*, **27** ,207-215.
- Choquet, R., Rouan, L. and Pradel, R. (2009) Program E-surge: a software application for fitting multievent models, in: D.L. Thomson, E.G. Cooch, M.J. Conroy (Eds.), Modeling demographic processes in marked populations, Springer Series: Environmental and Ecological Statistics Environmental and Ecological Statistics, vol. 3, Springer, Dunedin, 845-865.
- Choquet, R. and Cole, D.J. (2012) A Hybrid Symbolic-Numerical Method for Determining Model Structure. *Mathematical Biosciences*, **236**, 117-125
- Cole, D. J., Morgan, B. J. T and Titterington, D. M. (2010) Determining the Parametric Structure of Non-Linear Models. *Mathematical Biosciences*, **228**, 16-30.
- Cole, D. J. and Morgan, B. J. T. (2010) A note on determining parameter redundancy in age-dependent tag return models for estimating fishing mortality, natural mortality and selectivity. *JABES*, **15**, 431-434.
- Cole, D.J. (2012) Determining Parameter Redundancy of Multi-state Mark-recapture Models for Sea Birds. *Journal of Ornithology* , 152 (Suppl 2), 305-315.
- Cole, D. J., Morgan, B.J.T., McCrea, R.S, Pradel, R., Gimenez, O. and Choquet, R. (2014) Does your species have memory? Capture-Recapture Data with Memory Models. *Ecology and Evolution*, 4, 2124-2133.
- Cole, D. J. and McCrea, R. M. (2016). Parameter Redundancy in Discrete State-Space and Integrated Models. *Biometrical Journal*, 58, 1071-1090.
- Cooch and Evans (2017) Program Mark. A Gentle Introduction. 17th Edition. http://www.phidot.org/software/mark/docs/book/
- Dawid, A. P. (1979). Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society, Series B*, **41**, 1–31.
- Garrett, E. S. and S. L. Zeger (2000). *Latent class model diagnosis. Biometrics*, **56**, 1055–1067
- Gelfand, A. E. and K. Sahu (1999) Identifiability, improper priors, and Gibbs sampling for generalized linear models*. Journal of the American Statistical Association*, **94**, 247–253.
- Gimenez, O., Viallefont, A., Catchpole, E. A., Choquet, R. and Morgan, B. J. T. (2004). Methods for investigating parameter redundancy. *Animal Biodiversity and Conservation* **27**, 1–12.
- Gimenez, O., Morgan, B. J. T. and Brooks, S. P. (2009). Weak Identifiability in Models for Mark-Recapture-Recovery Data. In D. Thomson, E. Cooch and M. Conroy (eds.), *Modeling demographic processes in marked populations.*, pp. 8–48, Springer series.
- Hunter, C. M. and H. Caswell (2009). Rank and redundancy of multistate mark-recapture models for seabird populations with unobservable states. In D. Thomson, E. Cooch and M. Conroy (eds.), *Modeling demographic processes in marked populations.*, pp. 797–826, Springer series.
- Jiang, H., K. H. Pollock, C. Brownie, J. E. Hightower, J. E. Hoenig and W. S. Hearn (2007), Age-dependent tag return models for estimating fishing mortality, natural mortality and selectivity. *Journal of Agricultural, Biological, and Environmental Statistics* **12**, 177–194.
- Kendall, W. L. and Nichols, J. D. (2002). Estimating state transition probabilities for unobservable states using capture-recapture/resighting data. *Ecology* **83**, 3276–3284.
- Lele, S. R., Dennis, B. and Lutscher, F. (2007). Data Cloning: Easy Maximum Likelihood Estimation for Complex Ecological Models Using Bayesian Markov Chain Monte Carlo Methods. *Ecology Letters* **105**, 551–563/
- Lindley, D. V. (1971), *Bayesian Statistics: A Review.* SIAM, Philadelphia, PA.
- Neath, A. A. and F. J. Samaniego (1997), On the efficacy of Bayesian inference for non-identifiable models. *American Statistician* **51**, 225–232.
- Newman, K.B., Buckland, S.T., Morgan, B.J.T., King, R., Borchers, D.L. and Cole, D.J. and Besbeas, P. and Gimenez, O. and Thomas, L. (2014) Modelling Population Dynamics: model formulation, fitting and assessment using state-space methods. Springer.
- Rannala, B. (2002) Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Systematic Biology* **51**, 754-760.
- Rothenberg, T. J. (1971), Identification in parametric models. *Econometrica*, **39**, 577–591
- Viallefont, A., et al. (1998) Parameter identifiability and model selection in capture-recapture models: A numerical approach. *Biometrical Journal*, **40**, 313-325.