Group 2: Investigating Identifiability Using Winbugs

This worksheet is about investigating identifiability using data cloning and examining the overlap between prior and posterior. R2WinBUGS is used to access Winbugs. Winbugs will need to be installed, and you will need to know the directory where it is installed. The worksheet assumes that Rstudio is used to run the R code – however with suitable adaptations other R platforms would also work.

Depending on where Winbugs is installed and your set up, you may need to run Rstudio as an administrator. Otherwise you may get errors stating 'Permission denied'. One way to run as administrator right click on the Rstudio icon and choose 'run as administrator' option, then you may have to state yes to the question about permissions.

If you haven't already, you will need to install the following packages:
- Coda
- R2WinBugs

The following files are needed:
- datacloning.R
- overlap.R
- eg2bugs.R
- eg3bugs.R

If several laptops are available with WinBUGS installed, datacloning.R and overlap.R could be run on different machines simultaneously.

A. Data Cloning

Open the R file datacloning.R and run the commands listed. This will perform data cloning on the introductory Lapwing example with parameters ϕ_1, ϕ_a, ρ.

Most of the instructions are given in the comments, but note the following:
- At line 7 you will need to change the directory to match the directory where WinBUGS has been installed.
- Lines 15-45 give the WinBUGS model. This is based on Brooks *et al* (2014)'s WinBUGS code using a normal approximation. Using a Poisson / Binomial model is also possible, and produces similar results, but is slower to run. Uniform priors are used on all parameters.

B. Data Cloning

Open the R file overlap.R and run the commands listed. This will investigate the overlap between the prior and posterior in the introductory Lapwing example with parameters ϕ_1, ϕ_a, ρ.

As before most of the instructions are given in the comments, but note the following:
- At line 7 you will need to change the directory to match the directory where Winbugs has been installed.
- `overlap` is a function that will calculate the proportion overlap between the prior and posterior using a kernel density to approximate the prior. It uses the function kernel to find the kernel density, based on a similar program in Morgan (2009).
- Lines 15-45 give the Winbugs model. This is based on Brooks et al (2014) Winbugs code using a normal approximation. Using a Poisson / Binomial model is also possible, and produces similar results, but is slower to run. Uniform priors are used on all parameters.

Further Examples

Now look at the two further examples on the next pages.

Example 2: Immigration Model
Abadi *et al* (2010) examine an integrated population model for a population of Little Owls, where the population is affected by immigration. The integrated population model consists of census data and capture-recapture data (there is also productivity data which is not considered here). The model for the census data is:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & 0.5\rho\phi_1 \\ \phi_a + imm & \phi_a + imm \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{a,t} \end{bmatrix}$$

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

where $x_{1,t}$ is the number of 1-year old females, $x_{2,t}$ is the number of females older than a year, y_t is the total number of females. The parameters are: ϕ_j juvenile survival probability, ϕ_a adult survival, ρ productivity and imm the immigration rate. (η_t and ϵ_i are error processes.) The capture-recapture data consists of separate data for juveniles and adults. A CJS model is used with parameters ϕ_j, ϕ_a and p, where p is the recapture probability. (Note here we ignore the data on males, and only include the female data – this does not change the identifiability results).

The WinBUGS model is given in eg2bugs.R. This is adapted from the WinBUGS code given in Abadi *et al* (2010).

Is the model with census data alone identifiable? Is the integrated model with census and capture-recapture data identifiable? Which parameters can be estimated in both cases?

Note that data cloning is very slow and may not run with in the time allowed. It is possible to reduce the number of samples or use smaller values of K. However this may not produce ideal results. Results for a longer run (90000 samples, burnin 10000) are given below:

Integrated Census and Capture-Recapture
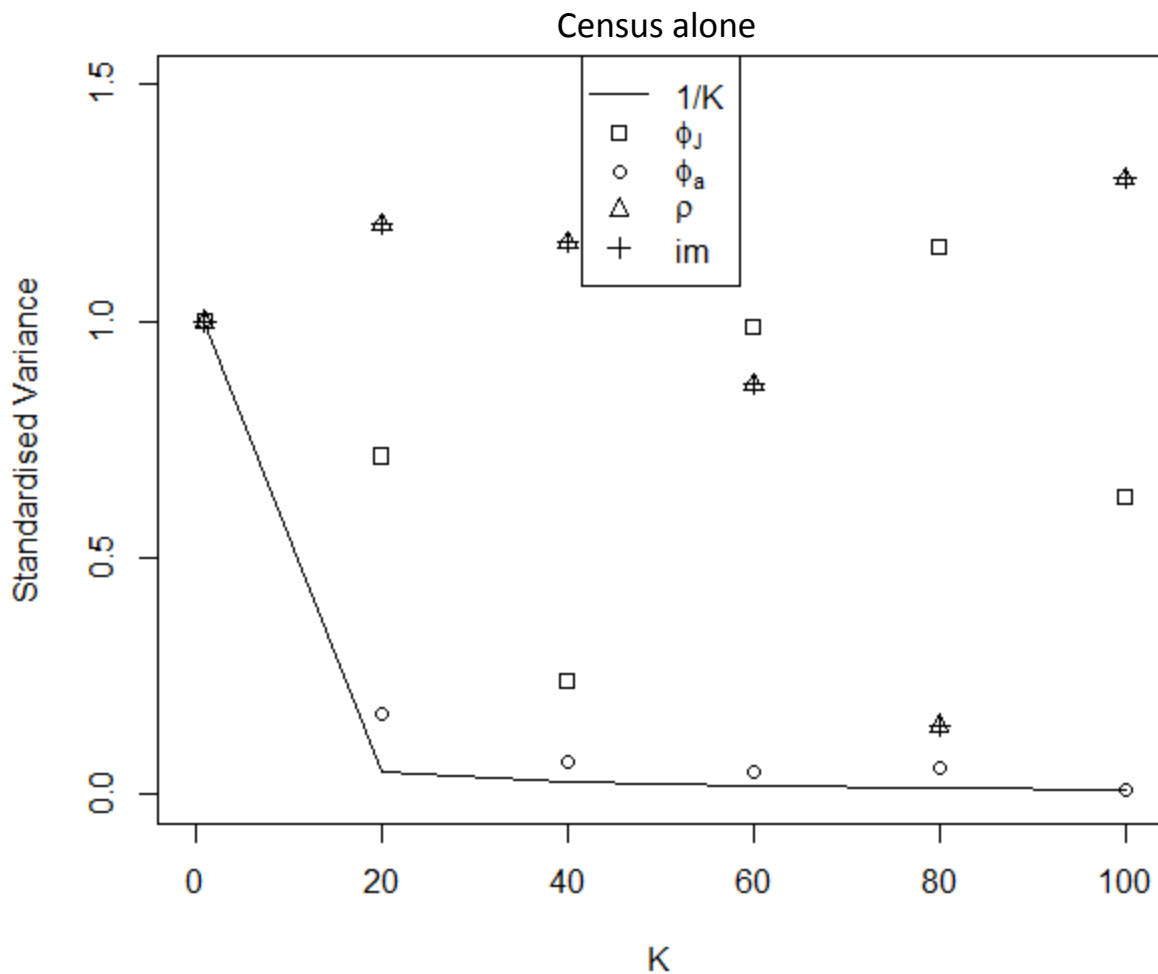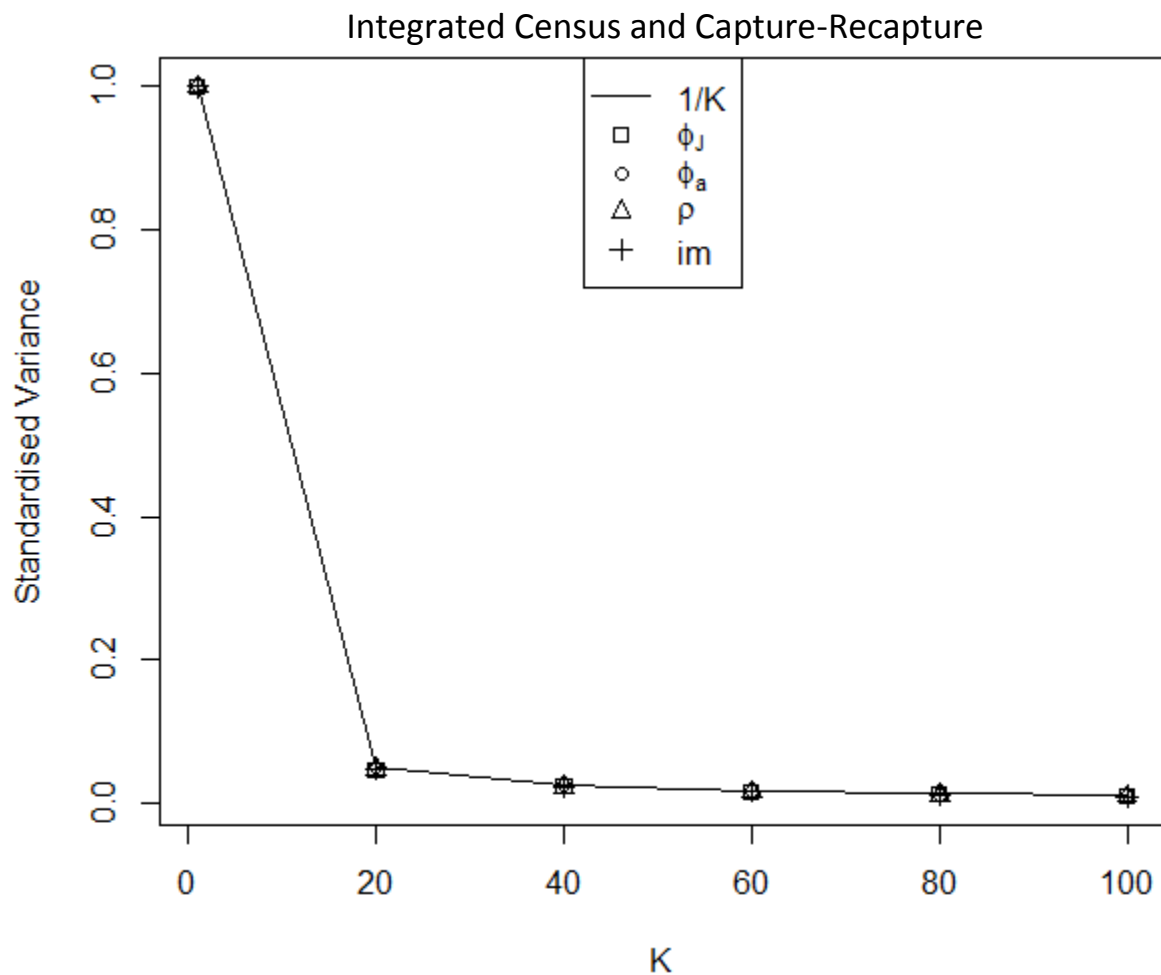
Example 3: Salmon State-Space Model

This example is from Newman *et al* (2014). The number of juvenile Salmon alive in year *t* in a certain river can be modelled as

$$N_t|N_{t-1} \sim Po(\alpha N_{t-1} \exp(-\beta N_{t-1})), \qquad \alpha > 0, \beta > 0.$$

The parameter α models survival and fecundity, the parameter β models density dependence and the parameter $n_0 = N_0$. The observation process is

$$y_t|N_t \sim lognormal\left(\log(N_t) - \frac{\sigma^2}{2}, \sigma^2\right).$$

σ^2 is assumed to be a fixed known value.

The winbugs code given in eg3bugs.m is based on winbugs code given in Newman et al (2014).

Is the model identifiable? When considering the posterior and prior overlap consider what happens if you increase the range of uniform prior distributions on alpha and beta.

What if the parameter α was time dependent?

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.

Brooks, S. P., King, R. and Morgan, B. J. T. (2004) A Bayesian Approach to Combining Animal Abundance and Demographic Data. *Animal Biodiversity and Conservation* **27**, 515-529.

Morgan, B. J. T. (2009), *Applied Stochastic Modelling*. Chapman and Hall.