

Analysis of dependent data

Population growth, extinction and viability analysis

Subhash R. Lele

Department of Mathematical Sciences
University of Alberta
Canada

Email: slele@ualberta.ca

July 14, 2016

Population growth models

You are now quite familiar with the use of hierarchical models for independent data. We have looked at how hierarchical models arise because of sampling error but they can arise in many other situations as well. We will now show you how hierarchical models arise in another area of interest to the conservation biologists: Population Viability Analysis.

This naturally leads to analysis of time series, spatial and spatio-temporal data. The main difference is theoretical in terms of how one writes the likelihood function. However, computational aspects are not substantially different than what we have learnt so far. We will also learn about how to deal with **missing data in time series** and how to obtain **predictions and prediction intervals**.

Modelling single species population growth

General model: Future value N_{t+1} depends on the current status $g(N_t; \theta)$ only. This is a first order Markovian property. One can easily extend this to dependence of higher order by including the term involving N_{t-1} etc.

$$N_{t+1} = f(N_t; \theta) \quad (1)$$

Many times, we model the growth in the population, rather than the population itself. Hence

$$\log(N_{t+1}) - \log(N_t) = g(N_t; \theta) \quad (2)$$

We will be using the growth models in the rest of this lecture.

The two most commonly used models for population growth are:

Examples:

Ricker or Discrete time Logistic growth: $g(N_t; \theta) = a + bN_t$

Gompertz growth: $g(N_t; \theta) = a + b \log(N_t)$

These are strictly deterministic model. However, in practice, we have stochasticity in the growth rates due to environmental factors or demographic factors.

We can incorporate environmental stochasticity in these models by using additive noise on the log-scale. The general structure of such models is:

$$\log(N_{t+1}) - \log(N_t) = g(N_t; \theta) + \varepsilon_{t+1} \quad (3)$$

Examples:

Ricker or Discrete time Logistic growth: $g(N_t; \theta) = a + bN_t + \varepsilon_{t+1}$

Gompertz growth: $g(N_t; \theta) = a + b \log(N_t) + \varepsilon_{t+1}$

where $\varepsilon_{t+1} \sim N(0, \sigma^2)$

Now we will see how to use JAGS and 'dclone' to fit this simple, stochastic, time series model, in either Bayesian or Frequentist paradigm.

Now we will see how to use JAGS and 'dclone' to fit this simple, stochastic, time series model, in either Bayesian or Frequentist paradigm.

Missing data: In practice, we may have a few years where we did not or could not collect the data. If the data are missing completely at random (that is, we did not skip the survey because we knew the populations were too low etc.), then the above program can be easily modified as follows.

Population Viability Analysis

For conservation purposes, we may not want simply the parameter values but we may also want to predict how the population would behave in future. We may want to project the population sizes for the next few years and compute the prediction intervals, extinction probability or other population viability related metrics.

Population Viability Analysis

For conservation purposes, we may not want simply the parameter values but we may also want to predict how the population would behave in future. We may want to project the population sizes for the next few years and compute the prediction intervals, extinction probability or other population viability related metrics.

We can modify the above program to do the population projection very easily as follows.

Population Viability Analysis

For conservation purposes, we may not want simply the parameter values but we may also want to predict how the population would behave in future. We may want to project the population sizes for the next few years and compute the prediction intervals, extinction probability or other population viability related metrics.

We can modify the above program to do the population projection very easily as follows.

Once you have the collection of the future trajectories, computing various metrics is a simple task.

Sampling variability

Above discussion and illustrations assumed that we knew the population sizes without any error. But, in practice, these are usually estimated using some kind of sampling as we have seen earlier. If we do not take into account the variability of these **estimate** population sizes, our statistical inferences are bound to be incorrect. This is similar to not taking into account detection error.

Sampling variability

Above discussion and illustrations assumed that we knew the population sizes without any error. But, in practice, these are usually estimated using some kind of sampling as we have seen earlier. If we do not take into account the variability of these **estimate** population sizes, our statistical inferences are bound to be incorrect. This is similar to not taking into account detection error.

We can modify the model quite easily using the hierarchical modelling ideas. We will assume Poisson distribution for the sampling variability for illustration purpose but it could be anything else (with the constraint that the parameters are estimable).

Observation model: $\hat{N}_t \sim \text{Poisson}(N_t)$

Observation model: $\hat{N}_t \sim \text{Poisson}(N_t)$

Process model: $\log(N_{t+1}) - \log(N_t) = g(N_t; \theta) + \varepsilon_{t+1}$

Observation model: $\hat{N}_t \sim \text{Poisson}(N_t)$

Process model: $\log(N_{t+1}) - \log(N_t) = g(N_t; \theta) + \varepsilon_{t+1}$

Prior: $\theta \sim \pi(\theta)$

We can easily modify the previous JAGS model to accommodate the sampling error as follows.

We can easily modify the previous JAGS model to accommodate the sampling error as follows.

Modification to data cloning based MLE is similar to the previous lecture. As an added bonus, we also get diagnostics for estimability of the parameters.

SUMMARY

1. Writing the JAGS model for dependent data structure is very similar to the independent data structure.
2. If you can write a JAGS model, you can do the Bayesian analysis quite easily.
3. If you can do the Bayesian analysis successfully, you can modify the JAGS model slightly and also can get the likelihood based analysis.
4. We are only showing you how to obtain the MLE and its standard error. However, one can also get Likelihood ratio based confidence intervals, Confidence intervals that are valid under smaller sample sizes and model selection using information criteria. You can also get prediction intervals for future projections.

1. An important point to remember when fitting models is that just because one can get answers from these programs does not mean they are valid answers.

1. An important point to remember when fitting models is that just because one can get answers from these programs does not mean they are valid answers.
2. We have to make sure the parameters are estimable using the available data. This is a major issue that has been ignored in ecological analyses.

1. An important point to remember when fitting models is that just because one can get answers from these programs does not mean they are valid answers.
2. We have to make sure the parameters are estimable using the available data. This is a major issue that has been ignored in ecological analyses.
3. We have to make sure that the MCMC algorithm has converged. This is never guaranteed but can be judged to some extent using various diagnostic procedures.

1. An important point to remember when fitting models is that just because one can get answers from these programs does not mean they are valid answers.
2. We have to make sure the parameters are estimable using the available data. This is a major issue that has been ignored in ecological analyses.
3. We have to make sure that the MCMC algorithm has converged. This is never guaranteed but can be judged to some extent using various diagnostic procedures.
4. Parameterization (form of the model) can have a major impact on the convergence of MCMC algorithm. Some parameterizations may lead to easy convergence while others may not. This can lead to some thorny scientific issues as we will discuss in the next lecture.

1. An important point to remember when fitting models is that just because one can get answers from these programs does not mean they are valid answers.
2. We have to make sure the parameters are estimable using the available data. This is a major issue that has been ignored in ecological analyses.
3. We have to make sure that the MCMC algorithm has converged. This is never guaranteed but can be judged to some extent using various diagnostic procedures.
4. Parameterization (form of the model) can have a major impact on the convergence of MCMC algorithm. Some parameterizations may lead to easy convergence while others may not. This can lead to some thorny scientific issues as we will discuss in the next lecture.
5. We will now show you how to analyze spatially dependent data. This involves some difficult issues of how to specify the priors on the spatial dependence parameter.