A few Bayesian lectures for the Uninitiated — Part 6: state-space models (= hidden Markov models)

Carsten F. Dormann

Biometry & Environmental System Analysis, University of Freiburg

20. April 2022

State-space model

- combines a process model and an observation model

State-space model

- combines a process model and an observation model
- the process model is unobservable = latent = hidden

State-space model

- combines a process model and an observation model
- the process model is unobservable = latent = hidden
- all inferences about the process are through the lens of the observation model

An SSM is also a hierarchical model

Process model

$N_{\text{true}} = \text{birth rate} * N - \text{death rate} * N = (b - d)N = rN$

r is to be estimated, with uncertainty (= error)

An SSM is also a hierarchical model

Process model

$N_{\text{true}} = \text{birth rate} * N - \text{death rate} * N = (b - d)N = rN$

r is to be estimated, with uncertainty (= error)

Observation model

$y \sim \text{NegBin}(\mu = e^{N_{\text{true}}}, \sigma = c)$ (alternative formulation of NegBin!)

y is observed with some (measurement) error

An SSM is also a hierarchical model

Process model

$N_{\text{true}} = \text{birth rate} * N - \text{death rate} * N = (b - d)N = rN$

r is to be estimated, with uncertainty (= error)

Observation model

$y \sim \text{NegBin}(\mu = e^{N_{\text{true}}}, \sigma = c)$ (alternative formulation of NegBin!)

y is observed with some (measurement) error

$\longrightarrow$ Two sources of variation: in the process, and in the observation.

Time series example

$$y_t \sim D_1(\text{mean} = D_2(f(N_{t-1}, r), \text{scale=d}), \text{scale} = c)$$
$$y_{t-1} \sim D_1(\text{mean} = D_2(f(N_{t-2}, r), \text{scale=d}), \text{scale} = c)$$

Time series example

$y_t \sim D_1(\text{mean} = D_2(f(N_{t-1}, r), \text{scale=d}), \text{scale} = c)$
$y_{t-1} \sim D_1(\text{mean} = D_2(f(N_{t-2}, r), \text{scale=d}), \text{scale} = c)$
Thus y_t depends on how y_{t-1} was estimated.

Time series example

$y_t \sim D_1(\text{mean} = D_2(f(N_{t-1}, r), \text{scale=d}), \text{scale} = c)$
$y_{t-1} \sim D_1(\text{mean} = D_2(f(N_{t-2}, r), \text{scale=d}), \text{scale} = c)$
Thus y_t depends on how y_{t-1} was estimated.

This leads to an ugly conditional, recursive expression for the likelihood of observing y_t, along the lines of:
$P(y_t) = P(y_t|y_{t-1})P(y_{t-1}) = P(y_t|y_{t-1})P(y_{t-1}|P(y_{t-2}))P(y_{t-2}) = \ldots$

Time series example

$y_t \sim D_1(\text{mean} = D_2(f(N_{t-1}, r), \text{scale=d}), \text{scale} = c)$
$y_{t-1} \sim D_1(\text{mean} = D_2(f(N_{t-2}, r), \text{scale=d}), \text{scale} = c)$
Thus y_t depends on how y_{t-1} was estimated.

This leads to an ugly conditional, recursive expression for the likelihood of observing y_t, along the lines of:
$P(y_t) = P(y_t|y_{t-1})P(y_{t-1}) = P(y_t|y_{t-1})P(y_{t-1}|P(y_{t-2}))P(y_{t-2}) = \ldots$

That means:

- the process error (estimation of r) cannot be simply estimated as we all do as $y \sim D(\hat{y}, error)$

Time series example

$y_t \sim D_1(\text{mean} = D_2(f(N_{t-1}, r), \text{scale=d}), \text{scale} = c)$
$y_{t-1} \sim D_1(\text{mean} = D_2(f(N_{t-2}, r), \text{scale=d}), \text{scale} = c)$
Thus y_t depends on how y_{t-1} was estimated.

This leads to an ugly conditional, recursive expression for the likelihood of observing y_t, along the lines of:
$P(y_t) = P(y_t|y_{t-1})P(y_{t-1}) = P(y_t|y_{t-1})P(y_{t-1}|P(y_{t-2}))P(y_{t-2}) = \ldots$

That means:

- the process error (estimation of r) cannot be simply estimated as we all do as $y \sim D(\hat{y}, error)$
- the estimate for the final data points is affected by the probability of getting to the previous data point, *and* the probability of observing its correct value.

Time series example

$y_t \sim D_1(\text{mean} = D_2(f(N_{t-1}, r), \text{scale=d}), \text{scale} = c)$
$y_{t-1} \sim D_1(\text{mean} = D_2(f(N_{t-2}, r), \text{scale=d}), \text{scale} = c)$
Thus y_t depends on how y_{t-1} was estimated.

This leads to an ugly conditional, recursive expression for the likelihood of observing y_t, along the lines of:
$P(y_t) = P(y_t|y_{t-1})P(y_{t-1}) = P(y_t|y_{t-1})P(y_{t-1}|P(y_{t-2}))P(y_{t-2}) = \ldots$

That means:

- the process error (estimation of r) cannot be simply estimated as we all do as $y \sim D(\hat{y}, error)$
- the estimate for the final data points is affected by the probability of getting to the previous data point, *and* the probability of observing its correct value.
- Both process error and observation error thus affect y_{t-k} and hence conditionally each following value.

SSM in reality (Clark et al. 2007 Ecol Appl)

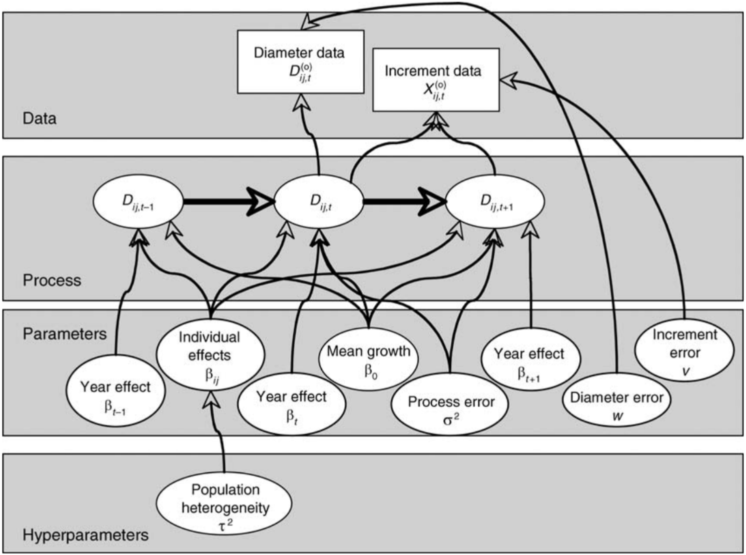

FIG. 1. Graphical representation of the model. The main process to be modeled is represented by the change in diameter ("Process" box). For each tree and year there may be diameter data, increment data, or both ("Data" box). Diameter growth depends not only on data, but also on parameters ("Parameters" box), allowing for population heterogeneity ("Hyperparameters" box).

In JAGS

That's the entire point of BUGS/JAGS/STAN: help developing the long conditional likelihood, without writing it down explicitly!

```
# State process
for (t in 1:(T-1)){
        r[t] ~ dnorm(mean.r, tau.proc)
        logN.est[t+1] <- logN.est[t] + r[t]
}

# Observation process
for (t in 1:T){
        y[t] ~ dnorm(logN.est[t], tau.obs)
}
```

Will this never end?

It does now (after the R-bit), **but Bayes will go on forever ...**

Useful references:

Kéry, M., & Royle, J. A. (2015). Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS. Volume 1: Prelude and Static Models. Academic Press.

Kéry, M., & Royle, J. A. (2020). Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS: Volume 2: Dynamic and Advanced Models. Academic Press.

Kéry, M., & Schaub, M. (2011). Bayesian Population Analysis using WinBUGS: A Hierarchical Perspective. Academic Press.