Stability properties for MAR(1) models

Mark D. Scheuerell

Northwest Fisheries Science Center

National Oceanic and Atmospheric Administration

Seattle, WA USA

mark.scheuerell@noaa.gov

Background

There is growing interest in the use of first-order vector autoregressive, or VAR(1), models in ecology where they are often referred to as multivariate autoregressive, or MAR(1), models (*e.g.*, Ives *et al.* 2003 *Ecological Monographs* 73:301–330).

Assume a MAR(1) model of the general form

$$\mathbf{x}_t = \mathbf{a} + \mathbf{B}(\mathbf{x}_{t-1} - \mathbf{a}) + \mathbf{w}_t$$

where $\mathbf{x}_t$ is an $n \times 1$ vector of state variates at time t, $\mathbf{a}$ is an $n \times 1$ vector of underlying levels (means) for each of the states, $\mathbf{B}$ is an $n \times n$ interaction matrix, and $\mathbf{w}_t$ is an $n \times 1$ vector of multivariate normal process errors; $\mathbf{w}_t \sim \mathrm{MVN}(\mathbf{0}, \mathbf{Q})$.

Of particular interest here is that the variance-covariance matrix of the stationary distribution for $\mathbf{x}_t$ as $t \to \infty$ gives an indication of the relative stability of the system.

State-space form

This MAR(1) model can be used within a state-space framework, wherein $\mathbf{a}$ is instead included as part of a second model for the observed data $\mathbf{y}$, such that

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{w}_t,$$
$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{a} + \mathbf{v}_t,$$

and $\mathbf{v}_t$ is an $n \times 1$ vector of observation errors, the statistical distribution of which does not affect the variance of the stationary distribution.

Variance of the stationary distribution

We will restrict this treatment to stationary models wherein all of the eigenvalues of $\mathbf{B}$ lie within the unit circle. Because $t = (t - 1)$ as $t \to \infty$, under assumptions of stationarity we can write the process equation from the above state-space model as

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_t + \mathbf{w}_t.$$

From this, it follows that

$$\text{Var}(\mathbf{x_t}) = \mathbf{B}\text{Var}(\mathbf{x_t})\mathbf{B}^\top + \text{Var}(\mathbf{w_t}).$$

If we define $\mathbf{\Sigma} = \text{Var}(\mathbf{x_t})$, then

$$\mathbf{\Sigma} = \mathbf{B}\mathbf{\Sigma}\mathbf{B}^\top + \mathbf{Q}.$$

Unfortunately, however, there is no closed-form solution for $\mathbf{\Sigma}$ when written in this form.

The *vec* operator

It turns out that we can use the *vec* operator to derive an explicit solution for $\mathbf{\Sigma}$. The *vec* operator converts an $i \times j$ matrix into an $(ij) \times 1$ column vector. For example, if

$$\mathbf{M} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix},$$

then

$$vec(\mathbf{M}) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}.$$

Solution

Thus, if $\mathbf{I}$ is an $n \times n$ identity matrix, and we define $\mathcal{I} = (\mathbf{I} \otimes \mathbf{I})$ and $\mathcal{B} = (\mathbf{B} \otimes \mathbf{B})$, then

$$vec(\mathbf{\Sigma}) = (\mathcal{I} - \mathcal{B})^{-1} vec(\mathbf{Q}).$$

Importance of species interactions to stability

Among the many ways of classifying stability, I am interested in the extent to which community interactions, relative to environmental forcing, contribute to the overall variance of the stationary distribution. In a stable system, any perturbation affecting one or more of the community members does not amplify as it moves throughout the community as a whole, such that the variances in log-densities over time would be driven almost entirely by random environmental variation. Given that same magnitude of random environmental variation, less stable systems are characterized by greater variances in the temporal dynamics of their constituents.

Here we will use determinants to measure the "volume" of a matrix. Looking back to the matrix form of the equation for the variance of the stationary distribution, we see that

$$\mathbf{\Sigma} = \mathbf{B}\mathbf{\Sigma}\mathbf{B}^\top + \mathbf{Q},$$

and hence

$$\mathbf{\Sigma} - \mathbf{Q} = \mathbf{B}\mathbf{\Sigma}\mathbf{B}^\top.$$

Therefore, the volume of the difference $\mathbf{\Sigma} - \mathbf{Q}$ provides a measure of how much species interactions contribute to the variance of the stationary distribution. Taking determinants of both sides, we get

$$\begin{aligned}
\det(\mathbf{\Sigma} - \mathbf{Q}) &= \det(\mathbf{B}\mathbf{\Sigma}\mathbf{B}^\top) \\
&= \det(\mathbf{B})\det(\mathbf{\Sigma})\det(\mathbf{B}^\top) \\
&= \det(\mathbf{B})\det(\mathbf{\Sigma})\det(\mathbf{B}).
\end{aligned}$$

The proportion $\pi_\mathbf{B}$ of the volume of $\mathbf{\Sigma}$ attributable to species interactions is then

$$\pi_\mathbf{B} = \frac{\det(\mathbf{\Sigma} - \mathbf{Q})}{\det(\mathbf{\Sigma})} = \det(\mathbf{B})^2.$$

Intra- versus inter-specific interactions

Now that we have a means for identifying the proportion $\pi_\mathbf{B}$ of the volume of $\mathbf{\Sigma}$ attributable to species interactions, we can go further and compare the relative contribution of intra- versus inter-specific interactions. To do so, we will decompose $\mathbf{B}$ into 2 matrices representing the two kinds of interactions, such that

$$\mathbf{B} = \mathbf{\Delta} + \mathbf{\Omega}.$$

The matrix $\mathbf{\Delta}$ has density-dependent effects for each community member along the diagonal and zeroes elsewhere. The matrix $\mathbf{\Omega}$ contains zeroes along the diagonal and (potentially) non-zero entries in the off-diagonals.

We can again use determinants to measure the relative contributions of intra- ($\mathbf{\Delta}$) versus inter-specific ($\mathbf{\Omega}$) interactions to the volume of $\mathbf{B}$, where

$$\begin{aligned}
\det(\mathbf{B}) &= \det(\mathbf{\Delta} + \mathbf{\Omega}) \\
&= \det(\mathbf{\Delta})\det(\mathbf{I} + \mathbf{\Omega}\mathbf{\Delta}^{-1}).
\end{aligned}$$

Therefore, the proportion $\pi_\mathbf{\Delta}$ of the volume of $\mathbf{B}$ attributable to density-dependent interactions is then

$$\pi_\mathbf{\Delta} = \frac{\det(\mathbf{\Delta})}{\det(\mathbf{B})} = \frac{1}{\det(\mathbf{I} + \mathbf{\Omega}\mathbf{\Delta}^{-1})},$$

where $\mathbf{I}$ is again an $n \times n$ identity matrix. The proportion of the volume of $\mathbf{B}$ attributable to competition, predator-prey interactions, etc is then simply $\pi_\mathbf{\Omega} = 1 - \pi_\mathbf{\Delta}$.

The effects of covariates

Often we would like to examine the potential effects of covariates on community dynamics. In that case, the MARSS(1) model becomes

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t,$$
$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{a} + \mathbf{v}_t,$$

where $\mathbf{C}$ is an $n \times p$ matrix of covariate effects, and $\mathbf{c}_t$ is a $p \times 1$ vector of covariates at time t. Although the variance of the stationary distrbution will be affected by any covariate effects, the method described above for estimating the relative effects of species interactions is unaffected. That is, if $\boldsymbol{\Xi}$ is the variance-covariance matrix of the covariates in $\mathbf{c}$, then analogous to above the variance of the stationary distrbution will be

$$\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top + \mathbf{C}\boldsymbol{\Xi}\mathbf{C}^\top + \mathbf{Q},$$
$$\boldsymbol{\Sigma} - \mathbf{C}\boldsymbol{\Xi}\mathbf{C}^\top - \mathbf{Q} = \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top,$$

and hence the proportion $\pi_\mathbf{B}$ of the volume of $\boldsymbol{\Sigma}$ attributable to species interactions is, as before, given by

$$\pi_\mathbf{B} = \frac{\det(\boldsymbol{\Sigma} - \mathbf{C}\boldsymbol{\Xi}\mathbf{C}^\top - \mathbf{Q})}{\det(\boldsymbol{\Sigma})} = \det(\mathbf{B})^2.$$

An example

As an example, I will fit a MARSS(1) model to some plankton data from Lake Washington (USA). These data were made publicly available by Dr. Daniel Schindler of the University of Washington, and they have been included as part of the `MARSS` package for **R**. In particular, I will focus on the time period from 1977 onward because *Daphnia* species were largely absent from the water column prior to then.

```
library(MARSS)
## load the data (2 datasets here)
data(lakeWAplankton)
raw <- as.data.frame(lakeWAplanktonRaw)
## keep only 1977-
dat <- subset(raw, Year>=1977)
## contents
colnames(dat)
```

```
##  [1] "Year"              "Month"
##  [3] "Temp"              "TP"
##  [5] "pH"                "Cryptomonas"
##  [7] "Diatoms"           "Greens"
##  [9] "Bluegreens"        "Unicells"
## [11] "Other.algae"       "Conochilus"
```

```
## [13] "Cyclops"               "Daphnia"
## [15] "Diaptomus"             "Epischura"
## [17] "Leptodora"             "Neomysis"
## [19] "Non.daphnid.cladocerans" "Non.colonial.rotifers"
```

```r
## phytoplankton names (omitting bluegreens)
phyto_names <- c("Cryptomonas", "Diatoms", "Greens", "Unicells", "Other.algae")
## zooplankton names (omitting Leptodora & Neomysis)
zoops_names <- c("Cyclops", "Diaptomus", "Epischura",   # copepods
                 "Daphnia", "Non.daphnid.cladocerans",  # cladocerans
                 "Conochilus", "Non.colonial.rotifers") # rotifers
all_names <- c(phyto_names,zoops_names)
## number of taxa
nP <- length(phyto_names)
nZ <- length(zoops_names)
nn <- nP + nZ
```

In addition to the biotic interactions, I will include the effects of water temperature on the plankton.

Now we need to set up the various vectors and matrices that define our MARSS(1) model. Here I will make the following assumptions:

1. independent and identically distributed observation errors for all of the taxa within each of the two plankton groups;
2. a block-diagonal form for $\mathbf{Q}$ wherein all of the taxa within a plankton group have the same variance and covariance of process errors;
3. missing covariates are imputed as averages of the values at $t - 1$ and $t + 1$; and
4. (possibly) different effects of covariates on each taxa.

The following code also makes use of various shorthand character description of vectors and matrics in `MARSS`.

```r
mod_list <- list()
## process model
mod_list$B <- "unconstrained"
mod_list$U <- "zero"
## include temperature effect
mod_list$C <- "unconstrained"
temp <- matrix(dat$Temp, nrow=1)
ii <- which(is.na(temp))
temp[ii] <- mean(temp[ii-1], temp[ii+1])
mod_list$c <- round(temp, 2)
## var-cov of process (environmental errors)
mod_list$Q <- matrix(list(0), nn, nn)
mod_list$Q[1:nP,1:nP] <- "g_phyto"
mod_list$Q[1:nZ+nP,1:nZ+nP] <- "g_zoops"
diag(mod_list$Q) <- c(rep("q_phyto", nP), rep("q_zoops", nZ))
## observation model
mod_list$Z <- "identity"
mod_list$A <- "scaling"
mod_list$D <- "zero"
mod_list$d <- "zero"
## observation var-cov
mod_list$R <- matrix(list(0), nn, nn)
```

```
diag(mod_list$R) <- c(rep("r_phyto", nP), rep("r_zoops", nZ))
## initial state
mod_list$tinitx <- 1
```

The model we are fitting is based upon log-densities, so we'll need to transform the data before proceeding. In this case, zeroes in the data set are indicative of sampling program and therefore I will convert them to NA.

```
yy <- dat[,all_names]
yy[yy==0] <- NA
## log-transform variates
## data must also be n x T, so transpose
yy <- t(log(yy))
## fit model
mod <- MARSS(yy, model=mod_list, control=list(maxit=2000))
```

```
## Stopped at iter=21 in MARSSkem at B update. denom is not invertible.
##  par, kf, states, iter, loglike are the last values before the error.
##  Try control$safe=TRUE which uses a slower but slightly more robust algorithm.
##  Use control$trace=1 to generate a more detailed error report. See user guide for insight.
```