Priorcovmatrix

Ignacio Alvarez-Castro

Center of Statistical Research (IESTA)
Universidad de la República, Uruguay.

LatinR2018
3rd - 5th September 2018. Universidad de Palermo, Argentina.

1. Introduction
2. Covariance matrix estimation
3. R packages with STAN code (simulation)
4. Visualize covariance matrix distribution

Introduction

Covariance matrix estimation

- Multivariate normal sampling models

- random-intercept, random-slope models

$$
\begin{aligned}
y_{ij} &= \beta_{0j} + \beta_{1j} x_{ij} + \beta_{2j} z_{ij} + \varepsilon_{ij} \\
\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \\ \beta_{2j} \end{pmatrix} &\sim N\left(\begin{pmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma \right), \quad \varepsilon_{ij} \sim N(0, \sigma^2)
\end{aligned}
$$

Bird counts on Superior forests

The Natural Resources Research Institute (University of Minnesota Duluth) carry out monitoring program for study regional population trends of forest birds.

Want to study population trend over time

$$E(y_{st}) = \beta_{0s} + \beta_{1s} t + \beta_{2t} t^2$$

Point count for :

- 19 years (1995 - 2013)
- 3 forest
- 73 bird species

Response y_{st}

- Total count in logs
- Average count accros sites
- Total count

Quadratic trend model

Hierarchical linear model with *IW* prior.

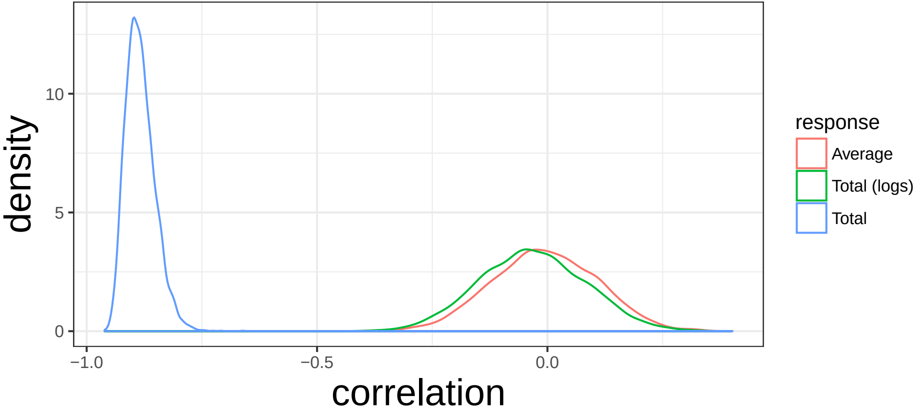
$$
\begin{aligned}
y_{st} &\sim N(\beta_{0s} + \beta_{1s} t + \beta_{2t} t^2, \sigma^2) \\
\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \\ \beta_{2j} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma \right) \\
\Sigma &\sim IW(d+1, I)
\end{aligned}
$$

Correlation among coefficient

$$\rho = \Sigma_{23} / \sqrt{\Sigma_{22} \Sigma_{33}}$$

Quadratic trend model

Hierarchical linear model with IW prior. Posterior $p(\rho|y)$



Multivariate normal model

Alvarez-Castro et al. (2014) compare alternative priors for Σ in this for multivariate normal data.

$$Y_i \sim N_d(0, \Sigma)$$

Alternative Σ priors

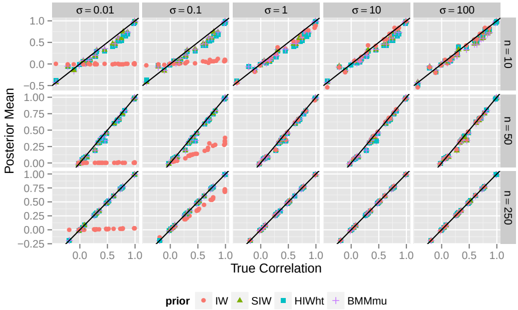
- Inverse Wishart: $IW(\nu, \Lambda)$: $p(\Sigma) \propto |\Sigma|^{-(\nu+d+1)/2} e^{-\frac{1}{2} tr(\Lambda \Sigma^{-1})}$
- Scaled Inverse Wishart
- Hierarchical inverse Wishart
- Separation strategy

Asses prior inpact on posterior inference:

- using simulations
- with the bird count data set (not shown)

Impact on posterior inference

Inference for ρ_{12}



When standard deviation is small, $\sigma = 0.01$ or $\sigma = 0.1$ the IW prior heavily shrinks the posterior correlation towards 0 even if the true correlation is close to 1.

Libraries with STAN code

STAN: general purpuse sofware for Bayesian inference

Set up package structure:

```
rstan_package_skeleton(path = 'PriorCovmatrix')
```

or

```
rstan_package_skeleton(path = 'PriorCovmatrix',
                       stan_files ='list of .stan files')
```

Libraries with STAN code

STAN: general purpuse sofware for Bayesian inference

Install package for first time

```
devtools::install(args = "--preclean")
```

or

```
devtools::install()
```

Priorcovmatrix: starting point

little funcionality so far

```
rIW(n, d, R)

rSS(n, k prior_cor = 'lkj', prior_sg ='ln',
    eta = k+1, R = diag(k), sigma_mu=0, sigma_sc=1)

covmat(x, xnames = NULL, colvar = 'dep')
```

Visualize a covariance matrix **distribution**

- not just one covariance matrix
- reduce to bivariate pieces (Tokuda et al., 2011)
- use scalar measures (Peña and Rodrıguez, 2003)

Parallel coordinate plot

- all matrix entries in one plot
- use color to look for patterns

Simulated example

Simulate n four-dimensional covariance matrices: $\Sigma \sim IW(d, R)$.

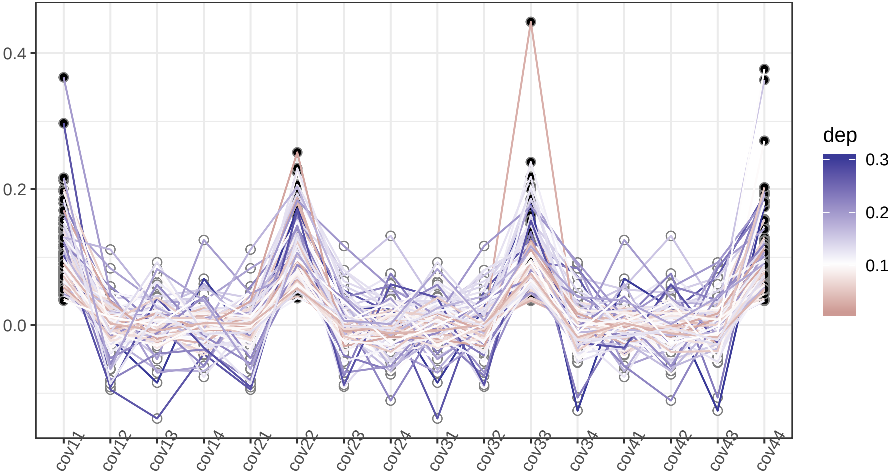
	parameter	values
sample size	n	100
degrees of freedom	d	5, 15
correlation	ρ	0, 0.8

Location matrix is constructed as:

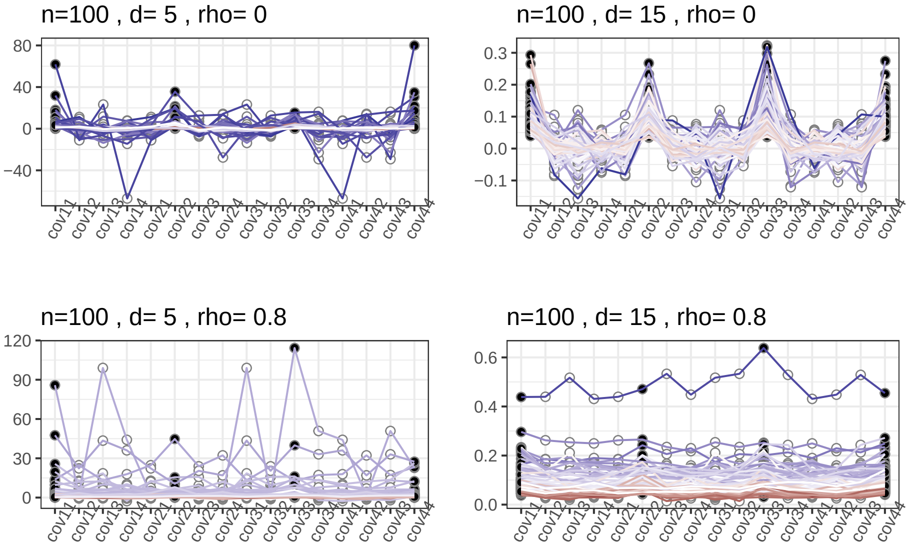
$$R = \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix}$$

Parallel coordinate plot: $\rho = 0 \;\; d = 15$

```
## Joining, by = "grp.var"
```



Parallel coordinate plot



Muchas Gracias !

- @nachalca
- nachalca@iesta.edu.uy
- https://github.com/nachalca/PriorCovmatrix

References

Alvarez-Castro, I., Niemi, J., and Simpson, M. (2014), "Bayesian Inference for a Covariance Matrix," in *Annual Conference on Applied Statistics in Agriculture*, New Prairie Press, pp. 71 – 82.

Peña, D. and Rodrıguez, J. (2003), "Descriptive measures of multivariate scatter and linear dependence," *Journal of Multivariate Analysis*, 85, 361–374.

Tokuda, T., Goodrich, B., Van Mechelen, I., and Gelman, A. (2011), "Visualizing Distributions of Covariance Matrices," \url{http://www.stat.columbia.edu/˜gelman/research/unpublished/Visualization.pdf}.