Bayesian inference for covariance matrix

Ignacio Alvarez

IESTA, Universidad de la Repblica, Uruguay

ISEC 2018

Introduction

Covariance matrix estimation

- Multivariate normal sampling models

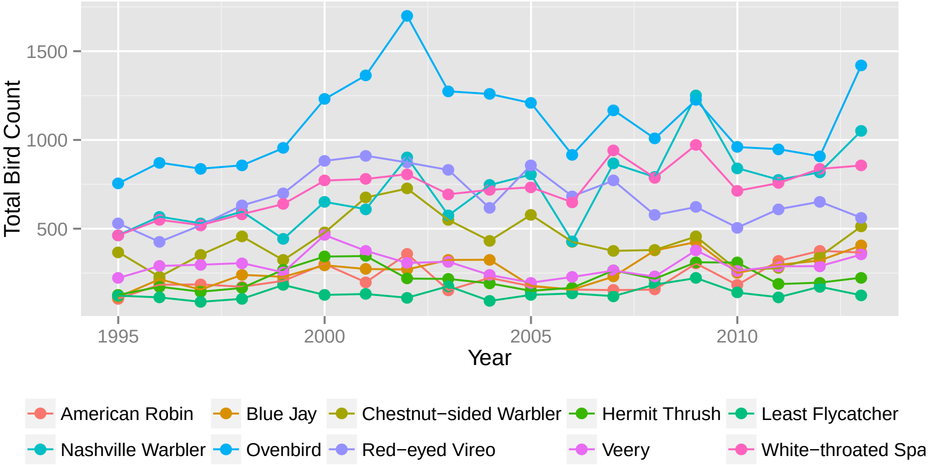
- random-intercept, random-slope models

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

- Ovaskainen et al. (2017) includes covariance matrix priors for comunity data hierarchical models

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.

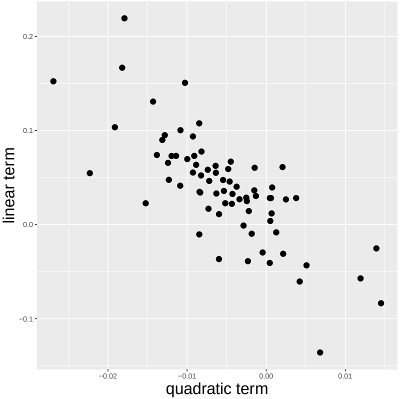


Quadratic trend model

- y_{st}: bird count for spacies s in year t.
- OLS regression model for each species

$$y_{st} \sim N(\beta_{0s} + \beta_{1s} t + \beta_{2t} t^2, \sigma^2)$$

- $\text{Corr}(\hat{\beta}_{1s}, \hat{\beta}_{2s}) = -.77$



Quadratic trend model

Hierarchical linear model with IW prior.

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

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

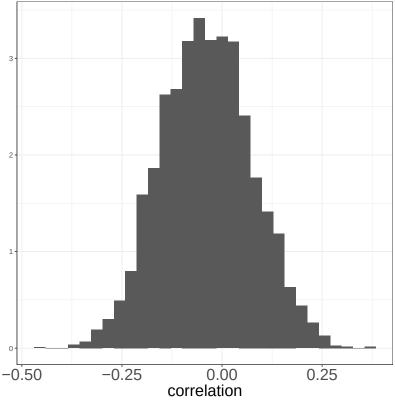
Quadratic trend model

Hierarchical linear model with *IW* prior.

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

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

How does $p(\rho|y)$ look like?



Multivariate normal model

Consider n observations from $Y_i \sim N_d(0, \Sigma)$ distribution.

Likelihood:

$$p(y|\mu, \Sigma) \propto |\Sigma|^{-n/2} e^{-\frac{1}{2}\sum_{i=1}^{n} y_i' \Sigma^{-1} y_i} = |\Sigma|^{-n/2} e^{-\frac{1}{2} tr(\Sigma^{-1} S_0)}$$

where $y_i \in R^d$ a realization of Y_i, $S_0 = \sum_{i=1}^{n} y_i y_i'$, individual entries of Σ are $\Sigma_{ii} = \sigma_i^2$ and $\Sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$.

Multivariate normal model

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

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

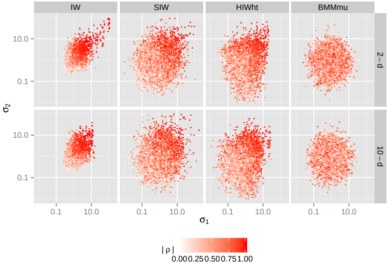
Alternative Σ priors

- 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

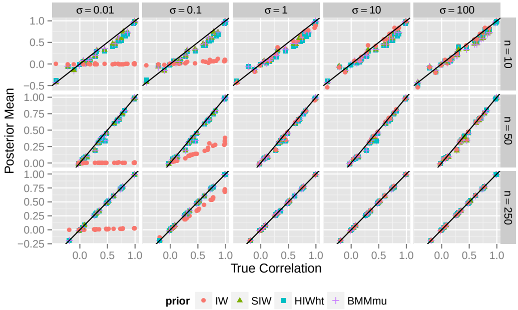
Samples from prior



Positive relationship among σ_1 and σ_2, also large $|\rho_{12}|$ values appear when the two variances are high.

Impact on posterior inference

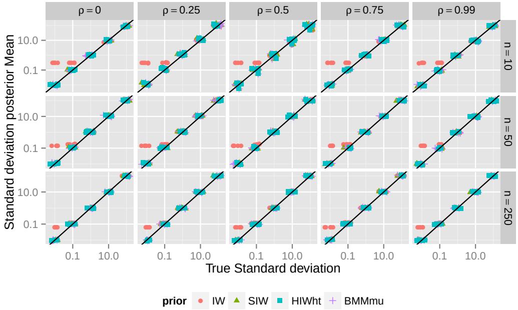
Inference for ρ_{12}



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.

Impact on posterior inference

Inference for σ_1



IW prior overestimate the standard deviation when its true value is very low

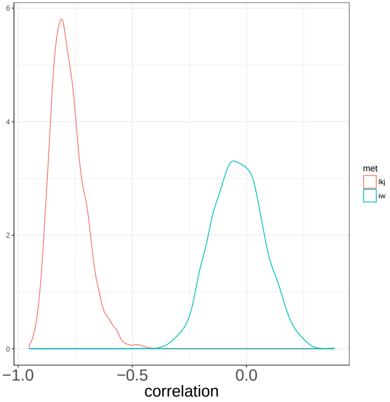
Separation strategy in linear model

$$
\begin{aligned}
y_{st} &\sim N(\beta_{0s} + \beta_{1s} t + \beta_{2t} t^2, \sigma^2) \\
\begin{pmatrix} \beta_{0s} \\ \beta_{1s} \\ \beta_{2s} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma \right) \\
\Sigma &= \mathrm{diag}(\lambda)\,\Omega\,\mathrm{diag}(\lambda) \\
\lambda_k &\sim Ca^+(0,1) \\
\Omega &\sim LKJ(3)
\end{aligned}
$$

Separation strategy in linear model

$$
\begin{aligned}
y_{st} &\sim N(\beta_{0s} + \beta_{1s} t + \beta_{2t} t^2, \sigma^2) \\
\begin{pmatrix} \beta_{0s} \\ \beta_{1s} \\ \beta_{2s} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma \right) \\
\Sigma &= \mathrm{diag}(\lambda)\, \Omega\, \mathrm{diag}(\lambda) \\
\lambda_k &\sim Ca^+(0,1) \\
\Omega &\sim LKJ(3)
\end{aligned}
$$

Posterior $p(\rho|y)$



Discussion

Prior choice (based on multivariate normal data results)

1. Separation strategy gives modeling flexibility and good inferences properties. (Barnard et al., 2000)
2. In Gibbs base samplers a prior which maintain conjugacy might be preferable (scaled inverse Wishart or hierarchical inverse Wishart)
3. If we are constraint to use IW, we may recommend to scale the data first.

Future steps

Different Model Hierarchical linear model context.

Different Priors Use LKJ prior for the correlation matrix.
Use other distributions for IW parameters ν and Λ.

References

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

Barnard, J., McCulloch, R., and Meng, X.-L. (2000), "Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage," *Statistica Sinica*, 10, 1281–1312.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., and Abrego, N. (2017), "How to make more out of community data? A conceptual framework and its implementation as models and software," *Ecology letters*, 20, 561–576.