

Following Makalic and Schmidt (2015), the horseshoe model is:

$$\begin{aligned} y \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2 &\sim \text{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n), \\ \beta_j \mid \lambda_j^2, \tau^2, \sigma^2 &\sim \text{N}(0, \lambda_j^2 \tau^2 \sigma^2), \\ \sigma^2 &\sim \sigma^{-2} d\sigma^2, \\ \lambda_j &\sim \text{C}^+(0, 1), \\ \tau &\sim \text{C}^+(0, 1) \end{aligned}$$

They re-parameterize for Gibbs sampling as:

$$\begin{aligned} y \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2 &\sim \text{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n), \\ \beta_j \mid \lambda_j^2, \tau^2, \sigma^2 &\sim \text{N}(0, \lambda_j^2 \tau^2 \sigma^2), \\ \sigma^2 &\sim \sigma^{-2} d\sigma^2, \\ \lambda_j^2 \mid \nu_j &\sim \text{IG}(1/2, 1/\nu_j), \\ \tau^2 \mid \xi &\sim \text{IG}(1/2, 1/\xi), \\ \nu_1, \dots, \nu_p, \xi &\mid \sim \text{IG}(1/2, 1). \end{aligned}$$

Where the inverse-Gamma distribution with pdf:

$$p(z \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{-\alpha-1} \exp\left(-\frac{\beta}{z}\right)$$

BSFG uses the following model for each column of \mathbf{Y}

$$y \mid \mathbf{F}, \boldsymbol{\lambda}_j, \sigma^2 \sim \text{N}(\mathbf{F}\boldsymbol{\lambda}_j, \sigma_j^2 \boldsymbol{\Sigma}_j).$$

A horseshoe model for $\boldsymbol{\lambda}_j$ could be:

$$\begin{aligned} \lambda_{kj} \mid \phi_{kj}^2, \tau_k^2, \sigma_j^2 &\sim \text{N}(0, \phi_{kj}^2 \tilde{\omega}_k^2 \sigma_j^2), \\ \sigma_j^2 &\sim \text{IG}(a, b), \\ \phi_{kj} &\sim \text{C}^+(0, 1), \\ \tilde{\omega}_k &\sim \text{C}^+(0, \tau_k^{-1/2}), \\ \tau_k &= \prod_{h=1}^k \delta_h, \\ \delta_h &\sim \text{Ga}(a_\delta, b_\delta), h \geq 1, \delta_1 = 1. \end{aligned}$$

The column-shrinkage is accomplished by ω_k , which is given a half-Cauchy prior with scale parameter stochastically decreasing controlled by the sequence δ_h which is stochastically increasing (could be set so that $\delta_h \geq 1$).

This model maintains the form of the horseshoe for each column of $\mathbf{\Lambda}$, but adds an increasingly strong prior on the number of non-zero entries.

This can be re-parameterized as:

$$\begin{aligned}
y \mid \mathbf{F}, \boldsymbol{\lambda}_j, \sigma_j^2 &\sim \text{N}(\mathbf{F}\boldsymbol{\lambda}_j, \sigma_j^2 \boldsymbol{\Sigma}_j), \\
\lambda_{kj} \mid \phi_{kj}^2, \tau_k^2, \sigma_j^2 &\sim \text{N}(0, \phi_{kj}^2 (\omega_k^2 \tau_k^{-1}) \sigma_j^2), \\
\sigma_j^2 &\sim \text{IG}(a, b), \\
\phi_{kj}^2 \mid \nu_{kj} &\sim \text{IG}(1/2, 1/\nu_{kj}), \\
\omega_k^2 \mid \xi_k &\sim \text{IG}(1/2, 1/\xi_k), \\
\nu_{kj}, \xi_k &\mid \sim \text{IG}(1/2, 1), \\
\tau_k &= \prod_{h=1}^k \delta_h, \\
\delta_h &\sim \text{Ga}(a_\delta, b_\delta), h \geq 1, \delta_1 = 1.
\end{aligned}$$

with $\tilde{\omega}_k = \omega_k^2 \tau_k^{-1}$.

I wonder if the half-Cauchy prior on $\tilde{\omega}_k$ is necessary. Another possibility would be to replace δ_1 with $\tilde{\omega}$ (ie fixed over the whole matrix $\mathbf{\Lambda}$, and then just add the τ_k column-penalty on top. I think I'll start with this.

$$\begin{aligned}
y \mid \mathbf{F}, \boldsymbol{\lambda}_j, \sigma_j^2 &\sim \text{N}(\mathbf{F}\boldsymbol{\lambda}_j, \sigma_j^2 \boldsymbol{\Sigma}_j), \\
\lambda_{kj} \mid \phi_{kj}^2, \tau_k^2, \sigma_j^2 &\sim \text{N}(0, \phi_{kj}^2 \omega^2 \tau_k^{-1} \sigma_j^2), \\
\sigma_j^2 &\sim \text{IG}(a, b), \\
\phi_{kj}^2 \mid \nu_{kj} &\sim \text{IG}(1/2, 1/\nu_{kj}), \\
\omega^2 \mid \xi &\sim \text{IG}(1/2, 1/\xi), \\
\nu_{kj}, \xi &\mid \sim \text{IG}(1/2, 1), \\
\tau_k &= \prod_{h=1}^k \delta_h, \\
\delta_h &\sim \text{Ga}(a_\delta, b_\delta), h \geq 1, \delta_1 = 1.
\end{aligned}$$