

We propose an “Sparse Infinite Factor Model” based on the horseshoe prior. The horseshoe prior is a popular choice for inducing approximate sparsity in the coefficients of a regression model. Bhattacharya and Dunson (2011) proposed an “Infinite factor model” by ordering the latent factors from most-to-least influential, and then modeling the decrease in factor importance from one factor to the next as a stochastic process. Using this strategy, they showed that as long as they included a sufficiently large number of factors are included, their model could automatically prioritize only the most important factors, allowing the insignificant ones to be truncated.

In their implementation of the “Infinite Factor Model,” Bhattacharya and Dunson (2011) used the Normal-InverseGamma prior on the coefficients of the loadings matrix  $\mathbf{\Lambda}$ , and used a sequence generated by the product of gamma-distributed to induce additional shrinkage on the coefficients of each column of the matrix. Here, we propose two alterations to this model: i) We use the horseshoe prior on the individual coefficients of  $\mathbf{\Lambda}$  to induce more complete shrinkage. ii) We use a half-Cauchy distribution with stochastically decreasing scale parameter to model the column-shrinkage, and parameterize this distribution based on prior belief in the proportion of non-zero entries in each column. In particular, we place a prior on the proportion of non-zero values in the first column of  $\mathbf{\Lambda}$ , and then a second prior on the change in odds of being non-zero for each coefficient in each subsequent column. Together, these modifications allow us to parameterize our model in terms of the effective number of non-zero coefficients in each factor.

Below, we show our new factor model and derive a Gibbs sampler for the parameters. Following Makalic and Schmidt (2015) and Piironen and Vehtari (2017), the horseshoe model for a one-dimensional trait  $\mathbf{y}$  with  $n$  observations and  $\mathbf{X}$  known is:

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

In our factor model, each column of the data matrix  $\mathbf{Y}$  is independent conditioned on the factor matrix  $\mathbf{F}$ , with distribution:

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

We propose the following model for the coefficients of  $\mathbf{\Lambda} = [\boldsymbol{\lambda}_j], j \in 1, \dots, p$ :

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

This model implements the local-global shrinkage property of the horseshoe prior on each column of  $\mathbf{\Lambda}$ , but with a stochastically decreasing global shrinkage parameter given by the synthetic parameter  $\tilde{\omega}^2 = \omega^2 \tau_k^{-1}$ . We can interpret this parameter as follows:

Piironen and Vehtari (2017) show that the effective number of parameters in a horseshoe model 1 can be approximated based on the shrinkage factors  $\kappa_j = (1 + n\sigma^{-2}\tau^2\lambda_j^2)$  as:

$$m_{\text{eff}} = \sum_{j=1}^D (1 - \kappa_j).$$

They calculate the mean of  $m_{\text{eff}}$  as:

$$\mathbb{E}[m_{\text{eff}} \mid \tau, \sigma] = \frac{\sigma^1 \tau \sqrt{n}}{1 + \sigma^{-1} \tau \sqrt{n}} D.$$

and suggest a prior for  $\tau$  with most of its mass near the value:

$$\tau_o = \frac{p_o}{1 - p_o} \frac{\sigma}{\sqrt{n}},$$

where  $p_o$  is the prior guess at the number of non-zero coefficients. The first term on the RHS of this equation can be interpreted as the odds of inclusion for any coefficient. A reasonable choice for this prior is thus  $\tau \sim \text{C}^+(0, \tau_o^2)$ . Using a similar argument, we can set:

$$\omega_o = \frac{p_o}{1 - p_o} \frac{\sigma}{\sqrt{n}},$$

where  $\frac{p_o}{1 - p_o}$  is the odds of inclusion for each coefficient *in the first column of  $\mathbf{\Lambda}$* , and  $\sigma$  is the prior guess for the typical residual standard deviation across traits. This follows because  $\tilde{\omega}_1^2 = \omega^2 \tau_1^{-1} = \omega^2$ .

Now, for the  $k$ th column of  $\mathbf{\Lambda}$ , we have

$$\tilde{\omega}_k^2 = \omega^2 \tau_k^{-1} = \omega^2 \prod_{h=1}^k \delta_h = \omega_{k-1}^2 \delta_k^{-1}.$$

If  $\tilde{\omega}_1$  is centered around  $\omega_o$ , then  $\tilde{\omega}_2$  is centered around:

$$\omega_o \frac{1}{\sqrt{\delta_2}} = \left( \frac{p_o}{1 - p_o} \frac{1}{\sqrt{\delta_2}} \right) \frac{\sigma}{\sqrt{n}}$$

Therefore, the expected odds of inclusion for a coefficient in the 2nd column of  $\mathbf{\Lambda}$  is reduced by a factor of  $\sqrt{\delta_2}$  relative to the first. The same result will hold for the odds of inclusion for each coefficient in each subsequent column, conditional on the odds of inclusion for each coefficient of the previous column. This provides a tool for calibrating the prior on  $\delta_k$  based on the expected rate of decline of the number of non-zero coefficients of each subsequent factor in the model. A prior that places most mass on  $\delta_k > 1$  will have stochastically increasing shrinkage on higher-order columns.

To implement this ‘‘Sparse Infinite Horseshoe Factor Model,’’ we re-parameterize it for Gibbs sampling following Makalic and Schmidt (2015) as:

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

All priors in this model are conjugate, allowing Gibbs updates.