

In the paper, *Handling Sparsity via the Horseshoe*, Carvalho and company present the Horseshoe prior. This prior was created for the cases in supervised-learning where there are sparse signals. The model can be represented as follows:

$$\begin{aligned} y_{ij} &\stackrel{\text{ind}}{\sim} N(x_i\theta_i, \sigma^2) \\ \theta_i &\stackrel{\text{ind}}{\sim} N(0, \lambda_i^2\tau_i^2) \\ \lambda_i &\sim Ca^+(0, 1) \\ \tau_i &\sim Ca^+(0, 1) \\ \pi(\sigma) &\propto 1/\sigma \end{aligned}$$

The setup for this model is fairly straight forward: set of coefficients,  $\theta$ , that explain the relationship between your response,  $y$ , and some explanatory variables,  $x$ . This prior was developed specifically when some  $\theta_i$  are really small and/or 0. There are some other ways to handle these situations with the most obvious being a point mass prior. The big upside to the horseshoe prior is that computationally it is easier to implement than the point mass prior, and the Horseshoe conforms very well to the gold standard of Bayesian Model-Averaging under heavy-tailed discrete-mixture model.

Graphics are included which really highlight the benefits of this prior compared to others with regard to its shrinkage capabilities. They plot the densities of shrinkage coefficients using a Laplace, Students-t, Strawderman-Berger, and Horseshoe priors. The Laplace prior is the only prior that can compete with the horseshoe prior for having good shrinkage capabilities. However, further in the paper, the shortcomings of the Laplace prior are presented. Most importantly, the Laplace places too much emphasis on the global shrinkage parameter,  $\tau$ , while not putting enough power into the local shrinkage parameters,  $\lambda_i$ . It is here where the Horseshoe prior really shines; in giving both these shrinkage parameters equal weight, it allows  $\tau$  to dictate the overall shrinkage while still allowing for individual  $\theta_i$  to get drawn in with  $\tau$  or to resist it.

I liked that in the paper they argued for a fully Bayesian approach to using their prior giving three reasons why using a fully bayes approach will help you avoid problems.