High-dimensional Bayes

Debdeep Pati
Florida State University

November 8, 2016

Horseshoe (Carvalho, Polson and Scott, 2010, Biometrika)

- $\theta_j \mid \lambda_j, \tau \sim N(0, \lambda_j^2 \tau^2), \lambda_j \sim \text{Ca}^+(0,1), \tau \sim \text{Ca}^+(0,1)$
- The horseshoe prior has two interesting features that make it particularly useful as a shrinkage prior for sparse problems.
- Its flat, Cauchy-like tails allow strong signals to remain large (that is, un-shrunk) a posteriori.
- Yet its infinitely tall spike at the origin provides severe shrinkage for the zero elements of θ.

Horseshoe for fixed τ

- Let $y_i = \theta_i + \epsilon_i, i = 1, \ldots, n, \epsilon_i \sim N(0, 1)$.
- Assume for now that $\tau = 1$, and define $\kappa_i = 1/(1 + \lambda_i^2)$.
- κ_i is a random shrinkage coefficient, and can be interpreted as the amount of weight that the posterior mean for θ_i places on 0 once the data y have been observed.

$$E(\theta_i \mid y_i, \lambda_i) = \frac{\lambda_i^2}{1 + \lambda_i^2} y_i + \frac{1}{1 + \lambda_i^2} 0 = (1 - \kappa_i) y_i$$

- Since $\kappa_i \in [0, 1]$, this is clearly finite, and so by Fubini's Theorem

$$\begin{aligned} E(\theta_i | y) &= \int_0^1 (1 - \kappa_i) y_i \pi(\kappa_i \mid y_i) d\kappa_i \\ &= (1 - E(\kappa_i \mid y_i)) y_i. \end{aligned}$$

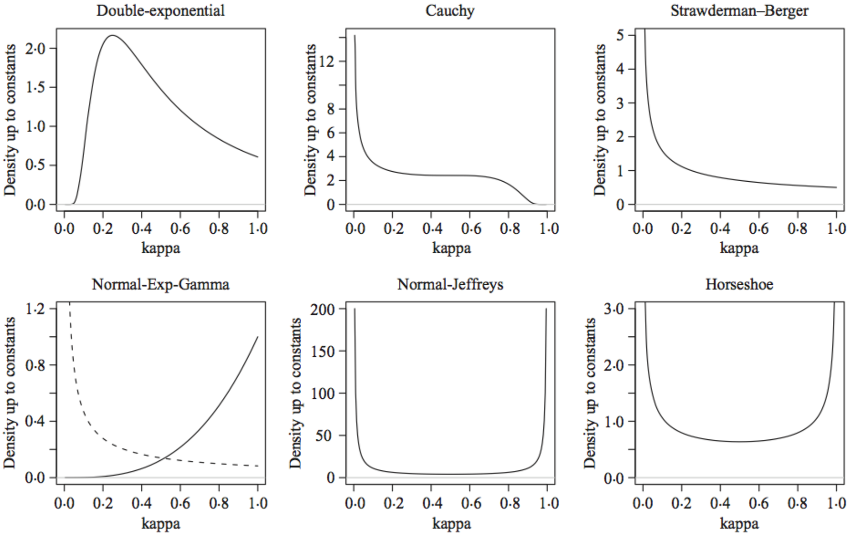
- If $\lambda_i \sim \mathrm{Ca}^+(0, 1)$, $\kappa_i \sim \mathrm{Beta}(1/2, 1/2)$.

κ_i for various priors

Table 1. *Priors for λ_i and κ_i associated with some common local shrinkage rules. For the normal-exponential-gamma prior, it is assumed that $d = 1$. Densities are given up to constants.*

Prior for θ_i	Density for λ_i	Density for κ_i
Double-exponential	$\lambda_i \exp(-\lambda_i^2/2)$	$\kappa_i^{-2} \exp\{-1/(2\kappa_i)\}$
Cauchy	$\lambda_i^{-2} \exp\{1/(2\lambda_i^2)\}$	$\kappa_i^{-1/2}(1-\kappa_i)^{-3/2} \exp[-\kappa_i/\{2/(1-\kappa_i)\}]$
Strawderman–Berger	$\lambda_i\,(1+\lambda_i^2)^{-3/2}$	$\kappa_i^{-1/2}$
Normal-exponential-gamma	$\lambda_i\,(1+\lambda_i^2)^{-(c+1)}$	κ_i^{c-1}
Normal-Jeffreys	λ_i^{-1}	$\kappa_i^{-1}(1-\kappa_i)^{-1}$
Horseshoe	$(1+\lambda_i^2)^{-1}$	$\kappa_i^{-1/2}(1-\kappa_i)^{-1/2}$

Distribution of κ_i for various priors



Strengths of the Horseshoe prior

- It is highly adaptive both to unknown sparsity and to unknown signal-to-noise ratio.
- It is robust to large, outlying signals.
- One can do variable selection by thresholding κ_i.

Horseshoe density

- The density $\pi_H(\theta_i)$ is not expressible in closed form, but very tight upper and lower bounds in terms of elementary functions are available.

Theorem

The Horseshoe prior satisfies the following:

1. $\lim_{\theta \to 0} \pi_H(\theta) = \infty$.
2. *For* $\theta \neq 0$,

$$\frac{K}{2} \log\left(1 + \frac{4}{\theta^2}\right) < \pi_H(\theta) < K \log\left(1 + \frac{2}{\theta^2}\right).$$

where $K = 1/\sqrt{2\pi^3}$.

Comparison with various priors

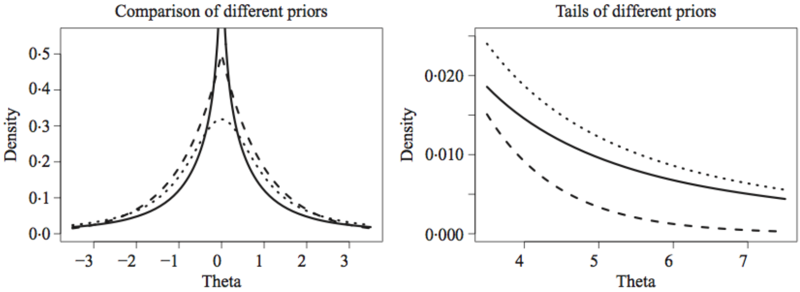


Fig. 1. Comparison of the horseshoe (solid), Cauchy (dotted) and double-exponential (dashed) densities.

Properties of Horseshoe

- It is symmetric about zero.
- It has heavy, Cauchy like tails that decay like θ_i^2.
- It has an infinitely tall spike at 0, in the sense that the density approaches infinity logarithmically fast asfrom either side.
- The priors flat tails allow each θ_i to be large if the data warrant such a conclusion, and yet its infinitely tall spike at zero means that the estimate can also be quite severely shrunk.