The horseshoe prior for nearly black vectors

S.L. van der Pas

Joint with B.J.K. Kleijn and A.W. van der Vaart

Bayes Club, April 4, 2014

Outline

The normal means model

We observe a vector $Y \in \mathbb{R}^n$ such that

$$Y_i = \theta_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where

- $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d.
- $\theta \in \ell_0[p_n]$, i.e. $\#(1 \leq i \leq n : \theta_i \neq 0) \leq p_n$.

Assumption: $p_n = o(n)$, $p_n \to \infty$ as $n \to \infty$.

Minimax risk

[Donoho, Johnstone, Hoch and Stern (1992)]

As $n, p_n \to \infty$:

$$\inf_{\widehat{\theta}\in\mathbb{R}^n} \sup_{\theta\in\ell_0[p_n]} \mathbb{E}_\theta\|\theta - \widehat{\theta}\|^2 = 2\sigma^2 p_n \log \frac{n}{p_n}(1 + o(1)),$$

where $\|.\|$ denotes the ℓ_2 norm.

Spike-and-slab

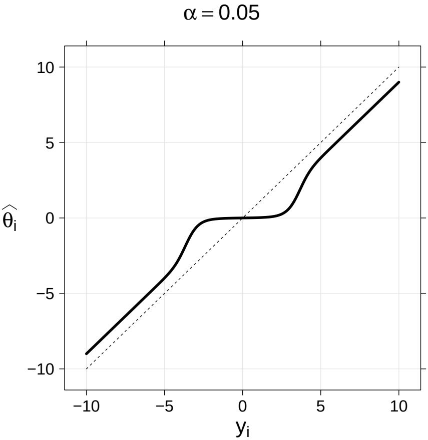
Mixture of a Dirac measure at zero and a univariate distribution g:

$$\theta_i \sim (1-\alpha)\delta_0 + \alpha g.$$

- Castillo and Van der Vaart (2012): good posterior concentration for several combinations of g and priors on α.
- Requires exploration of model space of size 2^n.

Johnstone and Silverman (2004)

- Empirical Bayes version of the spike-and-slab.
- Mixing weight obtained by marginal maximum likelihood.
- Coordinatewise posterior median or posterior mean.



Lasso

- MAP estimator, Laplace prior with common parameter on each θ_i.
- ℓ_2 risk of order $p_n \log n$ (Bickel, Ritov and Tsybakov (2009)).
- Full posterior distribution contracts at much slower rate than the mode (Castillo, Schmidt-Hieber and Van der Vaart).

The horseshoe prior

Introduced by Carvalho, Polson and Scott (2010). For $i = 1, \ldots, n$:

$$\theta_i | \lambda_i, \tau \sim \mathcal{N}\left(0, \sigma^2 \tau^2 \lambda_i^2\right), \quad \lambda_i \sim C^+(0, 1).$$

The horseshoe prior

Introduced by Carvalho, Polson and Scott (2010). For $i = 1, \ldots, n$:

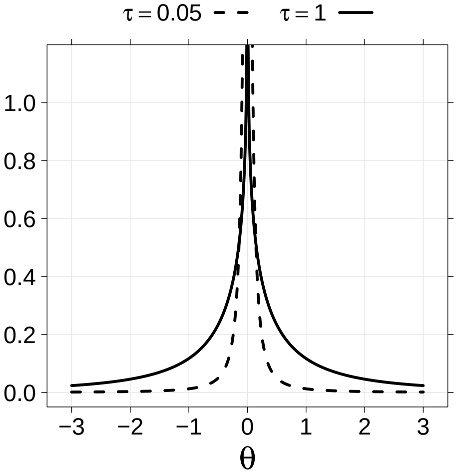
$$\theta_i | \lambda_i, \tau \sim \mathcal{N}\left(0, \sigma^2 \tau^2 \lambda_i^2\right), \quad \lambda_i \sim C^+(0, 1).$$

Density:

- increases logarithmically around zero;
- tails decay like θ_i^{-2}.

The horseshoe prior

$$\theta_i | \lambda_i, \tau \sim \mathcal{N}\left(0, \sigma^2 \tau^2 \lambda_i^2\right), \quad \lambda_i \sim C^+(0, 1).$$



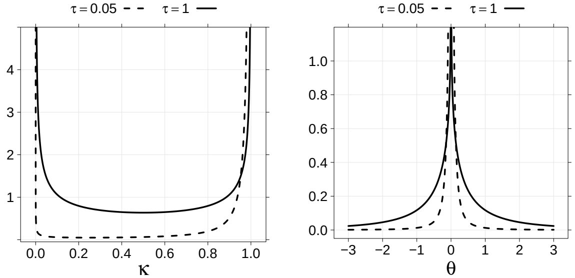
Origin of name

With $\kappa_i = \frac{1}{1+\tau^2\lambda_i^2}$:

$$\mathbb{E}[\theta_i | y_i, \tau] = (1 - \mathbb{E}[\kappa_i | y_i, \tau]) y_i.$$

Prior on κ_i:

$$p(\kappa_i) = \frac{\tau}{\pi} \frac{1}{1 - (1 - \tau^2)\kappa_i} (1 - \kappa_i)^{-\frac{1}{2}} \kappa_i^{-\frac{1}{2}}.$$



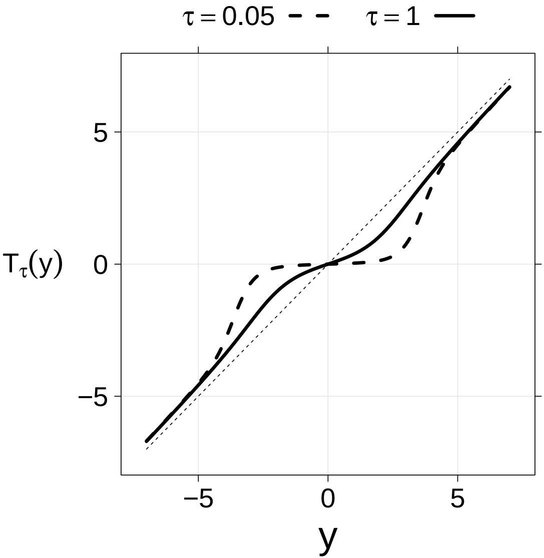
Posterior mean

By the identity $\mathbb{E}[\theta|y] = y + \frac{m'(y)}{m(y)}$, we find:

$$T_\tau(y_i) := \mathbb{E}[\theta|y_i, \tau] = y_i \left(1 - \frac{2\Phi_1\left(\frac{1}{2}, 1, \frac{5}{2}; \frac{y_i^2}{2\sigma^2}, 1 - \frac{1}{\tau^2}\right)}{3\Phi_1\left(\frac{1}{2}, 1, \frac{3}{2}; \frac{y_i^2}{2\sigma^2}, 1 - \frac{1}{\tau^2}\right)} \right),$$

where Φ_1 is the degenerate hypergeometric function of two variables.

Posterior mean



ℓ_2 risk

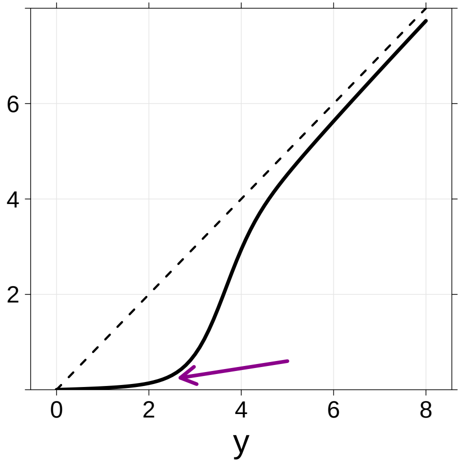
Theorem

Suppose $Y \sim \mathcal{N}(\theta_0, \sigma^2 I_n)$. Then the estimator $T_\tau(y)$ satisfies

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \| T_\tau(Y) - \theta_0 \|^2 \asymp p_n \log \frac{n}{p_n}$$

if $\tau = \frac{p_n}{n}$, as $n, p_n \to \infty$ and $p_n = o(n)$. The multiplicative constant before this rate is at most $4\sigma^2$

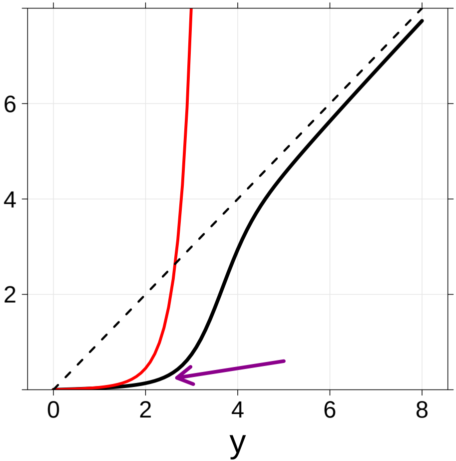
Bounds posterior mean



'Threshold' at $\sqrt{2\sigma^2 \log \frac{1}{\tau}}$

(cf. 'universal threshold' $\sqrt{2\sigma^2 \log n}$).

Bounds posterior mean

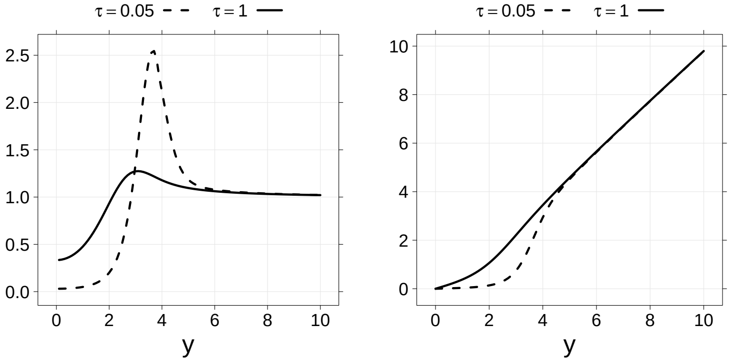


'Threshold' at $\sqrt{2\sigma^2 \log \frac{1}{\tau}}$

(cf. 'universal threshold' $\sqrt{2\sigma^2 \log n}$).

Posterior variance

$$\mathrm{Var}(\theta_i|y_i) = \frac{\sigma^2}{y_i} T_\tau(y_i) - (T_\tau(y_i) - y_i)^2 + y_i^2 \frac{8\Phi_1\left(\frac{1}{2}, 1, \frac{7}{2}; \frac{y_i^2}{2\sigma^2}, 1 - \frac{1}{\tau^2}\right)}{15\Phi_1\left(\frac{1}{2}, 1, \frac{3}{2}; \frac{y_i^2}{2\sigma^2}, 1 - \frac{1}{\tau^2}\right)}.$$



Upper bound on posterior concentration

Theorem

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \sum_{i=1}^{n} \mathrm{Var}(\theta_{0i} | Y_i) \lesssim p_n \log \frac{n}{p_n},$$

if $\tau = \frac{p_n}{n}$ and $p_n = o(n)$. The multiplicative constant before this rate is at most $2\sigma^2$.

Upper bound on posterior concentration

Theorem

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \sum_{i=1}^{n} \mathrm{Var}(\theta_{0i} | Y_i) \lesssim p_n \log \frac{n}{p_n},$$

if $\tau = \frac{p_n}{n}$ and $p_n = o(n)$. The multiplicative constant before this rate is at most $2\sigma^2$.

Theorem

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \Pi \left(\theta : \|\theta - T_\tau(Y)\|^2 > M_n p_n \log \frac{n}{p_n} \middle| Y \right) \to 0,$$

for every $M_n \to \infty$ as $n \to \infty$.

Choice of τ

- The risk result holds for $\tau = \left(\frac{p_n}{n}\right)^\alpha$, $\alpha \geq 1$. The constant is at most $4\alpha\sigma^2$.
- The posterior contraction result holds for $\tau = \left(\frac{p_n}{n}\right)^\alpha$, $\alpha \geq 1$. The constant is at most $2\alpha\sigma^2$.

Lower bound on posterior variance

Theorem
Suppose $Y \sim \mathcal{N}(\theta_0, \sigma^2 I_n)$, $\theta_0 \in \ell_0[p_n]$. Then:

$$\mathbb{E}_{\theta_0} \sum_{i=1}^{n} \mathrm{Var}(\theta_{0i} | Y_i) \gtrsim \frac{p_n^\alpha}{n^{\alpha-1}} \sqrt{\log \frac{n}{p_n}}$$

if $\tau = \left(\frac{p_n}{n}\right)^\alpha$, $\alpha > 0$, and $p_n = o(n)$, as $n \to \infty$.

Hence, $\tau = \frac{p_n}{n}$ seems to be optimal.

Conditions on $\widehat{\tau}$

Theorem

If $\widehat{\tau} \in (0,1)$ satisfies:

1. $\mathbb{P}_\theta(\widehat{\tau} > c\frac{p_n}{n}) \lesssim \frac{p_n}{n}$ for a constant $c \geq 1$ such that $\frac{p_n}{n} \leq \frac{1}{c}$;
2. There exists a function $g : \mathbb{N} \times \mathbb{N} \to (0,1)$ such that $\widehat{\tau} \geq g(n, p_n)$ with probability one and $-\log(g(n,p_n))\mathbb{P}_\theta\left(\widehat{\tau} \leq \frac{p_n}{n}\right) \lesssim \log\frac{n}{p_n}$,

then:

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \| T_{\widehat{\tau}}(Y) - \theta_0 \|^2 \asymp p_n \log \frac{n}{p_n}$$

as $n, p_n \to \infty$ and $p_n = o(n)$. If only the first condition can be verified for an estimator $\widehat{\tau}$, then $\sup\{\frac{1}{n}, \widehat{\tau}\}$ will have an ℓ_2 risk of at most order $p_n \log n$.

Conditions on $\widehat{\tau}$

Theorem

If $\widehat{\tau} \in (0,1)$ satisfies:

1. $\mathbb{P}_\theta(\widehat{\tau} > c\frac{p_n}{n}) \lesssim \frac{p_n}{n}$ for a constant $c \geq 1$ such that $\frac{p_n}{n} \leq \frac{1}{c}$;
2. There exists a function $g : \mathbb{N} \times \mathbb{N} \to (0,1)$ such that $\widehat{\tau} \geq g(n, p_n)$ with probability one and $-\log(g(n,p_n))\mathbb{P}_\theta\left(\widehat{\tau} \leq \frac{p_n}{n}\right) \lesssim \log \frac{n}{p_n}$,

then:

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \| T_{\widehat{\tau}}(Y) - \theta_0 \|^2 \asymp p_n \log \frac{n}{p_n}$$

as $n, p_n \to \infty$ and $p_n = o(n)$. If only the first condition can be verified for an estimator $\widehat{\tau}$, then $\sup\{\frac{1}{n}, \widehat{\tau}\}$ will have an ℓ_2 risk of at most order $p_n \log n$.

The factor $\log \frac{n}{p_n}$

- Not much to be gained if θ_0 is very sparse.
- If $p_n = n^\alpha$, $\alpha \in (0, 1)$, then $p_n \log \frac{n}{p_n} = (1 - \alpha) p_n \log n$.

Example

$$\widehat{\tau} = \frac{\#\{|y_i| \geq \sqrt{c_1 \sigma^2 \log n}, i = 1, \ldots, n\}}{c_2 n}$$

- $c_1 > 2, c_2 > 1$: $p_n \to \infty$.
- $c_1 = 2, c_2 > 1$: $p_n \gtrsim \log n$.

Conclusions

- A one-component prior can yield the $p_n \log \frac{n}{p_n}$ rate.
- Horseshoe prior has promising posterior contraction properties.
- Rate $p_n \log n$ easy to get in practice.