The horseshoe estimator: posterior concentration around nearly black vectors

S.L. van der Pas

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

ERCIM 2014

Outline

The normal means model

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

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

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

Goal

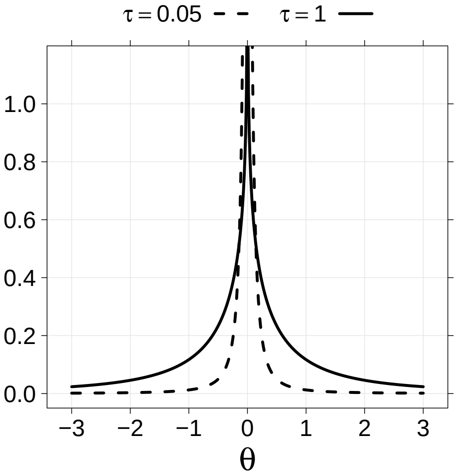
Assumption: there is some true θ_0 generating the data.

Goal: recovery and uncertainty quantifcation.

The horseshoe prior

Introduced by Carvalho, Polson and Scott (2010).

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



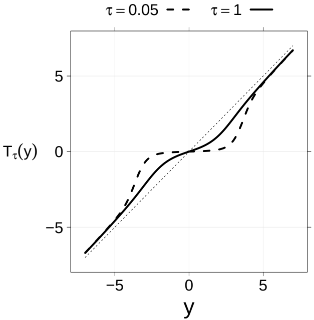
Posterior mean

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

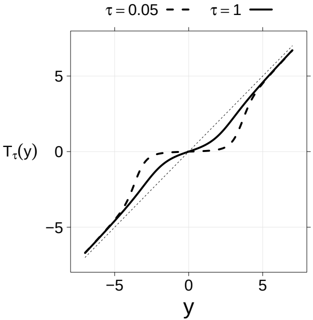
$$T_\tau(y_i) := \mathbb{E}[\theta|y_i,\tau] = y_i \left(1 - \frac{2\Phi_1\left(\frac{1}{2}, 1, \frac{5}{2}; \frac{y_i^2}{2\sigma^2}, 1 - \frac{1}{\tau^2}\right)}{3\Phi_1\left(\frac{1}{2}, 1, \frac{3}{2}; \frac{y_i^2}{2\sigma^2}, 1 - \frac{1}{\tau^2}\right)} \right),$$

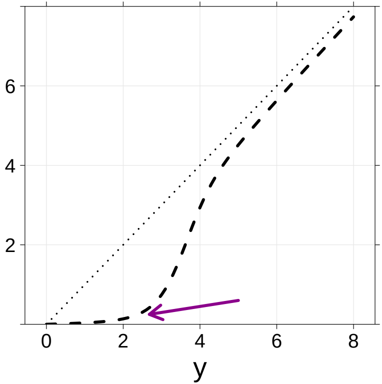
where Φ_1 is the degenerate hypergeometric function of two variables.

Posterior mean - threshold effect



Posterior mean - threshold effect





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

ℓ_2 risk

Theorem
Suppose $Y \sim \mathcal{N}(\theta_0, \sigma^2 I_n)$. Then, if $\tau = \frac{p_n}{n}$, as $n, p_n \to \infty$ and $p_n = o(n)$:

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

The multiplicative constant before this rate is at most $4\sigma^2$

Remark: the theorem holds for other choices of τ as well.

Posterior concentration

Theorem
With $\tau = \frac{p_n}{n}$:

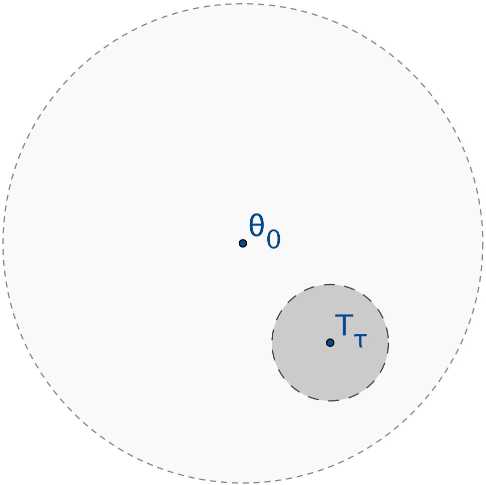
$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \Pi_\tau \left(\theta : \|\theta - \theta_0\|^2 > M_n p_n \log \frac{n}{p_n} \,\middle|\, Y \right) \to 0,$$

and

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

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

Mismatch?



Posterior variance

Theorem

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

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

if $\tau = \frac{p_n}{n}\sqrt{\log \frac{n}{p_n}}$, and $p_n = o(n)$, as $n \to \infty$.

Posterior variance

Theorem

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

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

if $\tau = \frac{p_n}{n}\sqrt{\log \frac{n}{p_n}}$, and $p_n = o(n)$, as $n \to \infty$.

For smaller values of τ, there are θ_0 for which $\mathbb{E}_{\theta_0}\| T_\tau(Y) - \theta_0 \|^2$ is of larger order than $\mathbb{E}_{\theta_0} \sum_{i=1}^{n} \mathrm{Var}(\theta_{0i} \mid Y_i)$.

Empirical Bayes - conditions on $\widehat{\tau}$

Theorem

If $\widehat{\tau} \in (0,1)$ satisfies, for $\tau = \frac{p_n}{n}$ or $\tau = \frac{p_n}{n}\sqrt{\log(n/p_n)}$:

1. $\widehat{\tau}$ does not overestimate τ by too much.

   $\mathbb{P}_\theta(\widehat{\tau} > c\tau) \lesssim \frac{p_n}{n}$ for a constant $c \geq 1$ such that $\tau \leq 1/c$

2. $\widehat{\tau}$ does underestimate τ by too much.

   $\exists\, g : \mathbb{N} \times \mathbb{N} \to (0,1)$ such that $\widehat{\tau} \geq g(n, p_n)$ w.p. 1, and

   $-\log(g(n,p_n))\mathbb{P}_\theta\left(\widehat{\tau} \leq \tau\right) \lesssim \log \frac{n}{p_n}$,

then:

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

Empirical Bayes - conditions on $\widehat{\tau}$

Theorem

If $\widehat{\tau} \in (0,1)$ satisfies, for $\tau = \frac{p_n}{n}$ or $\tau = \frac{p_n}{n}\sqrt{\log(n/p_n)}$:

1. $\widehat{\tau}$ does not overestimate τ by too much.

   $\mathbb{P}_\theta(\widehat{\tau} > c\tau) \lesssim \frac{p_n}{n}$ for a constant $c \geq 1$ such that $\tau \leq 1/c$

2. $\widehat{\tau}$ does underestimate τ by too much.

   $\exists\, g : \mathbb{N} \times \mathbb{N} \to (0,1)$ such that $\widehat{\tau} \geq g(n, p_n)$ w.p. 1, and

   $-\log(g(n,p_n))\mathbb{P}_\theta\left(\widehat{\tau} \leq \tau\right) \lesssim \log \frac{n}{p_n}$,

then:

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

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 $p_n \log n$.

Example

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

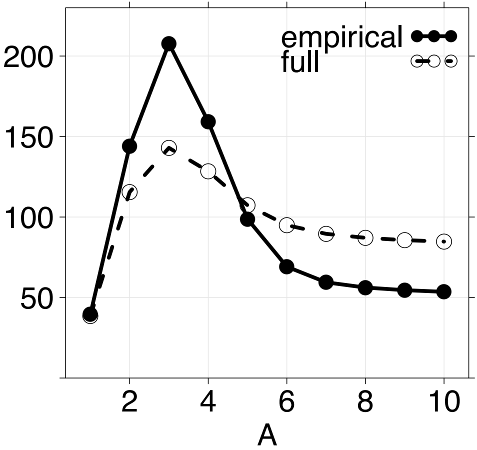
- Satisfies first condition if $c_1 > 2, c_2 > 1$: $p_n \to \infty$.

Simulation - full Bayes vs empirical Bayes

$$Y \sim \mathcal{N}(\theta, I_{400}), \quad \theta = (\underbrace{A, \dots, A}_{40 \text{ times}}, \underbrace{0, \dots, 0}_{360 \text{ times}}), \quad A \in \{1, \dots, 10\}.$$

Simulation - full Bayes vs empirical Bayes

$$Y \sim \mathcal{N}(\theta, I_{400}), \quad \theta = (\underbrace{A, \ldots, A}_{40 \text{ times}}, \underbrace{0, \ldots, 0}_{360 \text{ times}}), \quad A \in \{1, \ldots, 10\}.$$



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.