How many needles in the haystack?
Adaptive inference and uncertainty quantification for the horseshoe.

S.L. van der Pas, B. Szabó and A.W. van der Vaart

Leiden University, The Netherlands

JSM, August 1st, 2016
Conclusion

The horseshoe posterior distribution is suitable for uncertainty quantification, even when the number of signals is unknown.

This is true for

- empirical Bayes;
- hierarchical Bayes

under some conditions, with some exceptions.

For empirical Bayes, we recommend the MMLE.
The sparse normal means problem

Nearly black vector $\theta \in \ell_0[p]$.

At most $p$ nonzeros (signals).

Observe

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

where $\varepsilon_i \sim \mathcal{N}(0, 1)$, i.i.d.

Assume: $p \to \infty$, $p/n \to 0$ as $n \to \infty$. 

$$\theta = \begin{pmatrix}
0 \\
0 \\
\vdots \\
0 \\
\text{nonzero}
\end{pmatrix}$$
The horseshoe

Introduced by Carvalho, Polson and Scott (2010).

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

- Empirical Bayes: plug in estimate $\hat{\tau}_n$ for $\tau$.
- Hierarchical Bayes: $\tau \sim \pi_n$. 

![Graph showing two distributions with labels $\tau = 0.05$ and $\tau = 1$.]
The horseshoe works really well

- **Great performance in many simulation studies.** [e.g. Carvalho, Polson, Scott (2009, 2010), Polson, Scott (2010, 2012), Armagan, Dunson, Lee (2013), Bhattacharya, Pati, Pillai, Dunson (2015)]

- **Posterior mean achieves minimax rate.** [vdP, Kleijn, van der Vaart (2014)]

- **Posterior contracts** at the minimax rate. [vdP, Kleijn, van der Vaart (2014)]

**Correct choice of** $\tau$ **is essential**

- $\tau$ can be at most of order $\frac{p}{n} \sqrt{\log(n/p)}$. 
Adaptivity

The number of signals $p$ is **not** assumed to be known.
Contributions of this work

1. Characterize behaviour of the maximum marginal likelihood estimator (MMLE);

2. Establish contraction rates for hierarchical and empirical Bayes;

3. Study capability of the posterior distribution for uncertainty quantification (balls and intervals) for hierarchical and empirical Bayes.
The MMLE

\[ \hat{\tau}_M = \arg \max_{\tau \in \left[ \frac{1}{n}, 1 \right]} \prod_{i=1}^{n} \int_{-\infty}^{\infty} \varphi(y_i - \theta) g_{\tau}(\theta) d\theta, \]

where

- \( \varphi(\cdot) \) is the standard normal density;
- \( g_{\tau}(\cdot) \) is the marginal prior density for \( \theta_i \).

\[ g_{\tau}(\theta) = \int_0^{\infty} \varphi \left( \frac{\theta}{\lambda \tau} \right) \frac{1}{\lambda \tau} \frac{2}{\pi(1+\lambda^2)} d\lambda. \]

**Why \( \tau \in \left[ \frac{1}{n}, 1 \right] \)?**

- Interpretation of \( \tau \).
- Computational.
One-dimensional optimization

\( n = 100, p = 1 \)

\( n = 100, p = 5 \)

\( n = 100, p = 15 \)

\( n = 100, p = 40 \)
Main contribution about MMLE

The MMLE satisfies all of our conditions.

- posterior contraction at the (near) minimax-rate;
- honest and adaptive coverage.
For given positive constants $k_S, k_M, k_L$ and sequence $f_n \uparrow \infty$, define three regions:

$$\mathcal{S} := \{ i : |\theta_{0,i}| \leq k_S/n \},$$
$$\mathcal{M} := \{ i : f_n \tau_n \leq |\theta_{0,i}| \leq k_M \sqrt{2 \log(1/\tau_n)} \},$$
$$\mathcal{L} := \{ i : k_L \sqrt{2 \log n} \leq |\theta_{0,i}| \}.$$

$$\tau_n = (p/n) \sqrt{\log(n/p)}.$$
Credible intervals and posterior contraction

Under some conditions:

**Theorem [adaptive coverage, empirical or hierarchical Bayes]**

- The fraction of parameters in $S$ and $L$ that is contained in the credible sets **converges to 1**, with probability one.
- For every parameter in $M$, the probability that it is **not covered** converges to 1.

Hierarchical Bayes requires $p \geq C \log n$ for sufficiently large $C$.

**Theorem [adaptive contraction, empirical or hierarchical Bayes]**

The posterior contracts at the **near-minimax rate** $\sqrt{p \log n}$. 
Write $\tau_n = (p/n)\sqrt{\log(n/p)}$. Let $\pi_n$ be the prior density on $\tau$.

**Empirical Bayes condition:**
There exists a constant $C > 0$ such that $\hat{\tau}_n \in [1/n, C\tau_n]$, with $P_{\theta_0}$-probability tending to one, uniformly in $\theta_0 \in \ell_0[p]$. 

**Hierarchical Bayes conditions:**
- $\pi_n$ is supported inside $[1/n, 1]$.
- $\int t_n^2 \pi_n(\tau) d\tau \gtrsim e^{-cp}$, for some large enough $c$, where $t_n = C\tau_n$. 

Weaker version of Condition HB-2 leads to rate of $\sqrt{p \log n}$.
Write $\tau_n = (p/n)\sqrt{\log(n/p)}$. Let $\pi_n$ be the prior density on $\tau$.

**Empirical Bayes condition:**
There exists a constant $C > 0$ such that $\hat{\tau}_n \in [1/n, C\tau_n]$, with $P_{\theta_0}$-probability tending to one, uniformly in $\theta_0 \in \ell_0[p]$.

**Hierarchical Bayes conditions:**
- $\pi_n$ is supported inside $[\frac{1}{n}, 1]$.  
- $\int_{t_n/2}^{t_n} \pi_n(\tau) d\tau \gtrsim e^{-cp}$, for some large enough $c$, where $t_n = C_u\pi^{3/2}\tau_n$.

Weaker version of Condition HB-2 leads to rate of $\sqrt{p} \log n$. 
Credible balls - we cannot have it all

Honest uncertainty quantification is irreconcilable with adaptation to sparsity [Nickl, van de Geer (2013)].

Honesty of confidence set $C_n(Y^n)$ relative to $\Theta \subset \mathbb{R}^n$:

$$\liminf_{n \to \infty} \inf_{\theta_0 \in \Theta} P_{\theta_0}(\theta_0 \in C_n(Y^n)) \geq 1 - \alpha.$$  

Adaptation to a partition $\Theta = \bigcup_{p \in P} \Theta_p$. For every $p \in P$:

$$\liminf_{n \to \infty} \inf_{\theta_0 \in \Theta_p} P_{\theta_0} (\text{diam}(C_n(Y^n)) \leq r_{n,p}) = 1,$$

with $r_{n,p}$ the (near) minimax rate relative to $\Theta_p$.

Adaptive results for credible balls are restricted to vectors satisfying the excessive-bias restriction, following Belitser and Nurushev (2015).
Self-similarity

A vector $\theta_0 \in \ell_0[p]$ is \textit{self-similar} if

$$\# \left( i : |\theta_0,i| \geq A \sqrt{2 \log(n/p)} \right) \geq \frac{p}{C}$$

for constants $A, C > 1$.

\textbf{Stronger} than the excessive-bias restriction.
The excessive-bias restriction

A vector $\theta_0 \in \ell_0[p]$ satisfies the excessive-bias restriction if there exists an integer $q \geq 1$ with

1. $\# \left( i : |\theta_0,i| \geq A\sqrt{2 \log(n/q)} \right) \geq \frac{q}{C_1}$;

2. $\sum_{i : |\theta_0,i| < A\sqrt{2 \log(n/q)}} \theta^2_{0,i} \leq C_2 q \log(n/q)$.

for constants $A > 1$ and $C_1, C_2 > 0$. 

Notation

▶ $\Theta[p]$ denotes the set of all such vectors $\theta_0$, for fixed constants $A$, $C_1$, $C_2$.

▶ $\tilde{p}(\theta_0)$ denotes $\# \left( i : |\theta_0,i| \geq A\sqrt{2 \log(n/q)} \right)$, for the smallest possible $q$. 

The excessive-bias restriction

A vector $\theta_0 \in \ell_0[p]$ satisfies the excessive-bias restriction if there exists an integer $q \geq 1$ with

1. $\# \left( i : |\theta_0,i| \geq A\sqrt{2 \log(n/q)} \right) \geq \frac{q}{C_1}$;

2. $\sum_{i : |\theta_0,i| < A\sqrt{2 \log(n/q)}} \theta_{0,i}^2 \leq C_2 q \log(n/q)$.

for constants $A > 1$ and $C_1, C_2 > 0$.

Notation

- $\Theta[p]$ denotes the set of all such vectors $\theta_0$, for fixed constants $A, C_1, C_2$.

- $\tilde{p} = \tilde{p}(\theta_0)$ denotes $\# \left( i : |\theta_0,i| \geq A\sqrt{2 \log(n/q)} \right)$, for the smallest possible $q$. 
Credible balls - results

**Theorem** [adaptive coverage, empirical or hierarchical Bayes]
The credible balls of the horseshoe are honest and have rate adaptive size, with \( r_{n,p} = \sqrt{\tilde{p}} \log(n/\tilde{p}) \), uniformly in \( \theta_0 \in \Theta[p] \) and \( \tilde{p}(\theta_0) \geq \tilde{p}_n \) for given \( \tilde{p}_n \rightarrow \infty \).

**Empirical Bayes condition:**
The estimator \( \hat{\tau}_n \) satisfies, for some constant \( C > 1 \):

\[
\inf_{\theta_0 \in \Theta[p]} P_{\theta_0} \left( C^{-1} \tau_n(\tilde{p}) \leq \hat{\tau}_n \leq C \tau_n(\tilde{p}) \right) \rightarrow 1.
\]

\( \tau_n(\tilde{p}) = (\tilde{p}/n) \sqrt{\log(n/\tilde{p})} \).

**Hierarchical Bayes conditions:**
- \( \pi_n \) is supported on \([1/n, 1]\) and bounded away from zero.
- \( \tilde{p} \geq C \log n \) for sufficiently large \( C \).
Simulation study take-home messages

▶ Behaviour of $\tau$ is key. Larger values lead to larger intervals and better coverage but possibly worse estimates.

▶ Empirical Bayes with the MMLE and hierarchical Bayes with truncated Cauchy perform best, and closely mimic each other.

Included in comparison: EB with ‘simple’ estimator, HB with (non-truncated) Cauchy.

▶ Theoretical tool for analyzing hierarchical Bayes.
▶ Practical benefit.
Conclusion

Credible balls and intervals based on the horseshoe prior have

- good coverage;
- optimal size,

adapting to the sparsity level, both in the empirical and in the hierarchical Bayes setting, under some conditions.

For empirical Bayes, we recommend the MMLE.