Asymptotic analysis of Bayesian methods for sparse regression

Aad van der Vaart

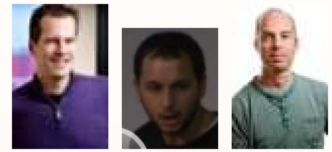
Universiteit Leiden

O'Bayes, December 2013

Contents

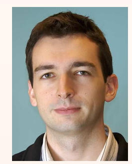
Co-authors

RNA sequencing



Mark van de Wiel, Gwenael Leday, Wessel van Wieringen

Sequence model & Regression



Ismael Castillo

Regression

Johannes Schmidt-Hieber

Horsehoe

Stéphanie van der Pas, Bas Kleijn

Sparsity

Bayesian sparsity

A sparse model has many parameters, but most of them are thought to be (nearly) zero.

Bayesian sparsity

A sparse model has many parameters, but most of them are thought to be (nearly) zero.

We express this in the prior,

and apply the standard (full or empirical) Bayesian machine.

Bayesian sparsity

A sparse model has many parameters, but most of them are thought to be (nearly) zero.

We express this in the prior,

and apply the standard (full or empirical) Bayesian machine.

In the remainder of this talk consider two simple models:

- Sequence model. Data $Y \sim N_n(\theta, I)$.
- Regression model. Data $Y \sim N_n(X_{n\times p}\theta, I)$.

In both cases θ is known to have many (almost) zero coordinates, and p and n are large.

Bayesian sparsity — RNA sequencing

$Y_{i,j}$: RNA expression count of tag $i = 1, \ldots, p$ in tissue $j = 1, \ldots, n$,
x_j: covariates of tissue j.

$Y_{i,j} \sim$ (zero-inflated) *negative binomial*, with

$$\mathrm{E}Y_{i,j} = e^{\alpha_i + \beta_i x_j}, \qquad \mathrm{var}\, Y_{i,j} = \mathrm{E}Y_{i,j}\big(1 + \mathrm{E}Y_{i,j} e^{-\phi_i}\big).$$

Many tags i are thought to be unrelated to x_j: $\beta_i = 0$ for most i.

[Smyth & Robinson et al., van der Wiel & vdV et al., 12]

Model selection prior

Constructive definition of prior Π for $\theta \in \mathbb{R}^p$:

(1) Choose s from prior π on $\{0, 1, 2, \ldots, p\}$.

(2) Choose $S \subset \{0, 1, \ldots, p\}$ of size $|S| = s$ at random.

(3) Choose $\theta_S = (\theta_i : i \in S)$ from density g_S on $\mathbb{R}^S$ and set $\theta_{S^c} = 0$.

Model selection prior

Constructive definition of prior Π for $\theta \in \mathbb{R}^p$:

(1) Choose s from prior π on $\{0, 1, 2, \ldots, p\}$.

(2) Choose $S \subset \{0, 1, \ldots, p\}$ of size $|S| = s$ at random.

(3) Choose $\theta_S = (\theta_i \colon i \in S)$ from density g_S on $\mathbb{R}^S$ and set $\theta_{S^c} = 0$.

We are particularly interested in π.

Model selection prior

Constructive definition of prior Π for $\theta \in \mathbb{R}^p$:

(1) Choose s from prior π on $\{0, 1, 2, \ldots, p\}$.

(2) Choose $S \subset \{0, 1, \ldots, p\}$ of size $|S| = s$ at random.

(3) Choose $\theta_S = (\theta_i : i \in S)$ from density g_S on $\mathbb{R}^S$ and set $\theta_{S^c} = 0$.

We are particularly interested in π.

EXAMPLE *(Slab and spike)*

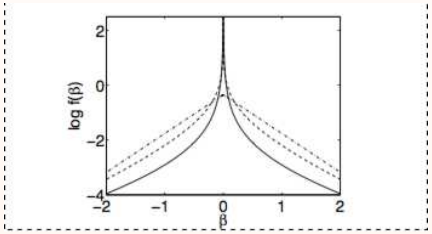
- Choose $\theta_1, \ldots, \theta_p$ i.i.d. from $\tau\delta_0 + (1-\tau)G$.
- Put a prior on τ, e.g. Beta$(1, p+1)$.

This gives binomial π and product densities $g_S = \otimes_{i \in S} g$.

[Mitchell & Beachamp (88), George, George & McCulloch, Yuan, Berger, Johnstone & Silverman, Richardson et al., Johnson & Rossell,...]

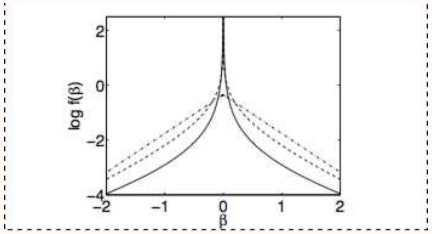
Other sparsity priors

Rather than distribution with a point mass at zero, one may use a continuous prior with a density that peaks near zero.



Other sparsity priors

Rather than distribution with a point mass at zero, one may use a continuous prior with a density that peaks near zero.



- *Bayesian LASSO*: $\theta_1, \ldots, \theta_p$ iid from a mixture of Laplace (λ) distributions over $\lambda \sim \sqrt{\Gamma(a, b)}$).
- *Bayesian bridge*: Same but with Laplace replaced with a density $\propto e^{-|\lambda y|^\alpha}$.
- *Normal-Gamma*: $\theta_1, \ldots, \theta_p$ iid from a Gamma scale mixture of Gaussians. *Correlated multivariate normal-Gamma*: $\theta = C\phi$ for a $p \times k$-matrix C and ϕ with independent normal-Gamma $(a_i, 1/2)$ coordinates.
- *Horseshoe*: Normal-Root Cauchy with Cauchy scale.
- *Normal spike*.
- *Scalar multiple of Dirichlet*.
- *Nonparametric Dirichlet*.
- ...

[Park & Casella 08, Polson & Scott, Griffin & Brown 10, 12, Carvalho & Polson & Scott, 10, George& Rockova 13, Bhattacharya et al. 12,...]

LASSO is not Bayesian!

$$\underset{\theta}{\operatorname{argmin}}\left[\|Y - X\theta\|^2 + \lambda \sum_{i=1}^{p} |\theta_i|\right].$$

The LASSO is the *posterior mode* for prior $\theta_i \overset{\text{iid}}{\sim} \mathrm{Laplace}(\lambda)$, but the full posterior distribution is useless, even with hyperprior on λ.
Trouble:
λ must be large to shrink θ_i to 0, but small to model nonzero θ_i.

Frequentist Bayes

Frequentist Bayes

Assume data Y follows a given parameter θ_0 and consider the posterior $\Pi(\theta \in \cdot \mid Y)$ as a *random measure* on the parameter set.

Frequentist Bayes

Assume data Y follows a given parameter θ_0 and consider the posterior $\Pi(\theta \in \cdot \,|\, Y)$ as a *random measure* on the parameter set.

We like $\Pi(\theta \in \cdot \,|\, Y)$:

- to put "most" of its mass near θ_0 for "most" Y.
- to have a spread that expresses "remaining uncertainty".
- to select the model defined by the nonzero parameters of θ_0.

Frequentist Bayes

Assume data Y follows a given parameter θ_0 and consider the posterior $\Pi(\theta \in \cdot \,|\, Y)$ as a *random measure* on the parameter set.

We like $\Pi(\theta \in \cdot \,|\, Y)$:

- to put "most" of its mass near θ_0 for "most" Y.
- to have a spread that expresses "remaining uncertainty".
- to select the model defined by the nonzero parameters of θ_0.

We evaluate this by probabilities or expectations, given θ_0.

Benchmarks for recovery — sequence model

$Y^n \sim N_n(\theta, I)$, for $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$.

$$\|\theta\|_0 = \#(1 \le i \le n\colon \theta_i \ne 0),$$
$$\|\theta\|_q^q = \sum_{i=1}^n |\theta_i|^q, \qquad 0 < q \le 2.$$

Benchmarks for recovery — sequence model

$$Y^n \sim N_n(\theta, I), \text{ for } \theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n.$$

$$\|\theta\|_0 = \#(1 \leq i \leq n \colon \theta_i \neq 0),$$

$$\|\theta\|_q^q = \sum_{i=1}^n |\theta_i|^q, \qquad 0 < q \leq 2.$$

Frequentist benchmarks: minimax rate relative to $\|\cdot\|_2$ over:

- black bodies $\{\theta \colon \|\theta\|_0 \leq s_n\}$:

$$\sqrt{s_n \log(n/s_n)}.$$

[(if $s_n \to \infty$ with $s_n/n \to 0$.) Donoho & Johnstone, Golubev, Johnstone and Silverman, Abramovich et al.,. . .]

Benchmarks for recovery — sequence model

$Y^n \sim N_n(\theta, I)$, for $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$.

$$\|\theta\|_0 = \#(1 \le i \le n \colon \theta_i \ne 0),$$
$$\|\theta\|_q^q = \sum_{i=1}^n |\theta_i|^q, \qquad 0 < q \le 2.$$

Frequentist benchmarks: minimax rate relative to $\|\cdot\|_q$ over:

- black bodies $\{\theta\colon \|\theta\|_0 \le s_n\}$:

$$s_n^{1/q}\sqrt{\log(n/s_n)}.$$

Benchmarks for recovery — sequence model

$Y^n \sim N_n(\theta, I)$, for $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$.

$$\|\theta\|_0 = \#(1 \le i \le n \colon \theta_i \ne 0),$$
$$\|\theta\|_q^q = \sum_{i=1}^n |\theta_i|^q, \qquad 0 < q \le 2.$$

Frequentist benchmarks: minimax rate relative to $\|\cdot\|_q$ over:

- black bodies $\{\theta \colon \|\theta\|_0 \le s_n\}$:

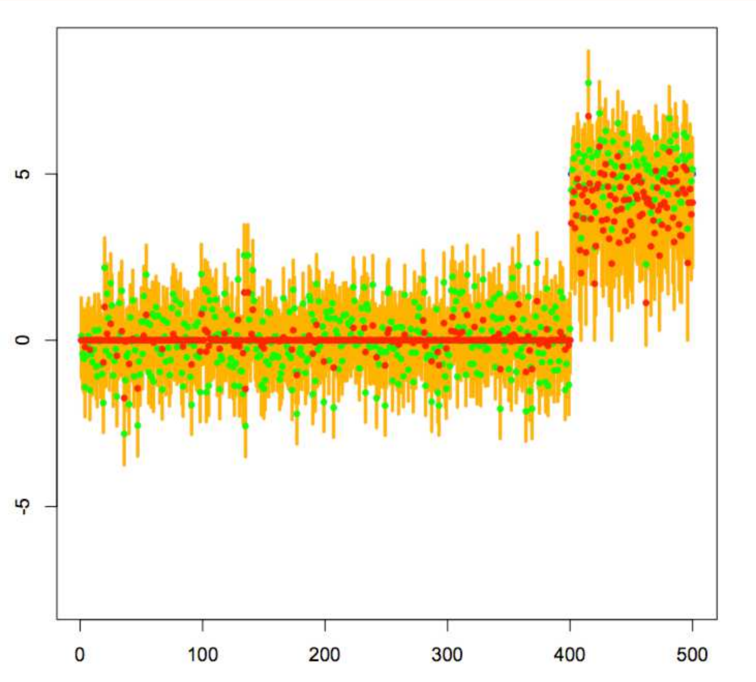
$$s_n^{1/q}\sqrt{\log(n/s_n)}.$$

- weak ℓ_r-balls $m_r[s_n] := \{\theta \colon \max_i i|\theta_{[i]}|^r \le n(s_n/n)^r\}$:

$$n^{1/q}(s_n/n)^{r/q}\sqrt{\log(n/s_n)}^{1-r/q}.$$

[(if $s_n \to \infty$ with $s_n/n \to 0$.) Donoho & Johnstone, Golubev, Johnstone and Silverman, Abramovich et al.,. . .]

Uncertainty quantification



Single data with $\theta_0 = (0, \ldots, 0, 5, \ldots, 5)$ and $n = 500$ and $\|\theta_0\|_0 = 100$.
Red dots: marginal posterior medians
Orange: marginal credible intervals
Green dots: data points.

Sequence model

Model selection prior

Prior Π_n on $\theta \in \mathbb{R}^n$:

(1) Choose s from prior π_n on $\{0, 1, 2, \ldots, n\}$.

(2) Choose $S \subset \{0, 1, \ldots, n\}$ of size $|S| = s$ at random.

(3) Choose $\theta_S = (\theta_i \colon i \in S)$ from density g_S on $\mathbb{R}^S$ and set $\theta_{S^c} = 0$.

Model selection prior

Prior Π_n on $\theta \in \mathbb{R}^n$:

(1) Choose s from prior π_n on $\{0, 1, 2, \ldots, n\}$.
(2) Choose $S \subset \{0, 1, \ldots, n\}$ of size $|S| = s$ at random.
(3) Choose $\theta_S = (\theta_i : i \in S)$ from density g_S on $\mathbb{R}^S$ and set $\theta_{S^c} = 0$.

Assume

- $\pi_n(s) \leq c\, \pi_n(s-1)$ for some $c < 1$, and every (large) s.
- g_S is product of densities e^h for uniformly Lipschitz $h: \mathbb{R} \to \mathbb{R}$ and with finite second moment.
- $s_n, n \to \infty$, $s_n/n \to 0$.

EXAMPLES:
- *complexity prior*: $\pi_n(s) \propto e^{-as \log(bn/s)}$.
- *slab and spike*: $\theta_i \overset{\text{iid}}{\sim} \tau\delta_0 + (1-\tau)G$ with $\tau \sim B(1, n+1)$.

Gaussian g is excluded. More general g_S are possible, e.g. (weak) dependence or grouping of coordinates.

Dimensionality

There exists M such that THEOREM *(black body)*

$$\sup_{\|\theta_0\|_0 \le s_n} \mathrm{E}_{\theta_0} \Pi_n\big(\theta\colon \|\theta\|_0 \ge M s_n \,|\, Y^n\big) \to 0.$$

Outside the space in which θ_0 lives, the posterior is concentrated in low-dimensional subspaces along the coordinate axes.

Dimensionality

There exists M such that THEOREM *(black body)*

$$\sup_{\|\theta_0\|_0 \le s_n} \mathrm{E}_{\theta_0} \Pi_n\big(\theta \colon \|\theta\|_0 \ge M s_n \,|\, Y^n\big) \to 0.$$

Outside the space in which θ_0 lives, the posterior is concentrated in low-dimensional subspaces along the coordinate axes.

THEOREM *(weak ball)*
For complexity prior π_n, any $r \in (0,2)$ and large M,

$$\sup_{\theta_0 \in m_r[s_n]} \mathrm{E}_{\theta_0} \Pi_n\big(\|\theta\|_0 > M s_n^* \,|\, Y^n\big) \to 0,$$

for "effective dimension": $s_n^* := n(s_n/n)^r / \log^{r/2}(n/s_n)$.

[Assume s_n not too small: $s_n^* \gtrsim 1$.]

Recovery

THEOREM *(black body)*

For every $0 < q \leq 2$ and large M,

$$\sup_{\|\theta_0\|_0 \leq s_n} \mathrm{E}_{\theta_0} \Pi_n\big(\theta\colon \|\theta - \theta_0\|_q > M r_n s_n^{1/q-1/2} | Y^n\big) \to 0,$$

for $r_n^2 = s_n \log(n/s_n) \vee \log(1/\pi_n(s_n))$.

If $\pi_n(s_n) \geq e^{-a s_n \log(n/s_n)}$ minimax rate is attained.

Recovery

THEOREM *(black body)*
For every $0 < q \leq 2$ and large M,

$$\sup_{\|\theta_0\|_0 \leq s_n} \mathrm{E}_{\theta_0} \Pi_n\big(\theta\colon \|\theta - \theta_0\|_q > M r_n s_n^{1/q-1/2} \,|\, Y^n\big) \to 0,$$

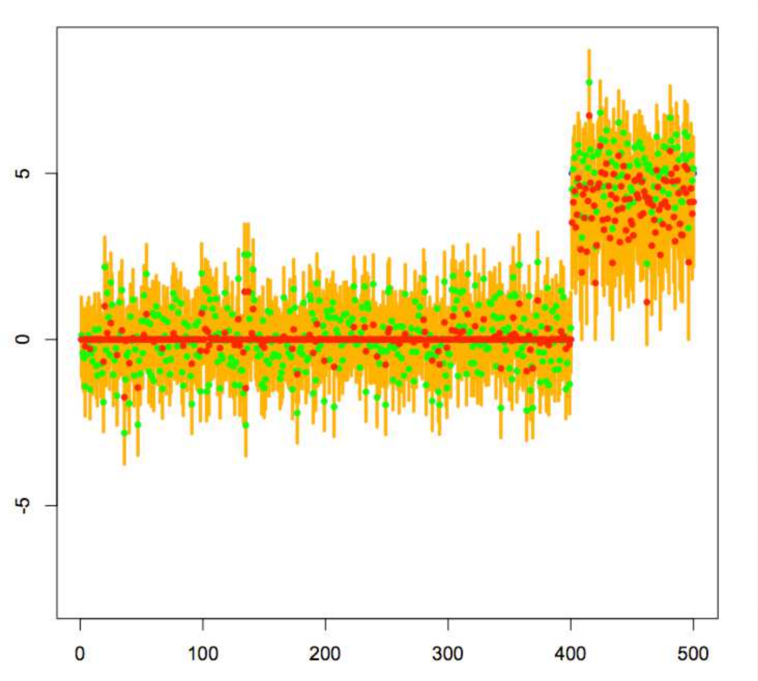
for $r_n^2 = s_n \log(n/s_n) \vee \log(1/\pi_n(s_n))$.

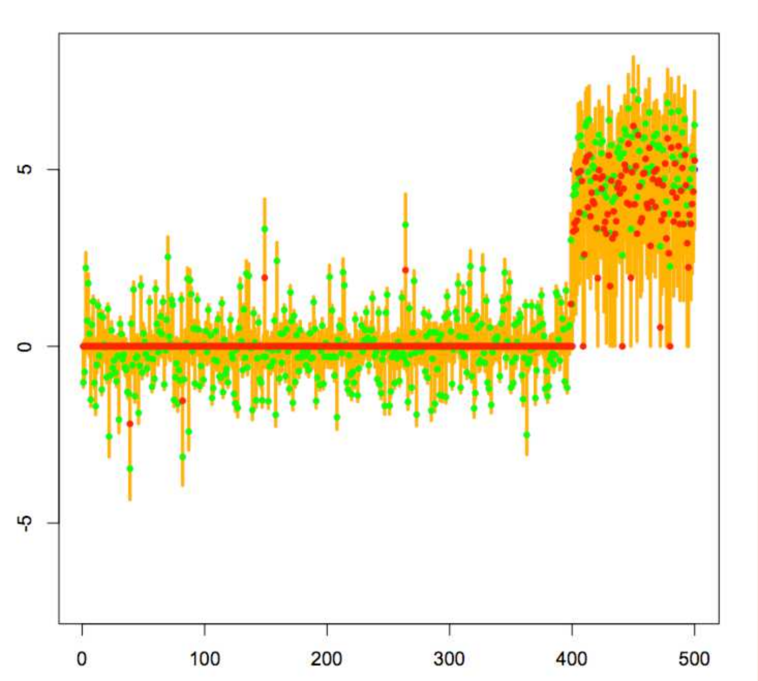
If $\pi_n(s_n) \geq e^{-a s_n \log(n/s_n)}$ minimax rate is attained.

THEOREM *(weak ball)*
For complexity prior π_n, any $r \in (0,2)$, any $q \in (r,2)$, the minimax rate $\mu^*_{n,r,q}$, and large M

$$\sup_{\theta_0 \in m_r[p_n]} \mathrm{E}_{\theta_0} \Pi_n\big(\theta\colon \|\theta - \theta_0\|_q > M \mu^*_{n,r,q} \,|\, Y^n\big) \to 0.$$

Illustration





Single data with $\theta_0 = (0, \ldots, 0, 5, \ldots, 5)$ and $n = 500$ and $\|\theta_0\|_0 = 100$.

Red dots: marginal posterior medians

Orange: marginal credible intervals

Green dots: data points.

g standard Laplace density.

$\pi_n(k) \propto \binom{2n-k}{n}^{\kappa}$ for $\kappa_1 = 0.1$ (left) and $\kappa_1 = 1$ (right).

Sequence model II

Horseshoe prior

Prior Π_n on $\mathbb{R}^n$:

(1) Choose "sparsity level" τ: empirical Bayes or $\tau \sim \mathrm{Cauchy}^+(0,1)$.

(2) Generate $\sqrt{\psi_1}, \ldots, \sqrt{\psi_n}$ iid from $\mathrm{Cauchy}^+(0,\tau)$.

(3) Generate independent $\theta_i \sim N(0,\psi_i)$.

Horseshoe prior

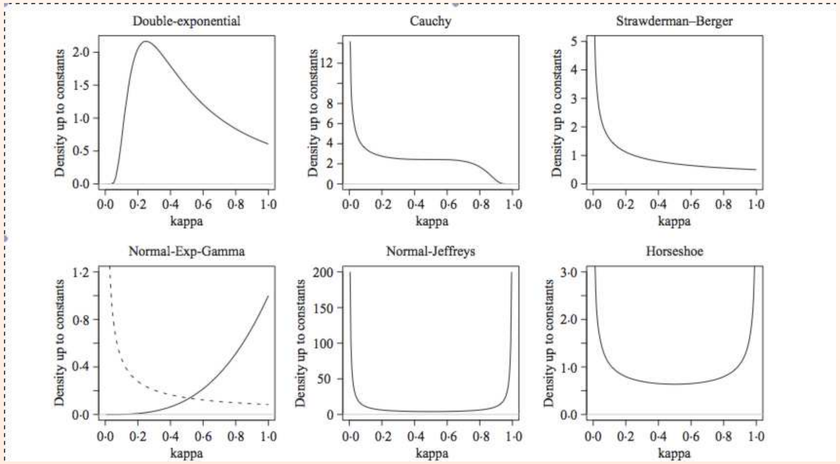
Prior Π_n on $\mathbb{R}^n$:

(1) Choose "sparsity level" τ: empirical Bayes or $\tau \sim \mathrm{Cauchy}^+(0,1)$.

(2) Generate $\sqrt{\psi_1}, \ldots, \sqrt{\psi_n}$ iid from $\mathrm{Cauchy}^+(0,\tau)$.

(3) Generate independent $\theta_i \sim N(0,\psi_i)$.

MOTIVATION: if $\theta \sim N(0,\psi)$ and $Y|\,\theta \sim N(\theta,1)$,
then $\theta|\,Y \sim N\big((1-\kappa)Y, 1-\kappa\big)$ for $\kappa = 1/(1+\psi)$.
This suggests a prior for κ that concentrates near 0 or 1.



[Carvalho & Polson & Scott, 10.]

Recovery

THEOREM *(black body)*

If $(s_n/n)^c \leq \hat{\tau}_n \leq C s_n/n$ for some $c, C > 0$, then for every $M_n \to \infty$,

$$\sup_{\|\theta_0\|_0 \leq s_n} \mathrm{E}_{\theta_0} \Pi_n\big(\theta\colon \|\theta - \theta_0\|_2 > M_n s_n \log(n/s_n) \,|\, Y^n\big) \to 0.$$

Minimax rate $s_n \log(n/p_n)$ is attained,
τ can be interpreted as sparsity level.

Recovery

THEOREM *(black body)*
If $(s_n/n)^c \leq \hat{\tau}_n \leq C s_n/n$ for some $c, C > 0$, then for every $M_n \to \infty$,

$$\sup_{\|\theta_0\|_0 \leq s_n} \mathrm{E}_{\theta_0} \Pi_n\big(\theta\colon \|\theta - \theta_0\|_2 > M_n s_n \log(n/s_n) \,|\, Y^n\big) \to 0.$$

Minimax rate $s_n \log(n/p_n)$ is attained,
τ can be interpreted as sparsity level.

- Posterior spread is (nearly?) of the same order.
- Easy to construct *some* $\hat{\tau}$.
- Hierarchical choice of τ not considered.

Regression

Regression model

$Y^n \sim N_n(X_{n \times p}\theta, I)$ for X known, $\theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$, $p \geq n$.

Regression model

$Y^n \sim N_n(X_{n\times p}\theta, I)$ for X known, $\theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$, $p \geq n$.

Summary of next 7 slides:

- similar results as in sequence model,
  under *sparse identifiability conditions* on the regression matrix.
- allow scaling of prior on zero elements.

Compatibility and coherence

$$\|X\|:=\max_j \|X_{\cdot,j}\|.$$

Compatibility number $\phi(S)$ for $S\subset\{1,\ldots,p\}$ is: $\displaystyle\inf_{\|\theta_{S^c}\|_1\le 7\|\theta\|_1}\frac{\|X\theta\|_2\sqrt{|S|}}{\|X\|\|\theta_S\|_1}$.

Compatibility in s_n-sparse vectors means: $\displaystyle\inf_{\theta:\|\theta\|_0\le 5s_n}\frac{\|X\theta\|_2\sqrt{|S_\theta|}}{\|X\|\|\theta\|_1}\gg 0.$

Strong compatibility in s_n-sparse vectors means: $\displaystyle\inf_{\theta:\|\theta\|_0\le 5s_n}\frac{\|X\theta\|_2}{\|X\|\|\theta\|_2}\gg 0.$

Mutual coherence means: $\displaystyle s_n\max_{i\ne j}\left|\mathrm{cor}(X_{\cdot i},X_{\cdot j})\right|\ll 1.$

Write $\phi(\theta)=\phi(S_\theta)$ for the set $S_\theta=\{i:\theta_i\ne 0\}$.

Model selection prior

Prior Π_n for $\theta \in \mathbb{R}^p$:

(1) Choose s from prior π_n on $\{0, 1, 2, \ldots, p\}$.

(2) Choose $S \subset \{0, 1, \ldots, s\}$ of size $|S| = s$ at random.

(3) Choose $\theta_S = (\theta_i \colon i \in S)$ from density g_S on $\mathbb{R}^S$ and set $\theta_{S^c} = 0$.

Model selection prior

Prior Π_n for $\theta \in \mathbb{R}^p$:

(1) Choose s from prior π_n on $\{0, 1, 2, \ldots, p\}$.
(2) Choose $S \subset \{0, 1, \ldots, s\}$ of size $|S| = s$ at random.
(3) Choose $\theta_S = (\theta_i \colon i \in S)$ from density g_S on $\mathbb{R}^S$ and set $\theta_{S^c} = 0$.

Assume

- $\pi_p(s) \leq p^{-2}\pi_n(s-1)$ and $\pi_n(s) \geq c^{-s}e^{-as\log(bp)}$.
- g_S is product of Laplace (λ) densities.
- $p^{-1} \leq \lambda/\|X\| \leq 2\sqrt{\log p}$, for $\|X\| := \max_j \|X_{\cdot,j}\|$.

λ can be fixed or even $\lambda \to 0$.
Scenario 1: sequence model: $\|X\| = 1$, $\lambda \geq p^{-1}$.
Scenario 2: each $(Y_i, X_{i.})$ instance of fixed equation: $\|X\| \sim \sqrt{n}$, $\lambda \gtrsim \sqrt{n}/p$.
Scenario 3: sequence model with $N(0, \sigma_n^2)$ errors: $\lambda \gtrsim \sigma_n^{-1}/n$.

Dimensionality

THEOREM

For any $s_n, c_0 > 0$

$$\sup_{\|\theta_0\|_0 \le s_n, \phi(\theta_0) \ge c_0} \mathrm{E}_{\theta_0} \Pi_n(\theta\colon \|\theta\|_0 > 4s_n \,|\, Y^n) \to 0.$$

Outside the space in which θ_0 lives, the posterior is concentrated in low-dimensional subspaces along the coordinate axes.

Recovery

THEOREM *(black body)*

Given compatibility of s_n-sparse vectors, for every $c_0 > 0$,

$$\sup_{\|\theta_0\|_0 \le s_n, \phi(\theta_0) \ge c_0} \mathrm{E}_{\theta_0} \Pi_n(\theta\colon \|X(\theta - \theta_0)\|_2 \gtrsim \sqrt{s_n \log p} \,|\, Y^n) \to 0,$$

$$\sup_{\|\theta_0\|_0 \le s_n, \phi(\theta_0) \ge c_0} \mathrm{E}_{\theta_0} \Pi_n(\theta\colon \|\theta - \theta_0\|_1 \gtrsim s_n \sqrt{\log p}/\|X\| \,|\, Y^n) \to 0.$$

Recovery

THEOREM *(black body)*

Given compatibility of s_n-sparse vectors, for every $c_0 > 0$,

$$\sup_{\|\theta_0\|_0 \le s_n, \phi(\theta_0) \ge c_0} \mathrm{E}_{\theta_0} \Pi_n(\theta: \|X(\theta - \theta_0)\|_2 \gtrsim \sqrt{s_n \log p} | Y^n) \to 0,$$

$$\sup_{\|\theta_0\|_0 \le s_n, \phi(\theta_0) \ge c_0} \mathrm{E}_{\theta_0} \Pi_n(\theta: \|\theta - \theta_0\|_1 \gtrsim s_n \sqrt{\log p} / \|X\| | Y^n) \to 0.$$

Minimax rates (almost) attained.

Recovery

THEOREM *(black body)*

Given compatibility of s_n-sparse vectors, for every $c_0 > 0$,

$$\sup_{\|\theta_0\|_0 \le s_n, \phi(\theta_0) \ge c_0} \mathrm{E}_{\theta_0} \Pi_n(\theta\colon \|X(\theta-\theta_0)\|_2 \gtrsim \sqrt{s_n \log p} | Y^n) \to 0,$$

$$\sup_{\|\theta_0\|_0 \le s_n, \phi(\theta_0) \ge c_0} \mathrm{E}_{\theta_0} \Pi_n(\theta\colon \|\theta-\theta_0\|_1 \gtrsim s_n \sqrt{\log p}/\|X\| | Y^n) \to 0.$$

Minimax rates (almost) attained.

THEOREM *(oracle, weak norms)*

Under same conditions

$$\sup_{\substack{\|\theta_0\|_0 \le s_n, \phi(\theta_0) \ge c_0 \\ \|\theta_*\|_0 \le s_*, \phi(\theta_*) \ge c_0}} \mathrm{E}_{\theta_0} \Pi_n(\theta\colon \|X(\theta-\theta_0)\|_2^2 + \sqrt{\log p}\|X\| \|\theta-\theta_0\|_1$$

$$\gtrsim \|X(\theta_*-\theta_0)\|_2^2 + s_* \log p | Y^n) \to 0.$$

Selection

THEOREM *(No supersets)*

Given strong compatibility of s_n sparse vectors,

$$\sup_{\|\theta_0\|_0 \le s_n, \phi(\theta_0) \ge c_0} \mathrm{E}_{\theta_0} \Pi_n(\theta: S_\theta \supset S_{\theta_0}, S_\theta \neq S_{\theta_0} \mid Y^n) \to 0.$$

Selection

THEOREM *(No supersets)*

Given strong compatibility of s_n sparse vectors,

$$\sup_{\|\theta_0\|_0\le s_n, \phi(\theta_0)\ge c_0} \mathrm{E}_{\theta_0}\Pi_n(\theta: S_\theta \supset S_{\theta_0}, S_\theta \neq S_{\theta_0} \,|\, Y^n) \to 0.$$

THEOREM *(Finds big signals)*

- Given compatibility of s_n sparse vectors,

$$\inf_{\|\theta_0\|_0\le s_n, \phi(\theta_0)\ge c_0} \mathrm{E}_{\theta_0}\Pi_n\Big(\theta: S_\theta \supset \{i: |\theta_{0,i}| \gtrsim s_n\sqrt{\log p}/\|X\|\} \,|\, Y^n\Big) \to 1.$$

Selection

THEOREM *(No supersets)*

Given strong compatibility of s_n sparse vectors,

$$\sup_{\|\theta_0\|_0 \le s_n, \phi(\theta_0) \ge c_0} \mathrm{E}_{\theta_0} \Pi_n(\theta : S_\theta \supset S_{\theta_0}, S_\theta \neq S_{\theta_0} \,|\, Y^n) \to 0.$$

THEOREM *(Finds big signals)*

- Given compatibility of s_n sparse vectors,

$$\inf_{\|\theta_0\|_0 \le s_n, \phi(\theta_0) \ge c_0} \mathrm{E}_{\theta_0} \Pi_n\Big(\theta : S_\theta \supset \{i : |\theta_{0,i}| \gtrsim s_n \sqrt{\log p} / \|X\|\} \,|\, Y^n\Big) \to 1.$$

- Under strong compatibility s_n can be replaced by $\sqrt{s_n}$.
- Under mutual coherence s_n can be replaced by a constant.

Selection

THEOREM *(No supersets)*

Given strong compatibility of s_n sparse vectors,

$$\sup_{\|\theta_0\|_0 \le s_n, \phi(\theta_0)\ge c_0} \mathrm{E}_{\theta_0}\Pi_n(\theta\colon S_\theta \supset S_{\theta_0}, S_\theta \neq S_{\theta_0} \,|\, Y^n) \to 0.$$

THEOREM *(Finds big signals)*

- Given compatibility of s_n sparse vectors,

$$\inf_{\|\theta_0\|_0 \le s_n, \phi(\theta_0)\ge c_0} \mathrm{E}_{\theta_0}\Pi_n\bigl(\theta\colon S_\theta \supset \{i\colon |\theta_{0,i}| \gtrsim s_n\sqrt{\log p}/\|X\|\} \,|\, Y^n\bigr) \to 1.$$

- Under strong compatibility s_n can be replaced by $\sqrt{s_n}$.
- Under mutual coherence s_n can be replaced by a constant.

Corollary: if *all* nonzero $|\theta_{0,i}|$ are suitably big, then posterior probability of true model S_{θ_0} tends to 1.

Bernstein-von Mises theorem

Assume 'flat priors':

$$\frac{\lambda}{\|X\|} s_n \sqrt{\log p} \to 0.$$

Bernstein-von Mises theorem

Assume 'flat priors':

$$\frac{\lambda}{\|X\|} s_n \sqrt{\log p} \to 0.$$

THEOREM

Given compatibility of s_n-sparse vectors,

$$\mathrm{E}_{\theta_0} \Big\| \Pi_n(\cdot \,|\, Y^n) - \sum_S \hat{w}_S N(\hat{\theta}_{(S)}, \Gamma_S^{-1}) \otimes \delta_{S^c} \Big\| \to 0,$$

for $\hat{\theta}_{(S)}$ the LS estimator for model S, Γ_S^{-1} its covariance, and

$$\hat{w}_S \propto \frac{\pi_p(s)}{\binom{p}{s}} \Big(\frac{\lambda\sqrt{2\pi}}{2}\Big)^s |\Gamma_S|^{-1/2} e^{\frac{1}{2}\|X_S \widehat{\theta}_{(S)}\|_2^2} 1_{|S| \le 4s_n, \|\theta_{0,S^c}\|_1 \lesssim s_n \sqrt{\log p}/\|X\|}.$$

THEOREM

Given consistent model selection, mixture can be replaced by $N(\hat{\theta}_{(S_{\theta_0})}, \Gamma^{-1}_{S_{\theta_0}})$.

Credible set

THEOREM

Given consistent model selection, credible sets for individual parameters are asymptotic confidence sets.

Credible set

THEOREM

Given consistent model selection, credible sets for individual parameters are asymptotic confidence sets.

Open questions:

- What if true model is not consistently selected?
- Do credible sets for multiple parameters control for multiplicity correction?