Asymptotic analysis of Bayesian methods for sparse regression

Aad van der Vaart

Universiteit Leiden

O’Bayes, December 2013
Sparsity
Frequentist Bayes
Sequence model
Sequence model II
Regression
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 $\theta$ 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 $j$.

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

$$EY_{i,j} = e^{\alpha_i + \beta_i x_j}, \quad \text{var } Y_{i,j} = EY_{i,j} (1 + EY_{i,j} e^{-\phi_i}).$$

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

1. Choose $s$ from prior $\pi$ 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 $\theta_{Sc} = 0$. 
Model selection prior

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

(1) Choose $s$ from prior $\pi$ 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 $\theta_{Sc} = 0$.

We are particularly interested in $\pi$. 
Model selection prior

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

1. Choose $s$ from prior $\pi$ 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 $\theta_{Sc} = 0$.

We are particularly interested in $\pi$.

**EXAMPLE**  *(Slab and spike)*

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

This gives binomial $\pi$ 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.
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 ($\lambda$) 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 $\phi$ 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!

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

The LASSO is the posterior mode for prior \( \theta_i \overset{iid}{\sim} \text{Laplace}(\lambda) \), but the full posterior distribution is useless, even with hyperprior on \( \lambda \).

Trouble:
\( \lambda \) must be large to shrink \( \theta_i \) to 0, but small to model nonzero \( \theta_i \).
Frequentist Bayes
Frequentist Bayes

Assume data $Y$ follows a given parameter $\theta_0$ and consider the posterior $\Pi(\theta \in \cdot | Y)$ as a random measure on the parameter set.
Frequentist Bayes

Assume data $Y$ follows a given parameter $\theta_0$ 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 $\theta_0$ for “most” $Y$.
- to have a spread that expresses “remaining uncertainty”.
- to select the model defined by the nonzero parameters of $\theta_0$. 
Frequentist Bayes

Assume data $Y$ follows a given parameter $\theta_0$ 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 $\theta_0$ for “most” $Y$.
- to have a spread that expresses “remaining uncertainty”.
- to select the model defined by the nonzero parameters of $\theta_0$.

We evaluate this by probabilities or expectations, given $\theta_0$. 
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: \theta_i \neq 0),

\|\theta\|_q = \sum_{i=1}^{n} |\theta_i|^q, \quad 0 < q \leq 2.
Benchmarks for recovery — sequence model

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

\[
\|\theta\|_0 = \#(1 \leq i \leq n: \theta_i \neq 0),
\]

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

**Frequentist benchmarks:** minimax rate relative to \( \| \cdot \|_2 \) over:

- **black bodies** \( \{ \theta: \|\theta\|_0 \leq s_n \} \):

\[
\sqrt{s_n \log(n/s_n)}.
\]

[(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

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

\[
\|\theta\|_0 = \#(1 \leq i \leq n: \theta_i \neq 0),
\]

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

**Frequentist benchmarks:** minimax rate relative to \( \| \cdot \|_q \) over:

- **black bodies** \( \{ \theta: \|\theta\|_0 \leq s_n \} \):
  \[
s_n^{1/q} \sqrt{\log(n/s_n)}.\]
Benchmarks for recovery — sequence model

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

\[ \|\theta\|_0 = \#(1 \leq i \leq n: \theta_i \neq 0), \]

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

**Frequentist benchmarks:** minimax rate relative to \( \| \cdot \|_q \) over:

- **black bodies** \( \{ \theta: \|\theta\|_0 \leq s_n \} \):
  \[ s_n^{1/q} \sqrt{\log(n/s_n)}. \]

- **weak \( \ell_r \)-balls** \( m_r[s_n] := \{ \theta: \max_i |\theta[i]|^r \leq n(s_n/n)^r \} \):
  \[ n^{1/q} (s_n/n)^{r/q} \sqrt{\log(n/s_n)^{1-r/q}}. \]

([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 $\Pi_n$ on $\theta \in \mathbb{R}^n$:

1. Choose $s$ from prior $\pi_n$ 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 $\theta_{Sc} = 0$. 
Model selection prior

Prior \( \Pi_n \) on \( \theta \in \mathbb{R}^n \):

1. Choose \( s \) from prior \( \pi_n \) 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 \( \theta_{Sc} = 0 \).

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 \( h: \mathbb{R} \rightarrow \mathbb{R} \) and with finite second moment.
- \( s_n, n \rightarrow \infty, s_n/n \rightarrow 0 \).

EXAMPLES:

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

$$
\sup_{\|\theta_0\|_0 \leq s_n} \mathbb{E}_{\theta_0} \Pi_n \left( \theta : \|\theta\|_0 \geq M s_n \mid Y^n \right) \rightarrow 0.
$$

Outside the space in which $\theta_0$ lives, the posterior is concentrated in low-dimensional subspaces along the coordinate axes.
**Dimensionality**

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

\[
\sup_{\|\theta_0\|_0 \leq s_n} \mathbb{E}_{\theta_0} \Pi_n(\theta: \|\theta\|_0 \geq Ms_n | Y^n) \to 0.
\]

Outside the space in which \( \theta_0 \) lives, the posterior is concentrated in low-dimensional subspaces along the coordinate axes.

**THEOREM** *(weak ball)*

For complexity prior \( \pi_n \), any \( r \in (0, 2) \) and large \( M \),

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

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

[Assume \( s_n \) not too small: \( s_n^* \gtrsim 1 \).]
Recovery

**THEOREM (black body)**

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

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

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

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

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

$$
\sup_{\|\theta_0\|_0 \leq s_n} \mathbb{E}_{\theta_0} \Pi_n(\theta: \|\theta - \theta_0\|_q > Mr_n s_n^{1/q-1/2} | Y^n) \to 0,
$$

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

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

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

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

$\pi_n(k) \propto \left(\frac{2n-k}{n}\right)^\kappa$ for $\kappa_1 = 0.1$ (left) and $\kappa_1 = 1$ (right).
Sequence model II
Horseshoe prior

Prior $\Pi_n$ on $\mathbb{R}^n$:

1. Choose “sparsity level” $\tau$: empirical Bayes or $\tau \sim \text{Cauchy}^+(0, 1)$.
2. Generate $\sqrt{\psi_1}, \ldots, \sqrt{\psi_n}$ iid from $\text{Cauchy}^+(0, \tau)$.
3. Generate independent $\theta_i \sim N(0, \psi_i)$. 
Horseshoe prior

Prior $\Pi_n$ on $\mathbb{R}^n$:

1. Choose “sparsity level” $\tau$: empirical Bayes or $\tau \sim \text{Cauchy}^+(0, 1)$.
2. Generate $\sqrt{\psi_1}, \ldots, \sqrt{\psi_n}$ iid from $\text{Cauchy}^+(0, \tau)$.
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 $\theta | Y \sim N((1 - \kappa)Y, 1 - \kappa)$ for $\kappa = 1/(1 + \psi)$. This suggests a prior for $\kappa$ that concentrates near 0 or 1.

[Carvalho & Polson & Scott, 10.]
Recovery

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

\[
\sup_{\|\theta_0\|_0 \leq s_n} E_{\theta_0} \Pi_n(\theta: \|\theta - \theta_0\|_2 > M_n s_n \log(n/s_n)| Y^n) \to 0.
\]

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

**THEOREM** *(black body)*

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

\[
\sup_{\|\theta_0\|_0 \leq s_n} E_{\theta_0} \Pi_n(\theta: \|\theta - \theta_0\|_2 > M_n s_n \log(n/s_n) | Y^n) \to 0.
\]

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

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

\[ Y^n \sim N_n(X_{n \times p} \theta, I) \text{ for } X \text{ known, } \theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p, p \geq n. \]
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 \textit{sparse identifiability conditions} on the regression matrix.
- allow scaling of prior on zero elements.
Compatibility and coherence

\[ \|X\| := \max_j \|X_{.,j}\|. \]

Compatibility number \( \phi(S) \) for \( S \subset \{1, \ldots, p\} \) is:
\[ \inf_{\|\theta_S\|_1 \leq 7\|\theta\|_1} \frac{\|X\theta\|_2 \sqrt{|S|}}{\|X\| \|\theta_S\|_1}. \]

Compatibility in \( s_n \)-sparse vectors means:
\[ \inf_{\theta: \|\theta\|_0 \leq 5s_n} \frac{\|X\theta\|_2 \sqrt{|S\theta|}}{\|X\| \|\theta\|_1} \gg 0. \]

Strong compatibility in \( s_n \)-sparse vectors means:
\[ \inf_{\theta: \|\theta\|_0 \leq 5s_n} \frac{\|X\theta\|_2}{\|X\| \|\theta\|_2} \gg 0. \]

Mutual coherence means:
\[ s_n \max_{i \neq j} \left| \text{cor}(X_{.,i}, X_{.,j}) \right| \ll 1. \]

Write \( \phi(\theta) = \phi(S_{\theta}) \) for the set \( S_{\theta} = \{i: \theta_i \neq 0\} \).
Model selection prior

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

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

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

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

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 ($\lambda$) densities.
- $p^{-1} \leq \lambda/\|X\| \leq 2\sqrt{\log p}$, for $\|X\| := \max_j \|X_{.,j}\|$.

$\lambda$ 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: $\lambda \gtrsim \sigma_n^{-1}/n$. 
THEOREM

For any $s_n, c_0 > 0$

$$\sup_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} \mathbb{E}_{\theta_0} \Pi_n(\theta: \|\theta\|_0 > 4s_n | Y^n) \to 0.$$ 

Outside the space in which $\theta_0$ lives, the posterior is concentrated in low-dimensional subspaces along the coordinate axes.
Theorem (black body)

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

$$
\sup_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} E_{\theta_0} \Pi_n(\theta : \|X(\theta - \theta_0)\|_2 \gtrsim \sqrt{s_n \log p} | Y^n) \to 0,
$$

$$
\sup_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} E_{\theta_0} \Pi_n(\theta : \|\theta - \theta_0\|_1 \gtrsim s_n \sqrt{\log p / \|X\|} | Y^n) \to 0.
$$
Recovery

**THEOREM**  (*black body*)

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

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

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

Minimax rates (almost) attained.
Recovery

**THEOREM (black body)**

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

\[
\sup_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} \mathbb{E}_{\theta_0} \Pi_n(\theta; \|X(\theta - \theta_0)\|_2) \gtrsim \sqrt{s_n \log p} \|Y^n\) \to 0,
\]

\[
\sup_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} \mathbb{E}_{\theta_0} \Pi_n(\theta; \|\theta - \theta_0\|_1) \gtrsim s_n \sqrt{\log p/\|X\| \|Y^n\}) \to 0.
\]

Minimax rates (almost) attained.

**THEOREM (oracle, weak norms)**

Under same conditions

\[
\sup_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0, \|\theta_*\|_0 \leq s_*, \phi(\theta_*) \geq c_0} \mathbb{E}_{\theta_0} \Pi_n(\theta; \|X(\theta - \theta_0)\|_2 + \sqrt{\log p} \|X\| \|\theta - \theta_0\|_1 \gtrsim \|X(\theta_* - \theta_0)\|_2 + s_\ast \log p \|Y^n\) \to 0.
\]
Selection

THEOREM  *(No supersets)*

Given strong compatibility of $s_n$ sparse vectors,

$$\sup_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} E_{\theta_0} \Pi_n (\theta: S_{\theta} \supset S_{\theta_0}, S_{\theta} \neq S_{\theta_0} | Y^n) \to 0.$$
THEOREM (No supersets)
Given strong compatibility of \( s_n \) sparse vectors,

\[
\sup_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} \mathbb{E}_{\theta_0} \Pi_n(\theta: S_\theta \supset S_{\theta_0}, S_\theta \neq S_{\theta_0} | Y^n) \to 0.
\]

THEOREM (Finds big signals)

• Given compatibility of \( s_n \) sparse vectors,

\[
\inf_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} \mathbb{E}_{\theta_0} \Pi_n(\theta: S_\theta \supset \{i: |\theta_0, i| \gtrsim s_n \sqrt{\log p/\|X\|}\} | Y^n) \to 1.
\]
**Selection**

**THEOREM**  *(No supersets)*

Given strong compatibility of \( s_n \) sparse vectors,

\[
\sup_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} \mathbb{E}_{\theta_0} \prod_n (\theta: S_\theta \supset S_{\theta_0}, S_\theta \neq S_{\theta_0} | Y^n) \to 0.
\]

**THEOREM**  *(Finds big signals)*

- Given compatibility of \( s_n \) sparse vectors,

\[
\inf_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} \mathbb{E}_{\theta_0} \prod_n (\theta: S_\theta \supset \{i: |\theta_0,i| \geq s_n \sqrt{\log p/\|X\|}\} | Y^n) \to 1.
\]

- Under strong compatibility \( s_n \) can be replaced by \( \sqrt{s_n} \).
- Under mutual coherence \( s_n \) can be replaced by a constant.
THEOREM  \((No \ sups)ets)\n
Given strong compatibility of \(s_n\) sparse vectors, \[ \sup_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} \mathbb{E}_{\theta_0} \Pi_n(\theta: S_\theta \supset S_{\theta_0}, S_\theta \neq S_{\theta_0} \mid Y^n) \to 0. \]

THEOREM  \((Finds \ big \ signals)\n
• Given compatibility of \(s_n\) sparse vectors, \[ \inf_{\|\theta_0\|_0 \leq s_n, \phi(\theta_0) \geq c_0} \mathbb{E}_{\theta_0} \Pi_n(\theta: S_\theta \supset \{i: |\theta_0,i| \geq s_n \sqrt{\log p/\|X\|}\} \mid Y^n) \to 1. \]

• 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 \(S_{\theta_0}\) tends to 1.
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,

\[
E_{\theta_0} \left\| \Pi_n (\cdot | Y^n) - \sum_S \hat{w}_S N(\hat{\theta}(S), \Gamma_S^{-1}) \otimes \delta_{S^c} \right\| \to 0,
\]

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

\[
\hat{w}_S \propto \frac{\pi_p(s)}{p^s} \left( \frac{\lambda \sqrt{2\pi}}{2} \right)^s |S|^{-1/2} e^{\frac{1}{2} \|X_S \hat{\theta}(S)\|_2^2} 1_{|S| \leq 4s_n, \|\theta_0, S^c\|_1 \lesssim s_n \sqrt{\log p}/\|X\|}.
\]

**THEOREM**
Given consistent model selection, mixture can be replaced by \(N(\hat{\theta}(S_{\theta_0}), \Gamma_{S_{\theta_0}}^{-1})\).
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?