LASSO is not Fully Bayesian

Behrad Moniri
bemoniri@ee.sharif.edu

LASSO is not fully Bayesian.

- Data Generation Model: $Y = X\beta_0 + \epsilon$ where $\epsilon \sim \mathcal{N}(0, I)$.
- The LASSO optimization problem:

$$\hat{\beta}_{\lambda}^{\text{LASSO}} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \left[\|Y - X\beta\|_2^2 + 2\lambda\|\beta\|_1 \right] \tag{1}$$

- The posterior mode for the prior $\beta_i \sim \text{Laplace}(\lambda)$. Hence it has a Bayesian flavor.
- **Frequentist Optimality**: Can attain the (near) minimax rate $O(s \log n)$ for the square Euclidean loss over s-sparse signals, if $\lambda \approx \sqrt{2\log(n)}$.

Not really Bayesian!

- We will show that the full posterior does not contract at the same speed as its mode!
- Useless for uncertainty quantification, the central idea of Bayesian inference.

Not really Bayesian!

- We will show that the full posterior does not contract at the same speed as its mode!
- Useless for uncertainty quantification, the central idea of Bayesian inference.
- The good behavior of the LASSO estimators must be due to the sparsity-inducing form of the posterior mode, **not** the Bayesian connection.

The Main Result

Theorem (7) from [CSV15]

Assume that $X = I$. For any $\lambda = \lambda_n$ such that $\sqrt{n}/\lambda_n \to \infty$, there exists $\delta > 0$ such that, as $n \to \infty$,

$$\mathbb{E}_{\beta^0=0}\Pi_{\lambda_n}^{\mathrm{LASSO}}\left(\beta : \|\beta\|_2 \leq \delta\sqrt{n}\left(\frac{1}{\lambda_n} \wedge 1\right)\middle| Y\right) \to 0.$$

- Let $\lambda_n = \sqrt{2\log(n)}$. The posterior places **no** weight on the ball $\|\beta\|_2 = O\left(\sqrt{\frac{n}{\log n}}\right)$ which is substantially larger than the minimax rate $\sqrt{s\log(n)}$ (unless the signal is extremely dense).

We will first state two lemmas:

Lemma (5.2) form [CV12]

For any prior probability distribution Π_n on $\mathbb{R}^n$, any positive measure $\tilde{\Pi}_n$ with $\tilde{\Pi}_n \leq \Pi_n$, and any β_0,

$$\int \frac{p_\beta}{p_{\beta_0}}(Y)\, d\Pi_n(\beta) \geq \|\tilde{\Pi}_n\| e^{-\tilde{\sigma}^2/2+\tilde{\mu}^T(Y-\beta_0)},$$

where $\tilde{\mu} = \int (\beta - \beta_0)\, d\tilde{\Pi}_n(\beta)/\|\tilde{\Pi}_n\|$ and $\tilde{\sigma}^2 = \int \|\beta - \beta_0\|_2^2\, d\tilde{\Pi}_n(\beta)/\|\tilde{\Pi}_n\|$. Also, for any $r > 0$,

$$P_{\beta_0}\left(\int \frac{p_\beta}{p_{\beta_0}}(Y)\, d\Pi_n(\beta) \geq e^{-r^2}\Pi_n\left(\beta : \|\beta - \beta_0\|_2 < r\right)\right) \geq 1 - e^{-r^2/8}.$$

Proof of Lemma (5.2): First Assertion

We can write:

$$\log\left(\int \frac{p_\beta}{p_{\beta_0}}(Y)\frac{d\Pi_n(\beta)}{\|\tilde{\Pi}_n\|}\right) \geq \int \log\left(\frac{p_\beta}{p_{\beta_0}}(Y)\right)\frac{d\Pi_n(\beta)}{\|\tilde{\Pi}_n\|} \quad \text{(Jensen's)}$$

$$\geq \int \log\left(\frac{p_\beta}{p_{\beta_0}}(Y)\right)\frac{d\tilde{\Pi}_n(\beta)}{\|\tilde{\Pi}_n\|} \quad (\tilde{\Pi}_n \leq \Pi_n)$$

$$= -\frac{\tilde{\sigma}^2}{2} + \tilde{\mu}^\top(Y-\beta) \quad \text{(Gaussianity)}$$

Where $\tilde{\mu} = \int(\beta-\beta_0)\frac{d\tilde{\Pi}_n}{\|\tilde{\Pi}_n\|}$ and $\tilde{\sigma}^2 = \int \|\beta-\beta_0\|_2^2\frac{d\tilde{\Pi}_n}{\|\tilde{\Pi}_n\|}$.
Hence:

$$\int \frac{p_\beta}{p_{\beta_0}}(Y)\,d\Pi_n(\beta) \geq \|\tilde{\Pi}_n\|e^{-\tilde{\sigma}^2/2+\tilde{\mu}^T(Y-\beta_0)}.$$

Proof of Lemma (5.2): Second Assertion

- Let $\tilde{\Pi}_n$ be Π_n restricted to the set $\{\beta : \|\beta - \beta_0\|_2 \le r\}$. Trivially $\tilde{\Pi}_n \le \Pi_n$, $\|\tilde{\mu}\| \le r$, $\tilde{\sigma} \le r$, and $\|\tilde{\Pi}_n\| = \Pi_n(\beta : \|\beta - \beta_0\|_2 < r)$.
- Under P_{β_0}, we have that $\tilde{\mu}^\top (Y - \beta_0)$ has the same distribution as $\|\tilde{\mu}\| Z$ where $Z \sim \mathcal{N}(0, I_n)$.
- Note that

$$P\Big[Zr \le -r^2 + r^2/2\Big] \le \exp\big(-r^2/8\big).$$

- Hence:

$$P_{\beta_0}\left(\int \frac{p_\beta}{p_{\beta_0}}(Y)\, d\Pi_n(\beta) \ge e^{-r^2}\Pi_n\big(\beta : \|\beta - \beta_0\| < r\big)\right) \ge 1 - e^{-r^2/8}.$$

Which concludes the proof. □

Lemma (7.1) from [CV12]

Lemma (7.1) from [CV12]

We have $\mathrm{E}_{\beta_0}\Pi_n(\beta : \|\beta - \beta_0\| < s_n|Y) \to 0$, for any s_n for which there exist r_n such that

$$\frac{\Pi_n(\beta : \|\beta - \beta_0\| < s_n)}{\Pi_n(\beta : \|\beta - \beta_0\| < r_n)} = o\left(e^{-r_n^2}\right).$$

Based on the previous lemma, there exists and event $\mathcal{A}_n$ with $\Pr[A_n] \geq 1 - \exp(-r_n^2/8)$ such that

$$\int \frac{p_\beta}{p_{\beta_0}}(Y)\, d\Pi_n(\beta) \geq e^{-r^2} \Pi_n\big(\beta : \|\beta - \beta_0\| < r\big).$$

$$\begin{aligned}
\mathrm{E}_{\beta_0}\Big[\Pi_n(\beta : \|\beta - \beta_0\| < s_n | Y) 1_{\mathcal{A}_n}\Big] &= \mathrm{E}_{\beta_0}\left[\frac{\int_{\beta : \|\beta-\beta_0\| < s_n} p_\beta(Y)\, d\Pi_n(\beta)}{\int p_\beta(Y) d\Pi_n(\beta)} 1_{\mathcal{A}_n}\right] \\
&\leq \mathrm{E}_{\beta_0}\left[\frac{\int_{\beta : \|\beta-\beta_0\| < s_n} \frac{p_\beta}{p_{\beta_0}}(Y)\, d\Pi_n(\beta)}{e^{-r_n^2} \Pi_n(\beta : \|\beta - \beta_0\| \leq r_n)}\right] \\
&\leq \frac{\Pi_n(\beta : \|\beta - \beta_0\| < s_n)}{e^{-r_n^2} \Pi_n(\beta : \|\beta - \beta_0\| \leq r_n)} \to 0.
\end{aligned}$$

Which concludes the proof. □

Proof of the Main Theorem

Assume that $\beta_0 = 0$. We will use lemma (7.1) to prove this theorem. We will need to show that for proper sequences r_n and s_n,

$$\frac{\Pi_n(\beta : \|\beta\|_2 < s_n)}{\Pi_n(\beta : \|\beta\|_2 < r_n)} = o\big(e^{-r_n^2}\big).$$

- Under the prior, $|\beta_i| \sim \text{Exp}(\lambda)$ and $\|\beta\|_1 \sim \text{Gamma}(\lambda, n)$. Thus,

$$\begin{aligned}\Pi_n(\beta : \|\beta\|_2 < s_n) &\le \Pi_n(\beta : \|\beta\|_1 < s_n\sqrt{n}) \\ &= \int_0^{\sqrt{n}s_n} \lambda^n u^{n-1} \frac{e^{-\lambda u}}{\Gamma(n)} du \le \frac{(\lambda\sqrt{n}s_n)^n}{\Gamma(n+1)}.\end{aligned}$$

- We can write,

$$\Pi_n(\beta : \|\beta\|_2 < r_n) = \left(\frac{\lambda}{2}\right)^n \int_{\|b\|_2 \le r_n} e^{-\lambda\|b\|_1} \mathrm{d}b \ge \left(\frac{\lambda}{2}\right)^n e^{-\lambda\sqrt{n}r_n} v_n r_n^n,$$

  where v_n is the area of unit l_2 ball in $\mathbb{R}^n$ and the last inequality follows from $\|\beta\|_1 \le \sqrt{n}\|\beta\|_2$.

By substituting $v_n = \pi^{\frac{n}{2}}/\Gamma(1+\frac{n}{2})$, and last two inequalities,

$$\frac{\Pi(\|\beta\|_2 < s_n)}{\Pi(\|\beta\|_2 < r_n)} \leq \Big(\frac{2}{\sqrt{\pi}}\Big)^n \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(n+1)} n^{n/2} \exp\Big(\lambda\sqrt{n} r_n - n\log\big(\frac{r_n}{s_n}\big)\Big)$$
$$\leq \exp\left(\lambda\sqrt{n} r_n - n\log(c\frac{r_n}{s_n})\right).$$

Set $r_n = \sqrt{n}\big(\max(\lambda^{-1}, 1)\big)$ and $s_n = \delta r_n$. Under these assumptions, the conditions of lemma (7.1) hold and we have:

$$\mathrm{E}_{\beta_0}\Pi_n(\beta : \|\beta - \beta_0\| < s_n|Y,) \to 0,$$

equivalently,

$$\mathbb{E}_{\beta^0=0}\Pi_{\lambda_n}\left(\beta : \|\beta\|_2 \leq \delta\sqrt{n}\left(\frac{1}{\lambda_n} \wedge 1\right)\Big| Y\right) \to 0.$$

Which concludeds the proof of the main theorem. □

References

CSV15 Ismaël Castillo, Johannes Schmidt-Hieber, and Aad van der Vaart,
Bayesian Linear Regression with Sparse Priors, *Annals of Statistics*, 2015.

CV12 Ismaël Castillo, and Aad van der Vaart,
Needles and Straw in a Haystack: Posterior Concentration for Possibly Sparse Sequences, *Annals of Statistics*, 2012.