*Biometrika* (2013), **xx**, x, *pp.* 1–9

*Printed in Great Britain*

# Posterior consistency in linear models under shrinkage priors

By A. ARMAGAN
*SAS Institute Inc., Cary, North Carolina 27513, USA*
artin.armagan@sas.com

and D. B. DUNSON
*Department of Statistical Science, Duke University, Durham, North Carolina 27708, USA*
dunson@stat.duke.edu

and J. LEE
*Department of Statistics, Seoul National University, Seoul, 151-747, Korea*
leejyc@gmail.com

and W. U. BAJWA
*Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854, USA*
waheed.bajwa@rutgers.edu

and N. STRAWN
*Department of Mathematics, Duke University, Durham, North Carolina 27708, USA*
nstrawn@math.duke.edu

## Summary

We investigate the asymptotic behavior of posterior distributions of regression coefficients in high-dimensional linear models as the number of dimensions grows with the number of observations. We show that the posterior distribution concentrates in neighborhoods of the true parameter under simple sufficient conditions. These conditions hold under popular shrinkage priors given some sparsity assumptions.

*Some key words*: Bayesian Lasso; Generalized double Pareto prior; Heavy tails; High-dimensional data; Horseshoe prior; Posterior consistency; Shrinkage estimation.

## 1. Introduction

Consider the linear model $y_n = X_n\beta_n^0 + \varepsilon_n$, where $y_n$ is an $n$-dimensional vector of responses, $X_n$ is the $n \times p_n$ design matrix, $\varepsilon_n \sim \mathrm{N}\left(0, \sigma^2 I_n\right)$ with known $\sigma^2$, and some of the components of $\beta_n^0$ are zero. Let $\mathcal{A}_n = \{j : \beta_{nj}^0 \neq 0, j = 1, \ldots, p_n\}$ and $|\mathcal{A}_n| = q_n$ denote the set of indices and number of nonzero elements in $\beta_n^0$.

In studying the behavior of regression methods in high-dimensional settings, it is increasingly common to allow the number of candidate predictors $p_n$ to grow with sample size $n$. This is realistic in many applications. In genomics the number of predictors tends to be larger by design for studies with more subjects. In collecting single nucleotide polymorphisms, gene expression, proteomics and so on, one can obtain an immense number of candidate predictors. However, when $n$ is small, attempting to measure and include all such predictors in the statistical analysis seems unreasonable, so that one tends to collect and analyze increasing subsets of an effectively unbounded number of candidate predictors as sample size increases. In such applications, we are often interested in inferences on the model parameters as much as building a predictive model in order to understand the associations between the response and the candidate predictors.

Our setup is not new, and we follow Ghosal (1999) who also focused on asymptotic properties of the posterior on the regression coefficients assuming known $\sigma^2$ and growing $p_n$. The increasing $p_n$ paradigm induces some challenges relative to the traditional literature on posterior consistency in that growing dimension of $\beta_n^0$ results in a changing $\ell_2$ neighborhood around $\beta_n^0$. This makes it more challenging to show that the posterior assigns all such neighborhoods probability converging to one. One way to bypass this issue is to focus on the predictive distribution of $y_n$ given $X_n$ as in Jiang (2007). However, this does not address the common interest in inferences on the regression coefficients. Ghosal (1999) and Bontemps (2011) provide results on asymptotic normality of the posteriors in linear models for $p_n^4 \log p_n = o(n)$ and $p_n \leq n$, respectively. As a corollary, Ghosal (1999) states posterior consistency results in linear models when $p_n^3 \log n/n \to 0$ under the usual assumptions on $X_n$. However, both Ghosal (1999) and Bontemps (2011) require Lipschitz conditions ensuring that the prior is sufficiently flat in a neighborhood of the true $\beta_n^0$. Such conditions are restrictive when using shrinkage priors that are designed to concentrate on sparse $\beta_n$ vectors.

Our main contribution is providing a simple sufficient condition on the prior concentration to achieve the desired asymptotic posterior behavior when $p_n = o(n)$. Our particular focus is on shrinkage priors, including the Laplace, Student's $t$, generalized double Pareto, and horseshoe-type priors (Johnstone & Silverman, 2004; Carvalho et al., 2010; Armagan et al., 2011, 2013). There is a rich methodological and applied literature supporting such priors but a lack of theoretical results.

## 2. Sufficient Conditions for Posterior Consistency

Our results on posterior consistency rely on the following assumptions as $n \to \infty$:

(A1) Let $p_n = o(n)$;
(A2) Let $\Lambda_{n\min}$ and $\Lambda_{n\max}$ be the smallest and the largest singular values of $X_n$, respectively. Then $0 < \Lambda_{\min} < \liminf_{n\to\infty} \Lambda_{n\min}/\sqrt{n} \leq \limsup_{n\to\infty} \Lambda_{n\max}/\sqrt{n} < \Lambda_{\max} < \infty$;
(A3) Let $\sup_{j=1,\ldots,p_n} |\beta_{nj}^0| < \infty$;
(A4) Let $q_n = o\{n^{1-\rho/2}/(\sqrt{p_n} \log n)\}$ for $\rho \in (0, 2)$;
(A5) Let $q_n = o(n/\log n)$.

Assumptions (A4) and (A5) will be used in different settings.

Lemma 1. *Let $\mathcal{B}_n := \{\beta_n : \|\beta_n - \beta_n^0\| > \epsilon\}$ where $\epsilon > 0$. To test $H_0 : \beta_n = \beta_n^0$ vs $H_1 : \beta_n \in \mathcal{B}_n$, we define a test function $\Phi_n(y_n) = I(y_n \in \mathcal{C}_n)$ where the critical region is $\mathcal{C}_n :=$*

$\{y_n : \|\hat{\beta}_n - \beta_n^0\| > \epsilon/2\}$ *and* $\hat{\beta}_n = (X_n^T X_n)^{-1} X_n^T y_n$. *Then, under assumptions (A1) and (A2), as* $n \to \infty$,

1. $E_{\beta_n^0}(\Phi_n) \le \exp\{-\epsilon^2 n \Lambda_{\min}^2/(16\sigma^2)\}$,
2. $\sup_{\beta_n \in \mathcal{B}_n} E_{\beta_n}(1 - \Phi_n) \le \exp\{-\epsilon^2 n \Lambda_{\min}^2/(16\sigma^2)\}$.

THEOREM 1. *Given Lemma 1, the posterior of* $\beta_n$ *under prior* $\Pi_n(\beta_n)$ *is strongly consistent, that is, for any* $\epsilon > 0$, $\Pi_n(\mathcal{B}_n|y_n) = \Pi_n(\beta_n : ||\beta_n - \beta_n^0|| > \epsilon|y_n) \to 0$ $pr_{\beta_n^0}$*–almost surely as* $n \to \infty$, *if*

$$\Pi_n\left(\beta_n : \|\beta_n - \beta_n^0\| < \frac{\Delta}{n^{\rho/2}}\right) > \exp(-dn)$$

*for all* $0 < \Delta < \epsilon^2 \Lambda_{\min}^2/(48\Lambda_{\max}^2)$ *and* $0 < d < \epsilon^2 \Lambda_{\min}^2/(32\sigma^2) - 3\Delta\Lambda_{\max}^2/(2\sigma^2)$ *and some* $\rho > 0$.

Theorem 1 provides a simple sufficient condition on the concentration of the prior around sparse $\beta_n^0$. We use Theorem 1 to provide conditions on $\beta_n^0$ under which specific shrinkage priors achieve posterior consistency focusing on priors that assume independent and identically distributed elements of $\beta_n$.

### 2·1. *Laplace Prior*

THEOREM 2. *Under assumptions (A1)–(A4), the Laplace prior* $f(\beta_{nj}|s_n) = (1/2s_n)\exp(-|\beta_{nj}|/s_n)$ *with scale parameter* $s_n$ *yields a strongly consistent posterior if* $s_n = C/(\sqrt{p_n} n^{\rho/2} \log n)$ *for finite* $C > 0$.

### 2·2. *Student's t Prior*

The density function for the scaled Student's $t$ distribution is

$$f(\beta_j|s, d_0) = \frac{1}{s\sqrt{d_0}\mathrm{B}(1/2, d_0/2)}\left(1 + \frac{\beta_j^2}{s^2 d_0}\right)^{-(d_0+1)/2},$$

with scale $s$, degrees of freedom $d_0$, and $\mathrm{B}(\cdot)$ denoting the beta function.

THEOREM 3. *Under assumptions (A1)–(A3) and (A5), the scaled Student's t prior with parameters* $s_n$ *and* $d_{0n}$ *yields a strongly consistent posterior if* $d_{0n} = d_0 \in (2, \infty)$ *and* $s_n = C/(\sqrt{p_n} n^{\rho/2} \log n)$ *for finite* $\rho > 0$ *and* $C > 0$.

### 2·3. *Generalized Double Pareto Prior*

As defined by Armagan et al. (2013), the generalized double Pareto density is given by

$$f(\beta_j|\alpha, \eta) = \frac{\alpha}{2\eta}\left(1 + \frac{|\beta_j|}{\eta}\right)^{-(\alpha+1)}, \quad \alpha, \eta > 0.$$

THEOREM 4. *Under assumptions (A1)–(A3) and (A5), the generalized double Pareto prior with parameters* $\alpha_n$ *and* $\eta_n$ *yields a strongly consistent posterior if* $\alpha_n = \alpha \in (2, \infty)$ *and* $\eta_n = C/(\sqrt{p_n} n^{\rho/2} \log n)$ *for finite* $\rho > 0$ *and* $C > 0$.

### 2·4. *Horseshoe-like Priors*

As defined in Armagan et al. (2011), generalized beta scale mixtures of normals are obtained by the following three equivalent representations:

$$\beta_j \sim \mathrm{N}(0, 1/\varrho_j - 1), f(\varrho_j) = \frac{\Gamma(a_0 + b_0)}{\Gamma(a_0)\Gamma(b_0)} \xi^{b_0} \varrho_j^{b_0 - 1} (1 - \varrho_j)^{a_0 - 1} \{1 + (\xi - 1)\varrho_j\}^{-(a_0+b_0)} \quad (1)$$

$$\beta_j \sim \mathrm{N}(0, \tau_j), \tau_j \sim \mathrm{Ga}(a_0, \lambda_j), \lambda_j \sim \mathrm{Ga}(b_0, \xi)$$

$$\beta_j \sim \mathrm{N}(0, \tau_j), f(\tau_j) = \frac{\Gamma(a_0 + b_0)}{\Gamma(a_0)\Gamma(b_0)} \xi^{-a_0} \tau^{a_0 - 1} (1 + \tau_j/\xi)^{-(a_0+b_0)}$$

where $a_0, b_0, \xi > 0$. Due to the representation in (1) and the work by Carvalho et al. (2010), we refer to these priors as *horseshoe-like*. The above formulation yields a general family that covers special cases discussed in Johnstone & Silverman (2004), a technical report by Griffin & Brown (2007) and Carvalho et al. (2010). The resulting marginal density on $\beta_j$ is

$$f(\beta_j | a_0, b_0, \xi) = \frac{\Gamma(b_0 + 1/2)\Gamma(a_0 + b_0)\mathrm{U}\{b_0 + 1/2, 3/2 - a_0, \beta_j^2/(2\xi)\}}{(2\pi\xi)^{1/2}\Gamma(a_0)\Gamma(b_0)}, \quad (2)$$

where $\mathrm{U}(\cdot)$ denotes the confluent hypergeometric function of the second kind.

THEOREM 5. *Under assumptions (A1)–(A3) and (A5), the prior in (2) with parameters $a_{0n} = a_0 \in (0, \infty)$, $b_{0n} = b_0 \in (1, \infty)$ and $\xi_n$ yields a strongly consistent posterior if $\xi_n = C/(p_n n^\rho \log n)$ for finite $\rho > 0$ and $C > 0$.*

## 3. FINAL REMARKS

Our analysis is heavily dependent on the construction of good tests. Results can be extended utilizing appropriate tests relying on an estimator with asymptotically vanishing probability of being outside of a *shrinking* neighborhood of the truth. For instance, one could use results similar to Bickel et al. (2009) given additional conditions on $X_n$. Theorem 7.2 of Bickel et al. (2009) states that

$$\mathrm{pr}_{\beta_n^0}\left( \|\hat{\beta}_{nL} - \beta_n^0\|_2^2 > M \frac{a_n \log p_n}{n} \right) \leq p_n^{1 - a_n^2/8} \quad (3)$$

for $a_n > 2\sqrt{2}$ and for some $M > 0$, where $\hat{\beta}_{nL}$ denotes the Lasso estimator. Hence using (3), in a similar fashion to Lemma 1, we can obtain consistent tests with an $\epsilon$-neighborhood contracting at a rate $\mathcal{O}\{(a_n \log p_n)^{1/2}/\sqrt{n}\}$. Assuming $q_n < \infty$ for simplicity and letting $a_n = \mathcal{O}(\log n)$, following Theorems 1, 3, 4 and 5, we anticipate that under the Student's $t$, generalized double Pareto and horseshoe-like priors, a *near-optimal* contraction rate of $\mathcal{O}\{(\log n \log p_n)^{1/2}/\sqrt{n}\}$ is possible.

As in almost all of the Bayesian asymptotic literature, we have focused on sufficient conditions. Our conditions are practically appealing in allowing priors to be screened for their usefulness in high-dimensional settings. However, it would be of substantial interest to additionally provide theory allowing one to rule out the use of certain classes of priors in particular settings.

## 4. Technical Details

*Proof of Lemma 1.* Noting that $\hat{\beta}_n = (X_n^{\mathrm{T}} X_n)^{-1} X_n^{\mathrm{T}} y_n$, $E_{\beta_n^0}(\Phi_n) = \mathrm{pr}_{\beta_n^0}(\|\hat{\beta}_n - \beta_n^0\| > \epsilon/2) \leq \mathrm{pr}_{\beta_n^0}\{\chi^2_{p_n} > \epsilon^2 n \Lambda^2_{\min}/(4\sigma^2)\}$ where $\chi^2_p$ is a chi-squared distributed random variable with $p$ degrees of freedom. The inequality is attained using assumption (A2). Similarly, $\sup_{\beta_n \in \mathcal{B}_n} E_{\beta_n}(1 - \Phi_n) \leq \sup_{\beta_n \in \mathcal{B}_n} \mathrm{pr}_{\beta_n}(|\|\hat{\beta}_n - \beta_n\| - \|\beta_n^0 - \beta_n\|| \leq \epsilon/2) \leq \sup_{\beta_n \in \mathcal{B}_n} \mathrm{pr}_{\beta_n}(\|\hat{\beta}_n - \beta_n\| \geq -\epsilon/2 + \|\beta_n^0 - \beta_n\|) = \mathrm{pr}_{\beta_n}(\|\hat{\beta}_n - \beta_n\| \geq \epsilon/2) \leq \mathrm{pr}_{\beta_n^0}\{\chi^2_{p_n} > \epsilon^2 n \Lambda^2_{\min}/(4\sigma^2)\}$. Simplifying the inequality $\mathrm{pr}\{\chi^2_p - p \geq 2(px)^{1/2} + 2x\} \leq \exp(-x)$ by Laurent & Massart (2000), we state that $\mathrm{pr}(\chi^2_p \geq x) \leq \exp(-x/4)$ if $x \geq 8p$. Then, using assumption (A1), as $n \to \infty$,

$$E_{\beta_n^0}(\Phi_n) \leq \exp\{-\epsilon^2 n \Lambda^2_{\min}/(16\sigma^2)\},$$
$$\sup_{\beta_n \in \mathcal{B}_n} E_{\beta_n}(1 - \Phi_n) \leq \exp\{-\epsilon^2 n \Lambda^2_{\min}/(16\sigma^2)\}.$$

This completes the proof. □

*Proof of Theorem 1.* Our proof relies on a technique originally devised by Schwartz (1965). The posterior probability of $\mathcal{B}_n$ is given by

$$\begin{aligned}
\Pi_n(\mathcal{B}_n | y_n) &= \frac{\int_{\mathcal{B}_n} \{f(y_n|\beta_n)/f(y_n|\beta_n^0)\}\Pi(d\beta_n)}{\int \{f(y_n|\beta_n)/f(y_n|\beta_n^0)\}\Pi(d\beta_n)} \\
&\leq \Phi_n + \frac{(1-\Phi_n)J_{\mathcal{B}_n}}{J_n} \\
&= I_1 + I_2/J_n,
\end{aligned}$$

where $J_{\mathcal{B}_n} = \int_{\mathcal{B}_n} \{f(y_n|\beta_n)/f(y_n|\beta_n^0)\}\Pi(d\beta_n)$ and $J_n = J_{\Re^{p_n}}$. We need to show that $I_1 + I_2/J_n \to 0$ $\mathrm{pr}_{\beta_n^0}$–almost surely as $n \to \infty$. Let $b = \epsilon^2 \Lambda^2_{\min}/(16\sigma^2)$. For sufficiently large $n$, $\mathrm{pr}_{\beta_n^0}\{I_1 \geq \exp(-bn/2)\} \leq \exp(bn/2) E_{\beta_n^0}(I_1) = \exp(-bn/2)$ using Lemma 1. This implies that $\sum_{n=1}^{\infty} \mathrm{pr}_{\beta_n^0}\{I_1 \geq \exp(-bn/2)\} < \infty$ and hence by the Borel–Cantelli lemma $\mathrm{pr}_{\beta_0}\{I_1 \geq \exp(-bn/2)$ infinitely often$\} = 0$. We next look at the behavior of $I_2$:

$$\begin{aligned}
E_{\beta_n^0}(I_2) &= E_{\beta_n^0}\{(1-\Phi_n)J_{\mathcal{B}_n}\} \\
&= E_{\beta_n^0}\left\{(1-\Phi_n)\int_{\mathcal{B}_n} \frac{f(y_n|\beta_n)}{f(y_n|\beta_n^0)}\Pi_n(d\beta_n)\right\} \\
&= \int_{\mathcal{B}_n}\int (1-\Phi_n) f(y_n|\beta_n) dy_n \Pi_n(d\beta_n) \\
&\leq \Pi_n(\mathcal{B}_n) \sup_{\beta_n \in \mathcal{B}_n} E_{\beta_n}(1-\Phi_n) \\
&\leq \exp(-bn)
\end{aligned}$$

Then for sufficiently large $n$, $\mathrm{pr}_{\beta_n^0}\{I_2 \geq \exp(-bn/2)\} \leq \exp(-bn/2)$ using Lemma 1. Again $\sum_{n=1}^{\infty} \mathrm{pr}_{\beta_n^0}\{I_2 \geq \exp(-bn/2)\} < \infty$ and hence by the Borel–Cantelli lemma $\mathrm{pr}_{\beta_0}\{I_2 \geq \exp(-bn/2)$ infinitely often$\} = 0$.

We have shown that both $I_1$ and $I_2$ tend towards zero exponentially fast. Now we analyze the behavior of $J_n$. To complete the proof, we need to show that $\exp(bn/2)J_n \to$

$\infty$ $\text{pr}_{\beta_n^0}$–almost surely as $n \to \infty$.

$$\exp(bn/2)J_n = \exp(bn/2)\int \exp\left\{-n\frac{1}{n}\log\frac{f(y_n|\beta_n^0)}{f(y_n|\beta_n)}\right\}\Pi_n(d\beta_n)$$
$$\geq \exp\{(b/2-\nu)n\}\Pi_n(\mathcal{D}_{n,\nu}) \quad (4)$$

where $\mathcal{D}_{n,\nu} = \{\beta_n : n^{-1}\log\{f(y_n|\beta_n^0)/f(y_n|\beta_n)\} < \nu\} = \{\beta_n : n^{-1}(\|y_n - X_n\beta_n\|^2 - \|y_n - X_n\beta_n^0\|^2) < 2\sigma^2\nu\}$ for any $0 < \nu < b/2$. Then $\Pi_n(\mathcal{D}_{n,\nu}) \geq \Pi_n\{\beta_n : n^{-1}|\|y_n - X_n\beta_n\|^2 - \|y_n - X_n\beta_n^0\|^2| < 2\sigma^2\nu\}$. Using the identity $x^2 - x_0^2 = 2x_0(x-x_0) + (x-x_0)^2$ for all $x, x_0 \in \Re$,

$$\Pi_n(\mathcal{D}_{n,\nu}) \geq \Pi_n\left\{\beta_n : n^{-1}\left|2\|y_n - X_n\beta_n^0\|(\|y_n - X_n\beta_n\| - \|y_n - X_n\beta_n^0\|)\right.\right.$$
$$\left.\left. + (\|y_n - X_n\beta_n\| - \|y_n - X_n\beta_n^0\|)^2\right| < 2\sigma^2\nu\right\}$$
$$\geq \Pi_n\left\{\beta_n : n^{-1}(2\|y_n - X_n\beta_n^0\|\|X_n\beta_n - X_n\beta_n^0\| + \|X_n\beta_n - X_n\beta_n^0\|^2) < 2\sigma^2\nu\right\}$$
$$\geq \Pi_n\left(\beta_n : n^{-1}\|X_n\beta_n - X_n\beta_n^0\| < \frac{2\sigma^2\nu}{3\kappa_n}, \|X_n\beta_n - X_n\beta_n^0\| < \kappa_n\right) \quad (5)$$

given that $\|y_n - X_n\beta_n^0\| \leq \kappa_n$. For $\kappa_n = n^{(1+\rho)/2}$ with $\rho > 0$ and $\kappa_n^2/\sigma^2 \geq 8n$, $\text{pr}_{\beta_n^0}(y_n : \|y_n - X_n\beta_n^0\|^2 > \kappa_n^2) = \text{pr}_{\beta_n^0}(y_n : \chi_n^2 > \kappa_n^2/\sigma^2) \leq \exp\{-\kappa_n^2/(4\sigma^2)\}$. Since $\sum_{n=1}^{\infty}\text{pr}_{\beta_n^0}(y_n : \|y_n - X_n\beta_n^0\| > \kappa_n) < \infty$, by the Borel–Cantelli lemma $\text{pr}_{\beta_n^0}(y_n : \|y_n - X_n\beta_n^0\| > \kappa_n$ infinitely often$) = 0$. Following from (5) and the fact that $\kappa_n \to \infty$, as $n \to \infty$, for sufficiently large $n$, $\Pi_n(\mathcal{D}_{n,\nu}) \geq \Pi_n\{\beta_n : n^{-1}\|X_n\beta_n - X_n\beta_n^0\| < 2\sigma^2\nu/(3\kappa_n)\} \geq \Pi_n(\beta_n : \|\beta_n - \beta_n^0\| < \Delta/n^{\rho/2})$, where $\Delta = 2\sigma^2\nu/(3\Lambda_{\max})$. Hence following (4), $\Pi_n(\mathcal{B}_n|y_n) \to 0$ $\text{pr}_{\beta_n^0}$–almost surely as $n \to \infty$ if $\Pi_n(\beta_n : \|\beta_n - \beta_n^0\| < \Delta/n^{\rho/2}) > \exp(-dn)$ for all $0 < d < b/2 - \nu$. This completes the proof. □

*Proof of Theorem 2.* We need to calculate the probability assigned to the region $\{\beta_n : \|\beta_n - \beta_n^0\| < \Delta/n^{\rho/2}\}$ under the Laplace prior.

$$\Pi_n\left(\beta_n : \|\beta_n - \beta_n^0\| < \frac{\Delta}{n^{\rho/2}}\right) = \Pi_n\left\{\beta_n : \sum_{j\in\mathcal{A}_n}(\beta_{nj} - \beta_{nj}^0)^2 + \sum_{j\notin\mathcal{A}_n}\beta_{nj}^2 < \frac{\Delta^2}{n^\rho}\right\}$$
$$\geq \prod_{j\in\mathcal{A}_n}\left\{\Pi_n\left(\beta_{nj} : |\beta_{nj} - \beta_{nj}^0| < \frac{\Delta}{\sqrt{p_n}n^{\rho/2}}\right)\right\}$$
$$\times\Pi_n\left\{\beta_n^{j\notin\mathcal{A}} : \sum_{j\notin\mathcal{A}_n}\beta_{nj}^2 < \frac{(p_n - q_n)\Delta^2}{p_n n^\rho}\right\}$$
$$\geq \prod_{j\in\mathcal{A}_n}\left\{\Pi_n\left(\beta_{nj} : |\beta_{nj} - \beta_{nj}^0| < \frac{\Delta}{\sqrt{p_n}n^{\rho/2}}\right)\right\}\left\{1 - \frac{p_n n^\rho E\left(\sum_{j\notin\mathcal{A}_n}\beta_{nj}^2\right)}{(p_n - q_n)\Delta^2}\right\} \quad (6)$$

where $E(\beta_{nj}^2)$ can verified to be $2s_n^2$. Following from (6)

$$\Pi_n\left(\beta_n : \|\beta_n - \beta_n^0\| < \frac{\Delta}{n^{\rho/2}}\right) \geq$$
$$\left\{\frac{\Delta}{\sqrt{p_n}n^{\rho/2}s_n}\exp\left(-\frac{\sup_{j\in\mathcal{A}_n}|\beta_{nj}^0|}{s_n} - \frac{\Delta}{s_n\sqrt{p_n}n^{\rho/2}}\right)\right\}^{q_n}\left(1 - \frac{2p_n n^\rho s_n^2}{\Delta^2}\right). \quad (7)$$

Taking the negative logarithm of both sides of (7) and letting $s_n = C/(\sqrt{p_n} n^{\rho/2} \log n)$ for some $C > 0$, we obtain

$$-\log \Pi_n \left( \beta_n : \|\beta_n - \beta_n^0\| < \frac{\Delta}{n^{\rho/2}} \right) \leq -q_n \log \Delta + q_n \log C - q_n \log\log n$$
$$-\log \left\{ 1 - \frac{2C^2}{\Delta^2 (\log n)^2} \right\} + \frac{q_n \Delta \log n}{C} + \frac{q_n \sqrt{p_n} n^{\rho/2} \log n \sup_{j \in \mathcal{A}_n} |\beta_{nj}^0|}{C} \tag{8}$$

as $n \to \infty$. It is easy to see that the dominating term in (8) is the last one and $-\log \Pi_n(\beta_n : \|\beta_n - \beta_n^0\| < \Delta/n^{\rho/2}) < dn$ for all $d > 0$. This completes the proof. □

*Proof of Theorem 3.* $E(\beta_{nj}^2)$, in this case, is given by $d_0 s_n^2/(d_0 - 2)$. For the sake of simplicity, we let $d_0 = 3$. Then following from (6)

$$\Pi_n \left( \beta_n : \|\beta_n - \beta_n^0\| < \frac{\Delta}{n^{\rho/2}} \right) \geq \left( 1 - \frac{3 p_n n^\rho s_n^2}{\Delta^2} \right)$$
$$\times \left[ \frac{2\Delta}{\sqrt{p_n} n^{\rho/2} s_n \sqrt{3} \mathrm{B}(1/2, 3/2)} \left\{ 1 + \frac{2 \sup_{j \in \mathcal{A}_n} (\beta_{nj}^0)^2}{3 s_n^2} + \frac{2\Delta^2}{3 s_n^2 p_n n^\rho} \right\}^{-2} \right]^{q_n}. \tag{9}$$

Taking the negative logarithm of both sides of (9) and letting $s_n = C/(\sqrt{p_n} n^{\rho/2} \log n)$ for some $C > 0$, we obtain

$$-\log \Pi_n \left( \beta_n : \|\beta_n - \beta_n^0\| < \frac{\Delta}{n^{\rho/2}} \right) \leq q_n \log \left\{ \frac{\sqrt{3} C \mathrm{B}(1/2, 3/2)}{2\Delta} \right\} - q_n \log\log n$$
$$-\log \left\{ 1 - \frac{C^2}{\Delta^2 (\log n)^2} \right\} + 2 q_n \log \left\{ 1 + \frac{2 p_n n^\rho \log n \sup_{j \in \mathcal{A}_n} (\beta_{nj}^0)^2}{3C^2} + \frac{2\Delta^2 (\log n)^2}{3C^2} \right\} \tag{10}$$

as $n \to \infty$. It is easy to see that the dominating term in (10) is the last one and $-\log \Pi_n(\beta_n : \|\beta_n - \beta_n^0\| < \Delta/n^{\rho/2}) < dn$ for all $d > 0$. The result can be easily shown to hold for all $d_0 \in (2, \infty)$. This completes the proof. □

*Proof of Theorem 4.* $E(\beta_{nj}^2)$, in this case, can verified to be $2\eta_n^2/(\alpha^2 - 3\alpha + 2)$ for $\alpha > 2$. For the sake of simplicity, we let $\alpha = 3$. Then following from (6)

$$\Pi_n \left( \beta_n : \|\beta_n - \beta_n^0\| < \frac{\Delta}{n^{\rho/2}} \right) \geq$$
$$\left\{ \frac{3\Delta}{\sqrt{p_n} n^{\rho/2} \eta_n} \left( 1 + \frac{\sup_{j \in \mathcal{A}_n} |\beta_{nj}^0|}{\eta_n} + \frac{\Delta}{\eta_n \sqrt{p_n} n^{\rho/2}} \right)^{-4} \right\}^{q_n} \left( 1 - \frac{p_n n^\rho \eta^2}{\Delta^2} \right). \tag{11}$$

Taking the negative logarithm of both sides of (11) and letting $\eta_n = C/(\sqrt{p_n} n^{\rho/2} \log n)$ for some $C > 0$, we obtain

$$-\log \Pi_n \left( \beta_n : \|\beta_n - \beta_n^0\| < \frac{\Delta}{n^{\rho/2}} \right) \leq -q_n \log 3\Delta - 3 q_n \log C - q_n \log\log n$$
$$-\log \left\{ 1 - \frac{C^2}{\Delta^2 (\log n)^2} \right\} + 4 q_n \log \left( C + \Delta \log n + \sqrt{p_n} n^{\rho/2} \log n \sup_{j \in \mathcal{A}_n} |\beta_{nj}^0| \right) \tag{12}$$

as $n \to \infty$. It is easy to see that the dominating term in (12) is the last one and $-\log \Pi_n(\beta_n : \|\beta_n - \beta_n^0\| < \Delta/n^{\rho/2}) < dn$ for all $d > 0$. The result can be easily shown to hold for all $\alpha \in (2, \infty)$. This completes the proof. □

*Proof of Theorem 5.* Similarly to the previous cases, we can show that $E(\beta_{nj}^2) = \xi_n \Gamma(a_0+1)\Gamma(b_0-1)/\{\Gamma(a_0)\Gamma(b_0)\}$. Then following from (6)

$$\Pi_n\left(\beta_n : \|\beta_n - \beta_n^0\| < \frac{\Delta}{n^{\rho/2}}\right) \geq \left\{1 - \frac{p_n n^\rho E(\beta_{nj}^2)}{\Delta^2}\right\}\left(\frac{2\Delta}{\sqrt{p_n}n^{\rho/2}}\right)^{q_n}$$
$$\times\left[\frac{\mathrm{U}\{b_0+1/2, 3/2-a_0, \sup_{j\in\mathcal{A}_n}(\beta_{nj}^0)^2/\xi_n + \Delta/(p_n n^\rho \xi_n)\}}{(2\pi\xi_n)^{1/2}\Gamma(a_0)\Gamma(b_0)\Gamma(b_0+1/2)^{-1}\Gamma(a_0+b_0)^{-1}}\right]^{q_n}. \quad (13)$$

We can use the expansion $\mathrm{U}(a,b,z) = z^{-a}\{\sum_{m=0}^{R-1}(a)_m(1+a-b)_m(-z)^m/m! + \mathcal{O}(|z|^{-R})\}$ for large $z$, where $(a)_m = a(a+1)\ldots(a+m-1)$ and $R$th term is the smallest in the expansion (Abramowitz & Stegun, 1972). Letting $R = 1$, for sufficiently large $n$, (13) can be further bounded as

$$\Pi_n\left(\beta_n : \|\beta_n - \beta_n^0\| < \frac{\Delta}{n^{\rho/2}}\right) > \left\{1 - \frac{p_n n^\rho E(\beta_{nj}^2)}{\Delta^2}\right\}$$
$$\times\left[\frac{\sqrt{2}\Delta\Gamma(b_0+1/2)\Gamma(a_0+b_0)}{\sqrt{p_n}n^{\rho/2}\sqrt{\xi_n}\sqrt{\pi}\Gamma(a_0)\Gamma(b_0)\{\sup_{j\in\mathcal{A}_n}(\beta_{nj}^0)^2/\xi_n + \Delta/(p_n n^\rho \xi_n)\}^{(b_0+1/2)}}\right]^{q_n}. \quad (14)$$

Taking the negative logarithm of both sides of (14) and letting $\xi_n = C/(p_n n^\rho \log n)$ for some $C > 0$, we obtain

$$-\log \Pi_n\left(\beta_n : \|\beta_n - \beta_n^0\| < \frac{\Delta}{n^{\rho/2}}\right) <$$
$$-q_n \log\left\{\frac{\sqrt{2}\Delta\Gamma(b_0+1/2)\Gamma(a_0+b_0)}{\sqrt{C}\sqrt{\pi}\Gamma(a_0)\Gamma(b_0)}\right\} - \log\left\{1 - \frac{C\Gamma(a_0+1)\Gamma(b_0-1)}{\log n\Delta\Gamma(a_0)\Gamma(b_0)}\right\}$$
$$-\frac{q_n}{2}\log\log n + q_n\left(b_0 + \frac{1}{2}\right)\log\left\{\frac{p_n n^\rho \log n \sup_{j\in\mathcal{A}_n}(\beta_{nj}^0)^2}{C} + \frac{\Delta\log n}{C}\right\} \quad (15)$$

as $n \to \infty$. It is easy to see that the dominating term in (15) is the last one and $-\log \Pi_n(\beta_n : \|\beta_n - \beta_n^0\| < \Delta/n^{\rho/2}) < dn$ for all $d > 0$. This completes the proof. □

## Acknowledgements

This work was supported by the National Institute of Environmental Health Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Environmental Health Sciences or the National Institutes of Health. Jaeyong Lee was supported by Advanced Research Center Program (S/ERC), the National Research Foundation of Korea grant funded by the Korean government (MSIP). Waheed U. Bajwa was supported in part by the NSF. Nate Strawn was supported by DARPA Mathematics of Sensing, Exploitation, and Execution (MSEE) program (managed by Dr. Tony Falcone).

## References


Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* New York: Dover.

Armagan, A., Dunson, D. B. & Clyde, M. (2011). Generalized beta mixtures of Gaussians. *Advances in Neural Information Processing Systems (NIPS)* .

Armagan, A., Dunson, D. B. & Lee, J. (2013). Generalized double Pareto shrinkage. *Statistica Sinica* **23**, 119–143.

Bickel, P. J., Ritov, Y., Alexandre & Tsybakov, B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Annals of Statistics* **37**, 1705–1732.

Bontemps, D. (2011). Bernstein–von Mises theorems for Gaussian regression with increasing number of regressors. *Annals of Statistics* **39**, 2557–2584.

Carvalho, C. M., Polson, N. G. & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97**, 465–480.

Ghosal, S. (1999). Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli* **5**, 315–331.

Jiang, W. (2007). Bayesian variable selection for high dimensional generalized linear models: Convergence rates of the fitted densities. *The Annals of Statistics* **35**, 1487–1511.

Johnstone, I. M. & Silverman, B. W. (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Annals of Statistics* **32**, 1594–1649.

Laurent, B. & Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics* **28**, 1302–1338.

Schwartz, L. (1965). On Bayes procedures. *Zeitschrift für wahrscheinlichkeitstheorie und verwandte gebiete* **4**, 10–26.