High-dimensional properties for empirical priors in linear regression with unknown error variance

Xiao Fang\textsuperscript{1}* and Malay Ghosh\textsuperscript{1}

Department of Statistics, University of Florida, Griffin-Floyd Hall, Gainesville, 32611, FL, USA.

*Corresponding author(s). E-mail(s): xiaofang@ufl.edu; ghoshm@ufl.edu;

Abstract

We study full Bayesian procedures for high-dimensional linear regression. We adopt data-dependent empirical priors introduced in [1]. In their paper, these priors have nice posterior contraction properties and are easy to compute. Our paper extend their theoretical results to the case of unknown error variance. Under proper sparsity assumption, we achieve model selection consistency, posterior contraction rates as well as Bernstein von-Mises theorem by analyzing multivariate t-distribution.

Keywords: Bernstein von-Mises theorem, model selection consistency, multivariate t-distribution, posterior contraction rate

1 Introduction

In a series of articles, Ryan Martin and his colleagues introduced empirical priors in sparse high dimensional linear regression models (see for example, [2], [1], [3] and [4]). These priors are all data dependent and achieve nice posterior contraction rates, specifically, concentration of the parameter of interest around the true value at a very rapid rate. Moreover, these priors are quite satisfactory for both estimation and prediction as pointed out in their articles.

While [1] introduced priors both when the error variance $\sigma^2$ in a linear regression model is known and unknown, their theoretical results were proved
only for known $\sigma^2$. The objective of this paper is to fill in the gap and obtain theoretical properties, namely posterior contraction rates for unknown $\sigma^2$. The technical novelty of this approach is that unlike the former, our algebraic manipulations require handling of multivariate t-distributions rather than multivariate normal distributions. In addition to the above, we have established model selection consistency as well as a Bernstein von-Mises theorem in our proposed set up.

The outline of this paper is as follows. We have introduced the model and certain basic lemmas needed for the rest of the paper in Section 2. Posterior concentration as well as model selection consistency results are stated in Section 3. Bernstein von-Mises theorem and its application are stated in Section 4. All the proofs are given in Section 5. Some final remarks are made in Section 6.

2 The Model

Consider the standard linear regression model

$$y = X\beta + \sigma z,$$ (1)

where $y$ is a $(n \times 1)$ vector of response variables, $X$ is $n \times p$ design matrix, $\beta$ is a $(p \times 1)$ regression parameter, $\sigma > 0$ is an unknown scale parameter and $z \sim N_n(0, I_n)$.

Let $S \subseteq \{1, 2, \cdots, p\}$, $|S| \leq R$, $|S|$ denotes the cardinality of $S$. $R(\leq n \ll p)$ is the rank of $X$. Also let $X_S$ $(n \times |S|)$ denote a submatrix of the column vectors of $X$ corresponding to the elements of $S$. It is assumed that $X_S^TX_S$ is nonsingular. The corresponding elements of the regression vector $\beta$ is denoted by $\beta_S$. Also, let $\hat{\beta}_S = (X_S^TX_S)^{-1}X_S^Ty$ denote the least square estimator of $\beta_S$.

2.1 The prior

The prior considered is as follows:

(i) $\beta_S|S, \sigma^2 \sim N_{|S|}(\hat{\beta}_S, \sigma^2 \gamma^{-1}(X_S^TX_S)^{-1})$, $\beta_{S^c} = 0$ with probability 1.

(ii) $\sigma^2 \sim IG(a_0, b_0)$, i.e an inverse gamma distribution with shape and scale parameters $a_0$ and $b_0$.

(iii) Marginal priors for $S$: $\pi(S) = \binom{p}{|S|}^{-1} f_n(s)$, where $s = |S|$, and $f_n(s) \propto (c p^n)^{-s}$ for $s = 0, 1, \cdots, R$ and $f_n(s) = 0$ for $s = R + 1, \cdots, p$. 

Then the empirical joint prior for \((\beta, \sigma^2)\) is
\[
\Pi_n(\beta, \sigma^2) = \sum_S \pi(S) N(S | \hat{\beta}_S, \hat{\sigma}_S^2, \gamma^{-1}(X_S^T X_S)^{-1}) \delta_0(\beta_{S^c})
\]
\[
\times \exp\left(-\frac{b_0}{\sigma^2}\right)(\sigma^2)^{-a_0-1}b_0^a_0 / \Gamma(a_0),
\]
(2)

where \(\delta_0\) denotes the Dirac-Delta function.

### 2.2 The posterior distribution

Following [1], we consider also a fractional likelihood
\[
L^\alpha(\beta, \sigma^2) = (2\pi \sigma^2)^{\frac{n}{2}} \exp\left(-\frac{\alpha}{2\sigma^2} \|y - X\beta\|_2\right), 0 < \alpha < 1.
\]

Then the posterior for \(\beta\) conditional on \(\sigma^2\) and \(S\) is
\[
\pi^n(\beta | \sigma^2, S)
\]
\[
\propto \exp\left[-\frac{\alpha}{2\sigma^2} \|y - X_S \beta_S - X_{S^c} \beta_{S^c}\|_2^2 - \frac{\gamma}{2\sigma^2} (\beta_S - \hat{\beta}_S)^T X_S^T X_S (\beta_S - \hat{\beta}_S)\right]
\]
\[
\times (\sigma^2)^{-\left(\frac{\alpha}{2} + a_0 + 1\right)} \delta_0(\beta_{S^c}).
\]

Consider the identity
\[
\alpha \|y - X_S \beta_S - X_{S^c} \beta_{S^c}\|_2^2 + \gamma (\beta_S - \hat{\beta}_S)^T X_S^T X_S (\beta_S - \hat{\beta}_S)
\]
\[
= (\alpha + \gamma)(\beta_S - \hat{\beta}_S)^T X_S^T X_S (\beta_S - \hat{\beta}_S) + \alpha \|y - \hat{y}_S\|^2 + 
\]
\[
+ \beta_{S^c}^T X_{S^c}^T X_{S^c} \beta_{S^c} - 2 \beta_{S^c}^T X_{S^c}^T (y - X_S \beta_S),
\]

where \(\hat{y}_S = X_S \hat{\beta}_S\). Now recalling that \(\beta_{S^c}\) is concentrated at 0 with prior probability 1, it follows that
\[
\pi^n(\beta | \sigma^2, S) = \pi^n(\beta | \sigma^2, S) \delta_0(\beta_{S^c}),
\]
\[
\pi^n(\beta_S | \sigma^2) = N_0(S | \hat{\beta}_S, \sigma^2 (X_S^T X_S)^{-1}).
\]

(4)

Also the conditional posterior for \(\sigma^2\) given \(S\) is
\[
\pi^n(\sigma^2 | S) \sim IG(a_0 + \frac{\alpha n}{2}, b_0 + \frac{\alpha}{2} \|y - \hat{y}_S\|^2).
\]

(5)

Thus the conditional distribution for \(\beta_S\) given \(S\) is
\[
\pi^n(\beta_S | S) = D^n_S(\beta_S) \frac{(\alpha + \gamma)^{|S|/2} (b_0 + \frac{\alpha}{2} \|y - \hat{y}_S\|^2)^{-a_0+\alpha n/2}}{\pi(S) \gamma^{|S|/2} \Gamma(a_0 + \alpha n/2)},
\]

(6)
where

$$D^n_S(\beta_S) = \frac{\pi(S)(2\pi)^{-|S|/2} \Gamma \left(\frac{n\alpha}{2} + \frac{|S|}{2} + a_0\right) |\gamma(X^T_S X_S)|^{\frac{1}{2}}}{(2b_0 + (\beta_S - \hat{\beta}_S)^T(\gamma + \alpha)(X^T_S X_S)(\beta_S - \hat{\beta}_S) + \alpha \|y - \hat{y}_S\|^2)^{\frac{n\alpha}{2} + \frac{|S|}{2} + a_0}}. \quad (7)$$

Finally, the marginal posterior of $S$ is

$$\pi^n(S) \propto \pi(S)\left(\frac{\gamma}{\alpha + \gamma}\right)^{|S|/2} \left\{ b_0 + \frac{\alpha}{2} \|y - \hat{y}_S\|^2 \right\}^{-(a_0 + \alpha n/2)}. \quad (8)$$

**Remark 1** All our results except Corollary 2 are obtained for all $0 < \alpha < 1$ and all $\gamma > 0$. However, for higher-order properties, such as credible probability of set, some conditions on $\alpha$ and $\gamma$ are required. For example, in Corollary 2 related to uncertainty, we assume $\alpha + \gamma \leq 1$. So we can always set $\alpha$ close to 1, $\gamma$ close to 0, then the fractional likelihood is almost the same as the normal likelihood, and $\beta_S|S, \sigma^2$ is like non-informative prior.

### 2.3 Tail bounds for the Chi squared distribution

**Lemma 1** For any $a > 0$, we have

$$P(|\chi^2_p - p| > a) \leq 2 \exp(-\frac{a^2}{4p}).$$

Proof: See Lemma 4.1 of [5].

**Lemma 2**

(i) For any $c > 0$, we have

$$P(\chi^2_p(\lambda) - (p + \lambda) > c) \leq \exp\left(-\frac{p}{2} \left\{ \frac{c}{p + \lambda} - \log(1 + \frac{c}{p + \lambda}) \right\}\right).$$

(ii) For $\omega < 1$ then

$$P(\chi^2_p(\lambda) \leq \omega \lambda) \leq c_1 \lambda^{-1} \exp\{-\lambda(1 - \omega)^2/8\},$$

where $c_1 > 0$ is a constant.

(iii) For any $c > 0$, $P(\chi^2_p(\lambda) - p \leq -c) \leq \exp(-\frac{c^2}{4p}).$

Proof: See [5] Lemma 4.2 for (i) and for (ii) on [6]. (iii) follows from the fact if $U \sim \chi^2_p(\lambda)$, then $P(U > c)$ is strictly increasing in $\lambda$ for fixed $p$ and $c > 0$. Hence

$$P(\chi^2_p(\lambda) - p \leq -c) \leq P(\chi^2_p - p \leq -c) \leq \exp(-\frac{c^2}{4p}).$$
2.4 Notations

We define \( S_\beta = \{ j : \beta_j \neq 0 \} \).

Assume the true model is

\[
y \sim N_n(X\beta^*, \sigma_0^2 I_n)
\]

and let \( S^* = S_{\beta^*}, s^* = |S^*| \).

We use \( g_n \preceq M \) to denote \( g_n \leq M \) for sufficiently large \( n \), \( g_n \succeq M \) to denote \( g_n \geq M \) for sufficiently large \( n \).

3 Posterior concentration rates

Define empirical Bayes posterior probability of event \( B \subset \mathbb{R}^p \) as

\[
\Pi_n(B) = \int_B \int_{\sigma^2 > 0} L^\alpha(\beta, \sigma^2) \Pi_n(d\beta, d\sigma^2) = \frac{\sum_S \int_B D_n^S(d\beta S) \delta_{\sigma^2}(d\sigma^2 \cap d\beta S^c)}{\sum_S \int_{R^p} D_n^S(d\beta S) \delta_{\sigma^2}(d\beta S^c)}
\]

and let \( D_n = \sum_S \int_{R^p} D_n^S(d\beta S) \delta_{\sigma^2}(d\beta S^c) \).

Then recalling the pdf of a multivariate \( t \) distribution, we have

\[
D_n = \sum_S \pi(S) \frac{\nu(S)}{2} \left( \frac{\gamma}{\gamma + \alpha} \right)^{\frac{s^*}{2}} \left( \frac{1}{b_0 + \frac{\alpha \|y - \hat{y}_S\|^2}{2}} \right)^{\frac{n\alpha}{2} + a_0} \geq \pi(S^*) \frac{\nu(S^*)}{2} \left( \frac{\gamma}{\gamma + \alpha} \right)^{\frac{s^*}{2}} \left( \frac{1}{b_0 + \frac{\alpha \|y - \hat{y}_{S^*}\|^2}{2}} \right)^{\frac{n\alpha}{2} + a_0}.
\]

Since \( p \gg n \), we have to add some regularity conditions to get posterior concentration results for our model.

**regularity conditions:**

(A1) There exist constants \( d_1, d_2 \), such that \( 0 < d_1 < \sigma_0^2 < d_2 < \infty \),

(A2) \( s^* \log p = o(n) \).

(A3) \( R \log p = o(n) \).

For a given \( \Delta \), define \( B_n(\Delta) = \{ \beta \in R^p : |S_\beta| \geq \Delta \} \), i.e the set of \( \beta \) vectors with no less then \( \Delta \) non-zero entries.

The following theorem implies that the posterior distribution is actually concentrated on a space of dimension close to \( s^* \).

**Theorem 1** Let \( s^* \leq R \), and assume conditions (A1)-(A3) to hold. Then there exists constant \( C > 1, G > 0 \) such that

\[
E_{\beta^*} \{ \Pi_n(B_n(\Delta_n)) \} \leq \exp \{-Gs^* \log(p/s^*)\} \to 0
\]

with \( \Delta_n = Cs^* \), uniformly in \( \beta^* \) as \( n \to \infty \).
Empirical priors in linear regression with unknown error variance

To get posterior concentration results, we first establish model selection result. The following theorem demonstrates that asymptotically our empirical Bayesian posterior will not include any unnecessary variables.

**Theorem 2** Let $s^* \leq R$, and assume conditions (A1)-(A3) to hold. Also, let the constant $a$ in the marginal prior of $S$ satisfy $p^a \gg s^*$. Then $E_{\beta^*}(\Pi^n(\beta : S_\beta \supset S^*)) \to 0$, uniformly in $\beta^*$.

To get model selection consistency, it remains to show our empirical Bayesian posterior will asymptotically not miss any true variables.

Define

$$\kappa(s) = \kappa_X(s) = \inf_{\beta : 0 < |S_\beta| \leq s} \frac{\|X\beta\|_2}{\|\beta\|_2}, \quad s = 1, 2, \cdots p. \quad (11)$$

Then $\kappa(s)$ is a non-increasing function of $s$. By [1], for any $\beta, \beta'$, $\kappa(|S_{\beta_\beta} - \beta'|) \geq \kappa(|S_{\beta_\beta}| + |S_{\beta'}|)$.

**Theorem 3** Assume $s^* \leq R$, conditions (A1)-(A3) to hold and

$$\min_{j \in S^*} |\beta^*_j| \geq \rho_n := M'(\log p)^{1/2},$$

where $M' = \frac{1}{\kappa(C' s^*)} \left\{ \frac{12 M d_2}{\alpha} \right\}^{1/2}$. We also assume $\kappa(C' s^*) > 0$, $C' = C + 1$ for $C$ as in Theorem 1 and $M$ is a constant with $M > 1 + a$, $a > 0$ being the parameter in the prior of $S$. Then

$$E_{\beta^*}(\Pi^n(\beta : S_\beta \not\supset S^*)) \to 0.$$

**Remark 2** Intuitively, we are not able to distinguish between 0 and a very small non-zero. Hence, in Theorem 3 we define a cutoff $\rho_n$, which is similar to Theorem 5 in [1] and Theorem 5 in [7].

**Corollary 1** (Selection consistency) Assume that the conditions of Theorem 3 hold. Then $E_{\beta^*}(\Pi^n(\beta : S_\beta = S^*)) \to 1$.

**Proof:** Follows immediately from Theorems 2 and 3.

We now state our posterior concentration result. It is similar to the posterior concentration theorem in [1]. But our proof is completely different from theirs. They apply Holder’s inequality and Renyi divergence formula, while the key to our proof is using the model selection consistency result.

Set

$$B_{\epsilon_n} = \{ \beta \in \mathbb{R}^p : \|X(\beta - \beta^*)\|_2 > \epsilon_n \},$$
Theorem 4 Assume conditions (A1)-(A3) hold, then there exist constant $M, G > 0$ such that

$$E_{\beta^*} \{ \Pi^n(B_{M\epsilon_n}) \} \preceq \exp(-G\epsilon_n) \to 0$$

uniformly in $\beta^*$ as $n \to \infty$, where $\epsilon_n = s^* \log(p/s^*)$.

By adding some conditions on $X$, we are able to separate $\beta$ from $X\beta$ so we can get posterior consistency result for $\beta$.

Set

$$B_{\delta_n}^\delta = \{ \beta \in \mathbb{R}^p : \| \beta - \beta^* \|^2 > \delta_n \},$$

where $\delta_n$ is a positive sequence of constants to be specified.

Theorem 5 There exists a constant $M$ such that

$$E_{\beta^*} \{ \Pi^n(B_{M\delta_n}^\delta) \} \preceq \exp(-G\epsilon_n) \to 0$$

uniformly in $\beta^*$ as $n \to \infty$, where

$$\epsilon_n = s^* \log(p/s^*) \kappa(C's^*)^2,$$

$G > 0$ is a constant, $C' = C + 1$, $C$ being the constant in Theorem 1.

Proof: Same as the proof of Theorem 3 in [1].

Remark 3 If for any $|S| < R$ we have

$$0 < k_1 < \lambda_{\min}(X_S^T X_S/n) < \lambda_{\max}(X_S^T X_S/n) < k_2 < \infty,$$

where $k_1, k_2$ are positive constant, then we get posterior consistency for the coefficient $\beta$ under $l_2$ norm.

4 Bernstein von-Mises Theorem

In this section, we show that the posterior distribution of $\beta$ is asymptotically normal and this property leads to many interesting results.

Theorem 6 (Bernstein von-Mises Theorem) Let $H$ denote Hellinger distance, $d_{TV}$ denote the total variation distance and $\Pi^n$ be empirical Bayes posterior distribution for $\beta$. If $s^* = o(\sqrt{n}/\log n)$ and $E_{\beta^*} \pi^n(S^*) \to 1$, then

$$E_{\beta^*} H^2(\Pi^n, N_{s^*}(\hat{\beta}_{s^*}, \frac{\sigma_0^2}{\alpha + \gamma}(X_{S^*}^T X_{S^*})^{-1}) \otimes \delta_{0}S^*) \to 0$$

uniformly in $\beta^*$ as $n \to \infty$, where $\delta_{0}S^*$ denotes the point mass distribution for $\beta_{S^*}$ concentrated at the origin.

Since $d_{TV}(\cdot) \leq H^2(\cdot)$, we also have

$$E_{\beta^*} d_{TV}(\Pi^n, N_{s^*}(\hat{\beta}_{s^*}, \frac{\sigma_0^2}{\alpha + \gamma}(X_{S^*}^T X_{S^*})^{-1}) \otimes \delta_{0}S^*) \to 0$$

uniformly in $\beta^*$ as $n \to \infty$. 

Empirical priors in linear regression with unknown error variance

Define $H_n(t) = \Pi^n(\{\beta : x^T \beta \leq t\})$, $t_\gamma = \inf\{t : H_n(t) \geq 1 - \gamma\}$.

**Corollary 2 (valid uncertainty quantification)** If $\alpha + \gamma \leq 1$, then Eq. (13) implies

$$P_{\beta^*}(t_\gamma \geq x^T \beta^*) \geq 1 - \gamma + o(1), \quad n \to \infty,$$

which validates uncertainty quantification.

Consider a pair $(\tilde{X}, \tilde{y})$ where $\tilde{X} \in \mathbb{R}^{d \times p}$ is a given matrix of explanatory variable values at which we want to predict the corresponding response $\tilde{y} \in \mathbb{R}^d$. Let $f^n_X(\tilde{y}|S)$ be the conditional posterior predictive distribution of $\tilde{y}$, given $S$. $f^n_X(\tilde{y}|S)$ is a t-distribution, with $2a + \alpha n$ degree of freedom, location $\tilde{X}_S \hat{\beta}_S$, and the scale matrix

$$b + (\alpha/2)\|y - \hat{y}_S\|^2 / (a + \alpha n/2)(I_d + \frac{1}{\alpha + \gamma} \tilde{X}_S (\tilde{X}_S^T \tilde{X}_S)^{-1} \tilde{X}_S^T).$$

Then the predictive distribution of $\tilde{y}$ is

$$f^n_X(\tilde{y}) = \sum_S \pi^n(S)f^n_X(\tilde{y}|S).$$

For this predictive distribution, we have similar Bernstein von-Mises Theorem as Theorem 6. The proof is similar.

**Theorem 7** If $s^* = o(\sqrt{n}/\log n)$ and $E_{\beta^*}\pi^n(S^*) \to 1$, then

$$E_{\beta^*} H^2(f^n_X, N_d(\tilde{X}_S^*, \hat{\beta}_S^*, \sigma_0^2[I_d + \frac{1}{\alpha + \gamma} \tilde{X}_S^* (\tilde{X}_S^T \tilde{X}_S^*)^{-1} \tilde{X}_S^T])) \to 0$$

uniformly in $\beta^*$ as $n \to \infty$.

Since $d_{TV}(\cdot) \leq H^2(\cdot)$, we also have

$$E_{\beta^*} d_{TV}(f^n_X, N_d(\tilde{X}_S^*, \hat{\beta}_S^*, \sigma_0^2[I_d + \frac{1}{\alpha + \gamma} \tilde{X}_S^* (\tilde{X}_S^T \tilde{X}_S^*)^{-1} \tilde{X}_S^T])) \to 0$$

uniformly in $\beta^*$ as $n \to \infty$.
5 Proofs

Proof of Theorem 1:
Let \( u_S = |S \cup S^*| \), \( R_{n,S} = \{ y : y^T(I - P_{S \cup S^*})y/\sigma_0^2 \geq \frac{\alpha+1}{2}(n - u_S) \} \).

\[
\Pi_n(B_n(\Delta_n)) = \sum_{\Delta_n \leq |S| \leq R} \int_{|S|} D_n^2(d\beta_S)
\]

\[
\sum_{\Delta_n \leq |S| \leq R} \pi(S) \Gamma\left(\frac{\alpha}{2} + a_0\right) \left(\frac{\gamma}{\gamma + \alpha}\right)^{|S|/2} \left[ \frac{1}{b_0 + \frac{1}{2} \| y - \hat{y}_S \|^2} \right] \frac{\alpha}{2} y^T(P - P_{S^*})y \frac{\alpha}{2} y^T(I - P_{S \cup S^*})y \leq \sum_{\Delta_n \leq |S| \leq R} \frac{\alpha}{2} y^T(P_{S \cup S^*} - P_{S^*})y \frac{\alpha}{2} y^T(I - P_{S \cup S^*})y
\]

(17)

Let \( U_1, U_2 \) denote the two terms on the right hand side of (17). It suffices to show that there exists constant \( C > 1, G > 0 \) such that \( E_{\beta^*} U_i \leq \exp\{-G s^* \log(p/s^*)\} \) with \( \Delta_n = Cs^* \) uniformly in \( \beta^* \) for \( i = 1, 2 \).

Since

\[
y^T(I - P_{S \cup S^*})y/\sigma_0^2 \sim \chi_n^2 - u_S,
\]

by Lemma 1 we have

\[
E_{\beta^*} U_1 = \sum_{\Delta_n \leq |S| \leq R} P(R_{n,S}^c) \leq \sum_{\Delta_n \leq |S| \leq R} 2 \exp\{-\frac{(1 - \alpha)^2(n - u_S)}{16}\}
\]

\[
\leq \sum_{\Delta_n \leq |S| \leq R} 2 \exp\{-\frac{(1 - \alpha)^2}{20}n\} \leq \frac{p}{R} 2 \exp\{-\frac{(1 - \alpha)^2}{20}n\}
\]

(19)

\[
\leq 2R p^R \exp\{-\frac{(1 - \alpha)^2}{20}n\} \leq \exp\{-\frac{(1 - \alpha)^2}{30}n\}.
\]
Let $Z_1 = y^T(P_{S \cup S^*} - P_{S^*})y/\sigma^2_0$, then $Z_1 \sim \chi^2_{u_S - s^*}$. By $1 + x \leq \exp(x)$,

$$[1 + \frac{y^T(P_{S \cup S^*} - P_{S^*})y/\sigma^2_0}{\frac{n\alpha}{2}(n - u_S)}]^{n\alpha + a_0} \leq \exp\{Z_1 (\frac{n\alpha/2 + a_0}{\frac{n\alpha}{2}(n - u_S)})\} \leq \exp\{\frac{\alpha Z_1}{1 + \alpha}\}. \tag{20}$$

Using the mgf of a chi-squared distribution,

$$E \beta^* U_2 \leq \sum_{\Delta_n \leq |S| \leq R} \frac{\pi(S)}{\pi(S^*)} (\frac{\gamma}{\alpha + \gamma})^{|S|/2 - s^*/2} E \beta^* \exp\{\frac{\alpha Z_1}{1 + \alpha}\}$$

$$= \sum_{\Delta_n \leq |S| \leq R} \frac{\pi(S)}{\pi(S^*)} (\frac{\gamma}{\alpha + \gamma})^{|S|/2 - s^*/2} (1 - \frac{2\alpha}{1 + \alpha})^{-|S| - |S^*|}$$

$$\leq \sum_{\Delta_n \leq |S| \leq R} \frac{\pi(S)}{\pi(S^*)} (\frac{\gamma}{\alpha + \gamma})^{-s^*/2} (1 - \frac{\alpha}{1 + \alpha})^{-|S|}$$

$$= \exp(cs^*) \frac{(p)}{f_n(s^*)} \sum_{s=\Delta_n} R \phi^s f_n(s), \tag{21}$$

where $\phi = (\frac{1 - \alpha}{1 + \alpha})^{-1} > 1$.

By [1] we have $\log\{\frac{(p)}{f_n(s^*)} \sum_s \phi^s f_n(s)\} = O(s^* \log(p/s^*))$. Also since $\phi > 1$, $\sum_s \phi^s f_n(s) > 1$. Hence

$$E \beta^* U_2 \leq \exp\{cs^* + s^* \log(p/s^*)\} \sum_{s=\Delta_n} R \phi^s f_n(s). \tag{22}$$

From the expression of $f_n(s)$, we get

$$\sum_{s=\Delta_n} R \phi^s f_n(s) = O((\frac{\phi}{cp^a})^{\Delta_n + 1}). \tag{23}$$

and

$$((\frac{\phi}{cp^a})^{\Delta_n + 1} = \exp\{-(\Delta_n + 1)[a \log p + \log(c/\phi)]\}.\$$

So when $\Delta_n = C s^*$, $C > a^{-1}$ and $G' = \frac{aC - 1}{2}$,

$$E \beta^* U_2 \leq \exp\{-G' s^* \log(p/s^*)\} \to 0 \tag{24}$$
as \( n \to \infty \).

By (19) and (24), when \( \Delta_n = Cs^* \), \( C > a^{-1} \) and \( G = \frac{aC-1}{4} \), we have

\[
E_{\beta^*}\{\Pi^n(B_n(\Delta_n))\} \leq \exp\{-Gs^* \log(p/s^*)\} \to 0
\]

with \( \Delta_n = Cs^* \), uniformly in \( \beta^* \) as \( n \to \infty \).

**Proof of Theorem 2:**

Since

\[
E_{\beta^*}\{\Pi^n(\beta : S_\beta \supset S^*)\} = \sum_{S : S \supset S^*} E_{\beta^*}\pi^n(S),
\]

by Theorem 1, it suffices to show that

\[
\sum_{S : S \supset S^*, |S| \leq Cs^*} E_{\beta^*}\pi^n(S) \to 0. \tag{25}
\]

Let \( \tilde{R}_{n,S} = \{y : y^T(I - P_S)y/\sigma_0^2 \geq \frac{a+1}{2}(n - s)\} \). For \( S \supset S^* \), by \( 1 + x \leq \exp(x) \),

\[
\pi^n(S) \leq \frac{\pi^n(S)}{\pi^n(S^*)} \mathbb{1}_{\tilde{R}_{n,S}} \leq \frac{\pi(S)}{\pi(S^*)}(\frac{\gamma}{\alpha + \gamma})^{|S|-s^*/2} \exp\{[y^T(P_S - P_{S^*})y/\sigma_0^2] \frac{n\alpha/2 + a_0}{\frac{a+1}{2}(n - s)}\}. \tag{26}
\]

Since \( y^T(P_S - P_{S^*})y/\sigma_0^2 \sim \chi^2_{s^* - s}^2 \) and \( y^T(I - P_S)y/\sigma_0^2 \sim \chi^2_{n - s}^2 \), by Lemma 1 and mgf formula for chi-squared distribution, we have

\[
E_{\beta^*}\pi^n(S) \leq P(\tilde{R}_{n,S}^c) + \frac{\pi(S)}{\pi(S^*)}(\frac{\gamma}{\alpha + \gamma})^{|S|-s^*/2} \frac{1 - \alpha}{1 + \alpha} - (|S|-s^*). \tag{27}
\]
Empirical priors in linear regression with unknown error variance

Let $z = \left( \frac{\gamma}{\alpha + \gamma} \right)^{1/2} (\frac{1-\alpha}{1+\alpha})^{-1}$. Hence

$$\sum_{S : S \supset S^* , |S| \leq C s^* } E_{\beta^*} \pi^n (S) \leq \sum_{S : S \supset S^* , |S| \leq C s^* } P(\tilde{R}_{n,S}^c) + \sum_{S : S \supset S^* , |S| \leq C s^* } \frac{\pi(S)}{\pi(S^*)} e^{-|S| - s^*}$$

$$\leq \exp\left\{ -(1 - \alpha)^2 \right\}_{30} n + \sum_{s^* < s \leq C s^* } \frac{(p - s)(p)}{(p)} \left( \frac{z}{c p a} \right)^{s - s^*}$$

$$= \exp\left\{ -(1 - \alpha)^2 \right\}_{30} n + \sum_{s^* < s \leq C s^* } \left( \frac{s}{1 - s^*} \right) \left( \frac{z}{c p a} \right)^{s - s^*}$$

$$\leq \exp\left\{ -(1 - \alpha)^2 \right\}_{30} n + \sum_{s^* < s \leq C s^* } \left( C z s^* \right)^{s - s^*}$$

$$\leq \exp\left\{ -(1 - \alpha)^2 \right\}_{30} n + \frac{C z s^*}{c p a} \times O(1).$$

(28)

Then RHS goes to 0, when $p^a \gg s^*$.

**Proof of Theorem 3:**

Since

$$E_{\beta^*} \{ \Pi^n (\beta : S_{\beta} \not\supset S^*) \} = \sum_{S : S \not\supset S^*} E_{\beta^*} \pi^n (S),$$

by Theorem 1, it suffices to show

$$\sum_{S : S \not\supset S^* , |S| \leq C s^* } E_{\beta^*} \pi^n (S) \to 0. \quad \text{(29)}$$

Define

$$L_{n,S}$$

$$= \{ y : \left( \frac{\alpha + 1}{2} (n - |S|) \right) \leq \| y - \hat{y}_S \|^2 / \sigma_0^2 \leq 2(n - |S|) + 2 \| (I - P_S) X \beta^* \|^2 / \sigma_0^2 \},$$

(30)

$$\pi^n (S) \leq 1 L_{n,S} + \frac{\pi^n (S)}{\pi^n (S^*)} 1 L_{n,S}$$

$$= 1 L_{n,S} + \frac{\pi(S)}{\pi(S^*)} (\frac{\gamma}{\alpha + \gamma})^{\frac{|S|/2 - s^*/2}{\alpha}} \left[ b_0 + \frac{\gamma}{2} \| y - \hat{y}_S \|^2 \right]^{\alpha + \gamma} 1 L_{n,S}.$$  

(31)
By $1 + x \leq \exp(x)$,

$$
\frac{b_0 + \frac{\alpha}{2} \| y - \hat{y}s \|}{b_0 + \frac{\alpha}{2} \| y - \hat{y}s \|^2} \leq [1 + \frac{\alpha}{2} y^T (P_S - P_{S^*}) y \exp(\frac{\alpha}{2} y^T (P_S - P_{S^*}) y)}{b_0 + \frac{\alpha}{2} \| y - \hat{y}s \|^2}]^{n\alpha/2 + a_0} \mathbb{1}_{n,S} 
$$

$$
\leq \exp\{\frac{\alpha}{2} y^T (P_S - P_{S^*}) y \frac{n\alpha/2 + a_0}{b_0 + \frac{\alpha}{2} \| y - \hat{y}s \|^2}\} \mathbb{1}_{n,S} 
$$

$$
\leq \exp\{y^T (P_S - P_{S^*}) y \frac{\alpha(n\alpha/2 + a_0)}{2b_0 + \alpha(n + 1)\sigma_0^2(n - |S|)}\} + \exp\{y^T (P_S - P_{S^*}) y \frac{\alpha(n\alpha/2 + a_0)}{2b_0 + 2\alpha\sigma_0^2(2(n - |S|) + 2 \| (I - P_S)X\beta^* \|^2 / \sigma_0^2)\}.
$$

(32)

Since

$$
\| y - \hat{y}s \|^2 / \sigma_0^2 \sim \chi_{n-s}^2(\| (I - P_S)X\beta^* \|^2 / \sigma_0^2),
$$

by Lemma 2 we have

$$
\sum_{|S| \leq C_s^*} P(L_{n,S}^c) \leq \sum_{|S| \leq C_s^*} \{\exp\{-(n - |S|)} + 2 \exp\{-(1 - \alpha)^2(n - |S|)/16\}\}
$$

$$
\leq \exp\{-\frac{(1 - \alpha)^2n}{20}\} \sum_{s \leq C_s^*} \binom{p}{s} \leq \exp\{-\frac{(1 - \alpha)^2n}{20}\} C_s^* p C_s^* \leq \exp\{-\frac{(1 - \alpha)^2n}{30}\} \to 0,
$$

as $n \to \infty$.

Plugging $y = X\beta^* + \sigma_0 \epsilon$ into $y^T (P_S - P_{S^*}) y$, where $\epsilon \sim N_n(0, I)$, we get

$$
- \| (I - P_S)X\beta^* \|^2 - 2\sigma_0 \epsilon^T (I - P_S)X\beta^* + \sigma_0^2 \epsilon^T (P_S - P_{S^*}) \epsilon.
$$

Bound the right-most quadratic form above as follows,

$$
\epsilon^T (P_S - P_{S^*}) \epsilon = \epsilon^T (P_S - P_{S_{S^*}}) \epsilon - \epsilon^T (P_{S^*} - P_{S_{S^*}}) \epsilon \leq \epsilon^T (P_S - P_{S_{S^*}}) \epsilon,
$$

so

$$
y^T (P_S - P_{S^*}) y \leq - \| (I - P_S)X\beta^* \|^2 - 2\sigma_0 \epsilon^T (I - P_S)X\beta^* + \sigma_0^2 \epsilon^T (P_S - P_{S_{S^*}}) \epsilon.
$$

We also observe that $(I - P_S)(P_S - P_{S_{S^*}}) = 0$, which implies that

$$
\epsilon^T (I - P_S)X\beta^* \perp \epsilon^T (P_S - P_{S_{S^*}}) \epsilon.
$$
Empirical priors in linear regression with unknown error variance

Then by the mgf of normal and chi-squared distribution, we have

\[
E_{\beta^*} \exp \left\{ \frac{l}{2\sigma_0^2} \{ y^T (P_S - P_{S^*}) y \} \right\} \leq (1 - l)^{-\frac{1}{2}|S| - |S \cap S^*|} \exp \left\{ -\frac{l(1 - l)}{2\sigma_0^2} \| (I - P_S) X \beta^* \|^2 \right\}.
\]

Hence,

\[
E_{\beta^*} \left\{ \left[ \frac{b_0 + \frac{\alpha}{2}}{2b_0 + \frac{\alpha}{2}} \right] \| y - \hat{y}_{S^*} \|^2 \right\} \leq E_{\beta^*} \exp \left\{ y^T (P_S - P_{S^*}) y \frac{\alpha(n\alpha/2 + a_0)}{2b_0 + \alpha(\alpha + 1)\sigma_0^2(n - |S|)} \right\}
\]

\[
+ E_{\beta^*} \exp \left\{ y^T (P_S - P_{S^*}) y \frac{\alpha(n\alpha/2 + a_0)}{2b_0 + 2\alpha\sigma_0^2[2(n - |S|) + 2\| (I - P_S) X \beta^* \|^2/\sigma_0^2]} \right\}
\]

\[
\leq \sum_{i=1}^{2} (1 - \alpha_i)^{-\frac{1}{2}|S| - |S \cap S^*|} \exp \left\{ -\frac{\alpha_i(1 - \alpha_i)}{2\sigma_0^2} \| (I - P_S) X \beta^* \|^2 \right\},
\]

where \( \alpha_1 = \frac{2\sigma_0^2\alpha(n\alpha/2 + a_0)}{2b_0 + \alpha(\alpha + 1)\sigma_0^2(n - |S|)} \leq \frac{\alpha}{1 + \alpha} \), \( \alpha_2 = \frac{2\sigma_0^2\alpha(n\alpha/2 + a_0)}{2b_0 + 2\alpha\sigma_0^2[2(n - |S|) + 2\| (I - P_S) X \beta^* \|^2/\sigma_0^2]} \leq \frac{\alpha}{4} \).

Since

\[\| (I - P_S) X \beta^* \|^2 = \| (I - P_S) X_{S \cup S^*} \beta_{S \cup S^*} \|^2,\]

it follows from Lemma 5 of [8] that

\[\| (I - P_S) X \beta^* \|^2 \geq \kappa(S \cup S^*)(|S^*| - |S \cap S^*|)p_n^2.\]

Next, we have \(|S \cup S^*| \leq |S| + |S^*| \leq C'|S^*|\). By the monotonicity of \( \kappa \) we also have

\[\| (I - P_S) X \beta^* \|^2 \geq \kappa(C's^*)^2(|S^*| - |S \cap S^*|)p_n^2 \geq (M')^2 \kappa(C's^*)^2(|S^*| - |S \cap S^*|) \log p \]

Then

\[
\frac{\alpha_1(1 - \alpha_1)}{2\sigma_0^2} \| (I - P_S) X \beta^* \|^2 \geq \frac{\alpha}{2(\alpha + 1)^2d_2} (M')^2 \kappa(C's^*)^2(|S^*| - |S \cap S^*|) \log p \geq M(|S^*| - |S \cap S^*|) \log p,
\]
\[
\frac{\alpha_2(1 - \alpha_2)}{2\sigma_0^2} \| (I - P_S)X \beta^* \|^2 \geq \frac{3\alpha_2}{8\sigma_0^2} \| (I - P_S)X \beta^* \|^2
\]
\[
\geq \frac{3\sigma_0^2 \alpha(n\alpha/2 + a_0)(M')^2 \kappa(C's^*)^2(|S^*| - |S \cap S^*|) \log p}{8b_0 + 8\alpha\sigma_0^2[2(n - |S|) + 2(M')^2 \kappa(C's^*)^2(|S^*| - |S \cap S^*|) \log p/\sigma_0^2]}
\]
\[
\geq \frac{(3\alpha^2/2)(M')^2 \kappa(C's^*)^2(|S^*| - |S \cap S^*|) \log p}{18\alpha\sigma_0^2[n + s^* \log p \cdot (24d_2M)/(d_1\alpha)]}
\]
\[
\geq \frac{\alpha}{12d_2} (M')^2 \kappa(C's^*)^2(|S^*| - |S \cap S^*|) \log p = M(|S^*| - |S \cap S^*|) \log p.
\]

By (31)(32)(33)(35), we have
\[
\sum_{S: S \not\subseteq S^*, |S| \leq C_s^*} E_{\beta^*} \pi^n(S) \leq 
\sum_{S: S \not\subseteq S^*, |S| \leq C_s^*} E_{\beta^*} \{1_{\mathbb{L}^c_{n,s}} + \frac{\pi(S)}{\pi(S^*)} \left(\frac{\gamma}{\alpha + \gamma}\right)^{|S|/2 - s^*/2} \frac{b_0 + \alpha}{b_0 + \alpha} \| y - \hat{y}_{s^*} \|^2 \}^{n\alpha/2 + a_0} \mathbb{1}_{L_{n,s}} \}
\leq \exp\left\{-\frac{(1 - \alpha)^2 n}{30}\right\} + \sum_{S: S \not\subseteq S^*, |S| \leq C_s^*} 2 \frac{\pi(S)}{\pi(S^*)} \nu^{s^* - |S|}(\sqrt{2p - M})\delta |S^*| - |S \cap S^*|,
\]

where \( \nu = \left(\frac{2\gamma}{\alpha + \gamma}\right)^{-1/2}. \)
Plug in the prior of $S$, and let $t$ be the number of variables in $S \cap S^\ast$. We get

$$\sum_{S: S \not\subseteq S^\ast, |S| \leq Cs^\ast} \frac{\pi(S)}{\pi(S^\ast)} \nu^{|S| - |S^\ast|} (\sqrt{2p - M}) |S^\ast| - |S \cap S^\ast|$$

$$\leq \sum_{s=0}^{Cs^\ast \min \{s, s^\ast\}} \sum_{t=1}^{s^\ast - 1} \left( \frac{p^s}{s} \right) \left( \frac{p - s^\ast}{s^\ast - t} \right) \left( \nu c p^a \right) s^\ast - s (\sqrt{2p - M}) s^\ast - t$$

$$= \sum_{s=0}^{Cs^\ast \min \{s, s^\ast\}} \sum_{t=1}^{s^\ast - 1} \left( \frac{S}{t} \right) \left( \frac{p - s}{s^\ast - t} \right) \left( \nu c p^a \right) s^\ast - s (\sqrt{2p - M}) s^\ast - t$$

$$= \sum_{s=0}^{Cs^\ast \min \{s, s^\ast\}} \sum_{t=1}^{s^\ast - 1} \left( \frac{p^s}{s} \right) \left( \frac{p - s^\ast}{s^\ast - t} \right) \left( \nu c p^a \right) s^\ast - s (\sqrt{2p - M}) s^\ast - t$$

$$= \sum_{s=0}^{s^\ast - 1} \sum_{t=1}^{s^\ast} (\nu c p^a / s) s^\ast - s (\sqrt{2sp^{1-M}}) s^\ast - t + \sum_{s=s^\ast}^{Cs^\ast} \sum_{t=1}^{s^\ast - 1} (\nu c p^a / s) s^\ast - s (\sqrt{2sp^{1-M}}) s^\ast - t$$

$$\leq \sum_{s=0}^{s^\ast - 1} (\nu c p^{1+a-M} / s) s^\ast - s + s^\ast p^{1-M} \sum_{s=s^\ast}^{Cs^\ast} (\nu c p^a / s) s^\ast - s \to 0,$$

which implies $\sum_{S: S \not\subseteq S^\ast, |S| \leq Cs^\ast} E_{\beta^\ast} \pi^n(S) \to 0$.

As we see, $\| (I - P_S) X \beta^\ast \|^2$ plays an important role in the proof of Theorem 3. We modify this proof to get a useful result.

**Lemma 3** Assume conditions (A1)-(A3) hold, define

$$H_3^\beta = \{ \beta^\ast : \| (I - P_S) X \beta^\ast \|^2 \geq K s^\ast \log(p/s^\ast) \},$$

$K > (2a + 3) \frac{12d_2}{\alpha}$ is a constant, with $a > 0$ the constant in marginal prior of $S$, $C$ is a constant as in Theorem 1. Then there exists constant $G_1 > 0$, such that

$$\sum_{S: S \not\subseteq S^\ast, |S| \leq Cs^\ast} 1_{H_3^\beta} \cdot E_{\beta^\ast} \pi^n(S) \leq \exp \{- G_1 s^\ast \log(p/s^\ast) \} \to 0$$

uniformly in $\beta^\ast$ as $n \to \infty$. 
Empirical priors in linear regression with unknown error variance

Proof By (31), (33), (35), for $S \in \{ S : S \not\subseteq S^*, |S| \leq Cs^* \}$,

$$1_{H_S^*} \cdot E_{\beta^*} \cdot \pi^n(S) \leq 1_{H_S^*} \cdot E_{\beta^*} [\pi^n(S)1_{Q_n}] + P(Q_n^c)$$

$$\leq 1_{H_S^*} \cdot \frac{\pi(S)}{\pi(S^*)} (\frac{\gamma}{\alpha + \gamma}) \frac{1}{2} \sum_{i=1}^{2|S| - s^*} (1 - \alpha_i)^{-\frac{1}{2}} (|S| - |S \cap S^*|) \exp\{- \frac{\alpha_i (1 - \alpha_i)}{2\sigma_0^2} \| (I - P_S)X\beta^* \|^2 \} + \exp\{- \frac{(1 - \alpha)^2 n}{20}\}.$$  

(38)

Then on $H_S^*$,

$$\frac{\alpha_1 (1 - \alpha_1)}{2\sigma_0^2} \| (I - P_S)X\beta^* \|^2 \geq \frac{K\alpha}{2(\alpha + 1)^2 d_2} s^* \log(p/s^*),$$

$$\frac{\alpha_2 (1 - \alpha_2)}{2\sigma_0^2} \| (I - P_S)X\beta^* \|^2 \geq \frac{K\alpha}{12d_2} s^* \log(p/s^*).$$

So

$$1_{H_S^*} \cdot E_{\beta^*} [\pi^n(S)1_{Q_n}] \leq 2 \frac{\pi(S)}{\pi(S^*)} \nu^{s^* - |S|} |\sqrt{2(p/s^*)} - M_0| s^*,$$

(39)

where $\nu = \left( \frac{2(\gamma+1)}{\alpha+\gamma} \right)^{-1/2}$, $M_0 = \frac{K\alpha}{12d_2}$. □

Set

$$B_{\epsilon_n} = \{ \beta \in \mathbb{R}^p : \|X(\beta - \beta^*)\|^2 \geq \epsilon_n \},$$

(40)

where $\epsilon_n$ is a positive sequence of constants to be specified later.

Proof of Theorem 4:

Since

$$\Pi^n(B_{M\epsilon_n}) = \sum_S \pi^n(B_{M\epsilon_n}|S)\pi^n(S) \leq \sum_{|S| \leq Cs^*} \pi^n(B_{M\epsilon_n}|S)\pi^n(S) + \sum_{|S| > Cs^*} \pi^n(S),$$

(41)

by Theorem 1 it suffices to show that

$$E_{\beta^*} \sum_{|S| \leq Cs^*} \pi^n(B_{M\epsilon_n}|S)\pi^n(S) \leq \exp(-G_2 \epsilon_n).$$

(42)

Let $\beta_{S^+}$ be a $p$-vector by augmenting $\beta_S$ with $\beta_j = 0$ for all $j \in S^c$, and $B_{M\epsilon_n}(S)$ be the set of all $\beta_S$ such that $\beta_{S^+} \in B_n$, $\pi^n(B_{M\epsilon_n}|S) = \pi^n(B_{M\epsilon_n}(S)|S)$.

Define

$$Q_{n,S} = \{ \beta^* \in \mathbb{R}^p : \| (I - P_S)X\beta^* \|^2 \geq M\epsilon_n/2 \}.$$
Empirical priors in linear regression with unknown error variance

Then
\[
\sum_{|S| \leq C_s} \pi^n(B_{M \epsilon_n}(S)|S) \pi^n(S) \\
= \sum_{S: \not\subseteq S, |S| \leq C_s} \pi^n(B_{M \epsilon_n}(S)|S) \pi^n(S) + \sum_{S: \not\subseteq S, |S| \leq C_s} \pi^n(B_{M \epsilon_n}(S)|S) \pi^n(S) \\
\leq W_1 + W_2 + W_3, 
\]
(43)

where \( W_1 = \sum_{S: \not\subseteq S, |S| \leq C_s} \pi^n(S) \mathbb{1}_{Q_{n,s}} \), \( W_2 = \sum_{S: \not\subseteq S, |S| \leq C_s} \pi^n(B_{M \epsilon_n}(S)|S) \mathbb{1}_{Q_{n,s}} \), and \( W_3 = \sum_{S: \not\subseteq S, |S| \leq C_s} \pi^n(B_{M \epsilon_n}(S)|S) \). Thus it suffices to show that \( E_\beta \cdot W_i \leq \exp(-G_\epsilon_n), i = 1, 2, 3 \) for some \( G > 0 \).

For \( W_1 \), let \( M > (2a + 3) \frac{2d^2}{\alpha} \). Then by Lemma 3, we have \( E_\beta \cdot W_1 \leq \exp(-G_1 \epsilon_n) \).

Next we consider \( W_2 \). By (6) and generalized Holder’s inequality, we have
\[
E_\beta \cdot \pi^n(B_{M \epsilon_n}(S)|S) \\
= \frac{(a + \gamma/2)|S|/2}{2\pi} |X_S^T X_S|^{1/2} \frac{\Gamma(a_0 + \alpha n/2 + |S|/2)}{\Gamma(a_0 + \alpha n/2)} \\
\times \int_{B_{M \epsilon_n}(S)} \frac{\{b_0 + \frac{\alpha}{2} \|y - \hat{y}_S\|^2\}^{(a_0+\alpha n/2)}}{\{b_0 + \frac{\alpha}{2} \|y - \hat{y}_S\|^2 + \frac{\gamma+\alpha}{2} \|X_S(\beta_S - \hat{\beta}_S)\|^2\}^{(a_0+\alpha n/2+|S|/2)}} d\beta_S \\
\leq \frac{(a + \gamma/2)|S|/2}{2\pi} |X_S^T X_S|^{1/2} \frac{\Gamma(a_0 + \alpha n/2 + |S|/2)}{\Gamma(a_0 + \alpha n/2)} \\
\times \int_{B_{M \epsilon_n}(S)} \left[ E_\beta \cdot \frac{\{b_0 + \frac{\alpha}{2} \|y - \hat{y}_S\|^2\}^{(a_0+\alpha n/2-1)/2}}{\{b_0 + \frac{\alpha}{2} \|y - \hat{y}_S\|^2 + \frac{\gamma+\alpha}{2} \|X_S(\beta_S - \hat{\beta}_S)\|^2\}^{(a_0+\alpha n/2-1)/2}} \right]^{1/2} \\
\cdot \left[ E_\beta \cdot \frac{1}{\{b_0 + \frac{\alpha}{2} \|y - \hat{y}_S\|^2 + \frac{\gamma+\alpha}{2} \|X_S(\beta_S - \hat{\beta}_S)\|^2\}^{(|S|/2+1)/2}} \right]^{1/4} \\
\cdot \left[ E_\beta \cdot \frac{1}{\{b_0 + \frac{\alpha}{2} \|y - \hat{y}_S\|^2 + \frac{\gamma+\alpha}{2} \|X_S(\beta_S - \hat{\beta}_S)\|^2\}^{(|S|/2+1)/2}} \right]^{1/4} d\beta_S. 
\]
(44)

We now observe that
\[
\|y - \hat{y}_S\|^2 \sim \sigma_0^2 \chi_{\alpha n-|S|}^2 \left( \frac{(I - P_S) X \beta^* \| \beta^* \|_2}{\sigma_0^2} \right) 
\]
and
\[
\|X_S(\beta_S - \hat{\beta}_S)\|^2 |\beta_S| \sim \sigma_0^2 \chi_{|S|}^2 \left( \frac{P_S X (\beta_S + \beta^*) \| \beta^* \|_2}{\sigma_0^2} \right). 
\]
Also
\[
\| X(\beta S_+ - \beta^*) \|^2_2 = \| (I - P_S)X(\beta S_+ - \beta^*) \|^2_2 + \| P_S X(\beta S_+ - \beta^*) \|^2_2 \\
= \| (I - P_S)X\beta^* \|^2_2 + \| P_S X(\beta S_+ - \beta^*) \|^2_2.
\] (45)

Hence by Lemmas 1 and 2, for \(|S| \leq Cs^*\), on \(Q^c_{n,S} \cap B_{M\epsilon_n}(S)\), we have
\[
\| P_S X(\beta S_+ - \beta^*) \|^2_2 \geq M\epsilon_n/2, \quad P(\|y - \hat{y}_S\|^2 > 2\sigma_0^2 n) \leq \exp(-n/10),
\] (46)
and
\[
P(\| X_S(\beta \hat{S} - \hat{\beta}_S) \|^2 \leq M\epsilon_n/4| \beta_S) \\
\leq P(\| X_S(\beta \hat{S} - \hat{\beta}_S) \|^2 \leq \| P_S X(\beta S_+ - \beta^*) \|^2_2 /2| \beta_S) \\
\leq \frac{C'\sigma_0^2}{\| P_S X(\beta S_+ - \beta^*) \|^2_2} \exp\{-\frac{\| P_S X(\beta S_+ - \beta^*) \|^2_2}{32\sigma_0^2}\} \\
\leq \frac{2c_2 C'}{M\epsilon_n} \exp\{-\frac{\| P_S X(\beta S_+ - \beta^*) \|^2_2}{32\sigma_0^2}\} \leq \frac{2c_2 C'}{M\epsilon_n} \exp\{-\frac{M\epsilon_n}{64\sigma_0^2}\},
\] (47)
where \(C'\) is a constant.

Let
\[
A = \{y : \|y - \hat{y}_S\|^2 \geq 2\sigma_0^2 n\}, B_1 = \{(y, \beta_S) : \| X_S(\beta \hat{S} - \hat{\beta}_S) \|^2 \leq M\epsilon_n/4\}, \\
B_2 = \{(y, \beta_S) : \| X_S(\beta \hat{S} - \hat{\beta}_S) \|^2 \leq \| P_S X(\beta S_+ - \beta^*) \|^2_2/2\}.
\] (48)

Hence on \(Q^c_{n,S} \cap B_{M\epsilon_n}(S)\), we have
\[
[E_{\beta^*}(b_0 + \frac{\alpha}{2}\|y - \hat{y}_S\|^2)^{4/2}]^{1/4} \\
= [b_0^2 + \alpha\sigma_0^2 b_0(n - |S|) + \frac{\| (I - P_S)X\beta^* \|^2_2}{\sigma_0^2} + \frac{\alpha^2\sigma_0^4}{4}(2(n - |S|)) \\
+ \frac{4\| (I - P_S)X\beta^* \|^2_2}{\sigma_0^2}|n - |S| + \frac{\| (I - P_S)X\beta^* \|^2_2}{\sigma_0^2}]^{1/4} \\
\leq 2n^{1/2}.
\] (49)
20  Empirical priors in linear regression with unknown error variance

Also by $1 + x \leq \exp(x)$,

\[
\begin{align*}
\{b_0 + \frac{\alpha}{2} \| y - \hat{y}_S \|^2 \}^{(a_0 + \alpha n/2 - 1/2)^2} & \{b_0 + \frac{\alpha + \alpha n}{2} \| X_S (\beta_S - \hat{\beta}_S) \|^2 \}^{(a_0 + \alpha n/2 - 1/2)^2} \\
\leq & \left[ E_{\beta^*} \left( \frac{b_0 + \alpha \sigma_0^2 n}{b_0 + \alpha \sigma_0^2 n + \frac{\alpha + \alpha n}{8} M \epsilon_n} \right)^{(a_0 + \alpha n/2 - 1/2)^2} + E_{\beta^*} \cdot 1_{A \cup B_1} \right]^{1/2} \\
\leq & \exp\left\{ - \frac{\alpha + \gamma}{8d_2} M \epsilon_n \right\} + \exp\left\{ - n/10 \right\} + \frac{2c_2 C'}{M \epsilon_n} \exp\left\{ - \frac{M \epsilon_n}{64 \sigma_0^2} \right\}^{1/2} \\
\succeq & \exp\left\{ - \frac{\alpha}{128d_2} M \epsilon_n \right\}
\end{align*}
\]

and

\[
\begin{align*}
\{b_0 + \frac{\alpha}{2} \| y - \hat{y}_S \|^2 + \frac{\alpha + \alpha n}{2} \| X_S (\beta_S - \hat{\beta}_S) \|^2 \}^{(|S|/2 + 1/2)^4} & \\
\leq & \left[ E_{\beta^*} \left( \frac{b_0 + \alpha \sigma_0^2 n}{b_0 + \alpha \sigma_0^2 n + \frac{\alpha + \alpha n}{8} M \epsilon_n} \right)^{(|S|/2 + 1/2)^4} + E_{\beta^*} \cdot 1_{B_2} \right]^{1/4} \\
\leq & \left[ E_{\beta^*} \left\{ b_0 + \frac{\alpha + \alpha n}{4} \| P_S X (\beta_{S^*} - \beta^*) \|_2 \right\}^{(|S|/2 + 1/2)^4} \\
& + \frac{2c_2 C'}{M \epsilon_n} \exp\left\{ - \frac{\| P_S X (\beta_{S^*} - \beta^*) \|_2^2}{32 \sigma_0^2} \right\} \right]^{1/4} \\
\succeq & \left\{ b_0 + \frac{\alpha + \alpha n}{4} \| P_S X (\beta_{S^*} - \beta^*) \|_2 \right\}^{(|S|/2 + 1/2)} + C' \exp\left\{ - \frac{\| P_S X (\beta_{S^*} - \beta^*) \|_2^2}{128 \sigma_0^2} \right\}. \\
\end{align*}
\]
Then by (44), (49), (50), (51) and \( \| P_S X (\beta_S + \beta^* ) \|_2^2 = \| X_S (\beta_S - (X_S^T X_S)^{-1} X_S^T \beta^*) \|_2^2 \), we have

\[
E_\beta \cdot W_2 
\leq \sum_{S: \| S \|_2 \leq C_S} \left\{ \frac{(\alpha + \gamma)}{2\pi} \| S \|_2 / 2 \| X_S^T X_S \|^{1/2} \frac{\Gamma(a_0 + \alpha n / 2 + |S| / 2)}{\Gamma(a_0 + \alpha n / 2)} \right\} 
\times \int_{R^{|S|}} \left\{ b_0 + \frac{\alpha}{2} \| y - \hat{y}_S \|_2^2 \frac{\Gamma(a_0 + \alpha n / 2 + |S| / 2)}{\Gamma(a_0 + \alpha n / 2)} \right\} \frac{2n^{1/2}}{\exp\left\{ - \frac{\alpha}{128 d_2} M \epsilon \right\}} d\beta_S
\]

\[
+ \sum_{S: \| S \|_2 \leq C_S} \left\{ \frac{(\alpha + \gamma)}{2\pi} \| S \|_2 / 2 \| X_S^T X_S \|^{1/2} \frac{\Gamma(a_0 + \alpha n / 2 + |S| / 2)}{\Gamma(a_0 + \alpha n / 2)} \right\} \frac{2C' n^{1/2}}{\exp\left\{ - \frac{\alpha}{128 d_2} M \epsilon \right\}} d\beta_S
\]

\[
\leq C_S^* \left( \frac{p}{C_S^*} \right)^n \Gamma(a_0 + \alpha n / 2 + |S| / 2) \exp\left\{ - \frac{\alpha}{128 d_2} M \epsilon \right\} \leq \exp\left\{ - C' \epsilon n \right\}
\]

for some \( C' > 0 \), when \( M > C d_2 \).

For \( W_3 \), by (6) and Holder’s inequality, we have

\[
E_\beta \cdot \pi^n (B_{M \epsilon} (S) \| S \)
\]

\[
= \left( \frac{\alpha + \gamma}{2\pi} \right) \| S \|_2 / 2 \| X_S^T X_S \|^{1/2} \frac{\Gamma(a_0 + \alpha n / 2 + |S| / 2)}{\Gamma(a_0 + \alpha n / 2)} \]

\[
\times \int_{B_{M \epsilon} (S)} \left\{ b_0 + \frac{\alpha}{2} \| y - \hat{y}_S \|_2^2 \frac{\Gamma(a_0 + \alpha n / 2 + |S| / 2)}{\Gamma(a_0 + \alpha n / 2)} \right\}^{1/2} \frac{d\beta_S}{\left\{ b_0 + \frac{\alpha}{2} \| y - \hat{y}_S \|_2^2 \right\}^{1/2}}
\]

\[
\leq \left( \frac{\alpha + \gamma}{2\pi} \right) \| S \|_2 / 2 \| X_S^T X_S \|^{1/2} \frac{\Gamma(a_0 + \alpha n / 2 + |S| / 2)}{\Gamma(a_0 + \alpha n / 2)} \]

\[
\times \left\{ b_0 + \frac{\alpha}{2} \| y - \hat{y}_S \|_2^2 \right\}^{1/2} \frac{d\beta_S}{\left\{ b_0 + \frac{\alpha}{2} \| y - \hat{y}_S \|_2^2 \right\}^{1/2}}
\]

\[
\cdot E_\beta \left( \frac{\alpha}{2} \| y - \hat{y}_S \|_2^2 \right)^{1/4} \cdot d\beta_S.
\]
When \( S^* \subseteq S \),
\[
\|y - \hat{y}_S\|^2 \sim \sigma_0^2 \chi_{n-|S|}^2
\]
and
\[
\|X_S(\beta_S - \hat{\beta}_S)\|^2 \|\beta_S \sim \sigma_0^2 \chi_{|S|}^2 (\|X(\beta_{S^*} + \beta^*)\|^2 / \sigma_0^2).
\]
Hence by Lemmas 1 and 2, for \(|S| \leq Cs^*\), on \(B_{M\epsilon_n}(S)\), we have
\[
\|P_S X(\beta_{S^*} - \beta^*)\|^2 \geq M\epsilon_n, \quad P(\|y - \hat{y}_S\|^2 > 2\sigma_0^2 n) \leq \exp(-n/10),
\]
and
\[
\begin{align*}
P(\|X_S(\beta_S - \hat{\beta}_S)\|^2 \leq M\epsilon_n / 4|\beta_S|) \\
\leq P(\|X_S(\beta_S - \hat{\beta}_S)\|^2 \leq \|P_S X(\beta_{S^*} - \beta^*)\|^2 / 2|\beta_S|) \\
\leq \frac{C'\sigma_0^2}{\|P_S X(\beta_{S^*} - \beta^*)\|^2} \exp\left\{-\frac{1}{32}\frac{\|P_S X(\beta_{S^*} - \beta^*)\|^2}{\sigma_0^2} \right\} \\
\leq \frac{2c_2C'}{M\epsilon_n} \exp\left\{-\frac{1}{32}\frac{\|P_S X(\beta_{S^*} - \beta^*)\|^2}{\sigma_0^2} \right\} \leq \frac{2c_2C'}{64\sigma_0^2} \exp\left\{-\frac{M\epsilon_n}{64\sigma_0^2} \right\}.
\end{align*}
\]
Recalling the definitions of \(A, B_1\) and \(B_2\), by the same technique used for \(E\beta_1W_2\), we can prove that when \(M > \frac{256Cd_1}{\alpha}\), \(E\beta_1W_3 \leq \exp(-G\epsilon_n)\), for some \(G > 0\).

We prove the Theorem by taking \(M \geq \max\{(2a + 3)\frac{24d_2}{\alpha}, \frac{256Cd_2}{\alpha}\}\).

**Proof of Theorem 6:**

Since
\[
\pi^n(\beta) = \sum_S \pi^n(\beta_S | S)\pi^n(S).
\]
Hence by convexity of \(H^2\) and \(H^2 \leq 2\), we have
\[
\begin{align*}
H^2(\Pi^n, \{N_{s^*}(\beta_{S^*}, \frac{\sigma_0^2}{\alpha + \gamma}(X_{S^*}^T X_{S^*})^{-1}) \circ \delta_{0S^*}\}) \\
= H^2(\sum_S \pi^n(\beta_S | S)\pi^n(S), N_{s^*}(\beta_{S^*}, \frac{\sigma_0^2}{\alpha + \gamma}(X_{S^*}^T X_{S^*})^{-1}) \circ \delta_{0S^*}) \\
\leq \sum_S \pi^n(S) H^2(\pi^n(\beta_S | S), N_{s*}(\beta_{S^*}, \frac{\sigma_0^2}{\alpha + \gamma}(X_{S^*}^T X_{S^*})^{-1}) \circ \delta_{0S^*}) \\
\leq \sum_{S \neq S^*} 2\pi^n(S) + H^2(\pi^n(\beta_S | S^*), N_{s*}(\beta_{S^*}, \frac{\sigma_0^2}{\alpha + \gamma}(X_{S^*}^T X_{S^*})^{-1}) \circ \delta_{0S^*}) \\
= \sum_{S \neq S^*} 2(1 - \pi^n(S^*)) + H^2(\pi^n(\beta_S | S^*), N_{s*}(\beta_{S^*}, \frac{\sigma_0^2}{\alpha + \gamma}(X_{S^*}^T X_{S^*})^{-1}) \circ \delta_{0S^*}).
\end{align*}
\]
Empirical priors in linear regression with unknown error variance

We also have $E_{\beta^*} \pi^n(S^*) \to 1$ by dominated convergence theorem. So it suffices to show that

$$E_{\beta^*} H^2(\pi^n(\beta_S^*|S^*), N_s^*(\hat{\beta}_S^*, \frac{\sigma_0^2}{\alpha + \gamma}(X_{S^*}^T X_{S^*})^{-1}) \otimes \delta_{0S^*}) \to 0.$$ 

This we prove by showing expectation of the Hellinger affinity

$$E_{\beta^*} \int_{R^d} \sqrt{\pi^n(\beta_S^*|S^*) \cdot N_s^*(\beta_S^*|\hat{\beta}_S^*, \frac{\sigma_0^2}{\alpha + \gamma}(X_{S^*}^T X_{S^*})^{-1})} d\beta_S^* \to 1. \quad (53)$$

To this end, let

$$Q = \{ y : (n-s^*)-\sqrt{n-s^*} \log(n-s^*) \leq \|y-\hat{y}_S^*\|^2/\sigma_0^2 \leq (n-s^*)+\sqrt{n-s^*} \log(n-s^*) \}.$$
Then by $1 + x \leq \exp(x)$,

$$E_{\beta^*} \int_{R^{s^*}} \sqrt{n}(\beta_S^* | S^*) \cdot N_{s^*}(\beta_S^* | \hat{\beta}_S^*, \frac{\sigma_0^2}{\alpha + \gamma} (X_S^T X_S)^{-1}) d\beta_S^*$$

$$\geq (\sigma_0^2)^{-s^*/4} \frac{(\alpha + \gamma)^{s^*/2}}{2\pi} |X_S^T X_S|^{1/2} \frac{\Gamma^{1/2}(a_0 + \alpha/2 + s^*/2)}{\Gamma^{1/2}(a_0 + \alpha/2)}$$

$$\times E_{\beta^*} \int_{R^{s^*}} \frac{1}{Q} \cdot (b_0 + \frac{\alpha}{2} \|y - \hat{y}_S\|^2)^{-s^*/4} \cdot \exp\left\{-\frac{(\alpha + \gamma) \|X_S^* (\beta_S^* - \hat{\beta}_S^*)\|^2}{4\sigma_0^2}\right\}$$

$$\times \exp\left\{-\frac{(\alpha + \gamma) \|X_S^* (\beta_S^* - \hat{\beta}_S^*)\|^2}{4b_0 + 2\alpha\sigma_0^2(n - s^*)/\sqrt{n} - s^*/\log(n - s^*)}\right\} d\beta_S^*$$

$$\geq (\sigma_0^2)^{-s^*/4} \frac{(\alpha + \gamma)^{s^*/2}}{2\pi} |X_S^T X_S|^{1/2} \frac{\Gamma^{1/2}(a_0 + \alpha/2 + s^*/2)}{\Gamma^{1/2}(a_0 + \alpha/2)}$$

$$\times \frac{\exp\left\{-\frac{(\alpha + \gamma) \|X_S^* (\beta_S^* - \hat{\beta}_S^*)\|^2}{2\sigma_0^2(n - 2\sqrt{n} \log n)}\right\} d\beta_S^*}{\exp\left\{-\frac{(\alpha + \gamma) \|X_S^* (\beta_S^* - \hat{\beta}_S^*)\|^2}{2\sigma_0^2(n - 2\sqrt{n} \log n)}\right\} d\beta_S^*}$$

$$= \frac{\Gamma^{1/2}(a_0 + \alpha/2 + s^*/2)}{\Gamma^{1/2}(a_0 + \alpha/2)(\alpha/2)^{s^*/4}} \left[1 + \frac{\log n}{\alpha \sqrt{n}}\right]^{-s^*/4} \left[1 + \frac{\sqrt{n}/(2 \log n) + \alpha \sqrt{n} \log n}{\alpha(n - 2\sqrt{n} \log n)}\right]^{-s^*/2}$$

$$\geq \sqrt{\frac{a_0 + \alpha/2}{\alpha n/2}} \frac{a_0 + \alpha/2 + 1}{\alpha n/2} \ldots \frac{a_0 + \alpha/2 + s^*/2 - 1}{\alpha n/2}$$

$$\geq \left(\frac{a_0 + \alpha/2}{\alpha n/2}\right)^{s^*/4} \rightarrow 1.$$  

(54)

**Proof of Corollary 2:**

Let $\psi = x^T \beta$. Similar to the proof of Theorem 6 we have

$$E_{\beta^*} d_{TV}(\Pi_{\psi}, N(x_S^T \hat{\beta}_S^*, \frac{\sigma_0^2}{\alpha + \gamma} x_S^T (X_S^T X_S)^{-1} x_S)) \rightarrow 0,$$
Empirical priors in linear regression with unknown error variance

where $\Pi^n_\psi$ is the derived posterior distribution of $\psi = x^T \beta$. Hence

$$\sqrt{\frac{\sigma_0^2}{\alpha + \gamma} x^T S^* (X^T S^* X S^*)^{-1} x S^*} \tilde{\beta}_{S^*} + t_{\gamma}$$

is asymptotically equals $\Phi^{-1}(\gamma)$. Then writing $Z$ as a

$\Phi^{-1}(\gamma)$ random variable,

$$P_{\beta^*}(t_{\gamma} \geq x^T \beta^*) = P_{\beta^*}(\frac{x^T \tilde{\beta}_{S^*} - x^T \beta^*}{\sqrt{\frac{\sigma_0^2}{\alpha + \gamma} x^T S^* (X^T S^* X S^*)^{-1} x S^*}} \geq \frac{x^T \tilde{\beta}_{S^*} - t_{\gamma}}{\sqrt{\frac{\sigma_0^2}{\alpha + \gamma} x^T S^* (X^T S^* X S^*)^{-1} x S^*}})$$

$$= P_{\beta^*}(Z \geq \frac{x^T \tilde{\beta}_{S^*} - t_{\gamma}}{\sqrt{\frac{\sigma_0^2}{\alpha + \gamma} x^T S^* (X^T S^* X S^*)^{-1} x S^*}}) \geq 1 - \gamma + o(1), \quad n \to \infty.$$

6 Final Remarks

The paper extends the work of Ryan Martin and his colleagues who proposed empirical prior for linear regression models. The contribution of this article is to extend their work for unknown error variance. The theoretical advancement is the derivation of new results related to model selection consistency, posterior contraction rates as well as a new Bernstein von-Mises theorem in our framework. An important open question is whether similar results can be found in the set up of [7] who proposed Laplace priors for a linear regression problem with known error variance.

References

[1] Martin, R., Mess, R., Walker, S.: Empirical bayes posterior concentration in sparse high-dimensional linear models. Bernoulli 23(3), 1822–1847 (2017)

[2] Martin, R., Walker, S.: Asymptotically minimax empirical bayes estimation of a sparse normal mean vector. Electron. J. Stat 8(2), 2188–2206 (2014)

[3] Martin, R., Ning, B.: Empirical priors and coverage of posterior credible sets in a sparse normal mean model. The Indian Journal of Statistics 82-A, 477–498 (2020)

[4] Martin, R., Tang, Y.: Empirical priors for prediction in sparse high-dimensional linear regression. Journal of Machine Learning Research 21, 1–30 (2020)

[5] Cao, X., Khare, K., Ghosh, M.: High-dimensional posterior consistency for hierarchical non-local priors in regression. Bayesian Anal 15(1), 241–262 (2020)

[6] Shin, M., Bhattacharya, A., Johnson, V.: Functional horseshoe priors for subspace shrinkage. Journal of the American Statistical Association 115(532), 1784–1797 (2019)
Empirical priors in linear regression with unknown error variance

[7] Castillo, I., Schmidt-Hieber, J., van der vaart, A.: Bayesian linear regression with sparse priors. Ann. Statist 43(5), 1986–2018 (2015)

[8] Arias-Castro, E., Lounici, K.: Estimation and variable selection with exponential weights. Electron. J. Stat 8(1), 328–354 (2014)