Bayesian Regression Based on Principal Components for High Dimensional Data

Jaeyong Lee and Hee-Seok Oh

Seoul National University

March 24, 2010

Abstract

Motivated by a climate prediction problem, we consider high dimensional Bayesian regression where the number of covariates is much larger than the number of observations. To reduce the dimension of the covariate, the response is regressed on the principal components obtained from the covariates, and it is argued that the PCA regression is equivalent to the original model in terms of prediction. In the PCA regression setting under the sparsity condition, we examine large sample properties of two different modeling strategies: regression with and without covariate selection. For the regression without covariate selection, we obtain the consistency results of the estimators and posteriors with normal priors with constant and decreasing variances, and James-Stein estimator; for the regression with covariate selection, we obtain convergence rates of Bayesian model averaging (BMA) and median probability model (MPM) estimators, and the posterior with variable selection prior. Based on the large sample properties, we conclude that variable selection is essential in high dimensional Bayesian regression. A simulation study also confirms the conclusion. The methodologies are applied to a climate prediction problem. [1]

[1] *AMS 2000 subject classifications*: Primary 62C10; secondary 62J05

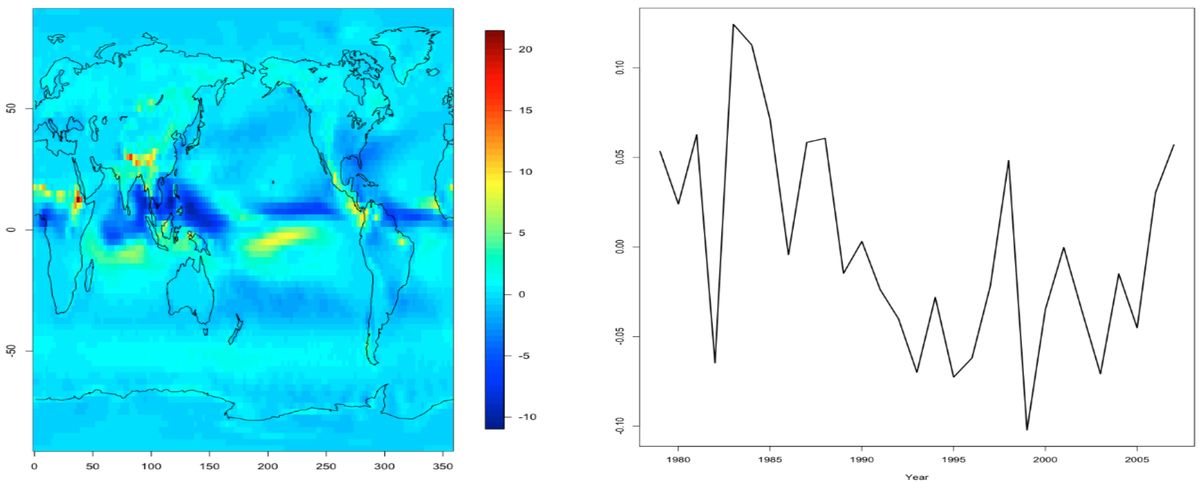
Key words and phrases: High dimensional Bayesian regression; Stein estimator; Bayesian model averaging; median probability model; posterior consistency; posterior convergence rate

1 Introduction

1.1 Motivation

Every year Korea Meteorological Administration (KMA) provides the predicted summer – the period of June, July and August, in short JJA season – precipitation of the next year over the region of Korea. It is an important prediction, for many business and government policy decisions are based on this prediction. The prediction is made through a complicated process, but for our discussion, we can reasonably simplify the problem in the following way. The prediction is based on the past observations and the general circulation model – computer model simulating global climate variables, in short GCM – output covering the whole globe. The past observations are the precipitation of the JJA season over Korea, which we denote by an n-dimensional vector y_n with each coordinate representing the mean precipitation anomaly of a year. The GCM output is denoted by an $n \times q_n$ matrix X_n whose ith row consists of the values of the climate variables covering the whole globe generated from the GCM for the ith year. For the data set we have in mind, y_n is the vector of the JJA precipitation anomalies of 29 years from 1979 to 2007 and X_n contains the GCM precipitation anomalies of 29 years covering the globe by $2.5° \times 2.5°$ grids with total $10,512$ grid points; thus $n = 29$ and $q_n = 10,512$. Figure 1 shows pictorial summaries of the data set. Panel (a) is the JJA season precipitation anomalies generated from GDAPS, a GCM developed by KMA, output of year 2007. Each row of the design matrix can be represented by such plot. Panel (b) of Figure 1 shows the time series plot of the observed precipitation anomalies over Korea, representing the response vector y_n.

We wish to find the regression relation between the observed mean precipitation over Korea and the GCM output covering the whole globe based on the data, y_n and X_n. As a reliable regression equation is obtained, we feed the the GCM output for the next year – note the future GCM output can be obtained by running the GCM for one more year – to the regression equation and make a prediction by the estimated mean value at the future GCM output.



(a) (b)

Figure 1: (a) The GDAPS precipitation anomalies of year 2007; (b) The time series plot of the observed JJA season precipitation anomalies over Korea from year 1979 to year 2007.

For the problem described above, we consider the following linear model

$$y_n = X_n\beta_n + \epsilon_n,\ \ \epsilon_n \sim N(0, \sigma^2 I_n), \tag{1}$$

where β_n is the q_n-dimensional vector of regression coefficients, ϵ_n is the n-dimensional vector of errors, I_n is the $n \times n$ identity matrix, and the error variance σ^2 is assumed to be known for the sake of simplicity.

Since the prediction is based on the regression relation between y_n and X_n, a reliable estimation of the mean of y_n is essential in the process. When we considered this problem, the first question that occurred to us was "when we have only 29 observations with 10,512 covariates, can we have reliable posterior or Bayes estimate of the mean response for prediction?" This is the main motivation of the current study. To answer this question, we formulate the problem in the following way: as $n \to \infty$ and $q_n >> n$, can we have consistent posterior of the mean vector $\mu_n = \mathrm{E}(y_n) = X_n\beta_n$? If so, what is the convergence rate? Since the prediction is made through the mean function of the response, the natural norm for this problem is the Euclidean norm of the mean vector

$$\frac{1}{n}||\hat{\mu}_n - \mu_n||^2 \tag{2}$$

with $\hat{\mu}_n$ being an estimator of μ_n.

In the typical regression setting with fixed number of parameters, it is well known that the posterior is asymptotically optimal: under reasonable conditions, the posterior is consistent, and has $1/\sqrt{n}$ convergence rate. But, when $q_n >> n$, not much of the large sample properties are known for the posterior. Exceptions are Ghosal (1997, 1999) and Jiang (2007). Ghosal (1997, 1999) proved the asymptotic normality of the posterior for linear models if $q_n^4 \log(q_n)/n \to 0$ (e.g. $q_n = O(n^{1/5})$) as $n \to \infty$. Since the goal was the asymptotic normality of the posterior, the growth rate for q_n is quite slow, and it hardly fits the problem we consider. The results of Jiang (2007) is perhaps more relevant to our problem. He considered the generalized linear model with variable selection prior and proved the convergence rate of the posterior under the Hellinger distance is near parametric rate even when $q_n >> n$. However, for our problem, the Euclidean norm of the mean vector is deemed more natural than the Hellinger distance of the the densities.

1.2 PCA Regression

In this section, we describe how the problem posed in the previous section can be transformed to the PCA regression setting. It is clear that the 10,512 regression coefficients can not be estimated based on the 29 observations unless additional strong assumptions are made. In such situations, PCA is often performed to reduce the dimension of the covariates. In our problem, PCA is especially relevant, because climatologists often explain the variations in the climate by principal components and give physical interpretations to them. Interestingly, the problem we posed in the previous subsection can be transformed to the PCA regression setting without any compromise.

Suppose

$$X_n^T X_n = A \Lambda A^T,$$

where $A = [a_1, a_2, \ldots, a_{q_n}]$ is a $q_n \times q_n$ orthogonal matrix, $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{q_n})$ is a $q_n \times q_n$ diagonal matrix with $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_{p_n} > 0 = \lambda_{p_n+1} = \ldots = \lambda_{q_n}$ and (a_i, λ_i) are the ith pair of eigenvector and eigenvalue of $X_n^T X_n$ for $i = 1, 2, \ldots, q_n$. Note that p_n is the rank of X_n and the number of nonzero eigenvalues of $X_n^T X_n$; thus, $p_n \leq \min(n, q_n)$, and we assume $p_n > 0$. The nonzero eigenvalues provide natural partitions of A and Λ such that

$$\begin{aligned} A &= [A_1, A_2] \\ \Lambda &= \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}, \end{aligned}$$

where $A_1 = [a_1, a_2, \ldots, a_{p_n}]$, $A_2 = [a_{p_n+1}, a_{p_n+2}, \ldots, a_{q_n}]$, $\Lambda_1 = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{p_n})$ and $\Lambda_2 = \text{diag}(\lambda_{p_n+1}, \lambda_{p_n+2} \ldots, \lambda_{q_n}) = 0$. Thus,

$$X_n^T X_n = [A_1, A_2] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} A_1^T \\ A_2^T \end{bmatrix} = A_1 \Lambda_1 A_1^T + A_2 \Lambda_2 A_2^T = A_1 \Lambda_1 A_1^T.$$

If all of λ_i's are nonzeros, the response vector y_n is regressed on, $Z_n = X_n A$, the principal components of X_n [Jolliffe (2002)]. For the problem of estimating μ_n under norm (2), it suffices to use the principal components with nonzero eigenvalues as regressors. Multiplying $\Lambda_1^{-1/2} A_1^T X_n^T / \sqrt{n}$ to (1), we obtain

$$t_n = \eta_n + \xi_n, \ \xi_n \sim N\big(0, \frac{\sigma^2}{n} I_{p_n}\big), \tag{3}$$

where

$$\begin{aligned} t_n &= \frac{1}{\sqrt{n}} \Lambda_1^{-1/2} A_1^T X_n^T y_n, \\ \eta_n &= \frac{1}{\sqrt{n}} \Lambda_1^{-1/2} A_1^T X_n^T X_n \beta_n = \frac{1}{\sqrt{n}} \Lambda_1^{1/2} A_1^T \beta_n, \\ \xi_n &= \frac{1}{\sqrt{n}} \Lambda_1^{-1/2} A_1^T X_n^T \epsilon_n. \end{aligned}$$

An interesting point is that norm (2) depends only on η_n. For an estimator $\hat{\beta}_n$ of β_n, let $\hat{\mu}_n = X_n \hat{\beta}_n$ and $\hat{\eta}_n = \Lambda^{1/2} A_1^T \hat{\beta}_n / \sqrt{n}$. Then,

$$\begin{aligned} \frac{1}{n} ||\hat{\mu}_n - \mu||^2 &= \frac{1}{n} ||X_n \hat{\beta}_n - X_n \beta_n||^2 \\ &= \frac{1}{n} (\hat{\beta}_n - \beta_n)^T A_1 \Lambda_1 A_1^T (\hat{\beta}_n - \beta_n) \\ &= ||\hat{\eta}_n - \eta_n||^2. \end{aligned} \tag{4}$$

Thus, consistency and convergence rate under model (3) with norm (4) are equivalent to those under model (1) with norm (2), and this equivalence is independent of q_n. Conversely, for a given estimator $\hat{\eta}_n$ of η, we can also reconstruct the estimator of β_n by

$$\hat{\beta}_n = \sqrt{n} A_1 \Lambda_1^{-1/2} \hat{\eta}_n. \tag{5}$$

Note that with the reconstructed estimator $\hat{\beta}_n$ of (5) we have the same norm, i.e.,

$$\frac{1}{n} ||X_n \hat{\beta}_n - X_n \beta_n||^2 = ||\hat{\eta}_n - \eta_n||^2.$$

For the rest of the paper, we consider model (3) with norm (4).

In order to investigate large sample properties of estimators and posteriors, we assume a sequence of true parameter values (η_n^0) satisfies sparsity condition: for some fixed integer $r > 0$, $\eta_{ni}^0 = 0$ for all $i > r$. The probability and expectation under the true model are denoted by P_n^0 and E_n^0. An estimator $\hat{\eta}_n$ of η_n is said to be consistent in probability (a.s.) or in L_2-norm, if, as $n \to \infty$, $||\hat{\eta}_n - \eta_n^0||^2 \to 0, P_n^0 - \mathit{probability}$ $(P_n^0 - \mathit{a.s.})$ or $\mathrm{E}_n^0[||\hat{\eta}_n^0 - \eta_n^0||^2] \to 0$, respectively. The prior is denoted by π_n and the posterior probability and expectation are denoted by $\pi_n(\cdot|y_n)$ (or $P_n(\cdot|y_n)$) and $\mathrm{E}_n(\cdot|y_n)$, respectively. The posterior is said to be consistent at (η_n^0) P_n^0-a.s. (or in P_n^0-probability) if, for all $\epsilon > 0$, $P_n^0(||\eta_n - \eta_n^0||^2 > \epsilon|y_n) \to 0$, P_n^0-a.s. (or in P_n^0-probability), as $n \to \infty$. The posterior convergence rate is defined as the smallest possible rate ϵ_n tending to 0 such that, as $n \to \infty$, $P_n(||\eta_n - \eta_n^0|| > \epsilon_n|y_n) \to 0$, in P_n^0 - probability.

1.3 Methodologies and Main Results

After the regression model (1) is transformed to the PCA regression model (3) with norm (4), the question we initially asked is posed in the question of priors. Especially we are interested in the following questions: "Can we use the usual prior ignoring the fact p_n increases as $n \to \infty$?" and "Is there evidence that the Bayesian model averaging is better in the performance?" To answer these questions, we consider Bayesian methodologies in two categories: regressions without covariate selection and with covariate selection.

First, we consider the regression without covariate selection. In this case, we examine the consistency of posteriors with normal priors having constant variances and decreasing variances. The prior with constant variances produces consistent posterior only when $p_n/n \to 0$, and is inadequate for the case p_n grows faster rate than that. The prior with decreasing variances does render the consistency of posterior if the prior variance decreases faster than $1/k^2$, where k is the index of covariate. But, its practical implication is that we need to know the covariates' order of importance before we fit the regression model; thus, the prior with decreasing variances is not appropriate for many problems where there is no apparent ordering of covariates. Additionally, we considered James-Stein estimator [James and Stein (1961)] which was successful in many applications. We were curious if the James-Stein estimator can handle the case $p_n \to \infty$. It turns out that James-Stein estimator performs similarly to the Bayesian model with constant variances and it is not suitable for our situation.

Next, we consider the variable selection prior. Under this prior, we considered two estimators: Bayesian model averaging (BMA) estimator [Hoeting et al. (1999)] and median probability model (MPM) estimator [Barbieri and Berger (2004)]. Under a sparsity assumption, these two estimators and the posterior are all consistent even for the case $p_n = n$ and have similar convergence rates. Bayesians have argued for long time that the BMA is superior for the prediction, but there seldom exists theoretical support for it. We believe this paper is one of first few theoretical support for BMA.

In conclusion, if $p_n/n \to 0$ as $n \to \infty$, as long as the consistency is concerned, the Bayesian modeler can ignore the fact $p_n \to \infty$ and use the priors as if p_n is fixed; but if p_n increases faster and $p_n/n \not\to 0$, one must include the variable selection in the data analysis.

The plan for the paper is as follows. In sections 2 and 3, we present theoretical results of the cases regression without covariate selection and with covariate selection, respectively. A simulation study is shown in section 4 to support the theoretical results in sections 2 and 3. The brief result of the analysis of the climate problem we described is presented in section 5. All the proofs are given in section 6.

2 Regression without Covariate Selection

2.1 Normal Priors with Constant Variances

In this subsection, we consider the model (3) with prior

$$\eta_n \sim N\big(0, \frac{\sigma^2}{\alpha_n} I_{p_n}\big), \tag{6}$$

where α_n is positive for all n. Each coordinate η_{ni} has the same variance but it can vary as n grows. If $p_n = q_n$, this prior is the same as the g-type prior of β_n [Zellner (1986)]

$$\beta_n \quad \sim \quad N\big(0, \frac{\sigma^2}{\alpha_n}(\frac{1}{n}X_n^T X_n)^{-1}\big).$$

Thus, this prior with $\alpha_n = \alpha > 0$ corresponds to the Bayesian model fitting ignoring the large number of covariates.

For the simplicity of the exposition, we consider three cases for the sequence (α_n): $\lim_{n\to\infty} \alpha_n/n = 0$, $\lim_{n\to\infty} \alpha_n/n = \alpha^*$ for some $\alpha^* > 0$, and $\lim_{n\to\infty} \alpha_n/n = \infty$. In addition, we consider two cases of p_n: $\lim_{n\to\infty} p_n/n = 0$, and $\lim_{n\to\infty} p_n/n = p^*$ with $0 < p^* \leq 1$.

Theorem 1. The following are equivalent.

(a) The Bayes estimator $\tilde{\eta}_n$ is consistent, P_n^0-a.s.

(b) The posterior of η_n is consistent, P_n^0-a.s.

(c) One of the following three cases holds: (i). $\lim_{n\to\infty} \alpha_n/n = 0$ and $\lim_{n\to\infty} p_n/n = 0$; (ii). $\eta_{ni}^0 = 0$ for all i and $\lim_{n\to\infty} \alpha_n/n = \infty$; and (iii). $\eta_{ni}^0 = 0$ for all i and $\lim_{n\to\infty} p_n/n = 0$.

Remark 1. When the number of covariates p is fixed, one typically assumes $X_n^T X_n/n$ converges to a positive definite square matrix. Thus, the fixed number of covariates case corresponds to case (i) of (c) in Theorem 1. As far as the consistency of the Bayes estimator is concerned, the increasing sequence of p can be treated the same as the fixed p if $p/n \to 0$ as $n \to \infty$.

Remark 2. If $p/n \not\to 0$ and some of η_{ni}^0 are nonzeros, the usual Bayes estimator is inconsistent for whatever sequence of α. This means that if we ignore the large number of covariates and fit the Bayesian regression model casually, even the consistency is not warranted.

2.2 Normal Priors with Decreasing Variance Priors

In nonparametric regression model with infinite number of basis functions, the prior of the regression coefficients often has decreasing variances [Lenk (1999)]. We examine the following prior for our problem with the same spirit

$$\eta_n \sim N(0, \sigma^2 V_n^{-1}),$$

where $V_n = \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_p)$ and $\alpha_k = \alpha(k)$ for some positive increasing function α defined on R_+. The posterior of η_n given y_n is that $\eta_{n1}, \ldots, \eta_{np}$ are independent and

$$\eta_{ni}|y_n \sim N\Big(\frac{n}{n+\alpha_i} t_{ni}, \frac{\sigma^2}{n+\alpha_i}\Big),\ i = 1, 2, \ldots, p.$$

Theorem 2. If $\alpha(k)$ increases faster rate than k^2, the Bayes estimator and the posterior are consistent, P_n^0-a.s.

Remark 3. Although we obtain consistency for the normal prior with decreasing variances, the prior variance structure $V_n = \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is not satisfactory in practice, because it suggests that before we see the data we already know the importance of the covariates and we can order the covariates by its importance. Thus, this prior variance structure can not be applied in most applications where the importance of the covariates are equal.

2.3 Stein Estimation

Multiplying $\sqrt{n}/\sigma$ to the both sides of (3), we obtain

$$\frac{\sqrt{n}}{\sigma} t_n \sim N\Big(\frac{\sqrt{n}}{\sigma}\eta_n, I_p\Big). \tag{7}$$

Setting $X = \frac{\sqrt{n}}{\sigma} t_n$ and $\theta = \frac{\sqrt{n}}{\sigma}\eta_n$, we obtain the standard multivariate normal model with known variance:

$$X \sim N(\theta, I_p). \tag{8}$$

The estimation problem of θ in model (8) has been the focal point of the celebrated James-Stein estimation. The Stein estimation has been successfully applied to many problems and we are curious about the effect of shrinkage of the James-Stein estimator on the regression model with increasing number of covariates. In this subsection, we study the large sample behavior of the James-Stein estimator in the regression model setting. The James-Stein estimator of θ is given by

$$\hat{\theta}^{JS} = \Big(1 - \frac{p-2}{||X||^2}\Big)X,$$

from which we obtain the James-Stein estimator for η_n,

$$\begin{aligned} \hat{\eta}_n^{JS} &= \Big(1 - \frac{p-2}{||X||^2}\Big)\frac{\sigma}{\sqrt{n}}X \\ &= \Big(1 - \frac{(p-2)\sigma^2}{n||t_n||^2}\Big)t_n. \end{aligned} \tag{9}$$

Theorem 3. Suppose $A = \lim_{n\to\infty} A_n$ with $A_n = ||\eta_n^0||^2/\sigma^2$.

(a) Suppose $\lim_{n\to\infty} p/n = c \in (0, 1]$. The James-Stein estimator $\hat{\eta}_n^{JS}$ is inconsistent in L_2-norm if $A > 0$, and is consistent if $A = 0$.

(b) Suppose $\lim_{n\to\infty} p/n = 0$. The James-Stein estimator $\hat{\eta}_n^{JS}$ is consistent in L_2-norm.

(c) The consistency and inconsistency in L_2-norm in the parts (a) and (b) can be replaced by consistency and inconsistency in probability, respectively.

Remark 4. The performance of the James-Stein estimator is similar to the Bayes estimator of the normal prior with constant variances: if $p/n \to 0$, it is consistent; but if $p/n \not\to 0$, it is not.

3 Regression Methods with Covariate Selection

In this section, we consider the variable selection prior on η_n. Specifically, we consider the prior

$$\eta_{ni} \sim \pi_0\delta_0 + (1-\pi_0)N\Big(0, \frac{\sigma^2}{\alpha}\Big),\ i = 1, 2, \ldots, p, \tag{10}$$

and $\eta_{n1}, \eta_{n2}, \ldots, \eta_{np}$ are independent a priori. Here δ_0 is the degenerate probability measure at 0, and $\pi_0 \in (0,1)$ is the prior probability of η_{ni} being 0, which is typically set to 1/2. With prior (10) and model (3), the posterior of η_n given y_n is defined as follows: $\eta_{n1}, \eta_{n2}, \ldots, \eta_{np}$ are independent and

$$\eta_{ni}|y_n \sim \pi_0(t_{ni})\delta_0 + \pi_1(t_{ni}) N\Big(\frac{n}{\alpha+n} t_{ni}, \frac{\sigma^2}{\alpha+n}\Big), \ i = 1, 2, \ldots, p, \tag{11}$$

where $\pi_0(t)$ and $\pi_1(t)$ are function defined on $t \in R$ such that $\pi_1(t) = 1 - \pi_0(t)$ and

$$\pi_0(t) = \frac{\pi_0 \frac{\sqrt{n}}{\sigma}\phi\big(\frac{\sqrt{n}}{\sigma}t\big)}{\pi_0 \frac{\sqrt{n}}{\sigma}\phi\big(\frac{\sqrt{n}}{\sigma}t\big) + \pi_1 \frac{1}{\sigma\sqrt{1/n+1/\alpha}}\phi\big(\frac{1}{\sigma\sqrt{1/n+1/\alpha}}t\big)}.$$

Let

$$B_{01}(t) = \frac{\frac{\sqrt{n}}{\sigma}\phi\big(\frac{\sqrt{n}}{\sigma}t\big)}{\frac{1}{\sigma\sqrt{1/n+1/\alpha}}\phi\big(\frac{1}{\sigma\sqrt{1/n+1/\alpha}}t\big)}$$

and $B_{10}(t) = 1/B_{01}(t)$. We now consider the performance of the posterior (11) and two Bayes estimators associated with variable selection: the Bayesian model averaging (BMA) estimator and median probability model (MPM) estimator. The BMA estimator of η_n, $\hat{\eta}_n^{BMA}$ is the posterior mean of η_n:

$$\hat{\eta}_n^{BMA} = \pi_1(t_n)\frac{n}{\alpha+n} t_n, \tag{12}$$

where $\pi_i(t_n) = (\pi_i(t_{n1}), \pi_i(t_{n2}), \ldots, \pi_i(t_{np}))^T$ for $i = 0, 1$. Since the inclusion probability of variable i is $\pi_1(t_{ni})$, the MPM estimator of η_n, $\hat{\eta}_n^{MPM}$ is

$$\hat{\eta}_n^{MPM} = \frac{n}{\alpha+n} t_n I(\pi_1(t_n) > 1/2), \tag{13}$$

where the values of the indicator function and π_i with a vector argument is the vector whose coordinates are the functions applied to the vector. With the assumption of $\pi_0 = \pi_1 = 1/2$,

$$\hat{\eta}_n^{MP} = \frac{n}{\alpha+n} t_n I(B_{01}(t_n) < 1).$$

Theorem 4. Under model (3) with prior (10), the following results hold:

(a) As $n \to \infty$, $\displaystyle\sum_{i=1}^{p} (\hat{\eta}_{ni}^{BMA} - \eta_{ni}^0)^2 = O_p\Big(\frac{1}{n}\Big) + O_p\Big(\frac{p(\log n)^{3/8}}{n^{3/2}}\Big)$.

(b) As $n \to \infty$, $\sum_{i=1}^{p} (\hat{\eta}_{ni}^{MPM} - \eta_{ni}^{0})^2 = O_p\Big(\frac{1}{n}\Big) + O_p\Big(\frac{p(\log n)^{1/2}}{n^{3/2}}\Big)$.

(c) For a sequence of positive real numbers, (c_n) converging to 0,

$$\pi(||\eta - \eta^0||^2 > c_n | y_n) = \frac{1}{c_n}\Big(O_p(1/n) + O_p\Big(\frac{p(\log n)^{3/8}}{n^{3/2}}\Big)\Big).$$

Remark 5. Unlike the methods of the regression without covariate selection, the BMA and MPM estimators, and the posterior with variable selection prior are all consistent in P_n^0-probability for $p_n \leq n$.

Remark 6. The BMA and MPM estimators and the posterior have similar convergence rates up to the small difference of $\log(n)$ factor.

Remark 7. If roughly $p = o(\sqrt{n})$ — to be exact, $p = O_p\Big(\frac{\sqrt{n}}{(\log n)^{3/8}}\Big)$ for the BMA and the posterior and $p = O_p\Big(\frac{\sqrt{n}}{(\log n)^{1/2}}\Big)$, the BMA and MPM estimators and the posterior have the parametric rate $1/\sqrt{n}$.

Remark 8. If $p = n$, the convergence rates of the BMA and MPM estimators and the posterior are roughly $1/n^{1/4}$ — to be exact, they are $\frac{(\log n)^{3/16}}{n^{1/4}}$ for the BMA estimator and the posterior and $\frac{(\log n)^{1/4}}{n^{1/4}}$ for the MPM estimator.

We finish the section by studying the asymptotic behavior of the posterior probability of choosing the true model,

$$\delta_n = \prod_{i=1}^{r} \pi_1(t_{ni}) \prod_{i=r+1}^{p} \pi_0(t_{ni}).$$

Moreno, Girón and Casella (2010) provides similar result for two competing models when $p \to \infty$ as $n \to \infty$. Our result covers the case when there are 2^p competing models, but with a restricted assumption on the model.

Theorem 5. Consider model (3) with prior (10). If $p = o\big(\frac{\sqrt{n}}{(\log n)^{1/8}}\big)$, the posterior probability of choosing the true model, δ_n converges to 1 in probability.

Remark 9. If roughly $p = o(\sqrt{n})$ — to be exact, $p = o\Big(\frac{\sqrt{n}}{(\log n)^{1/8}}\Big)$, the posterior choose the true model eventually.

4 A Simulation Study

In this section, we show the finite sample performance of the procedures through a simulation study. The data set is generated 1000 times from the following orthogonal model

$$t_n = \eta_n + \xi_n,\ \ \xi_n \sim N\Big(0, \frac{\sigma^2}{n} I_p\Big),$$

where t_n, ξ_n are p-dimensional real values vectors, and σ^2 is fixed at 1. The true parameter η is set at $(3, 2, 1, 0, \ldots, 0)$. The first three are $(3, 2, 1)$ and the rest are zeros. In actual simulation, the η is randomly permuted. For each data set, six estimators are computed and for each estimator,

$$MSE = ||\hat{\eta}_n - \eta||^2$$

is computed. The MSE is averaged over 1000 data sets. The Bayes estimator for the normal prior with constant variances is

$$\hat{\eta}_n^{CV} = \frac{n}{n+\alpha} t_n,$$

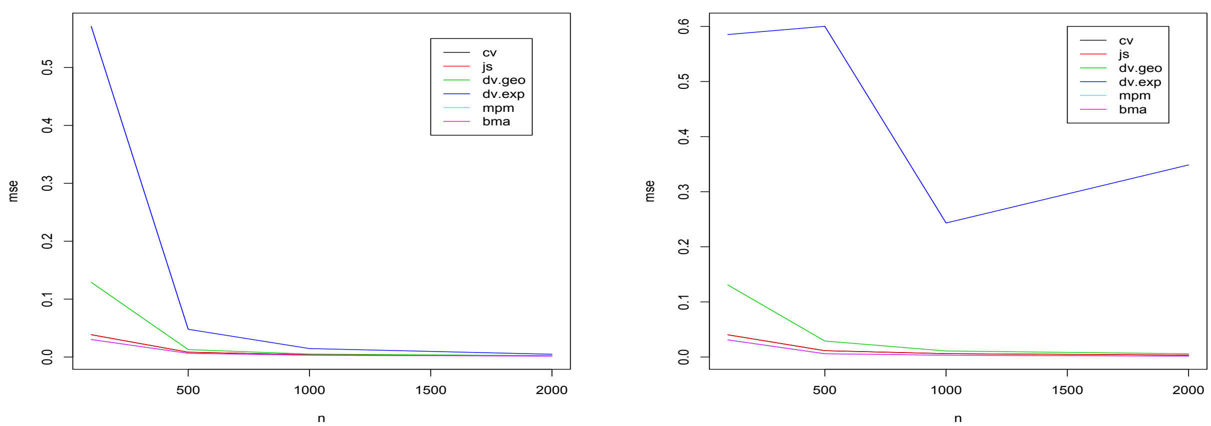
where $\alpha = 0.1$, and the James-Stein estimator is

$$\hat{\eta}_n^{JS} = \Big(1 - \frac{(p-2)\sigma^2}{n||t_n||^2}\Big) t_n.$$

For the normal prior with decreasing variances, we take $\alpha(x) = \alpha x^3$ and αe^x with $\alpha = 0.1$ which represent geometric and exponential decay rates of variances, respectively. The corresponding Bayes estimators are

$$\begin{aligned}
\hat{\eta}_{ni}^{DVGeo} &= \frac{n}{n + \alpha i^3} t_{ni}, \\
\hat{\eta}_{ni}^{DVExp} &= \frac{n}{n + \alpha e^i} t_{ni}.
\end{aligned}$$

The numbers of observations n are chosen as 100, 500, 1000, and 2000, to emulate the situation with increasing n. The MSEs are summarized in Figures 2 - 5. The number of covariates are chosen to represent the five different rates of p, $p = 4$, $\log(n)$, $\sqrt{n}$, $n/2$ and n. For the three rates of $p = 4, \log(n)$, and $\sqrt{n}$ with $p/n \to 0$, the MSE of all estimators decreases as $n \to \infty$; however, for the last two cases of $p = n/2$, and n, it is apparent that the MSEs of the MPM and BMA estimators tend to zero, but the MSEs of the other estimators do not converge to zero. In all cases, the MPM and BMA estimators are better than the others, and these two provide almost identical results. Although the theory indicates consistency of the Bayes estimators with the prior with decreasing variances, the simulation results show no signs of consistency for these estimators, which is because we randomly permute η_n^0 before we generate the data. This shows the danger of priors with decreasing variances. If one does not have a good idea of importance of covariates, priors with decreasing variances are not recommended.

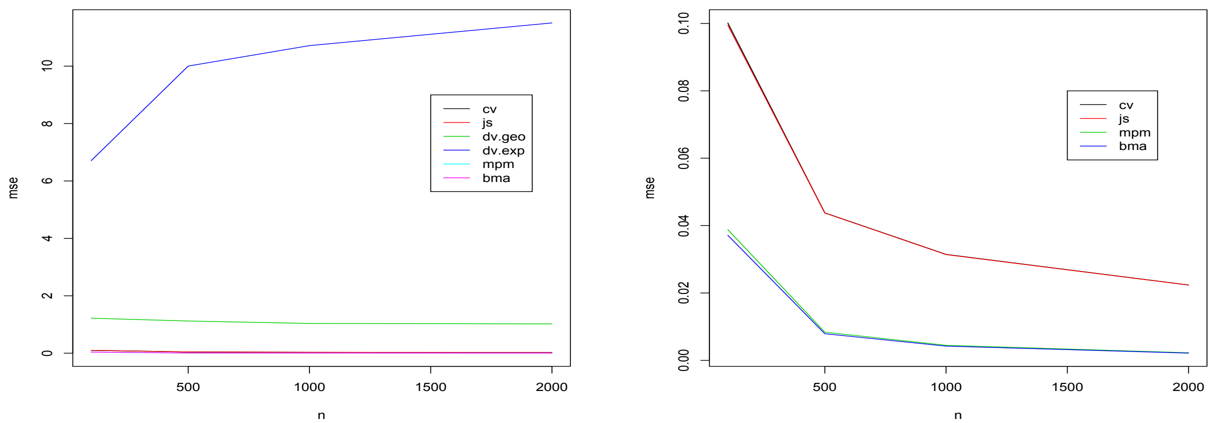


(a) (b)

Figure 2: The scatter plots of MSEs of 6 estimators with sample sizes $n = 100, 500, 1000, 2000$ when $p = 4$ (panel (a)) and when $p = \log(n)$ (panel (b)).

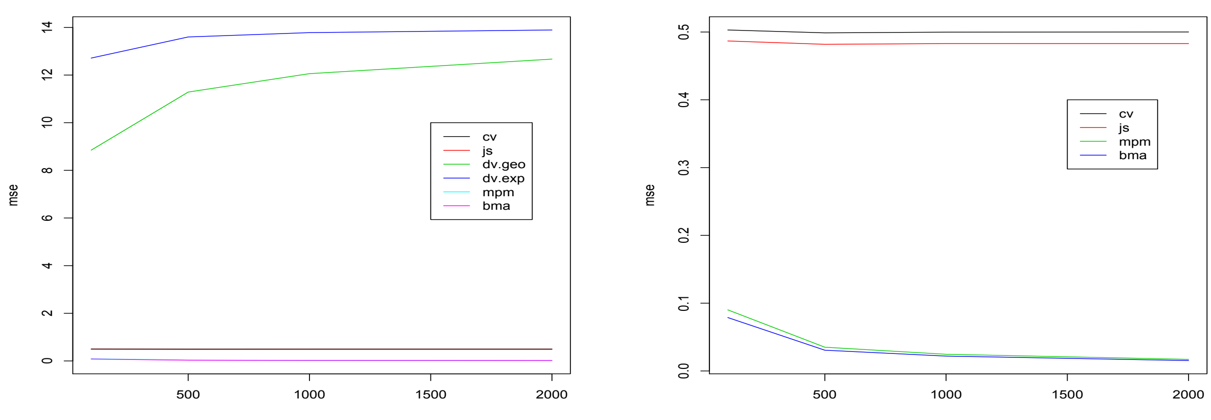
5 Predicting Summer Precipitation over Korea

We have applied the seven predictors – six predictors shown in the simulation study together with the GCM predictor whose predicted value is obtained by directly using the GCM ouput – to the prediction problem we described in Section 1. In the actual application, we have 20 ensemble runs from the GCM; thus, for each prediction method, we have 20 predicted values each of which is obtained by applying the prediction method to a



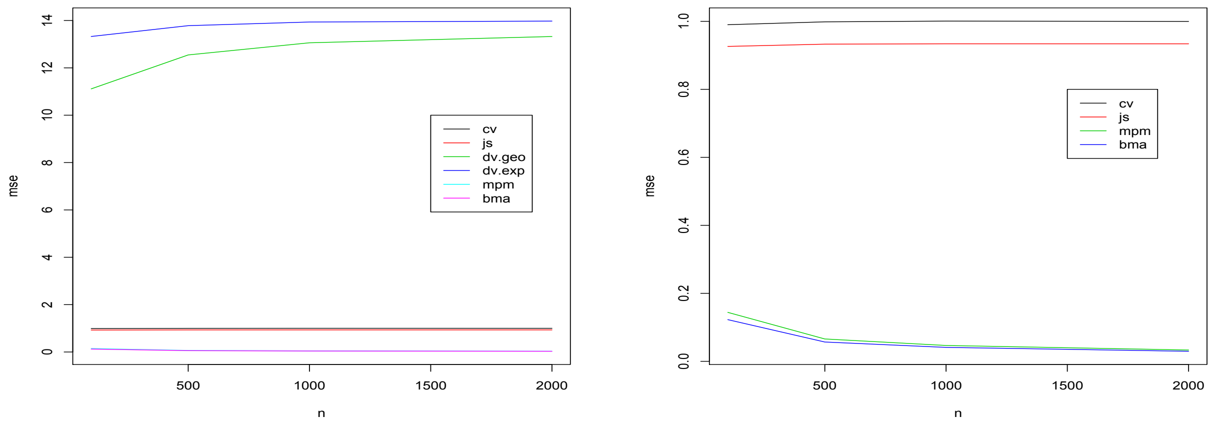
(a) (b)

Figure 3: (a) MSEs of 6 estimators when $n = 100, 500, 1000, 2000$ when $p = \sqrt{n}$; (b) the same plot showing MSEs of 4 estimators.



(a) (b)

Figure 4: (a) MSEs of 6 estimators when $n = 100, 500, 1000, 2000$ when $p = n/2$; (b) the same plot showing MSEs of 4 estimators.



(a) (b)

Figure 5: (a) MSEs of 6 estimators when $n = 100, 500, 1000, 2000$ when $p = n$; (b) the same plot showing MSEs of 4 estimators.

pair of one ensemble run and the actual observations of precipitation. The actual predicted value is obtained by averaging the 20 predicted values. To compare the performance of the prediction methods, we used cross-validation. Table 1 shows the cross-validation error of the seven prediction methods. The comparison result is not as obvious as the simulation result, and James-Stein MPM and BMA estimators show similar performance.

Table 1: Cross-validation errors of seven prediction methods applied to the prediction problem of the summer precipitation over Korea

	GCM	CV	DV Geo	DV Exp	James-Stein	BMA	MPM
cv error	0.5980	0.5221	0.4931	0.4890	0.4556	0.4660	0.4532

6 Proofs of Main Results

In this section, we present the proofs of the theorems given in sections 2 and 3. Unless stated otherwise, Z with subscript denotes sequences (or arrays) of independent standard normal random variables.

6.1 Proof of Theorem 1

In this subsection, we consider model (3) with prior (6). The posterior is $\eta_n|y_n \sim N\big(\frac{n}{\alpha_n+n}t_n, \frac{\sigma^2}{\alpha_n+n}I_{p_n}\big)$. Before we proceed to the proof of Theorem 1, we need the following lemma.

Lemma 1. The sequence of the posterior $(||\eta_n - \eta_n^0||^2|y_n)$ is uniformly integrable $P_n^0 - a.s.$

Proof. It suffices to show

$$\sup_n \mathrm{E}(||\eta_n - \eta_n^0||^4|y_n) < \infty,\ P_n^0 - a.s.$$

We bound the expectation by

$$\begin{aligned}\mathrm{E}(||\eta_n - \eta_n^0||^4|y_n) &\leq 8\mathrm{E}\big(||\eta_n - \mathrm{E}(\eta_n|y_n)||^4|y_n\big) + 8||\mathrm{E}(\eta_n|y_n) - \eta_n^0||^4 \\ &=: 8I_1 + 8I_2.\end{aligned}$$

Since

$$\begin{aligned}I_1 &\leq 2\sum_{i=1}^{p}\mathrm{E}\Big((\eta_{ni} - \mathrm{E}(\eta_{ni}|y_n))^4|y_n\Big) \\ &= \frac{6p\sigma^4}{(\alpha_n+n)^2} \leq \frac{6p\sigma^4}{n^2},\end{aligned}$$

$\sup_n I_1 < \infty$ with P_n^0-a.s. Now, set $t_{ni} = \eta_{ni}^0 + \frac{\sigma}{\sqrt{n}}Z_{ni}$ with P_n^0 probability 1 where $Z_{ni} \sim N(0,1)$. Note

$$\begin{aligned}I_2 &\leq 2\sum_{i=1}^{p}\Big(\frac{n}{\alpha_n+n}(\eta_{ni}^0 + \frac{\sigma}{\sqrt{n}}Z_{ni}) - \eta_{ni}^0\Big)^4 \\ &\leq 16\sum_{i=1}^{p}\frac{\alpha_n^4}{(\alpha_n+n)^4}(\eta_{ni}^0)^4 + 16\sum_{i=1}^{p}\frac{n^2\sigma^4}{(\alpha_n+n)^4}Z_{ni}^4 \\ &\leq 16\sum_{i=1}^{r}(\eta_{ni}^0)^4 + 16\frac{p\sigma^4}{n^2}\frac{1}{p}\sum_{i=1}^{p}Z_{ni}^4.\end{aligned}$$

Thus, $\sup_n I_2 < \infty$ with P_n^0-a.s. This completes the proof. □

Proof of Theorem 1. Equivalence of (a) and (c). Denote $\xi_{ni} = \frac{\sigma}{\sqrt{n}} Z_{ni}$ for all n and i where Z_{ni} are independent standard normal random variables. We have

$$\begin{aligned} ||\tilde{\eta}_n - \eta_n^0||^2 &= \sum_{i=1}^{r} \Big(- \frac{\alpha/n}{1+\alpha/n} \eta_{ni}^0 + \frac{1}{1+\alpha/n} \frac{\sigma}{\sqrt{n}} Z_{ni} \Big)^2 + \sum_{i=r+1}^{p} \Big(\frac{1}{1+\alpha/n} \frac{\sigma}{\sqrt{n}} Z_{ni} \Big)^2 \\ &=: I_1 + I_2. \end{aligned}$$

Note

$$\lim_{n\to\infty} I_1 = \begin{cases} 0, & \text{if } \lim_{n\to\infty} \alpha/n = 0 \\ \sum_{i=1}^{r} \left(\frac{\alpha^*}{1+\alpha^*} \right)^2 (\eta_{0i}^0)^2, & \text{if } \lim_{n\to\infty} \alpha/n = \alpha^* \\ \sum_{i=1}^{r} (\eta_{0i}^0)^2, & \text{if } \lim_{n\to\infty} \alpha/n = \infty, \end{cases}$$

where the limit of the LHS is P_n^0-a.s. Thus, $I_1 \to 0$, P_n^0-a.s if and only if either $\alpha/n \to 0$ or $\eta_{0i}^0 = 0$ for all i. Note

$$I_2 = \frac{1}{(1+\alpha/n)^2} \sigma^2 \frac{p}{n} \frac{1}{p} \sum_{i=r+1}^{p} Z_{ni}^2.$$

By the almost sure convergence result of a triangular array of independent random variables [Hu, Móricz and Taylor (1989); Li, Rao, and Tomkins (1995)], $\frac{1}{p} \sum_{i=r+1}^{p} Z_{ni}^2 \to 1$, P_n^0-a.s. Thus, $I_2 \to 0$, P_n^0-a.s. if and only if either $p/n \to 0$ or $\alpha/n \to \infty$. Combining the results, we get the conclusion.

Equivalence of (b) and (c). By Lemma 1, the P_n^0-a.s convergence of $\mathrm{E}(||\eta_n - \eta_n^0||^2 | y_n)$ is equivalent to the convergence of the posterior. Note

$$\mathrm{E}(||\eta_n - \eta_n^0||^2 | y_n) \;=\; tr\left(\mathrm{Var}(\eta_n | y_n)\right) + ||\tilde{\eta}_n - \eta_n^0||^2$$

and

$$tr\left(\mathrm{Var}(\eta_n | y_n)\right) = \frac{p\sigma^2}{\alpha + n} \le \frac{p/n}{\alpha/n + 1} \sigma^2 \le \max\left(\frac{p}{n}, \frac{1}{\alpha/n}\right) \sigma^2.$$

For all the cases of Theorem 1, $tr\left(\mathrm{Var}(\eta_n | y_n)\right) \to 0$, P_n^0-a.s. Thus, the convergence of $||\tilde{\eta}_n - \eta_n^0||^2$ is equivalent to the convergence of the posterior. This completes the proof. □

6.2 Proof of Theorem 2

For the proof of Theorem 2, we need the following lemma.

Lemma 2. If $\alpha(k)$ increases faster rate than k^a for some $a > 2$,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{n}{(n+\alpha_i)^2} Z_{ni}^2 = 0, \ P_n^0 - a.s.$$

Proof. Let $k_n = n^b$ with $0 < b < 1$ and $ab - 2 > 0$. Then,

$$\sum_{i=1}^{k_n} \frac{n}{(n+\alpha_i)^2} Z_{ni}^2 \le \frac{k_n}{n} \frac{1}{k_n} \sum_{i=1}^{k_n} Z_{ni}^2 \to 0, \ P_n^0 - a.s.$$

We will show

$$\sum_{i=k_n+1}^{n} \frac{n}{(n+\alpha_i)^2} Z_{ni}^2 \to 0, \ P_n^0 - a.s.$$

It suffices to show, by the Borel-Cantelli Lemma, for all $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P\big(\sum_{i=k_n+1}^{n} \frac{n}{(n+\alpha_i)^2} Z_{ni}^2 > \epsilon \big) < \infty.$$

Using a bound for the tail probability of the normal distribution, we have

$$\begin{aligned}
P\big(\sum_{i=k_n+1}^{n} \frac{n}{(n+\alpha_i)^2} Z_{ni}^2 > \epsilon \big) &\le 2 \sum_{i=k_n+1}^{n} P\Big(Z_{ni} > \sqrt{\frac{\epsilon}{(n-k_n)n}} (n+\alpha_i) \Big) \\
&\le 2(n-k_n) \frac{1}{\sqrt{2\pi}} \frac{\sqrt{(n-k_n)n}}{\sqrt{\epsilon}(n+\alpha(k_n))} \exp\Big(- \frac{\epsilon(n+\alpha(k_n))^2}{2n(n-k_n)} \Big) \\
&\le \sqrt{\frac{2}{\pi\epsilon}} n \exp\big(- \frac{\epsilon}{2} \frac{\alpha^2(k_n)}{n^2} \big) \\
&\le \sqrt{\frac{2}{\pi\epsilon}} n \exp\big(- \frac{\epsilon}{2} n^{ab-2} \big).
\end{aligned}$$

Since $ab > 2$,

$$\sum_{n=1}^{\infty} P\big(\sum_{i=k_n+1}^{n} \frac{n}{(n+\alpha_i)^2} Z_{ni}^2 > \epsilon \big) \le \sum_{n=1}^{\infty} \sqrt{\frac{2}{\pi\epsilon}} n \exp\big(- \frac{\epsilon}{2} n^{ab-2} \big) < \infty.$$

This completes the proof. □

Proof of Theorem 2. (a) Note

$$||\tilde{\eta}_n - \eta_n^0||^2 = \sum_{i=1}^{r} \Big(-\frac{\alpha_i}{n+\alpha_i} \eta_{ni}^0 + \frac{n}{n+\alpha_i} \frac{\sigma}{\sqrt{n}} Z_{ni} \Big)^2 + \sigma^2 \sum_{i=r+1}^{p} \frac{n}{(n+\alpha_i)^2} Z_{ni}^2.$$

Since r is a fixed integer, the first term converges to 0, P_n^0-a.s. The second term also converges to 0, P_n^0-a.s. by Lemma 2.

(b) Since

$$\mathrm{E}(||\eta_n - \eta_n^0||^2|y_n) = tr\,(\mathrm{Var}(\eta_n|y_n)) + ||\tilde{\eta}_n - \eta_n^0||^2,$$

and by Theorem 2, it suffices to show, as $n \to \infty$,

$$tr\,(\mathrm{Var}(\eta_n|y_n)) = \sigma^2 \sum_{i=1}^{p} \frac{1}{n+\alpha_i} \to 0.$$

Let $k_n = n^b$ with $1/a < b < 1$. Then,

$$\sum_{i=1}^{n} \frac{1}{n+\alpha_i} = \sum_{i=1}^{k_n} \frac{1}{n+\alpha_i} + \sum_{i=k_n+1}^{n} \frac{1}{n+\alpha_i}.$$

Note, as $n \to \infty$,

$$\sum_{i=1}^{k_n} \frac{1}{n+\alpha_i} \leq \frac{k_n}{n} \to 0$$

and

$$\sum_{i=k_n+1}^{n} \frac{1}{n+\alpha_i} \leq \frac{n-k_n}{\alpha(k_n)} \leq \frac{n}{n^{ab}} \to 0.$$

This completes the proof. □

6.3 Proof of Theorem 3

For the proof of Theorem 3, we need the following two lemmas.

Lemma 3. Let $V_n \sim \chi_n^2(\lambda_n)$ where $\chi_n^2(\lambda_n)$ is the chisquare distribution with degree freedom n and noncentrality parameter λ_n, and n is a positive integer and $\lambda_n \geq 0$. Then, for a nonnegative integer k,

$$\mathrm{E}(V^{-k}) \leq \frac{1}{(n-2)(n-4)\dots(n-2k)}.$$

Proof. Let $J_n \sim Poisson(\lambda_n/2)$ and $V_n|J_n \sim \chi^2_{n+2J_n}$ with convention that $J_n = 0$ with probability 1 if $\lambda_n = 0$. Then, marginally $V_n \sim \chi_n^2(\lambda_n)$. Thus,

$$\begin{aligned} \mathrm{E}(V^{-k}) &= \mathrm{E}\left(\frac{1}{(n+2J_n-2)(n+2J_n-4)\dots(n+2J_n-2k)}\right) \\ &\leq \frac{1}{(n-2)(n-4)\dots(n-2k)}. \end{aligned}$$

The last inequality follows from the fact that J_n is a nonnegative integer. □

Lemma 4. $||\hat{\eta}_n^{JS} - \eta_n^0||^2$ is uniformly integrable.

Proof. Set $t_{ni} = \eta_{ni}^0 + \frac{\sigma}{\sqrt{n}} Z_{ni}$ where Z_{ni} are independent standard normal random variables. Let $V_n = \frac{n||t_n||^2}{\sigma^2}$ and $U_n = \sum_{i=1}^p Z_{ni}^2$. Note $V_n \sim \chi_p^2(\lambda_n)$ with

$$\lambda_n = \sum_{i=1}^p \Big(\frac{\sqrt{n}}{\sigma}\eta_{ni}^0\Big)^2 = \frac{n}{\sigma^2} B_n$$

and $U_n \sim \chi_p^2$, where $B_n = \sum_{i=1}^p (\eta_{ni}^0)^2$. We have

$$\begin{aligned} ||\hat{\eta}_n^{JS} - \eta_n^0||^2 &= \sum_{i=1}^p \Big(-\frac{(p-2)}{V_n}\eta_{ni}^0 + \frac{\sigma}{\sqrt{n}} Z_{ni} - \frac{(p-2)}{V_n}\frac{\sigma}{\sqrt{n}} Z_{ni}\Big)^2 \\ &\le 2\frac{(p-2)^2}{V_n^2} B_n + 2\frac{\sigma^2}{n} U_n + 2\frac{(p-2)^2}{V_n^2}\frac{\sigma^2}{n} U_n. \end{aligned}$$

This implies

$$\mathrm{E}||\hat{\eta}_n^{JS} - \eta_n^0||^4 \le 8(n-2)^4 B^2 \mathrm{E} V_n^{-4} + 8\frac{\sigma^4}{n^2}\mathrm{E}U_n^2 + 8\frac{(n-2)^4\sigma^4}{n^2}\mathrm{E}(V_n^{-4}U_n^2). \tag{14}$$

Since $\mathrm{E}U_n^2 = O(n^2)$ and by Lemma 3, $\mathrm{E}(V_n^{-4}) = O(n^{-4})$, the first two terms of (14) are bounded. Note

$$\mathrm{E}(V_n^{-4}U_n^2) \le (\mathrm{E}(V_n^{-8}))^{1/2}(\mathrm{E}U_n^4)^{1/2}.$$

Again, by Lemma 3 and the moments of the chisquare distribution, $\mathrm{E}(V_n^{-4}U_n^2) = O(n^{-2})$. Thus, the last term of (14) is also bounded. This implies

$$\sup_n \mathrm{E}||\hat{\eta}_n^{JS} - \eta_n^0||^4 < \infty.$$

This completes the proof. $\square$

Proof of Theorem 3. From an equality from Hoffman (2000),

$$\mathrm{E}||\hat{\eta}_n^{JS} - \eta_n^0||^2 = \frac{\sigma^2}{n}\Big(p - \frac{(p-2)^2}{2}\int_0^1 (1-t)^{p/2-2} e^{-\frac{nA_n t}{2}} dt\Big). \tag{15}$$

We first consider case (b). Note

$$\int_0^1 (1-t)^{p/2-2} e^{-\frac{nA_n t}{2}} dt \ \ge \ \frac{2}{p-2} e^{-\frac{nA_n}{2}}.$$

Thus,

$$\mathrm{E}||\hat{\eta}_n^{JS} - \eta_n^0||^2 \;\; \leq \;\; \frac{\sigma^2}{n}\Big(p - (p-2)e^{-\frac{nA_n}{2}}\Big) \to 0,$$

as $n \to \infty$.

Now, we consider case (a). If $A > 0$, we can set $A_{n0} \in (0, c)$ and $A_{n1} > 0$ such that $A_n = A_{n0} + A_{n1}$, $\lim_{n\to\infty} A_{n0} = A_0 \in (0, c)$ and $\lim_{n\to\infty} A_{n1} = A_1 \geq 0$. If $A = 0$, define similarly with $A_0 = 0$. Since $e^{-x} \leq 1 - x + x^2/2$ for $x \geq 0$,

$$\begin{aligned}
& \int_0^1 (1-t)^{p/2-2} e^{-\frac{nA_n t}{2}} dt \\
\leq \;\; & \int_0^1 (1-t)^{p/2-2} e^{-\frac{nA_{n0} t}{2}} dt \\
\leq \;\; & \int_0^1 (1-t)^{p/2-2} \Big(1 - \frac{nA_{n0}t}{2} + \frac{n^2 A_{n0}^2 t^2}{8}\Big) dt \\
= \;\; & \frac{1}{p/2-1} - \frac{nA_{n0}}{2}\frac{1}{p/2(p/2-1)} + \frac{n^2 A_{n0}^2}{8}\frac{2}{(p/2+1)p/2(p/2-1)}.
\end{aligned}$$

Thus,

$$\begin{aligned}
& \mathrm{E}||\hat{\eta}_n^{JS} - \eta_n^0||^2 \\
\geq \;\; & \frac{\sigma^2 p}{n}\Big(1 - \frac{p-2}{p} + A_{n0}\frac{n}{p}\frac{p-2}{p} - A_{n0}^2 \frac{n^2}{p^2}\frac{p-2}{p+2}\Big) \\
\to \;\; & \sigma^2 c \frac{A_0}{c}\Big(1 - \frac{A_0}{c}\Big) > 0,
\end{aligned}$$

as $n \to \infty$. When $A = 0$,

$$\lim_{n\to\infty} \mathrm{E}||\hat{\eta}_n^{JS} - \eta_n||^2 \;\; = \;\; \lim_{n\to\infty} \frac{\sigma^2}{n}\Big(p - \frac{(p-2)^2}{2}\frac{1}{p/2-1}\Big) = \lim_{n\to\infty}\frac{2\sigma^2}{n} = 0.$$

The conclusion (c) follows from Lemma 4. This completes the proof. □

6.4 Proofs of Theorems 4 and 5

We need the following lemmas.

Lemma 5. Suppose $t = \xi + \frac{\sigma}{\sqrt{n}} Z$ with $Z \sim N(0,1)$.

(a) If $\xi = 0$, $B_{01}(t) \leq 2\sqrt{\frac{n}{\alpha}}$ for sufficiently large n and $\pi_1(t) \leq \frac{\pi_1}{\pi_0}\sqrt{\frac{\alpha}{n}} e^{\frac{1}{2}Z^2}$.

(b) if $\xi \neq 0$, for sufficiently large n,

$$B_{01}(t) \le 2\sqrt{\frac{n}{\alpha}} \exp\Big(-\frac{n}{8\sigma^2}\xi^2\Big) \text{ and } \pi_0(t) \le 2\frac{\pi_0}{\pi_1}\sqrt{\frac{n}{\alpha}} \exp\Big(-\frac{n}{8\sigma^2}\xi^2\Big).$$

Proof. Note

$$B_{01}(t) = \sqrt{1+\frac{n}{\alpha}} \exp\Big(-\frac{n}{2\sigma^2}\frac{n}{n+\alpha}\Big(\xi + \frac{\sigma}{\sqrt{n}}Z\Big)^2\Big).$$

First, consider the case $\xi = 0$. For sufficiently large n,

$$\begin{aligned} B_{01}(t) &= \sqrt{1+\frac{n}{\alpha}} \exp\Big(-\frac{1}{2}\frac{n}{n+\alpha}Z^2\Big) \\ &\le 2\sqrt{\frac{n}{\alpha}}. \end{aligned}$$

Also,

$$\begin{aligned} \pi_1(t) &= \frac{\pi_1}{\pi_0\sqrt{1+n/\alpha}\exp(-\frac{1}{2}\frac{n}{n+\alpha}Z^2) + \pi_1} \\ &\le \frac{\pi_1}{\pi_0}\sqrt{\frac{\alpha}{n}} e^{\frac{1}{2}Z^2}. \end{aligned}$$

Now, consider the case $\xi \neq 0$. For sufficiently large n,

$$B_{01}(t) \le 2\sqrt{\frac{n}{\alpha}} \exp\Big(-\frac{n}{2\sigma^2}\frac{1}{2}\frac{\xi^2}{2}\Big) = 2\sqrt{\frac{n}{\alpha}} \exp\Big(-\frac{n}{8\sigma^2}\xi^2\Big)$$

and

$$\begin{aligned} \pi_0(t) &= \frac{\pi_0 B_{01}(t)}{\pi_1 + \pi_0 B_{01}(t)} \\ &\le 2\frac{\pi_0}{\pi_1}\sqrt{\frac{n}{\alpha}} \exp\Big(-\frac{n}{8\sigma^2}\xi^2\Big). \end{aligned}$$

This completes the proof. □

Lemma 6. Suppose $Z \sim N(0,1)$, $b > 1$ and $c > 0$. Then,

(a) $\mathrm{E}Z^2 I(Z^2 > b) \le \sqrt{\frac{8b}{\pi}} e^{-b/2}$;

(b) $\mathrm{E}e^{Z^2/2} I(|Z| \le c) = \sqrt{\frac{2}{\pi}} c^{1/4}$;

(c) $\mathrm{E}Z^2 e^{Z^2/2} I(|Z| \le c) = \dfrac{\sqrt{2}}{3\sqrt{\pi}} c^{3/4}$.

Proof. The proofs can be obtained by the integration by part and the fact that $Z^2 \sim \chi^2_1$. □

Lemma 7. Suppose $Z, Z_{ni} \sim N(0,1)$ and independent for $n \ge 1$ and $1 \le i \le p$, where $1 \le p \le n$ is an integer depending on n. Let $t_{ni} = \frac{\sigma}{\sqrt{n}} Z_{ni}$ for each n and i. Then,

(a) $\displaystyle\sum_{i=1}^p \pi_1(t_{ni}) = O_p\Big(\frac{p(\log n)^{1/8}}{\sqrt{n}}\Big)$;

(b) $\displaystyle\sum_{i=1}^p Z_{ni}^2 \pi_1(t_{ni}) = O_p\Big(\frac{p(\log n)^{3/8}}{\sqrt{n}}\Big)$;

(c) $\displaystyle\sum_{i=1}^p Z_{ni}^2 \pi_1^2(t_{ni}) = O_p\Big(\frac{p(\log n)^{3/8}}{\sqrt{n}}\Big)$.

Proof. We begin with a simple bound of $\pi_1(t_{ni})$:

$$\pi_1(t_{ni}) \quad \le \quad \frac{e^{Z_{ni}^2/2}}{e^{Z_{ni}^2/2} + a\sqrt{n}},$$

where $a = \pi_0/(\pi_1\sqrt{\alpha})$.
(a) Let $c_n = \frac{p(\log n)^{1/8}}{\sqrt{n}}$. To show (a), it suffices to show

$$\sup_{n\ge 2} P\Big(\sum_{i=1}^p \pi_1(t_{ni}) > c_n M\Big) \to 0, \text{ as } M \to \infty.$$

Let

$$\begin{aligned}
& P\Big(\sum_{i=1}^p \pi_1(t_{ni}) > c_n M\Big) \\
\le\ & P\Big(\sum_{i=1}^p \frac{e^{Z_{ni}^2/2}}{e^{Z_{ni}^2/2} + a\sqrt{n}} > c_n M\Big) \\
\le\ & P\Big(\sum_{i=1}^p \frac{e^{Z_{ni}^2/2}}{e^{Z_{ni}^2/2} + a\sqrt{n}} I(|Z_{ni}| \le \sqrt{\log n} M\Big) > c_n M) + P(|Z_{ni}| > \sqrt{\log n} M \text{ for some } 1 \le i \le p) \\
=:\ & I_1 + I_2.
\end{aligned}$$

By a bound for tail probability of standard normal,

$$
\begin{aligned}
I_2 &\leq \sum_{i=1}^{p} P(|Z_{ni}| > M\sqrt{\log n}) \\
&\leq \frac{1}{\sqrt{2\pi}} \frac{p}{Mn^{M^2/2}\sqrt{\log n}}.
\end{aligned}
$$

For $M \geq 2$,

$$
\sup_{n\geq 2} I_2 \leq \sup_{n\geq 2} \frac{1}{\sqrt{2\pi}} \frac{p}{Mn^2\sqrt{\log n}} \leq \frac{1}{\sqrt{2\pi}} \frac{1}{M2\sqrt{\log 2}}.
$$

Thus, as $M \to \infty$, $\sup_{n\geq 2} I_2 \to 0$.

Now we consider $\sup_{n\geq 2} I_1$. By Lemma 6,

$$
\begin{aligned}
I_1 &\leq \frac{p}{c_n M} \mathrm{E}\left[\frac{e^{Z^2/2}}{e^{Z^2/2} + a\sqrt{n}} I(|Z| \leq M\sqrt{\log n})\right] \\
&\leq \frac{\sqrt{2}}{a\Gamma(1/2)} \frac{1}{M^{3/4}}.
\end{aligned}
$$

Thus, as $M \to \infty$, $\sup_{n\geq 2} I_1 \to 0$.

(b) Similarly, the proof can be obtained with $c_n = \frac{p(\log n)^{3/8}}{\sqrt{n}}$.

(c) The result follows from (b) and the fact $\pi_1(t_{ni}) \leq 1$. □

Proof of Theorem 4. First, write $t_{ni} = \eta_{ni}^0 + \sigma/\sqrt{n} Z_{ni}$ $i = 1, 2, \ldots, p$ for independent standard normal random variables Z_{ni}, $1 \leq i \leq p$ and $n = 1, 2, \ldots$.

(a) Set

$$
\sum_{i=1}^{p} (\hat{\eta}_{ni}^{BMA} - \eta_{ni}^0)^2 = \sum_{i=1}^{r} (\hat{\eta}_{ni}^{BMA} - \eta_{ni}^0)^2 + \sum_{i=r+1}^{p} (\hat{\eta}_{ni}^{BMA})^2 = I_1 + I_2.
$$

Now we investigate I_1 and I_2 one by one. We bound I_1 by $I_{11} + I_{12}$ as follows:

$$
\begin{aligned}
I_1 &\leq 2\sum_{i=1}^{r} \pi_1(t_{ni})^2 \big(\frac{n}{\alpha+n}\big)^2 \frac{\sigma^2}{n} Z_{ni}^2 + 2\sum_{i=1}^{r} \big(1 - \pi_1(t_{ni})\frac{n}{\alpha+n}\big)^2 (\eta_{ni}^0)^2 \\
&:= I_{11} + I_{12}.
\end{aligned}
$$

Note

$$
\mathrm{E} I_{11} \leq 2\frac{\sigma^2}{n} \sum_{i=1}^{r} \mathrm{E}(Z_{ni}^2) = 2r\frac{\sigma^2}{n} = O\big(\frac{1}{n}\big).
$$

For each $i = 1, 2, \ldots, r$,

$$\big(1 - \pi_1(t_{ni})\frac{n}{\alpha+n}\big)^2 \ \leq \ 2\pi_0(t_{ni})^2 + O\big(\frac{1}{n^2}\big)$$

and by Lemma 5

$$\begin{aligned}\pi_0(t_{ni})^2 &= 4\big(\frac{\pi_0}{\pi_1}\big)^2\frac{n}{\alpha}\exp\Big(-\frac{n}{4\sigma^2}(\eta_{ni}^2)^2\Big)\\ &\leq 4\big(\frac{\pi_0}{\pi_1}\big)^2\frac{n}{\alpha}\exp\Big(-\frac{n}{4\sigma^2}\min_{1\leq i\leq r}(\eta_{ni}^2)^2\Big).\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{E}I_{12} &\leq 16\big(\frac{\pi_0}{\pi_1}\big)^2\frac{n}{\alpha}\exp\Big(-\frac{n}{4\sigma^2}\min_{1\leq i\leq r}(\eta_{ni}^2)^2\Big)\sum_{i=1}^{r}(\eta_{ni}^0)^2 + 2\sum_{i=1}^{r}(\eta_{ni}^0)^2 O\big(\frac{1}{n^2}\big)\\ &= O\big(\frac{1}{n^2}\big).\end{aligned}$$

Now, we consider I_2. By Lemma 5,

$$\begin{aligned}I_2 &\leq \frac{\sigma^2}{n}\sum_{i=r+1}^{p} Z_{ni}^2\pi_1^2(t_{ni})\\ &= O_p\Big(\frac{p(\log n)^{3/8}}{n^{3/2}}\Big).\end{aligned}$$

Thus,

$$\sum_{i=1}^{p}(\hat{\eta}_{ni}^{BMA} - \eta_{ni}^0)^2 = O_p\big(\frac{1}{n}\big) + O_p\Big(\frac{p(\log n)^{3/8}}{n^{3/2}}\Big).$$

(b) Note $\pi_1(t_{ni}) > 1/2$ is equivalent to the posterior odds, $\pi_0(t_{ni})/\pi_1(t_{ni})$, smaller than 1 and the posterior odds is

$$\frac{\pi_0}{\pi_1}\sqrt{1+\frac{n}{\alpha}}\exp\Big(-\frac{n}{2\sigma^2}\frac{n}{n+\alpha}t_{ni}^2\Big).$$

As in the proof of (a), we set

$$\sum_{i=1}^{p}(\hat{\eta}_{ni}^{MP} - \eta_{ni}^0)^2 = \sum_{i=1}^{r}(\hat{\eta}_{ni}^{MP} - \eta_{ni}^0)^2 + \sum_{i=r+1}^{p}(\hat{\eta}_{ni}^{MP})^2 =: J_1 + J_2.$$

Consider J_2 first. For $i > r$, $t_{ni} = \sigma Z_{ni}/\sqrt{n}$ and the posterior odds is

$$\frac{\pi_0}{\pi_1}\sqrt{1+\frac{n}{\alpha}}\exp\Big(-\frac{n}{2(n+\alpha)}Z_{ni}^2\Big).$$

Solving $\pi_1(t_{ni}) > 1/2$, we obtain

$$Z_{ni}^2 > b_n,$$

where

$$b_n = \log\Big[\Big(\frac{n+\alpha}{\alpha}\Big)^{(n+\alpha)/n}\Big(\frac{\pi_0}{\pi_1}\Big)^{2(n+\alpha)/n}\Big].$$

Note $b_n = O(\log n)$ and

$$\begin{aligned} e^{-b_n/2} &\leq \alpha^{(n+\alpha)/(2n)}(\pi_1/\pi_0)^{(n+\alpha)/n}\frac{1}{(n+\alpha)^{1/2}} \\ &= O\big(\frac{1}{\sqrt{n}}\big). \end{aligned}$$

Using the above facts, we obtain

$$\begin{aligned} \mathrm{E}J_2 &\leq \frac{\sigma^2}{n}p\sqrt{\frac{8b_n}{\pi}}e^{-b_n/2} \\ &= O\Big(\frac{p\sqrt{\log n}}{n^{3/2}}\Big). \end{aligned}$$

Now consider J_1.

$$\begin{aligned} J_1 &\leq 2\sum_{i=1}^{r}\Big(\frac{n}{n+\alpha}I\big(\pi_1(t_{ni}) > \frac{1}{2}\big) - 1\Big)^2(\eta_{ni}^0)^2 + 2\sum_{i=1}^{r}\frac{n^2}{(n+\alpha)^2}\frac{\sigma^2}{n}Z_{ni}^2 I\big(\pi_1(t_{ni}) > \frac{1}{2}\big) \\ &:= J_{11} + J_{12}. \end{aligned}$$

Taking the expectation, we have

$$\mathrm{E}J_{12} \leq \frac{2r}{n}\sigma^2 = O\big(\frac{1}{n}\big).$$

By Lemma 5, for sufficiently large n,

$$\Big(\frac{n}{n+\alpha}I(\pi_1(t_{ni}) > 1/2) - 1\Big)^2 \leq 2\Big(\frac{\alpha}{n+\alpha}\Big)^2.$$

Thus,

$$\mathrm{E}J_{11} = 2rO\big(\frac{1}{n^2}\big)\sum_{i=1}^{r}(\eta_{ni}^0)^2 = O\big(\frac{1}{n^2}\big).$$

In summary, we have

$$J = O_p\big(\frac{1}{n}\big) + O_p\Big(\frac{p\sqrt{\log n}}{n^{3/2}}\Big).$$

(c) For a sequence of positive real numbers, (c_n) converging to 0, we have

$$\begin{aligned}\pi(||\eta-\eta^0||^2 > c_n|y_n) &\leq \frac{1}{c_n}\mathrm{E}(||\eta-\eta^0||^2|y_n)\\ &= \frac{1}{c_n}\left(tr\mathrm{Var}(\eta|y_n) + ||\hat{\eta}^{BMA}-\eta^0||^2\right).\end{aligned}$$

By (a), $||\hat{\eta}^{BMA}-\eta^0||^2 = O_p(1/n) + O_p\big(\frac{p(\log n)^{3/8}}{n^{3/2}}\big)$; thus, it suffices to show that

$$tr\mathrm{Var}(\eta|y_n) = O_p(1/n) + O_p\Big(\frac{p(\log n)^{3/8}}{n^{3/2}}\Big).$$

We decompose $tr(\mathrm{Var}(\eta_n|y_n))$ as follows:

$$\begin{aligned}tr(\mathrm{Var}(\eta_n|y_n)) &= \sum_{i=1}^{p}\Big(\pi_1(t_{ni})\frac{\sigma^2}{\alpha+n} + \frac{n^2}{(\alpha+n)^2}t_{ni}^2\pi_1(t_{ni})\pi_0(t_{ni})\Big)\\ &= \frac{\sigma^2}{\alpha+n}\sum_{i=1}^{r}\pi_1(t_{ni}) + \frac{\sigma^2}{\alpha+n}\sum_{i=r+1}^{p}\pi_1(t_{ni})\\ &\quad+\frac{n^2}{(\alpha+n)^2}\sum_{i=1}^{r}\big(\eta_{ni}^0 + \frac{\sigma}{\sqrt{n}}Z_{ni}\big)^2\pi_1(t_{ni})\pi_0(t_{ni}) + \frac{n^2}{(\alpha+n)^2}\frac{\sigma^2}{n}\sum_{i=r+1}^{p}Z_{ni}^2\pi_1(t_{ni})\pi_0(t_{ni})\\ &=: I_1 + I_2 + I_3 + I_4.\end{aligned}$$

Since $\pi_1(t_{ni}) \leq 1$, as $n \to \infty$,

$$\mathrm{E}I_1 \leq \sigma^2\frac{r}{n} = O\big(\frac{1}{n}\big).$$

By Lemma 7, as $n \to \infty$,

$$I_2 = O_p\Big(\frac{p(\log n)^{1/8}}{n^{3/2}}\Big).$$

Note

$$\begin{aligned}I_3 &\leq \sum_{i=1}^{r}\big(\eta_{ni}^0 + \frac{\sigma}{\sqrt{n}}Z_{ni}\big)^2\pi_0(t_{ni})\\ &\leq 2\frac{\pi_0}{\pi_1}\sqrt{\frac{n}{\alpha}}e^{-\frac{n}{8\sigma^2}\min_{1\leq i\leq r}(\eta_{ni}^0)^2}\sum_{i=1}^{r}\Big(2(\eta_{ni}^0)^2 + 2\frac{\sigma^2}{n}Z_{ni}\Big).\end{aligned}$$

For sufficiently large n and for some positive constant C,

$$\mathrm{E}I_3 \leq C\sqrt{n}e^{-\frac{n}{8\sigma^2}\min_{1\leq i\leq r}(\eta_{ni}^0)^2}.$$

Thus, $\mathrm{E}I_3$ is exponentially decreasing. Finally, by Lemma 7, we have

$$\begin{aligned} I_4 &\leq \frac{\sigma^2}{n}\sum_{i=r+1}^{p} Z_{ni}^2\pi_1(t_{ni}) \\ &= O_p\Big(\frac{p(\log n)^{3/8}}{n^{3/2}}\Big). \end{aligned}$$

This completes the proof. □

Proof of Theorem 5. By Lemma 5, as $n \to \infty$,

$$\begin{aligned} \prod_{i=1}^{r}\pi_1(t_{ni}) &\geq \prod_{i=1}^{r}\Big(1 - 2\frac{\pi_0}{\pi_1}\sqrt{\frac{n}{\alpha}}e^{-\frac{n}{8\sigma^2}(\eta_{ni}^0)^2}\Big) \\ &\longrightarrow 1 \text{ in } P_n^0 - probability. \end{aligned}$$

It suffices to show $\sum_{i=r+1}^{p}\log(1-\pi_1(t_{ni})) \to 0$ in P_n^0-probability. For $\epsilon > 0$,

$$\begin{aligned} & P\Big(\sum_{i=r+1}^{p}\log(1-\pi_1(t_{ni})) < -\epsilon\Big) \\ \leq\ & P\Big(\sum_{i=r+1}^{p}\log(1-\pi_1(t_{ni}))I(\pi_1(t_{ni}) < 1/2) < -\epsilon\Big) + P(\pi_1(t_{ni}) \geq 1/2 \text{ for some } r+1 \leq i \leq p) \\ =:\ & I_1 + I_2. \end{aligned}$$

For $Z \sim N(0,1)$, we have

$$\begin{aligned} I_2 &\leq pP\Big(\frac{e^{Z^2/2}}{e^{Z^2/2}+a\sqrt{n}} \geq \frac{1}{2}\Big) \\ &\leq pP\big(|Z| \geq \sqrt{\log n}\big) \\ &\leq \frac{1}{\sqrt{2\pi}}\frac{p}{\sqrt{n\log n}}. \end{aligned}$$

By the assumption, $I_2 \to 0$, as $n \to \infty$.

By the Taylor theorem, for $0 \leq x < 1$, $\log(1-x) = -x/(1-x')$ for some x' between 0 and x. Thus, $\log(1-x) \geq -2x$ for $0 \leq x \leq 1/2$. Thus,

$$\begin{aligned} I_1 &\leq P\Big(-2\sum_{i=r+1}^{p}\pi_1(t_{ni}) < -\epsilon\Big) \\ &= P\Big(\sum_{i=r+1}^{p}\pi_1(t_{ni}) > \frac{\epsilon}{2}\Big). \end{aligned}$$

By Lemma 5 and the assumption,

$$\sum_{i=r+1}^{p} \pi_1(t_{ni}) = O_p\Big(\frac{p(\log n)^{1/8}}{\sqrt{n}}\Big) = o_p(1).$$

Thus, $I_2 \to 0$, as $n \to \infty$. This completes the proof. $\square$

References

[1] Barbieri, M. M. and Berger, J. O. (2004). Optimal predictive model selection. *Ann. Statist.* **32** 870–897.

[2] Ghosal, S. (1997). Normal approximation to the posterior distribution for generalized linear models with many covariates. *Math. Methods Statist.* **6** 332–348.

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

[4] Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial (with doscussion). *Statist. Sci.* **14** 382–417.

[5] Hoffmann, K. (2000). Stein estimation—a review. *Statist. Papers* **41** 127–158.

[6] Hu, T. C., Móricz, F. and Taylor, R. L. (1989). Strong laws of large numbers for arrays of rowwise independent random variables. *Acta Math. Hungar.* **54** 153–162.

[7] James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, 361–379. Univ. California Press, Berkeley, Calif.

[8] Jiang, W. (2007). Bayesian variable selection for high dimensional generalized linear models: convergence rates of the fitted densities. *Ann. Statist.* **35** 1487–1511.

[9] Jolliffe, I. T. (2002). *Principal Component Analysis.* Springer-Verlag, New York, Second edition.

[10] Lenk, P. J. (1999). Bayesian inference for semiparametric regression using a Fourier representation. *J. R. Stat. Soc. Ser.* B **61** 863–879.

[11] Li, D. L., Rao, M. B. and Tomkins, R. J. (1995). A strong law for B-valued arrays. *Proc. Amer. Math. Soc.* **123** 3205–3212.

[12] Moreno, E., Girón, F. J. and Casella, G. (2010). Consistency of objective Bayes factors as the model dimension grows. *To appear in Ann. Statist.*

[13] Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques*, Vol. 6 of *Stud. Bayesian Econometrics Statist.*, 233–243 North-Holland, Amsterdam.