

A comparison study of Bayesian high-dimensional linear regression models

Ju-Won Shin^a, Kyoungjae Lee^{1,a}

^aDepartment of Statistics, Inha University

Abstract

We consider linear regression models in high-dimensional settings ($p \gg n$) and compare various classes of priors. The spike and slab prior is one of the most widely used priors for Bayesian regression models, but its model space is vast, resulting in a bad performance in finite samples. As an alternative, various continuous shrinkage priors, including the horseshoe prior and its variants, have been proposed. Although each of the above priors has been investigated separately, exhaustive comparative studies of their performance have been conducted very rarely. In this study, we compare the spike and slab prior, the horseshoe prior and its variants in various simulation settings. The performance of each method is demonstrated in terms of the regression coefficient estimation and variable selection. Finally, some remarks and suggestions are given based on comprehensive simulation studies.

Keywords: Bayesian regression, spike and slab prior, continuous shrinkage prior

1. 서론

다음과 같은 회귀분석 모형,

$$Y_i = X_i^T \beta + \epsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

$Y_i \in \mathbb{R}$, $X_i = (X_{i1}, \dots, X_{ip})^T \in \mathbb{R}^p$, $\beta = (\beta_1, \dots, \beta_p)^T \in \mathbb{R}^p$ 그리고 $\epsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ 에서, 자료의 갯수 n 보다 변수의 갯수 p 가 큰 고차원 상황을 고려하자. 이 경우 p 차원 회귀계수 β 에 추가적인 가정이 없다면, 모수의 갯수가 자료의 갯수보다 크기 때문에 일치성을 가지는 추론이 불가능하다. 모수의 차원을 줄이기 위해 흔하게 사용되는 가정 중 하나는, 회귀계수 중 대부분이 0이라는 희박성(sparsity)을 가정하는 것이다 (Martin 등, 2017) 희박성 가정 하에서 고차원 회귀분석 모형을 다루기 위하여, 연구자들은 크게 두 가지 전략을 사용해오고 있다. 첫 번째 전략은 변수선택과 추정을 동시에 진행하는 것으로, 먼저 0이 아닌 회귀계수들을 선택한 후에 그들의 값을 추정하는 방법이다. 빈도론의 lasso 추정량 (Tibshirani 1996)이 대표적인 예이다. 두 번째 전략은 p 차원 회귀계수 벡터를 그대로 추정하되, 0이라고 생각되는 회귀계수들을 0에 매우 가까운 값으로 축소시키는 방법이다. 빈도론의 Ridge 추정량 (Hoerl과 Kennard, 1970)이 대표적인 예이다.

The National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT)(No. 2019R1F1A1059483).

¹ Corresponding author: Department of Statistics, Inha University, 100 Inha-ro Michuhol-gu, Incheon 22212, Korea.
E-mail: leekjstat@gmail.com

고차원 상황에서 쓸 수 있는 대표적인 베이지안 추론 방법은 spike and slab 사전분포 (George와 McCulloch, 1993; Ishwaran과 Rao, 2005)가 있다. Spike and slab 사전분포는 사후분포 수렴속도 등의 이론적 성질들이 알려져 있고 변수 선택을 명확하게 한다는 장점이 있지만, 차원이 매우 커지는 상황에서는 2^p 개의 모델 공간을 탐험해야 하는 계산적인 문제가 있다. 이러한 문제를 해결하기 위해 다양한 연속 수축 사전분포(continuous shrinkage prior)가 제안되어 왔고, 그 중 가장 대표적인 사전분포는 horseshoe 사전분포 (Carvalho 등, 2010)이다. 또한, horseshoe 사전분포를 보완한 regularized horseshoe 사전분포 (Piironen과 Vehtari, 2017)와 horseshoe plus 사전분포 (Bhadra 등, 2017)가 있다. 이러한 연속 수축 사전분포들은 정확하게 0 값을 가지는 사후분포 표본을 주지는 않지만, 모델 공간의 크기가 spike and slab 사전분포에 비해 매우 작아서 Markov chain Monte Carlo (MCMC) 표본의 mixing이 빠를 것이라는 기대가 있다.

현재 고차원 베이지안 회귀분석을 위해 다양한 spike and slab 사전분포와 연속 수축 사전분포들이 개발되어 있지만, 이들의 실질적인 성능에 대한 비교 결과는 찾아보기 힘들다. 비록 (Van 등, 2019)에서 spike and slab 사전분포와 다양한 수축 사전분포들의 성능을 비교하였지만, 다양한 희박 정도(sparsity level)와 신호의 강도에 따른 회귀계수 추정과 변수선택 측면에서의 비교 연구는 진행되지 않았다. 이는 실제 분석에 사용할 사전분포를 선택해야 하는 사용자 입장에서는 매우 중요한 결과라 할 수 있다. 따라서, 본 논문에서는 몇 가지 대표적인 사전분포들을 선택하여, 다양한 모의실험 셋팅에서 이들의 성능을 비교하는 연구를 진행하였다.

본 논문은 다음과 같이 구성된다. 2장에서는 고차원 회귀분석을 위해 제안된 다양한 베이지안 방법론들을 소개한다. 다음으로 3장에서는, 2장에서 소개된 베이지안 방법론들의 성능을 다양한 셋팅에서 비교한다. 이 때 2장에서는 편의상 오차항의 분산 σ^2 을 상수로 생각하고 회귀계수 β 에 대한 사전분포에 집중하지만, 3장에서는 σ^2 에 대한 사전분포까지 고려한 추론을 통해 비교연구를 진행하려 한다. 마지막으로, 4장에서는 시뮬레이션 결과를 바탕으로 몇 가지 제안을 하며 논문을 마무리한다.

2. 고차원 회귀 모형

2.1. Spike and slab 사전분포

Spike and slab 사전분포의 기본적인 형태는 다음과 같다.

$$\begin{aligned} \beta_j | \gamma_j &\stackrel{\text{iid}}{\sim} (1 - \gamma_j)N(0, \nu_0) + \gamma_j N(0, \nu_1), \quad \text{for } 0 \leq \nu_0 \ll \nu_1, \\ \gamma_j &\stackrel{\text{iid}}{\sim} \text{Ber}(\theta), \quad \text{for } j = 1, \dots, p. \end{aligned} \quad (2.1)$$

이 때, ν_0 값은 매우 작은 값으로 설정하고, ν_1 값은 큰 값으로 설정한다. 본 논문에서는 $\nu_0 = 0.001$ 로 두고 ν_1 에는 절반 코시(half-Cauchy) 분포를 사전분포로 설정하였고, 이에 해당하는 사전분포의 확률밀도함수를 Figure 1에 나타내었다. Figure 1 처럼 분산을 ν_0 로 가지는 정규분포는 0에서의 밀도가 매우 높은 뾰족한 모양을 하고 있고, 분산을 ν_1 으로 가지는 정규분포는 비교적 평평한(slab) 모양을 가진다. Spike and slab 사전분포의 spike와 slab 부분은 각각 잡음과 신호를 가려내어 회귀계수를 추정하는 역할을 한다. 여기서 잡음은 0이거나 0과 다름 없어서 유의미하지 않은 회귀계수를 의미하고, 신호는 유의미한 회귀계수를 의미한다. 이렇게 명확한 변수 선택이 가능하기 때문에, spike and slab 사전분포는 베이지안 고차원 회귀분석에서 사용되는 대표적인 사전분포로 자리하고 있다.

Spike and slab 사전분포는 MCMC 방법 중 Gibbs sampling을 사용하지만, 만약 ν_0 값을 0으로 설정하게 되면 MCMC 시행마다 γ 의 차원이 변하는 문제가 발생한다. 이 때는 Gibbs sampling이 불가능 해지므로, γ 샘플을 얻기위해 β_j 를 적분한 $f(\gamma_j | \sigma^2, y)$ 를 구해 MCMC를 진행한다. 모형 (1.1)과 spike and slab 사전분포에

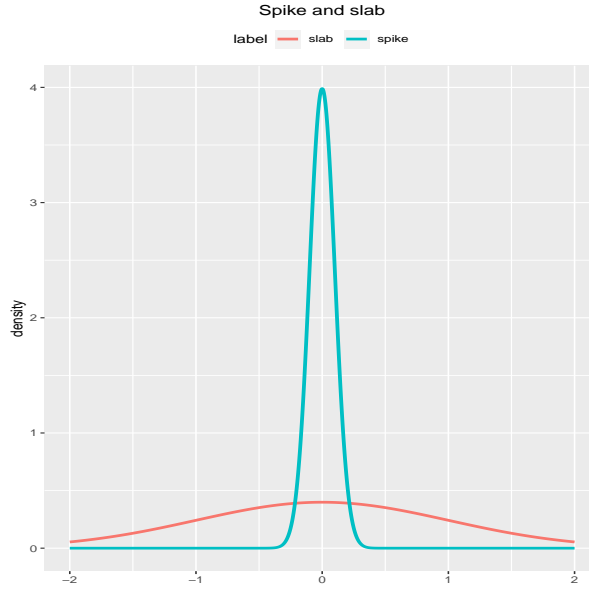


Figure 1: The probability density function of a spike and slab prior.

기반한 (γ, β) 의 완전조건부 사후분포는 다음과 같다. 모든 $j = 1, \dots, p$ 에 대해,

$$\begin{aligned} \beta \mid \text{rest} &\sim N_p(A_\gamma X^T Y \sigma^{-2}, A_\gamma), \\ \gamma_j \mid \text{rest} &\sim \text{Ber}\left(\frac{a}{a+b}\right), \end{aligned} \tag{2.2}$$

여기서 $a = f(\beta_j \mid \gamma_j = 1)\theta$, $b = f(\beta_j \mid \gamma_j = 0)(1 - \theta)$ 이다.

위의 식에서, $Y = (Y_1, \dots, Y_n)^T$, $X = (X_1, \dots, X_n)^T \in \mathbb{R}^{n \times p}$, $A_\gamma = (\sigma^{-2} X^T X + D_\gamma^{-1})^{-1}$, $D_\gamma = \text{diag}((1 - \gamma_1)v_0 + \gamma_1 v_1, \dots, (1 - \gamma_p)v_0 + \gamma_p v_1)$ 이다. 위 사후분포 식 (2.2)을 이용하면 Gibbs sampling을 이용한 사후분포 추론을 진행할 수 있다 (George와 McCulloch, 1993). 최종적으로 변수선택을 하는 방법은 m 번의 MCMC 시행 이후에 얻을 수 있는 γ 의 최빈값(mode)을 찾는 방법과 p 개의 $\gamma_j \in [0, 1]^p$ 들을 독립으로 보고 적절한 분계점(threshold)보다 높으면 유의미하다고 판단하는 방법 (Barbieri와 Berger, 2004)이 있다. 본 논문에서는 분계점을 이용한 변수선택을 사용하였고, 이를 표현하면 다음과 같다. 모든 $j = 1, \dots, p$ 에 대해,

$$\gamma_j = \begin{cases} 1, & p_j \geq p_t, \\ 0, & \text{o.w.,} \end{cases} \tag{2.3}$$

여기서 $p_j = m_j^*/m$ 이고, m_j^* 는 m 개의 사후분포 표본 중 $\gamma_j = 1$ 이 등장한 횟수이며, p_t 는 분계점을 의미한다.

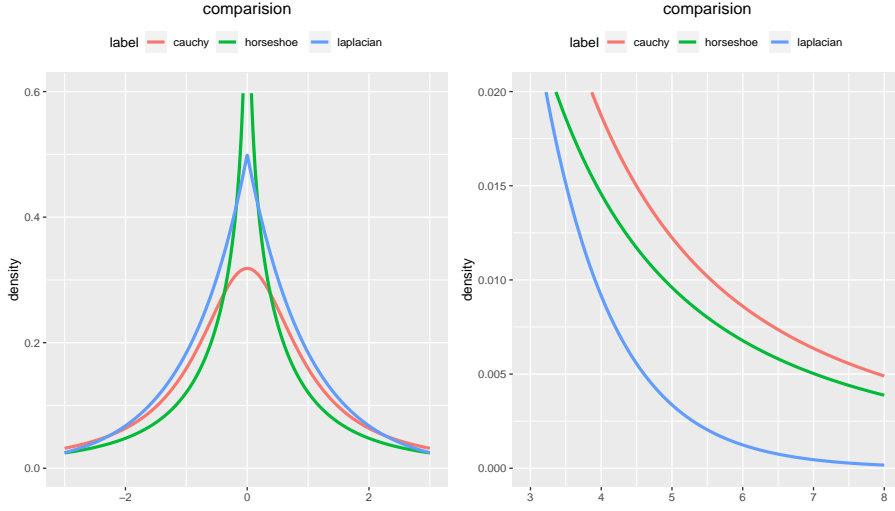


Figure 2: Probability density functions of Horseshoe, Cauchy and Laplacian distributions.

2.2. Horseshoe 사전분포

Horseshoe 사전분포는 β_j 에 다음과 같은 사전분포를 정의한다.

$$\begin{aligned}\beta_j \mid \lambda_j, \tau, \sigma^2 &\sim N(0, \lambda_j^2 \tau^2 \sigma^2), \\ \lambda_j &\sim C^+(0, 1), \\ \tau &\sim C^+(0, 1).\end{aligned}\tag{2.4}$$

위 식 (2.4)에서 $C^+(0, 1)$ 는 절반 코시분포를 의미한다. 일반적으로 τ 에 절반 코시분포를 사전분포로 설정하지만, cross validation 방법을 이용하여 τ 를 설정하거나, 임의로 설정하는 방법도 있다 (Piironen과 Vehtari, 2017). Horseshoe 사전분포에 기반하여 (β, τ, Λ) 의 사후분포 샘플을 얻기 위한 Gibbs sampling 알고리즘은 다음과 같다 (Makalic과 Schmidt, 2015).

$$\begin{aligned}\beta \mid \text{rest} &\sim N_p(AX^T Y, \sigma^2 A), \\ \tau^2 \mid \text{rest} &\sim \text{IG}\left(\frac{p+1}{2}, \frac{1}{\xi} + \frac{1}{2\sigma^2} \beta^T \Lambda^{-1} \beta\right), \\ \xi \mid \text{rest} &\sim \text{IG}\left(1, 1 + \frac{1}{\tau^2}\right), \\ \lambda_j^2 \mid \text{rest} &\sim \text{IG}\left(1, \frac{1}{\eta_j} + \frac{\beta_j^2}{2\sigma^2 \tau^2}\right), \quad j = 1, \dots, p, \\ \eta_j \mid \text{rest} &\sim \text{IG}\left(1, 1 + \frac{1}{\lambda_j^2}\right), \quad j = 1, \dots, p, \\ \sigma^2 \mid \text{rest} &\sim \text{IG}\left(\frac{n+p}{2}, \frac{1}{2} \|y - X\beta\|_2^2 + \frac{1}{2\tau^2} \beta^T \Lambda^{-1} \beta\right).\end{aligned}$$

위의 식에서 $A = (X^T X + 1/\tau^2 \Lambda^{-1})^{-1}$, $\Lambda = \text{diag}(\lambda_1^2, \dots, \lambda_p^2)$ 이며, $\text{IG}(a, b)$ 는 형상모수가 a , 척도모수가 b 인 역감마 분포를 나타낸다. 위에서 τ^2 과 λ_j^2 의 완전조건부 사후분포를 유도하기 위해, 잠재 변수(latent variable) ξ 와 η_j

를 도입하였다. 이 때 절반 코시분포의 성질을 이용하였는데, 예를 들어 $\lambda_j \sim C^+(0, 1)$ 와 $\lambda_j^2 | \eta_j \sim \text{IG}(1/2, 1/\eta_j)$, $\eta_j \sim \text{IG}(1/2, 1)$ 가 동치임을 이용하였다. 이에 따르면 λ_j^2 의 완전조건부 사후분포를 유도할 수 있다.

$$\pi(\lambda_j^2 | \text{rest}) \propto (\lambda_j^2)^{\frac{1}{2}} \exp\left(-\frac{\beta_j^2}{2\sigma^2\tau^2\lambda_j^2}\right) (\lambda_j^2)^{\frac{1}{2}-1} \exp\left(-\frac{1}{\eta_j\lambda_j^2}\right) \propto (\lambda_j^2)^{-1/2} \exp\left\{-\left(\frac{1}{\eta_j} + \frac{\beta_j^2}{2\sigma^2\tau^2}\right)\frac{1}{\lambda_j^2}\right\}.$$

따라서, λ_j^2 의 완전조건부 사후분포가 $\text{IG}(1, 1/\eta_j + \beta_j^2/(2\sigma^2\tau^2))$ 임을 확인할 수 있고, η_j , τ^2 그리고 ξ 의 완전조건부 사후분포를 유도하는 과정도 이와 유사한 계산을 통해 보일 수 있다.

Figure 2는 β 의 주변분포(marginal distribution)와 코시분포(Cauchy distribution), 라플라시안(Laplacian) 분포를 비교한 그림이다. β 의 주변분포가 비교적 0에서 매우 높은 밀도를 가지고, 두꺼운 꼬리를 가지고 있는 것을 확인할 수 있는데, 이는 spike and slab 사전분포에서 spike와 slab에 대응한다. Horseshoe 사전분포에서 τ 는 전체의 희박한 정도(global sparsity level)를 조절하고 p 개의 λ_j 가 지역별 희박한 정도(local sparsity level)를 조절한다. 그러나 spike and slab 사전분포처럼 정확한 변수 선택을 해주지는 않으므로, β 의 신용 구간(credible interval)을 이용하거나, 수축 인자(shrinkage factor)를 이용하여 변수 선택을 해야 한다. 본 연구에서는 수축 인자를 이용하여 변수 선택을 하였다. 다음과 같은 모형 $f(y | \beta) = N(y | \beta, \sigma^2 I)$ 에서, 수축 인자(k_j)는 아래와 같이 구해진다 (Piironen과 Vehtari, 2017).

$$E(\beta_j | Y, \tau^2 \lambda_j^2) = \left(\frac{\tau^2 \lambda_j^2}{\tau^2 \lambda_j^2 + 1}\right) y_j + \left(\frac{1}{\tau^2 \lambda_j^2 + 1}\right) 0 = (1 - k_j) y_j, \quad \text{where } k_j = \frac{1}{1 + \sigma^{-2} \tau^2 \lambda_j^2}. \quad (2.5)$$

위 식 (2.5)에 따르면, k_j 값이 1에 가까우면 β_j 의 사후분포 평균이 0에 가깝게 수축되는 효과가 있고, 반대로 k_j 값이 0에 가까우면 β_j 의 사후분포 평균이 y_j 에 가깝게 정해진다. 따라서, 수축 인자인 k_j 가 적절한 분계점보다 크면 유의한 변수로 선택하는 방식으로 변수 선택을 진행할 수 있다.

2.3. Horseshoe plus 사전분포

Horseshoe plus 사전분포는 다음과 같다. 모든 $j = 1, \dots, p$ 에 대해,

$$\begin{aligned} \beta_j | \lambda_j, \tau, \sigma^2 &\sim N(0, \lambda_j^2 \tau^2 \sigma^2), \\ \lambda_j | \phi_j &\sim C^+(0, \phi_j), \\ \phi_j &\sim C^+(0, 1), \\ \tau &\sim C^+(0, 1). \end{aligned}$$

Horseshoe plus 사전분포는 꼬리가 일반적인 horseshoe 사전분포의 꼬리보다 두꺼운 특징을 가지고 있다. 이러한 특징 때문에 실제 회귀계수 벡터가 매우 희박한 상황, 즉 극소수를 제외하고는 회귀계수 성분이 모두 0인 상황에서는, 일반적인 horseshoe 사전분포보다 더 좋은 성능을 얻을 수 있다고 기대되는 측면이 있다. 이 경우에도 수축 인자(k_j)에 기반한 변수 선택을 진행할 수 있고, 수축 인자는 (2.5)와 같은 모양으로 구해진다.

Horseshoe plus 사전분포에 기반하여 모수들의 완전조건부 사후분포를 계산하기 위해서는, 다음의 성질을 이용한다. (a) $\lambda_j | \phi_j \sim C^+(0, \phi_j)$, $\phi_j \sim C^+(0, 1)$ 는 (b) $\lambda_j^2 | \eta_j \sim \text{IG}(1/2, 1/\eta_j)$, $\eta_j | \phi_j^2 \sim \text{IG}(1/2, 1/\phi_j^2)$, $\phi_j^2 | \xi_j \sim \text{IG}(1/2, 1/\xi_j)$, $\xi_j \sim \text{IG}(1/2, 1)$ 과 동치이다. 이를 이용하면 모수들의 완전조건부 사후분포를 계산할 수 있고, horseshoe 사전분포와 비교했을 때 완전조건부 사후분포가 달라지는 부분만 적으면 아래와 같다

(Makalic과 Schmidt, 2016).

$$\begin{aligned}\eta_j | \text{rest} &\sim \text{IG}\left(1, \frac{1}{\phi_j^2} + \frac{1}{\lambda_j^2}\right), \\ \phi_j^2 | \text{rest} &\sim \text{IG}\left(1, \frac{1}{\eta_j} + \frac{1}{\xi_j}\right), \\ \xi_j | \text{rest} &\sim \text{IG}\left(1, 1 + \frac{1}{\phi_j^2}\right).\end{aligned}$$

Horseshoe plus 사전분포에 기반한 사후분포 추론은 R package rstan을 통해 no-u-turn sampler (NUTS) 알고리즘을 사용하였다.

2.4. Regularized horseshoe 사전분포

Regularized horseshoe 사전분포는 horseshoe 사전분포에서 β_j 의 조건부 사전분포를 다음과 같이 수정한 사전분포이다. 모든 $j = 1, \dots, p$ 에 대해,

$$\beta_j | \lambda_j, \tau, c, \sigma^2 \sim N\left(0, \tilde{\lambda}_j^2 \tau^2 \sigma^2\right), \quad \text{where } \tilde{\lambda}_j^2 = \frac{c^2 \lambda_j^2}{c^2 + \tau^2 \lambda_j^2}. \quad (2.6)$$

이 때, $c > 0$ 는 주어진 상수이다. 위의 식 (2.6)에서 $\tau^2 \lambda_j^2 \ll c^2$ 이면 $\tilde{\lambda}_j^2 \rightarrow \lambda_j^2$ 이 되므로, β_j 에 대한 사전분포가 일반적인 horseshoe 사전분포에 가까워진다. 반면에 $\tau^2 \lambda_j^2 \gg c^2$ 이면 $\tilde{\lambda}_j^2 \rightarrow c^2/\tau^2$ 이 되므로, β_j 에 대한 사전분포가 $N(0, c^2)$ 에 가까워진다. 따라서 사전분포 (2.6)은 잡음에 해당하는 β_j 에는 작은 $\tau^2 \lambda_j^2$ 값을 부여하여 horseshoe 사전분포와 유사하게 0으로 축소시키고, 신호에 해당하는 β_j 에도 큰 $\tau^2 \lambda_j^2$ 값을 통해 c^2 의 분산을 가지는 정규분포처럼 모형화하여 어느 정도 0으로 축소시키는 특징이 있다. 즉, 일반적인 horseshoe 사전분포와 비슷하게 작동하지만, 신호들을 더 잘 정칙화(regularization)하는 특징이 있다. Regularized horseshoe 사전분포에서는 λ_j, c^2 등의 완전조건부 사후분포를 계산하기 힘들기 때문에, 사후분포 추론은 R package rstan을 통해 진행하였다.

Regularized horseshoe 사전분포에서는 일반적인 horseshoe 사전분포와는 다른 수축 인자를 정의한다 (Pironen과 Vehtari, 2017). 공변량 벡터 $X_i = (X_{i1}, \dots, X_{ip})^T$ 의 성분들이 서로 독립이고 $\text{Var}(X_{ij}) = s_j^2$, $f(Y | X, \beta, \sigma) = N(X\beta, \sigma^2 I)$, $\pi(\beta_j | \tau, \lambda_j) = N(0, \tau^2 \lambda_j^2)$ 일 경우, 수축 인자는 다음과 같이 구할 수 있다.

$$E(\beta_j | \tau, \sigma^2, \Lambda, Y) = (1 - k_j)\hat{\beta}_j, \quad \text{where } k_j = \frac{1}{1 + a_j^2 \lambda_j^2} \quad \text{and } a_j^2 = n\sigma^{-2} s_j^2.$$

이 때, $\hat{\beta}_j$ 는 β_j 의 최대가능도추정량이 존재할 때, 그것의 j 번째 성분이다. 위 식을 식 (2.6)에 적용하면, 다음과 같은 수축 인자를 구할 수 있다.

$$\tilde{k}_j = \frac{1}{1 + a_j^2 \tilde{\lambda}_j^2} \approx \frac{(1 - b_j)\lambda_j^2}{1 + b_j a_j^2 \lambda_j^2} = (1 - b_j)k_j + b_j.$$

따라서, regularized horseshoe의 수축 인자 \tilde{k}_j 는 $(b_j, 1)$ 사이의 값을 가지고, 이 때 $b_j = 1/(1 + n\sigma^{-2} s_j^2 c^2)$ 이다.

Figure 3은 horseshoe, horseshoe plus 그리고 regularized horseshoe 사전분포에서 수축 인자의 확률밀도 함수를 나타낸 그림이다. 그림을 보면, $\tau = 0.01$ 일 경우 0 근처에서 horseshoe 사전분포의 수축 인자가 다른 사전분포의 수축 인자보다 매우 낮은 밀도함수 값을 가지는 것을 확인할 수 있다. 이는 전체의 희박 정도를 조절하는 τ 값이 작게 추정되는 상황에서 horseshoe 사전분포가 붕괴하는 문제를 일으킬 수 있다 (Lee 등, 2020).

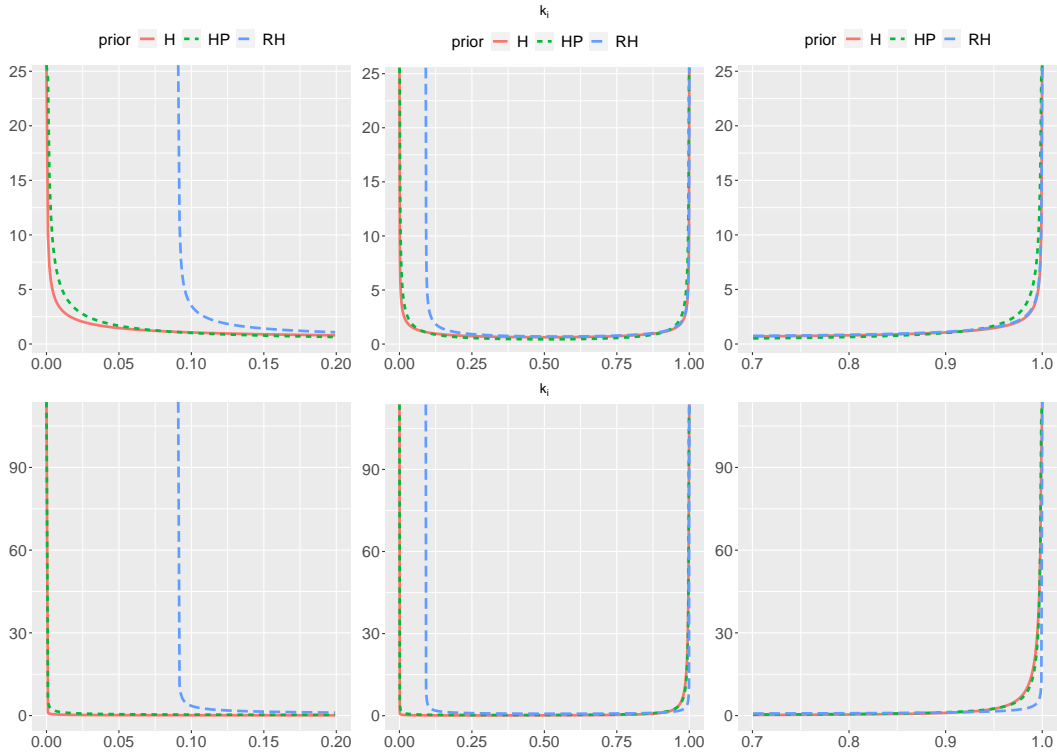


Figure 3: Probability density functions of various priors with $\tau = 1$ (top row) and $\tau = 0.01$ (bottom row). In the above, H , HP and RH represent horseshoe, horseshoe plus and regularized horseshoe priors, respectively.

그 이유는 horseshoe 사전분포의 수축 인자가 0 근처의 값을 가지지 못해 신호와 잡음을 구분하지 못하고, 모든 회귀계수의 수축 인자를 1로 추정하여 잡음으로 판단하기 때문이다. 반면에 다른 사전분포들은 $\tau = 0.01$ 일 때, 0 근처에서 horseshoe 사전분포보다 비교적 높은 밀도함수 값을 유지하고 있다. 고차원 상황에서는 τ 값이 매우 작게 추정되기 때문에 horseshoe 사전분포가 언제든지 붕괴될 위험이 있다. 이러한 문제를 해결할 수 있는 간단한 방법은 τ 에 대한 사전분포를 절반 코시분포가 아닌 절단된 코시분포

$$\tau \sim TC^+(0, 1)_{\left(\frac{1}{p}, \infty\right)}, \tag{2.7}$$

로 설정하여 τ 값이 매우 작아지는 것을 방지하는 것이나, 이 방법도 완벽하게 문제를 해결하지는 못한다.

3. 시뮬레이션 자료를 통한 비교연구

초모수의 선택 문제에서 벗어나기 위해 모든 모형에서 σ^2 에 Jeffrey's 사전분포인,

$$\pi(\sigma^2) \propto \frac{1}{\sigma^2},$$

를 설정하였다. 또한, 기존의 horseshoe 사전분포에서 τ 에 절단된 코시분포 (2.7)을 설정한 모형을 추가하였고, spike and slab 사전분포 (2.1)에서는 θ 값을 0.5로 고정된 모형과 균일분포를 사전분포로 설정한 모형을 적합

Table 1: Confusion matrix

	True signal	True noise
Predicted signal	TP	FP
Predicted noise	FN	TN

하였다. 표기의 편의성을 위해, horseshoe, truncated horseshoe, regularized horseshoe, horseshoe plus, spike and slab($\theta \sim \text{Unif}(0, 1)$), spike and slab($\theta = 0.5$)을 각각 H , TH, RH, HP, SS, SS_{ber}로 표기하였다.

참 회귀계수를 β_0 로 표기하고, 다음과 같은 상황을 고려한다. $\beta_0 = (\beta_{\text{signal}}^T, \beta_{\text{noise}}^T)^T \in \mathbb{R}^p$, $\beta_{\text{signal}} = (\beta_{\text{signal}, j}) \in \mathbb{R}^D$ 그리고 $\beta_{\text{noise}} = (\beta_{\text{noise}, j}) \in \mathbb{R}^q$. 이 때, D 는 신호의 개수, q 는 잡음의 개수이다. 관측치의 개수는 n 으로 표기하였고, 공변량 X 는 다변량 정규분포 $N_p(0, I)$ 에서 추출하여 생성하였다.

본 비교연구에서는 크게 두 가지 셋팅을 고려하였다. 첫 번째 셋팅인 scenario 1에서는, signal-to-noise ratio (SNR)과 0이 아닌 β 의 성분값이 고정된 상황에서 희박정도에 따른 사전분포들의 성능 비교를 목적으로 두었다. 이 때, SNR과 희박정도(sparsity level; sp)는 다음과 같이 정의된다.

$$\text{SNR} = \frac{\text{Var}(X_i^T \beta_0)}{\sigma^2}, \quad \text{sp} = \frac{D}{p}.$$

두 번째 셋팅인 scenario 2에서는, 희박정도와 오차항의 분산이 고정된 상황에서 0이 아닌 β 의 성분값 변화에 따른 사전분포들의 성능 비교를 목적으로 두었다. Scenario 1과 2에 대한 구체적인 설명은 아래와 같다.

- Scenario 1: $(n, p) = (50, 100), (200, 500), (300, 1000)$ 으로 변화시키고, 각 차원마다 sp 값을 0, 0.05, 0.1, 0.15, 0.2로 변화시킨다. 이 때, $\beta_{\text{noise}} = 0$, $\beta_{\text{signal}, j} = 2$ for all $j = 1, \dots, D$, SNR = 10으로 고정한다.
- Scenario 2: $(n, p) = (50, 100), (200, 500), (300, 1000)$ 으로 변화시키고, sp = 0.05, $\sigma^2 = 3^2$, $\beta_{\text{noise}} = 0$ 으로 고정한다. 이 때, β_{signal} 의 성분을 생성하는 분포를 다음과 같이 변화시킨다. $\beta_{\text{signal}, j} \sim \text{Unif}(Z_j \mu, 2Z_j \mu)$ for all $j = 1, \dots, D$, where $p(Z_j = 1) = P(Z_j = -1) = 0.5$ and $\mu = 0.1, 0.5, 1, 2$.

결과를 평가하기 위해 mean squared error (MSE)를 사용하며, 다음과 같이 신호와 잡음에 대한 MSE를 나누어 표현한다.

$$\text{MSE}_S = \frac{1}{D} \sum_{j=1}^D (\hat{\beta}_j - \beta_{\text{signal}, j})^2, \quad \text{MSE}_N = \frac{1}{q} \sum_{j=D+1}^p (\hat{\beta}_j)^2. \quad (3.1)$$

Scenario2에서는 신호의 크기가 달라지는 것을 보정해주기 위해, 신호에 대해서는 다음과 같이 변형된 MSE를 사용한다.

$$\text{MSE}_S = \frac{1}{D\mu^2} \sum_{j=1}^D (\hat{\beta}_j - \beta_{\text{signal}, j})^2. \quad (3.2)$$

또한 각 사전분포의 변수 선택 결과를 비교하기 위해 Table1의 값을 이용하여 민감도(sensitivity)와 특이도(specificity)를 다음과 같이 정의하였다.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}.$$

이 때, TP, TN, FP 그리고 FN의 정의는 Table 1에 정리되어 있다. 이를 이용하여 receiver operating characteristic (ROC) 곡선을 그린 후 area under the curve (AUC)값을 구해 정리하였는데, 이 때 ROC 곡선을 그리기 위해서 spike and slab 사전분포에서는 식 (2.3)을, horseshoe 사전분포에서는 수축 인자를 사용하였다.

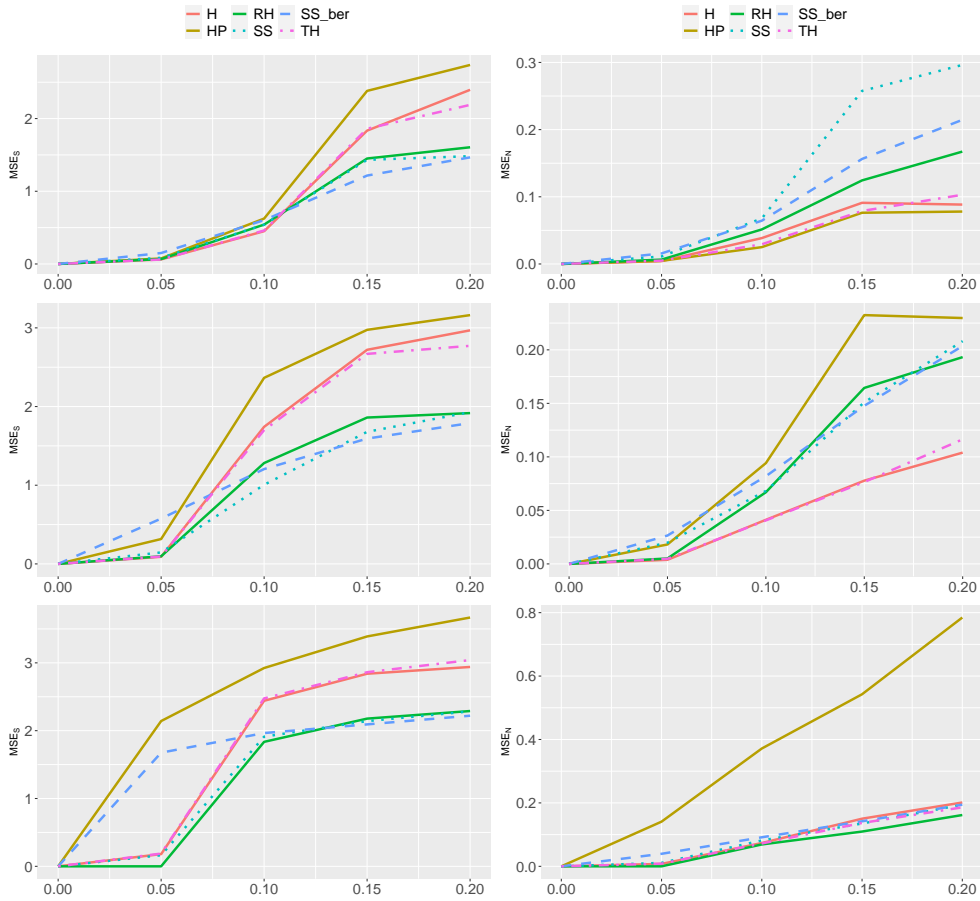


Figure 4: MSE_S (left column) and MSE_N (right column) for Scenario 1 with $p = 100$ (top row), $p = 500$ (middle row) and $p = 1000$ (bottom row). The x-axis represents the level of sparsity (sp).

Figure 4는 scenario 1의 결과이고, x축과 y축의 값은 각각 sp 값과 식 (3.2)의 MSE를 나타낸다. 신호를 추론하는 상황에서 가장 좋은 결과를 보인 사전분포는 RH, SS, SS_{ber} 이고, 이 사전분포들은 MSE_S 와 MSE_N 두 개의 측면에서 거의 비슷한 결과를 보인다. 이는 RH가 일반적인 horseshoe 사전분포보다 spike and slab 사전분포의 특징을 더 잘 따라가기 때문이다. 또한 sp 값이 작을 때 SS_{ber} 가 SS보다 MSE_S 값이 더 높은 경향이 있는 것을 확인할 수 있다. 이는 사전분포 SS_{ber} 에서는 0이 아닌 회귀 계수들의 비율인 θ 값이 0.5로 고정되어, 실제로는 0인 회귀 계수들이 매우 많을 때에도 불필요하게 많은 회귀 계수들을 유의하게 추론하기 때문으로 해석할 수 있다. TH와 H는 두 개의 MSE 측면에서 매우 비슷한 결과를 보이고 MSE_N 측면에서 차원에 상관없이 매우 좋은 결과를 보인다. 이를 통해, TH와 H는 horseshoe 사전분포의 특징인 잡음을 0으로 수축시키는 성질을 잘 수행하고 있는 것을 알 수 있다. 본 논문의 비교실험에서 HP는 대부분의 상황, 특히 sp 값이 매우 작은 희박한 상황에서도 MSE 값이 매우 크게 나오는 경향이 있었다. 이것은 다소 의외의 결과였는데, 왜냐하면 HP는 실제 회귀계수가 매우 희박한 상황에서 더 좋은 성능을 보일 거라고 기대되었기 때문이다. HP의 좋지 않은 성능은 (Lee 등, 2020) Supplementary Material 4장의 시뮬레이션 결과에서도 공통적으로 관찰되는데, 이러한 현상의 이유에 대해서는 포괄적인 실험을 통해 더욱 깊은 연구가 필요할 것으로 보인다.

Table 2: AUC values for Scenario 1 with $p = 500$. sp represents the level of sparsity

sp	SS	SS _{ber}	H	TH	HP	RH
0.05	0.99	0.99	1	1	0.99	1
0.10	0.92	0.92	0.5	0.77	0.5	0.88
0.15	0.86	0.86	0.5	0.65	0.5	0.69
0.20	0.75	0.77	0.53	0.53	0.5	0.58

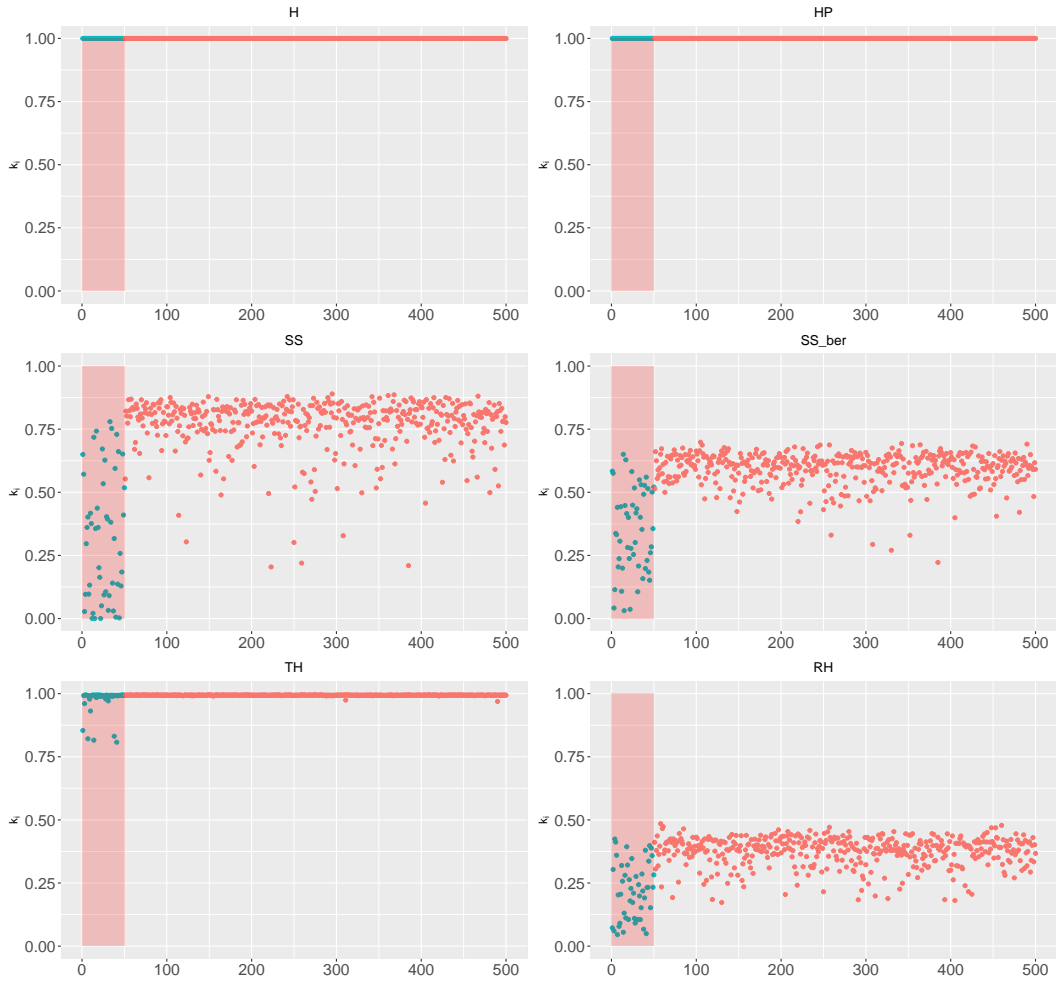
Figure 5: The estimated shrinkage factors for true signals (green dots) and true signals (red dots) when $p = 500$ and $sp = 0.1$. The x -axis represents the indices for p variables, while red region means the indices for true signals.

Table 2는 각 사전분포에 따라 변수선택의 결과를 보기 위해 ROC 곡선의 넓이인 AUC 값을 정리한 표이다. 전반적으로 sp 값이 증가함에 따라 AUC 값이 줄어드는 경향이 있는데, 이는 신호가 많아질 수록 변수선택이 어려워지는 것을 의미한다. 또한 sp 값이 0.1 이상이면 SS_{ber} 와 SS 이 다른 사전분포 H , TH , HP , RH 보다 좋은 변수 선택 성능을 보이는 경향이 있었다. 특히 sp 값이 0.1 이상일 때 H 와 HP 의 AUC 값이 대략 0.5가

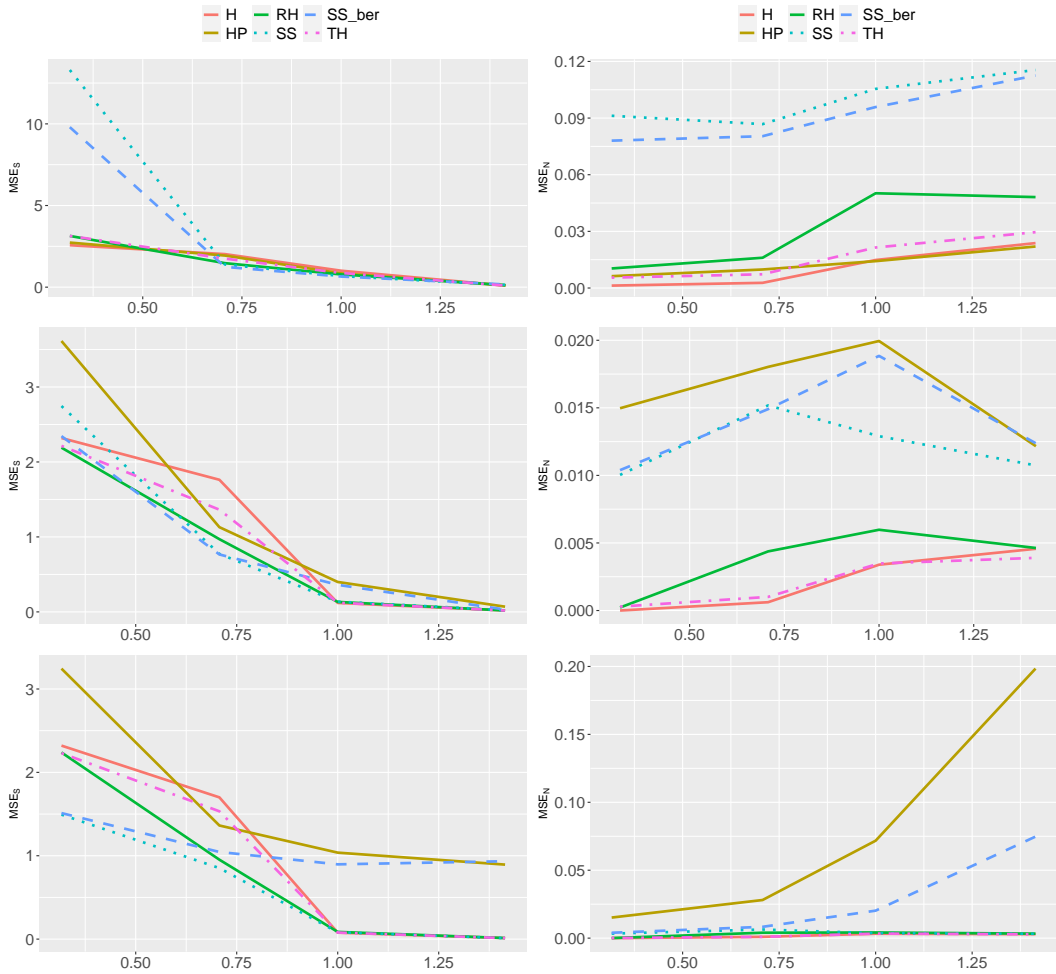


Figure 6: MSE_S (left column) and MSE_N (right column) for Scenario 2 with $p = 100$ (top row), $p = 500$ (middle row) and $p = 1000$ (bottom row). The x-axis represents the value of $\sqrt{\mu}$.

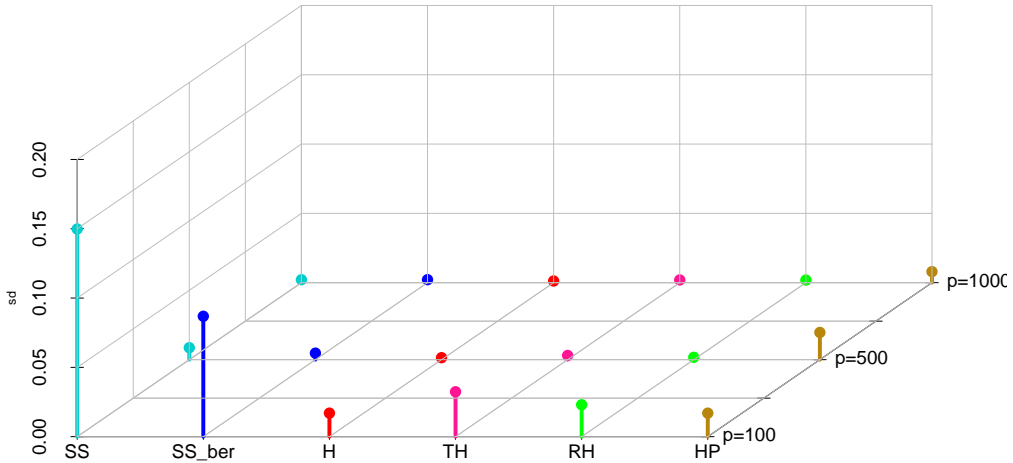
나왔는데, 이는 2.4장에서 언급한 horseshoe 사전분포의 붕괴 현상 때문으로 보인다. 반면에 RH와 TH에서는 붕괴하는 현상이 일어나지 않았다.

Figure 5에서는, Scenario1에서 $sp = 0.1$ 이고 $p = 500$ 일 때 2.4장에서 언급한 horseshoe 사전분포가 붕괴하는 현상을 묘사하였다. 빨간 영역은 유의한 변수들의 영역으로 y 축의 k_i 값이 0에 가까울 수록 이상적이고, 빨간 영역이 아닌 부분은 k_i 값이 1에 가까울 수록 이상적이라 할 수 있다. 이 때 spike and slab 사전분포의 k_i 는 1에서 식 (2.3)의 p_j 를 뺀 값을 의미한다. 하지만 추론 결과를 보면, H, HP에서 모든 k_i 값이 정확하게 1로 추정된 것을 알 수 있고, TH도 이들과 비슷한 형태를 보인다. 이는 전체적인 희박 정도를 조절하는 τ 값이 매우 작게 추정되어 붕괴현상이 일어나는 것으로 해석된다 (Lee 등, 2020). 반면에, 다른 사전분포들은 붕괴하는 현상은 일어나지 않았다.

Figure 6은 scenario 2의 결과이고, x 축과 y 축의 값은 각각 $\sqrt{\mu}$ 값과 (3.2)의 MSE를 나타낸다. Table 3는 각 사전분포에 따라 변수 선택의 결과를 보기 위해 ROC curve의 넓이인 AUC 값을 정리한 표이다. 모든 사전분포

Table 3: AUC values for Scenario 2 with $p = 500$

μ	SS	SS _{ber}	H	TH	HP	RH
0.1	0.53	0.51	0.50	0.52	0.61	0.55
0.5	0.90	0.88	0.86	0.79	0.90	0.81
1.0	0.99	0.99	1.00	0.91	0.50	1.00
2.0	1.00	1.00	1.00	1.00	1.00	1.00

Figure 7: Standard deviations of MSE_S for each prior when $\mu = 0.1$.

가 μ 값이 증가할 때 MSE_S 값이 감소하는 것을 알 수 있고, $\mu = 0.1$ 일 때는 대부분의 사전분포가 큰 MSE_S 값을 가진다. 특히 $p = 100$ 이고 μ 값이 작을 때, spike and slab 사전분포들이 매우 큰 MSE_S 값을 가지는 경향을 확인할 수 있었다. 이러한 현상의 이유를 살펴보기 위해, Figure 7에 $\mu = 0.1$ 일 때 각 사전분포에 기반한 MSE_S 의 표준편차를 나타내었다. Scenario 2에서는 차원 p 가 증가할수록 SNR 값이 증가하도록 설정되었다는 것을 유의하자. 그림에서 $p = 100$ 일 때 SS와 SS_{ber}의 표준편차가 매우 높은 것을 알 수 있는데, 이는 SNR 값이 매우 작아지는 상황에서 spike and slab 사전분포의 변동이 심하다는 것을 의미한다. 결과적으로 SNR 값이 매우 작은 상황에서는 horseshoe 사전분포가 안정적이고, spike and slab 사전분포는 불안정하다는 것을 의미한다.

Table 3를 보면 전반적으로 μ 값이 증가함에 따라 AUC 값이 증가하므로, 신호가 커질수록 변수 선택이 쉬워지는 것을 알 수 있다. 특히 $\mu = 0.5$ 일 때는 대부분의 사전분포의 AUC 값이 0.8 이상이 된다. 이 때 SS와 SS_{ber} 모두 0.9 근처의 AUC 값을 가지는데, 이는 잡음과 신호를 어느 정도 구분할 수 있는 상황에서는 spike and slab 사전분포가 좋은 변수 선택 성능을 가지는 것으로 해석될 수 있다. 비록 μ 값이 1 이상이 될 때 horseshoe 사전분포가 미세하게 더 좋은 결과를 보이지만, 크게 의미가 있는 차이는 아닌 것으로 해석된다. 다만 spike and slab 사전분포는 크기가 큰 회귀 계수도 0 쪽으로 축소시키기 때문에, 이것이 영향을 주었을 가능성도 배제할 수는 없다. 또한, $p = 500, 1000$ 이고 μ 값이 1 이상일 때, SS_{ber}와 HP의 MSE_S 는 다른 사전분포보다 완만하게 감소하는 것으로 보아, 신호가 일정 이상 커지는 상황에서 두 사전분포의 성능 개선이 상대적으로 느리게 나타나는 것으로 생각된다. Table 4는 본 논문의 모의실험 결과를 정리한 표이다.

Table 4: Summary of numerical studies in Section 3

Scenario1	Summary
SS	Reasonable in terms of MSE_S and variable selection
SS_{ber}	Similar to SS but having large MSE_S when $sp = 0.05$
H	Small AUC & large MSE_S when $sp \geq 0.1$
TH	Small AUC & large MSE_S but slightly better than H or HP
HP	Small AUC & large MSE_S
RH	Having similar performance to that of SS but slightly worse AUC
Scenario2	Summary
SS	Reasonable in terms of MSE_S and variable selection except signals are small ($\mu = 0.1$)
SS_{ber}	Similar to SS but having large MSE_S when signals are large ($\mu \geq 1$) and $p = 1000$
H	Slightly better than SS when signals are large ($\mu \geq 1$)
TH	Slightly better than SS when signals are large ($\mu \geq 1$)
HP	Slightly worse than other horseshoe type priors in terms of MSE_S
RH	Reasonable in terms of MSE_S and variable selection

4. 결론

본 연구에서는 희박정도와 신호의 크기에 따른 spike and slab 사전분포와 horseshoe 사전분포의 추론 성능을 비교하였다. 비교연구 결과에 따르면, 회귀계수 추정과 변수선택 측면에서 spike and slab 사전분포가 준수한 결과를 보였기 때문에, 신호의 크기나 SNR에 대한 사전적인 정보가 없다면 θ 에 $Unif(0, 1)$ 사전분포를 설정한 spike and slab 사전분포를 사용하는 것이 좋은 성능을 보여주었다. 반면에 신호의 크기가 일정한 기준점 이상일 때는 잡음과 신호 측면에서 horseshoe 사전분포가 더 좋은 결과를 보였고, SNR값이 매우 작아지는 상황에서 spike and slab 사전분포는 불안정한 결과를 보였다. 그러므로, 신호의 크기가 매우 큰 상황이나 매우 작은 상황에서는 spike and slab 사전분포보다 horseshoe 사전분포를 사용하는 것이 더욱 안정적인 결과를 얻을 수 있다. Horseshoe 타입의 사전분포 중에서도, 비교연구 결과에 따르면 horseshoe plus와 horseshoe 사전분포는 이러한 상황에서 붕괴하는 경향이 있으므로, 붕괴 위험이 적은 regularized horseshoe를 사용하는 것이 더 좋은 결과를 줄 수 있을 것으로 보인다.

References

Barbieri MM, Berger JO (2004). Optimal predictive model selection, *The Annals of Statistics*, **32**, 870–897.

Bhadra A, Datta J, Polson NG, and Willard B (2017). The horseshoe+ estimator of ultra-sparse signals, *Bayesian Analysis*, **12**, 1105–1131.

Carvalho CM, Polson NG, and Scott JG (2010). The horseshoe estimator for sparse signals, *Biometrika*, **97**, 465–480.

Van ES, Oberski DL, and Mulder J (2019). Shrinkage priors for Bayesian penalized regression, *Journal of Mathematical Psychology*, **89**, 31–50.

George EI and McCulloch RE (1993). Variable selection via gibbs sampling, *Journal of the American Statistical Association*, **88**, 881–889.

Hoerl AE and Kennard RW (1970). Ridge regression: applications to nonorthogonal problem, *Technometrics*, **12**, 69–82.

Ishwaran H and Rao JS (2005). Spike and slab variable selection: frequentist and Bayesian strategies, *The Annals of Statistics*, **33**, 730–773.

- Lee SY, Pati D, and Mallick BK (2020). *Continuous Shrinkage Prior Revisited: A Collapsing Behavior and Remedy*, arXiv preprint arXiv:2007.02192.
- Makalic E and Schmidt DF (2015). A simple sampler for the horseshoe estimator, *IEEE Signal Processing Letters*, **23**, 179–182.
- Makalic E and Schmidt DF (2016). *High-Dimensional Bayesian Regularised Regression with the Bayesreg Package*, arXiv preprint arXiv:1611.06649.
- Martin R, Mess R, and Walker SG (2017). Empirical bayes posterior concentration in sparse high-dimensional linear models, *Bernoulli*, **23**, 1822–1847.
- Piironen J and Vehtari A (2017). On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 905–913.
- Piironen J and Vehtari A (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors, *Electronic Journal of Statistics*, **11**, 5018–5051.
- Tibshirani R (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, 267–288.

Received May 5, 2021; Revised May 27, 2021; Accepted June 4, 2021

베이지안 고차원 선형 회귀분석에서의 비교연구

신주원^a, 이경재^{1,a}

^a인하대학교 통계학과

요약

본 연구에서는, 고차원상황($p \gg n$)에서의 회귀분석 모형을 고려하여 다양한 베이지안 회귀분석 방법들을 비교하였다. Spike and slab 사전분포는 고차원 베이지안 회귀분석에서 가장 많이 사용되는 사전분포 중 하나이지만, 탐험해야 하는 모형 공간이 너무 크기 때문에 유한 표본에서 좋지 않은 성능을 보일 수 있다는 문제가 있다. 이에 대한 대안으로, horseshoe 사전분포를 비롯한 다양한 연속 수축사전분포들이 제안되어 사용되고 있다. 비록 위 사전분포들 각각에 대해서는 많은 연구들이 진행되고 있지만, 이들에 대한 포괄적인 비교연구는 매우 드물게 진행되고 있다. 따라서 본 연구에서는, spike and slab 사전분포와 다양한 연속수축사전분포들을 다양한 상황에서 비교하는 연구를 진행 하였다. 각 방법의 성능은 회귀계수 추정 측면과 변수선택 측면을 나누어 비교하였다. 최종적으로, 본 연구에서 진행된 시뮬레이션 연구에 기반하여, 사용시 몇 가지 주의점과 제안들을 제시하였다.

주요용어: 베이지안 회귀모형, spike and slab 사전분포, 연속 수축사전분포

¹교신저자: (22212) 인천광역시 미추홀구 인하로 100, 인하대학교 통계학과. E-mail: leekjstat@gmail.com