

Comparison of Laplace and Double Pareto Penalty: LASSO and Elastic Net

Minjung Kyung^{a,1}

^aDepartment of Statistics, Duksung Women's University

(Received September 22, 2014; Revised October 14, 2014; Accepted October 21, 2014)

Abstract

Lasso (Tibshirani, 1996) and Elastic Net (Zou and Hastie, 2005) have been widely used in various fields for simultaneous variable selection and coefficient estimation. Bayesian methods using a conditional Laplace and a double Pareto prior specification have been discussed in the form of hierarchical specification. Full conditional posterior distributions with each priors have been derived. We compare the performance of Bayesian lassos with Laplace prior and the performance with double Pareto prior using simulations. We also apply the proposed Bayesian hierarchical models to real data sets to predict the collapse of governments in Asia.

Keywords: Lasso, Elastic net, hierarchical models, scale mixture of normals, Laplace prior, double Pareto prior.

1. 서론

일반 선형모형에서 모형의 편의(bias)를 줄이기 위해 표본의 크기가 작더라도 많은 예측변수를 사용하는 경우가 있다. 그러나 이러한 경우 모형에 포함된 예측변수들이 많아서 변수들 간에 높은 선형 관계가 존재하는 다중공선성의 문제가 발생하게 된다. 그러므로 많은 수의 예측변수를 포함한 모형보다는 적은 수이나 설명력이 큰 중요한 예측변수들이 포함된 모형을 선호한다. 이러한 문제점들을 해결하고 정확한 예측성을 가진 적절한 예측변수들을 포함한 모형을 찾기 위해 선형 모형에서는 다양한 종류의 벌점을 사용한 최소제곱법이 발달하였다. 다양한 벌점을 사용한 벌점회귀모형은 변수 선택과 계수 추정을 동시에 활용할 수 있다는 특성 때문에 다양한 분야에서 활발하게 사용되고 있다.

n 개의 관측치에 대해 반응변수 Y 로의 p 개의 예측변수들의 선형모형은 다음과 같다.

$$\mathbf{y} = \mu \mathbf{1}_n + \mathbf{X}\beta + \epsilon, \quad (1.1)$$

여기에서 $\mathbf{y} = (y_1, \dots, y_n)'$, $\mathbf{X} = (X_1, X_2, \dots, X_p)$ 은 표준화된 예측변수들의 $n \times p$ 행렬이다. $\beta = (\beta_1, \dots, \beta_p)'$ 은 각 예측변수에 대한 회귀계수의 벡터이고, $\epsilon \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$ 는 오차벡터이다. 벌점회귀 모형은 보통 예측변수의 수가 관측치의 갯수보다 적은 경우($p < n$)에 사용되지만, 예측변수의 수가 표

This research was supported by the Duksung Women's University research grants 3000001920.

¹Department of Statistics, Duksung Women's University, 419 Samyang-ro 144 Gil 33, Dobong-Gu, Seoul 132-714, Korea. E-mail: mkyung@duksung.ac.kr

본크기보다 더 클때($p \gg n$)에도 사용한다. 만일 예측변수의 수가 표본의 크기보다 더 크면 잔차제곱합 $RSS = (\tilde{\mathbf{y}} - \mathbf{X}\beta)'(\tilde{\mathbf{y}} - \mathbf{X}\beta)$ ($\tilde{\mathbf{y}} = \mathbf{y} - \bar{y}\mathbf{1}_n$)을 최소화 시키는 최소제곱추정량(ordinary least-squares estimator)이 유일하지 않고 분산 또한 커져 모형의 예측도가 낮아진다. 이에 벌점을 적용한 회귀모형은 적절한 예측변수를 선택하는 방법을 제시함으로써 적은 수의 예측변수를 포함하지만 예측도가 높은 모형을 선택할 때에 사용한다.

대표적인 벌점회귀모형으로는 Hoerl과 Kennard (1970)가 제안한 능형회귀(Ridge regression)와 Tibshirani (1996)의 Least Absolute Shrinkage and Selection Operator(Lasso) 등이 있다. 능형회귀모형은 변수에 다중공선성이 존재할 때 유용한 모형으로 회귀모수에 L_2 norm의 벌점을 사용한다. 한편 Lasso의 경우 축소 추정과 변수 선택에 유리한 L_1 norm을 사용한다. 그러나 Tibshirani (1996)는 다중공선성이 존재할 때, $p > n$ 인 경우 예측성능에서는 능형회귀모형의 예측도가 Lasso보다 더 높고, Lasso는 최대 n 개의 변수까지만 선택할 수 있으며, 순서성이나 그룹변수 및 범주형 변수의 경우 개별적인 혹은 카테고리 별로 변수를 선택해야 하는 단점이 있다고 설명하였다. 이후 이러한 단점이 보완된 모형이 제시되었는데, 예측변수의 순서성이 고려된 벌점회귀모형으로는 Fused Lasso (Tibshirani 등, 2005)가 있고, 그룹 및 범주형 변수 전체를 하나로 묶어 변수선택법에 적용하는 Group Lasso (Yuan과 Lin, 2006)와 Blockwise Sparse Regression (Kim 등, 2006)등이 있다. 이 외에 Bridge regression (Frank와 Friedman, 1993), Smoothly Clipped Absolute Deviation(SCAD) (Fan과 Li, 2001) 등 다양한 벌점을 사용한 벌점회귀모형들이 제시 되었다.

이들 중 Lasso는 가장 많이 사용되는 모형이다. Lasso는 회귀모수에 미분불가능한 L_1 norm을 적용하여 잔차제곱합(RSS)을 최소화 시키는 벌점최소제곱법으로, Lasso 추정량은 다음과 같이 표현된다.

$$\hat{\beta}_L = \arg \min_{\beta} (\tilde{\mathbf{y}} - \mathbf{X}\beta)'(\tilde{\mathbf{y}} - \mathbf{X}\beta) + \lambda \sum_{j=1}^p |\beta_j|. \quad (1.2)$$

$\tilde{\mathbf{y}} = \mathbf{y} - \bar{y}\mathbf{1}_n$ 는 중심화된 반응값의 벡터이고, \mathbf{X} 는 표준화된 회귀변수의 행렬이며, $\lambda \geq 0$ 는 축소량을 조절하는 조율모수(tuning parameter)이다. 이러한 Lasso의 추정법에는 Osborne 등 (2000)이 제안한 두 가지 방법과 최근 Efron 등 (2004)이 발표한 Least Angle Regression Selection(LARS)알고리즘이 있다.

또한, 알려지지 않은 변수들의 집단과 예측변수들의 다중공선성 문제를 해결하기 위해 Zou와 Hastie (2005)는 정칙화된 Lasso의 형태인 Elastic Net(EN)을 제안하였다. Elastic net 추정량은 다음과 같이 표현된다.

$$\hat{\beta}_{EN} = \arg \min_{\beta} (\tilde{\mathbf{y}} - \mathbf{X}\beta)'(\tilde{\mathbf{y}} - \mathbf{X}\beta) + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p |\beta_j|^2,$$

λ_1 은 축소량을 조절하는 조율모수이고, λ_2 는 다중공선성을 설명하는 조율모수이다. 이러한 EN은 안정화된(stabilized) Lasso라고 해석되며, 개별변수의 변수 선택과 동시에 그룹변수들의 그룹 선택에도 가능한 장점이 있다. 그리고 EN은 $p \gg n$ 일 때에도 좋은 추정량을 제공한다.

한편, Tibshirani (1996)는 수식 (1.2)에서 Lasso 추정량을 회귀계수 β_j 의 사전분포로 독립인 라플라스(Laplace) 혹은 이중 지수(Double Exponential) 분포를 사용한 베이저안 사후 최빈값으로 설명하였다. 이에 Park과 Casella (2008)는 라플라스 분포를 정규분포의 척도모수에 지수분포를 혼합하여 표현한 베이저안 계층모형으로 깁스 표본(Gibbs sampling) 방법을 제안하였다. 그들은 다음과 같은 조건부 라플라스 사전분포를 사용하여 완전 베이즈 사후분포를 서술하였다.

$$\pi(\beta|\sigma^2) = \prod_{j=1}^p \frac{\lambda}{2\sigma} e^{-\frac{\lambda|\beta_j|}{\sigma}}, \quad (1.3)$$

$\pi(\sigma^2) = 1/\sigma^2$ 인 무정보적 척도불변성 주변 사후분포를 사용하였고, 단봉형인(unimodal) 사후분포를 확인하기 위해 σ^2 이 주어졌다는 조건부 분포의 중요성을 설명하였다. 왜냐하면 단봉형의 사후분포가 아니면 깃스 표본의 수렴이 늦어지며, 점추정값이 의미 없어지기 때문이다.

Lasso 추정량을 사용한 추론을 위해 0 회귀모수(zero coefficient)의 표준오차 추정량을 얻는 다양한 방법들이 제안되었다. Kyung 등 (2010)은 벌점회귀모형에서 표준오차의 신뢰성(reliability)에 문제가 있음을 지적하며, 근사 공분산행렬과 붓스트랩 방법도 유한표본의 경우 불일치(inconsistent) 성질로 Lasso 추정량들의 표준오차 추정량이 불안정함을 설명하였고, 만약 $\beta_j = 0$ 인 경우 Lasso는 초효율적(Knight과 Fu, 2002)임을 이용하여 Lasso 모수의 붓스트랩 추정량이 불일치함을 증명하였다. 이에 Kyung 등 (2010)은 기하 에르고딕 마코프 연쇄(geometrically ergodic Markov chain)를 바탕으로한 베이저안 방법론을 이용하여 여러 종류의 Lasso의 모수에 대한 완전 사후 분포(full posterior distribution)로 부터의 추정값과 표준오차를 구하는 방법을 제시하였다. Lasso는 모형선택을 위한 점추정 방법이지만, MCMC를 활용한 베이저안 방법에서는 변수선택법 대신 정확한 예측과 의미있는 예측변수들을 결정하는데 사후분포를 활용하여 점추정값 뿐 아니라 사후추정량의 의미있는 표준오차를 제공하였다.

최근 Armagan 등 (2013)은 선형모형에서 베이저안 축소 추정과 검정을 위해 일반화된 이중 파레토 사전분포(a generalized double Pareto prior)를 제안하였다. 이중 파레토 사전분포도 라플라스 혹은 정규분포의 척도모수 혼합법을 이용하여 표현하였고, 이 사전분포는 라플라스 사전분포와 정규-제프리의 사전분포(Normal-Jeffreys' priors)로 표현될 수 있는 일반화된 분포임을 증명하였다. 이중 파레토 분포는 라플라스 분포와 같이 0값에서 정상(spike)값을 갖고, 스튜던트 t -분포와 같은 꼬리의 성질도 갖고 있다. Armagan 등 (2013)은 MAP(maximum a posterior) 추정량이 희박한 추정(sparse estimation)으로 여러 문제에서 중요한 역할을 하고, 안정된 정칙 과정을 통해서 추정되며, 점근적 성질이 있음을 증명하였다.

이 논문에서는 선형모형에 L_1 벌점을 사용한 Lasso와 L_1 과 L_2 벌점을 같이 사용한 EN에 대해 깃스 표본법을 이용한 베이저안 분석을 적용하고, 사전분포로는 라플라스 분포와 이중 파레토 분포를 적용하여 비교한다. 일반 Lasso모형에 이중 파레토 분포를 적용하면 라플라스 분포를 적용한 것보다 희박성(sparsity)은 클 것이라 예측되지만, 정확한 예측과 의미있는 예측변수를 결정하는 방법론에는 어떤 결과가 나올지 비교 분석하였다. 그리고 라플라스 분포와 이중 파레토 분포를 EN에 있는 L_1 벌점에 적용했을 때, 추정량에 어떤 변화가 있는지도 비교하였다. 이에 라플라스 사전분포를 사용한 Lasso와 EN은 본래의 이름인 베이저안 Lasso와 베이저안 EN으로 사용하고, 이중 파레토 분포를 적용한 Lasso와 EN은 베이저안 이중 파레토 Lasso(BDP Lasso)와 베이저안 이중 파레토 EN(BDP EN)으로 사용한다.

본 논문의 구성은 다음과 같다. 2절에서는 라플라스 사전분포와 이중 파레토 분포를 사용한 계층적 모형에 대해서 설명하고, 3절에서는 각 계층적 모형의 표본추출법과 깃스 표본의 수렴에 대해서 설명한다. 그리고 4절에서는 모의실험과 데이터 분석을 통해 제안된 모형을 적용하며, 5절에서는 요약과 결론으로 끝을 맺는다.

2. 계층적 모형

이 절에서는 라플라스 사전분포와 이중파레토 분포의 정규 혼합법에 대해서 논하고, 완전 베이저안 사후

분포를 이용하기 위해 계층적 모형의 구현에 대해 설명한다.

2.1. 라플라스 사전 분포

별점회귀모형의 베이지안 방법론은 정규분포의 척도모수(scale parameter)에 지수분포를 혼합(mixture)하여 라플라스 분포(Laplace distribution)를 구현하는 Andrews와 Mallows (1974)의 정규 혼합모형을 응용한 계층적 모형으로 표현할 수 있다.

$$\frac{a}{2} \exp(-a|z|) = \int_0^\infty \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{z^2}{2s}\right) \frac{a^2}{2} \exp\left(-\frac{a^2}{2}s\right) ds.$$

이러한 라플라스 사전분포를 Figueriedo (2003)는 EM을 이용한 지도학습(supervised learning)모형에 사용하였고, Bae와 Mallick (2004)은 유전자 선택 문제에서 계층구조를 실현한 마코프체인몬테카를로(MCMC)에 사용하였다.

Park과 Casella (2008)에서 구현한 Lasso는 다음과 같은 계층적 모형으로 나타난다.

$$\begin{aligned} \mathbf{y} | \mu, \mathbf{X}, \beta, \sigma^2 &\sim N_n(\mu \mathbf{1}_n + \mathbf{X}\beta, \sigma^2 \mathbf{I}_n), \\ \beta | \sigma^2, \tau_1^2, \dots, \tau_p^2 &\sim N_p(\mathbf{0}_p, \sigma^2 \mathbf{D}_\tau), \quad \mathbf{D}_\tau = \text{diag}(\tau_1^2, \dots, \tau_p^2), \\ \tau_1^2, \dots, \tau_p^2 &\sim \prod_{j=1}^p \frac{\lambda^2}{2} e^{-\frac{\lambda^2 \tau_j^2}{2}} d\tau_j^2, \quad \tau_1^2, \dots, \tau_p^2 > 0, \\ \sigma^2 &\sim \pi(\sigma^2) d\sigma^2, \quad \sigma^2 > 0. \end{aligned} \quad (2.1)$$

모수 μ 는 독립인 편편사전분포(flat prior)로도 사후분포의 특성을 유지한다. 위의 식에서 $\tau_1^2, \dots, \tau_p^2$ 를 적분하면 원하는 형식인 식 (1.3)에서와 같은 β 에 대한 조건부 사전분포가 된다. σ^2 에는 공액사전분포인 역감마분포를 사용해도 되지만 부적절 사전분포(improper prior)를 사용하여도 역감마함수 사후분포가 된다. 위와 같이 $\tau_1^2, \dots, \tau_p^2$ 에 모수가 $\lambda^2/2$ 인 지수분포를 가정하면, 결과적으로 β 에는 평균이 0이고 분산이 $\sigma^2 \lambda^{-2}$ 인 조건부 라플라스 분포를 적용하는 것과 같다.

한편, Elastic net의 사후분포를 이끌어내기 위하여 회귀계수에 다음과 같은 조건부 사전분포를 적용한다.

$$\pi(\beta | \sigma^2) \propto \exp\left\{-\frac{\lambda_1}{\sqrt{\sigma^2}} \sum_{j=1}^p |\beta_j| - \frac{\lambda_2}{2\sigma^2} \sum_{j=1}^p \beta_j^2\right\}.$$

이를 위하여 계층적 모형을 다음과 같이 나타낼 수 있다 (Kyung, 등, 2010).

$$\begin{aligned} \mathbf{y} | \mu, \mathbf{X}, \beta, \sigma^2 &\sim N_n(\mu \mathbf{1}_n + \mathbf{X}\beta, \sigma^2 \mathbf{I}_n), \\ \beta | \sigma^2, \tau_1^2, \dots, \tau_p^2 &\sim N_p(\mathbf{0}_p, \sigma^2 \Sigma_\beta), \\ \tau_i^2 | \lambda_1 &\sim \text{Exp}(\lambda_1^2), \quad i = 1, \dots, p, \\ \sigma^2 &\sim \pi(\sigma^2) d\sigma^2, \quad \sigma^2 > 0, \end{aligned} \quad (2.2)$$

여기서 Σ_β 는 Lasso의 계층적 모형과 다른 공분산 행렬로

$$\Sigma_\beta = \text{Diagonal}(\tau_i^{-2} + \lambda_2)^{-1}, \quad i = 1, \dots, p$$

이다. 회귀모수벡터 β 와 λ_1, λ_2 는 τ 가 주어졌을 때 조건적 독립이다. 그리고 Lasso와 또 다른 점은 λ_2 가 Σ_β 의 대각원소에 나타나며, 이는 김스표분방법으로 쉽게 추정할 수 있다는 것이다.

다음은 조율모수에 대해서 살펴보고자 한다. Park과 Casella (2008)는 Lasso에 조건부 라플라스 분포를 활용하면서 조율모수(Tuning Parameter)의 추정을 위해 EM(Expectation-Maximization) 알고리즘을 사용하였고, 완벽 깃스 표본을 위해 감마 분포를 사전분포로 사용하였다. 이 연구에서는 Park과 Casella (2008)에서 사용한 조율모수의 제곱, 즉 λ^2 에 감마 사전분포를 사용한다.

$$\pi(\lambda^2) = \frac{\delta^r}{\Gamma(r)} (\lambda^2)^{r-1} e^{-\delta\lambda^2}, \quad r > 0, \delta > 0. \quad (2.3)$$

EN에는 두개의 조율모수가 있는데, L_1 벌점에 적용되는 조율모수 λ_1^2 에는 Lasso에 적용한 감마분포를 적용하고, L_2 벌점의 조율모수 λ_2 에 다른 모수를 갖는 감마분포를 적용한다 (Kyung 등, 2010).

$$\begin{aligned} \pi(\lambda_1^2) &= \frac{\delta_1^{r_1}}{\Gamma(r_1)} (\lambda_1^2)^{r_1-1} e^{-\delta_1\lambda_1^2}, \quad r_1 > 0, \delta_1 > 0, \\ \pi(\lambda_2) &= \frac{\delta_2^{r_2}}{\Gamma(r_2)} (\lambda_2)^{r_2-1} e^{-\delta_2\lambda_2}, \quad r_2 > 0, \delta_2 > 0. \end{aligned} \quad (2.4)$$

2.2. 이중 파레토 사전분포

일반화된 이중 파레토 분포(Generalized Double Pareto Distribution)는 Armagan 등 (2013)에 의해 선형 모형의 회귀모수에 대한 베이저안 축소 추정량(Bayesian shrinkage estimator)의 사전분포로 제안되었다. 일반화된 이중 파레토 분포는 다음과 같은 분포함수를 갖는다.

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

ϕ 는 척도모수이고 α 는 형태모수이다. Armagan 등 (2013)은 명제 1에서 일반화된 이중 파레토 분포를 정규분포의 척도모수(scale parameter)에 지수분포를 혼합하고 지수분포의 모수의 제곱근에 감마분포를 혼합하여 구현하였다. 즉, $\theta \sim N(0, \tau)$ 와 $\tau \sim \text{Exp}(\lambda^2/2)$, 그리고 $\lambda \sim \text{Ga}(\alpha, \eta)$ 를 혼합하여 θ 에 대한 주변분포를 구하는데 이는 $\theta \sim \text{GDP}(\phi = \eta/\alpha, \alpha)$ 이 된다. 이것은 역 가우스 분포(Inverse Gaussian distribution)를 이용한 변수변환으로 쉽게 증명할 수 있다.

이중 파레토 사전분포를 사용하는 Lasso의 계층적 모형은 다음과 같다.

$$\begin{aligned} \mathbf{y} | \mu, \mathbf{X}, \beta, \sigma^2 &\sim N_n(\mu\mathbf{1}_n + \mathbf{X}\beta, \sigma^2\mathbf{I}_n), \\ \beta | \sigma^2, \tau_1^2, \dots, \tau_p^2 &\sim N_p(\mathbf{0}_p, \sigma^2\mathbf{D}_\tau), \quad \mathbf{D}_\tau = \text{diag}(\tau_1^2, \dots, \tau_p^2), \\ \tau_1^2, \dots, \tau_p^2 &\sim \prod_{j=1}^p \frac{\lambda^2}{2} e^{-\frac{\lambda^2\tau_j^2}{2}} d\tau_j^2, \quad \tau_1^2, \dots, \tau_p^2 > 0, \\ \lambda &\sim \text{Ga}(\alpha, \eta) d\lambda, \quad \lambda > 0, \\ \sigma^2 &\sim \pi(\sigma^2) d\sigma^2, \quad \sigma^2 > 0, \end{aligned} \quad (2.5)$$

그리고 이중 파레토 사전분포를 사용한 Elastic Net의 계층적 모형은 다음과 같다.

$$\begin{aligned} \mathbf{y} | \mu, \mathbf{X}, \beta, \sigma^2 &\sim N_n(\mu\mathbf{1}_n + \mathbf{X}\beta, \sigma^2\mathbf{I}_n), \\ \beta | \sigma^2, \tau_1^2, \dots, \tau_p^2 &\sim N_p(\mathbf{0}_p, \sigma^2\mathbf{\Sigma}_\beta), \\ \tau_1^2, \dots, \tau_p^2 &\sim \prod_{j=1}^p \frac{\lambda_1^2}{2} e^{-\frac{\lambda_1^2\tau_j^2}{2}} d\tau_j^2, \quad \tau_1^2, \dots, \tau_p^2 > 0, \end{aligned}$$

$$\begin{aligned}\lambda_1 &\sim \text{Ga}(\alpha_1, \eta_1) d\lambda_1, \quad \lambda_1 > 0, \\ \lambda_2 &\sim \text{Ga}(\alpha_2, \eta_2) d\lambda_2, \quad \lambda_2 > 0, \\ \sigma^2 &\sim \pi(\sigma^2) d\sigma^2, \quad \sigma^2 > 0.\end{aligned}$$

이중 파레토 사전분포를 사용한 Lasso와 EN의 계층적 모형은 라플라스 사전분포를 사용한 Lasso의 계층적 모형과 같지만, 조율모수인 λ^2 혹은 λ_1^2 이 아닌 λ 혹은 λ_1 에 감마사전분포를 사용한다는 차이점이 있다. 사전분포를 λ 에 가정하느냐 λ^2 에 적용하느냐의 차이는 큰 차이가 없는 것 같지만, 계층적 모형의 경우 사후분포가 달라지며, 특히 별점을 사전분포로 가정한 모형들의 경우 계층적 모형을 합친 주변분포가 라플라스 사전분포이거나 이중 파레토 사전분포가 되는 큰 차이점이 있기 때문에 표본추출 및 모형적용에 어떤 영향을 미치는지 비교할 필요가 있다.

3. 표본추출 구조

이 절에서는 2절에서 표현한 계층적 모형으로 라플라스 사전분포와 이중 파레토 사전분포 각각의 경우의 조건부 사후분포를 도출하고, 깃스표본의 수렴에 대하여 서술한다. 분산 σ^2 에 대한 사전분포로는 부적절한 사전분포(improper prior) $\pi(\sigma^2) = 1/\sigma^2$ 를 사용하였다.

3.1. 라플라스 사전 분포를 사용한 조건부 사후분포

라플라스 사전분포를 사용한 계층적 모형으로부터 도출한 Lasso의 사후분포는 다음과 같이 주어진다.

$$\begin{aligned}\beta \mid \mu, \sigma^2, \tau_1^2, \dots, \tau_p^2, \mathbf{X}, \mathbf{y} &\sim N_p \left((\mathbf{X}'\mathbf{X} + \mathbf{D}_\tau^{-1})^{-1} \mathbf{X}'\tilde{\mathbf{y}}, \sigma^2 (\mathbf{X}'\mathbf{X} + \mathbf{D}_\tau^{-1})^{-1} \right), \\ \frac{1}{\tau_j^2} = \gamma_j \mid \mu, \beta, \sigma^2, \mathbf{X}, \mathbf{y} &\sim \text{inverse Gaussian} \left(\frac{\lambda_j^2 \sigma}{|\beta_j|}, \lambda_j^2 \right) \mathbf{I}(\gamma_j > 0) \quad \text{for } j = 1, \dots, p, \\ \sigma^2 \mid \mu, \beta, \tau_1^2, \dots, \tau_p^2, \mathbf{X}, \mathbf{y} &\sim \text{inverted gamma} \left(\frac{n-1+p}{2}, \frac{1}{2} (\tilde{\mathbf{y}} - \mathbf{X}\beta)' (\tilde{\mathbf{y}} - \mathbf{X}\beta) + \frac{\lambda}{2} \beta' \mathbf{D}_\tau^{-1} \beta \right) \\ \lambda^2 \mid \beta, \sigma^2, \tau_1^2, \dots, \tau_p^2, \mathbf{X}, \tilde{\mathbf{y}} &\sim \text{gamma} \left(p + \alpha, \frac{1}{2} \sum_{j=1}^p \tau_j^2 + \eta \right).\end{aligned} \quad (3.1)$$

여기서 \mathbf{D}_τ 는 대각원호가 (τ_i^2) , $i = 1, \dots, p$ 인 대각행렬이다. β , σ^2 와 λ^2 의 완전 조건부 사후분포(fully conditional posterior distribution)는 공액류(conjugate class)를 만족하지만, $1/\tau_j^2 = \gamma_j$ 는 역가우스 분포가 된다. 사후분포의 도출은 계산이 어렵지 않으며, 자세한 사항은 생략한다.

λ_1 과 λ_2 가 주어졌을 때, 라플라스 사전분포를 사용한 EN의 완전 조건부 사후분포는 다음과 같다.

$$\begin{aligned}\beta \mid \sigma^2, \tau_1^2, \dots, \tau_p^2, \mathbf{X}, \tilde{\mathbf{y}} &\sim N_p \left((\mathbf{X}'\mathbf{X} + \mathbf{D}_\tau^{*-1})^{-1} \mathbf{X}'\tilde{\mathbf{y}}, \sigma^2 (\mathbf{X}'\mathbf{X} + \mathbf{D}_\tau^{*-1})^{-1} \right), \\ \frac{1}{\tau_j^2} = \gamma_j \mid \beta, \sigma^2, \mathbf{X}, \tilde{\mathbf{y}} &\sim \text{inverse Gaussian} \left(\sqrt{\frac{\lambda_1^2 \sigma^2}{\beta_j^2}}, \lambda_1^2 \right) \mathbf{I}(\gamma_j > 0) \quad \text{for } j = 1, \dots, p, \\ \sigma^2 \mid \beta, \tau_1^2, \dots, \tau_p^2, \mathbf{X}, \tilde{\mathbf{y}} &\sim \text{inverted gamma} \left(\frac{n-1+p}{2}, \frac{1}{2} (\tilde{\mathbf{y}} - \mathbf{X}\beta)' (\tilde{\mathbf{y}} - \mathbf{X}\beta) + \frac{1}{2} \beta' \mathbf{D}_\tau^{*-1} \beta \right),\end{aligned} \quad (3.2)$$

여기서 \mathbf{D}_τ^* 는 대각원호가 $(\tau_i^{-2} + \lambda_2)^{-1}$, $i = 1, \dots, p$ 인 대각행렬이다. 마찬가지로 β 와 σ^2 의 완전 조건부 사후분포는 공액류의 성질을 갖지만, $1/\tau_j^2 = \gamma_j$ 는 역가우스 분포가 된다.

감마 사전분포 $Ga(r_h, \delta_h)$, ($h = 1, 2$)를 적용한 λ_1^2 과 λ_2 의 완전 조건부 공액 사후분포는 다음과 같다.

$$\lambda_1^2 | \beta, \sigma^2, \tau_1^2, \dots, \tau_p^2, \lambda_2, \mathbf{X}, \tilde{\mathbf{y}} \sim \text{gamma} \left(p + \alpha_1, \frac{1}{2} \sum_{j=1}^p \tau_j^2 + \eta_1 \right),$$

$$\lambda_2 | \beta, \sigma^2, \tau_1^2, \dots, \tau_p^2, \lambda_1, \mathbf{X}, \tilde{\mathbf{y}} \sim \text{gamma} \left(\frac{p}{2} + \alpha_2, \frac{1}{2\sigma^2} \sum_{j=1}^p \beta_j^2 + \eta_2 \right).$$

λ_1^2 에 대한 사후분포는 Lasso의 λ^2 에 대한 사후분포를 유도하는 과정과 같다. 그러나 λ_2^2 에 대한 사후분포는 전자와는 달리 완전사후분포로부터 유도하는 방법으로 계산한다.

3.2. 이중 파레토 사전분포를 사용한 조건부 사후 분포

이중 파레토 사전분포를 사용한 Lasso의 완전 조건부 사후분포는 다음과 같다.

$$\beta | \mu, \sigma^2, \tau_1^2, \dots, \tau_p^2, \mathbf{X}, \mathbf{y} \sim N_p \left((\mathbf{X}'\mathbf{X} + \mathbf{D}_\tau^{-1})^{-1} \mathbf{X}'\tilde{\mathbf{y}}, \sigma^2 (\mathbf{X}'\mathbf{X} + \mathbf{D}_\tau^{-1})^{-1} \right),$$

$$\frac{1}{\tau_j^2} = \gamma_j | \mu, \beta, \sigma^2, \mathbf{X}, \mathbf{y} \sim \text{inverse Gaussian} \left(\frac{\lambda\sigma}{|\beta_j|}, \lambda^2 \right) \mathbf{I}(\gamma_j > 0) \quad \text{for } j = 1, \dots, p, \quad (3.3)$$

$$\sigma^2 | \mu, \beta, \tau_1^2, \dots, \tau_p^2, \mathbf{X}, \mathbf{y} \sim \text{inverted gamma} \left(\frac{n+p}{2}, \frac{1}{2} (\tilde{\mathbf{y}} - \mathbf{X}\beta)' (\tilde{\mathbf{y}} - \mathbf{X}\beta) + \frac{1}{2} \beta' \mathbf{D}_\tau^{-1} \beta \right),$$

$$\lambda | \beta, \sigma^2, \tau_1^2, \dots, \tau_p^2, \mathbf{X}, \tilde{\mathbf{y}} \sim \text{gamma} \left(p + \alpha, \frac{1}{2} \sum_{j=1}^p \frac{|\beta_j|}{\sigma} + \eta \right).$$

여기서 \mathbf{D}_τ 는 대각원호가 (τ_i^2) , $i = 1, \dots, p$ 인 대각행렬이다. β , σ^2 와 λ 의 완전 조건부 공액 사후분포를 만족하지만, $1/\tau_j^2 = \gamma_j$ 는 역가우스 분포가 된다.

한편, λ_1 과 λ_2 가 주어진 조건 하에서 이중 파레토 조건부 사전분포를 고려한 Elastic Net의 사후분포는 다음과 같다.

$$\beta | \sigma^2, \tau_1^2, \dots, \tau_p^2, \mathbf{X}, \tilde{\mathbf{y}} \sim N_p \left((\mathbf{X}'\mathbf{X} + \mathbf{D}_\tau^{-1} + \lambda_2 \mathbf{I})^{-1} \mathbf{X}'\tilde{\mathbf{y}}, \sigma^2 (\mathbf{X}'\mathbf{X} + \mathbf{D}_\tau^{-1} + \lambda_2 \mathbf{I})^{-1} \right),$$

$$\frac{1}{\tau_j^2} = \gamma_j | \mu, \beta, \sigma^2, \mathbf{X}, \mathbf{y} \sim \text{inverse Gaussian} \left(\frac{\lambda_1 \sigma}{|\beta_j|}, \lambda_1^2 \right) \mathbf{I}(\gamma_j > 0) \quad \text{for } j = 1, \dots, p, \quad (3.4)$$

$$\sigma^2 | \beta, \tau_1^2, \dots, \tau_p^2, \mathbf{X}, \tilde{\mathbf{y}} \sim \text{inverted gamma} \left(\frac{n}{2} + p, \frac{1}{2} (\tilde{\mathbf{y}} - \mathbf{X}\beta)' (\tilde{\mathbf{y}} - \mathbf{X}\beta) + \frac{1}{2} \beta' (\mathbf{D}_\tau^{-1} + \lambda_2 \mathbf{I}) \beta \right),$$

여기서 \mathbf{D}_τ^* 는 대각원호가 $(\tau_i^{-2} + \lambda_2)^{-1}$, $i = 1, \dots, p$ 인 대각행렬이다. 마찬가지로 β 와 σ^2 의 완전 조건부 사후분포는 공액류의 성질을 갖지만, $1/\tau_j^2 = \gamma_j$ 는 역가우스 분포가 된다.

λ_h , $h = 1, 2$ 에 대해 감마 사전분포를 사용한 완전 조건부 사후분포는 다음과 같다.

$$\lambda_1 | \beta, \sigma^2, \tau_1^2, \dots, \tau_p^2, \lambda_2, \mathbf{X}, \tilde{\mathbf{y}} \sim \text{gamma} \left(p + \alpha_1, \frac{1}{2} \sum_{j=1}^p \frac{|\beta_j|}{\sigma} + \eta_1 \right),$$

$$\lambda_2 | \beta, \sigma^2, \tau_1^2, \dots, \tau_p^2, \lambda_1, \mathbf{X}, \tilde{\mathbf{y}} \sim \text{gamma} \left(\frac{p}{2} + \alpha_2, \frac{1}{2\sigma^2} \sum_{j=1}^p \beta_j^2 + \eta_2 \right).$$

λ_1 과 λ_2 의 완전 조건부 공액 사후분포는 감마분포가 되며, Lasso의 경우와 마찬가지로, λ_2 의 사후분포 유도는 완전 사후분포로부터 시작한다.

라플라스 사전분포와 이중 파레토 사전분포로부터 유도한 사후분포의 다른점은 벌점모수 λ 의 사후분포에 대한 모수가 달라진다는 것이다. Lasso모형의 라플라스 사전분포 경우 λ^2 의 사후분포의 척도모수가 $\frac{1}{2} \sum_{j=1}^p \tau_j^2 + \eta$ 이지만, 이중 파레토 분포 경우 λ 의 사후분포의 척도모수가 $\frac{1}{2} \sum_{j=1}^p \frac{|\beta_{j1}|}{\sigma} + \eta$ 가 된다. 이는 라플라스 사전분포의 경우 벌점모수의 척도모수가 회귀계수의 척도모수의 함수라면, 이중 파레토 사전분포의 벌점모수의 척도모수는 회귀계수의 0값으로 부터의 거리의 절대치의 합의 함수가 된다. EN의 경우에도 벌점모수 중 λ_1 의 사후분포가 Lasso모형의 경우와 같이 라플라스 사전분포의 경우와 이중 파레토 사전분포의 경우가 다그다. 다음 절에서 이러한 모수의 차이가 회귀계수의 추정 및 예측 오차에 어떤 영향을 미치는가를 모의실험 및 데이터 분석을 통해 설명한다.

3.3. 깃스 표본의 수렴

위에서 유도한 사후분포를 통해 얻는 베이저안 Lasso와 EN, 그리고 베이저안 이중 파레토 Lasso와 EN의 깃스 표본의 수렴에 대해 살펴보자. 이는 Liu 등 (1994) 과 Hobert와 Geyer (1998)의 계층모형에서의 결합모형과 주변모형의 수렴성에 대한 관계로부터 생각할 수 있다. Liu 등 (1994)은 결합분포에서의 깃스표본보다는 주변분포에서의 깃스표본이 정상분포로 더 빨리 수렴한다는 것을 증명하였다. 그리고 Hobert와 Geyer (1998)는 일원임의효과모형을 통해 공액사전분포족의 경우 회귀계수와 척도모수의 블록깃스 표본의 기하학적 에르고딕(geometrically ergodic) 성질을 증명하였다. 이를 활용하여 Kyung 등 (2010)은 회귀모형은 일원임의효과모형의 임의효과를 적분한 주변확률모형이므로 총변동(total variation) norm을 통해 라플라스 사전분포를 사용한 베이저안 Lasso의 깃스 표본은 기하학적 에르고딕 마코프 체인(geometrically ergodic Markov chain)임을 증명하였다.

이중 파레토 사전분포를 사용한 베이저안 이중 파레토 Lasso는 라플라스 사전분포를 사용한 계층적 모형과 같지만, 벌점모수의 사후분포에 대한 모수가 달라진다. 그러므로 이중 파레토 사전분포를 사용한 계층적 모형 역시 일원임의효과모형의 임의효과를 적분한 주변확률모형이므로 Hobert와 Geyer (1998)의 블록깃스 표본의 기하학적 에르고딕 성질을 사용할 수 있다. 그러므로 총변동 norm을 사용하여 베이저안 이중 파레토 Lasso역시 깃스 표본은 기하학적 에르고딕 마코프 체인임을 설명할 수 있다.

베이저안 EN 및 베이저안 이중 파레토 EN의 깃스 표본은 라플라스 또는 이중 파레토 사전분포를 사용한 Lasso의 사후분포로부터의 깃스 표본에 L_2 벌점의 조율모수 λ_2 에 대한 표본과정의 더해진 표본이라 할 수 있겠다. 이 모형은 일원임의효과모형의 임의효과를 적분하지 않는 확장된 형태의 결합확률모형으로 깃스표본의 수렴 속도는 주변확률모형보다 느리지만 총변동 norm을 통해 기하학적 에르고딕 마코프 체인임을 예상한다. 즉, Kyung 등 (2010)에서 증명한 베이저안 Lasso의 모수 블록 (β, μ) 와 (λ, τ) 를 (β, μ) 와 $(\lambda_1, \lambda_2, \tau)$ 으로 확장한 깃스표본은 기하학적 에르고딕 마코프 체인이라 할 수 있다.

4. 적용

이 절에서는 제안된 사전분포의 성능을 비교하기 위한 모의 실험을 진행한다. 이 실험에서는 다른 Lasso논문들에서 사용한 설정을 사용함으로써 제안된 모형이 모의실험에 어떻게 적용되는지를 파악하려고 한다. 그리고 예측정확도를 판단하기 위해 실제 데이터에 제안한 모형을 적용하였는데, 실제데이터는 Kyung 등 (2010)이 사용한 아시아 국가 실패(state failure)데이터이다.

4.1. 모의실험

모의실험을 위해 다음과 같은 참모형에서 모의자료를 만들어낸다.

Table 4.1. Mean and standard errors of prediction errors for Examples 1–3 and three methods, based on 100 replications: Bayesian lassos with Laplace prior vs. Bayesian lassos with double Pareto prior vs. LARS-EN from Zou and Hastie (2005), standard errors in parentheses.

알고리즘	모형	모의자료 1	모의자료 2	모의자료 3	
베이저안	라플라스	Lasso	1.26(0.52)	1.29(0.55)	48.36(13.21)
		Elastic Net	1.36(0.50)	1.33(0.55)	39.29(12.00)
	이중파레토	Lasso	1.23(0.52)	1.27(0.57)	40.62(13.71)
		Elastic Net	1.38(0.54)	1.39(0.61)	43.90(14.04)
LARS	Lasso	2.03(0.96)	1.78(0.77)	68.80(11.24)	
	Elastic Net	2.28(0.77)	2.43(1.08)	64.76(17.19)	

$$\mathbf{y} = \mathbf{X}\beta + \sigma\epsilon, \quad \epsilon_i \sim \text{iid } N(0, 1) \text{ for } i = 1, \dots, n.$$

총 세 가지의 다른 설정이 사용되었는데, 이 모형들은 원 Lasso 논문 (Tibshirani, 1996)과 elastic net 논문 (Zou 와 Hastie, 2005)에서 적용된 모의 자료의 설정이다.

모의자료 1: 8개의 예측변수를 포함한 참모형에서, 모형을 적용하기 위해 $n = 20$ 개의 모의데이터를 생성한다. 참모형의 회귀계수는 $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)'$ 이고 표준편차는 $\sigma = 3$ 으로 고정한다. 두 예측변수 사이의 짝상관관계(pairwise correlation)는 $\text{corr}(i, j) = 0.5^{|i-j|}$, $i, j = 1, 2, \dots, 8$ 이다.

모의자료 2: 8개의 예측변수를 포함한 참모형에서, 모형을 적용하기 위해 $n = 20$ 개의 모의데이터를 생성한다. 참모형의 회귀계수는 $\beta = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)'$ 이고 표준편차는 $\sigma = 3$ 으로 고정한다. 두 예측변수 사이의 짝상관관계(pairwise correlation)는 $\text{corr}(i, j) = 0.5^{|i-j|}$, $i, j = 1, 2, \dots, 8$ 이다.

모의자료 3: 40개의 예측변수를 포함한 모형에서 모형 적용을 위해 $n = 100$ 개의 모의자료를 생성하고, 예측오차의 비교를 위해 $n = 400$ 개의 실험자료를 생성한다. 참모형의 회귀계수는 $\beta = (\mathbf{0}', \mathbf{2}', \mathbf{0}', \mathbf{2}')$ 으로 벡터 $\mathbf{0}$ 은 길이가 10인 0 벡터이고, 벡터 $\mathbf{2}$ 는 길이가 10인 2 벡터이다. 표준편차는 $\sigma = 15$ 이고, 짝상관관계는 $\text{corr}(i, j) = 0.5$ 으로 고정한다.

총 세 가지 설정에서 생성된 각각의 모의자료에 라플라스 사전분포를 사용한 Lasso와 EN, 그리고 이중 파레토 사전분포를 사용한 Lasso와 EN을 적용하고 비교하였다. 벌점모수 λ 에는 형태모수 $a = 1$ 과 척도모수 $b = 0.1$ 인 비교적 평평(flat)하고 최대우도추정량(MLE) 근처에서 최대 사후 확률을 갖는 감마분포를 사전분포로 사용하였다. 분산 σ^2 에 대한 사전분포로는 부적절한 사전분포(improper prior)를 사용하였다. 베이저안 추정값으로 10,000번의 burn-in 반복 후 10,000개의 깃스 표본으로부터 계산된 사후 평균을 사용하였다. 모의 모형에 대한 예측오차의 평균값을 계산하기 위해, 각각 주어진 모형의 추정을 100번 반복하였다.

Table 4.1은 각 설정의 모의자료를 라플라스 사전분포를 사용한 베이저안 Lasso와 베이저안 EN과 이중 파레토 사전분포를 사용한 베이저안 이중 파레토 Lasso와 베이저안 이중 파레토 EN, 그리고 LARS 알고리즘과 LARS-EN 알고리즘 방법을 적용하여 분석하는 과정을 100번 반복하며 얻은 계산된 예측오차의 평균값과 표준오차를 요약한 내용이다. 모의자료 1과 모의자료 2의 설정의 경우 이중 파레토 사전분포를 사용한 Lasso의 예측오차가 가장 작았지만, 라플라스 사전분포를 사용한 모형들의 예측오차와 큰 차이는 없었다. 예측오차에 대한 값은 평균제곱합의 경우와 달리 LARS 알고리즘을 통해 구한 모형들의 경우 더 큰 값을 보였다. 모의자료 3의 경우는 Lasso나 EN의 모형보다는 Fused Lasso가 더 적

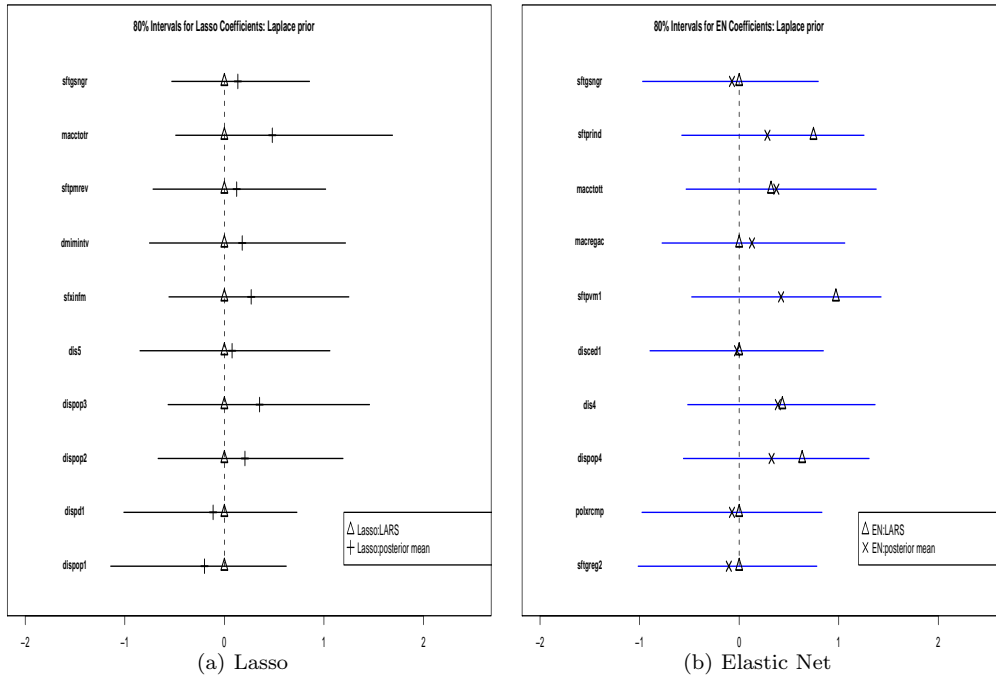


Figure 4.1. Top ten effects of posterior estimates (absolute value) by Bayesian Lasso and Bayesian EN with Laplace prior. The solid line is the 80% posterior credible interval.

합한 모형이지만, 라플라스 사전분포를 사용한 베이시안 EN 모형과 이중파레토 분포를 사용한 Lasso가 다른 모형에 비해 작은 예측오차값을 갖는다.

별점모수인 λ 는 깃스 표본으로부터 사후 평균값을 점추정값으로 사용하였고, LARS 알고리즘에 이 값을 사용하여 회귀모수를 추정하였다. 이러한 별점모수의 추정치를 고려할 때, 전체적으로 이중 파레토 사전분포를 사용한 Lasso 모형이 다른 모형이나 방법들에 비해 예측오차 값이 작다는 것을 알 수 있었다. LARS 알고리즘을 통해 추정된 모형의 예측오차에는 제거된 설명변수들이 사용되지 않지만, 베이시안 방법을 통해 추정된 모형의 예측오차에는 비록 작은 추정치라도 모든 설명변수들이 사용된다. 그러므로 LARS 알고리즘을 적용한 분석법의 예측오차는 평균적으로 베이시안 방법을 적용한 방법의 예측오차보다 값이 작다고 설명할 수 있다.

4.2. 데이터 분석

State Failure Task Force(SFTF; Esty 등, 1999)는 미국 정부가 지원하는 학제간(interdisciplinary) 연구원들에 의해 수집된 자료로, 폭력과 혼란으로 붕괴된 정부를 이해하고 예측하기 위해 사용되었다. 원 자료는 아시아, 아프리카, 남미, 그리고 중동지방의 네 지역별 데이터로 결측치(missing value) 및 중복된 관측치를 포함하고 있고, 전체 지역의 자료를 하나의 큰 데이터로 사용하여 분석하는 등 많은 문제점이 지적된 자료이다. 자료에 대한 자세한 사항은 Kyung 등 (2010)에서 확인할 수 있다.

여기에서는 아시아지역 데이터만을 선별하여 분석한다. 원래의 아시아지역 데이터는 $p = 128$ 개의 예측변수와 정부의 붕괴상태에 대한 반응변수로 이루어졌지만, 관측값의 70% 이상 결측치인 4개의 예측변

Table 4.2. 95% posterior credible intervals of the top ten effects of posterior estimates (absolute value) by Bayesian Lasso and Bayesian EN with Laplace prior.

Variable	Bayesian Lasso Quantiles			Variable	Bayesian EN Quantiles		
	0.05	0.50	0.95		0.05	0.50	0.95
didpop1	-1.5354	-0.1237	0.9404	sftgreg2	-1.3487	-0.0844	1.0906
dispd1	-1.4080	-0.0754	1.0654	polxrcmp	-1.2956	-0.0584	1.1476
dispop2	-0.9867	0.1250	1.6272	dispop4	-0.8404	0.2629	1.6279
dispop3	-0.8576	0.2218	1.9722	dis4	-0.8069	0.3394	1.7305
dis5	-1.2311	0.0368	1.5027	disced1	-1.2152	-0.0223	1.1814
sfxinfm	-0.8403	0.1615	1.6582	sftpvm1	-0.7551	0.3538	1.7712
dmimintv	-1.1011	0.1016	1.7030	macregac	-1.0844	0.1039	1.3749
sftpmrev	-1.0526	0.0786	1.4125	macctott	-0.8154	0.3154	1.7176
macctotr	-0.8055	0.3190	2.2769	sftprind	-0.8613	0.2237	1.5865
sftgsngr	-0.7634	0.0969	1.1077	sftgsngr	-1.2802	-0.0527	1.1222

수와 중복된 관측치를 가진 예측변수를 제거한 후 $p = 100$ 개의 예측변수를 포함한 $n = 19$ 개의 자료로 분석하였다.

Figure 4.1는 라플라스 사전분포를 사용한 Lasso와 EN의 회귀모수(β) 중 각 모형의 추정치 중 절대값이 가장 큰 10개의 예측변수의 회귀모수에 대한 사후분포의 80% 신뢰구간으로, LARS 알고리즘을 통해 얻은 추정치(Δ)를 표시하였다. 라플라스 사전분포를 사용하였을 때 별점모수들의 추정치는 Lasso의 경우 $\hat{\lambda} = 0.857$ 이고 EN의 경우 $\hat{\lambda}_1 = 5.152$ 와 $\hat{\lambda}_2 = 27.935$ 로 각 별점모수의 깃스 표본의 중앙값(median)으로 얻은 것이다. LARS 알고리즘을 통해 선택된 Lasso 모형의 변수는 **sftpmag**(magnitude of failure event), **sftpvm3**(revolutionary war magnitude), **sftpmmax**(maximum yearly magnitude score)의 세 가지 이지만, 이 세 변수는 Figure 4.1의 80% 신뢰구간에 포함되지 않았다. 그리고 LARS-EN 알고리즘을 통해 선택된 EN 모형에서는 Lasso LARS의 세 변수 **sftpmag**, **sftpvm3**와 **sftpmmax**를 포함한 총 43개의 변수를 선택하였다. 그리고 이것을 통해 베이지안 방법론과 LARS 알고리즘은 서로 다른 결과를 나타내고 있음을 확인하였다. LARS 알고리즘의 경우 별점회귀 선형모형의 계수추정법을 통해 Lasso의 경우 최대 n 의 예측계수까지 추정가능한 방법이지만, 베이지안 Lasso와 EN은 별점을 사전분포로 사용하여 회귀계수의 사후분포를 얻는 방법론이다. 그러나 LARS-EN 알고리즘을 통해 선택된 43개의 변수 중, **sftprind**(regime crisis indicator), **macctott**(total of the annual score for for Ethnic War, Geno/Politicide, Revolution and Colonial Wars), **sftpvm1**(revolutionary war magnitude), **dis4**(highest discrimination value for either economic discrimination or political discrimination)와 **dispop4**(population proportion- largest group)은 베이지안 EN방법을 통해 얻은 10개의 변수에도 포함되어 있었다. 그러므로 모형에 대한 차이도 생각할 수 있다.

Table 4.2는 라플라스 사전분포를 사용한 Lasso와 EN의 회귀모수(β) 중 각 모형의 추정치 중 절대값이 가장 큰 10개의 예측변수의 회귀모수에 대한 중앙값과 사후분포의 95% 신뢰구간이다. 베이지안 Lasso가 선택한 10개의 회귀계수와 베이지안 EN이 선택한 회귀계수 중 **sftgsngr**만 동시에 선택했고, 나머지 9개의 회귀계수는 모두 다르다. 별점모수들의 추정치는 Lasso의 경우 $\hat{\lambda} = 0.857$ 이고 EN의 경우 $\hat{\lambda}_1 = 5.152$ 와 $\hat{\lambda}_2 = 27.935$ 로 별점모수의 추정치가 큰 차이를 보이므로 선택한 예측변수들이 다르다고 생각할 수 있다. 즉 회귀계수의 사후분포에 영향을 미치는 별점모수들의 차이로 서로 다른 결과를 내었다 할 수 있다. 95% 신뢰구간에 모두 0값이 포함되어 있지만, LARS-EN의 결과와 베이지안 EN의 변수 선택의 결과 중 5개의 변수가 서로 중복된다. 그러므로 라플라스 사전분포를 사용한 베이지

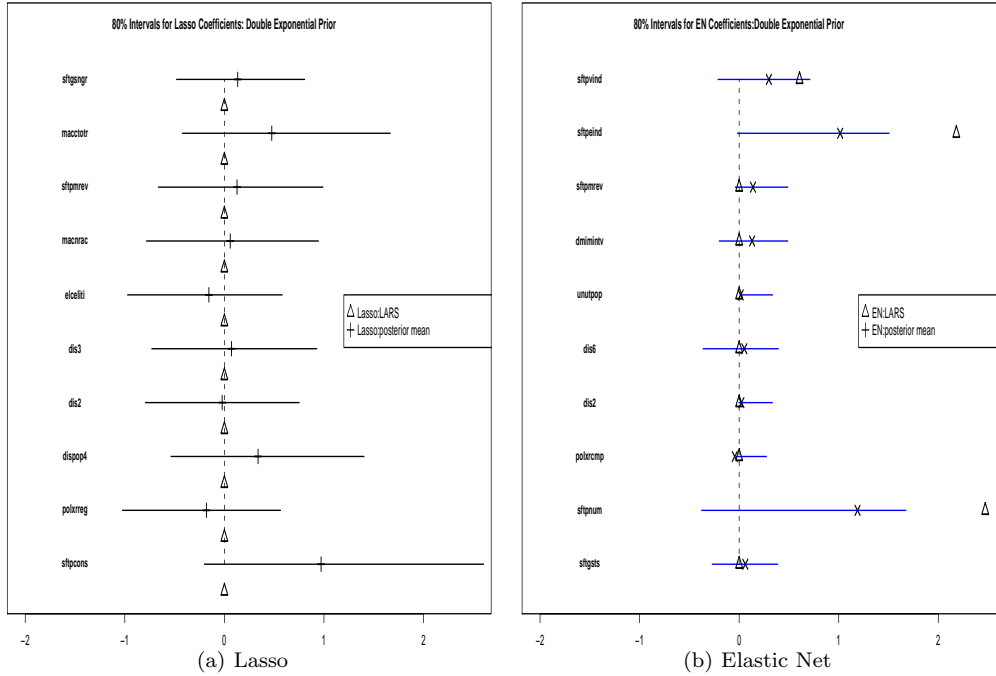


Figure 4.2. Top ten effects of posterior estimates (absolute value) by Bayesian Lasso and Bayesian EN with double Pareto prior. The solid line is the 80% posterior credible interval.

안 Lasso보다는 베이시안 EN이 LARS 알고리즘의 변수 선택의 결과를 포함하고 있다고 할 수 있다.

Figure 4.2는 이중 파레토 사전분포를 사용한 Lasso와 EN의 회귀모수(β) 중 각 모형의 추정치 중 절대값이 가장 큰 10개의 예측변수의 회귀모수에 대한 사후분포의 80% 신뢰구간이다. 기존 방법인 LARS와 LARS-EN 알고리즘을 통해 얻은 추정치(Δ)도 신뢰구간 위에 표시하였다. 이중 파레토 사전분포를 사용하였을 때 별점모수들의 추정치는 Lasso의 경우 $\hat{\lambda} = 0.950$ 이고 EN의 경우 $\hat{\lambda}_1 = 2.027$ 과 $\hat{\lambda}_2 = 1.895$ 이다. LARS 알고리즘에 의해 선택된 세 변수는 Figure 4.2의 80% 신뢰구간에 포함되지 않았다. 그러나 LARS-EN 알고리즘을 통해 선택된 43개의 변수 중 **sftpvind**(revolutionary war indicator)와 **sftpeind**(ethnic war indicator), 그리고 **sftpnun**(event no. in each country)의 세 변수가 이중 파레토 사전분포를 사용한 EN모형으로 선택한 10개의 예측변수에 포함되어 있다.

Table 4.3은 이중 파레토 사전분포를 사용한 Lasso와 EN의 회귀모수(β) 중 각 모형의 추정치 중 절대값이 가장 큰 10개의 예측변수의 회귀모수에 대한 중앙값과 사후분포의 95% 신뢰구간이다. 베이시안 Lasso가 선택한 10개의 회귀계수와 베이시안 EN이 선택한 회귀계수 중 **sftpmrev**만 동시에 선택 했고, 나머지 9개의 회귀계수는 모두 다르다. 별점모수들의 추정치는 Lasso와 EN의 경우 값에 큰 차이가 없었지만, 선택된 10개의 예측변수는 달랐다. 베이시안 EN의 경우 변수 **sftpnun**와 **sftpeind**의 95% 사후 신뢰구간에는 0값이 포함되어 있지 않았고, 이 변수들은 LARS-EN을 통해서도 선택된 변수들이다. LARS 알고리즘에 의해 회귀계수가 0으로 추정된 예측변수들의 경우, 수리적으로는 모형에 영향력이 없지만 모형에 포함되었을 때 모형 전체를 설명하거나 반응변수를 설명하는 개연성을 증가시키는 역할을 할 수 있다. 그러므로 베이시안 EN의 경우 모형에 매우 큰 의미가 있는 회귀계수들은 95% 신뢰구간

Table 4.3. 95% posterior credible intervals of the top ten effects of posterior estimates (absolute value) by Bayesian Lasso and Bayesian EN with double Pareto prior.

Variable	Bayesian Lasso Quantiles			Variable	Bayesian EN Quantiles		
	0.05	0.50	0.95		0.05	0.50	0.95
sftpcons	-0.4566	0.6917	3.3216	sftgsts	-0.3414	0.0470	0.5044
polxrreg	-1.3785	-0.1121	0.8433	sftpnum	0.5641	1.1895	1.8123
dispop4	-0.8555	0.2087	1.9097	polxrcmp	-0.4974	-0.0335	0.3865
dis2	-1.1252	-0.0160	1.0818	dis2	-0.3904	0.0183	0.4434
dis3	-1.0872	0.0399	1.2758	dis6	-0.3726	0.0320	0.5260
elceliti	-1.3170	-0.1000	0.8521	unutpop	-0.4111	0.0083	0.4453
macnrac	-1.1285	0.0244	1.3614	dmimintv	-0.2916	0.1055	0.6297
sftpmrev	-1.0012	0.0805	1.3580	sftpmrev	-0.2568	0.1092	0.6147
macctotr	-0.7104	0.2929	2.1883	sftpeind	0.3771	1.0172	1.6428
sftgsngr	-0.7213	0.1013	1.0599	sftpvind	-0.1375	0.2603	0.8504

을 통해서 찾을 수 있고, 인과관계를 설명하는 다른 모수들도 포함하는 효과가 있다.

Table 4.2과 Table 4.3를 비교해보면, 라플라스 사전분포를 사용한 Lasso와 이중 파레토 사전분포를 사용한 Lasso의 각각의 10개의 변수들 중 sftpmrev, macctotr와 sftgsngr는 두 경우 모두 선택되었고 추정치의 값들도 크게 다르지 않다. 이 세 변수는 다른 사전분포를 사용하여도 베이저안 Lasso방법에서는 중요한 의미를 갖는 변수들이며, 반응변수를 설명하는데에도 중요한 변수라 할 수 있다.

라플라스 사전분포를 사용한 EN과 이중 파레토 사전분포를 사용한 EN의 절대값이 가장 큰 10개의 변수들을 살펴보면, polxrcmp는 두 경우에서 모두 선택되었다. 그러나 변수들 간의 다중공선성이 고려된 EN 모형의 경우 영향력 있는 변수를 찾을 수 있고 인과관계를 설명하는 다른 모수들을 포함하고 있어 모형적인 방법론으로나 분석된 자료를 설명하는데 더 적합하다 하겠다. 사회과학 자료들을 분석하면 많은 변수들이 이론적 근거에 의해 중요한 의미를 갖지만, 통계적 예측의 면에서는 큰 도움이 되지 않는 경우가 많다. 그러므로 위의 실험 결과에서 나타난 바와 같이 변수들 간의 상관관계도 포함되는 베이저안 EN 모형의 결과가 영향력의 값은 작아보이지만 보다 많은 내용을 설명할 수 있다.

5. 결론

이 논문에서는 선형모형에 L_1 벌점을 사용한 Lasso와 L_1 과 L_2 벌점을 같이 사용한 EN에 대해 깃스 표본법을 이용한 베이저안 분석에 대해 논술하고, 정규분포의 척도모수에 분포를 가정하여 섞는 정규혼합모형을 통해 라플라스 분포와 이중 파레토 분포를 적용한 공액계층모형의 완전 조건 사후분포들을 도출하였다.

점추정값을 찾는 LARS 알고리즘에 비해 라플라스 분포와 이중 파레토 분포를 적용했을 때, 사후분포로부터 의미있는 예측변수와 회귀계수들의 사후 신뢰구간을 구할 수 있었다. 여기에서 Lasso의 추정값이 0인 회귀계수에 대해 유효한 신뢰구간을 계산할 수 없다는 단점에 대한 대응책과 0 값으로 추정된 회귀계수를 얼마나 확신할 수 있는지에 대한 확인이 가능하다는 것이 본 연구의 성과이다.

그리고, 일반 Lasso모형에 이중 파레토 분포를 적용하면 라플라스 분포를 적용한 것보다 희박성(sparsity)의 크기의 차이 때문에 서로 다른 모수를 선택하는 것을 확인할 수 있었다. 벌점모수인 λ 의 추정치가 라플라스 분포와 이중 파레토 분포의 경우 크게 다르지 않았지만, 모형의 차이로 인해 선택되는 모수에 차이가 있었다. EN의 경우도 마찬가지로 라플라스 분포를 사용했을 때와 이중 파레토 분포를 사용했을 때의 회귀계수에 대한 추정치가 매우 달랐다. 그러나 많은 변수들이 통계적 예측 면에서는 큰 도움

이 되지 않지만 이론적 근거에 의해 중요한 의미를 갖고 변수들 사이의 관계성도 중요한 경우 라플라스 사전분포를 사용한 모형보다는 이중 파레토 사전분포를 사용한 모형이 더 합리적인 모형을 선택함을 알 수 있었다.

제안된 방법론은 다양한 사전분포를 사용하여 비교 가능하다. 라플라스나 이중 파레토 분포 외에 희박성(sparsity)이 더 큰 사전분포를 사용한다면, 선택되는 변수나 계산된 추정값은 달라질 것이다. 그리고 추후 fused lasso나 grouped lasso의 별점에 대해 라플라스 사전분포 대신 이중 파레토 분포를 사용하는 방법도 새로운 연구주제가 될 수 있을 것이다.

References

- Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions, *Journal of the Royal Statistical Society, Series B*, **36**, 99–102.
- Armagan, A., Dunson, D. B. and Lee, J. (2013). Generalized double Pareto shrinkage, *Statistica Sinica*, **23**, 119–143.
- Bae, K. and Mallick, B. K. (2004). Gene selection using a two-level hierarchical Bayesian model, *Bioinformatics*, **20**, 3423–3430.
- Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression, *The Annals of Statistics*, **32**, 407–451.
- Esty, D. C., Goldstone, J. A., Gurr, T. R., Harff, B., Levy, M., Dabelko, G. D., Surko, P. T. and Unger, A. N. (1999). *State Failure Task Force Report: Phase II Findings Environmental Change & Security Project Report 5*, Summer.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**, 1348–1360.
- Figueiredo, M. A. T. (2003). Adaptive sparseness for supervised learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25** 1150–1159.
- Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools, *Technometrics*, **35**, 109–135.
- Hobert, J. P. and Geyer, C. J. (1998). Geometric ergodicity of Gibbs and Block Gibbs samplers for a hierarchical random effects model, *Journal of Multivariate Analysis*, **67** 414–430.
- Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Applications to nonorthogonal problems, *Technometrics*, **12**, 55–68.
- Kim, Y., Kim, J. and Kim, Y. (2006). Blockwise sparse regression, *Statistica Sinica*, **16**, 375–390.
- Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators, *The Annals of Statistics*, **28**, 1356–1378.
- Kyung, M., Gill, J., Ghosh, M. and Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos, *Bayesian Analysis*, **5**, 369–412.
- Liu, J. S., Wong, W. H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes, *Biometrika*, **81**, 27–40.
- Osborne, M. R., Presnell, B. and Turlach, B. A. (2000). A new approach to variable selection in least squares problems, *IMA Journal of Numerical Analysis*, **20**, 389–404.
- Park, T. and Casella, G. (2008). The Bayesian lasso, *Journal of the American Statistical Association*, **103**, 681–686.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso, *Journal of the Royal Statistical Society, Series B*, **67**, 91–108.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society, Series B*, **68**, 49–67.
- Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society, Series B*, **67**, 301–320.

라플라스와 이중 파레토 벌점의 비교: LASSO 와 Elastic Net

경민정^{a,1}

^a덕성여자대학교 정보통계학과

(2014년 9월 22일 접수, 2014년 10월 14일 수정, 2014년 10월 21일 채택)

요약

연속적인 변수 선택과 계수 추정을 동시에 활용할 수 있다는 특성 때문에 LASSO (Tibshirani, 1996)와 Elastic Net (Zou와 Hastie, 2005)은 다양한 분야에서 활발하게 사용되고 있다. 조건부 라플라스와 이중 파레토 사전분포를 적용한 공액계층모형을 표현하였고, 각각의 사전분포에 대한 완전 조건 사후분포를 도출하였다. 제안된 사전분포를 적용한 벌점회귀모형을 비교하기 위한 모의 실험을 진행하였고, 예측정확도를 판단하기 위해 아시아 국가 실패(the collapse of governments in Asia)의 실제 데이터에 제안한 모형을 적용하였다.

주요어: LASSO, Elastic net, 계층 모형, 정규분포의 척도 혼합물 모형, 라플라스 사전분포, 이중 파레토 사전분포.

이 논문은 덕성여자대학교 교내연구비 3000001920 지원을 받아 수행되었습니다.

¹(132-714) 서울시 도봉구 삼양로 144길 33, 덕성여자대학교 정보통계학과. E-mail: mkyung@duksung.ac.kr