Analysis of the High-Dimensional Spatial Autoregressive Models with Bayesian Regularization

Seyede Narges Malaeke

Mohsen Mohammadzadeh

Tarbiat Modares University

4th Seminar on Spatial Statistics and Its Applications

n.malaeke@modares.ac.ir

Contents

1 Introduction

2 Bayesian Hierachical

3 SAR and MESS

4 Shrinkage Priors

5 Simulation Study

n.malaeke@modares.ac.ir

Introduction

High Dimensional

That is: **p>n**

- **Over parametrization**
- **Complexity**

Spatial Auto Regressive

Bayesian Model Averaging(Lesage and Pace(2007)

Bayesian Approach

Prior Information

n.malaeke@modares.ac.ir

Bayesian Hierachical

Posterior

$$\pi(\beta,\sigma^2|Y) = \frac{\pi(Y|\beta,\sigma^2)\pi(\beta,\sigma^2)}{\int \pi(Y|\beta,\sigma^2)\pi(\beta,\sigma^2)d(\beta,\sigma^2)}$$

- $\boldsymbol{Y|\beta,\sigma^2}$
- **Data**
- **Posterior**
- **Posterior Log**

Log

$$\log\pi(\beta,\sigma^2|Y,X) \propto \log\pi(Y|X\beta,\sigma^2) + \log\pi(\beta|\sigma^2,\lambda) + \log\pi(\sigma^2) + \log\pi(\lambda)$$

- **Hyperparameters Controls Complexity**
- **Maximum a Posterior Estimator**
- $\boldsymbol{\log\pi(\beta|\sigma^2,\lambda)}$
- **Shrinkage Priors**
- **Spike and Slab Priors**

n.malaeke@modares.ac.ir

Bayesian Hierachical

Advantages

01 Credible Interval by MCMC

02 Reduce Complexity of Estimate by Prior Information

03 Computation Power

04 Estimate Regularization Parameter

01 Gibbs Variable Selection

02 Bayesian Model Averaging

03 Search Stochastic Variable Selection

n.malaeke@modares.ac.ir

Spatial Autoregressive Model

Matrix Exponential Spatial Specification

SAR

$$Y = \alpha WY + X\beta + \epsilon, \qquad \epsilon \sim N(0, \sigma^2 I_n), \qquad W_{n\times n}, \qquad |\alpha| < 1$$

MESS

$$S(\alpha) = I - \alpha W, \qquad S(\alpha)Y = X\beta + \epsilon$$

$$S(\rho) = \exp(\rho W), \qquad \rho \in (-\infty, \infty), \qquad S(\rho)Y = X\beta + \epsilon$$

Lesage & Pace(2007)

Advantages

$S(\rho)$: NonSingular

$$S(\rho)^{-1} = \exp(\rho W)^{-1} = \exp(-\rho W)$$

$$\det(\exp(\rho W)) = \exp(tr(\rho W))$$

Correspondence

$$\alpha \approx 1 - \exp(\rho)$$

Lesage & Pace(2007)

Global-Local Shrinkage Priors

$$S(\rho)Y = X\beta + \epsilon\,, \qquad \epsilon \sim N(0,\Omega)$$

$$f\left(Y \middle| X\beta, \sigma^2, \rho\right) = (2\pi)^{-\frac{n}{2}} \det(\Omega)^{-\frac{1}{2}} \exp\left(-\frac{e^{T}\Omega^{-1}e}{2}\right)$$

$$\beta_j|\psi_j \sim N\left(0, \psi_j\right)$$

$$\psi_j \sim D$$

Griffin & Brown(2010)

More hierachical layers:

More flexibility and adaptive shrinkage

Pfarrhofer & Piribauer (2019)

None

$$\underline{\beta} = 0 \,, \underline{\Sigma} = 1000\, \boldsymbol{I}_p$$

$$\rho \sim N(0, c), \qquad c = 10$$

Prior

$$\boldsymbol{\beta} | \cdot \sim N\left(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\Sigma}}\right),$$

$$\bar{\boldsymbol{\Sigma}} = \left(\underline{\Sigma}^{-1} + \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\Omega}^{-1} \boldsymbol{X}\right)^{-1},$$

$$\bar{\boldsymbol{\beta}} = \bar{\Sigma}(\underline{\Sigma}^{-1} \underline{\boldsymbol{\beta}} + \boldsymbol{X}^{T} \boldsymbol{\Omega}^{-1} S(\rho) Y)$$

$$\rho | \cdot \propto \exp\left(-\frac{e^{T}\Omega^{-1}e}{2}\right) \pi(\rho)$$

$$\min[1, \frac{g(\rho^{*}|\cdot)}{g(\rho_{t-1}|\cdot)}]$$

$$\rho^{*} \sim N(\rho_{t-1}, \tau)$$
$$\rho_t = \rho^{*}$$

Posterior

n.malaeke@modares.ac.ir

Normal-Gamma Shrinkage

Griffin & Brown (2010)

Prior

$\beta_j|\psi_j \sim N\left(0,2\lambda^{-2}\psi_j\right),$

$\psi_j \sim G(\theta,\theta),$

$\lambda^2 \sim G(d_0,d_1)$

$\psi_j|\lambda,\beta_j \sim GIG(\theta-\frac{1}{2},\beta_j^2,\theta\lambda^2)$

$\lambda^2|\psi \sim G\ (d_0+\theta P, d_1+2^{-1}\theta\sum\psi_j)$

$\text{diag}\left(\underline{\boldsymbol{\Sigma}}\right)=\boldsymbol{\psi}$

Posterior

n.malaeke@modares.ac.ir

Dirichlet-Laplace Shrinkage

Bhattacharya & et al (2015)

Prior

$$\beta_j \sim N\left(0, \varphi_j \phi_j^2 \tau^2\right),$$

$$\varphi_j \sim \exp\left(\frac{1}{2}\right),$$

$$\phi \sim \mathrm{Dir}(a, \dots, a),$$

$$\tau \sim G\left(na, \frac{1}{2}\right)$$

$$\mu_j = \frac{\phi_j \tau}{|\beta_j|},$$

$$\tilde{\varphi}_j | \phi, \beta \sim \mathrm{GIG}\left(-\frac{1}{2}, 1, \mu_j^{-2}\right), \varphi_j = \frac{1}{\tilde{\varphi}_j}$$

$$\tau | \phi, \beta \sim \mathrm{GIG}\left(1 - p, 2\sum_{j=1}^{p} \frac{|\beta_j|}{\phi_j}, 1\right)$$

$$T_j \sim \mathrm{GIG}\left(a - 1, 2|\beta_j|, 1\right), \ \phi_j = \frac{T_j}{\sum_{j=1}^{P} T_j}$$

$$\mathrm{diag}(\underline{\boldsymbol{\Sigma}}) = \left(\varphi_1 \phi_1^2 \tau^2, \dots, \varphi_p \phi_p^2 \tau^2\right)^T$$

Posterior

n.malaeke@modares.ac.ir

Simulation

- $\mathbf{n} = \mathbf{100}$
  $\mathbf{p} = \mathbf{200}$

- $\boldsymbol{X_{n\times p}} \sim \boldsymbol{N_p(0,1)}$
  $\boldsymbol{W}$: **5-Nearest Neighbors**
  $\boldsymbol{\sigma^2 = 1}$

- $\boldsymbol{\beta_{p\times 1}}$: $\boldsymbol{\beta_{1:5} \sim N(0,5), \beta_{6:10} \sim N(0,1),\ \beta_{11:p} = 0}$
  $\boldsymbol{\rho \sim N(0,3)}$
  $\boldsymbol{Y = \exp(-\rho W)(X\beta + \epsilon),\ \epsilon \sim N(0, \sigma^2 I_n)}$

- $\mathbf{RMSE}(\hat{\boldsymbol{\theta}}) = \left(\sum_{j=1}^{M} (\hat{\boldsymbol{\theta}}_j - \tilde{\boldsymbol{\theta}}_j)^2 / M\right)^{1/2}$

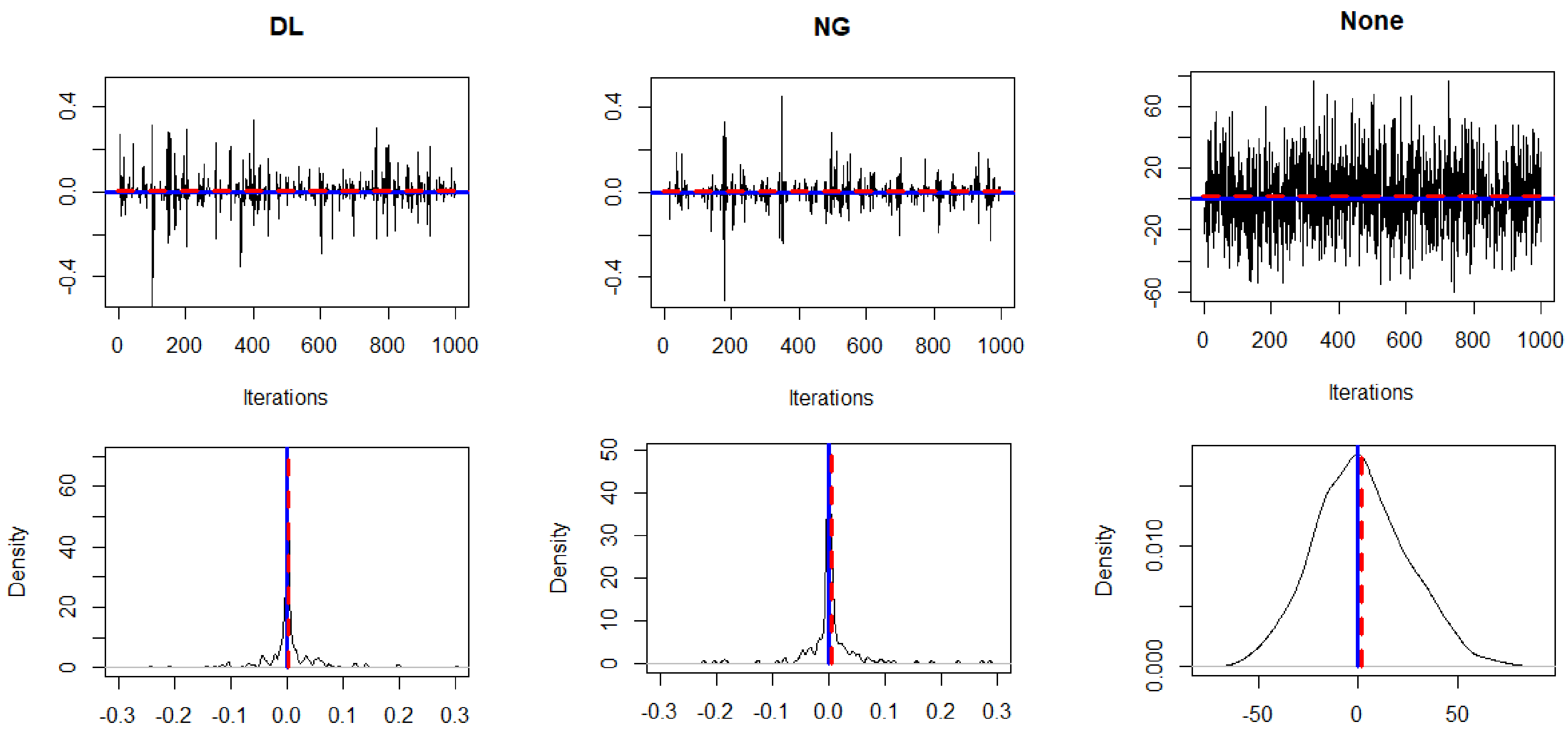
- **Hyperparameters**
  $\theta = 0.1, \qquad d_0 = d_1 = 0.01$
  $a = 1/p$

- **MCMC Algorithm**
  - Input initial value: θ_0
  - Generate random value: full conditional
  - Number iteration= 2000
  - Burn-in= 1000
  - ACF

n.malaeke@modares.ac.ir

N=100,P=200

Dirichlet-Laplace	Normal-Gamma	None



Simulation Results

	P=200		
	RMSE_{β}	RMSE_{ρ}	Time(min.)
Dirichlet-Laplace	**0.22**	**0.66**	**31.29**
Normal-Gamma	**0.29**	**0.66**	**31.78**
None	**0.39**	**0.94**	**30.59**

Disadvantages and Future Works

- Estimate hyperparameters by Monte Carlo EM algorithm
- Define another shrinkage prior and compare results
- Estimate posterior density
- Reduce costs (time)

References

- Bhattacharya, A., Patti, D., Pillai, N. S., and Dunson, D. B. (2015), Dirichlet–Laplace Priors for Optimal Shrinkage, *Journal of the American Statistical Association*, 110, 1479-1490.

- Buhlman, P., and Van De Geer, S. (2011), Statistics for High-Dimensional Data: Methods, theory and Applications. Springer Science and Business Media.

- Buhlman, P., Kalisch, M., and Meier, L. (2014), High-Dimensional Statistics with a View Toward Applications in Biology.

- Griffin, J. E., and Brown, P. J. (2010), Inference with Normal-Gamma Prior Distributions in Regression Problems, *Bayesian Analysis*, 5, 171-188.

- Giraud, C. (2014), Introduction to High-dimensional Statistics (Vol. 138), CRC Press.

- Himel Mallick, N. Y. (2013), Bayesian Methods for High Dimensional Linear Models, *Biostatistics*, 1, 005.

- Pfarrhofer, M., and Piribauer, P. (2019), Flexible Shrinkage in High-Dimensional Bayesian Spatial Autoregressive Models, *Spatial Statistics*, 29, 109-128.

- Park, T., and Casella, G. (2008), The Bayesian Lasso, *Journal of the American Statistical Association*, 103, 681-686.

THANKS

FOR

YOUR ATTENTION

n.malaeke@modares.ac.ir
mohsen_m@modares.ac.ir

@nargesmalaeke