Bayesian Variable Selection for Poisson Regression with Spatially Varying Coefficients

Jinjian Mu
University of Connecticut
Department of Statistics
215 Glenbrook Rd., Storrs, CT
USA
jinjian.mu@uconn.edu

Guanyu Hu
University of Connecticut
Department of Statistics
215 Glenbrook Rd., Storrs, CT
USA
guanyu.hu@uconn.edu

Lynn Kuo
University of Connecticut
Department of Statistics
215 Glenbrook Rd., Storrs, CT
USA
lynn.kuo@uconn.edu

Abstract: Poisson regression model is commonly used to model count data. In many scenarios, the data are collected from various locations so spatially varying coefficient Poisson regression model is developed to adjust for spatial dependence. We propose a Bayesian variable selection method for Poisson regression model with spatially varying coefficients. Considering computation efficiency we assign a conjugate multivariate log-gamma (MLG) prior to the regression coefficients and further incorporate the spatial information into the covariance matrix. We apply the horseshoe prior to facilitate a robust variable selection method with computational efficiency and build a MCMC algorithm for the posterior inference.

Key–Words: Spatial Poisson Model, Bayesian Variable Selection, MLG Prior, Horseshoe Prior, MCMC

1 Introduction

Poisson regression model is widely used for count data. Sometimes the data are collected from various locations such as the number of people died last year in every country, or the number of accidents that happened last month in every state in US. In this case, Poisson regression model needs to be adjusted for spatial variation thus Poisson regression model with spatially varying coefficients is developed.

The spatially varying coefficient model was first introduced by [5], and then was extended to different kinds of regression models. For spatially varying coefficient Poisson regression model, latent Gaussian process is used to model the dependency such as in [4]. However, the computation is very expensive for Gaussian process model, so [1] proposed a computationally efficient model by developing a multivariate log-gamma distribution (MLG) as the prior for Poisson regression coefficients.

In this paper, we propose a Bayesian variable selection method for Poisson regression model with spatially varying coefficients. We adopt the multivariate log-gamma distribution (MLG) as the prior for regression coefficients to achieve high computational efficiency. In Bayesian framework, the usual setting is to assign a normal prior for the regression coefficient, and we can approximate the multivariate normal distribution using the MLG distribution and further incorporate the spatial information into the corresponding covariance matrix. The horseshoe prior [2, 3] is used to do variable selection. We build a MCMC algorithm for our proposed model.

The rest of the paper is organized as follows. In section 2, we introduce the proposed model with its hierarchical structure. Section 3 shows the Bayesian computing algorithm. Some discussions about our model are in section 4.

2 Variable Selection for Poisson Regression Model with Spatially Varying Coefficients

Suppose there are n sites, p predictors and m_i observations in site i, $i = 1, \dots, n$. Then for $y_j(s_i)$, the count of j-th observation in site i, the Poisson regression model with spatially varying coefficients can be described as

$$y_j(s_i) \sim \text{Poisson}(\exp\{\boldsymbol{x}_j'(s_i)\boldsymbol{\beta}(s_i)\}) \quad (1)$$

where $\boldsymbol{x}_j(s_i) = (x_{j1}(s_i), \dots, x_{jp}(s_i))'$ is the corresponding covariate vector and $\boldsymbol{\beta}(s_i) = (\beta_1(s_i), \dots, \beta_p(s_i))'$ is the p-dimensional regression coefficients. We can fix the k-th regression coefficient across all locations and combine them as $\boldsymbol{\beta}_k = (\beta_k(s_1), \dots, \beta_k(s_n))', k = 1, \dots, p$. Then let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \dots, \boldsymbol{\beta}_p')'$, and we can write the model in matrix form as:

$$\boldsymbol{y} \sim \text{Poisson}(\exp\{\boldsymbol{X}\boldsymbol{\beta}\}) \quad (2)$$

where $\boldsymbol{y} = (y_1(s_1), \ldots, y_{m_1}(s_1), \ldots, y_1(s_n), \ldots, y_{m_n}(s_n))'$ is the outcome vector for all the observations in all sites, and matrix $\boldsymbol{X}$ contains all the covariates. See the format of $\boldsymbol{X}$ in [6].

Multivariate log-gamma (MLG) distribution is developed by [1] and it can be a conjugate prior for the regression coefficients of a multidimensional Poisson regression model. If $\boldsymbol{q} = \boldsymbol{c} + \boldsymbol{V}\boldsymbol{w}$, where $\boldsymbol{c} \in \mathbb{R}^m$, $\boldsymbol{V}$ is a $m \times m$ invertible matrix, and $\boldsymbol{w} = (w_1, \ldots, w_m)'$ are m mutually independent log-gamma random variables with $w_i \sim \text{LG}(\alpha_i, \kappa_i), i = 1, \ldots, m$, then we have $\boldsymbol{q} \sim \text{MLG}(\boldsymbol{c}, \boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)'$ and $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_m)'$. If $\boldsymbol{q} \sim \text{MLG}(\boldsymbol{c}, \alpha^{1/2}\boldsymbol{V}, \alpha\mathbf{1}, \alpha\mathbf{1})$, then $\boldsymbol{q}$ converges in distribution to a multivariate normal random vector with mean $\boldsymbol{c}$ and covariance matrix $\boldsymbol{V}\boldsymbol{V}'$ as α goes to infinity.

Based on the approximately normal property of the MLG distribution, it is a natural idea to assign an approximately normal MLG prior to the Poisson regression coefficients $\boldsymbol{\beta}$, which on the one hand is concordant to the usual normal setting in the Bayesian analysis, on the other hand keeps the conjugacy of the prior. In this way, $\boldsymbol{\beta}$ is specified to have a MLG prior, that is,

$$\boldsymbol{\beta} \sim \text{MLG}(\mathbf{0}_{np}, \alpha^{1/2}\boldsymbol{\Sigma}^{1/2}, \alpha\mathbf{1}_{np}, \alpha\mathbf{1}_{np}), \tag{3}$$

where $\boldsymbol{\Sigma}^{1/2}$ is the square root of a positive definite matrix $\boldsymbol{\Sigma}$. Here we choose α as 10000, thus $\boldsymbol{\beta} \sim \text{MVN}(\mathbf{0}_{np}, \boldsymbol{\Sigma})$ approximately.

Since $\boldsymbol{\Sigma}$ can be regarded as the covariance matrix of the approximately normal distribution, we can incorporate the spatial information by a spatial correlation matrix and employ the horseshoe prior [2, 3] for variable selection. For unknown sparsity patterns and large outlying signals, the horseshoe prior gives more robust results, so we apply the horseshoe prior here to do variable selection. The covariance matrix $\boldsymbol{V}$ can be expressed as

$$\begin{aligned}
\boldsymbol{\Sigma} &= \sigma^2\tau^2\text{diag}(\lambda_1^2, \ldots, \lambda_p^2) \otimes \boldsymbol{H} \\
1/\sigma^2 &\sim \text{Gamma}(3/2, 3/2) \\
\tau &\sim \text{Ca}^+(0, 1) \\
\lambda_k &\sim \text{Ca}^+(0, 1)
\end{aligned} \tag{4}$$

where σ^2 is the variance parameter following an inverse-gamma distribution with shape 3/2 and scale 3/2, τ is the global shrinkage parameter, $\lambda_k, k = 1, \ldots, p$ are the local shrinkage parameters allowing for local variation of the shrinkage, $\otimes$ is the Kronecker product, and $\boldsymbol{H}$ is the spatial correlation matrix with the (l, m)-th entry as $\exp\{-\frac{1}{b} \times dist(s_l, s_m)\}$ where $dist(s_l, s_m)$ is some kind of distance between site l and site m and b is a tuning parameter. To make $\boldsymbol{H}$ a valid spatial correlation matrix, we assign a uniform prior on b, $b \sim \text{Unif}(0, B)$ where B is the upper bound of b and depends on the maximum entry of the distance matrix. Combining the specification of $\boldsymbol{H}$ with 2, 3, and 4, the hierarchical structure of our model can be described as,

$$\begin{aligned}
\boldsymbol{y} &\sim \text{Poisson}(\exp\{\boldsymbol{X}\boldsymbol{\beta}\}) \\
\boldsymbol{\beta} &\sim \text{MLG}(\mathbf{0}_{np}, \alpha^{1/2}\boldsymbol{\Sigma}^{1/2}, \alpha\mathbf{1}_{np}, \alpha\mathbf{1}_{np}) \\
\boldsymbol{\Sigma} &= \sigma^2\tau^2\text{diag}(\lambda_1^2, \ldots, \lambda_p^2) \otimes \boldsymbol{H} \\
1/\sigma^2 &\sim \text{Gamma}(3/2, 3/2) \\
\tau &\sim \text{Ca}^+(0, 1) \\
\lambda_k &\sim \text{Ca}^+(0, 1) \\
\boldsymbol{H}_{(l,m)} &= \exp\{-\frac{1}{b} \times dist(s_l, s_m)\} \\
b &\sim \text{Unif}(0, B)
\end{aligned} \tag{5}$$

From 5, we can make posterior inference based on local shrinkage parameters λ_k's. According to [2, 3], if the posterior mean of λ_k is less than 1, we can decide that the corresponding k-th predictor is not significant.

3 Bayesian Computation

We build a MCMC algorithm for our proposed model to obtain posterior samples. From the joint posterior distribution, we can derive full conditional posterior distributions for all the parameters. Since the MLG prior on $\boldsymbol{\beta}$ is conjugate, we can derive a cMLG distribution as the full conditional posterior distribution for $\boldsymbol{\beta}$ and Gibbs sampler can be applied here. For other parameters, the full conditional posterior distributions are not in standard forms, so we use Metropolis-Hastings algorithms to draw posterior samples. The detailed MCMC algorithm is described as follows. To simplify the notation, denote $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)'$.

- **Step 1.** Start with initial values $\boldsymbol{\beta}^{(0)}, \sigma^{(0)}, \tau^{(0)}, \boldsymbol{\lambda}^{(0)}, b^{(0)}$. Let t be the iteration index and set $t = 1$.

- **Step 2.** Draw samples of $\boldsymbol{\beta}$ from the cMLG distribution

$$\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \sigma, \tau, \boldsymbol{\lambda}, b \sim \text{cMLG}(\boldsymbol{H}_\beta, \boldsymbol{\alpha}_\beta, \boldsymbol{\kappa}_\beta) \tag{6}$$

where $\boldsymbol{H}_\beta = (\boldsymbol{X}', \alpha^{-1/2}\boldsymbol{\Sigma}^{-1/2})', \boldsymbol{\alpha}_\beta = (\boldsymbol{y}', \alpha\mathbf{1}'_{np})', \boldsymbol{\kappa}_\beta = (\mathbf{1}'_N, \alpha\mathbf{1}'_{np})', N = \sum_i^n m_i$. $\boldsymbol{y}$ can contain 0's since they are count values, but $\boldsymbol{\alpha}_\beta$ corresponds to the shape parameters of gamma distributions thus can't be 0's. In this case, we can simply add a small value like 0.5 to $\boldsymbol{y}$.

- **Step 3.** Update σ using Metropolis-Hastings algorithm with proposal distribution Trunc-Normal$(\sigma^{(t-1)}, 0.1^2)$. The full conditional of σ is

$$\begin{aligned}&\pi(\sigma|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \tau, \boldsymbol{\lambda}, b) \propto \\ &\sigma^{-np-1} \exp\{\alpha^{1/2}\mathbf{1}'_{np}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\beta} \\ &- \alpha\mathbf{1}'_{np} \exp[\alpha^{-1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\beta}] - 2/3\sigma^{-2}\}\end{aligned} \quad (7)$$

  where $\boldsymbol{\Sigma} = \sigma^2\tau^2\text{diag}(\lambda_1^2, \ldots, \lambda_p^2) \otimes \boldsymbol{H}$ also involves σ.

- **Step 4.** Update τ using Metropolis-Hastings algorithm with proposal distribution Trunc-Normal$(\tau^{(t-1)}, 0.07^2)$. The full conditional of τ is

$$\begin{aligned}&\pi(\tau|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \sigma, \boldsymbol{\lambda}, b) \propto \\ &\tau^{-np}(1+\tau^2)^{-1} \exp\{\alpha^{1/2}\mathbf{1}'_{np}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\beta} - \\ &\alpha\mathbf{1}'_{np} \exp[\alpha^{-1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\beta}]\}\end{aligned} \quad (8)$$

  where $\boldsymbol{\Sigma} = \sigma^2\tau^2\text{diag}(\lambda_1^2, \ldots, \lambda_p^2) \otimes \boldsymbol{H}$ also involves τ.

- **Step 5.** Update $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)'$ independently and simultaneously using Metropolis-Hastings algorithm. For each $\lambda_k, k = 1, \ldots, p$, the corresponding proposal distribution is Trunc-Normal$(\lambda_k^{(t-1)}, 0.02^2)$. The full conditional distribution of $\boldsymbol{\lambda}$ is

$$\begin{aligned}&\pi(\boldsymbol{\lambda}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \sigma, \tau, b) \propto \\ &\left[\prod_k \lambda_k^{-n}(1+\lambda_k^2)^{-1}\right] \exp\{\alpha^{1/2}\mathbf{1}'_{np}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\beta} - \\ &\alpha\mathbf{1}'_{np} \exp[\alpha^{-1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\beta}]\}\end{aligned} \quad (9)$$

  where $\boldsymbol{\Sigma} = \sigma^2\tau^2\text{diag}(\lambda_1^2, \ldots, \lambda_p^2) \otimes \boldsymbol{H}$ also involves $\boldsymbol{\lambda}$.

- **Step 6.** Update b using Metropolis-Hastings algorithm with proposal distribution Normal$(b^{(t-1)}, 4000^2)$. The full conditional of b is

$$\begin{aligned}&\pi(b|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \sigma, \tau, \boldsymbol{\lambda}) \propto \\ &\det(\exp\{-\frac{1}{b}\boldsymbol{D}\})^{-p/2} \exp\{\alpha^{1/2}\mathbf{1}'_{np}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\beta} - \\ &\alpha\mathbf{1}'_{np} \exp[\alpha^{-1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\beta}]\}\end{aligned} \quad (10)$$

  where $\boldsymbol{D}$ is the distance matrix with the (l, m)-th entry as $dist(s_l, s_m)$ and $\boldsymbol{\Sigma} = \sigma^2\tau^2\text{diag}(\lambda_1^2, \ldots, \lambda_p^2) \otimes \boldsymbol{H}$ also involves b.

- **Step 7.** If $t < M$, repeat step 2 to step 5 and $t = t + 1$. M is the number of posterior samples we desire for posterior inference.

- **Step 8.** Calculate the posterior estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ using the posterior means.

4 Discussion

In this paper, we proposed a Bayesian variable selection method for Poisson regression model with spatially varying coefficients. For the conjugacy of the prior, we assigned a MLG prior to the regression coefficients. We combined the spatial correlation matrix and horseshoe prior for variable selection.

In the future work, we will conduct several simulation studies with multiple replicates and evaluate the variable selection performance using sensitivity and specificity. Different simulation scenarios will also be considered such as generating the regression coefficients from MLG instead of MVN and including some spatial constant coefficients. Five years of data (2010—2014) from the Office of the Connecticut Medical Examiner and the Connecticut Hospital Inpatient Discharge Database will be analyzed to further illustrate the proposed model.

References:

[1] J. R. Bradley, S. H. Holan, C. K. Wikle, et al. Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with discussion). *Bayesian Analysis*, 13(1):253–310, 2018.

[2] C. M. Carvalho, N. G. Polson, and J. G. Scott. Handling sparsity via the horseshoe. In *International Conference on Artificial Intelligence and Statistics*, pages 73–80, 2009.

[3] C. M. Carvalho, N. G. Polson, and J. G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010.

[4] P. J. Diggle, J. A. Tawn, and R. Moyeed. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350, 1998.

[5] A. E. Gelfand, H.-J. Kim, C. Sirmans, and S. Banerjee. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396, 2003.

[6] J. Mu, G. Hu, and L. Kuo. Bayesian variable selection for poisson regression with spatially varying coefficients with applications to connecticut adolescent suicide data. unpublished, N.D.