Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Tianyu Ding Date

Spatial Bayesian Variable Selection with Nonlocal Priors

By

Tianyu Ding

Master of Science of Public Health

Emory University

Rollins School of Public Health

Department of Biostatistics and Bioinformatics

Jian Kang

Committee Chair

Tianwei Yu

Committee Member

Spatial Bayesian Model Selection with Nonlocal Prior

By

Tianyu Ding

B.S., South China University of Technology, 2013

MPH, Emory University

Rollins School of Public Health

2015

Thesis Committee Chair: Jian Kang, PhD

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Department of Biostatistics and Bioinformatics

2015

Bayesian Spatial Variable Selection with Nonlocal Priors

By Tianyu Ding

A common objective of Bayesian spatial variable selection for regression models is to select important spatially distributed predictors that are strongly associated with the outcome. Many Bayesian variable selection procedures use local priors which do not have posterior variable selection consistency. We propose a novel spatial Bayesian model selection procedure based on nonlocal prior with incorporating spatial dependence of predictors. It show been shown that the Bayesian model selection procedure with nonlocal prior enjoys good theoretical properties and achieves better performance than existing methods. In this thesis, we show that incorporating the spatial dependence between the predictors can improve the nonlocal prior based Bayesian variable selection procedure in terms of both selection accuracy and prediction accuracy. We demonstrate the advantages of our method via simulation studies and an analysis of the brain imaging data from an Autism study.

Spatial Bayesian Model Selection with Nonlocal Prior

By

Tianyu Ding

B.S., South China University of Technology, 2013

MPH, Emory University

Rollins School of Public Health

2015

Thesis Committee Chair: Jian Kang, PhD

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Department of Biostatistics and Bioinformatics

2015

Acknowledgements

I want to thank the faculty, advisors, and staff of the Department of Biostatistics and Bioinformatics at Rollins School of Public Health for the dynamic two years of learning that I have had. This thesis is only a sample of the vast knowledge that was attained and applied through my two years here at Rollins. I would especially like to thank Dr. Jian Kang for all of his advice and support to help me write this thesis. Also a special thanks to Tianwe Yu for taking the time to read my thesis.

Lastly I want to give a special thanks to my parents for their loving support and encouragement that motivated me to pursue a degree at the graduate level.

Contents

1 Introduction

In most recent Bayesian model selection procedures, local prior densities which are positive at null parameter values have been applied frequently. However, local priors are not much competitive in many cases because sometimes a posterior probability of true model is too low to support it as a true model. There are some Bayesian model selection procedures local prior specifications based on the intrinsic Bayes factor (Berger and Pericchi 1996), the fractional Bayes factor (O'Hagan 1995) and g-priors (Liang et al. 2008) have this limitation.

Most Bayesian model selection methods with local priors usually do not report the posterior probability assigned to the true models obtained by their algorithms because sometimes, the probability is not strong enough to support the true models they find, especially in some high-dimension settings. In order to eliminate this deficiency, Johnson and Rossell (2012) proposed to use two nonlocal prior specifications on model parameters. Compared with local priors, nonlocal prior densities are density functions that are identically zero whenever a model parameter is equal to its null value, which is typically 0 in model selection settings. They also compare their non-local Bayesian model selection procedures with some usual frequentist methods including smoothly clipped absolute deviation(SCAD) algorithm (Fan and Li 2001), adaptive LASSO (Zou 2006). The results have shown that their procedure is more competitive: the posterior probability of true model is almost 1 and the proportion of correct model selections is almost 1 either.

In this thesis, we proposed to extend the Bayesian variable selection method with non local priors in high dimension settings by incorporating the spatial dependence to improve the selection accuracy. In some high dimension settings, the spatial dependence between predictors can facilitate the variable selection in the regression model. In some studies, Bayesian variable selection procedure performs well for spatial data

even with local priors (Lee et al. 2014).

Our goals of this paper are described as following aspects: first, we apply their Bayesian model selection procedure to spatial dependent data in high dimension settings. Second, we implement our model with consideration of spatial dependence based on their non-local prior model which will be introduced later. Motivation of our algorithm is the spatial distributions of the predictors in model. In their simulation study, they generate simulation data without considering the spatial dependence, if we take this factor into account, can we get the same good results? In order to test the efficiency and accuracy of their model in different settings, we repeat their algorithm and implement our new algorithm to do the comparison.

In general, our construction of model space is similar with their model: if we restrict the models in linear format, and let $Y_n = (y_1, \ldots, y_n)'$ denote a vector, $X_n = (x_1, \ldots, x_n)$ denote a $n \times p$ matrix, where x_i denote a $1 \times p$ vector for $1 \leq i \leq n$, β denote a $p \times 1$ regression coefficients vector with ith component denote as β_i. Then, we have the following linear model:

$$Y_n \sim N(X_n\beta, \sigma^2 I_n). \tag{1}$$

Definitions of two types of nonlocal prior of β are formula (2): product momoent(pMOM) densities and formula (3): product inverse moment (piMOM):

$$\pi(\beta|\tau, \sigma^2, r) = d_p(2\pi)^{-p/2}(\tau\sigma^2)^{-rp-p/2}|A_p|^{1/2} \times exp[-\frac{1}{2\tau\sigma^2}\beta' A_p\beta] \prod_{i=1}^{p} \beta^{2r} \tag{2}$$

$$piMOM : \pi(\beta|\tau, \sigma^2, r) = \frac{(\tau\sigma^2)^{rp/2}}{\Gamma(r/2)^p} \prod_{i=1}^{p} |\beta|^{-(r+1)} exp(-\frac{\tau\sigma^2}{\beta_i^2}) \tag{3}$$

These two densities are nonlocal densities at 0 because when any component of β is 0, they are identically 0. We also describe this property by using a univariate setting of this densities shown in Figure 1 (Johnson and Rossell 2012).

Considering performance of these two types of prior settings, we will use pMOM for our following analysis. The parameter settings will be specified in following section. The innovative part of our study is the calculation of d_p which is the spatial dependence distance of data . d_p is inverse of an central moment of multivariate normal distribution: $MVN(0, A_p)$ where A_p represent the positions' relationships of data. Because it is hard to calculate d_p based on their demonstration, they set A_p as identity matrix to calculate the normalizing constant distance d_p. However, in many spatial dependent data analysis, there will exist some errors of final model if we just set A_p it as identity matrix. The primary innovation of this article is to calculate d_p without just setting A_p as identity matrix.

In the next section, we will briefly go over the nonlocal prior model selection procedures and give detailed description of calculation of A_p and impose it into the model selection procedures. In section 3, we describe simulation algorithms to explore the performance of the new algorithm with new definition of A_p. In section 4, we will apply this new algorithm to real data analysis with brain imaging data.

2 The Model

2.1 Nonlocal Priors with Spatial Dependence

In this section, we review the nonlocal prior model selection procedure by Johnson and Rossell (2012) and we follow their denotations if there is no ambiguity. As we discussed in the introduction, we consider the linear model which is given by

$$Y_n \sim N(X_n\beta, \sigma^2 I_n). \tag{4}$$

Now we define one component of β will be excluded from the true model if its value is 0. We define a model $j = j_1, \ldots, j_k (1 \leq j_1 < \ldots < j_k \leq p)$ where $\beta_{j_1} \neq 0, \ldots, \beta_{j_k} \neq 0$ and all other components of β are equal to 0. We define model $k \subset j$ if all components of β in k are included in components of j. We define t is true model. There are p

components of β, leading to a total of 2^p potential models. Denote by $\mathcal{J}$ the total model combination. Denote by X_j the design matrix for model j which is formed from columns of X_n. Then under each model k, the sampling density for the data is defined as:

$$Y_n|\beta_k, \sigma^2 \sim N(X_k\beta_k, \sigma^2 I_n). \tag{5}$$

The properties of posterior probabilities can be studied based on the specification of σ^2. If σ^2 is known. The marginal density of the data with a pMOM prior density of model k is given by:

$$m_k(y_n) = d_k(2\pi)^{-n/2}\tau^{-k/2-rk}(\sigma^2)^{-n/2-rk}[\frac{|A_k|}{|C_k|}]^{1/2} \times exp[-\frac{R_k}{2\sigma^2}]E_k(\prod_{i=1}^{k}\beta_{k_i}^{2r}) \tag{6}$$

where

$$C_k = X_k{}'X_k + \frac{1}{\tau}A_k, \qquad \widetilde{\beta_k} = C_k^{-1}X_k{}'y_n, \qquad R_k = y_n{}'(I_n - X_kC_k^{-1}X_k{}')y_n,$$

and E_k represents the expectation of a multivariate normal distribution with mean $\widetilde{\beta_k}$ and covariance matrix $\sigma^2C_k^{-1}$.

If σ^2 is unknown, a common inverse gamma density with shape and scale parameters $(\alpha;\psi)$ is introduced for the value of σ^2. The marginal density of the data under model k is:

$$m_k(y_n) = d_k(2\pi)^{-n/2}2^{v/2}\tau^{-k/2-rk}[\frac{|A_k|}{|C_k|}]^{1/2} \times \frac{\psi^\alpha}{\Gamma(\alpha)}(v_k s_k^2)^{-v_k/2}\Gamma(\frac{v_k}{2})E_k^T(\prod_{i=1}^{k}\beta_{k_i}^{2r}) \tag{7}$$

where

$$v_k = n + 2rk + 2\alpha, \qquad s_k^2 = \frac{2\psi + R_k}{v_k},$$

$$d_k = [\int_{\mathcal{R}^k}(2\pi)^{-k/2}|A_k|^{1/2}exp(-\frac{1}{2}\gamma' A_k\gamma)\prod_{i=1}^{k}\gamma_i^{2r}d\gamma]^{-1}$$

In order to calculate the marginal likelihood function easily, the Laplace approximation to it under model k is:

$$\frac{\Gamma(\frac{v_k}{2})\psi^\alpha 2^{\frac{v_k}{2}}(2\psi + y'y - \widetilde{\beta_k}{}'C_k\widetilde{\beta_k})^{-\frac{v_k}{2}}}{\Gamma(\alpha)d_k(2\pi)^{\frac{n}{2}}\tau^{\frac{k}{2}+rk}} \times \frac{(\prod_{i\in k}(\beta_i^*)^{2r})exp\{-\frac{v_k-2}{2v_k}(\beta_k^* - \widetilde{\beta_k})'\frac{C_k}{s_k^2}(\beta_k^* - \widetilde{\beta_k})\}}{|C_k + 2r\frac{v_k s_k^2}{v_k-2}D(\beta_k^*)|^{\frac{1}{2}}}, \tag{8}$$

where $D(\beta_k^*)$ is the diagonal matrix with positions $(i,i), i \in k$ is given by $1/(\beta_k^*)^2$ and

$$\beta_k^* = argmax_{\beta_k}\{N(\beta_k; \widetilde{\beta}_k, \frac{v_k}{v_k - 2} s_k^2 C_k^{-1}) \prod_{i \in k} \beta_i^{2r}\}. \tag{9}$$

Based on the unknown σ^2 situation, if we set $A_k = I_k$, then the normalization constant d_k is given by

$$d_k = [(2r-1)!!]^{-k} \tag{10}$$

However, our improvement of this method is to calculate d_k without just setting $A_k = I_k$. We will implement the calculation of d_k in next section.

The posterior probability of model t is defined as:

$$p(t|y_n) = \frac{p(t)m_t(y_n)}{\sum_{k \in \mathcal{J}} p(k)m_k(y_n)}, \tag{11}$$

where $p(K)$ is defined as the prior probability assigned to model K. This prior probability is defeined as:

$$p(k|\gamma) = \gamma^k(1-\gamma)^{n-k}, \qquad \gamma \sim Beta(\zeta_0, \zeta_1) \tag{12}$$

where we also define $\zeta_0 = \zeta_1 = 1$.

Based on the above results of Laplace approximation of marginal likelihoods of the data, we can use the following Metropolis-Hastings MCMC algorithm for exploring the model space.

1. Choose an initial model k^{curr}
2. For $i = 1, \ldots, p$,

(a) Define model k^{cand} by excluding or including β_i from model k^{curr}, according to whether β_i is currently included or excluded from k^{curr}.

(b) Compute

$$r = \frac{m_{k^{cand}}(y)p(k^{cand})}{m_{k^{cand}}(y)p(k^{cand}) + m_{k^{curr}}(y)p(k^{curr})} \tag{13}$$

based on the approximation of marginal likelihoods.

(c) Draw $u \sim U(0,1)$. If $r > u$, define $k^{curr} = k^{cand}$.

3. Repeat Step 2 until a suffciently long MCMC chain is obtained.

2.2 Normalizing Constant Computation

The distance d_k of model k is given by

$$d_k = \left[\int_{\mathcal{R}^k} (2\pi)^{-k/2} |A_k|^{1/2} exp(-\frac{1}{2}\gamma' A_k \gamma) \prod_{i=1}^{k} \gamma_i^{2r} d\gamma \right]^{-1} \tag{14}$$

where

$$\gamma \sim MVN(0, A_k) \tag{15}$$

which means

$$d_k^{-1} = E[(\gamma_1)^{2r}(\gamma_2)^{2r} \dots (\gamma_k)^{2r}] \tag{16}$$

where $E[\dots]$ denotes the expected value.

To derive this central moment, we can consider the derivatives of the distribution's characteristic function. Because the derivation is complex to be fully described in article, we just give central part of this derivation. The characteristic function of the multivariate normal distribution is(Lukacs (1942)):

$$E[e^{it^T X}] = e^{it^T \mu - \frac{1}{2} t^T \sum t} \tag{17}$$

where $t = (t_1, t_2, \dots, t_n)$. Because we want to calculate the centrol moments, the characteristic function has reduced as:

$$E[e^{it^T X}] = e^{-\frac{1}{2} t^T \sum t} \tag{18}$$

The moment is then, a $k_1, \dots, k_n$-order derivative of the characteristic function evaluated at $t = 0$:

$$m_{k1,\dots,k_n} = i^{-\sum_{i=1}^{n} k_i} \frac{d^{\sum_{i=1}^{n} k_i}}{d^{k_1} t_1 d^{k_2} t_2 \dots d^{k_n} t_n} E[e^{it^T X}]|_{t=0} \tag{19}$$

Because in our model, we just to calculate central moments with orders all equal to $2r$, the derivative becomes:

$$m_{2r,\ldots,2r} = i^{2nr} \frac{d^{2nr}}{d^{2r}t_1 d^{2r}t_2 \ldots d^{2r}t_n} E[e^{it^T X}]|_{t=0} \tag{20}$$

If we apply Taylor expansion on exponential item in (22), this formula becomes:

$$m_{2r,\ldots,2r} = i^{2nr} \frac{d^{2nr}}{d^{2r}t_1 d^{2r}t_2 \ldots d^{2r}t_n} \sum_{l=0}^{\inf} (-\frac{1}{2}(t^T \Sigma t))^l / l!|_{t=0} \tag{21}$$

Because, $2nr$ is even, the item $i^{2nr} = (-1)^{-nr}$. Σ is covariance matrix of this multivariate normal distribution and in our model, it has been specifed as A_n, then the expression in t specifed in our model is:

$$t^T \Sigma t = \sum_{ij} a_{ij} t_i t_j \tag{22}$$

In order to calculate:

$$(t^T \Sigma t)^l = (\sum_{ij} a_{ij} t_i t_j)^l = \sum_{ij} a_{ij} t_i t_j \ldots \sum_{ij} a_{ij} t_i t_j \tag{23}$$

we need to find the coefficient of $t_1^{l_1} t_2^{l_2} \ldots t_n^{l_n}$ in (23). Here l_{ij} is a certain number of times of choosing $a_{ij} t_i t_j$ into a particular product item. Because one item $a_{ij} t_i t_j$ will be chosen from each $\sum_{ij} a_{ij} t_i t_j$, we have $\sum_{ij} l_{ij} = l$. For any l_{ij} there are

$$\binom{l}{l_{ii} \ldots l_{nn}} \tag{24}$$

ways to choose the items. Then,

$$\begin{aligned} (\sum_{ij} a_{ij} t_i t_j)^l &= \sum_{(l_{ij})|\sum_{ij} l_{ij} = l} \binom{l}{l_{ii} \ldots l_{nn}} \prod_{ij} (a_{ij} t_i t_j)^{l_{ij}} \\ &= \sum_{(l_{ij})|\sum_{ij} l_{ij} = l} \binom{l}{l_{ii} \ldots l_{nn}} \prod_{ij} a_{ij}^{l_{ij}} \prod_{ij} (t_i t_j)^{l_{ij}} \end{aligned} \tag{25}$$

because $t_i t_j = t_j t_i$, each a_{ij} is combined with two t's, the total exponent in t is $2l$ which means an item $t_1^{l_1} t_2^{l_2} \ldots t_n^{l_n}$ has property: $\sum_{ij} l_{ij} = 2l$. For following derivation, we also define $l_1, \ldots l_n$:

$$\prod_{ij} (t_i t_j)^{l_{ij}} = t_1^{l_1} \ldots t_n^{l_n} \tag{26}$$

Then for any $t_k, k = 1, \ldots, n$, the exponent of t_k is:

$$\sum_{i=k, j\neq k} l_{ij} + \sum_{i\neq k, j=k} l_{ij} + 2l_{kk} \tag{27}$$

Then we have:

$$l_k = \sum_i l_{ik} + \sum_j l_{kj} = \sum_i (l_{ik} + l_{ki}) \tag{28}$$

Implement (28) into (25):

$$\begin{aligned} &(\sum_{ij} a_{ij} t_i t_j)^l = \\ &\sum_{(l_1,\ldots,l_n)|\sum_i l_i = 2l} \sum_{(l_{11},\ldots,l_{nn})|\sum_i (l_{ik}+l_{ki}) = l_k} (\binom{l}{l_{11}\ldots l_{nn}} \prod_{ij} a_{ij}^{l_{ij}}) \prod_i (t_i)^{l_i} \end{aligned} \tag{29}$$

Moreover, the deriviative terms in (21) can be expressed as:

$$\begin{aligned} &\frac{d^{2nr}}{d^{2r}t_1 d^{2r}t_2 \ldots d^{2r}t_n} (\sum_{ij} a_{ij} t_i t_j)^l) = \\ &\sum_{(l_1,\ldots,l_n)|\sum_i l_i = 2l} \sum_{(l_{11},\ldots,l_{nn})|\sum_i (l_{ih}+l_{hi}) = l_h, h=1,\ldots,n} (\binom{l}{l_{11}\ldots l_{nn}} \prod_{ij} a_{ij}^{l_{ij}}) \\ &\prod_i I\{2r \leq l_i\} \frac{l_i!}{(l_i - 2r)!} (t_i)^{l_i - 2r} \end{aligned} \tag{30}$$

Considering $t = 0$, only items with $l_i = 2r$ for all i will remain. Then, we get $l = nr$. After reduction, we can get the central moments:

$$m_{2r,\ldots,2r} = \frac{1}{2}^{nr} ((2r!)^n)/(nr)! \sum_{(l_{11},\ldots,l_{nn})|\sum_i (l_{ih}+l_{hi}) = 2r, h=1,\ldots,n} \binom{nr}{l_{11}\ldots l_{nn}} \prod_{ij} a_{ij}^{l_{ij}} \tag{31}$$

Now, (31) is the calculation result of central moments of multrivariate normal distribution. Then, we have:

$$d_k = m_{2r,\ldots,2r}^{-1} \tag{32}$$

with dimension k. Replace $d_k = [(2r_1)!!]^{-k}$ in (8). The Laplace approximation used in our model is:

$$\frac{\Gamma(\frac{v_k}{2}) \psi^\alpha 2^{\frac{v_k}{2}} (2\psi + y'y - \widetilde{\beta}_k{}' C_k \widetilde{\beta}_k)^{-\frac{v_k}{2}}}{\Gamma(\alpha) m_{2r,\ldots,2r}^{-1} (2\pi)^{\frac{n}{2}} \tau^{\frac{k}{2}+rk}} \times \frac{(\prod_{i\in k} (\beta_i^*)^{2r}) exp\{-\frac{v_k - 2}{2v_k} (\beta_k^* - \widetilde{\beta}_k)' \frac{C_k}{s_k^2} (\beta_k^* - \widetilde{\beta}_k)\}}{|C_k + 2r \frac{v_k s_k^2}{v_k - 2} D(\beta_k^*)|^{\frac{1}{2}}}, \tag{33}$$

We will use the same Metropolis-Hastings MCMC algorithm for exploring the model space and compare our results with Johnson and Rossell's results in simulation studies.

3 Simulation Studies

In this section, we proceeded the above algorithm in several simulation experiments, and we compare the results of our algorithm with Johnson and Rossell's algorithm. Considering the complexity of the algorithm, for each simulated dataset, we performed 500 burn-in iterations and 5000 subsequent updates for posterior inference.

Based on the main results of simulation study in Johnson and Rossell's article, we use following settings to generate simulation data. We set sample size $n = 200$, size of parameters $p = 1000$. We generate $X_{n\times p}$ from a multivaraite normal distribution:

$$X \sim MVN(0, \Sigma) \tag{34}$$

where Σ is a $p \times p$ matrix with diagonal elements 1 and off-diagonal elements $\rho = 0.25$ which means the variance of each column of X is 1 and correlation between each two columns of X is 0.25. Considering we will apply our algorithm in imaging areas, we generate X in a $10 \times 10 \times 10$ space. We set four components of the regression coefficients to nonzero values and others are zero. The four points that we set non-zero β are in the center of our data space which is $(5, 5, 5), (5, 5, 6), (6, 5, 5), (5, 5, 6)$. We have drawn a 3D scatter plot for how the nonzero β distribute in data space in Figure 2. In each simulation dataset, we will set the values of non-zero β in following fomula(Johnson (2014)):

$$\beta_i = (-1)^{u_i}(c_i log(n)/\sqrt{n} + |z_i|) \tag{35}$$

where u was a Bernoulli random variable with success probability 0.4, z was a standard normal deviate, and c will be set in 0.5,1.0,2.0,4.0 to do comparison. $Y = X \times \beta + \epsilon$, where ϵ is a $p \times 1$ vector and each component is generated from a standard normal

distribution. Because we are considering the process of unknown σ^2, we set a vague $IG(0.001, 0.001)$ prior for it in all procedures based on non local priors, i.e $\alpha = 0.001, \psi = 0.001$. To calculate A_p, because in our model space, the distance between each point is standard Eular distance, then the smallest distance is 1. We define each element of matrix A_p is:$e^{-\rho_1 \times d}$, where ρ_1 is a user defined parameter based on the distance definition and d is the square of Eular distance between two points. We also set the hyperparameter $\tau = 0.348$. We repeat the simulation process 10 times with the four different settings of c. The results are shown in Table 1.

The results in Table 1 has proved that our new algorithm has some improvements. In Table 1, MSE is means square error of fitted value $\hat{Y}$ and the true Y. Because we already know the positions and numbers of non-zero β, TP in table 1 means true postive cases i.e all the non-zero β are included in the final model, TN means true negative case, FP is false positive, FN is false negative. 1 means results of our algorithm and 2 means results of Johnson and Rossell's algorithm.

The results have shown that in all the cases, the final models obtained by our algorithm have competitive MSE compared with Johnson and Rossell's algorithm. Moreover, our algorithm can include all the non-zero β in our final model. We can find when c increases, their algorithm will choose a lot zero β into final model, even the MSE is small.

4 Real Data Studies

In this section, we have proceeded these two algorithms with Autism Brain Imaging Data Exchange data (ABIDE). The major goal of the ABIDE study is to explore association of brain activity with the autism spectrum disorder (ASD) which is a group of developmental disabilities characterized by atypical development in socialization, communication, and behavior (Rice 2009). There are 539 ASDs and 573

age-matched typical controls in this data set, aggregated 20 resting-state functional magnetic resonance imaging (R-fMRI) data sets from 17 different sites. In this work, we analyze the voxel-wise fALLF values over 1 specific region: ParaHippocampal_L, which includes 978 voxels that is a similar size with our simulation studies.

For the real data, we set FIQ standard score as our outcome Y which is developed from information gathered from patient reports, functional status instruments, and clinical observations. This instrument measures physical functioning, work status (missed days of work and job difficulty), depression, anxiety, morning tiredness, pain, stiffness, fatigue, and well-being over the past week. We have proceeded these two algorithms on both ASD group and control group. Due to the missing data of FIQ score, the final dataset includes 438 controls and 403 ASDs. We set image data of these subjects as X in our analysis. Because the image data are highly correlated, we apply Fisher z transformation ie. $log(fALLF/(1 - fALLF))$ on them and we center the outcome i.e. FIQ score for following analysis. All the parameters settings are same as described in simulation section.

When we proceed these two algorithms with ABIDE data, our novel algorithm can obtain a final model quickly with a stable MCMC chain and the MSE is pretty small to support our final model is accurate, however, the performance of Johnson and Rossell's algorithm is poor: with the same data, the final model their algorithm obtain is a model with all the coefficients are 0 which means we even can not calculate the MSE from their final model. The results are shown in Table 2 and we also plot a 3-dimension picture Figure 3 to show how the voxels we select distribute in the region.

5 Discussion

We have developed a spatial Bayesian model selection procedure with nonlocal prior considering the spatial dependence of data. In the process we demonstrated how to incorporate calculation of spatial distance to the nonlocal prior model. This work has the potential to have a broad and immediate application in spatial dependence data analysis in different areas like: brain imaging area (Lee et al. 2014), spatial lattice data (Song and Oliveira 2012)). For broader goals it would be of interest to extend the model to neighborhood selection (Assuncao and Krainski 2009)).

From simulation studies, we can find the algorithm we propose is quite competitive with the old one. When we proceed the real data studies, our results have proved that if the Bayesian model selection procedures ignore spatial dependence in spatial data, the procedures perform poor. Our novel algorithm provide a appropriate way to choose high post posterior probability model with spatial data.

However, such a solution presents numerous modeling and computational challenges. For example, the study of MCMC convergence in such high dimensional settings remains unexplored (present study included). Hence the consequences of having far more parameters than possible MCMC iterations remains unknown. Moreover, there are few methods of calculation of central moment in software (Phillips 2010) and the speed of calculation will decrease dramatically when the dimension of d_k increases which means the algorithm is appropriate for sparse spatial model recently.

Despite computational and modeling difficulties, we are confident that nonlocal prior Bayesian approaches represent an important direction in spatial dependence data due to its final will have high posterior probability and its competitive accuracy.

Reference

Assuncao, R. and Krainski, E. (2009), "Neighborhood Dependence in Bayesian Spatial Models," *Biometrical Journal*, 851-69, 10.1002.

Berger, J. O. and Pericchi, L. R. (1996), "The Intrinsic Bayes Factor for Model Selection and Prediction," *JASA*, 91, 109–122.

Fan, J. and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties," *JASA*, 96, 1348–1360.

Johnson, V. E. (2014), "On Numerical Aspects of Bayesian Model Selection in High and Ultrahigh-dimensional Settings," *Bayesian Analysis*, 8, 741–758.

Johnson, V. E. and Rossell, D. (2012), "Bayesian Model Selection in High-Dimensional Settings," *JASA*, 107:498, 649–660.

Lee, K.-J., Jones, G. L., Caffo, B. S., and Bassett, S. (2014), "Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data," *Bayesian Analysis*, 9, 699–732.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008), "Mixtures of g Priors for Bayesian Variable Selection," *JASA*, 103, 410–423.

Lukacs, E. (1942), "A Characterization of the Normal Distribution," *JASA*, 13, 91–93.

O'Hagan, A. (1995), "Fractional Bayes Factors for Model Comparison," *JRSSB*, 57, 99–138.

Phillips, K. (2010), "Neighborhood Dependence in Bayesian Spatial Models," *Journal of Statistical Software*, 33.

Rice, C. (2009), "Prevalence of Autism Spectrum Disorders: Autism and Developmental Disabilities Monitoring Network, United States, 2006. Morbidity and Mortality Weekly Report." *Centers for Disease Control and Prevention*, 58, 28.

Song, J. J. and Oliveira, V. D. (2012), "Bayesian model selection in spatial lattice models," *Statistical Methodology*, 9, 10.1016.

Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *JASA*, 101, 1418–1429.

Appendices

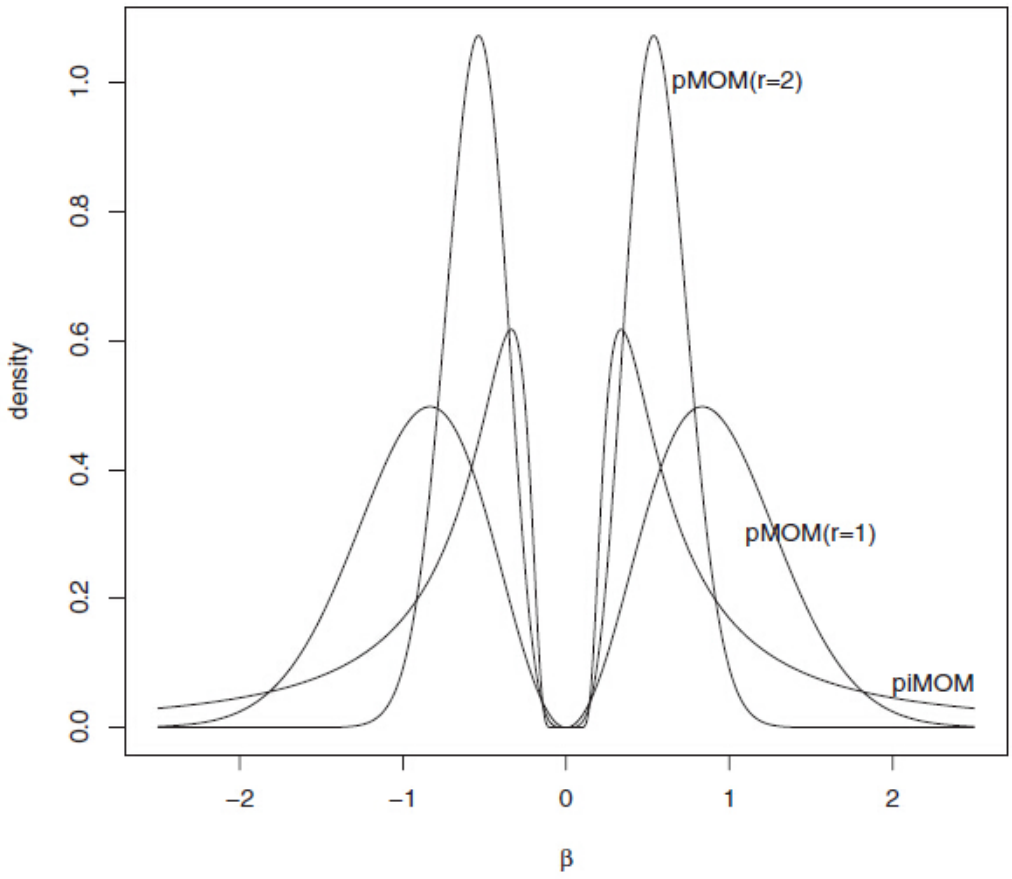

Figure 1: Nonlocal prior densities for a single regression coefficient.

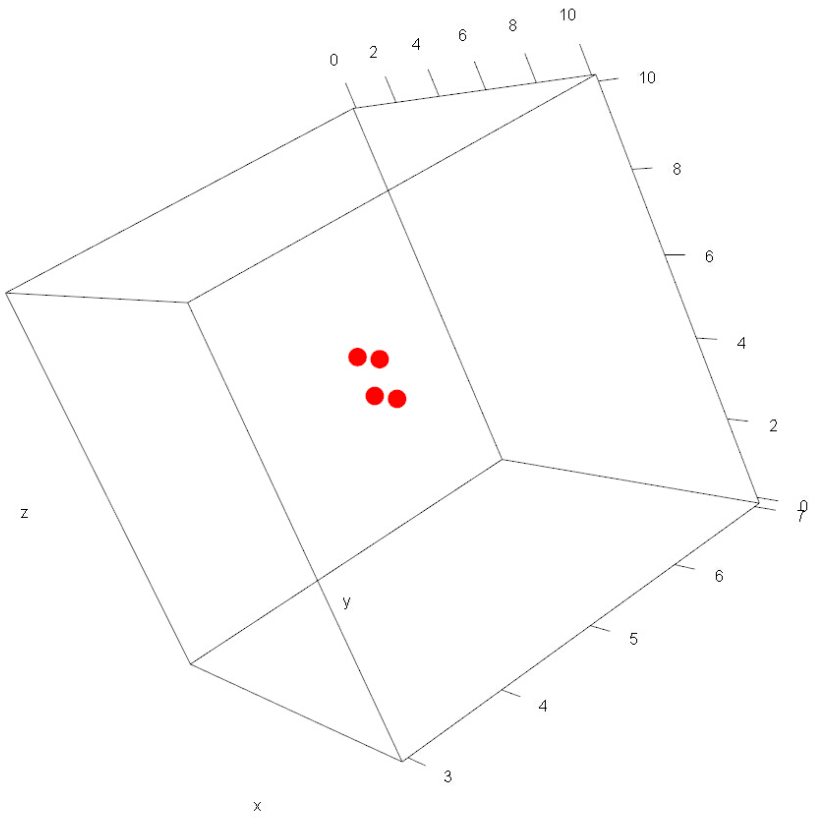

Figure 2: 3-dimension plot of simulation data with nonzero β

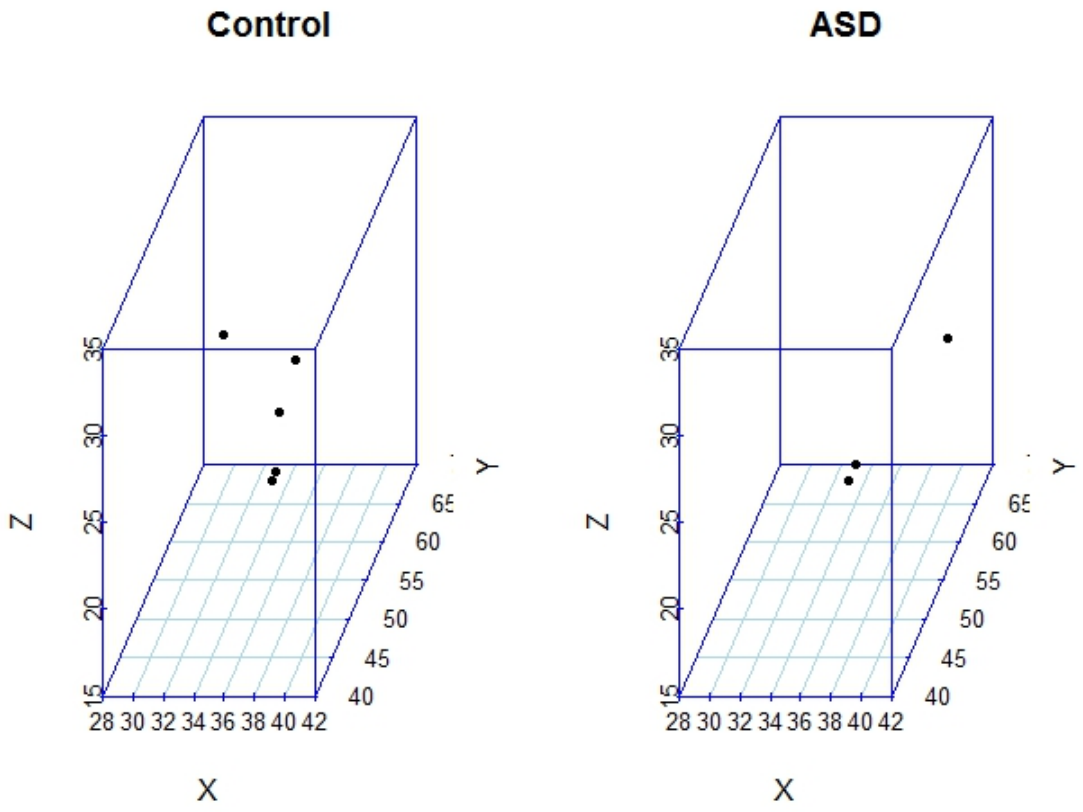

Figure 3: Spatial distribution of selected voxels in both groups

c	MSE1	MSE2	TP1	TP1	TN1	TN2	FP1	FP2	FN1	FN2
0.5	1.032	1.089	4	2	996	996	0	0	0	2
1.0	0.982	0.980	4	4	996	996	0	0	0	0
2.0	0.885	0.871	4	4	996	996	0	0	0	0
4.0	0.973	0.964	4	4	996	996	0	0	0	0
0.5	0.905	0.933	4	3	996	996	0	0	0	1
1.0	0.835	0.962	4	3	996	996	0	0	0	1
2.0	0.921	0.913	4	4	996	996	0	0	0	0
4.0	1.020	1.085	4	2	996	920	0	76	0	2
0.5	1.080	1.060	4	4	996	996	0	0	0	0
1.0	0.926	0.912	4	4	996	996	0	0	0	0
2.0	0.925	0.894	4	4	996	996	0	0	0	0
4.0	0.854	0.870	4	4	996	965	0	31	0	0
0.5	1.118	1.214	4	3	996	996	0	0	0	1
1.0	0.920	0.877	4	4	996	996	0	0	0	0
2.0	0.945	0.931	4	4	996	996	0	0	0	0
4.0	1.094	1.093	4	4	996	996	0	0	0	0
0.5	0.929	0.890	4	4	996	996	0	0	0	0
1.0	0.993	0.975	4	4	996	996	0	0	0	0
2.0	1.126	1.107	4	4	996	996	0	0	0	0
4.0	0.919	0.927	4	2	996	912	0	84	0	2
0.5	1.049	1.226	4	1	996	996	0	0	0	3
1.0	0.769	0.768	4	4	996	996	0	0	0	0
2.0	0.910	0.861	4	4	996	996	0	0	0	0
4.0	0.996	0.974	4	4	996	996	0	0	0	0
0.5	1.066	1.147	4	3	996	996	0	0	0	1
1.0	1.064	1.060	4	4	996	996	0	0	0	0
2.0	1.003	0.975	4	4	996	996	0	0	0	0
4.0	0.960	0.952	4	4	996	996	0	0	0	0
0.5	1.045	1.022	4	4	996	996	0	0	0	0
1.0	0.952	0.911	4	4	996	996	0	0	0	0
2.0	1.237	1.176	4	4	996	996	0	0	0	0
4.0	1.036	0.773	4	4	996	947	0	49	0	0
0.5	1.136	1.109	4	4	996	996	0	0	0	0
1.0	0.930	0.915	4	4	996	996	0	0	0	0
2.0	0.865	0.846	4	4	996	996	0	0	0	0
4.0	0.884	0.567	4	4	996	923	0	73	0	0
0.5	0.864	0.849	4	4	996	996	0	0	0	0
1.0	0.923	0.898	4	4	996	996	0	0	0	0
2.0	1.012	1.000	4	4	996	996	0	0	0	0
4.0	1.107	1.065	4	4	996	996	0	0	0	0

Table 1: 10 times Simulation results of our model and Johnson and Rossell's model

Group	Selected Voxels Position	MSE
Control	1 3 539 741 935	0.931
ASD	1 7 410	0.956

Table 2: Final model and MSE of both groups in region: ParaHippocampal_L