Spatial Bayesian Variable Selection with Application to Functional Magnetic Resonance Imaging (fMRI)

by

Ying Yang

Department of Statistical Science
Duke University

Date: ______________________

Approved:

________________________________
Fan Li, Supervisor

________________________________
Merlise A. Clyde

________________________________
Nicole Calakos

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Statistical Science in the Graduate School of Duke University 2011
ABSTRACT

Spatial Bayesian Variable Selection with Application to Functional Magnetic Resonance Imaging (fMRI)

by

Ying Yang

Department of Statistical Science
Duke University

Date: _____________________

Approved:

___________________________

Fan Li, Supervisor

___________________________

Merlise A. Clyde

___________________________

Nicole Calakos

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Statistical Science in the Graduate School of Duke University 2011
Copyright © 2011 by Ying Yang
All rights reserved except the rights granted by the Creative Commons Attribution-Noncommercial Licence
Abstract

Functional magnetic resonance imaging (fMRI) is a major neuroimaging methodology and have greatly facilitate basic cognitive neuroscience research. However, there are multiple statistical challenges in the analysis of fMRI data, including, dimension reduction, multiple testing and inter-dependence of the MRI responses. In this thesis, a spatial Bayesian variable selection (BVS) model is proposed for the analysis of multi-subject fMRI data. The BVS framework simultaneously account for uncertainty in model specific parameters as well as the model selection process, solving the multiple testing problem. A spatial prior incorporate the spatial relationship of the MRI response, accounting for their inter-dependence. Compared to the non-spatial BVS model, the spatial BVS model enhances the sensitivity and accuracy of identifying activated voxels.
Contents

Abstract iv
List of Tables vii
List of Figures viii
1 Introduction 1
  1.1 Functional magnetic resonance imaging (fMRI) 1
  1.2 fMRI data analysis 4
  1.3 Scope of this thesis 8
2 Multilevel fMRI data analysis with spatial Bayesian variable selection 9
  2.1 Two-level random effects model 9
  2.2 Spatial smoothing of voxel-specific effects via Ising prior 10
  2.3 A joint model for fMRI and behavioral/SNP data 12
3 Posterior inference 14
  3.1 Gibbs sampler for the spatial model 14
  3.2 Gibbs sampler for the spatial model with predictors for subject-specific effects 15
  3.3 Incorporating anatomical prior information 17
4 Simulation study 19
  4.1 Duke Neurogenetics Study 19
  4.2 Simulation Scheme 20
4.3 Selection of hyper-parameters ........................................ 25
4.4 Convergence Diagnostic ............................................. 26
4.5 Performance of the spatial BVS model .......................... 26

5 Conclusion .................................................................. 30

A R codes for Spatial Bayesian variable selection .................. 32
  A.1 Parameter calculation .................................................. 32
  A.2 Gibbs Sampler ......................................................... 34

Bibliography .................................................................. 40

Biography ................................................................... 42
List of Tables

4.1 Parameters to generate simulated data ............... 22

4.2 Performance of spatial BVS models (Area under ROC curve) under
different signal strength $\mu_L$ for individual response curve ........ 28
List of Figures

1.1 Canonical model of the HRF in fMRI data analysis ................. 3
1.2 Voxel BOLD TSeries as convolution of stimulus function with HRF . 4
4.1 Block design experimental scheme in Duke Neurogenetics Study . . 20
4.2 Design matrix for Duke Neurogenetics Study ...................... 21
4.3 Grid layout and responsive voxels in simulation .................... 22
4.4 Example MRI reading $Y_{ij}$ for three response curves; case 3 ........ 23
4.5 Simulated MRI reading $Y_{ij}$ from the same subject for three voxels that are Strongly Activated (a), Weakly Activated (b), and Unactivated (c). Response curve X2, case 3 ................................. 24
4.6 ROC curves for individual response curve, case 3 .................... 27
4.7 ROC curves and posterior amplitude maps; response curve X1, case 2 29
1 Introduction

1.1 Functional magnetic resonance imaging (fMRI)

To measure brain activity associated with discrete states of mind is the holy grail of cognitive neuroscience. One major advancement in realizing this dream is neuroimaging. Neuroimaging, particularly that based upon functional magnetic resonance, has experienced an explosive growth in recent years and become a dominant tool in cognitive neuroscience. In human subjects, fMRI has been used routinely to study sensory processing or control of action, and to investigate the neural mechanisms of cognitive capacities including recognition, memory and emotion (Logothetis (2008)).

Functional MRI is a noninvasive technique to study brain activity. Functional activation of the brain can be detected with MRI via local changes in the concentration of blood oxygen in the brain. When a group of neurons are active, they induce a regional increase in blood-oxygen-level-dependent (BOLD) contrast. The magnitude of this BOLD contrast depends on the strength of the scanner and the quantities used to generate the contrast, thus, is relative and not individually quantititative. As a result, fMRI studies generally compare BOLD contrasts in the presence or absence of some cognitive task or stimulus (Logothetis (2008)).
In a typical fMRI experiment, a series of brain images are acquired while the subject performs a set of tasks, and changes in the measured signal between individual images are used to make inferences regarding task-related activations in the brain (e.g., BOLD signal). The fMRI data comprise a sequence of magnetic resonance images (MRI), each consisting of a number of uniformly spaced volume elements, or voxels, that partition the brain into equally sized boxes. In an fMRI experiment for a single subject, images of this type are collected at T separate time point where T varies between 100-2000. For many fMRI human subject studies, three-dimensional (3D) images of the full brain volume are desired; the standard approach toward 3D imaging is to acquire a stack of adjacent slices (e.g., 20 - 30) in quick succession. A typical full brain volume image has dimension $64 \times 64 \times 30$ (with spatial resolution on the order of $3 \times 3 \times 5 \ mm^3$) and takes 2-3 seconds to obtain. In other words, a single subject fMRI experiment would generate data consisting of approximately 100,000 time series of length T, correlating spatially with each other. The same experiment may be repeated several times for the same subject, as well as for multiple subjects to facilitate population inference (Lindquist (2008)).

In most fMRI studies, BOLD signals from individual voxels are used to infer brain activity. BOLD imaging takes advantage of inherent differences between oxygenated and deoxygenated hemoglobin, each of which produces different local magnetic fields–deoxy-hemoglobin has the effect of suppressing the MR signal, whereas oxy-hemoglobin does not. The cerebral blood flow refreshes areas of the brain that are active during the execution of a mental task with oxygenated blood, thereby changing the local magnetic susceptibility and the measured MR signal in active brain regions. The usual signal increases reported in BOLD fMRI experiments are due to the fact that population neural activation induces a regional increase in cerebral blood flow that is always larger than the oxygen consumption rate, since oxygen uptake is diffusion-limited. The net effect of neural excitation is thus a seemingly
paradoxical drop in the deoxy-hemoglobin concentration, which in turn increases the MRI signal strength (Logothetis (2008)).

Due to the complex events linking regional neural activation to local hemodynamic change that generate the BOLD signal, hemodynamic response function (HRF) has been used to deconvolve the BOLD signal to obtain neural activation. Fig.1.1 shows the standard shape used to model the HRF. The positive rise in signal corresponding to the regional increase in blood flow has an onset approximately 2 seconds after the onset of population neural activity and peaks 5-8 seconds after the peak of population neural activity. Then the BOLD signal decreases to a below baseline level for roughly 10 seconds, due to the fact that blood flow decreases more rapidly than blood volume, allowing for a greater concentration of deoxy-hemoglobin in previously active brain regions (Lindquist (2008)). Because the time between acquisition of each individual image (repetition time, TR) in whole brain fMRI study is approximately 2 seconds, the fMRI time series is capable of capturing both the positive and the negative BOLD response.

![Figure 1.1: Canonical model of the HRF in fMRI data analysis](image)

The evoked BOLD response in fMRI depends both on the applied stimulus and the hemodynamic response to neural activation. In a linear system framework, the BOLD signal at time $t$, $x(t)$, is modeled as the convolution of a stimulus function
\(v(t)\) and the HRF \(h(t)\), that is,

\[
x(t) = \int h(u)v(t - u)du
\]  

(1.1)

\(h(t)\) is either assumed to take a canonical form, or alternatively modeled using a set of linear basis functions. Fig.1.2 gives an example of BOLD response of a voxel if the HRF takes the canonical form.

**Figure 1.2**: Voxel BOLD TSeries as convolution of stimulus function with HRF

1.2 fMRI data analysis

There are several common objectives in the analysis of fMRI data. These include localizing regions of the brain activated by a task, determining distributed networks that correspond to brain function, and making predictions about psychological or disease states. To achieve these objectives, it is necessary to construct statistical maps for every voxel in the brain—whether a voxel is activated by the cognitive task or stimulus (activation map), and what is the magnitude of its activation (amplitude map) (Logothetis (2008)).

In the generalized linear model (GLM) framework (Friston et al. (1995); Worsley and Friston (1995); Goutte et al. (2000)), single subject fMRI data is analyzed by fitting each voxel independently (i.e. univariate approach) as a linear combination of
independent variables, plus an error term. The voxel-wise GLM for a single subject is expresses as,

$$Y_j = X\beta_j + \epsilon_j$$  \hspace{1cm} (1.2)

where $Y_j$ is a column vector of $T$ rows (the number of collected time points) representing the time series BOLD signal associated with a single voxel $j$. $X$ represents the design matrix with $T$ rows $\times$ $K$ columns, each representing a regressor (i.e. an explanatory variable) as convolution of stimulus function with HRF (such as in Fig.1.2); such time series is also called hypothetical time series of the hemodynamic response. $\beta_j$ is a column vector with $K$ rows representing the unknown parameter associated to each regressor for voxel $j$. Finally, $\epsilon_j$ is a column vector with $T$ rows representing error.

Constructing activation map and amplitude map under the GLM framework is no easy task. First, a fMRI dataset for a single subject consists of roughly 100,000 voxels, which is tens to hundreds times more than the number of observations (small $n$ large $p$ problem; West (2003)). Second, the measured BOLD signal is corrupted by random noise and various nuisance components that arise from both hardware reasons and the subjects themselves. Due to the large amount of data and the relatively weak BOLD signal, there are many statistical challenges in fMRI data analysis and it is generally considered not feasible for a full spatiotemporal model for the data (Lindquist (2008)).

One statistical challenge in analyzing fMRI data is how to compress the massive image time series into low-dimensional quantitative summaries. To mitigate this challenge, many fMRI experiments use block designs (Logothetis (2008)). In a block design experiments, the different experimental conditions are separated into extended time intervals, or blocks. For example, one might repeat the process of interest (e.g., finger tapping) during an experimental block (A) and have the subject rest during a
control block (B). The statistical significance is determined from difference in BOLD signals between these blocks. In general, increasing the length of each block will lead to a larger evoked response during the task. This increases the separation in signal between blocks, which, in turn, leads to higher detection power and are robust to uncertainties in the shape of HRF. The robustness is due to the fact that the predicted response depends on the total activation caused by a series of stimuli, rendering it less sensitive to the variations in the shape of response to individual stimulus (see Fig.1.2). If the same mental processes are evoked throughout each block, then the massive time series can be compressed into an average and variance for each condition, and inference can be drawn based upon these quantities (Friston et al. (1995); Lindquist (2008)). Activation map is then constructed in a voxel-by-voxel manner, which involves large scale multiple testing—a second statistical challenge in analyzing fMRI data.

A conventional approach to reduce the scale of multiple testing is by region-of-interest (ROI) studies. ROIs are chosen based upon neuroanatomy and prior information on which brain regions might be related to the cognitive task. This approach involve computing an ROI average measure of brain activity in each scan and then performing statistical analysis on the summary statistics. This approach reduces the dimension of multiple testing to tens to hundreds.

Recent fMRI studies started to consider analyzing brain activity measured simultaneously at many locations. To do so, researches first set a permissive threshold to select activated voxels in a voxel-by-voxel manner without spatial or temporal smoothing, then perform multivariate analysis (Kamitani and Tong (2005)). Such multivariate analysis can often detect small differences between two task or stimulus conditions—differences that are not picked up by conventional univariate methods (Haynes and Rees (2006)). However, such analysis does not incorporate the spatial information of voxels, nor is it easily adaptable to analyze interactions of different
Bayesian inference has become increasingly popular in fMRI data analysis due to several attractive properties: first, the posterior inference offers direct probabilistic interpretation of the estimates; second, it eschews the multiple-comparison problem faced by classical inference; third, incorporating prior information is straightforward within the Bayesian framework. Gössl et al. (2001) and Woolrich et al. (2004) proposed Bayesian spatial-temporal models for single-subject fMRI data. Friston (2002) investigated a univariate hierarchical GLM for multi-subject fMRI data under the empirical Bayes framework. Bowman (2007), Bowman et al. (2008) and Derado et al. (2010) constructed Bayesian hierarchical spatial models for multi-subject data. Smith et al. (2003) and Smith and Fahrmeir (2007) proposed a Bayesian variable selection (BVS) approach to perform selection of active voxels and spatial smoothing simultaneously in the analysis of single-subject data. Within the BVS framework, model sparsity can be naturally achieved, which is crucial for computational efficiency, especially when Markov chain Monte Carlo (MCMC) calculations are involved. We will also adopt the BVS approach in this research, but with several major extensions as illustrated below.

Despite the rapid advances in Bayesian inference of fMRI data, however, to our knowledge, no research has been conducted in joint modelling of neuroimaging outcomes with subject-specific predictors for multi-subject analysis, while incorporating model sparsity as well as spatial smoothing. In this thesis, we propose a full Bayesian framework for joint modelling of fMRI data and behavioral or even high dimensional subject-specific data, such as single-nucleotide polymorphism (SNP) data. The novelty of the proposed approach are: (1) it enables simultaneous spatial smoothing and selection of active voxels for multi-subject data; (2) the proposed additive assumption substantially reduces the number of parameters to be estimated, thus vastly improves the computational feasibility of Bayesian hierarchical models in imaging.
analysis. Further, with our additive assumption, it is theoretically trivial to impose Bayesian variable selection priors on high dimensional subject-specific predictors, potentially allowing for simultaneously inference of the relationship between fMRI phenotypes and SNP genotypes and selection of important SNPs.

1.3 Scope of this thesis

This thesis is the preliminary work for a project focusing on novel statistical analysis of fMRI data. In this thesis, I described our multi-subject additive spatial Bayesian variable selection model (Chapter 2). I derived the Gibbs sampler for all parameters to be estimated in the model (Chapter 3). I then assessed the performance of the spatial BVS model (vs the non-spatial BVS model) with simulation, using the hemodynamic time series for the actual fMRI data as my design matrix (Chapter 4). Computation in this thesis is done with R.

The goal for the project is to analyze whole-brain fMRI data for multiple subjects with our additive spatial BVS model, and associate fMRI phenotype with behavioral phenotype and/or SNP genotype. However, implementation of the Gibbs sampling in the whole-brain fMRI is extremely intensive in computation. Because fMRI data are acquired slice by slice and spatial correlation among the voxels within a slice is much larger than that across the slices, future analysis shall be done by breaking the posterior computation by slices and implementing parallel computation, using fortran or C++. 
Multilevel fMRI data analysis with spatial Bayesian variable selection

2.1 Two-level random effects model

Let \( Y_i = (y_{i1}, \ldots, y_{iJ}) \) be the observed fMRI signal for subject \( i \) \((i = 1, \ldots, N)\), where \( y_{ij} = (y_{ij}(1), \ldots, y_{ij}(T))' \) is the time series of voxel \( j \). Let \( Y = (Y_1, \ldots, Y_n) \) denote all the fMRI data of the sample. Then the matrix form of the generalized linear model (GLM) for multi-subject fMRI data is

\[
y_{ij} = X_i \beta_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, N; \quad j = 1, \ldots, J,
\]

where \( X \) is a \( T \times K \) design matrix where \( X_k \)'s are the individual hypothetical time series of the hemodynamic response (for example, Fig.1.2, red trace in right panel), \( \beta_{ij} = (\beta_{ij1}, \ldots, \beta_{ijK})' \) is the \( K \times 1 \) coefficient measuring the hemodynamic magnitude of voxel \( j \) for subject \( i \), and \( \epsilon_{ij} \sim N(0, \sigma^2 \epsilon I_T) \) is a \( T \times 1 \) i.i.d. Gaussian error term with unknown variance \( \sigma^2 \epsilon \) that follows an inverse Gamma prior distribution. One can also assume an autoregressive structure for the error term (Penny et al. (2005)). In order to make the inference of model (2.4) computationally tractable, we propose
a key additive two-level model for the $\beta_{ijk}$'s:

$$
\beta_{ijk} = \alpha_{ik} + \lambda_{jk}, \quad \text{for } k = 1, \ldots, K,
$$

(2.2)

that is, we assume each of the subject-voxel specific effect $\beta_{ijk}$ is the sum of two independent terms: a subject-specific effect $\alpha_{ik}$ and a voxel-specific effect $\lambda_{jk}$. Moreover, it is reasonable to assume that the subject-specific effects do not vary by stimulus, i.e., $\alpha_{ik} = \alpha_i$. Then model (2.2) becomes

$$
\beta_{ijk} = \alpha_i + \lambda_{jk}, \quad \text{for } k = 1, \ldots, K.
$$

(2.3)

And the simplified model is

$$
y_{ij} = X\lambda_j + \alpha_i X1_K + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2_i I_T), \quad i = 1, \ldots, N; \quad j = 1, \ldots, J.
$$

(2.4)

where $\lambda_j = (\lambda_{j1}, \ldots, \lambda_{jK})'$, and $1_K = (1, \ldots, 1)'_K$. The additive assumption substantially reduces the number of parameters to be estimated ($NJK$ to $N + JK$), which is crucial for implementation in the massive neuroimaging data.

### 2.2 Spatial smoothing of voxel-specific effects via Ising prior

We assume a standard i.i.d. Gaussian prior for the subject-specific random effects, $\alpha = (\alpha_1, \ldots, \alpha_N)$: $\alpha \sim N(0, \sigma^2_\alpha I_N)$, and $\sigma^2_\alpha$ follows an inverse Gamma prior. More elaborated modelling will be considered in the next subsection. In fMRI experiments, usually only a small fraction of brain voxels are activated by the stimulus. Therefore, it is desirable to impose sparse prior on the voxel-specific random effects, $\lambda_j = (\lambda_{j1}, \ldots, \lambda_{jK})'$. This consideration motivates us to impose a “spike and slab” type of mixture prior (George and McCulloch (1993); Smith and Kohn (1996); Clyde and Parmigiani (1998)) for each $\lambda_j$. That is, we define a latent indicator $\gamma_j$ for each $\lambda_j$ with the conjugate setup for the variance (George and McCulloch (1997)):

$$
\lambda_j | \gamma_j = (1 - \gamma_j)\delta_0 I_K + \gamma_j N(0, \nu^2 \sigma^2_\alpha I_K),
$$

(2.5)
where $\delta_0$ is a point mass at 0 and $v$ is a pre-fixed hyperparameter. Note that above we can also assume different indicator $\gamma_{ik}$ independently for different $k$ and the discussion below will be similar for each $k$. For simplicity, our discussion here follows model (2.5). A key piece of information in fMRI data is the spatial structure among voxels. Formally, we can view the voxels along with their voxel-specific random effects, $\lambda = (\lambda_1, ..., \lambda_J)'$ lie on a 3-dimensional lattice. Then let $E = \{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq J\}$ be the set of all the neighboring pairs of voxels. Given $E$, let $a = (a_1, \ldots, a_J)'$ be a vector and $Q = (q_{j_1,j_2})_{J \times J}$ be a symmetric matrix of real numbers where $q_{j_1,j_2} = 0$ for all $(j_1, j_2) \notin E$. Then, we can assume an Ising prior distribution for $\gamma = (\gamma_1, ..., \gamma_J)'$ as proposed in Li and Zhang (2010):

$$\Pr(\gamma) = \exp(a'\gamma + \gamma'Q\gamma - \psi(a, Q)),$$

where $\psi(a, Q)$ is the normalizing constant: $\psi(a, Q) = \log\left(\sum_{\gamma \in \{0,1\}^J} \exp(a'\gamma + \gamma'Q\gamma)\right)$. If $Q = 0$, then $\psi(a, Q) = \sum_{j=1}^J \log(1 + e^{a_j})$, but in general there is no closed form for $\psi$. Similar priors were proposed in single-level GLM analysis by Smith et al. (2003) and Smith and Fahrmeir (2007).

In the Ising prior (2.6), the hyperparameter $a$ controls the sparsity of $\gamma$, and the entries in $Q$ control the smoothness of $\gamma$ over $E$ given $a$. Usually we do not want to favor a priori the inclusion of any voxels and this can be achieved by letting $a = a1_J$, where $1_J = (1, 1, \ldots, 1)' \in \mathbb{R}^J$. The hyperparameters $\{q_{j_1,j_2}\}$ represent the prior belief on the strength of coupling between the pairs of neighbors $(j_1, j_2)$, and it controls the smoothness of $\gamma$ over $E$ given $a$. Larger $q_{j_1,j_2}$ means tighter coupling. When $Q = 0$, the prior is back to i.i.d. Bernoulli. Without specific prior information of the strength of connection between each pair of neighbors, it is natural to assume $q_{j_1,j_2}$ to be constant. Then $(a, Q)$ reduce to two hyperparameters $(a, q)$, which can be either pre-fixed or assumed to follow certain hyperpriors.
2.3 A joint model for fMRI and behavioral/SNP data

In this section, we further propose a regression model to elaborate the relationship between the subject-specific effect $\alpha_i$ in (2.3) and the demographic and behavioral measurements as follows,

$$\alpha_i = Z'_i \zeta + e_i = \sum_{l=1}^{L} Z_{il} \zeta_l + e_i, \quad (2.7)$$

where $Z_i = (Z_{i1}, ..., Z_{iL})$ are the $L$ demographic and behavioral measurements of subject $i$, $\zeta = (\zeta_1, ..., \zeta_L)'$ are the corresponding coefficients, and $e_i$ is an error term and $e_i \sim N(0, \sigma^2_e)$. We assume a normal prior for the coefficients $\zeta$: $\zeta \sim N(\zeta_0, \sigma^2_\zeta I_L)$.

We propose two approaches to combine the inferences from the GLM (2.4) and model (2.7). The first approach is a two-stage procedure: first obtain a summary statistic, e.g., the posterior mode or mean $\tilde{\alpha}_i$ of the posterior draws of $\alpha_i$'s under the model introduced in the previous subsection; then take $\tilde{\alpha}_i$ as the response in model (2.7) with demographic and behavioral predictors, and conduct inference with linear model. Computationally less demanding, this approach is in fact a shortcut to the full Bayesian approach introduced below, where the models (2.4) and (2.7) are unified under one model. Notice that these two models together with the additive assumption (2.3), imply the following model:

$$y_{ij} = X\lambda_j + (Z'_i \zeta)X1_k + X1_k e_i + \epsilon_{ij}, \quad (2.8)$$

where $X1_k e_i + \epsilon_{ij} \sim N(0, \sigma^2_e X1_k 1_k' X' + \sigma^2_e I_T)$ can be viewed as a combined error term. Similar two-level model was proposed by Friston et al. (2002), but there are two major differences. First, here we introduce an explicit additive model that separates voxel- and subject-specific effects on imaging phenotypes; second, we simultaneously make inference on the subject-specific effects with behavioral predictors.
In the case of high dimensional covariates, e.g. SNP data, we can further impose sparse constraint by using a "spike and slab" prior on $\zeta_i$,

$$\zeta_i|\xi_i = (1 - \xi_i)\delta_0 + \xi_i N(0, \sigma^2_{\xi_i})$$

where $\delta_0$ is a point mass at 0.
3

Posterior inference

3.1 Gibbs sampler for the spatial model

Posterior distributions of the parameters $\alpha, \lambda, \gamma, \sigma_\epsilon, \sigma_\alpha$ can be obtained via Gibbs sampling (Gelfand and Smith (1990)), where parameters are sampled iteratively from their posterior distribution conditional on the rest of the parameters. Depending on specific study goals, various Gibbs samplers can be designed based on integrating out different selected subset of parameters. Below we present the posterior conditional distributions for the key parameters in a complete Gibbs sampler, where each and every parameter is sampled iteratively. The conditional distributions of the variances $\sigma_\epsilon$ and $\sigma_\alpha$ are standard inverse Gamma and are omitted here. We use the notation "\(\cdot\)" to denote all the rest of the parameters.

1. Subject-specific effects $\alpha_i$: $\alpha_i|\cdot \sim N(\mu_{\alpha_i}, \tau^2_{\alpha_i})$, with $\tau^2_{\alpha_i} = (\sigma^{-2}_\epsilon J_1^T X' X 1_K + \sigma^{-2}_\alpha)^{-1}$, and $\mu_{\alpha_i} = \tau^2_{\alpha_i} \sigma^{-2}_\epsilon \sum^J_{j=1} 1'_K X'(y_{ij} - X \lambda_j)$.

2. Voxel-specific effect $\lambda_j = (\lambda_{j1}, ..., \lambda_{jK})'$: $\lambda_j|\cdot \sim (1 - \gamma_j)\delta_0 + \gamma_j N(\mu_{\lambda_j}, \Omega_{\lambda_j})$, with $\Omega_{\lambda_j} = \sigma^2_\epsilon (NX'X + v^{-2} I_K)^{-1}$ and $\mu_{\lambda_j} = \sigma^2_\epsilon \Omega_{\lambda_j} \sum^N_{i=1} X'(y_{ij} - \alpha_i X 1_K)$.
3. The indicator of the voxel-specific effects $\gamma$’s: Let $\gamma_{(-j)} = \{\gamma_l : l \neq j\}$; $I_{(-j)}$ be the set of indices $\{l : \gamma_l = 1, \text{ and } q_{lj} \neq 0\}$. The prior probability of $\gamma_j$ is

$$\Pr(\gamma_j|\gamma_{(-j)}) = \exp[\gamma_j(a + q \sum_{l \in I_{(-j)}} \gamma_l)]/[1 + \exp(a + q \sum_{l \in I_{(-j)}} \gamma_l)]. \quad (3.1)$$

By the Bayes’ rule, the posterior probability of $\gamma_j$ given data and other parameters is

$$\Pr(\gamma_j = 1|Y, -) = \frac{\Pr(\gamma_j = 1|\gamma_{(-j)})}{\Pr(\gamma_j = 1|\gamma_{(-j)}) + F_{-1}^{-1}\cdot \Pr(\gamma_j = 0|\gamma_{(-j)})}$$

where $F_j$ is the Bayes factor:

$$F_j = \frac{\Pr(Y|\gamma_j = 1, -)}{\Pr(Y|\gamma_j = 0, -)} = \exp \left\{ -\frac{1}{2}\sigma^2 \left[ N\lambda'_jX'X\lambda_j + 2 \sum_{i=1}^N \lambda'_jX'(y_i - \alpha_iX1_K) \right] + \frac{1}{2}K \sum_{j=1}^J \sigma_{\epsilon}^2 \right\},$$

4. Inverse Gamma posteriors for $\sigma$’s: $\sigma_{\epsilon}^{-2}|- \sim G(a_{\epsilon}, b_{\epsilon})$, with $a_{\epsilon} = a_{\epsilon_0} + \frac{1}{2}TNJ + \frac{1}{2}K \sum_{j=1}^J \gamma_j$ and $b_{\epsilon} = b_{\epsilon_0} + \frac{1}{2} \sum_{j=1}^J \gamma_j \lambda'_j \lambda_j + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^J (y_{ij} - X\lambda_j - \alpha_iX1_K)'(y_{ij} - X\lambda_j - \alpha_iX1_K)$; $\sigma_{\alpha}^{-2}|- \sim G(a_{\alpha}, b_{\alpha})$, with $a_{\alpha} = a_{\alpha_0} + N\frac{N}{2}$ and $b_{\alpha} = b_{\alpha_0} + N\frac{1}{2} \sum_{i=1}^N \alpha_i^2$.

3.2 Gibbs sampler for the spatial model with predictors for subject-specific effects

Similar as before, posterior inference of model (2.8) can also be based on Gibbs sampling, where parameters are iteratively sampled from their posterior distributions conditional on all the others. Below we present the posterior conditional distributions for parameters in a complete Gibbs sampler for model (2.8), where each and every parameter is sampled iteratively.

1. Coefficients for demographic and behavioral predictors associated with subject-
specific effects \( \xi \): \( \xi \mid - \sim N(\mu_\xi, \Omega_\xi) \), with

\[
\begin{align*}
\Omega_\xi &= (\sigma_\varepsilon^{-2} J_1' K' X' X K \times Z Z' + \sigma_\zeta^{-2} I_L)^{-1}, \quad \text{and} \\
\mu_\xi &= \sigma^{-2}_\varepsilon \Omega_\xi \{ \sum_{i=1}^{N} \sum_{j=1}^{J} 1'_K X'(y_{ij} - X \lambda_j) - e_i J_1' K' X' X K \} + \zeta_0 \}.
\end{align*}
\]

2. Errors associated with subject-specific effects \( e_i \): \( e_i \mid - \sim N(\mu_{e_i}, \tau^2_{e_i}) \), with

\[
\begin{align*}
\tau^2_{e_i} &= (\sigma_\varepsilon^{-2} J_1' K' X' X K + \sigma_\varepsilon^{-2})^{-1}, \quad \text{and} \\
\mu_{e_i} &= \tau^2_{e_i} \sigma^{-2}_\varepsilon \sum_{j=1}^{J} 1'_K X'(y_{ij} - X \lambda_j) - (Z_i' \xi) J_1' K' X' X K].
\end{align*}
\]

3. Voxel-specific effect \( \lambda_j = (\lambda_{j1}, ..., \lambda_{jK})' \): \( \lambda_j \mid - \sim (1 - \gamma_j) \delta_0 + \gamma_j N(\mu_{\lambda_j}, \Omega_{\lambda_j}) \), with

\[
\begin{align*}
\Omega_{\lambda_j} &= \sigma^2_\varepsilon (N X' X + v^{-2} I_K)^{-1}, \quad \text{and} \\
\mu_{\lambda_j} &= \sigma^{-2}_\varepsilon \Omega_{\lambda_j} \sum_{i=1}^{N} [X' y_{ij} - (Z_i' \zeta) X' X K - e_i X' X K].
\end{align*}
\]

4. The indicator of the voxel-specific effects \( \gamma_j \)'s: the posterior probability of \( \gamma_j \) given data and other parameters is

\[
Pr(\gamma_j = 1|Y, -) = \frac{Pr(\gamma_j = 1|\gamma(-j))}{Pr(\gamma_j = 1|\gamma(-j)) + F_j^{-1}\cdot Pr(\gamma_j = 0|\gamma(-j))}
\]

where \( F_j \) is the Bayes factor:

\[
F_j = \frac{Pr(Y|\gamma_j = 1, -)}{Pr(Y|\gamma_j = 0, -)} = \exp \left\{ -\frac{1}{2\sigma^2_\varepsilon} [N \lambda_j' X' X \lambda_j + 2 \sum_{i=1}^{N} \lambda_j'[X' y_{ij} - (Z_i' \zeta) X' X K - e_i X' X K] \right\},
\]

16
and

\[ Pr(\gamma_j | \gamma_{(-j)}) = \frac{\exp[\gamma_j(a + q \sum_{l \in I_{(-j)}} \gamma_l)]}{1 + \exp(a + q \sum_{l \in I_{(-j)}} \gamma_l)} \]

5. Inverse Gamma posteriors for \( \sigma \)'s: \( \sigma_j^{-2} | \sim G(a_\sigma, b_\sigma) \), with

\[
a_\sigma = a_{\sigma 0} + \frac{1}{2} TNJ + \frac{1}{2} K \sum_j \gamma_j \\
b_\sigma = b_{\sigma 0} + \frac{1}{2} \sum_j \gamma_j \lambda_j' \lambda_j
\]

\[
+ \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^J \{ [y_{ij} - X\lambda_j - (Z'\zeta)X1K - e_iX1K]' x [y_{ij} - X\lambda_j - (Z'\zeta)X1K - e_iX1K] \};
\]

\[
\sigma_j^{-2} | \sim G(a_\sigma, b_\sigma), \text{ with } a_\sigma = a_{\sigma 0} + N/2 \text{ and } b_\sigma = b_{\sigma 0} + \frac{1}{2} \sum_i^N e_i^2;
\]

\[
\sigma_{\zeta}^{-2} | \sim G(a_\zeta, b_\zeta), \text{ with } a_\zeta = a_{\zeta 0} + L/2 \text{ and } b_\zeta = b_{\zeta 0} + \frac{1}{2} (\zeta - \zeta_0)'(\zeta - \zeta_0)
\]

To complete our spatial BVS model, I set \( a_{\sigma 0} = 1, b_{\sigma 0} = 0.1, a_{\zeta 0} = 1, b_{\zeta 0} = 0.1 \), resulting in vague or non-informative priors for the Inverse-Gamma distributions. These values generate densities that place large probability masses on large variance.

The rationale for using these vague priors is to ensure that the information in the data primarily governs the results. However, more informative priors may be employed when precise prior information is available. Note that the variance were very well estimated in our simulations, and our results were not sensitive to the choice of either diffuse or more informative priors.

3.3 Incorporating anatomical prior information

For fMRI studies, anatomical prior information can significantly facilitate data analysis and promote posterior inference. Local physiological responses upon neuronal
activation are known to be restricted to cortical and subcortical neuronal assemblies. BOLD responses associated with local increases in blood flow and volume are also mainly restricted to the grey matter and do not occur in white matter or ventricles. There are, however, some spatial inaccuracies at the cortical edges due to BOLD increases in draining veins and susceptibility-induced spatial misregistration (Logothetis (2008)). Thus, the spatial information about the distribution of cortical and subcortical gray matter may be helpful.

To incorporate anatomical prior information, define a grey matter indicator, \( g_j = 1 \) if voxel \( j \) is grey matter and \( g_j = 0 \) otherwise. Grey matter prior \( p(g_j = 1) \) encodes the certainty of whether a voxel is grey matter or not, and \( 0 < p(g_j = 1) < 1 \) at the presumed cortical edges. Assume also that the grey matter indicators are a priori independent across voxels, so that \( p(g_1, \ldots, g_J) = \Pi_{j=1}^J p_j \). Then it is known that \( p(\gamma_j = 1 | g_j = 0) = 0 \), because activation cannot occur outside grey matter. This incorporation would significantly reduce the model space for our spatial BVS model.
Simulation study

Simulation study is performed to assess the performance of spatial BVS model as compared to the non-spatial BVS model. Simulation is done using the hemodynamic time series for our fMRI application as design matrix $X$, therefore, serves as a preparation and interpretational basis for the actual fMRI data in our application.

4.1 Duke Neurogenetics Study

Our application, the Duke Neurogenetics Study, is aimed to link brain imaging phenotype with behavioral phenotype and/or SNP genotype. In this study, BOLD fMRI data, behavioral questionnaire data and SNP data of multiple subjects are collected. For the fMRI data, a variation of block design is used.

During one fMRI run of 175 repetition time (TR) of 2 seconds, each subject completed a standardized protocol (Fig.4.1) comprised of two blocks of a fearful facial expression matching (condition 2) and two blocks of an angry facial expression matching (condition 3) interleaved with four blocks of a shape-matching control task (condition 1). During each TR, a 3D brain image of dimension $79 \times 95 \times 70$ was acquired, resulting in an enormous 3D image time series of length 175 per subject.
The facial matching tasks are considered to be repulsive emotional stimuli, while the shape matching task is considered as a neutral stimulus. Since the experimental design involves more than one stimulus, 3 time series are provided as design matrix ($175 \times 3$; Fig. 4.2) for our multi-subject additive spatial BVS model (2.8).

![Block design experimental scheme in Duke Neurogenetics Study](image)

**Figure 4.1:** Block design experimental scheme in Duke Neurogenetics Study

### 4.2 Simulation Scheme

The design matrix $X$ is a $3 \times 175$ matrix, with 175 time points ($T=175$) per hypothetical time series and 3 time series ($K=3$) to the three different experimental blocks (Fig. 4.2). To examine the neural populations activated by each individual task, data is simulated with individual response curve ($K=1$) for 100 subjects ($N=100$) and 100 voxels ($J=100$). Grid layout and neighbor information are depicted in Fig. 4.3. The number in each grid reflects the voxel number $j$. Neighboring voxels are defined as grids adjacent to the one of interest.

To investigate how the performance of the spatial model may be affected by signal strength of the voxel specific effect, four datasets are generated for each individual
response curve with the same set of subject specific effect $\alpha_i$'s but different sets of voxel specific effect $\lambda_j$'s, following model (2.4). Conditions for the three cases are summarized in Table 4.1. Briefly, for responsive voxels, $\lambda_j \sim N(\mu_L, \sigma_L^2)$. The simulated BOLD signal for each subject at each voxel at each time point is $y_{ijt} = \alpha_i X_{i1} K + X_t \lambda_{jk} + \epsilon$, where $\epsilon \sim N(0, \sigma^2_e)$ is the error term associated with experimental noise in MRI readings. We assume similar experimental conditions for all subjects throughout the fMRI experiment in our application, thus constant noise. For all of our simulated dataset, $\sigma_e$ remains the same.

Fig. 4.4 depicts the simulated response (time series $Y_{ij}$) for an activated voxel from three simulated datasets for the three individual response curve (left panels), along with the corresponding $X_k$'s (right panels). These datasets are simulated from case 3. The responses are from the same subject, of voxel with the strongest amplitude.

Figure 4.2: Design matrix for Duke Neurogenetics Study
Table 4.1: Parameters to generate simulated data

<table>
<thead>
<tr>
<th></th>
<th>(\sigma_e)</th>
<th>(\mu_L)</th>
<th>(\sigma_L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>20</td>
<td>5.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Case 2</td>
<td>20</td>
<td>2.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Case 3</td>
<td>20</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Case 4</td>
<td>20</td>
<td>0.1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

of activation in each simulation.

Fig. 4.5 depicts the simulated response (time series \(Y_{ij}\)) for three different voxels from the same subject in Case 3 for response curve X2 (Fig.4.4(b)). The strongly activated voxel (a) has a voxel-specific effect \(\lambda_j = 0.91\) (same as in Fig.4.4c), the weakly activated voxel (b) has a voxel-specific effect \(\lambda_j = 0.37\), and the unactivated voxel (c) has no voxel-specific effect. This figure illustrates the low signal-to-noise ratio in individual voxel response in our simulation, which is comparable to actual fMRI data (Smith and Fahrmeir (2007)).
Figure 4.4: Example MRI reading $Y_{ij}$ for three response curves; case 3
Figure 4.5: Simulated MRI reading $Y_{ij}$ from the same subject for three voxels that are Strongly Activated (a), Weakly Activated (b), and Unactivated (c). Response curve X2, case 3.
4.3 Selection of hyper-parameters

There are three important hyper-parameters in our multi-subject spatial BVS model, and the choice of these hyper-parameters have implications in the model.

Hyperparameter $v$ encodes our prior knowledge about the variance of non-zero voxel specific effect $\lambda$ (Eq. 2.5), and links that to the variance of response $y_{ijt}$. A small $v$ reflects strong prior knowledge on the precision of $\lambda$’s. A flat prior (large $v$) becomes improper and MCMC becomes unbounded. In MCMC, hyperparameter $v$ greatly influence how far away $\lambda_j$ may deviate from its MLE in an iteration. A smaller $v$ leads to a smaller step on average, and more iterations for the MCMC to converge if initial values are far from the stationary distribution. Choice of $v$ depends on the data, and multiple values of $v$ were tested for each simulated dataset.

Hyperparameter $a$ encodes our knowledge about the percentage of non-zero voxels (or the prior probability of $\gamma_j = 1$) when hyperparameter $q = 0$ (Eq. 2.6 and 3.1). From Eq. 3.1, it is trivial to see that the prior probability of $\gamma_j = 1$ is positively correlated with hyperparameter $a$. Multiple values of hyperparameter $a$ were tested in one simulated dataset, and I found that the choice of hyperparameter $a$ does not significantly influence the convergence speed of MCMCs nor the results.

Hyperparameter $q$ is an important parameter in the Ising priors (Eq. 2.6). The term $q \sum_{l \in I_{(-j)}} \gamma_l$ is the interaction effect of the elements of $\gamma_l$ for all pairwise neighboring sites. Hyperparaneter $q$ controls the strength of spatial smoothing; when $q = 0$, the elements of $\gamma_l$ are independent and hence the non-spatial BVS model. Together with hyperparameter $a$, hyperparameter $q$ also influence the marginal probability of the percentage of non-zero voxels. Multiple values of hyperparameter $q$ are tested in this simulation.

For the Inverse-Gamma distributions, conventional non-informative priors are used, $a_{\alpha_0} = 1, b_{\alpha_0} = 0.1, a_{\beta_0} = 1, b_{\beta_0} = 0.1$. These values generate densities that
place large probability masses on large variance. When precise prior information is available, more informative priors can be employed. Our results are not sensitive to the choice of either diffuse or more informative priors. For example, I also used the values \( a_{\alpha 0} = 1, b_{\alpha 0} = 1, a_{\epsilon 0} = 1, b_{\epsilon 0} = 1 \) in one simulation, and the results were essentially identical to those obtained with the vague priors.

4.4 Convergence Diagnostic

2-3 MCMCs with different initial values are run for each models. Convergence of the Gibbs sampling algorithm is determined with multiple lines of evidence. 1) Behavior of the subject-specific effect \( \alpha_i \)’s. Three randomly chosen \( \alpha_i \) were monitored with trace plot. After 5,000-20,000 burnt-in iterations, \( \alpha_i \)’s converged to their true values; additional 30,000-50,000 iterations were run and used for subsequent analysis. 2) Behavior of the observed variance for response \( y_{ijt}, \sigma_\epsilon \). \( \sigma_\epsilon \) was monitored with trace plot and its convergence behavior is similar to those of \( \alpha_i \)’s. 3) Likelihood ratio (LR) statistics. The observed data likelihood is saved for each iteration. The LR statistics is based on the observed-data likelihood in the iteration \( L_i \) and the observed-data maximum likelihood calculated with the true values of parameters \( \hat{L} \). LR statistics is given by \(-2 \log(L_i/\hat{L})\). The distribution of LR statistics of each MCMC after apparent convergence (as estimated by \( \alpha_i \)’s) is not significantly different from \( \chi^2 \) distribution (two-sided Kolmogorov-Smirnov Test (ks.test in R), \( p > 0.05 \)). 4) Results from different MCMC are similar to each other after convergence. Therefore, iterations from multiple MCMC after convergence are pooled together to assess the performance of the spatial BVS model.

4.5 Performance of the spatial BVS model

To assess the performance of the spatial model, the posterior inclusion probability \( P(\gamma_j = 1|Y) \) is examined under the spatial and the non-spatial BVS model. To
Figure 4.6: ROC curves for individual response curve, case 3

better visualize and summarize the comparison between models, I compute the ROC curve as follows: only those covariates $\gamma_j$ with $P(\gamma_j = 1|Y)$ greater than a threshold are deemed positives, and those below the threshold are deemed negatives, then the ROC curve reflects the pair of (true positive rate, false positive rate) achieved by varying the calling threshold. The bigger area under the ROC curve (maximum 1), the better the discriminating power of the model.

Fig.4.6 is the ROC curves for the three simulated data set in case 3 of individual response curves. For response curve X1, multiple values of hyperparameter q are examined to determine the necessary degree of spatial smoothing (Fig.4.6a). All four spatial models ($q = 0.1, 0.25, 1, 2.5$) outperformed the non-spatial model ($q=0$) for response curve X1. For response curve X2 and X3, spatial model with $q = 1$ is compared with non-spatial model; spatial BVS models is comparable to (Fig.4.6b) or performs better than (Fig.4.6c) the non-spatial BVS model.

To better illustrate the performance of the spatial BVS model, posterior amplitude maps (Fig.4.7) are generated for simulation case 2 with response curve X1. The ROC curves for the spatial and non-spatial models are presented in the top panel. The true activation amplitude is depicted in the middle left panel. The posterior activation amplitudes for the non-spatial and spatial BVS models with cutoff criteria
$p(\gamma_i = 1|Y) > 0.175$ are depicted in the lower panel. Compared to the non-spatial BVS model, the spatial model thins out incorrectly identified voxels in isolation.

It is intuitive that model performance is correlated with signal strength. To compare model performance under different signal strength (four cases of simulations per response curve), I used the area under the ROC curve as a one-dimension summary statistics for model performance. Results from the spatial BVS model presented were obtained with $q = 1$. From Table 4.2, it is obvious that the spatial BVS model outperforms the non-spatial BVS model only when the signal is relatively weak. When signal strength is strong (i.e., case 1), data from multiple subject is sufficient to identify activated voxels in the non-spatial BVS model and spatial smoothing becomes not necessary. With weak signal, spatial smoothing greatly increase model performance.

Table 4.2: Performance of spatial BVS models (Area under ROC curve) under different signal strength $\mu_L$ for individual response curve

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th></th>
<th>X2</th>
<th></th>
<th>X3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-spatial</td>
<td>spatial</td>
<td>non-spatial</td>
<td>spatial</td>
<td>non-spatial</td>
</tr>
<tr>
<td>case 1</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>case 2</td>
<td>0.73</td>
<td>0.81</td>
<td>0.71</td>
<td>0.71</td>
<td>0.73</td>
</tr>
<tr>
<td>case 3</td>
<td>0.53</td>
<td>0.67</td>
<td>0.60</td>
<td>0.60</td>
<td>0.39</td>
</tr>
<tr>
<td>case 4</td>
<td>0.55</td>
<td>0.53</td>
<td>0.47</td>
<td>0.45</td>
<td>0.48</td>
</tr>
</tbody>
</table>
Figure 4.7: ROC curves and posterior amplitude maps; response curve X1, case 2
In this thesis, a multi-subject spatial Bayesian variable selection model is proposed for the analysis of multi-subject fMRI data. This model allows for simultaneous estimation of activation probability and amplitude of activation for all voxels and estimation of subject-specific predictors, solving the multiple testing problem in fMRI data analysis. Further, by incorporating spatial information through an Ising prior, the sensitivity in detecting activated voxels is increased, and the rate of false-positive is decreased.

A second important feature of the Bayesian approach is the ability to use prior information to inform the activation and amplitude maps. This is undertaken in a structured way using a prior distribution that, when combined with the information in the data in the form of the likelihood, informs the resulting posterior probabilities of activation. This allows for the possibility to use information from previous multi-subject studies to define anatomically informed priors for a new subject. This flexibility in assigning prior probabilities for activation is a distinct advantage of Bayesian method that is hard to replicate in non-Bayesian paradigms.

A third important feature of the Bayesian approach is the ability to perform
flexible inferences. The MCMC estimation procedures produce samples from the joint posterior distribution of all of the model parameters, which facilitates estimation of and inferences about functions of the model parameters. For example, by examination of the joint inclusion probability of voxel pairs or voxel groups in the posterior samples, it is possible to identify task-related functional connectivity while estimating task-specific activations.

The major advance of this thesis work is the proposal of multi-subject additive effect with spatial smoothing in fMRI data analysis. It is the first model to use spatial BVS framework for the analysis of multi-subject fMRI data. The proposed additive assumption substantially reduces the number of parameters to be estimated, and opens the possibility of high-dimensional variable selection for both the fMRI phenotype and subject-specific predictors such as genotype (SNPs).

One major drawback of the Bayesian approach is that it is computationally demanding. For my simulation of 100 voxels, an MCMC with 60,000 iterations takes more than six hours in R, with matrix calculation (esp. matrix inverse) as bottleneck. In order to analyze whole brain fMRI data (∼100,000 voxels), or fMRI data slice-by-slice (∼6,000 voxels/ slice), efficient computation in other language is warranted.

Another drawback of this thesis work is the lack of application to actual fMRI data. To compensate for this drawback, I used the the hemodynamic time series for the fMRI data in Duke Neurogenetics Study as design matrix $X$. Simulations performed in this thesis can then serve as a preparation and interpretational basis for analyzing the fMRI data in Duke Neurogenetics Study.
Appendix A

R codes for Spatial Bayesian variable selection

A.1 Parameter calculation

\[
XX = \text{t}(X)^\%*\%X;
\]

\[
YA=\text{array}(Y, \text{dim}=c(T, J, N));
\]

calculate.tausquare.alpha <- function(XX, phi.epsilon, phi.alpha)
{
  tausquare = 1/(phi.epsilon*J*sum(XX)+phi.alpha);
  return(tausquare);
}

calculate.mu.alpha
<- function(tausquare.alpha, phi.epsilon, YA, X, XX, lambda)
{
  sumlambda = apply(as.matrix(lambda), 1, sum);
  sumXlambda = as.vector(X%*%sumlambda);
  sumyj = apply(YA, c(1,3), sum);
}
Xk = apply(X, 1, sum);
sum = Xk%*%(sumyj-sumXlambda);
mu.alpha = tausquare.alpha*phi.epsilon*as.vector(sum);
return(mu.alpha)
}

calculate.omega.lambda.j <- function(phi.epsilon, XX)
{
  temp = solve(N*XX+diag(K)/v^2)/phi.epsilon;
  return(temp)
}

calculate.mu.lambda <- function(omega.lambda.j, phi.epsilon, X, XX, YA, alpha)
{
  sumyi = apply(YA, c(1,2), sum);
  sum1 = t(X) %*% sumyi;
  sum2 = apply(XX,2,sum)*sum(alpha);
  mu.lambda = omega.lambda.j%*%(sum1-sum2)*phi.epsilon;
  return(mu.lambda)
}

calculate.F.j <- function(phi.epsilon, lambda.j, XX, X, YA.j, alpha)
{
  sum1 = 2*sum(t(lambda.j)%*%t(X)%*%YA.j);
  sum2 = 2*t(lambda.j)%*%apply(XX,2,sum)*sum(alpha);
  temp3 = N*t(lambda.j)%*%XX%*%lambda.j;

\[ F = \exp(-\phi \epsilon / 2 \times (\text{temp3} - \text{sum1} + \text{sum2})); \]
return(F)
}

calculate.Pr.j <- function(phi.epsilon, lambda.j, XX, X, YA.j, alpha, gamma, gamma.ne, j, hyperq)
{
F.j = calculate.F.j(phi.epsilon, lambda.j, XX, X, YA.j, alpha);
neighbor = gamma.ne[j];
neighborin = neighbor[neighbor>0];
sum.ne.j = sum(gamma[neighborin]);
Pr0.j = 1/(1+exp(a+hyperq*sum.ne.j));
Pr1.j = 1-Pr0.j;
Pr.j = Pr1.j / (Pr1.j + Pr0.j/F.j);
return(Pr.j);
}

A.2 Gibbs Sampler

gibbs.alpha <- function(XX, phi.epsilon, phi.alpha, YA, X, lambda)
{
tausquare.alpha =
calculate.tausquare.alpha(XX, phi.epsilon, phi.alpha);
mu.alpha =
calculate.mu.alpha(tausquare.alpha, phi.epsilon,YA, X, XX, lambda);
alpha = rnorm(N, mu.alpha, sqrt(tausquare.alpha));
return(alpha)
}
gibbs.gamma <- function(phi.epsilon, lambda, XX, X, YA, alpha, gamma, gamma.ne, hyperq) {
  omega.lambda.j = calculate.omega.lambda.j(phi.epsilon, XX);
  mu.lambda =
    calculate.mu.lambda(omega.lambda.j, phi.epsilon, X, XX, YA, alpha);
  Pr=rep(0, J);
  for (j in 1:J) {
    YA.j = YA[,j,];
    lambda.j = lambda[,j,];
    Pr.j = calculate.Pr.j(phi.epsilon, lambda.j, XX, X, YA.j,
      alpha, gamma, gamma.ne, j, hyperq);
    Pr[j] = Pr.j;
    gamma[j] = rbinom(1, 1, Pr.j);
    mu.lambda.j=mu.lambda[,j,];
    lambda[,j,] = gamma[j]*rmvnorm(1, mu.lambda.j, omega.lambda.j)
      + (1-gamma[j])*rep(0, K);
  }
  return(list(lambda=lambda, gamma=gamma))
}

gibbs.phi.alpha <- function(alpha) {
  a.alpha = a.alpha0 + N/2;
  b.alpha = b.alpha0 + sum(alpha^2)/2;
phi.alpha = rgamma(1, shape=a.alpha, rate=b.alpha);
return(phi.alpha)
}

gibbs.phi.epsilon <- function(Y, YA, X, XX, alpha, lambda, gamma)
{
a.epsilon = a.epsilon0 + T*N*J/2 + sum(gamma)*K/2;
sum1 = sum(lambda^2)/v^2*0.5;
Xlambda = X%*%lambda;
NXlambda = array(rep(Xlambda,N), c(T, N*J));
Jalpha = rep(alpha, each=J);
X1K = apply(X, 1, sum);
Yresidual = Y - NXlambda - X1K%*%t(Jalpha);
RSS = sum(Yresidual^2);
b.epsilon = b.epsilon0 + sum1 + RSS/2;
phi.epsilon = rgamma(1, shape=a.epsilon, rate=b.epsilon);
logL = log(phi.epsilon)*N*J*T-phi.epsilon/2*RSS;
return(list(phi.epsilon=phi.epsilon, logL=logL))
}

gibbs1 <- function(para, Y, YA, X, XX, a, hyperq, v)
{
phi.alpha = para$phi.alpha;
phi.epsilon = para$phi.epsilon;
alpha = para$alpha;
lambda = para$lambda;
gamma = para$gamma;
alpha = gibbs.alpha(XX, phi.epsilon, phi.alpha, YA, X, lambda);
temp.gibbs = gibbs.gamma(phi.epsilon, lambda, XX, X, YA,
alpha, gamma, gamma.ne, hyperq);
lambda = temp.gibbs$lambda;
gamma = temp.gibbs$gamma;
phi.alpha = gibbs.phi.alpha(alpha);
phi.epsilon =
gibbs.phi.epsilon(Y, YA, X, XX, alpha, lambda, gamma)$phi.epsilon
logL = gibbs.phi.epsilon(Y, YA, X, XX, alpha, lambda, gamma)$logL
return(list(alpha=alpha, lambda=lambda, gamma=gamma,
phi.alpha=phi.alpha, phi.epsilon=phi.epsilon, logL=logL))

}
logLs = rep(0, num);
for(d in 1:(biter+riter))
{
    para = gibbs1(para, Y, YA, X, XX, a, hyperq, v);
    if(d>biter)
    {
        temp = d-biter;
        if(temp%%sparse==0)
        {
            count = temp/sparse;
            tag = count%%num;
            if(tag==0){tag=num};
            alphas[tag,] = para$alpha;
            lambdas[tag, ,] = para$lambda;
            gammas[tag,] = para$gamma;
            phi.alphas[tag] = para$phi.alpha;
            phi.epsilons[tag] = para$phi.epsilon;
            logLs[tag] = para$logL;
            if(count%%num==0)
            { temp2=seq(1, num, by=1);
                if(count/num==1)
                {
                    write.table(alphas[temp2,], file=paste("alpha", name), sep=" ");
                    write.table(gammas[temp2,], file=paste("gamma", name), sep=" ");
                    write.table(lambdas[temp2,,], file=paste("lambda", name), sep=" ");
                    write.table(phi.alphas[temp2],
                    file=paste("phi_alpha", name), sep=" ");
                }
            }
        }
    }
}
write.table(phi.epsilons[temp2],
    file=paste("phi_epsilon", name), sep=" ");
write.table(logLs[temp2], file=paste("logL", name), sep=" ");
}
else
{
write.table(alphas[temp2,,],file=paste("alpha",name),
    sep=" ", append=TRUE);
write.table(gammas[temp2,,],file=paste("gamma",name),
    sep=" ", append=TRUE);
write.table(lambdas[temp2,,],file=paste("lambda",name),
    sep=" ", append=TRUE);
write.table(phi.alphas[temp2],file=paste("phi_alpha",name),
    sep=" ", append=TRUE);
write.table(phi.epsilons[temp2],file=paste("phi_epsilon",name),
    sep=" ", append=TRUE);
write.table(logLs[temp2],file=paste("logL",name),sep=" ",append=TRUE);
}
}
}
}
}
}
return(list(alphas=alphas, lambdas=lambdas, gammas=gammas,
    phi.alphas=phi.alphas, phi.epsilons=phi.epsilons, logLs=logLs))
}
Bibliography


Biography

Ying Yang was born on December 15, 1982 and grew up in Guangzhou, Guangdong, China. Her parents are Wenhui Yang and Weiping Ye, who currently live in Guangzhou, Guangdong, China. Ying did her undergraduate work at Peking University, Beijing, China, where she majored in Biological Sciences and minored in Economics. After completing her Bachelors degree in June of 2005, she came directly to Duke University to study Neurobiology. During her graduate studies, she became interested in Statistics and studied in the department of Statistical Sciences. She completed her Ph.D degree in Neurobiology and her Master degree in Statistics in July of 2011.