BAYESIAN PARAMETRIC AND NONPARAMETRIC METHODS FOR MULTIPLE QTL MAPPING AND SNP-SET ANALYSIS

Wonil Chung

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2013

Approved by:

Fei Zou

Joseph G. Ibrahim

Fred A. Wright

Wei Sun

Tim Wiltshire

ABSTRACT

Wonil Chung: Bayesian Parametric and Nonparametric Methods for Multiple QTL Mapping and SNP-Set Analysis
(Under the direction of Professor Fei Zou)

Many complex traits and human diseases, such as blood pressure and body weight, are known to change over time. The genetic basis of such traits can be better understood by repeatedly collecting data over time. The resulting longitudinal data provide us useful resources for studying the joint action of multiple time-dependent genetic factors. In the first part of the dissertation, we extend two existing Bayesian multiple quantitative trait loci (QTL) mapping methods from univariate traits to longitudinal traits. Our first approach focuses on mapping genes with main effects and two-way gene-gene and gene-environment interactions. Multiple QTL are selected by a variable selection procedure based on the composite model space framework. Our second approach presents a Bayesian Gaussian process method to map multiple QTL without restricting to pairwise interactions. Rather than modeling each main and interaction term explicitly, the nonparametric Bayesian method measures the importance of each QTL, regardless whether it is mostly due to a main effect or some interaction effect(s), via an unspecified function. We assign a Gaussian process prior to this unknown function. For the unstructured covariance matrix, both approaches employ a modified Cholesky decomposition. For data where phenotype measurements are not collected at a fixed set of time points across all samples, we propose a grid-based approach which parsimoniously approximates the covariance matrix of each subject as a function of a covariance matrix defined on a set of pre-selected time points.

For most genome-wide association studies (GWAS), power to detect an association

between a single genetic variant, such as a single nucleotide polymorphism (SNP) and a complex trait is extremely low. Alternative strategies, such as regional SNP-set analysis have overcome some of the limitations of the standard single SNP analysis. Our third topic develops a Bayesian regional SNP-set analysis which extends the nonparametric Gaussian process model and simultaneously models multiple groups of rare and/or common SNP variants. Instead of assigning each SNP a hyperparameter, we assign a common hyperparameter to every SNP within each set to measure the cumulative effect of all SNPs in that set.

ACKNOWLEDGEMENTS

First of all, I would like to express my deep sense of gratitude to my advisor, Dr. Fei Zou, for her guidance and support throughout the period of my doctoral research. Her continuous encouragement and generous financial support have enabled me to accomplish my dissertation work. For more than four years, I have worked with her as a research assistant and have written my dissertation under the direction of her. Without her insightful advise and persistent help, this thesis would not have been possible.

I would like to give sincere thanks to my committee members: Dr. Joseph G. Ibrahim, Dr. Fred A. Wright, Dr. Wei Sun and Dr. Tim Wiltshire. My special thanks to Dr. Joseph G. Ibrahim for his extremely helpful insights and comments. I worked with him as a research assistant at LCCC in the first year and involved interesting research projects. I am deeply grateful to Dr. Fred A. Wright for giving me the opportunity to participate in the challenging project. From the project, I had learned how to analyze large human genomic data sets and how to communicate with other researchers. I would like to show my gratitude to Dr. Wei Sun for his valuable advice and helpful suggestions. I truly enjoyed working with him and studied how to analyze eQTL data and effectively present the results. I really appreciate Dr. Tim Wiltshire for his kind and valuable comments. I worked with him on my first genetic project at UNC and acquired information on how to analyze large mouse data sets and utilize statistical genetics software.

I would like to show special thanks to all my friends with whom I spent my Ph.D. years for their consistent help, advice and encouragement. Last, I would like to dedicate

this dissertation to all of my family members: my father Soonchik Chung, my mother Kimyoung Nam, my sisters Joo-Yeon Chung, Hye-Yeon Chung and my brother Won Seok Jung who have always supported and encouraged me with their endless patience and unconditional love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Statistical Methods for QTL Mapping

QTL Mapping

Quantitative traits are defined as phenotypes which vary in degree and usually determined by both genetic and environmental factors. There are three types of quantitative traits: continuous traits in which there is only a gradual change from one phenotype to another with no clear categories such as height, weight and blood pressure; meristic traits in which the range of phenotypes can be expressed by counting such as numbers of offsprings; threshold traits in which there are a small number of phenotypic classes such as complex diseases. Quantitative Trait Loci (QTL) refer to genomic regions that affect variation in quantitative traits. Identification of QTL is important for understanding of genetic nature of quantitative trait variation. QTL mapping is to identify genetic regions that affect phenotypic variation of quantitative traits, including the number of QTL, their genomic positions and associated genetic effects including main effects, gene-gene interactions and gene-environment interactions. QTL mapping is often conducted using molecular markers such as amplified fragment length polymorphism (AFLPs), or more commonly single-nucleotide polymorphism (SNPs).

Single Gene Model

The basic quantitative genetic model partitions the total variance in quantitative traits into genetic variance and environmental variance. For individual i, $P_i = G_i + E_i$ where P_i is phenotypic value, G_i is genetic value and E_i is environmental effect. Suppose two inbred parents (P1 and P2) differ in some quantitative traits. At locus q, the allele of parent P1 is labeled as B_q and the allele of P2 as b_q. An F1 generation is completely heterozygous with genotype $B_q b_q$, receiving one allele from each parent. A BC population is generated when F1 is crossed back with P1 (or P2). At locus q, every BC individual has equal probability of 1/2 to be $B_q b_q$ and $B_q B_q$ (or $b_q b_q$). If the average phenotypic value of P1 and P2 is m, the expected genetic values of $B_q b_q$ and $B_q B_q$ (or $b_q b_q$) can be defined as $m + d_q$ and $m + a_q$ (or $m - a_q$) where d_q is the dominance effect and a_q is the additive effect. The genetic value for BC can be expressed as $G_i = \mu + c_q x_{iq}$ where $\mu = m + \frac{1}{2}(d_q + a_q)$ (or $m + \frac{1}{2}(d_q - a_q)$), and $x_{iq} = s_{iq} - \frac{1}{2}$ if s_{iq} denotes the number of allele b_q (or B_q). An F2 population is generated when F1 individuals are crossed with each other and each F2 individual has probability of 1/4,1/2 and 1/4 to be $b_q b_q$, $B_q b_q$ and $B_q B_q$, respectively. The expected genetic values of $b_q b_q$, $B_q b_q$ and $B_q B_q$ can be defined as $m - a_q$, $m + d_q$ and $m + a_q$. The genetic value for F2 can be modeled as $G_i = \mu + e_q x_{iq} + f_q z_{iq}$ where $\mu = m + \frac{1}{2} d_q$, $x_{iq} = s_{iq} - 1$ and $z_{iq} = (1 + x_{iq})(1 - x_{iq}) - \frac{1}{2}$ if s_{iq} denotes the number of allele b_q.

Genetic Model for Epistasis

Fisher [1918] first partitioned genetic variances into additive, dominance and epistatic variances based on the least-squares principle. Cockerham [1954] further partitioned the two-gene epistatic variance into four variance components corresponding to additive × additive, additive × dominance, dominance × additive and dominance × dominance. Mather [1967] proposed other epistasis models, and Crow et al. [1970], Mather et al.

[1977], Haley et al. [1992] and Kearsey et al. [1998] applied the F_∞-metric model to study epistasis. Goodnight [2001] adopted an alternative model modified from Cockerham [1954] to the study of gene-gene interaction. Among them, Cockerham's model is more appropriate than the other models for studying epistasis and mapping QTL in the populations, such as BC and F2 [Kao and Zeng, 2002]. For the commonly used Cockerham epistatic model, it is assumed that there are two alleles affecting the traits of interest. At two loci q and q', the genotypes of P1 and P2 are $B_qB_qB_{q'}B_{q'}$ and $b_qb_qb_{q'}b_{q'}$, and all F1 individuals have genotype $B_qb_qB_{q'}b_{q'}$. Each BC population has equal probability of 1/2 for being $B_qb_qB_{q'}b_{q'}$ and $B_qB_qB_{q'}B_{q'}$ (or $b_qb_qb_{q'}b_{q'}$). The genetic values for BC can be expressed as $G_i = \mu + c_q x_{iq} + c_{q'} x_{iq'} + c_{qq'} x_{iqq'}$ where $x_{iq} = s_{iq} - \frac{1}{2}$ and $x_{iqq'} = x_{iq}x_{iq'}$ if s_{iq} denotes the number of allele b_q (or B_q). Each F2 individual has probability of 1/4,1/2 and 1/4 for being $b_qb_qb_{q'}b_{q'}$, $B_qb_qB_{q'}b_{q'}$ and $B_qB_qB_{q'}B_{q'}$, respectively. The genetic value for F2 can be modeled as $G_i = \mu + e_q x_{iq} + f_q z_{iq} + e_{q'} x_{iq'} + f_{q'} z_{iq'} + i_{aa} w_{iaa} + i_{ad} w_{iad} + i_{da} w_{ida} + i_{dd} w_{idd}$ where $x_{iq} = s_{iq} - 1$, $z_{iq} = (1 + x_{iq})(1 - x_{iq}) - \frac{1}{2}$, $w_{iaa} = x_{iq}x_{iq'}$, $w_{iad} = x_{iq}z_{iq'}$, $w_{ida} = z_{iq}x_{iq'}$ and $w_{idd} = z_{iq}z_{iq'}$. In the above model, i_{aa}, i_{ad}, i_{da} and i_{dd} are the epistatic effects between loci q and q', called additive × additive, additive × dominance, dominance × additive and dominance × dominance effects, respectively.

Single QTL Mapping

The QTL data include the phenotype values y_i $(i = 1, ..., n)$, the marker genotype values M_{ij} $(i, ..., n, j = 1, ..., m)$ located at certain positions λ_j where n is the sample size and m is the number of makers. The genotypes at a putative QTL are denoted by $\{qq, Qq, QQ\}$ to distinguish the QTL genotypes from the marker genotypes $\{mm, Mm, MM\}$.

The single QTL model assumes that there is only one QTL which is associated with the trait of interest [Lander and Botstein, 1989]. If the genotypes of QTL are observed,

QTL mapping become a simple linear regression problem. For BC population, the model can be specified as

$$y_i = \mu + \beta x_i + e_i \ (i = 1, ..., n), \tag{1.1}$$

where μ is the overall mean; β is the genetic effect; x_i is 1/2 if individual i has Qq genotype and -1/2 if individual i has QQ genotype; e_i is a random error with $e_i \sim N(0, \sigma^2)$. A test can be performed on β under $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$. For F2 population, the model can be constructed to test additive and dominance effects separately as

$$y_i = \mu + \beta x_i + \gamma z_i + e_i \ (i = 1, ..., n), \tag{1.2}$$

where β is the coefficient for additive effect; γ is the coefficient for dominance effect x_i is 1 for qq, 0 for Qq and -1 for QQ; z_i is 1/2 for Qq, -1/2 for qq and QQ. Linear regression can be conducted to test $H_0 : \beta = \gamma = 0$. If the effect of a marker is tested to be significant, that marker is claimed to be associated with one or more QTL. Although this single marker analysis is simple and captures candidate QTL, it cannot tell whether the markers are linked to one or more QTL and it does not estimate the putative positions of the QTL.

In practice, the QTL position is rarely known and the genotypes of QTL are usually unobserved, leading to all missing x_is and z_is. To solve this problem, interval mapping was introduced by Lander and Botstein [1989]. At any putative QTL position located in an interval between two flaking markers (M_{ij}, M_{ij+1}) of the individual i, the probabilities of the unobserved QTL genotypes (Q_i) for each individual are computed given genotypes at the pair of closest flaking markers. Let $P_{ik} = P(Q_i = k|M_{ij}, M_{ij+1}, \lambda)$ where $k = QQ$ or Qq for BC and $k = QQ$, Qq or qq for F2; λ is the testing position of putative QTL. P_{ik} can be calculated using recombination frequencies between two

markers or between a marker and a putative QTL. The distribution of the quantitative trait given the flanking maker genotypes follows a finite mixture model and the likelihood functions for BC and F2 are given by

$$L_{BC}(\mu,\beta,\sigma^2,\lambda)=\prod_{i=1}^{n}[P_{iQQ}\phi((y_i-\mu+\frac{1}{2}\beta)/\sigma)+P_{iQq}\phi((y_i-\mu-\frac{1}{2}\beta)/\sigma)], \tag{1.3}$$

$$L_{F2}(\mu,\beta,\gamma,\sigma^2,\lambda)=\prod_{i=1}^{n}[P_{iQQ}\phi((y_i-\mu+\beta+\frac{1}{2}\gamma)/\sigma)+P_{iQq}\phi((y_i-\mu-\frac{1}{2}\gamma)/\sigma)+P_{iqq}\phi((y_i-\mu-\beta+\frac{1}{2}\gamma)/\sigma)],$$

respectively, where $\phi(z)$ is the standard normal density function. The above likelihood functions can be maximized using EM algorithm to obtain MLE estimates $(\hat{\mu},\hat{\beta},\hat{\gamma},\hat{\sigma^2})$. Since the genotypes of QTL are treated as the missings, the observed data include only phenotypes and maker genotypes while the full data include phenotypes, marker genotypes and QTL genotypes. Test statistics are constructed using the LOD scores

$$LOD_{BC}(\lambda)=log_{10}\frac{L_{BC}(\hat{\mu},\hat{\beta},\hat{\sigma^2})}{L_{BC}(\tilde{\mu},0,\tilde{\sigma^2})},\ LOD_{F2}(\lambda)=log_{10}\frac{L_{F2}(\hat{\mu},\hat{\beta},\hat{\gamma},\hat{\sigma^2})}{L_{F2}(\tilde{\mu},0,0,\tilde{\sigma^2})}, \tag{1.4}$$

where $\tilde{\mu}$ and $\tilde{\sigma^2}$ are MLE estimates under the null hypothesis $H_0:\beta=0$ for BC, $H_0:\beta=\gamma=0$ for F2. The location with the maximum $LOD_{BC}(\lambda)$ or $LOD_{F2}(\lambda)$ is the estimate of the QTL position. Determining the threshold of test statistic (maximum $LOD_{BC}(\lambda)$ or $LOD_{F2}(\lambda)$) is quite complicated because many factors, such as the genome size, genetic map density and the proportion of missing data, could affect the distribution of the test statistic under the null hypothesis. The usual pointwise significance level on the basis of the chi-square approximation is inadequate since the entire genome is tested for the existence of QTL. With an infinitely dense-map and large samples, the LOD score can be approximated by Ornstein-Uhlenbeck diffusion processes for BC [Lander and Botstein, 1989] and F2 [Dupuis and Siegmund, 1999]. To obtain an empirical threshold, Churchill and Doerge [1994] proposed a permutation

procedure.

Multiple Interval Mapping

If there exist more than one QTL affecting the trait on the chromosome, single QTL method may fail to discover true QTL and instead identify ghost (false) QTL. Moreover, single QTL model may fail to detect QTL with high epistatic effect but low marginal effect. To solve this problem, multiple interval mapping [Kao and Zeng, 1997; Kao et al., 1999] was proposed. This method combines QTL mapping with the analysis of genetic architecture of quantitative traits through a search algorithm to search for the number and position of QTL and their genetic effects. Suppose there are p putative QTL and t significant pairwise epistatic effects. Note that the model only contains a subset (t pairs) of QTL pairs that each shows a significant epistatic effect since if all pairs of p QTL are fitted in the model, it can be overparameterized. Cockerham's genetic model is used to define the genetic parameters. The main advantage of Cockerham's model is that it has an orthogonal property in modeling genetic effects. For BC population, the model of multiple interval mapping can be expressed as

$$y_i = \mu + \sum_{r=1}^{p} \beta_r x_{ir} + \sum_{r \neq s}^{t} \beta_{rs} x_{ir} x_{is} + e_i = \mu + \boldsymbol{x}_i \boldsymbol{\beta} + e_i, \tag{1.5}$$

where β_r, β_{rs} are the marginal and epistatic effect; x_{ir} is 1/2 for Qq and -1/2 for QQ; $\boldsymbol{\beta}$ is the $(p+t) \times 1$ vector of marginal and epistatic effects; $\boldsymbol{x}_i$ is the $1 \times (p+t)$ vector of indicator variables. For F2 population, the model is given by

$$y_i = \mu + \sum_{r=1}^{p} (\beta_r x_{ir} + \gamma_r z_{ir}) + \sum_{r \neq s}^{t} (\beta_{rs} x_{ir} x_{is} + \delta_{rs} x_{ir} z_{is} + \xi_{rs} z_{ir} x_{is} + \gamma_{rs} z_{ir} z_{is}) + e_i = \mu + \boldsymbol{z}_i \boldsymbol{\gamma} + e_i, \tag{1.6}$$

where β_r, γ_r are the additive, dominance effect and β_{rs}, δ_{rs}, ξ_{rs}, γ_{rs} are additive × additive, additive × dominance, dominance × additive and dominance × dominance effects;

x_{ir} is 1 for qq, 0 for Qq, -1 for QQ and z_{ir} is 1/2 for Qq, -1/2 for QQ and qq; $\boldsymbol{\gamma}$ is the $(2p+4t)\times 1$ vector of all effects; $\boldsymbol{z}_i$ is the $1\times(2p+4t)$ vector of indicator variables. Even though the genotype of each putative QTL (Q_{ij}) in interval I_{ij} is unobserved, the probabilities of Q_{ij} can be inferred from the flanking markers of I_{ij} based on the recombination frequency between them. We have $P(Q_{i1},...,Q_{ip}|I_{i1},..,I_{ip}) = \prod_{j=1}^{p} P(Q_{ij}|I_{ij})$. We refer P_{ij} ($j = 1,...,2^p$ for BC, $j = 1,...,3^p$ for F2) as the conditional probabilities of all possible QTL genotypes of individual i. The likelihood functions of the multiple interval mapping for BC and F2 are the following mixture of normal distributions:

$$L_{BC}(\boldsymbol{\beta},\mu,\sigma^2) = \prod_{i=1}^{n}[\sum_{j=1}^{2^p} P_{ij}\phi(\frac{y_i-\mu-\boldsymbol{x}_{ij}\boldsymbol{\beta}}{\sigma})],\ L_{F2}(\boldsymbol{\gamma},\mu,\sigma^2) = \prod_{i=1}^{n}[\sum_{j=1}^{3^p} P_{ij}\phi(\frac{y_i-\mu-\boldsymbol{z}_{ij}\boldsymbol{\gamma}}{\sigma})], \tag{1.7}$$

where $\phi(z)$ is the standard normal density function. Again, for the MLE estimates of $(\boldsymbol{\beta},\boldsymbol{\gamma},\mu,\sigma^2)$, EM algorithm can be employed. The test for marginal effect is performed by LOD score for $H_0 : \beta_r = 0$ or $H_0 : \gamma_r = 0$. For testing epistatic effect, we use LOD score for $\beta_{rs} = 0$, $\delta_{rs} = 0$, $\xi_{rs} = 0$ and $\gamma_{rs} = 0$. In theory, multiple interval mapping can be applied to more than two QTL straightforwardly. However, since the search becomes multidimensional, there are some difficulties in parameter estimation and model identifiability to map more than two QTL simultaneously in practice.

Bayesian Interval Mapping

Several Bayesian methods for QTL mapping has been proposed. Satagopan et al. [1996] proposed a Bayesian methods to detect multiple QTL simultaneously using Markov chain Monte Carlo (MCMC) method. When the quantitative trait is explained by multiple genes (p genes) acting independently and their interactions, we have

$$y_i = \mu + \sum_{j=1}^{p}\beta_j x_{ij} + \sum_{j\neq k}^{t}\beta_{jk}x_{ij}x_{ik} + e_i = \mu + \boldsymbol{x}_i\boldsymbol{\beta} + e_i\ (i = 1,...,n), \tag{1.8}$$

where $\boldsymbol{x}_i$ is the p QTL genotypes and their interactions for the ith individual and $\boldsymbol{\beta}$ is the marginal and epistatic effects of the p loci. The genetic parameters are model unknowns ($\boldsymbol{\beta}$,μ,σ^2) and the QTL loci $\boldsymbol{\lambda} = \{\lambda_j\}_{j=1}^{p}$. For individual i, marker genotype $\boldsymbol{M}_i = \{M_{ik}\}_{k=1}^{m}$ and phenotypic trait $\boldsymbol{y} = (y_1, ..., y_n)^T$ are observed, but the genotypes of the putative QTL, $\boldsymbol{x}_i$ are not observed. However, the conditional distribution $P(\boldsymbol{x}_i|\boldsymbol{\lambda}, \boldsymbol{M}_i)$ can be obtained using recombination frequency between the putative QTL and the markers.

To implement Bayesian analysis, the prior distribution is required over the parameter space ($\boldsymbol{\lambda}$,$\boldsymbol{\beta}$,μ,σ^2). We assume prior independence of the parameters. That is, $P(\boldsymbol{\lambda}, \boldsymbol{\beta}, \mu, \sigma^2) = P(\boldsymbol{\lambda})P(\mu)P(\sigma^2)\prod_{j=1}^{p}\beta_j\prod_{j\neq k}^{t}\beta_{jk}$. A natural choice of prior for $\boldsymbol{\lambda}$ when there is no available information regarding the location would be a uniform distribution for p ordered variables on [0,D_m] where D_m is the length of the linkage group ($0 < \lambda_1 < ... < \lambda_p < D_m$). The prior for overall mean μ is a normal distribution centered at 0 with variance τ_μ^2 ($\mu \sim N(0, \tau_\mu^2)$). The phenotypic variance σ^2 is assumed to have an inverse gamma prior ($\sigma^2 \sim IG(u, v)$). The priors of QTL effect β_j, β_{jk} ($j, k = 1, ..., p$) are independent normal distributions with mean 0 and variance τ_β^2 ($\beta_j, \beta_{jk} \sim N(0, \tau_\beta^2)$).

The posterior distribution over all the unknown parameters ($\boldsymbol{\lambda}$,$\boldsymbol{\beta}$,μ,σ^2) is given by $P(\boldsymbol{\lambda}, \boldsymbol{\beta}, \mu, \sigma^2|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\beta}, \mu, \sigma^2)\ P(\boldsymbol{\lambda}, \boldsymbol{\beta}, \mu, \sigma^2)$. MCMC methods are used to obtain the posterior distribution. Specifically, Metropolis-Hastings algorithm is used to update QTL positions ($\boldsymbol{\lambda}$) and Gibbs sampler is used to sample model unknowns ($\boldsymbol{\beta}, \mu, \sigma^2$) as follows. For jth locus, a proposal position λ_j^* is drawn from a uniform distribution on the interval (max($\lambda_{j-1}, \lambda_j - d$),min($\lambda_{j+1}, \lambda_j + d$)) where d is the tuning parameter and λ_0=0 and $\lambda_{p+1} = D_m$. The proposal position is accepted with probability min(α,1) where $\alpha(\lambda_j, \lambda_j^*) = P(\lambda_j^*|\boldsymbol{\lambda}_{-j}, \boldsymbol{y})P(\lambda_j^*, \lambda_j)/P(\lambda_j|\boldsymbol{\lambda}_{-j}, \boldsymbol{y})\ P(\lambda_j, \lambda_j^*)$

where $\boldsymbol{\lambda}_{-j}$ represents all the elements of $\boldsymbol{\lambda}$ except λ_j. For model unknown parameters, we can directly sample $\boldsymbol{\beta}$, μ and σ^2 from their full conditionals. The full conditionals are given by $\beta_j|\boldsymbol{\lambda}, \boldsymbol{\beta}_{-j}, \mu, \sigma^2, \boldsymbol{y} \sim N(\eta_{\beta_j}, \tau^2_{\beta_j})$, $\beta_{jk}|\boldsymbol{\lambda}, \boldsymbol{\beta}_{-jk}, \mu, \sigma^2, \boldsymbol{y} \sim N(\eta_{\beta_{jk}}, \tau^2_{\beta_{jk}})$, $\mu|\boldsymbol{\lambda}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{y} \sim N(\eta_{\mu^*}, \tau^2_{\mu^*})$ and $\sigma^2|\boldsymbol{\lambda}, \boldsymbol{\beta}, \mu, \boldsymbol{y} \sim IG(u_{\sigma^2}, v_{\sigma^2})$ (see Satagopan et al. [1996] for more details). The marginal posterior density of $\boldsymbol{\beta}$, μ and σ^2 can be estimated since their full conditional densities are known completely. However, since the full conditional density of $\boldsymbol{\lambda}$ is not known, the density estimates of $\boldsymbol{\lambda}$ should be obtained from the MCMC samples by different kernel estimate methods (for example, the histogram estimator). Confidence intervals can be obtained as high posterior density (HPD) regions.

1.2 Gaussian Process Models

Bayesian Neural Networks

Over the last two decades, there has been much activity concerning the application of Gaussian process models to machine learning tasks [Rasmussen and Williams, 2006]. Gaussian process models in machine learning were aroused in the early 1990s. Many researchers realized that neural networks were not so easy to apply in practice. MacKay [1992] and Neal [1996] pursued the probabilistic framework using approximation methods and Markov chain Monte Carlo (MCMC) methods, respectively. Neal [1996] focused on sophisticated Markov chain methods for inference in large finite neural networks and he demonstrated that a large class of Bayesian models, based on neural networks, converged to a Gaussian process in the limit of an infinite network [Neal, 1996]. Gaussian processes can be also derived from the viewpoint of nonparametric Bayesian regression, by directly imposing Gaussian process prior on the regression function [MacKay, 1998]. Comparative studies have confirmed the better performance of Gaussian process

regression than other nonlinear models [Rasmussen, 1996].

The idea of using Gaussian process directly came from investigations by Neal [1996] into prior over weights for neural networks [Rasmussen, 1996]. We consider a neural network with J inputs, one layer of K *tanh* hidden units and one output unit. Both hidden and output units have weights and biases and the network is fully connected between consecutive layers:

$$\psi_k(\boldsymbol{x}) = tanh(\sum_{j=1}^{J} u_{jk}x_j + u_0), \qquad \eta(\boldsymbol{x}) = \sum_{k=1}^{K} v_k\psi_k(\boldsymbol{x}) + v_0. \tag{1.9}$$

The zero mean Gaussian priors are imposed on all weights and biases. That is, $u_{jk} \sim N(0, \sigma_u^2)$, $u_0 \sim N(0, \sigma_{u0}^2)$, $v_k \sim N(0, \sigma_v^2)$ and $v_0 \sim N(0, \sigma_{v0}^2)$. Given a specific input vector $\boldsymbol{x}_i$, we can derive the distribution of a output based on the priors on weights and biases. We have $E[v_k\psi_k(\boldsymbol{x}_i)] = E[v_k]E[\psi_k(\boldsymbol{x}_i)] = 0$ ($\because v_k \perp \psi_k(\boldsymbol{x}_i)$) and $E[(v_k\psi_k(\boldsymbol{x}_i))^2] = \sigma_v^2 E[(\psi_k(\boldsymbol{x}_i))^2]$ ($\because \psi_k(\boldsymbol{x}_i)$ is bounded). By Central Limit Theorem (CLT), as the number of hidden units K goes to infinity, the prior distribution of $\eta(\boldsymbol{x}_i)$ converges to a Gaussian distribution with mean 0, variance $c(\boldsymbol{x}_i) = \sigma_{v0}^2 + K\sigma_v^2 E[(\psi_k(\boldsymbol{x}_i))^2]$. If we choose σ_v^2 which scales inversely with K, a well defined prior can be obtained in the limit of infinite number of hidden units. Using the similar argument, the joint distribution for multiple inputs converges in the limit of K to a multivariate Gaussian with mean 0 and covariance $c(\boldsymbol{x}_i, \boldsymbol{x}_{i'}) = E[\eta(\boldsymbol{x}_i)\eta(\boldsymbol{x}_{i'})] = \sigma_{v0}^2 + K\sigma_v^2 E[\psi_k(\boldsymbol{x}_i)\psi_k(\boldsymbol{x}_{i'})]$.

Gaussian Process Models

A Gaussian process is a generalization of the Gaussian probability distribution. While a probability distribution describes random variables which are scalars or vectors, a stochastic process governs functions of input values. Formally, a Gaussian process is

a collection of random variables $\{\eta_x\}$ indexed by a set $x \in \boldsymbol{X}$, where any finite subset of η_x's has a joint multivariate Gaussian distribution. Often, Gaussian processes are defined over time, where the index set is time, but in our case, we index the random variables $\boldsymbol{\eta} = \{\eta_x\}$ by the input space $\boldsymbol{X}$.

Consider the case where a training set of n observations is available and thus data set $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)|i = 1, ..., n\} = \{\boldsymbol{x}, \boldsymbol{y}\}$, where $\boldsymbol{x}_i$ is a vector of p covariates and y_i is the scalar response. We will consider the following Bayesian regression model with Gaussian noise

$$y_i = \eta(\boldsymbol{x}_i) + e_i \ (i = 1, ..., n), \tag{1.10}$$

where η is an unknown function of p covariates that we modeled via a Gaussian process prior and e_i is a noise with distribution $N(0, \sigma_e^2)$.

The Gaussian process is completely specified by its mean function $m(\boldsymbol{x}_i)$ and covariance function $c(\boldsymbol{x}_i, \boldsymbol{x}_{i'})$. The Gaussian process can be expressed as $\eta \sim \mathcal{GP}(m(\boldsymbol{x}_i), c(\boldsymbol{x}_i, \boldsymbol{x}_{i'}))$. From here on, we will consider only Gaussian process model with a mean of zero, such that $\eta \sim \mathcal{GP}(0, c(\boldsymbol{x}_i, \boldsymbol{x}_{i'}))$. One of examples of a covariance function is

$$c(\boldsymbol{x}_i, \boldsymbol{x}_{i'}) = a_0 + a_1 \sum_{k=1}^{p} x_{ik} x_{i'k} + v_0 exp(-\frac{1}{2} \sum_{k=1}^{p} w_k (x_{ik} - x_{i'k})^2), \tag{1.11}$$

where $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})$ and $a_0, a_1, v_0, w_1, ..., w_p$ are hyperparameters. In this function, first two terms involving a_0 and a_1 control the scale of the bias and linear contribution to the covariance. The contribution from the linear terms in the covariance function may become large for inputs which are quite distant from the bulk of the output values. The exponential part defines the correlation between outputs and nearby inputs. The parameter w_k is multiplied by the coordinate-wise distance in input space and thus allows for different distance measures for each input dimension. For irrelevant inputs, the corresponding w_k should be small in order for the model to ignore these inputs while

parameter w_k get large for relevant inputs. The parameter v_0 defines the overall scale of the local correlations. The functions of Gaussian process are smooth and stationary. These are properties which are induced by the covariance function. In the Gaussian process models, the role of the kernel function and local model are both integrated in the covariance function. Like the kernel function, the covariance function is a function of the model inputs, it returns the covariance between the output corresponding to two inputs. The problem of learning in Gaussian processes is exactly the problem of finding suitable properties of the covariance function.

Marginal Likelihood

Let $\boldsymbol{\theta} = (a_0, a_1, v_0, w_1, ..., w_p, \sigma_e^2)$, $\eta_i = \eta(x_i)$ and $\boldsymbol{\eta} = (\eta_1, ..., \eta_n)$ such that $\boldsymbol{\eta} \sim \mathcal{GP}(0, \boldsymbol{\Sigma}_{\boldsymbol{\eta}})$. By the independence assumption, we have the likelihood as $P(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta}) = \prod_{i=1}^{n} P(y_i|\eta_i, \boldsymbol{\theta}) \stackrel{d}{=} N(\boldsymbol{\eta}, \sigma_e^2 I_n)$. The marginal likelihood of $\boldsymbol{y}$, $P(\boldsymbol{y}|\boldsymbol{\theta})$ is given by

$$P(\boldsymbol{y}|\boldsymbol{\theta}) = \int P(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta}) P(\boldsymbol{\eta}) d\boldsymbol{\eta} \stackrel{d}{=} N(0, \boldsymbol{\Sigma}) \text{ where } \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{\eta}} + \sigma_e^2 I_n. \tag{1.12}$$

The term marginal likelihood refers to the marginalization over the function value $\boldsymbol{\eta}$. The log likelihood of the hyperparameters and its partial derivatives are given by

$$\begin{aligned} log P(\boldsymbol{y}|\boldsymbol{\theta}) &= -\frac{1}{2} log|\boldsymbol{\Sigma}| - \frac{1}{2} \boldsymbol{y}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{y} - \frac{n}{2} log 2\pi, \\ \frac{\partial}{\partial \theta_i} log P(\boldsymbol{y}|\boldsymbol{\theta}) &= -\frac{1}{2} tr(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}) + \frac{1}{2} \boldsymbol{y}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}. \end{aligned} \tag{1.13}$$

To calculate the partial derivatives of the likelihood, it is necessary to invert the matrix $\boldsymbol{\Sigma}$, using, for example, Cholesky decomposition. Maximum likelihood can be implemented by several learning schemes such as a Monte Carlo method for integration over hyperparameters and maximum a posteriori (MAP) method.

Prior Specification

Let $\tau_{a_0} = 1/a_0$, $\tau_{a_1} = 1/a_1$, $\tau_{v_0} = 1/v_0$, $\tau_e = 1/\sigma_e^2$. We impose the Gamma priors on these four parameters: $\tau_{a_0} \sim Ga(\frac{\alpha_{a_0}}{2}, \frac{\alpha_{a_0}}{2\mu_{a_0}})$, $\tau_{a_1} \sim Ga(\frac{\alpha_{a_1}}{2}, \frac{\alpha_{a_1}}{2\mu_{a_0}})$, $\tau_{v_0} \sim Ga(\frac{\alpha_{v_0}}{2}, \frac{\alpha_{v_0}}{2\mu_{v_0}})$ and $\tau_e \sim Ga(\frac{\alpha_e}{2}, \frac{\alpha_e}{2\mu_e})$ where $\alpha_{a_0}, \alpha_{a_1}, \alpha_{v_0}, \alpha_e$ are positive shape parameters and $\mu_{a_0}, \mu_{a_1}, \mu_{v_0}, \mu_e$ are the means of $\tau_{a_0}, \tau_{a_1}, \tau_{v_0}, \tau_e$. The large values of α's produce priors concentrated near μ_α's. The priors for hyperparameter w_i are more complicated. As we expect the prior on the importance of hyperparameter w_i to be lower with increasing numbers of input (i.e. large p), we let $\tau_{w_i} = 1/w_i$ and put a gamma prior whose mean scales with the number of inputs p on τ_{w_i} as $\tau_{w_i} \sim Ga(\frac{\alpha_{w_i}}{2}, \frac{\alpha_{w_i}}{2\mu_{w_i}})$ where $\mu_{w_i} = \mu_0 p^{2/\alpha_{w_i}}$. Large p makes μ_{w_i} large and the mean of w_i small.

Hybrid Monte Carlo Method

The joint posterior distribution marginalized with respect to $\boldsymbol{\eta}$ is computed from the marginal likelihood multiplied by the prior: $P(\boldsymbol{\theta}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta})$. To obtain the posterior distribution, we need to integrate over the resulting posterior, but analytic integration is infeasible due to the complex form of the likelihood. The Hybrid Monte Carlo method [Duane et al., 1987] is appropriate for this case. When sampling from complicated multidimensional distributions, it is often advantageous to use gradient information to find regions of high probability when gradients can be obtained. The Hybrid Monte Carlo method avoids the random walk behavior by creating a virtual dynamic system where hyperparameter $\boldsymbol{\theta}$ plays the role of position variables, which are augmented by a set of momentum variables $\boldsymbol{\phi}$. The kinetic energy is a function of momentum variables: $\mathcal{K}(\boldsymbol{\phi}) = \frac{1}{2}\sum_{i=1}^{p+4}\phi_i^2$ where $\boldsymbol{\phi} = (\phi_1, ..., \phi_{p+4})$ is in one-to-one correspondence with the component of $\boldsymbol{\theta}$. The potential energy is defined as $\mathcal{E}(\boldsymbol{\theta}) = -logP(\boldsymbol{\theta}|\boldsymbol{y})$. The total energy $\mathcal{H}$ of the system which is called "Hamiltonian" function is the sum of the kinetic energy $\mathcal{K}$ and the potential energy $\mathcal{E}$, such that $\mathcal{H}(\boldsymbol{\phi}, \boldsymbol{\theta}) =$

$\mathcal{K}(\boldsymbol{\phi}) + \mathcal{E}(\boldsymbol{\theta})$. The dynamical system evolved through virtual time t is governed by the following Hamilton's differential equations:

$$\frac{d\theta_i}{dt} = \frac{\partial \mathcal{H}}{\partial \phi_i} = \phi_i, \qquad \frac{d\phi_i}{dt} = -\frac{\partial \mathcal{H}}{\partial \theta_i} = -\frac{\partial \mathcal{E}}{\partial \theta_i}. \tag{1.14}$$

Since the partial derivative of $\mathcal{E}$ with respect to θ is a complicated function, the above equation cannot be simulated exactly. We use the following leapfrog steps to approximate the dynamic system:

$$\begin{aligned} \phi_i(t + \frac{\epsilon}{2}) &= \phi_i(t) - \frac{\epsilon}{2}\frac{\partial \mathcal{E}}{\partial \theta_i}(\boldsymbol{\theta}(t)), \\ \theta_i(t + \epsilon) &= \theta_i(t) + \epsilon \phi_i(t + \frac{\epsilon}{2}), \\ \phi_i(t + \epsilon) &= \phi_i(t + \frac{\epsilon}{2}) - \frac{\epsilon}{2}\frac{\partial \mathcal{E}}{\partial \theta_i}(\boldsymbol{\theta}(t + \epsilon)), \end{aligned} \tag{1.15}$$

where ϵ is the step size for discretizing the dynamic system. The step size ϵ is set to the same value for all hyperparameters and is chosen to $\epsilon \propto n^{-1/2}$ since the magnitude of the gradients under the posterior are expected to be scale roughly as $n^{1/2}$ when the prior is vague. Rasmussen [1996] found that $\epsilon = 0.5n^{-1/2}$ performs reasonably well, in practice.

Maximum a Posteriori (MAP) Estimates

When there is a large number of observations, integration via Monte Carlo method is computationally infeasible and the maximum a posteriori (MAP) approach would be preferred. When the posterior is fairly narrow, the prediction for a MAP method may not differ much from the results of integrating over hyperparameters. To find the MAP estimates, conjugate gradient optimization technique [Rasmussen and Williams, 2006], simplex search method [Lagarias et al., 1998], finite element approach [Roberts

et al., 2003] and sparse grid approximation [Bungartz and Griebel, 2004] can be used [Hegland, 2007].

1.3 Bayesian Model Selection Methods

Bayesian Model Selection

A Bayesian approach to model selection is concerned with the following situation [Han and Carlin, 2001]. Suppose the observed data $\boldsymbol{y}$ are considered to have been generated by a model $m \in \mathcal{M}$, where $\mathcal{M}$ is the finite set of competing models. Corresponding to model m, there is a distinct unknown parameter vector $\boldsymbol{\theta}_m \in \Theta_m$ of dimension p where Θ_m is the set of all possible values for $\boldsymbol{\theta}_m$. Each model specifies the distribution of $\boldsymbol{y}$, $P(\boldsymbol{y}|m, \boldsymbol{\theta}_m)$. If $P(m)$ is the prior probability of model m, where $\sum_{m \in \mathcal{M}} P(m) = 1$, the posterior probability is given by $P(m|\boldsymbol{y}) = \frac{P(m)P(\boldsymbol{y}|m)}{\sum_{m \in \mathcal{M}} P(m)P(\boldsymbol{y}|m)}, m \in \mathcal{M}$ where $P(\boldsymbol{y}|m)$ is the marginal likelihood computed from $P(\boldsymbol{y}|m) = \int P(\boldsymbol{y}|m, \boldsymbol{\theta}_m)P(\boldsymbol{\theta}_m|m)d\boldsymbol{\theta}_m$ and $P(\boldsymbol{\theta}_m|m)$ is the model-specific conditional prior of $\boldsymbol{\theta}_m$. To compare two models, m and m', we often use the Bayes factor for model m over m':

$$BF_{mm'} = \frac{P(m|\boldsymbol{y})/P(m)}{P(m'|\boldsymbol{y})/P(m')} = \frac{P(\boldsymbol{y}|m)}{P(\boldsymbol{y}|m')}. \tag{1.16}$$

The Bayes factor $BF_{mm'}$ captures the change in the odds in favor of model m as we move from prior to posterior. Equation (1.16) shows that the Bayes factor for the comparison of two models can be obtained using the marginal likelihoods of two models. If the model-specific prior $P(\boldsymbol{\theta}_m|m)$ is improper, the marginal likelihood is necessarily improper as well and thus the Bayes factor (1.16) is not well defined. Various solutions have been proposed to this problem, including pseudo Bayes factor approaches [Berger and Pericchi, 1996]. However, from here on, we will only consider proper Bayes factors.

Several methods seek to estimate the marginal likelihood $\hat{P}(\boldsymbol{y}|m)$ directly for each

model and subsequently calculate the Bayes factor using equation (1.16). Chen and Shao [1998] developed an importance sampling approach to estimate the marginal likelihood using a technique of Chen and Shao [1997]. Newton and Raftery [1994] proposed the estimator for marginal likelihood, which is the harmonic mean of the likelihood values sampled from the stationary phase of the MCMC run. Chib [1995] and Chib and Jeliazkov [2001] provides a indirect method to estimate marginal likelihood in the context of Gibbs sampling and Metropolis-Hastings algorithm. All methods operate on a posterior sample that has already been produced by some noniterative or MCMC method, although the methods of Chib [1995] and Chib and Jeliazkov [2001] will often require multiple runs of slightly different version of the MCMC algorithm to produce the necessary output. However, for some complicated or high-dimensional models, these approaches are difficult to implement.

A slight more direct and more common approach to estimating posterior model probabilities using MCMC is to include the model indicator $\boldsymbol{\gamma}$ as a parameter in the sampling algorithm. This may complicate the initial sampling process, but has the clear benefit of producing a stream of samples $\{\gamma_i\}_{i=1}^{H}$ from the marginal posterior distribution of the model indicator, $P(\gamma|\boldsymbol{y})$. Once the sampler converges, the proportion of times the sampler visits model m is a simple estimate of each posterior model probability:

$$P(\gamma = m|\boldsymbol{y}) = \frac{\text{number of } \gamma_i = m}{\sum_{i=1}^{H} \text{number of } \gamma_i},\ m = 1, ..., K. \tag{1.17}$$

This estimate can be used to compute the Bayes factor between any two of the models, say m and m':

$$BF_{mm'} = \frac{P(\gamma = m|\boldsymbol{y})/P(\gamma = m)}{P(\gamma = m'|\boldsymbol{y})/P(\gamma = m')}. \tag{1.18}$$

To avoid the risk of increased correlations and slower convergence, it is sometimes possible to integrate the parameters $\boldsymbol{\theta}$ out of the model before sampling begins, yielding

a sampler that operates over the model space alone. Unfortunately, most model settings are too complicated to allow the entire parameter $\boldsymbol{\theta}$ to be integrated out of the model in a closed form, and thus require that the MCMC search be over the model and parameter space jointly. This joint search approach also permits posterior estimate of the parameters under each model $P(\boldsymbol{\theta}|\gamma = m, \boldsymbol{y})$, simply by conditioning on the samples produced when the chain is currently in state $\gamma = m$.

Stochastic Search Variable Selection

George and McCulloch [1993] proposed a model selection procedure which is called stochastic search variable selection (SSVS). This method introduces latent variables to determine whether particular regression coefficients may safely be estimated by 0 or not. We consider the following simple linear model

$$\boldsymbol{y} = \boldsymbol{x}\boldsymbol{\beta} + e,\ e \sim N(0, \sigma^2 I), \tag{1.19}$$

where $\boldsymbol{y} = (y_1, ..., y_n)^T$ is $n \times 1$, $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_p)$ is $n \times p$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ is $p \times 1$ and e is $n \times 1$. We assign prior for each β_i to be a mixture of two normal densities,

$$\beta_i|\gamma_i \sim (1 - \gamma_i)N(0, \tau_i^2) + \gamma_i N(0, c_i^2\tau_i^2), \tag{1.20}$$

where γ_i $(i = 1, ..., p)$ is a binary variable with $P(\gamma_i = 1) = 1 - P(\gamma_i = 0) = p_i$. When $\gamma_i = 0$, $\beta_i \sim N(0, \tau_i^2)$ and when $\gamma_i = 1$, $\beta_i \sim N(0, c_i^2\tau_i^2)$. When τ_i $(\tau_i > 0)$ is small and $\gamma_i = 0$, then β_i would probably be small that it could safely be estimated by 0. When c_i $(c_i > 1)$ is large and $\gamma_i = 1$, then a non-zero estimate of β_i is probably included in the final model. Based on this interpretation, p_i can be viewed as a prior probability that β_i is non-zero. The mixture prior for $\beta_i|\gamma_i$ can be written in vector form as $\boldsymbol{\beta}|\boldsymbol{\gamma} \sim N(0, \boldsymbol{D}_{\boldsymbol{\gamma}}\boldsymbol{R}\boldsymbol{D}_{\boldsymbol{\gamma}})$ where $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_p)$; $\boldsymbol{R}$ is the prior correlation matrix and $\boldsymbol{D}_{\boldsymbol{\gamma}} = diag(a_1\tau_i, ..., a_p\tau_p)$, where

$a_i = 1$ if $\gamma_i = 0$ and $a_i = c_i$ if $\gamma_i = 1$. We use the inverse gamma conjugate prior as the prior for σ^2: $\sigma^2|\boldsymbol{\gamma} \sim IG(\frac{\nu_\gamma}{2}, \frac{\nu_\gamma \lambda_\gamma}{2})$ where ν_γ and λ_γ are hyperparameters to be specified.

Product Space Search

Carlin and Chib [1995] proposed a Gibbs sampling method that avoids convergence difficulties and accommodates fairly general model settings. Suppose there are K candidate models, a distinct parameter vector $\boldsymbol{\theta}_m$ $(m = 1, ..., K)$ for each model, whose prior is assumed to be independent each other given the model indicator γ. Corresponding to model m, the likelihood is $P(\boldsymbol{y}|\boldsymbol{\theta}_m, \gamma = m)$ and the prior is $P(\boldsymbol{\theta}_m|\gamma = m)$. Since γ is an indicator of which $\boldsymbol{\theta}_m$ is relevant to $\boldsymbol{y}$, $\boldsymbol{y}$ is independent of $\boldsymbol{\theta}_{-m}$ given the model indicator γ where $\boldsymbol{\theta}_{-m}$ represents all elements of $\boldsymbol{\theta}$ except $\boldsymbol{\theta}_m$. Under the conditional independence assumption, we have

$$\begin{aligned} P(\boldsymbol{y}|\gamma = m) &= \int P(\boldsymbol{y}|\boldsymbol{\theta}_m, \gamma = m) P(\boldsymbol{\theta}|\gamma = m) d\boldsymbol{\theta} \\ &= \int P(\boldsymbol{y}|\boldsymbol{\theta}_m, \gamma = m) P(\boldsymbol{\theta}_m|\gamma = m) d\boldsymbol{\theta}_m, \end{aligned} \tag{1.21}$$

which has nothing to do with the pseudoprior $P(\boldsymbol{\theta}_m|\gamma \neq m)$. Therefore, a pseudoprior is only conveniently chosen liking density which is used to completely define the joint model specification. In order to implement the Gibbs sampler, the full conditional distribution of $\boldsymbol{\theta}_m$ is given by

$$P(\boldsymbol{\theta}_m|\boldsymbol{\theta}_{-m}, \gamma, \boldsymbol{y}) \propto \begin{cases} P(\boldsymbol{y}|\boldsymbol{\theta}_m, \gamma = m) P(\boldsymbol{\theta}_m|\gamma = m) & \text{if } \gamma = m \\ P(\boldsymbol{\theta}_m|\gamma \neq m) & \text{if } \gamma \neq m \end{cases} \tag{1.22}$$

Based on the equation (1.22), when $\gamma = m$, the parameter $\boldsymbol{\theta}_m$ is generated from the usual full conditional of model m; when $\gamma \neq m$, the parameter $\boldsymbol{\theta}_m$ is generated from the

pseudoprior. The full conditional distribution of the model indicator γ is given by

$$P(\gamma = m|\boldsymbol{\theta}, \boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{\theta}_m, \gamma = m)[\prod_{m' \epsilon \mathcal{M}} P(\boldsymbol{\theta}_{m'}|\gamma = m)]P(\gamma = m)}{\sum_{k=1}^{K}\{P(\boldsymbol{y}|\boldsymbol{\theta}_k, \gamma = k)[\prod_{m' \epsilon \mathcal{M}} P(\boldsymbol{\theta}_{m'}|\gamma = k)]P(\gamma = k)\}}. \tag{1.23}$$

Under the usual regularity conditions [Smith and Roberts, 1993], this Gibbs sampler will produce posterior samples from all conditional posterior distributions. When the Gibbs sampler converges, the posterior probability of model m can be estimated by

$$P(\gamma = m|\boldsymbol{y}) = \frac{1}{H}\sum_{i=1}^{H} I(\gamma_i = m), \tag{1.24}$$

which can be used to estimate the Bayes factors in favor of model m as

$$BF_{mm'} = \frac{P(\gamma = m|\boldsymbol{y})/P(\gamma = m)}{P(\gamma = m'|\boldsymbol{y})/P(\gamma = m')}. \tag{1.25}$$

Dellaportas et al. [2002] proposed a hybrid Gibbs-Metropolis version of product space search method. In this method, the model selection step is based on a proposal for a move from model m to m' with acceptance rate $\alpha_{mm'}$. This method is called Metropolized product space search method which proceeds as follows [Han and Carlin, 2001]: (1) Let the current state be $(m, \boldsymbol{\theta}_m)$, where $\boldsymbol{\theta}_m$ is of dimension p_m. (2) Propose a new model m' with probability $h(m, m')$. (3) Generate $\boldsymbol{\theta}_{m'}$ from a pseudoprior $P(\boldsymbol{\theta}_{m'}|\gamma \neq m')$ as in product space search method. (4) Accept the proposed move from model m to model m' with acceptance rate

$$\alpha_{mm'} = min\{1, \frac{P(\boldsymbol{y}|\boldsymbol{\theta}_{m'}, \gamma = m')P(\boldsymbol{\theta}_{m'}|\gamma = m')P(\boldsymbol{\theta}_m|\gamma = m')P(\gamma = m')h(m', m)}{P(\boldsymbol{y}|\boldsymbol{\theta}_m, \gamma = m)P(\boldsymbol{\theta}_m|\gamma = m)P(\boldsymbol{\theta}_{m'}|\gamma = m)P(\gamma = m)h(m, m')}\}. \tag{1.26}$$

When $m' = m$, the move become a Gibbs step. Posterior model probabilities and Bayes factors can be computed as before.

Reversible Jump MCMC

As the product space search method, the reversible jump MCMC method originally proposed by Green [1995], samples over the model and parameter space jointly but it avoids the full product space search at the cost of a less straightforward algorithm operating on the union space, $\mathcal{M} \times \bigcup_{m \in \mathcal{M}} \Theta_m$. This method generates a Markov chain that can jump between models with different dimensional parameter spaces, while retaining the aperiodicity, irreducibility, and detailed balance conditions necessary for MCMC convergence. The reversible jump MCMC algorithm proceeds as follows [Han and Carlin, 2001]: (1) Let the current state be $(m, \boldsymbol{\theta}_m)$, where $\boldsymbol{\theta}_m$ is of dimension p_m. (2) Propose a new model m' with probability $h(m, m')$. (3) Generate $\boldsymbol{u}$ from a proposal density $q(\boldsymbol{u}|\boldsymbol{\theta}_m, m, m')$. (4) Set $(\boldsymbol{\theta}_{m'}, \boldsymbol{u}') = g_{m,m'}(\boldsymbol{\theta}_m, \boldsymbol{u})$, where $g_{m,m'}$ is a deterministic function that is one-to-one and onto. This is a dimension-matching function, specified so that $p_m + dim(\boldsymbol{u}) = p_{m'} + dim(\boldsymbol{u}')$. (5) Accept the proposed move from model m to model m' with the acceptance rate

$$\alpha_{mm'} = min\{1, \frac{P(\boldsymbol{y}|\boldsymbol{\theta}_{m'}, \gamma = m')P(\boldsymbol{\theta}_{m'}|\gamma = m')P(\gamma = m')h(m', m)q(\boldsymbol{u}'|\boldsymbol{\theta}_{m'}, m', m)}{P(\boldsymbol{y}|\boldsymbol{\theta}_m, \gamma = m)P(\boldsymbol{\theta}_m|\gamma = m)P(\gamma = m)h(m, m')q(\boldsymbol{u}|\boldsymbol{\theta}_m, m, m')} |\frac{\partial g(\boldsymbol{\theta}_m, \boldsymbol{u})}{\partial(\boldsymbol{\theta}_m, \boldsymbol{u})}|\}. \tag{1.27}$$

When $m' = m$, the move can be either a standard Metropolis-Hastings or a Gibbs step. Posterior model probabilities and Bayes factors can be computed as described earlier. The dimension-matching aspect of this algorithm is a little obscure, so that further discussion is needed. Suppose we are comparing two models, for which $\theta_m \in \mathcal{R}^1$ and $\boldsymbol{\theta}_{m'} \in \mathcal{R}^2$ and θ_m is subvector of $\boldsymbol{\theta}_{m'}$. If we consider moving from model m to model m', we simply draw $u \sim q(u)$ and set $\boldsymbol{\theta}_{m'} = (\theta_m, u)$. In this case, the dimension-matching function g is identity function and $\boldsymbol{u}'$ should be ignored. We can set $h(m, m) = h(m, m') = h(m', m) = h(m', m') = \frac{1}{2}$ and the Jacobian of step 5 is equal to 1.

Composite Model Space Framework

Godsill [2001] proposed the composite model space framework, which essentially follows the setting of product space search method [Carlin and Chib, 1995] except parameters

are allowed to be shared between different models. If a standard Gibbs sampler is applied to this composite model space, it becomes the product space search method. However, a more sophisticated Metropolis-Hastings algorithm approach produces a version of the reversible jump algorithm that avoids the dimension matching step. The composite model space procedure is applicable when there exists a subvector $\boldsymbol{\beta}_{m'}$ of the parameter vector $\boldsymbol{\theta}_{m'}$ for the model m' such that $P(\boldsymbol{\theta}_{m'}|\boldsymbol{\theta}_{m'(-\boldsymbol{\beta}_{m'})}, \gamma = m', \boldsymbol{y})$ is available in closed form, and in the current model m, there exists an equivalent subvector $\boldsymbol{\theta}_{m(-\boldsymbol{\beta}_m)}$ of the same dimension as $\boldsymbol{\theta}_{m'(-\boldsymbol{\beta}_{m'})}$. The composite model space algorithm proceeds as follows [Han and Carlin, 2001]: (1) Let the current state be $(m, \boldsymbol{\theta}_m)$, where $\boldsymbol{\theta}_m$ is of dimension p_m. (2) Propose a new model m' with probability $h(m, m')$. (3) Set $\boldsymbol{\theta}_{m(-\boldsymbol{\beta}_m)} = \boldsymbol{\theta}_{m'(-\boldsymbol{\beta}_{m'})}$. (4) Accept the proposed move with the acceptance rate

$$\alpha_{mm'} = min\{1, \frac{P(m'|\boldsymbol{\theta}_{m'(-\boldsymbol{\beta}_{m'})}, \boldsymbol{y})h(m', m)}{P(m|\boldsymbol{\theta}_{m(-\boldsymbol{\beta}_m)}, \boldsymbol{y})h(m, m')}\}, \tag{1.28}$$

where $P(m|\boldsymbol{\theta}_{m(-\boldsymbol{\beta}_m)}, \boldsymbol{y}) = \int P(m, \boldsymbol{\beta}_m|\boldsymbol{\theta}_{m(-\boldsymbol{\beta}_m)}, \boldsymbol{y})d\boldsymbol{\beta}_m$. (5) If the model move is accepted, update the parameters of the new model $\boldsymbol{\beta}_{m'}$ and $\boldsymbol{\theta}_{m'(-\boldsymbol{\beta}_{m'})}$ using standard Gibbs or Metropolis-Hasting steps; otherwise, update the parameters of the old model $\boldsymbol{\beta}_m$ and $\boldsymbol{\theta}_{m(-\boldsymbol{\beta}_m)}$ using standard Gibbs or Metropolis-Hastings steps. Note that model move proposals from model m to model m always have acceptance rate 1, and thus when the current model is proposed, this algorithm could simplify to standard Gibbs or Metropolis-Hastings steps. Posterior model probabilities and Bayes factors can be computed as described earlier.

Variable selection is a special case of model selection. For variable selection, a natural parameterization for $\boldsymbol{\gamma}$ is as binary p-vector which is $\boldsymbol{\gamma} = \{\gamma_1, ..., \gamma_p\} \in \{0, 1\}^p$. That is, parameter $\boldsymbol{\gamma}$ is a vector of binary variables for indicating which covariates are included in $(\gamma_i = 1)$ or excluded from $(\gamma_i = 0)$ the model.

Deviance Information Criterion (DIC)

Suppose that $\boldsymbol{y} = (y_1, ..., y_n)^T$ is a vector of n observations generated from an unknown distribution $F(\boldsymbol{y})$ and a family of distributions with densities $\{P(\boldsymbol{y}|\boldsymbol{\theta})|\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \boldsymbol{R}^p\}$ is used to approximate the true distribution $F(\boldsymbol{y})$. The prior density and the posterior density are $P(\boldsymbol{\theta})$ and $P(\boldsymbol{\theta}|\boldsymbol{y})$, respectively. The predictive distribution for a future observation $\boldsymbol{z}$ is defined as $P(\boldsymbol{z}|\boldsymbol{y}) = \int P(\boldsymbol{z}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}$.

To measure the deviation of the predictive distribution $P(\boldsymbol{z}|\boldsymbol{y})$ from the true model $P(\boldsymbol{z})$, Spiegelhalter et al. [2002] considered the following posterior mean of the expected loglikelihood,

$$\xi = E_{\boldsymbol{z}}[E_{\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{z}|\boldsymbol{\theta})\}] = \int \{\int logP(\boldsymbol{z}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}\}dF(\boldsymbol{z}) \tag{1.29}$$

[Ando, 2007]. A natural estimator of ξ is as follows:

$$\hat{\xi} = \frac{1}{n}E_{\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\theta})\} = \frac{1}{n}\int logP(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}. \tag{1.30}$$

The estimator of ξ, $\hat{\xi}$, is generally positively biased since the same data $\boldsymbol{y}$ is used twice, one for constructing the posterior distribution $P(\boldsymbol{\theta}|\boldsymbol{y})$ and one for evaluating ξ. Note the expected bias of $\hat{\xi}$ is

$$b_{\boldsymbol{\theta}} = E_{\boldsymbol{y}}(\hat{\xi} - \xi) = \int (\hat{\xi} - \xi)dF(\boldsymbol{y}). \tag{1.31}$$

Let $\hat{b_{\boldsymbol{\theta}}}$ be an estimator of $b_{\boldsymbol{\theta}}$, then the bias-corrected estimator of ξ can be expressed as $\frac{1}{n}E_{\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\theta})\} - \hat{b_{\boldsymbol{\theta}}}$. Under this framework, Spiegelhalter et al. [2002] proposed the deviance information criterion (DIC),

$$DIC = -2E_{\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\theta})\} + P_D, \tag{1.32}$$

where P_D is an effective number of parameters, defined as $P_D = 2logP(\boldsymbol{y}|\bar{\boldsymbol{\theta}}) - 2E_{\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\theta})\}$ where $\bar{\boldsymbol{\theta}}$ is the posterior mean of $\boldsymbol{\theta}$.

Bayesian Predictive Information Criterion (BPIC)

Alternatively, Ando [2007] evaluated the asymptotic bias of $\hat{\xi}$ and proposed the Bayesian predictive information criterion:

$$BPIC = -2E_{\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\theta})\} + 2n\hat{b_{\boldsymbol{\theta}}}, \tag{1.33}$$

where $\hat{b_{\boldsymbol{\theta}}} = E[log\{P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta})\}] - log\{P(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_n)P(\hat{\boldsymbol{\theta}}_n)\} + tr\{J_n^{-1}(\hat{\boldsymbol{\theta}}_n)I_n(\hat{\boldsymbol{\theta}}_n)\} + p/2$, p is the dimension of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_n = argmax_{\boldsymbol{\theta}}P(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior mode and the matrices $I_n(\hat{\boldsymbol{\theta}}_n)$ and $J_n^{-1}(\hat{\boldsymbol{\theta}}_n)$ are given by $I_n(\hat{\boldsymbol{\theta}}_n) = \frac{1}{n}\sum_{i=1}^{n}\{\frac{\partial\xi_n(\boldsymbol{y}_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\xi_n(\boldsymbol{y}_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\}$, $J_n^{-1}(\hat{\boldsymbol{\theta}}_n) = -\frac{1}{n}\sum_{i=1}\{\frac{\partial^2\xi_n(\boldsymbol{y}_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\}$, respectively. Here, $\xi_n(\boldsymbol{y}_i,\boldsymbol{\theta}) = logP(\boldsymbol{y}_i|\boldsymbol{\theta}) + logP(\boldsymbol{\theta})/n$. Further, Ando [2011] proposed a simplified version of BPIC as

$$\text{Simplified } BPIC_1 = -2E_{\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\theta})\} + 2P_D, \tag{1.34}$$

where $P_D = n\hat{b}_{\boldsymbol{\theta}} = 2logP(\boldsymbol{y}|\bar{\boldsymbol{\theta}}_n) - 2E_{\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\theta})\}$. If we impose additional assumptions that (a) the prior is assumed to be dominated by the likelihood as n increases, that is, $logP(\boldsymbol{\theta}) = O(1)$, and (b) the specified models include the true model [Ando, 2011], then the estimated bias term $n\hat{b_{\boldsymbol{\theta}}}$ reduces to $n\hat{b_{\boldsymbol{\theta}}} \approx p$ and the BPIC can be further reduced to

$$\text{Simplified } BPIC_2 = -2E_{\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\theta})\} + 2p. \tag{1.35}$$

1.4 Bayesian Covariance Estimation

Bayesian Mixed Effects Models

The most common model for repeated measurements is the linear mixed effects model of Laird and Ware [1982]. For individual i with n_i repeated measurements, the mixed effects model is given by

$$\boldsymbol{y}_i = \boldsymbol{\mu}_i + \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \text{ i=1,...,n,} \tag{1.36}$$

where $\boldsymbol{y}_i$ is an $n_i \times 1$ response vector; $\boldsymbol{\mu}_i$ is the overall mean vector; $\boldsymbol{x}_i$ is an $n_i \times p$ matrix of fixed covariates; $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects; $\boldsymbol{z}_i$ is an $n_i \times k$ matrix of random covariates; $\boldsymbol{b}_i$ is a $k \times 1$ vector of random effects; $\boldsymbol{e}_i$ is an $n_i \times 1$ vector of random errors. It is standard to assume $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ are independent of each other and both are normally distributed with $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$ and $\boldsymbol{e}_i \sim N(0, \sigma^2 I_{n_i})$. Under this assumption, $\boldsymbol{y}_i|\boldsymbol{\mu}_i, \boldsymbol{\beta}, \boldsymbol{D}, \sigma^2 \sim N(\boldsymbol{\mu}_i + \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{b}_i, \sigma^2 I_{n_i})$. For Bayesian analysis of the random effects model [Zeger and Karim, 1991; Gilks et al., 1993], we specify the priors on $\boldsymbol{\mu}_i$, $\boldsymbol{\beta}$, $\boldsymbol{D}$ and σ^2 traditionally, as the prior for μ is chosen to be $P(\mu) \stackrel{d}{=} N(\mu_0, \sigma_\mu^2)$, the prior for $\boldsymbol{\beta}$ is chosen to be $P(\boldsymbol{\beta}) \stackrel{d}{=} N(0, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$ and the prior for σ^2 is chosen to be $P(\sigma^2) \stackrel{d}{=} IG(\frac{\delta_0}{2}, \frac{\gamma_0}{2})$. By far the most common approach to the prior on $\boldsymbol{D}$ is to use an inverse-Wishart prior, motivated by its conjugacy property. That is, $P(\boldsymbol{D}) \stackrel{d}{=} IW(n_0, \boldsymbol{C}_0)$. However, there are many other choices of prior on the covariance $\boldsymbol{D}$.

Modeling a covariance structure is one of the most difficult and important tasks in statistical analysis [Barnard et al., 2000]. A covariance matrix may have many parameters, which are constrained by a complex requirement that the matrix is nonnegative definite. There is no standard solution to the problem of choosing a prior on the covariance matrix in the mixed effects model, or hierarchical model [Kass and Natarajan, 2006]. Directly specifying a reasonable prior for a covariance matrix is not a easy task.

The usual inverse-Wishart prior is often inadequate because it is of restrictive form due to the common degrees of freedom for all the diagonal entries of $\boldsymbol{D}$. It is helpful to break the covariance matrix down into several components. There are several methods based on well-known matrix decompositions. Barnard et al. [2000] worked with the variance-correlation decomposition of the covariance matrix. Boik [2002] proposed a spectral decomposition on the matrix. Another approach is to use the Cholesky decomposition of the covariance matrix [Pourahmadi, 1999; Chen and Dunson, 2003].

Inverse Wishart Prior

Suppose $\boldsymbol{A}$ is a $p \times p$ positive definite random matrix. The Wishart distribution with n_0 degree of freedom is characterized by $\boldsymbol{A} \sim W(n_0, \boldsymbol{C}_0^{-1})$ if and only if $\boldsymbol{A} = \sum_{i=1}^{n_0} \boldsymbol{z}_i \boldsymbol{z}_i^T$, where p dimensional vectors $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$ are i.i.d. random samples from $N(0, \boldsymbol{C}_0^{-1})$. The diagonal elements of a Wishart matrix $\boldsymbol{A}$ are chi-square random variables. That is, $a_{ii} \sim \sigma_{ii}\chi_n^2$, where a_{ii} is the ith diagonal element of $\boldsymbol{A}$ and σ_{ii} is the ith diagonal element of $\boldsymbol{C}_0^{-1}$. If $\boldsymbol{D} = \boldsymbol{A}^{-1}$, $\boldsymbol{D}$ have an inverse Wishart distribution, denoted by $\boldsymbol{D} \sim IW(n_0, \boldsymbol{C}_0)$. The mean of inverse Wishart distribution is $E(\boldsymbol{D}) = \frac{\boldsymbol{C}_0}{n_0 - p - 1}$, which means large n_0 makes the prior relatively noninformative. Based on the inverse Wishart prior and other priors as described earlier, the conditional posterior distribution of $\boldsymbol{D}$ is $P(\boldsymbol{D}|\boldsymbol{y}, \boldsymbol{\mu}, \boldsymbol{b}, \sigma^2) \overset{d}{=} IW(n + n_0, \boldsymbol{C}_0 + \sigma^{-2} \sum_{i=1}^{n} \boldsymbol{b}_i \boldsymbol{b}_i^T)$ where $\boldsymbol{y} = (\boldsymbol{y}_i, ..., \boldsymbol{y}_n)^T$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_i, ..., \boldsymbol{\mu}_n)^T$ and $\boldsymbol{b} = (\boldsymbol{b}_i, ..., \boldsymbol{b}_n)^T$. The inverse Wishart prior results in a strong dependence between variance and correlation: high variance implies high correlation and low variance implies low or moderate correlation. To overcome this problem, the inverse Wishart prior for $\boldsymbol{D}$ where the scale matrix is determined from the variance can be considered [Kass and Natarajan, 2006].

Variance-Correlation Decomposition

The covariance matrix $\boldsymbol{D}$ can be decomposed as $\boldsymbol{D} = \boldsymbol{SRS}$ where $\boldsymbol{S}$ is the $k \times k$ diagonal matrix of standard deviations and $\boldsymbol{R}$ is the $k \times k$ correlation matrix [Barnard et al., 2000]. This decomposition has a strong practical appeal since these two factors of $\boldsymbol{D}$ are easily interpreted in terms of standard deviations and correlations. We can write the prior on $\boldsymbol{D}$ in terms of $(\boldsymbol{S}, \boldsymbol{R})$ as $P(\boldsymbol{S}, \boldsymbol{R}) = P(\boldsymbol{S})P(\boldsymbol{R}|\boldsymbol{S})$. Since $\boldsymbol{S}$ only have k-dimensional elements with component-wise nonnegativity as the only constraint, we can specify the prior for $\boldsymbol{S}$ as $P(log(\boldsymbol{s})) \overset{d}{=} N(\boldsymbol{\psi}, \boldsymbol{\Lambda})$ where $\boldsymbol{s} = (s_1, ..., s_n)$ are diagonal elements of $\boldsymbol{S}$ and $log(\boldsymbol{s}) = (log(s_1), ..., log(s_n))$. The choice of prior for $\boldsymbol{R}$ given $\boldsymbol{S}$ is more complicated due to the complexity of space of correlation matrices, and often a marginally uniform prior or jointly uniform prior are used for $\boldsymbol{R}$ [Barnard et al., 2000].

Spectral Decomposition

The spectral decomposition of a covariance matrix $\boldsymbol{D}$ is given by $\boldsymbol{D} = \boldsymbol{P\Lambda P}^T = \sum_{i=1}^{q} \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^T$, where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues with the ith element λ_i and $\boldsymbol{P}$ is the orthogonal matrix of normalized eigenvectors with the ith column, $\boldsymbol{e}_i$ [Pourahmadi, 2011]. There are three classes of priors on $\boldsymbol{\Lambda}$ and $\boldsymbol{P}$ and details can be found in [Leonard and Hsu, 1992; Yang and Berger, 1994; Daniels and Kass, 1999].

Cholesky Decomposition

The standard Cholesky decomposition of a positive-definite matrix is given by $\boldsymbol{D} = \boldsymbol{LL}^T$, where $\boldsymbol{L}$ is a unique lower-triangular matrix with positive diagonal elements. Statistical interpretation of the elements of $\boldsymbol{L}$ is difficult in the present form. However, rescaling $\boldsymbol{L}$ to unit lower-triangular matrices by the inverse of $\boldsymbol{\Delta} = diag(\delta_i, ..., \delta_k)$ makes the statistical interpretation of the diagonal elements of $\boldsymbol{L}$ and the components of the modified Cholesky decomposition easier. Pourahmadi [1999] proposed the following

modified Cholesky decomposition

$$\boldsymbol{D} = \boldsymbol{L}\boldsymbol{\Delta}^{-1}\boldsymbol{\Delta}\boldsymbol{\Delta}\boldsymbol{\Delta}^{-1}\boldsymbol{L}^T = \boldsymbol{\Psi}\boldsymbol{\Delta}^2\boldsymbol{\Psi}^T, \tag{1.37}$$

where $\boldsymbol{\Psi} = \boldsymbol{L}\boldsymbol{\Delta}^{-1}$. Chen and Dunson [2003] presented another modified Cholesky decomposition as follows.

$$\boldsymbol{D} = \boldsymbol{\Delta}\boldsymbol{\Delta}^{-1}\boldsymbol{L}\boldsymbol{L}^T\boldsymbol{\Delta}^{-1}\boldsymbol{\Delta} = \boldsymbol{\Delta}\boldsymbol{\Psi}\boldsymbol{\Psi}^T\boldsymbol{\Delta}, \tag{1.38}$$

where $\boldsymbol{\Psi} = \boldsymbol{\Delta}^{-1}\boldsymbol{L}$ is obtained from $\boldsymbol{L}$ by dividing the elements of its ith row by δ_i. In this decomposition, $\boldsymbol{\Delta}$ is a diagonal matrix with elements proportional to the square roots of the diagonal elements of $\boldsymbol{D}$ and $\boldsymbol{\Psi}$ is a unit lower-triangular matrix solely determining its correlation matrix. This total separation of variance and correlation is one of advantages over the more traditional modified Cholesky decomposition of Pourahmadi [1999] [Pourahmadi, 2007]. Given Chen and Dunson [2003]'s decomposition, the reparameterized mixed effects model can be written as

$$\boldsymbol{y}_i = \boldsymbol{\mu}_i + \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{\Delta}\boldsymbol{\Psi}\boldsymbol{c}_i + \boldsymbol{e}_i, \text{ i=1,...,n,} \tag{1.39}$$

where $\boldsymbol{c}_i = (c_1, ..., c_k)^T$ is $k \times 1$ vector of independent standard normal latent variables. In choosing priors for $\boldsymbol{\Delta}$ and $\boldsymbol{\Psi}$ and hence for $\boldsymbol{D}$, we permit the variance of random effects to have zero values [Chen and Dunson, 2003]. Conditionally conjugate prior distributions for $\boldsymbol{\Delta}$ and $\boldsymbol{\Psi}$ are chosen as $P(\boldsymbol{\Delta}, \boldsymbol{\psi}) = P(\boldsymbol{\psi}|\boldsymbol{\Delta})P(\boldsymbol{\Delta})$. The prior distribution for $\boldsymbol{\psi}$ conditional on $\boldsymbol{\Delta}$ is given by $P(\boldsymbol{\psi}|\boldsymbol{\Delta}) \propto N(\boldsymbol{\psi}|\boldsymbol{\psi}_0, \boldsymbol{R}_0)I(\boldsymbol{\psi} \in \boldsymbol{\mathfrak{R}}_{\boldsymbol{\Delta}})$ where $\boldsymbol{\Delta} = (\delta_1, ..., \delta_k)^T$, $\boldsymbol{\psi} = (\psi_{ml} : m = 2, ..., k; l = 1, ..., m-1)^T$ and $\boldsymbol{\mathfrak{R}}_{\boldsymbol{\Delta}} = \{\boldsymbol{\psi} : \psi_{ml} = \psi_{lm'} = 0 \text{ if } \delta_l = 0, l = 1, ..., k, m = l+1, ..., k, m' = 1, ..., l-1\}$. Here, the prior for $\boldsymbol{\psi}$, conditional on $\boldsymbol{\Delta}$, is proportional to a $N(\boldsymbol{\psi}_0, \boldsymbol{R}_0)$ multiplied by an indicator function that imposes

zero on the elements of $\boldsymbol{\psi}$ corresponding to zero $\boldsymbol{\Delta}$. For $\boldsymbol{\Delta}$, we assume that the δ's are independent of each other, so that $P(\boldsymbol{\Delta}) = \Pi_{l=1}^{k} P(\delta_l)$. Let $ZI - N^{+}(\pi, \mu, \sigma^2)$ denote the density of a zero inflated half normal distribution comprised of a point mass at zero with probability π and a $N(\mu, \sigma^2)$ density truncated below by zero. The prior distribution for $\boldsymbol{\Delta}$ is given by $P(\boldsymbol{\Delta}) = \Pi_{l=1}^{k} P(\delta_l) = \Pi_{l=1}^{k} ZI - N^{+}(\delta_l | p_{l0}, m_{l0}, s_{l0}^2)$ where p_{l0},m_{l0} and s_{l0}^2 are hyperparameters to be specified.

1.5 Outline of Thesis

In the present chapter, we have reviewed traditional QTL mapping, Bayesian QTL mapping methods and Gaussian process models. And we have introduced the literature on Bayesian model selection and Bayesian covariance estimation. In Chapter 2, we develop a Bayesian multiple QTL mapping method with a composite model space framework for longitudinal traits, and we apply the proposed method to the Genetic Analysis Workshop 18 (GAW18) longitudinal blood pressure data. The method is computationally efficient, but it only allows for pairwise gene-gene and gene-time/environment interactions and may miss genes with higher-order interactions. To overcome this difficulty, in Chapter 3, we propose a nonparametric Gaussian process model for longitudinal traits, which measures the importance of each QTL irrespective of whether it functions through main, epistatic effects, or interactions with environmental factors. Finally, in Chapter 4, we extend the nonparametric Gaussian process model to SNP-set analysis to map groups of rare and/or common variants.

CHAPTER 2

BAYESIAN MULTIPLE QTL MAPPING FOR LONGITUDINAL TRAITS

2.1 Introduction

Many complex traits, such as blood pressure, cholesterol level are time-dependent or longitudinal traits and are affected by genetic factors as well as nongenetic factors (e.g. age, sex or drug). It is crucial to consider the repeated measurements of traits for better understanding their genetic architectures. Over the past two decades, a diversity of statistical methodologies have been developed to map quantitative trait loci (QTL) for complex traits [Lander and Botstein, 1989; Zeng, 1993; Kao and Zeng, 1997; Yi et al., 2007]. Although these methods are effective in detecting QTL, they are not readily applicable to identify QTL with time-varying genetic effects. Recently, it becomes of great interest to study genes with time varying genetic effects through collecting time-dependent traits repeatedly over time.

Several different approaches are currently available for genetic analyses of longitudinal traits. For the data collected at the same time points across all individuals, the measured values at each time point can be treated as one variable and jointly analyzed by treating such longitudinal data as multivariate outcomes [Wu et al., 1999, 2002; Ma et al., 2002; Yap et al., 2009]. For the data collected at different time points across some or all individuals, the measured values can not be effectively grouped and thus the

multivariate analysis is not applicable. Alternatively, mixed effects models are applied to QTL mapping for longitudinal traits and a maximum-likelihood method is used for parameter estimation and statistical tests. [Yang et al., 2006]. Mixed effects models are flexible for modeling non-constant correlation among observations and unbalanced data. Moreover, the correlation of measurements at successive time points can be used to interpolate the adjacent time points. A flexible nonparametric time-varying coefficient QTL mapping method for recombinant inbred intercrosses (RIX) data models the varying genetic effects nonparametrically with the B-spline bases and models the polygenic effects via mixed effects model [Gong and Zou, 2012].

Most approaches mentioned above test one gene at a time and may have low power to map multiple genes that jointly affect the trait. Several Bayesian methods for multiple QTL mapping [Satagopan et al., 1996; Yi et al., 2002, 2003; Yi, 2004] have been proposed. Multiple QTL can be simultaneously detected by treating the number of QTL as a random variable using reversible jump Markov chain Monte Carlo (MCMC) method [Yi et al., 2002]. Alternatively, multiple QTL can be viewed as a variable selection problem [Yi et al., 2003; Yi, 2004] and Bayesian model selection method is used for identifying main, epistatic QTL [Yi et al., 2005] and QTL interacting with other covariates [Yi et al., 2007] based on the composite model space framework. These approaches use a fixed dimensional parameter space by setting a upper bound on the number of detectable QTL and introduce latent binary variables to indicate which factors should be included in or excluded from the model. It can reasonably reduce the model space and construct efficient MCMC algorithm. Banerjee et al. [2008] extended Bayesian variable selection method of Yi [2004] to multiple traits via the "seemingly unrelated regression" (SUR) model.

For Bayesian analysis under the mixed effects model framework [Zeger and Karim,

1991; Gilks et al., 1993], the covariance structure of random effects needs to be modeled. A large covariance matrix has numerous parameters, which are constrained by the fact that the covariance matrix is nonnegative definite. Thus, directly specifying a reasonable prior for a covariance matrix is not a simple task. There is no standard solution to the problem of choosing a prior on the covariance matrix in the mixed effects model [Kass and Natarajan, 2006]. The usual inverse-Wishart prior is often inadequate because of its restrictive form on the common degrees of freedom for all the diagonal entries. Moreover, the inverse Wishart prior leads to a strong dependence between variance and correlation: high variance implies high correlation and low variance implies low or moderate correlation. Alternatively, it may be helpful to decompose the covariance matrix into several components and model each component separately. Barnard et al. [2000] worked with a variance-correlation decomposition of the covariance matrix. Boik [2002] proposed to use a spectral decomposition. Another approach is based on the Cholesky decomposition [Pourahmadi, 1999; Chen and Dunson, 2003]. Chen and Dunson [2003]'s method has an advantage over the more traditional modified Cholesky decomposition of Pourahmadi [1999] thanks to the total separation of variance and correlation [Pourahmadi, 2007].

In this chapter, we develop a Bayesian multiple QTL model which extends the composite model space framework of Yi et al. [2007] to longitudinal traits. For data where phenotypes are not measured at a fixed set of times for all samples, we parsimoniously describe the covariance matrix of each subject as a covariance matrix predefined on a set of pre-selected time points. For those not measured at the pre-selected time points, we map each observed time point to two nearest adjacent grid time points via the linear interpolation. This approach only deals with a covariance matrix with a fixed dimension. The covariance matrix is modeled nonparametrically and we employ a modified Cholesky decomposition of Chen and Dunson [2003]. Such decomposition facilitates

the use of normal conjugate priors. The proposed method jointly models the main and pairwise interactions of all candidate genetic variants.

2.2 Bayesian Multiple QTL Model for Longitudinal Data

2.2.1 Bayesian Mixed Effects Model

Suppose there are n subjects, with subject i having phenotypes measured at n_i time points ($i = 1, ..., n$). First, we divide the entire genome into h loci, $\boldsymbol{\kappa} = (\kappa_1, ..., \kappa_h)^T$ and assume that all putative QTL are located at these fixed h loci. When the markers ($\boldsymbol{m}$) are dense enough, we set $\boldsymbol{\kappa}$ to the marker positions. If not, $\boldsymbol{\kappa}$ may contain points between markers in addition to the marker positions. Before mapping QTL, the conditional genotypic probabilities of $\boldsymbol{g}$ at loci $\boldsymbol{\kappa}$, $p(\boldsymbol{g}|\boldsymbol{\kappa}, \boldsymbol{m})$ can be computed from the observed marker data using multipoint method [Jiang and Zeng, 1997]. An upper bound on the number of QTL in the model is set to p which is usually much smaller than h. Based on the conditional probabilities of genotypes across the genome, we construct main effects (p terms), gene-gene interactions ($p(p-1)/2$ terms) and gene-time/gene-environment interactions (pq terms) where q is the number of time/environmental covariates [Yi et al., 2005].

Let $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_r)^T$ be the current positions of r putative QTL where $r = p + p(p-1)/2 + pq$. Each locus can affect the trait through its marginal effects (main effects) or two-way interactions with other loci (epistatic effects) or with environmental effects (gene-time/gene-environment interactions). We use latent binary variables $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_r)^T$ for indicating which effects are included in ($\gamma_i = 1$) or excluded from ($\gamma_i = 0$) the model. The vector of indicators and positions ($\boldsymbol{\gamma}$,$\boldsymbol{\lambda}$) determines the number and positions of QTL. For the ith sample, let $\boldsymbol{x}_{ti}$ denote the $n_i \times q$ design matrix of time/environmental covariates, $\boldsymbol{x}_{gi}$ denote the $n_i \times p$ design matrix of p putative QTL

genotypes, $\boldsymbol{x}_{ggi}$ denote the $n_i \times p(p-1)/2$ design matrix of epistatic effects, and $\boldsymbol{x}_{gti}$ denote the $n_i \times pq$ design matrix of gene-time/gene-environment interactions, which results in the design matrix $\boldsymbol{x}_i = (\boldsymbol{x}_{ti}, \boldsymbol{x}_{gi}, \boldsymbol{x}_{ggi}, \boldsymbol{x}_{gti})$.

Given $\boldsymbol{\gamma}, \boldsymbol{\lambda}$ and $\boldsymbol{x}_i$, we consider the following Bayesian mixed effects model:

$$\boldsymbol{y}_i = \boldsymbol{\mu}_i + \boldsymbol{x}_i \boldsymbol{\Gamma} \boldsymbol{\beta} + \boldsymbol{p}_i \boldsymbol{\nu}_i + \boldsymbol{e}_i \text{ (i=1,...,n)}, \tag{2.1}$$

where $\boldsymbol{y}_i = (y_{i1}, ..., y_{in_i})^T$ is the $n_i \times 1$ phenotype or trait vector of subject i; $\boldsymbol{\mu}_i = \mu \mathbf{1}_{n_i}$ is the $n_i \times 1$ vector of the overall mean; $\boldsymbol{\Gamma}$ is the diagonal matrix with diagonal elements $(\mathbf{1}_q,\ \boldsymbol{\gamma})$; $\boldsymbol{\beta} = (\boldsymbol{\beta}_t^T, \boldsymbol{\beta}_g^T, \boldsymbol{\beta}_{gg}^T, \boldsymbol{\beta}_{gt}^T)^T$ is the vector of genetic effects, time/environmental effects, epistatic effects, gene-time/gene-environment interactions; $\boldsymbol{e}_i$ is an $n_i \times 1$ vector of random error with $\boldsymbol{e}_i \sim N(0, \sigma_e^2 \boldsymbol{I}_{n_i})$. In order to model the correlation among the repeated measurements of the same individual, we first partition the observed time interval by k grid points, $\boldsymbol{t} = (t_1, ..., t_k)^T$. Then, we define $\boldsymbol{\nu}_i$ as a $k \times 1$ vector of random effects at the grid time points with $\boldsymbol{\nu}_i \sim N(0, \boldsymbol{D})$ where $\boldsymbol{D}$ is an $k \times k$ covariance matrix. If all traits are observed exactly on the k grid time points, matrix $\boldsymbol{p}_i$ is an identity matrix. If we have samples whose phenotypes are not measured on the grid points, an interpolation procedure (e.g. linear, polynomial or spline) can be applied for approximately modeling the within-subject correlations. For simplicity, we choose a linear interpolation. Let the incidence matrix $\boldsymbol{p}_i = (\boldsymbol{p}_{i1}^T, ..., \boldsymbol{p}_{in_i}^T)^T$. If jth time point of ith individual, x_{tij}, is between t_1 and t_2 $(t_1 \le x_{tij} \le t_2)$, then $\boldsymbol{p}_{ij} = (\frac{t_2 - x_{tij}}{t_2 - t_1}, \frac{x_{tij} - t_1}{t_2 - t_1}, \mathbf{0})$. If $x_{tij} = t_2$, then $\boldsymbol{p}_{ij} = (0, 1, \mathbf{0})$.

2.2.2 Reparameterized Model

For Bayesian estimation of mixed effects model (2.1), we conduct the factorization of the covariance matrix, $\boldsymbol{D}$, of the random effects following the modified Cholesky

decomposition of Chen and Dunson [2003]. Let $\boldsymbol{L}$ denote a $k \times k$ lower triangular Cholesky decomposition matrix which has nonnegative diagonal elements, such that $\boldsymbol{D} = \boldsymbol{L}\boldsymbol{L}^T$. Let $\boldsymbol{L} = \boldsymbol{\Delta}\boldsymbol{\Psi}$ where $\boldsymbol{\Delta} = \text{diag}(\delta_1, ..., \delta_k)$ and $\boldsymbol{\Psi}$ is a $k \times k$ matrix with the (l, m) element denoted by ψ_{lm}. To make $\boldsymbol{\Delta}$ and $\boldsymbol{\Psi}$ identifiable, we make the following assumptions:

$$\delta_l \geq 0, \ \psi_{ll} = 1 \text{ and } \psi_{lm} = 0, \text{ for } l = 1, ..., k; \ \ m = l + 1, ..., k. \tag{2.2}$$

These conditions make $\boldsymbol{\Delta}$ a nonnegative $k \times k$ diagonal matrix and $\boldsymbol{\Psi}$ a lower triangular matrix with 1's in the diagonal elements. This results in the following decomposition of $\boldsymbol{D}$,

$$\boldsymbol{D} = \boldsymbol{\Delta}\boldsymbol{\Psi}\boldsymbol{\Psi}^T\boldsymbol{\Delta}. \tag{2.3}$$

Based on the modified Cholesky decomposition of $\boldsymbol{D}$, we reparameterize model (2.1) as

$$\boldsymbol{y}_i = \boldsymbol{\mu}_i + \boldsymbol{x}_i\boldsymbol{\Gamma}\boldsymbol{\beta} + \boldsymbol{p}_i\boldsymbol{\Delta}\boldsymbol{\Psi}\boldsymbol{b}_i + \boldsymbol{e}_i \text{ (i=1,...,n)}, \tag{2.4}$$

where $\boldsymbol{b}_i = (b_{i1}, ...b_{ik})^T$ such that $b_{ij} \sim N(0, 1)$ and $b_{ij} \perp b_{ij'}$ $(j \neq j')$, $j = 1, ..., k$. For the later use, we define $\boldsymbol{v}_i = \boldsymbol{p}_i\boldsymbol{\Delta}\boldsymbol{\Psi} = (\boldsymbol{v}_{i1}, ..., \boldsymbol{v}_{in_i})^T$ and $\boldsymbol{v} = \text{diag}(\boldsymbol{v}_1, ..., \boldsymbol{v}_n)$.

2.2.3 Identifiability Problem of the Covariance

For statistical estimation and inference, the proposed Bayesian mixed effects model should be identifiable. However, the identifiability problem arises in the estimation of the covariance matrix of the phenotypes. Note that $\boldsymbol{y}[= (\boldsymbol{y}_1^T, ..., \boldsymbol{y}_n^T)^T]$ follows multivariate normal distribution. The covariance matrix of $\boldsymbol{y}$ is given by $\boldsymbol{P}\boldsymbol{\mathcal{D}}\boldsymbol{P}^T + \sigma^2\boldsymbol{I}_N$ where $\boldsymbol{P} = diag(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, $\boldsymbol{\mathcal{D}} = \boldsymbol{I}_n \otimes \boldsymbol{D}$ and $N = \sum_{i=1}^n n_i$. It is necessary to find the conditions that $\boldsymbol{P}\boldsymbol{\mathcal{D}}\boldsymbol{P}^T + \sigma^2\boldsymbol{I}_N$ is identifiable, which is equivalent to the fact that $\boldsymbol{P}\boldsymbol{\mathcal{D}}\boldsymbol{P}^T + \sigma^2\boldsymbol{I}_N = \boldsymbol{P}\hat{\boldsymbol{\mathcal{D}}}\boldsymbol{P}^T + \hat{\sigma^2}\boldsymbol{I}_N$ if and only if $\hat{\boldsymbol{\mathcal{D}}} = \boldsymbol{\mathcal{D}}$ and $\hat{\sigma^2} = \sigma^2$. Letting $\tilde{\boldsymbol{\mathcal{D}}} = \boldsymbol{\mathcal{D}} - \hat{\boldsymbol{\mathcal{D}}}$

and $\tilde{\sigma^2} = \sigma^2 - \hat{\sigma^2}$, we observe that $\boldsymbol{P}\boldsymbol{\mathcal{D}}\boldsymbol{P}^T + \sigma^2\boldsymbol{I}_N$ is identifiable if and only if the system of equations $\boldsymbol{P}\tilde{\boldsymbol{\mathcal{D}}}\boldsymbol{P}^T + \tilde{\sigma^2}\boldsymbol{I}_N = \boldsymbol{0}$ has no non-zero solutions for $\tilde{\boldsymbol{\mathcal{D}}}$ and $\tilde{\sigma^2}$.

Note that the jth row vector of $\boldsymbol{p}_i$ can be expressed as $\boldsymbol{p_{ij}} = a_{ij1}\boldsymbol{e}_1 + ... + a_{ijk}\boldsymbol{e}_k$ where $\sum_{r=1}^{k} a_{ijr} = 1$, $0 \le a_{ij1}, ..., a_{ijk} \le 1$ and $\boldsymbol{e}_r$ $(1 \le r \le k)$ is a $1 \times k$ unit row vector whose elements are all zero except the rth component, which is one. Note that one or only two adjacent a_{ijr} values are non-zero due to our linear interpolation. Further, let the (r, s)th element of $\tilde{\boldsymbol{D}}$ (i.e. $\tilde{\boldsymbol{\mathcal{D}}} = \boldsymbol{I}_n \otimes \tilde{\boldsymbol{D}}$) be $\tilde{d}_{r,s}$. The following lemma and theorem help to evaluate the identifiability problem and give us a useful information on selecting the number of grid points.

Lemma 1. *The system of equations* $\boldsymbol{P}\tilde{\boldsymbol{\mathcal{D}}}\boldsymbol{P}^T + \tilde{\sigma^2}\boldsymbol{I}_N = \boldsymbol{0}$ *where* $\tilde{\boldsymbol{\mathcal{D}}} = \boldsymbol{\mathcal{D}} - \hat{\boldsymbol{\mathcal{D}}}$ *and* $\tilde{\sigma^2} = \sigma^2 - \hat{\sigma^2}$ *is equivalent to the system of equations* $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{0}$ *where* $\boldsymbol{A}$ *is a* $\left[\frac{1}{2}\sum_{i=1}^{n} n_i(n_i+1)\right] \times \left[\frac{1}{2}k(k+1)+1\right]$ *matrix of constants and* $\boldsymbol{X}$ *is a* $\left[\frac{1}{2}k(k+1)+1\right] \times 1$ *vector containing all elements of matrix* $\tilde{\boldsymbol{\mathcal{D}}}$ *plus* $\tilde{\sigma^2}$.

Proof. Since only one or two adjacent a_{ijr} are nonzero, $\boldsymbol{p}_{ij}$ can be expressed as $\boldsymbol{p}_{ij} = a_{ij(c_{ij})}\boldsymbol{e}_{(c_{ij})} + a_{ij(c_{ij}+1)}\boldsymbol{e}_{(c_{ij}+1)}$ where $[c_{ij}, c_{ij}+1]$ refers the time interval on which the jth time of the ith subject falls and $1 \le c_{ij} \le k-1$. Note that the (j, j')th element of $\boldsymbol{p}_i\tilde{\boldsymbol{D}}\boldsymbol{p}_i^T + \tilde{\sigma^2}\boldsymbol{I}_{n_i}$ equals $\boldsymbol{p}_{ij}\tilde{\boldsymbol{D}}\boldsymbol{p}_{ij'}^T + \tilde{\sigma^2}I(j = j')$. Since $\boldsymbol{e}_r\tilde{\boldsymbol{D}}\boldsymbol{e}_s^T = \tilde{d}_{r,s}$ $(1 \le r, s \le k)$, $\boldsymbol{p}_{ij}\tilde{\boldsymbol{D}}\boldsymbol{p}_{ij'}^T + \tilde{\sigma^2}I(j = j') = a_{ij(c_{ij})}a_{ij'(c_{ij'})}\boldsymbol{e}_{(c_{ij})}\tilde{\boldsymbol{D}}\boldsymbol{e}_{(c_{ij'})}^T + a_{ij(c_{ij})}a_{ij'(c_{ij'}+1)}\boldsymbol{e}_{(c_{ij})}\tilde{\boldsymbol{D}}\boldsymbol{e}_{(c_{ij'}+1)}^T + a_{ij(c_{ij}+1)}a_{ij'(c_{ij'})}\boldsymbol{e}_{(c_{ij}+1)}\tilde{\boldsymbol{D}}\boldsymbol{e}_{(c_{ij'})}^T + a_{ij(c_{ij}+1)}a_{ij'(c_{ij'}+1)}\boldsymbol{e}_{(c_{ij}+1)}\tilde{\boldsymbol{D}}\boldsymbol{e}_{(c_{ij'}+1)}^T + \tilde{\sigma^2}I(j = j') = a_{ij(c_{ij})}a_{ij'(c_{ij'})}\tilde{d}_{(c_{ij}),(c_{ij'})} + a_{ij(c_{ij})}a_{ij'(c_{ij'}+1)}\tilde{d}_{(c_{ij}),(c_{ij'}+1)} + a_{ij(c_{ij}+1)}a_{ij'(c_{ij'})}\tilde{d}_{(c_{ij}+1),(c_{ij'})} + a_{ij(c_{ij}+1)}a_{ij'(c_{ij'}+1)}\tilde{d}_{(c_{ij}+1),(c_{ij'}+1)} + \tilde{\sigma^2}I(j = j')$. Therefore, $\boldsymbol{p}_i\tilde{\boldsymbol{D}}\boldsymbol{p}_i^T + \tilde{\sigma^2}\boldsymbol{I}_{n_i} = \boldsymbol{0}$ is equivalent to the system of equations, $\boldsymbol{A}_i\boldsymbol{X} = \boldsymbol{0}$ where $\boldsymbol{A}_i$ is a $\left[\frac{1}{2}n_i(n_i+1)\right] \times \left[\frac{1}{2}k(k+1)+1\right]$ matrix of constants and $\boldsymbol{X} = (\tilde{d}_{1,1}, \tilde{d}_{1,2}, \cdots, \tilde{d}_{1,k}, \tilde{d}_{2,2}, \cdots, \tilde{d}_{k,k}, \tilde{\sigma^2})^T$ is a $\left[\frac{1}{2}k(k+1)+1\right] \times 1$ vector. Therefore, we have $\boldsymbol{P}\tilde{\boldsymbol{\mathcal{D}}}\boldsymbol{P}^T + \tilde{\sigma^2}\boldsymbol{I}_N = \boldsymbol{0} \Leftrightarrow \boldsymbol{p}_i\tilde{\boldsymbol{D}}\boldsymbol{p}_i^T + \tilde{\sigma^2}\boldsymbol{I}_{n_i} = \boldsymbol{0}$ for $\forall i \Leftrightarrow \boldsymbol{A}\boldsymbol{X} = \boldsymbol{0}$ where $\boldsymbol{A} = (\boldsymbol{A}_1^T, ..., \boldsymbol{A}_n^T)^T$ is an $\left[\frac{1}{2}\sum_{i=1}^{n} n_i(n_i+1)\right] \times \left[\frac{1}{2}k(k+1)+1\right]$ matrix. This completes the proof. □

Theorem 1. *The proposed Bayesian mixed effects model (2.4) is identifiable if and*

only if rank($\boldsymbol{A}$)$=\frac{1}{2}k(k+1)+1$.

Proof. The proof follows from Lemma 1. See Schott [2005] for details. □

The above lemma and theorem enable us to check whether the model is identifiable. Below we show one toy example on how to use Theorem 1 to check the identifiability. Suppose there are 3 grid points which produce 2 time intervals. If the model is identifiable, the rank of $\boldsymbol{A}$ should be $\frac{1}{2}3(3+1)+1=7$ based on Theorem 1. If the phenotypes of all individuals are observed exactly on the 3 grid points, then $\boldsymbol{p}_i=\boldsymbol{I}_3$, so that $\boldsymbol{A}_1=\cdots=\boldsymbol{A}_n=\begin{pmatrix} 1&0&0&0&0&0&1\\ 0&1&0&0&0&0&0\\ 0&0&1&0&0&0&0\\ 0&0&0&1&0&0&1\\ 0&0&0&0&1&0&0\\ 0&0&0&0&0&1&1 \end{pmatrix}$ and $\boldsymbol{X}=\begin{pmatrix} \tilde{d}_{1,1}\\ \tilde{d}_{1,2}\\ \tilde{d}_{1,3}\\ \tilde{d}_{2,2}\\ \tilde{d}_{2,3}\\ \tilde{d}_{3,3}\\ \tilde{\sigma^2} \end{pmatrix}$ and the rank of $\boldsymbol{A}$ is $\frac{1}{2}3(3+1)=6$. Therefore, the model is non-identifiable. However, if one individual has phenotypes measures at different time points from the other individuals, the model becomes identifiable since the rank of $\boldsymbol{A}$ is now 7.

2.3 Prior Specifications

In order to complete Bayesian modeling, we need to specify priors for all unknown parameters. The proposed Bayesian mixed effects model contains parameters for both random and fixed effects. For the random effects, we follow the priors in Chen and Dunson [2003]. Specifically, we impose independent half normal priors on the diagonal elements of $\boldsymbol{\Delta}$ and normal priors on the lower triangular elements of $\boldsymbol{\Psi}$. For the fixed effects, we straightforwardly extend the priors presented in Yi et al. [2005, 2007].

Prior on $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$

We assume that all inclusion probabilities are independent of each other. Letting

$w_k = P(\gamma_k = 1)$ be the inclusion probability for the kth effect, we have the following independent prior on the indicator vector $\boldsymbol{\gamma}$: $P(\boldsymbol{\gamma}) = \prod_{k=1}^{r} w_k^{\gamma_k}(1-w_k)^{1-\gamma_k}$. The inclusion probability w_k is predetermined and varies according to whether it corresponds to a main effect, epistasis effect or gene-time/gene-environment interaction [Yi et al., 2005]. Setting w_k to a small value ensures that the model contains a small number of main effects, epistasis effects, gene-time/gene-environment interactions. We first specify the prior expected number of QTL with main effect, p_m, and all QTL, p_0 ($p_0 > p_m$), based on initial investigation and then choose a reasonably large upper bound, p, on the number of all QTL. We have the following hyperparameters: $w_m = 1 - [1 - \frac{p_m}{p}]^{1/g_1}$ for main effects, $w_e = 1 - [\frac{1-(p_0/p)}{(1-w_m)^{g_1}}]^{1/g_2(p-1)}$ for epistasis effects and $w_t = \frac{p_m}{p}$ for gene-time/gene-environment interactions where g_1 is the number of possible main effects for each QTL; g_2 (for example, we set $g_2 = g_1^2$ in this paper) is the number of possible two-way gene-gene interaction (see Yi et al. [2005] for details). To specify a prior on the QTL position vector $\boldsymbol{\lambda}$, we assume that the locations are independent and uniformly distributed over the h possible loci and there is at most one QTL within any given marker interval. Given the expected number of all QTL (p_0), the prior distribution of QTL location $\boldsymbol{\lambda}$ is given by $P(\boldsymbol{\lambda}) = \prod_{k=1}^{r} P(\lambda_k)$ where $P(\lambda_k) = \frac{p_0}{h}$.

Prior on b, Δ and Ψ

In model (2.4), we let the prior of b_{ij} follow a standard normal distribution, so that joint prior distribution for the latent variable $\boldsymbol{b} = (\boldsymbol{b}_1^T, ..., \boldsymbol{b}_n^T)^T$ is $P(\boldsymbol{b}) \stackrel{d}{=} N(0, \boldsymbol{I}_{nk})$. In order to specify priors for $\boldsymbol{\Delta}$ and $\boldsymbol{\Psi}$, we define two vectors $\boldsymbol{\delta} = (\delta_l : l = 1, ..., k)^T$ and $\boldsymbol{\psi} = (\psi_{ml} : m = 2, ..., k; l = 1, ..., m-1)^T$. For the prior for $\boldsymbol{\delta}$, we assume that the δ_l's are independent of each other, so $P(\boldsymbol{\delta}) = \Pi_{l=1}^{k} P(\delta_l)$. The prior distribution for $\boldsymbol{\delta}$ is $P(\boldsymbol{\delta}) = \Pi_{l=1}^{k} N^+(\delta_l | m_{l0}, s_{l0}^2)$ where $N^+(\delta_l | m_{l0}, s_{l0}^2)$ is the density of a half normal distribution which is a $N(\delta_l | m_{l0}, s_{l0}^2)$ density truncated below by zero. The prior distribution

for $\boldsymbol{\psi}$ is given by $P(\boldsymbol{\psi}) \stackrel{d}{=} N(\boldsymbol{\psi}_0, \boldsymbol{R}_0)$ where $\boldsymbol{\psi}_0$ and $\boldsymbol{R}_0$ are pre-specified hyperparameters.

Prior on $\boldsymbol{\beta}$, μ and σ^2

The genetic effects are first divided into several groups, corresponding to different types of effects (i.e. additive, dominance, additive-additive, additive-time/environment interactions, etc). Suppose the kth genetic effect belongs to group u. All effects in the same group u follow the same prior, $P(\beta_k|\gamma_k, \sigma_u^2) \stackrel{d}{=} N(0, \gamma_k\sigma_u^2)$ and the prior for variance σ_u^2 is an scaled inverse χ^2 distribution, $P(\sigma_u^2) \stackrel{d}{=} \text{inv-}\chi^2(\nu_u, s_u^2)$ whose expected value is $E(\sigma_u^2) = \nu_u s_u^2/(\nu_u - 2)$. The degree of freedom ν_u affects the skewness of the prior for σ_u^2 (we set $\nu_k = 6$) and the scale parameter s_u^2 controls the prior confidence region of the heritability which is the phenotypic variance explained by β_k divided by total phenotypic variance. Letting V_p be the total phenotypic variance and V_k be the sample variance for the column of $\boldsymbol{x}_i$ associated with the effect β_k, the heritability is calculated by $h_k = V_k\beta_k^2/V_p$. Setting $E(\sigma_u^2) = E(\beta_k^2)$, $s_u^2 = (\nu_u - 2)E(\sigma_u^2)/\nu_u = (\nu_u - 2)E(h_k)V_p/(\nu_u V_k)$ (expected effect heritabilities $E(h_k)$ is set to 0.1 for the analysis). The prior for the overall mean μ is given by $P(\mu) \stackrel{d}{=} N(\eta_0, \tau_0^2)$. We could empirically set $\eta_0 = \bar{y} = (1/N)\sum_{i=1}^{n}\sum_{j=1}^{n_i} y_{ij}$ and $\tau_0^2 = s_y^2 = (1/(N-1))\sum_{i=1}^{n}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2$ where $N = \sum_{i=1}^{n} n_i$. The prior for residual variance σ^2 can be chosen by $P(\sigma^2) \propto 1/\sigma^2$, which is noninformative prior for the residual variance σ^2 [Gelman et al., 2004].

2.4 MCMC Algorithm and Posterior Analysis

In this section, we describe the posterior distributions of all unknown parameters and MCMC algorithms. We then discuss how to summarize posterior samples. The joint posterior distribution is proportional to the product of the likelihood and the prior

distributions of all unknown parameters, which can be expressed as

$$P(\boldsymbol{\gamma}, \boldsymbol{\theta}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\theta})P(\boldsymbol{\gamma})P(\boldsymbol{\lambda})P(\boldsymbol{\beta}|\boldsymbol{\gamma})P(\boldsymbol{b})P(\boldsymbol{\delta})P(\boldsymbol{\psi})P(\mu)P(\sigma^2), \quad (2.5)$$

where $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\delta}, \boldsymbol{\psi}, \mu, \sigma^2)^T$. In order to obtain MCMC samples of all parameters, we use both Metropolis-Hastings and Gibbs sampling algorithms, alternately updating each unknown parameter conditional on all the other parameters and the observed data.

Posterior Calculation and MCMC Algorithm

For $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$, we utilize Metropolis-Hastings algorithm since their conditional distributions do not have known distribution forms. To update those parameters, we extend the Metropolis-Hastings algorithm proposed by Yi et al. [2007] for our Bayesian mixed effects model straightforwardly. For the other parameters, Gibbs sampling algorithm is applied. Specifically, since $\boldsymbol{b}$, $\boldsymbol{\delta}$ and $\boldsymbol{\psi}$ have multivariate normal or half normal priors, their conditional distributions are easy to derive by those conjugacy properties.

Conditional Posterior of $\boldsymbol{\gamma}$

The full conditional posterior distribution of the indicator variable γ_k can be expressed as

$$\begin{aligned} P(\gamma_k = 1|\boldsymbol{\gamma}_{-k}, \boldsymbol{\theta}_{-\beta_k}, \boldsymbol{y}) &= 1 - P(\gamma_k = 0|\boldsymbol{\gamma}_{-k}, \boldsymbol{\theta}_{-\beta_k}, \boldsymbol{y}) \\ &= \frac{w_k L_{k1}}{(1-w_k)L_{k0} + w_k L_{k1}}, \end{aligned} \quad (2.6)$$

where $\boldsymbol{\gamma}_{-k}$ represents all the elements in $\boldsymbol{\gamma}$ except γ_k, $\boldsymbol{\theta}_{-\beta_k}$ represents all the elements of $\boldsymbol{\theta}$ except β_k and $L_{km} = P(\boldsymbol{y}|\gamma_k = m, \boldsymbol{\gamma}_{-k}, \boldsymbol{\theta}_{-\beta_k})$ for $m = 1, 0$. Suppose that β_k belongs to group u (i.e. $P(\beta_k|\gamma_k, \sigma_u^2) \overset{d}{=} N(0, \gamma_k\sigma_u^2)$). Note that β_k is integrated out in L_{k1} and

L_{k0}. For L_{k1} and L_{k0}, we first derive the joint distribution of $\boldsymbol{y}$ and β_k.

$$
\begin{aligned}
& P(\boldsymbol{y}, \beta_k | \gamma_k = 1, \boldsymbol{\gamma}_{-k}, \boldsymbol{\theta}_{-\beta_k}) \propto P(\boldsymbol{y} | \gamma_k = 1, \boldsymbol{\gamma}_{-k}, \boldsymbol{\theta}) P(\beta_k | \sigma_u^2) \qquad (2.7) \\
\propto \;& exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} (y_{ij} - \mu - \boldsymbol{x}_{ij}\boldsymbol{\beta} - \boldsymbol{v}_{ij}\boldsymbol{b})^2) exp(-\frac{\beta_k^2}{2\sigma_u^2}) \\
\propto \;& exp(-\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{c_{ij}^2}{\sigma^2} - \frac{\beta_k^2}{\sigma_u^2}) \text{ where } c_{ij} = y_{ij} - \mu - \boldsymbol{x}_{ij}\boldsymbol{\beta} - \boldsymbol{v}_{ij}\boldsymbol{b} \\
\propto \;& exp(-\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{(c_{ij} + x_{ijk}\beta_k)^2 - 2x_{ijk}\beta_k(c_{ij} + x_{ijk}\beta_k) + x_{ijk}^2\beta_k^2}{\sigma^2} - \frac{\beta_k^2}{\sigma_u^2}) \\
\propto \;& exp(-\frac{1}{2}\{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}(c_{ij} + x_{ijk}\beta_k)^2}{\sigma^2} - (\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i} x_{ijk}(c_{ij} + x_{ijk}\beta_k)}{\sigma^2})^2 (\tilde{\sigma_u}^2)^{-1} + \frac{(\beta_k - \tilde{\mu_k})^2}{\tilde{\sigma_u}^2}\}),
\end{aligned}
$$

where $\tilde{\mu_k} = (\tilde{\sigma_u}^2)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} x_{ijk}(c_{ij} + x_{ijk}\beta_k)/\sigma^2$ and $\tilde{\sigma_u}^2 = \sigma_u^{-2} + \sigma^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} x_{ijk}^2$. Based on the joint distribution (2.7), L_{k1} and L_{k0} can be calculated as follows:

$$
\begin{aligned}
L_{k1} \;=\; & P(\boldsymbol{y} | \gamma_k = 1, \boldsymbol{\gamma}_{-k}, \boldsymbol{\theta}_{-\beta_k}) = \int_{\beta_k} P(\boldsymbol{y}, \beta_k | \gamma_k = 1, \boldsymbol{\gamma}_{-k}, \boldsymbol{\theta}_{-\beta_k}) d\beta_k \qquad (2.8) \\
\propto \;& (\tilde{\sigma_u}^2)^{-\frac{1}{2}} exp(-\frac{1}{2}\{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}(c_{ij} + x_{ijk}\beta_k)^2}{\sigma^2} - (\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i} x_{ijk}(c_{ij} + x_{ijk}\beta_k)}{\sigma^2})^2 (\tilde{\sigma_u}^2)^{-1}\}).
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
L_{k0} \;=\; & P(\boldsymbol{y} | \gamma_k = 0, \boldsymbol{\gamma}_{-k}, \boldsymbol{\theta}_{-\beta_k}) = \int_{\beta_k} P(\boldsymbol{y}, \beta_k | \gamma_k = 0, \boldsymbol{\gamma}_{-k}, \boldsymbol{\theta}_{-\beta_k}) d\beta_k \qquad (2.9) \\
\propto \;& (\sigma_u^2)^{-\frac{1}{2}} exp(-\frac{1}{2}\{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}(c_{ij} + x_{ijk}\beta_k)^2}{\sigma^2}\}).
\end{aligned}
$$

In order to update $\boldsymbol{\gamma}$, we use a Metropolis-Hastings scheme as described below [Yi et al., 2007]. Suppose the current γ_k is c(=0 or 1) and a new value d(=0 or 1) is proposed from the prior probability $P(\gamma_k = c)$. If c equals to d, the acceptance probability for Metropolis-Hastings scheme is set to 1, so that γ_k remains at c and no update needed.

Otherwise, we update γ_k from the current value c to $d = 1-c$ with acceptance probability

$$\alpha = min(1, (\frac{1-w_k}{w_k}R)^{1-2c}), \text{ where} \tag{2.10}$$

$$R = \frac{L_{k1}}{L_{k0}} = (\frac{\tilde{\sigma_u}^2}{\sigma_u^2})^{-\frac{1}{2}} exp(\frac{1}{2}(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i} x_{ijk}(c_{ij}+x_{ijk}\beta_k)}{\sigma^2})^2(\tilde{\sigma_u}^2)^{-1}).$$

Conditional Posterior of $\boldsymbol{\lambda}$

The full conditional posterior distribution for the kth QTL location is

$$P(\lambda_k|\boldsymbol{\gamma}, \boldsymbol{\lambda}_{-k}, \boldsymbol{y}) = \begin{cases} P(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\theta})P(\lambda_k) & \text{if } \gamma_k = 1 \\ P(\lambda_k) & \text{if } \gamma_k = 0 \end{cases} \tag{2.11}$$

Since this conditional distribution has a nonstandard form, the Metropolis-Hastings algorithm is needed to update λ_k. First, a new location λ_k^* is sampled from $q(\lambda_k^*; \lambda_k)$ which can be employed as the uniform distribution $U[\lambda_k - d, \lambda_k + d]$ where d is a predetermined tuning number (e.g., $d = 2$). And a proposal for the new location is accepted or rejected with the acceptance probability

$$\alpha = min(1, \frac{P(\lambda_k^*|\boldsymbol{\gamma}, \boldsymbol{\lambda}_{-k}, \boldsymbol{y})q(\lambda_k; \lambda_k^*)}{P(\lambda_k|\boldsymbol{\gamma}, \boldsymbol{\lambda}_{-k}, \boldsymbol{y})q(\lambda_k^*; \lambda_k)}). \tag{2.12}$$

Conditional Posterior of $\boldsymbol{b}$

Let $\boldsymbol{\mu} = \mu \mathbf{1}_N$ and $\boldsymbol{x} = (\boldsymbol{x}_i, ..., \boldsymbol{x}_n)^T$. The full conditional posterior distribution of the latent normal variable $\boldsymbol{b}$ is given by

$$\begin{aligned} P(\boldsymbol{b}|\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{-\boldsymbol{b}}) &\propto exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x}\boldsymbol{\beta}-\boldsymbol{v}\boldsymbol{b})^T(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x}\boldsymbol{\beta}-\boldsymbol{v}\boldsymbol{b})\}exp(-\frac{1}{2}\boldsymbol{b}^T\boldsymbol{b}) \\ &\propto exp\{-\frac{1}{2}(\boldsymbol{b}-\boldsymbol{b}^*)^T(\frac{1}{\sigma^2}\boldsymbol{v}^T\boldsymbol{v}+\boldsymbol{I}_{nk})(\boldsymbol{b}-\boldsymbol{b}^*)\}, \end{aligned} \tag{2.13}$$

where $\boldsymbol{b}^* = \frac{1}{\sigma^2}(\frac{1}{\sigma^2}\boldsymbol{v}^T\boldsymbol{v} + \boldsymbol{I}_{nk})^{-1}\boldsymbol{v}^T(\boldsymbol{y} - \boldsymbol{\mu} - \boldsymbol{x}\boldsymbol{\beta})$. That is, $\boldsymbol{b}|\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{-\boldsymbol{b}} \sim N_{nk}(\boldsymbol{b}^*, \boldsymbol{\Sigma}_b^*)$, where $\boldsymbol{\Sigma}_b^* = (\frac{1}{\sigma^2}\boldsymbol{v}^T\boldsymbol{v} + \boldsymbol{I}_{nk})^{-1}$ and $\boldsymbol{b}^* = \frac{1}{\sigma^2}\boldsymbol{\Sigma}_b^*\boldsymbol{v}^T(\boldsymbol{y} - \boldsymbol{\mu} - \boldsymbol{x}\boldsymbol{\beta})$.

Conditional Posterior of $\boldsymbol{\delta}$

In order to obtain the full conditional distribution of $\boldsymbol{\delta}$, we rewrite the model (2.4) as

$$y_{ij} = \mu + \boldsymbol{x}_{ij}\boldsymbol{\beta} + \sum_{l=1}^{k}\delta_l(p_{ijl}(b_{il} + \sum_{m=1}^{l-1} b_{im}\psi_{lm})) + e_{ij}, \tag{2.14}$$

and define the $k \times 1$ vector $\boldsymbol{w}_{ij} = (w_{ij1}, ..., w_{ijk})^T = (p_{ijl}(b_{il} + \sum_{m=1}^{l-1} b_{im}\psi_{lm}) : l = 1, ..., k)^T$ and $\xi_{ijl} = y_{ij} - \mu - \boldsymbol{x}_{ij}\boldsymbol{\beta} - \Sigma_{m\neq l} w_{ijm}\delta_m$. The full conditional distribution of $\boldsymbol{\delta}$ is given by

$$\begin{aligned} P(\boldsymbol{\delta}|\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{-\boldsymbol{\delta}}) &\propto exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{\mu} - \boldsymbol{x}\boldsymbol{\beta} - \boldsymbol{w}\boldsymbol{\delta})^T(\boldsymbol{y} - \boldsymbol{\mu} - \boldsymbol{x}\boldsymbol{\beta} - \boldsymbol{w}\boldsymbol{\delta})\} \\ &\quad \times \prod_{l=1}^{k}\{exp(-\frac{1}{2s_{l0}^2}(\delta_l - m_{l0})^2)I(\delta_l > 0)\} \end{aligned} \tag{2.15}$$

$$\begin{aligned} P(\delta_l|\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{-\delta_l}) &\propto exp\{-\frac{1}{2\sigma^2}(\boldsymbol{\xi}_l - \boldsymbol{w}_l\delta_l)^T(\boldsymbol{\xi}_l - \boldsymbol{w}_l\delta_l)\}\{exp(-\frac{1}{2s_{l0}^2}(\delta_l - m_{l0})^2)I(\delta_l > 0)\} \\ &\propto \{exp(-\frac{1}{2\sigma_l^{*2}}(\delta_l - \delta_l^*)^2)I(\delta_l > 0)\}, \end{aligned}$$

where $\boldsymbol{w} = (\boldsymbol{w}_{11}, ..., \boldsymbol{w}_{nn_n})^T$, $\boldsymbol{w}_l = (w_{11l}, ..., w_{nn_nl})^T$, $\boldsymbol{\xi}_l = (\xi_{11l}, ..., \xi_{nn_nl})^T$, $\sigma_l^{*2} = (\frac{1}{\sigma^2}\boldsymbol{w}_l^T\boldsymbol{w}_l + s_{l0}^{-2})^{-1}$ and $\delta_l^* = \sigma_l^{*2}(\frac{1}{\sigma^2}\boldsymbol{w}_l^T\boldsymbol{\xi}_l + s_{l0}^{-2}m_{l0})$. That is, $\delta_l|\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{-\delta_l} \sim N^+(\delta_l^*, \sigma_l^{*2})$, where $\sigma_l^{*2} = (\frac{1}{\sigma^2}\boldsymbol{w}_l^T\boldsymbol{w}_l + s_{l0}^{-2})^{-1}$ and $\delta_l^* = \sigma_l^{*2}(\frac{1}{\sigma^2}\boldsymbol{w}_l^T\boldsymbol{\xi}_l + s_{l0}^{-2}m_{l0})$.

Conditional Posterior of $\boldsymbol{\psi}$

In order to obtain the full conditional distribution of $\boldsymbol{\psi}$, we rewrite the model (2.4) as

$$y_{ij} = \mu + \boldsymbol{x}_{ij}\boldsymbol{\beta} + \sum_{l=1}^{k} b_{il}(\delta_l p_{ijl} + \sum_{m=l+1}^{k}\delta_m p_{ijm}\psi_{ml}) + e_{ij}, \tag{2.16}$$

and define the $k(k-1)/2 \times 1$ vector $\boldsymbol{u}_{ij} = (b_{il}\delta_m p_{ijm} : l = 1,..,k, m = l+1,...,k)^T$. The full conditional distribution of $\boldsymbol{\psi}$ is given by

$$\begin{aligned} P(\boldsymbol{\psi}|\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}_{-\boldsymbol{\psi}}) &\propto exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x\beta}-\boldsymbol{u\psi})^T(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x\beta}-\boldsymbol{u\psi})\} \\ &\quad \times exp(-\frac{1}{2}(\boldsymbol{\psi}-\boldsymbol{\psi}_0)^T\boldsymbol{R}_0^{-1}(\boldsymbol{\psi}-\boldsymbol{\psi}_0)) \\ &\propto exp\{-\frac{1}{2}(\boldsymbol{\psi}-\boldsymbol{\psi}^*)^T{\boldsymbol{\Sigma}_\psi^*}^{-1}(\boldsymbol{\psi}-\boldsymbol{\psi}^*)\}, \end{aligned} \tag{2.17}$$

where $\boldsymbol{u} = (\boldsymbol{u}_{11},...,\boldsymbol{u}_{nn_n})^T$, $\boldsymbol{\Sigma}_\psi^* = (\frac{1}{\sigma^2}\boldsymbol{u}^T\boldsymbol{u} + \boldsymbol{R}_0^{-1})^{-1}$ and $\boldsymbol{\psi}^* = \boldsymbol{\Sigma}_\psi^*(\frac{1}{\sigma^2}\boldsymbol{u}^T(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x\beta}) + \boldsymbol{R}_0^{-1}\boldsymbol{\psi}_0)$. That is, $\boldsymbol{\psi}|\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}_{-\boldsymbol{\psi}} \sim N(\boldsymbol{\psi}^*,\boldsymbol{\Sigma}_\psi^*)$ where $\boldsymbol{\Sigma}_\psi^* = (\frac{1}{\sigma^2}\boldsymbol{u}^T\boldsymbol{u} + \boldsymbol{R}_0^{-1})^{-1}$ and $\boldsymbol{\psi}^* = \boldsymbol{\Sigma}_\psi^*(\frac{1}{\sigma^2}\boldsymbol{u}^T(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x\beta}) + \boldsymbol{R}_0^{-1}\boldsymbol{\psi}_0)$.

Conditional Posterior of $\boldsymbol{\beta}$ and σ_u^2

Suppose that β_k belongs to group u (i.e. $P(\beta_k|\gamma_k,\sigma_u^2) \stackrel{d}{=} N(0,\gamma_k\sigma_u^2)$). If $\gamma_k = 0$, β_k=0. Otherwise, β_k is generated from its conditional posterior distribution:

$$\begin{aligned} & P(\beta_k|\gamma_k=1,\boldsymbol{\gamma}_{-k},\boldsymbol{\theta}_{-\beta_k},\boldsymbol{y}) \propto P(\boldsymbol{y}|\gamma_k=1,\boldsymbol{\gamma}_{-k},\boldsymbol{\theta})P(\beta_k|\gamma_k=1,\sigma_u^2) \\ \propto\ & exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{n_i}(y_{ij}-\mu-\boldsymbol{x}_{ij}\boldsymbol{\beta}-\boldsymbol{v}_{ij}\boldsymbol{b})^2)exp(-\frac{\beta_k^2}{2\sigma_u^2}) \\ \propto\ & exp(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{c_{ij}^2}{\sigma^2}-\frac{\beta_k^2}{\sigma_u^2}) \text{ where } c_{ij} = y_{ij}-\mu-\boldsymbol{x}_{ij}\boldsymbol{\beta}-\boldsymbol{v}_{ij}\boldsymbol{b} \\ \propto\ & exp(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{(c_{ij}+x_{ijk}\beta_k)^2-2x_{ijk}\beta_k(c_{ij}+x_{ijk}\beta_k)+x_{ijk}^2\beta_k^2}{\sigma^2}-\frac{\beta_k^2}{\sigma_u^2}) \\ \propto\ & exp(-\frac{(\beta_k-\tilde{\mu_k})^2}{2{\tilde{\sigma_u}}^2}). \end{aligned} \tag{2.18}$$

That is, $\beta_k|\gamma_k = 1,\boldsymbol{\gamma}_{-k},\boldsymbol{\theta}_{-\beta_k},\boldsymbol{y} \sim N(\tilde{\mu_k},{\tilde{\sigma_u}}^2)$ where $\tilde{\mu_k} = ({\tilde{\sigma_u}}^2)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i}x_{ijk}(c_{ij} + x_{ijk}\beta_k)/\sigma^2$ and ${\tilde{\sigma_u}}^2 = \sigma_u^{-2} + \sigma^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n_i}x_{ijk}^2$. For each type of genetic effect, the full

conditional posterior distribution of hyperparameter σ_u^2 is given by

$$\begin{aligned} P(\sigma_u^2|\beta_k) &\propto P(\beta_k|\sigma_u^2)P(\sigma_u^2) \\ &\propto (\sigma_u^2)^{-\frac{1}{2}} exp(-\frac{1}{2}\frac{\beta_k^2}{\sigma_u^2})(\sigma_u^2)^{-\frac{\nu_u}{2}-1} exp(-\frac{\nu_u s_u^2}{2\sigma_u^2}) \\ &\propto (\sigma_u^2)^{-\frac{\nu_u+1}{2}-1} exp(-\frac{\beta_k^2+\nu_u s_u^2}{2\sigma_u^2}). \end{aligned} \tag{2.19}$$

That is, $\sigma_u^2|\beta_k \sim \text{Inv-}\chi^2(\nu_u+1, (\beta_k^2+\nu_u s_u^2)/(\nu_u+1))$.

Conditional Posterior of μ and σ^2

The full conditional posterior distributions for μ is given by

$$\begin{aligned} P(\mu|\boldsymbol{\gamma}, \boldsymbol{\theta}_{-\mu}, \boldsymbol{y}) &\propto P(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\theta})P(\mu) \\ &\propto exp(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x\beta}-\boldsymbol{vb})^T(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x\beta}-\boldsymbol{vb})) \\ &\quad \times exp(-\frac{1}{2s_y^2}(\mu-\bar{y})^T(\mu-\bar{y})) \\ &\quad \text{where } \bar{y} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i} y_{ij} \text{ and } s_y^2 = \frac{1}{N-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 \\ &\propto exp(-\frac{1}{2}(\mu-\mu^*)^T(\frac{1}{\sigma^2}+\frac{1}{{s_y}^2})(\mu-\mu^*)) \\ &\quad \text{where } \mu^* = \frac{1}{\sigma^2}(\boldsymbol{y}-\boldsymbol{x\beta}-\boldsymbol{vb})^T(\boldsymbol{y}-\boldsymbol{x\beta}-\boldsymbol{vb}) + \frac{1}{s_y^2}\bar{y}. \end{aligned} \tag{2.20}$$

That is, $\mu|\boldsymbol{\gamma}, \boldsymbol{\theta}_{-\mu}, \boldsymbol{y} \sim N(\mu^*, \sigma_\mu^{2*})$ where $\sigma_\mu^{2*} = (\frac{1}{\sigma^2}+\frac{1}{{s_y}^2})^{-1}$ and $\mu^* = \frac{1}{\sigma^2}(\boldsymbol{y}-\boldsymbol{x\beta}-\boldsymbol{vb})^T(\boldsymbol{y}-\boldsymbol{x\beta}-\boldsymbol{vb}) + \frac{1}{s_y^2}\bar{y}$. The full conditional posterior distributions for σ^2 is given by

$$\begin{aligned} P(\sigma^2|\boldsymbol{\gamma}, \boldsymbol{\theta}_{-\sigma^2}, \boldsymbol{y}) &\propto P(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\theta})P(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{N}{2}} exp(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x\beta}-\boldsymbol{vb})^T(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x\beta}-\boldsymbol{vb}))\frac{1}{\sigma^2} \\ &\propto (\sigma^2)^{-\frac{N}{2}-1} exp(-\frac{N\hat{\sigma^2}}{2\sigma^2}) \\ &\quad \text{where } \hat{\sigma^2} = \frac{1}{N}(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x\beta}-\boldsymbol{vb})^T(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{x\beta}-\boldsymbol{vb}). \end{aligned} \tag{2.21}$$

That is, $\sigma^2|\boldsymbol{\gamma}, \boldsymbol{\theta}_{-\sigma^2}, \boldsymbol{y} \sim$ Inv-$\chi^2(N, \hat{\sigma^2})$ where $\hat{\sigma^2} = \frac{1}{N}(\boldsymbol{y} - \boldsymbol{\mu} - \boldsymbol{x\beta} - \boldsymbol{vb})^T(\boldsymbol{y} - \boldsymbol{\mu} - \boldsymbol{x\beta} - \boldsymbol{vb})$.

Posterior Analysis

The posterior samples can be used to estimate the posterior distribution of the parameters and the QTL positions. During the MCMC step, initial iterations are discarded as "burn-in" and the subsequent samples are thinned by keeping every kth simulation draw. The posterior inclusion probability of each possible locus κ_l can be calculated using its inclusion proportion in the MCMC samples as $p(\kappa_l|\boldsymbol{y}) = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{r} 1(\lambda_k^{(t)} = \kappa_l, \gamma_k^{(t)} = 1)$ where T is the total number of MCMC samples. With the prior $p(\kappa_l) = \frac{p_m}{h}$, the Bayes factor can be calculated to show evidence for inclusion of κ_l, against exclusion of κ_l as

$$BF(\kappa_l) = \frac{p(\kappa_l|\boldsymbol{y})/p(\kappa_l)}{(1 - p(\kappa_l|\boldsymbol{y}))/(1 - p(\kappa_l))} = \frac{p(\kappa_l|\boldsymbol{y})}{1 - p(\kappa_l|\boldsymbol{y})}\frac{1 - p(\kappa_l)}{p(\kappa_l)}. \tag{2.22}$$

The Bayes factor $BF(\kappa_l)$ captures the change in the odds in favor of the inclusion of κ_l as we move from prior to posterior. In the R/qtlbim manual [Yandell et al., 2007], the following criteria are suggested for judging the significance of each locus: weak support if $BF(\kappa_l)$ falls between 3 and 10; moderate support if $BF(\kappa_l)$ falls between 10 and 30; strong support if the $BF(\kappa_l)$ is larger than 30.

Choice of the Number of Grid Points

The critical issue with the proposed Bayesian mixed effects model is how to efficiently choose the number of grid points, k. To select the number of grid points, we evaluate the goodness of the predictive distributions of our Bayesian models. Spiegelhalter et al. [2002] proposed the deviance information criterion (DIC) as DIC $= -2E_{\boldsymbol{\gamma},\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\theta})\} + P_D$. The second term of DIC, P_D is the effective number of parameters, which is defined as $P_D = -2E_{\boldsymbol{\gamma},\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\theta})\} + 2logP(\boldsymbol{y}|\bar{\boldsymbol{\gamma}}, \bar{\boldsymbol{\theta}})$ where $\bar{\boldsymbol{\gamma}}$ and $\bar{\boldsymbol{\theta}}$ are the posterior means of $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$. Since $P(\boldsymbol{y}_i|\boldsymbol{\gamma}, \boldsymbol{\theta}) \stackrel{d}{=} N(\boldsymbol{\mu}_i + \boldsymbol{x}_i\boldsymbol{\Gamma\beta}, \boldsymbol{p_i D p_i}^T + \sigma^2\boldsymbol{I}_{n_i})$ in

model (2.1), the scores of DIC are easy to compute with MCMC samples. As stated by Robert and Titterington [2002], the observed data are used twice to calculate P_D, and thus the predictive distribution chosen by DIC overfits the observed data. To overcome the overfitting problem, Ando [2007] developed the following Bayesian predictive information criterion (BPIC) as BPIC = $-2E_{\boldsymbol{\gamma},\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\gamma},\boldsymbol{\theta})\} + 2n\hat{b}$ where $\hat{b}$ is the bias of the posterior mean of the expected loglikelihood. Under a certain mild regularity condition, the bias term can be approximately by $n\hat{b} \approx P_D$ [Ando, 2011], leading to the simplified BPIC = $2E_{\boldsymbol{\gamma},\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\gamma},\boldsymbol{\theta})\} + 2P_D$. Note that the penalty term of the simplified BPIC is twice of that of original DIC. To choose the optimal number of grid points for our Bayesian model, we first compute both DIC and simplified BPIC with several pre-selected numbers of grid points. We select the number of grid points with minimal DIC or simplified BPIC scores.

Implementation in R/qtlbimmixed

The proposed methods have been implemented in R/qtlbimmixed which is built on top of the widely used R packages, R/qtl [Broman et al., 2003] and R/qtlbim [Yandell et al., 2007]. The MCMC algorithm written in C and data manipulation procedure in R were modified for handling longitudinal data. For the choice of the optimal number of grid points, the R/qtlbimmixed provides both DIC and simplified BPIC scores for our Bayesian model.

2.5 Simulations Study and Real Data Analysis

2.5.1 Simulation I

In this section, we illustrate our Bayesian mixed effects model with simulations. We simulate a backcross population with 200 individuals and a single chromosome with

151 evenly spaced markers at 5cM intervals. The number of measurements for each individual ranges from three to seven and the total number of observations is 1000. Four different setups (Setups 1, 2, 3 and 4) are considered. We first simulate the data sets containing four QTL with only main effects (Setup 1). The four simulated QTL are located at markers 31, 61, 91 and 121, respectively. The simulated setup equals $\boldsymbol{y}_i = 0.5\cdot(\boldsymbol{x}_{31i}+\boldsymbol{x}_{61i}+\boldsymbol{x}_{91i}+\boldsymbol{x}_{121i}+\boldsymbol{t}_i)+\boldsymbol{p}_i\boldsymbol{\nu}_i+\boldsymbol{e}_i$, where the $\boldsymbol{x}_{ki}$ ($k = 31, 61, 91, 121$) are the genotype codes of the four simulated QTL and $\boldsymbol{t}_i = (t_{i1}, ..., t_{in_i})^T$ are the ith individual's time covariates generated from the uniform distribution $U[0, 1]$ and then standardized with mean 0 and variance 1. We set σ^2=1. The true number of grid points are set to 3 (i.e., true k = 3), and thus $\boldsymbol{p}_i$ can be calculated from $\boldsymbol{t}_i$ by linear interpolation. We set $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3) = (1, 1.2, 0.8)$ and $\boldsymbol{\psi} = (\psi_{21}, \psi_{31}, \psi_{32}) = (0.6, 0.4, 0.6)$. That is, $\boldsymbol{\nu}_i \sim N(0, \boldsymbol{D})$ with $\text{diag}(\boldsymbol{D}) = (1, 1.96, 0.97)$ and the lower triangle elements are $(d_{21}, d_{31}, d_{32}) = (0.72, 0.32, 0.81)$.

The prior distributions for the elements of $\boldsymbol{\delta}$ are chosen to be independent $N(0, 30)$ and the prior distributions for the elements of $\boldsymbol{\psi}$ are independent $N(0, 0.5)$. We set a relatively large variance for the prior of $\boldsymbol{\delta}$ and a somewhat diffused variance for the prior of $\boldsymbol{\psi}$. For all analyses, the MCMC algorithm ran for 4×10^5 iterations after discarding the first 1000 burn-in iterations. In order to reduce serial correlation in the stored samples, the chain was thinned for every 40 iteration, yielding 10^4 MCMC samples for posterior analysis.

To further investigate our Bayesian mixed effects model, we simulate additional data sets containing QTL that have one gene-gene interaction (Setup 2), or one gene-time interaction (Setup 3), or two gene-gene interactions (Setup 4). Specifically, we have $\boldsymbol{y}_i = 0.5 \cdot (\boldsymbol{x}_{31i} + \boldsymbol{x}_{61i} + \boldsymbol{x}_{91i} \cdot \boldsymbol{x}_{121i} + \boldsymbol{t}_i) + \boldsymbol{p}_i\boldsymbol{\nu}_i + \boldsymbol{e}_i$ for Setup 2, $\boldsymbol{y}_i = 0.5 \cdot (\boldsymbol{x}_{31i} + \boldsymbol{x}_{61i} + \boldsymbol{x}_{91i} + \boldsymbol{x}_{121i} \cdot \boldsymbol{t}_i) + \boldsymbol{p}_i\boldsymbol{\nu}_i + \boldsymbol{e}_i$ for Setup 3 and $\boldsymbol{y}_i = 0.5\cdot(\boldsymbol{x}_{31i}\cdot\boldsymbol{x}_{61i}+\boldsymbol{x}_{91i}\cdot\boldsymbol{x}_{121i}+\boldsymbol{t}_i)+\boldsymbol{p}_i\boldsymbol{\nu}_i+\boldsymbol{e}_i$ for Setup 4. For comparison, we run both R/qtlbimmixed and R/qtlbim with simulated data for

the four setups. Estimated marginal Bayes factor for each marker using R/qtlbimmixed with all time points and R/qtlbim with one randomly-selected time point are presented in Figures 2.1 and 2.2. The solid, dot-dashed and long-dashed lines represent main, epistatic effects and gene-time interaction, respectively. In general, R/qtlbimmixed has better power to detect QTL than R/qtlbim. Especially, for Setup 3, R/qtlbimmixed more clearly detects the gene-time interaction than R/qtlbim in Figure 2.2.

To evaluate the performance of our Bayesian mixed effects model, we further calculate the receiver operating characteristic (ROC) curves. For each setup, we conduct 100 simulations with uniformly generated QTL positions which are restricted to be at least 10cM apart. For a given cut-off of the Bayes factor, we calculate true and false positive findings as follows: a significant finding is claimed to be a true positive if it is located less than 10Mb from any one of the simulated causal SNPs; otherwise the finding is false. The ROC curves with the false positive rate less than 0.2 are presented in Figure 2.3. The solid line represents the result of R/qtlbimmixed where the measurements from all time points are analyzed. The dot-dashed line is from R/qtlbim where only one randomly selected measurement from each subject. We also apply R/qtlbim to all time measurements by (wrongly) assuming that all the measurements are independent. The results are summarized by the long-dashed line in Figure 2.3. ROC curves demonstrate that our R/qtlbimmixed with all measurements appears to perform better than R/qtlbim with both one and all measurements. That is, Bayesian mixed effects model with all longitudinal measurements which deals with the dependence of the data improves the mapping power/efficiency compared to ordinary Bayesian model with one or all measurements.

To diagnose convergence of MCMC samples, we run ten parallel chains with over-dispersed different initial values with respect to the true posterior distribution. Using 10^4 iterations, Geweke's Z-scores [Geweke et al., 1991] for each chain based on the

first 10% and last 50% of the samples indicate convergence of all parameters. Using ten chains, Gelman and Rubin's potential scale reduction factors [Gelman and Rubin, 1992] are calculated and upper limits are less than 1.01 for all parameters. Figure 2.4 presents the trace plots of σ^2, δ_1, δ_2, δ_3, ψ_{21}, ψ_{31} and ψ_{32} for each setup. The black lines represent the values of the draws for all parameters at each iteration and gray lines represent the true values of the parameters. Figure 2.4 shows that all chains move around the true values for every parameter, indicating a good convergence. We plot the marginal posterior and prior densities of the parameters for the random errors and the random effects in Figure 2.5. Estimated densities are based on 10000 random draws. It appears that the random draws approximately follow the normal density and the means of those are close to the true values. Figure 2.6 displays the 95% high posterior density (HPD) intervals of σ^2, δ_1, δ_2, δ_3, ψ_{21}, ψ_{31} and ψ_{32} for each setup. The blue dots represent the posterior means and blue lines represent the HPD intervals. It shows that most of the 95% HPD intervals contain the true values. Table 2.1 summaries the posterior estimates corresponding to the parameters of the random errors and random effects. The posterior means and medians are close to the true values and all the 95% HPD intervals contain the true ones, representing good performance of our algorithm.

2.5.2 Simulation II

We conduct another simulations and use DIC [Spiegelhalter et al., 2002] and simplified BPIC [Ando, 2007, 2011] to estimate the number of grid points. The settings are almost the same as in Simulation I except that the true number of grid points now varies from 2 to 4 (i.e., true $k = 2, 3, 4$). We simulate 100 data sets containing four QTL with only main effects. The four simulated QTL are randomly located with at least 10cM apart. The trait equals $\boldsymbol{y}_i = 0.5 \cdot (\boldsymbol{x}_{k_1 i} + \boldsymbol{x}_{k_2 i} + \boldsymbol{x}_{k_3 i} + \boldsymbol{x}_{k_4 i} + \boldsymbol{t}_i) + \boldsymbol{p}_i \boldsymbol{\nu}_i + \boldsymbol{e}_i$, where the $\boldsymbol{x}_{k_j i}$ (j = 1,2,3,4) are the genotype codes of the four simulated QTL and $\boldsymbol{t}_i = (t_{i1}, ..., t_{in_i})^T$

are the time points of the ith individual. We set $(\delta_1, \delta_2, \delta_3, \delta_4) = (1, 1.2, 0.8, 0.7)$ and $(\psi_{21}, \psi_{31}, \psi_{32}, \psi_{41}, \psi_{42}, \psi_{43}) = (0.6, 0.4, 0.6, 0.2, 0.4, 0.6)$. Table 2.2 shows the average DIC, average simplified BPIC scores and the percentages of correctly choosing the number of true grids. All average DIC and average simplified BPIC scores achieved the minimum scores at the true grid number and the percentages of selection of the right number of true grids are 97%, 91% and 94% for 2, 3 and 4 grid setups using DIC, 99%, 97% and 81% using simplified BPIC, repectively as shown in table 2.2.

2.5.3 Real Data Analysis

To further test our Bayesian mixed effects model, we analyze a real mouse data on age-related body weight in backcross mice of NZO/HILtJ and NON/ShiLtJ [Reifsnyder et al., 2000]. To identify diabetes-predisposing QTL, diabetes-prone (obese) NZO/HlLtJ mice were outcrossed with non-diabetic (non-obese) NONShiLtJ mice to generate F1 hybrids. F1 mice of both sexes were backcrossed (reciprocal) with non-diabetic (non-obese) NON/ShiLtJ, resulting in 204 male mice reaching 24 weeks of age. Body weights were measured in four-week intervals from 4-24 weeks of age, and each mouse has six body weight measurements at weeks 4, 8, 12, 16, 20 and 24, respectively. Total 84 microsatellite markers were genotyped and distributed ~ 20cM apart with higher concentration of markers in areas around suggestive QTL. Genomewide scans for body weight were performed in Reifsnyder et al. [2000], using one-way ANOVA. Significance of the ANOVA F-statistic was assessed by permutation test. Based on the separate analyses on phenotypes from weeks 8, 16 and 24, a large segment of chromosome 1 (D1Mit211-M1Mit76) were detected. We consider five body weights at 8, 12, 16, 20 and 24 weeks as longitudinal measurements while the first measurement and time are treated as fixed covariates. We set 84 marker positions as putative QTL locations.

For all analyses, we run 4×10^5 iterations after discarding the first 1000 burn-ins. The MCMC chain are thinned by one in forty, yielding 10^4 MCMC samples for the posterior analysis. The number of grid points is set to 3 since DIC and simplified BPIC achieve their minimal scores at $k = 3$. The genomewide profile of Bayes factors is presented in Figure 2.7, which shows a strong evidence of one QTL activity on chromosome 1 and suggestive QTL on chromosomes 10 and 15. Interestingly, our Bayesian analysis found a QTL on chromosome 11 with a weak gene-time interaction.

2.6 Analysis of GAW18 Longitudinal Blood Pressure Data

2.6.1 GAW18 Data and Analysis Plan

Genomewide association studies (GWAS) have been performed to examine genetic variants associated with blood pressure and hypertension [Levy et al., 2009; Padmanabhan et al., 2010]. As blood pressure changes over time, it is of great interest in collecting repeated blood measurements to study genes with time varying genetic effects. Genetic Analysis Workshop 18 (GAW18) data is a real human whole genome sequencing (WGS) study with systolic and diastolic blood pressure phenotypes plus age, sex, medication use and cigarette smoking [Almasy et al.]. The data are longitudinal, with three measurements for most participants at roughly 5-year intervals. In this section, we apply the proposed Bayesian mixed effects model to the GAW18 longitudinal blood pressure data. Due to the limited sample size, it is not feasible to include all available SNPs in our Bayesian analysis. To select a subset of top ranked SNPs, we extend the EMMA approach [Kang et al., 2008], an efficient algorithm which corrects for population structure and genetic relatedness in model organism association mapping to the data. We replace the kinship similarity matrix in EMMA with an estimated covariance matrix for dealing with the correlation among the multiple measurements from each sample.

Extended EMMA Method: For testing association between a given SNP and the phenotype, we fit the following mixed effects model

$$\boldsymbol{y}_i = \boldsymbol{\mu}_i + \boldsymbol{x}_i^{\boldsymbol{e}}\boldsymbol{\beta}^{\boldsymbol{e}} + \boldsymbol{x}_i^g\beta^g + \boldsymbol{u}_i + \boldsymbol{e}_i \ (i = 1, ..., n), \tag{2.23}$$

where $\boldsymbol{y}_i = (y_{i1}, ..., y_{in_i})^T$ is the $n_i \times 1$ phenotype vector of individual i; $\boldsymbol{\mu}_i = \mu\mathbf{1}_{n_i}$ with μ being the grand mean and $\mathbf{1}_{n_i}$ being the $n_i \times 1$ vector whose elements are all equal to 1; $\boldsymbol{x}_i^{\boldsymbol{e}}$ is the design matrix corresponding to non-genetic covariates (such as time) and $\boldsymbol{\beta}^{\boldsymbol{e}}$ is the associated non-genetic effects; $\boldsymbol{x}_i^g$ is the numerically coded genotype of individual i and β^g is the corresponding SNP effect. In the model, we assume random effect $\boldsymbol{u}_i \sim N(\mathbf{0}, \sigma_g^2\boldsymbol{K}_i)$ where $\boldsymbol{K}_i$ is an $n_i \times n_i$ matrix, and random error $\boldsymbol{e}_i \sim N(\mathbf{0}, \sigma_e^2\boldsymbol{I}_{n_i})$. The SNP effect can be tested as $H_0 : \beta^g = 0$ vs $H_1 : \beta^g \neq 0$ via the likelihood ratio test. For GWAS or WGS data, this test needs to be performed with a large number of SNPs which can be computationally intensive if we treat $\boldsymbol{K}_i$s as the unknowns and estimate them jointly with the fixed effects. EMMA [Kang et al., 2008] is an efficient algorithm originally developed for GWAS data where samples are potentially structured. EMMA models the structure effect via a similarity matrix. An R package that implements EMMA can either estimate the similarity matrix using genotype data or take any similarity matrix provided by users. We tweak EMMA for our purpose. We provide EMMA with the following similarity matrix $\boldsymbol{K} = diag(\hat{\boldsymbol{K}}_1, \hat{\boldsymbol{K}}_2, \cdots, \hat{\boldsymbol{K}}_n)$ where $\hat{\boldsymbol{K}}_i$s are the estimated correlation matrices from Model (2.23) in which β^g is set to 0. The idea of estimating $\boldsymbol{K}_i$s this way is not new and has been used in EMMAX [Kang et al., 2010], a fast version of EMMA. These estimates should be reasonable unless some SNPs have large effects, which is rare for most complex traits.

2.6.2 GAW18 Data Analysis

The original GAW18 data includes 849 individuals with both phenotype and imputed genotype data from a total 20 large pedigrees. Among them, there are 139 genetically unrelated individuals who have been measured for age, sex, current use of antihypertensive medication, current tobacco smoking status and blood pressure. Our analysis is based on the 139 unrelated individuals. The number of SBP (and DBP) measurements ranges from one to four. The whole genome sequence data provided by GAW18 has about 8.3M SNPs from the odd numbered autosomes, among which 5.5M have minor allele frequency (MAF) < 0.05. All SNPs provided have passed the initial quality control checking. However, for those SNPs with minor allele frequency > 0.05 (total of 2.8M), we have found that 17,463 of the SNPs deviate from the Hardy-Weinberg Equilibrium (HWE) with p-value $<0.05/2.8$M after the Bonferroni correction for multiple tests. We remove the SNPs with MAF < 0.05 and those not passing the HWE test, which results in about 2.8M SNPs for the subsequent analyses. For each SNP, we recoded the genotypes into the minor allele counts using PLINK [Purcell et al., 2007].

To check population outliers and potential population stratification, we generated a subset of SNPs that are in approximate linkage equilibrium with each other and performed the multidimensional scaling (MDS) analysis in PLINK [Purcell et al., 2007]. The pairwise scatter plots of the top four MDS scores from the genome-wide estimates of IBD sharing before and after removing three singletons from three pairs who have high IBD values are presented in Figure 2.8. In general, they show that the 139 individuals are quite homogeneous in terms of their ethnicities. However, several pairs have very high estimated IBD values. For example, the estimated IBD value between sample T2DG0400207 and sample T2DG0400247 is as high as 0.3. Though several individuals are likely to be related, we retained the 139 samples for all our analysis.

We applied the extended EMMA to the filtered GAW18 data on two log-transformed phenotypes: log(SBP) and log(DBP). We included five covariates (age, age^2, sex, medication and smoking status) in our analysis. We fit the data with different covariance matrices in SAS 9.2 and selected the spatial power covariance structure for the downstream analysis based on the AIC criteria. Specifically, we assume $cov(y_{ij}, y_{ij'}) = \sigma^2 \rho^{d_{i,jj'}}$ where $d_{i,jj'}$ is the time distance between the jth and j'th examinations for individual i. After obtaining the parameter estimates, we substituted the kinship matrix $\boldsymbol{K}$ in EMMA by $\boldsymbol{K} = diag(\hat{\boldsymbol{K}}_1, \hat{\boldsymbol{K}}_2, ..., \hat{\boldsymbol{K}}_n)$ where the jj' element of $\hat{K}_i$ equals $\hat{\rho}^{d_{i,jj'}}$ in which $\hat{\rho}$ is the parameter estimate of ρ. Figure 2.9 displays the Manhattan plots of the two phenotypes from the extended EMMA model. For SBP, one SNP on Chromosomes 5 reaches genomewide significance (p-value$< 5 \times 10^{-7}$ as in Burton et al. [2007]). For DBP, three SNPs on Chromosome 3, 17 and 21 exceeded a threshold of genomewide significance. The top ranked SNPs for SBP and DBP (with P-value $< 5 \times 10^{-7}$) are presented in Table 2.3.

For each phenotype, we selected a list of 3000 top ranked SNPs that are not highly correlated with each other (correlation < 0.95) for the proposed Bayesian mixed effects model. Again our Bayesian analysis included age, age^2, sex, medication, smoking status as covariates. Both SBP and DBP measurements were log-transformed. For all analyses, the MCMC algorithm ran for 4×10^5 iterations after discarding the first 1000 burn-in iterations. In order to reduce the serial correlation of the MCMC samples, the chain was thinned for every 40 iterations, yielding 10^4 MCMC samples for the posterior analysis. The posterior inclusion probability of each possible locus was calculated as its frequency in the MCMC samples. Each locus can be included in the model through its main effect and/or its interactions with other loci or age. We estimated the Bayes Factor (BF) and used it to judge the importance of any given locus to each phenotype (see Yi et al. [2005, 2007] for more details).

Figure 2.10 shows the one-dimensional genomewide profiles of $2log(BF)$ for the combined effects (main, epistasis and gene-age interactions). In the R/qtlbim manual [Yandell et al., 2007], the following criteria are suggested for judging the significance of each variable: weak support if BF falls between 3 and 10 ($2.2 < 2log(BF) < 4.6$); moderate support if BF falls between 10 and 30 ($4.6 < 2log(BF) < 6.8$); strong support if the BF is larger than 30 ($2log(BF) > 6.8$). Based on the criteria, we found several additional SNPs with strong signals on chromosomes 1, 3, 15 and 19 for SBP. No new SNPs was found for DBP. In particular, we found one SNP, chr3:197469358 has a very strong interaction with age. Top ranked SNPs for SBP (with $2log(BF) > 6.8$) are presented in Table 2.4. When comparing the results from the first and second methods in the genomic regions, we found that both results have high peaks in Figures 2.9 and 2.10 on the following genomic regions: chr1:17876090-17963920; chr5:3400160-3580412; chr9:15270545-15278594; chr17:79099414-79113680 for SBP and chr5:73777556-73777586; chr9:11305052-11326457 for DBP. It shows that there is reasonable concordance between both results (see Chung and Zou [2013] for more details).

2.7 Discussion

We have extended the Bayesian multiple QTL mapping model with a composite model space framework [Yi et al., 2007] for mapping longitudinal traits and identifying QTL with varying genetic effects. We have proposed new grid-based method to model covariance structure of the data. The proposed method for covariance estimation is parsimonious but with a reasonable number of grid points, it can approximate any type of covariance structure. The number of grid points is pre-set by users. In order to avoid the identifiability problem, we first need to compute the rank of $\boldsymbol{A}$ of Theorem 1 and then check if it equals to $\frac{1}{2}k(k+1)+1$ where k is the pre-set number of grid point. If so, our Bayesian model is identifiable. To choose the optimal number of grid

points, we utilize DIC and simplified version of BPIC through selecting the number of grid points which gives the minimum scores.

In order to obtain MCMC samples of all parameters, we use both Metropolis-Hastings and Gibbs sampling algorithms, alternately updating each unknown parameters conditional on all other parameters and the observed data. For conditional updates of the indicator variable vector $\boldsymbol{\gamma}$ and QTL position vector $\boldsymbol{\lambda}$, Metropolis-Hastings algorithm is applied. However, for obtaining MCMC samples of other parameters including the elements of decomposed matrix of covariance, Gibbs sampling algorithm is used via their conditional posterior distributions. The simulation study shows that proposed Bayesian method with all time points outperformed ordinary Bayesian method with one or all time points. To further test our method, we analyzed real mouse data on age-related body weight in backcross mice with all time points together and found strong evidence of QTL activity on chromosome 1, suggestive evidence on chromosome 10 and 15 and weak gene-time interaction on chromosome 11.

We further applied the proposed method to the GAW18 longitudinal blood pressure data. We first utilized the extended EMMA method. We replace the kinship similarity matrix in EMMA with an estimated correlation matrix for dealing with the dependent structure of the repeated measurements. And the proposed Bayesian method which models multiple SNPs simultaneously and allows for gene-gene interactions and gene-time interactions is applied. The GAW18 data contains extended pedigrees. Ideally, we should utilize all available data in our analysis. What complicates the analysis on longitudinal pedigree data is that both the correlation structure of the repeated measurements and the familial correlation structure of related individuals should be considered. We are currently extending the proposed methods for the GAW18 pedigree data. Moreover, our Bayesian model for GWAS data relies on a set of pre-selected putative SNPs. How to select a good set of putative SNPs, especially those with low

marginal effects but high interactions with other SNPs or environmental factors is challenging and deserves further investigations.

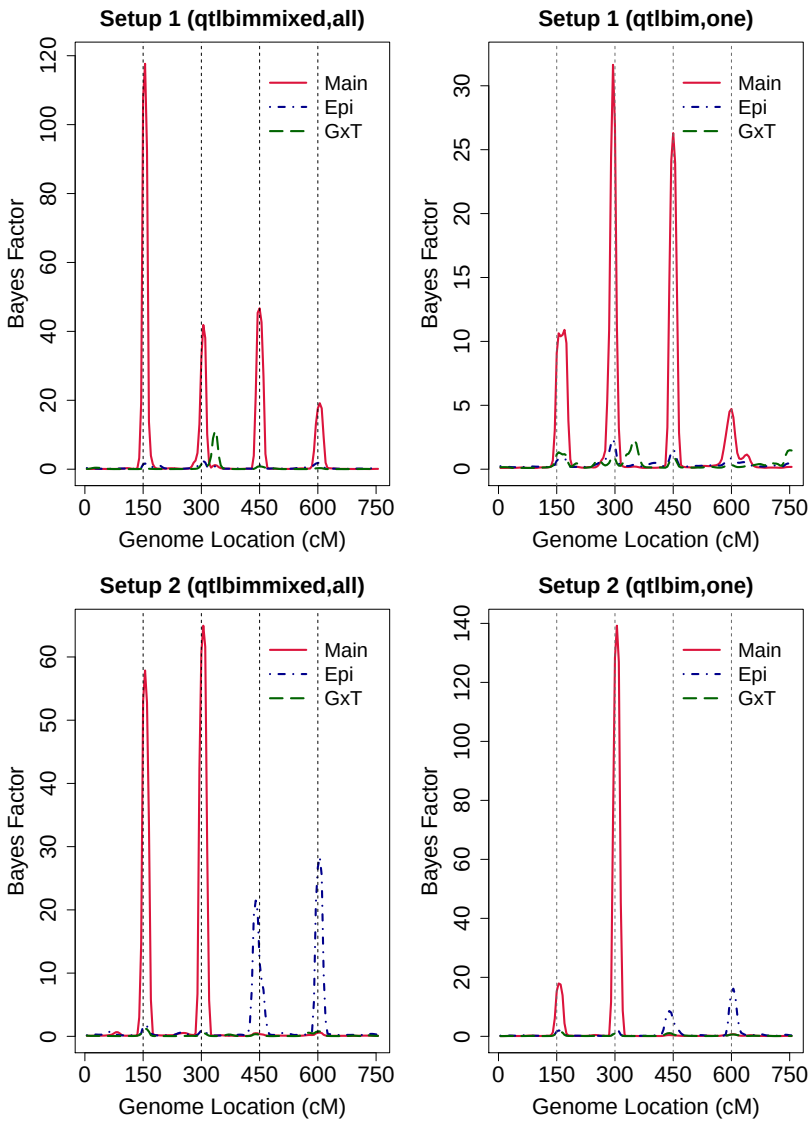

Figure 2.1: Estimated marginal Bayes factors for each marker from R/qtlbimmixed with all time points and R/qtlbim with one randomly selected time point for Setups 1 and 2. The solid (red), dot-dashed (blue) and long-dashed (green) lines represent main, epistatic effects and gene-time interaction, respectively.

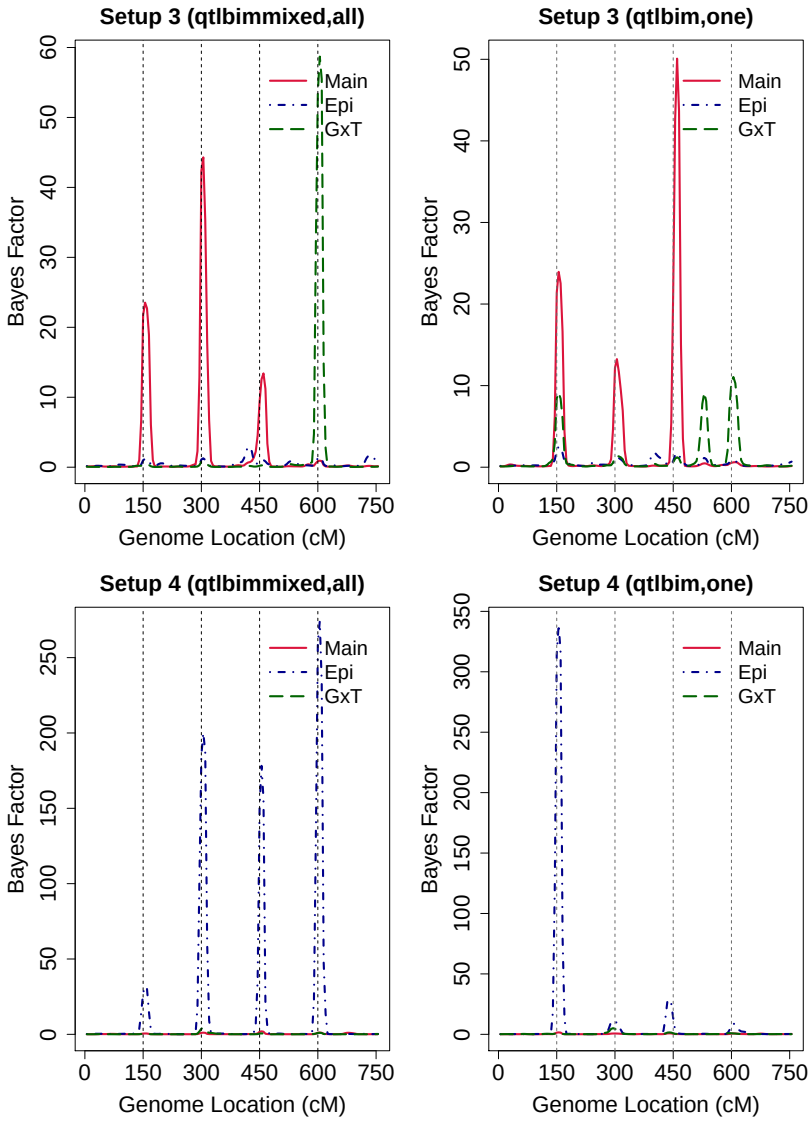

Figure 2.2: Estimated marginal Bayes factors for each marker from R/qtlbimmixed with all time points and R/qtlbim with one randomly selected time point for Setups 3 and 4. The solid (red), dot-dashed (blue) and long-dashed (green) lines represent main, epistatic effects and gene-time interaction, respectively.

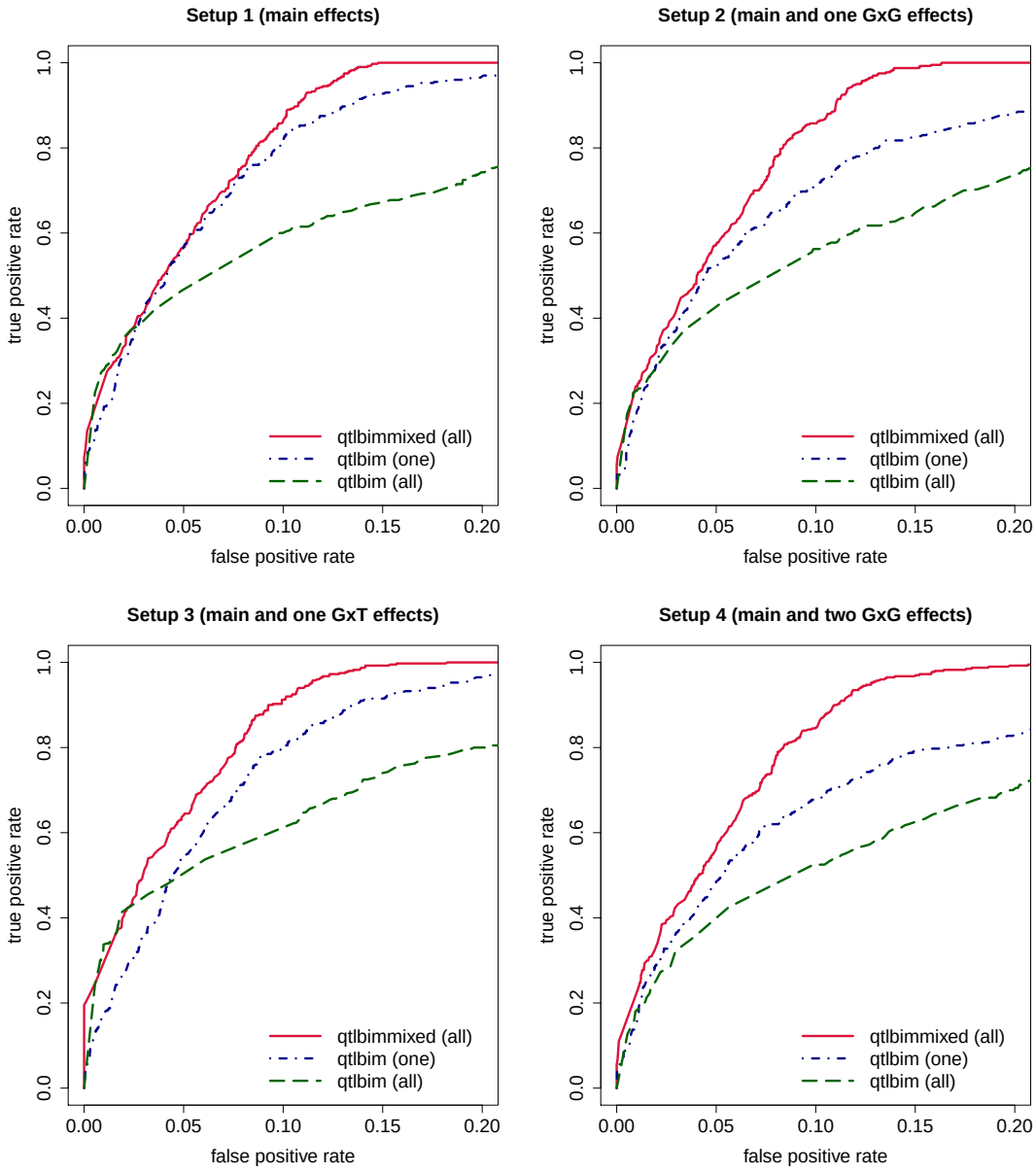

Figure 2.3: Estimated ROC curves for Setups 1, 2, 3 and 4: solid line (red) - proposed R/qtlbimmixed on all data; dot-dashed line (blue) - R/qtlbim on one randomly selected time point data; long-dashed lines (green) - R/qtlbim on all data.

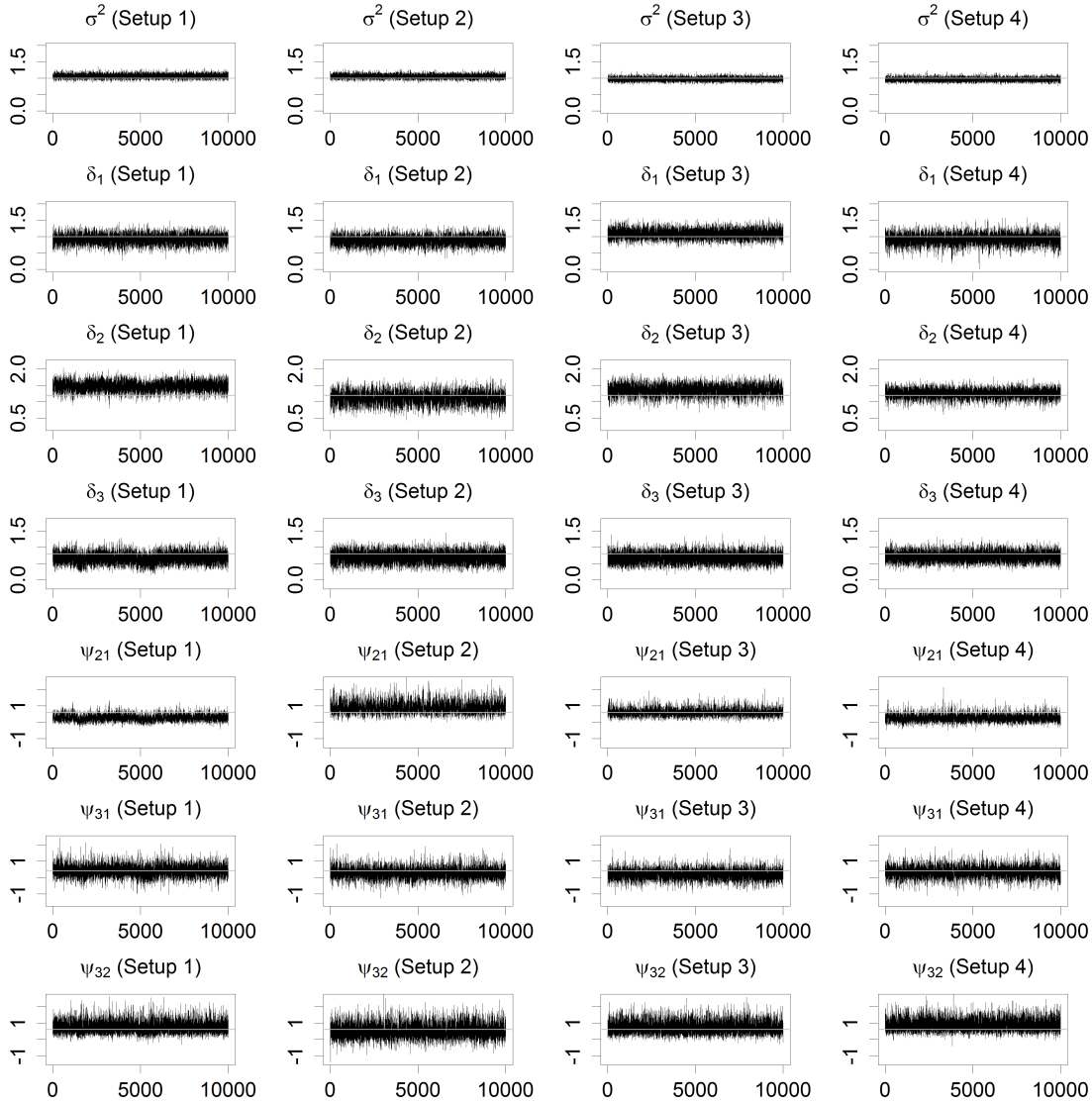

Figure 2.4: Trace plots of σ^2, δ_1, δ_2, δ_3, ψ_{21}, ψ_{31} and ψ_{32} for Setups 1,2,3 and 4 in the simulation study. The black lines represent the values of the draws for all parameters at each iteration and gray lines represent the true values of the parameters.

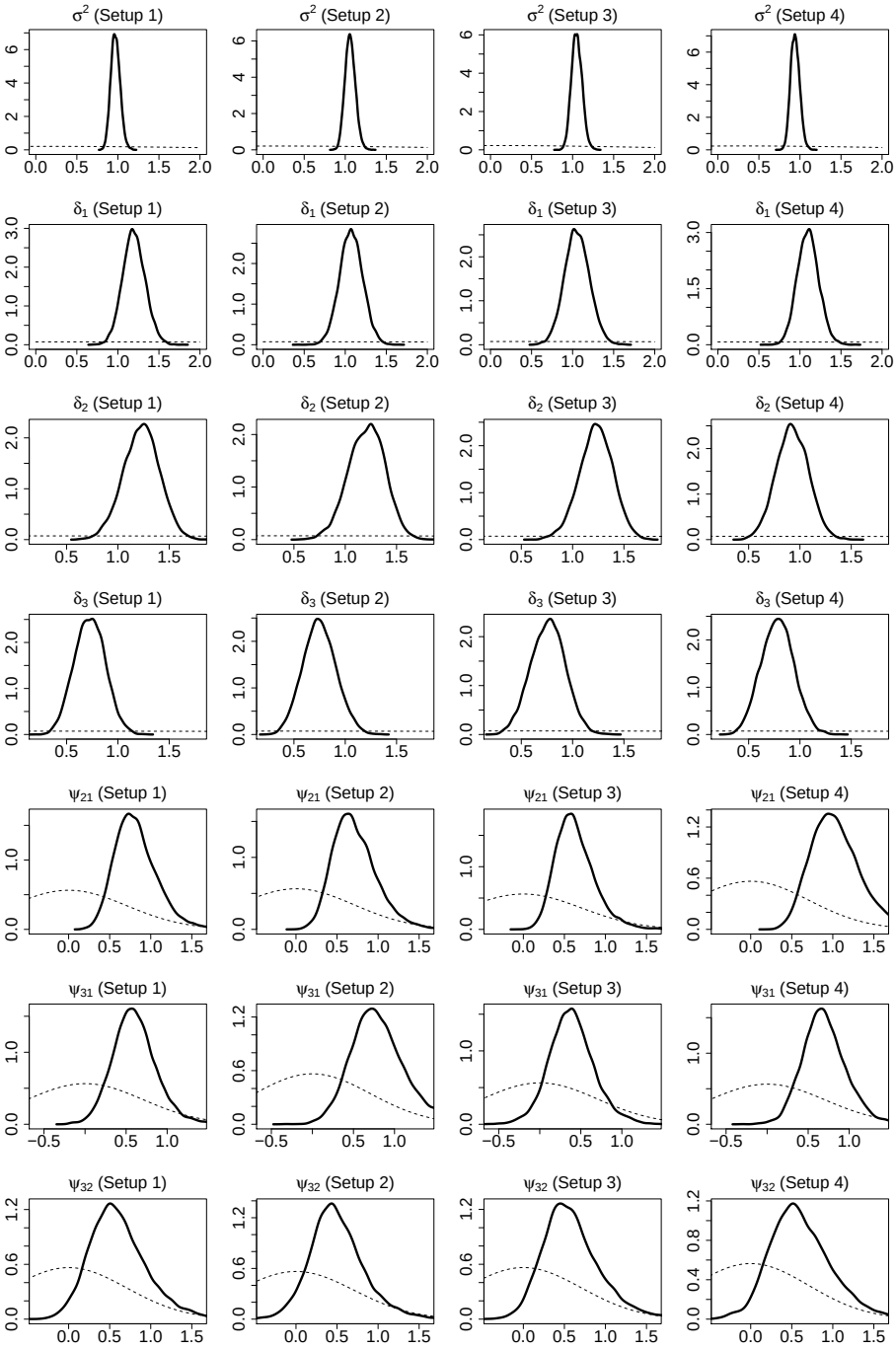

Figure 2.5: Posterior (solid line) and prior (dashed line) densities of the parameters for random errors and random effects for Setups 1,2,3 and 4. Estimated densities are based on 10000 random draws.

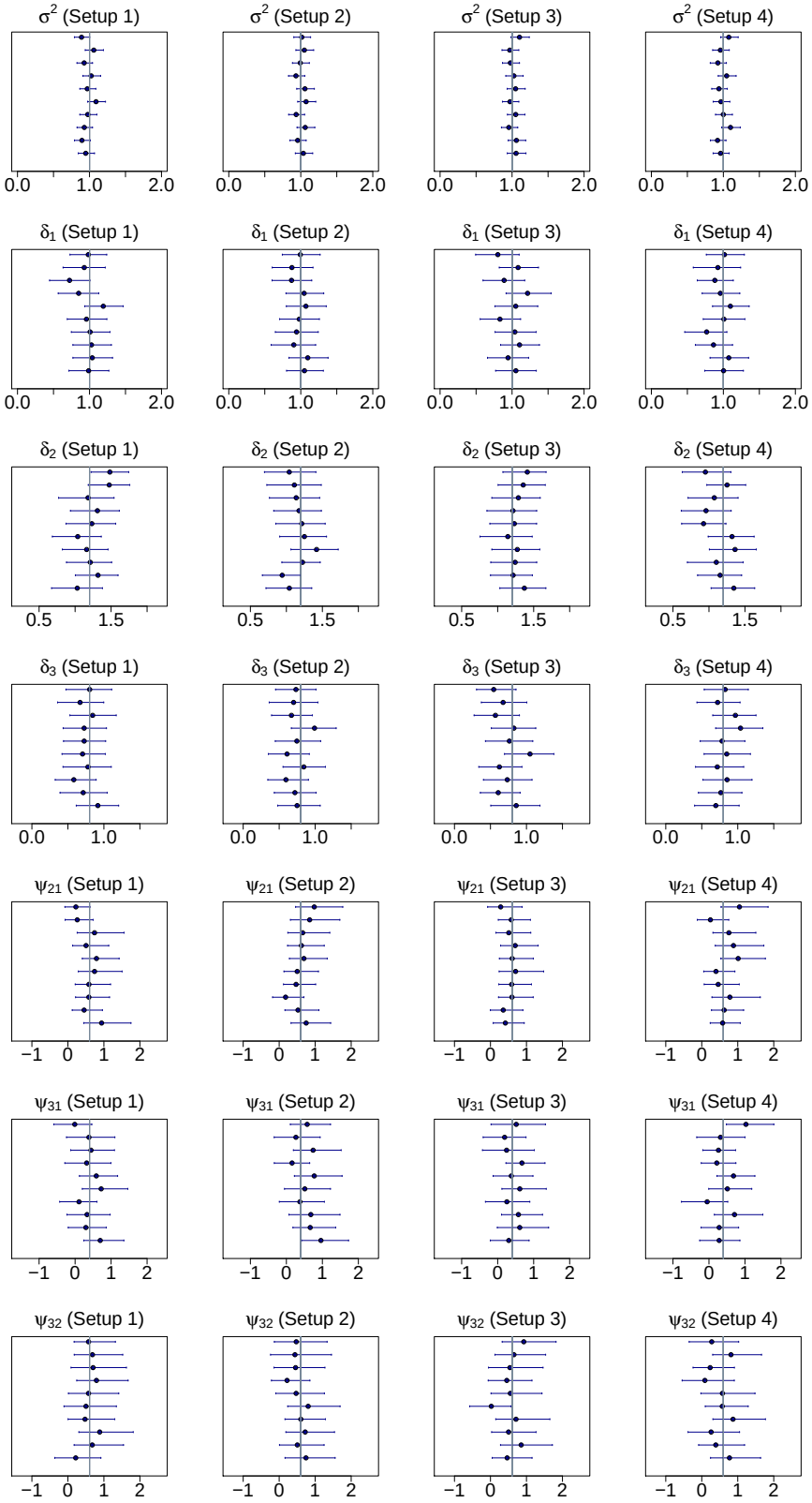

Figure 2.6: 95% HPD intervals of σ^2, δ_1, δ_2, δ_3, ψ_{21}, ψ_{31} and ψ_{32} for Setups 1,2,3 and 4. The blue dots represent the posterior means and blue lines represent HPD intervals.

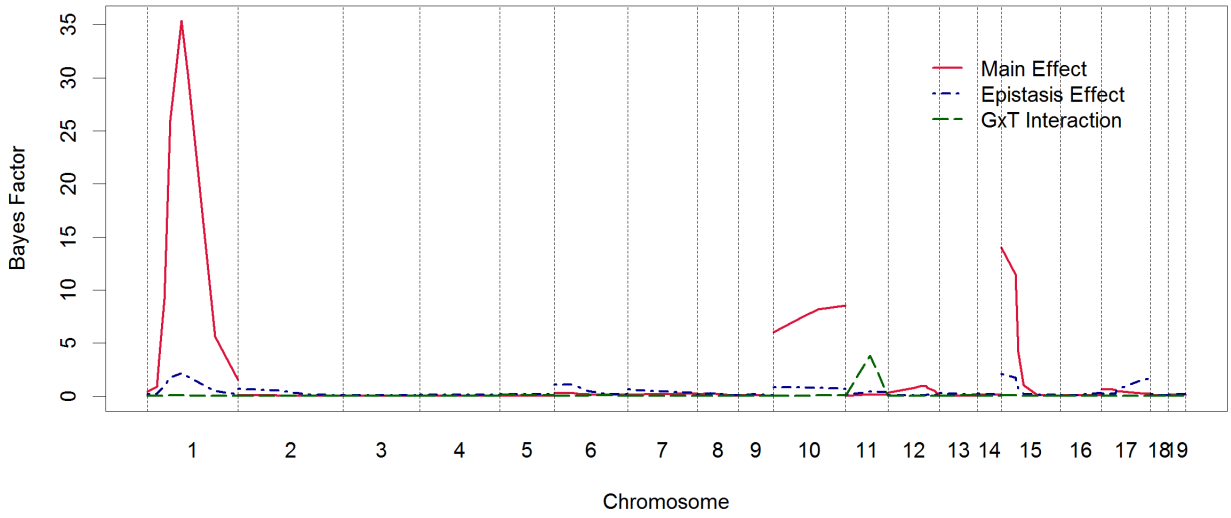

Figure 2.7: Genomewide profile of Bayes factors for body weight in backcross mice involving NZO/HILtJ and NON/ShiLtJ. The solid line (red) represents main effects, the dashed line (blue) represents epistasis effects and long-dashed line (green) represents gene-time interactions.

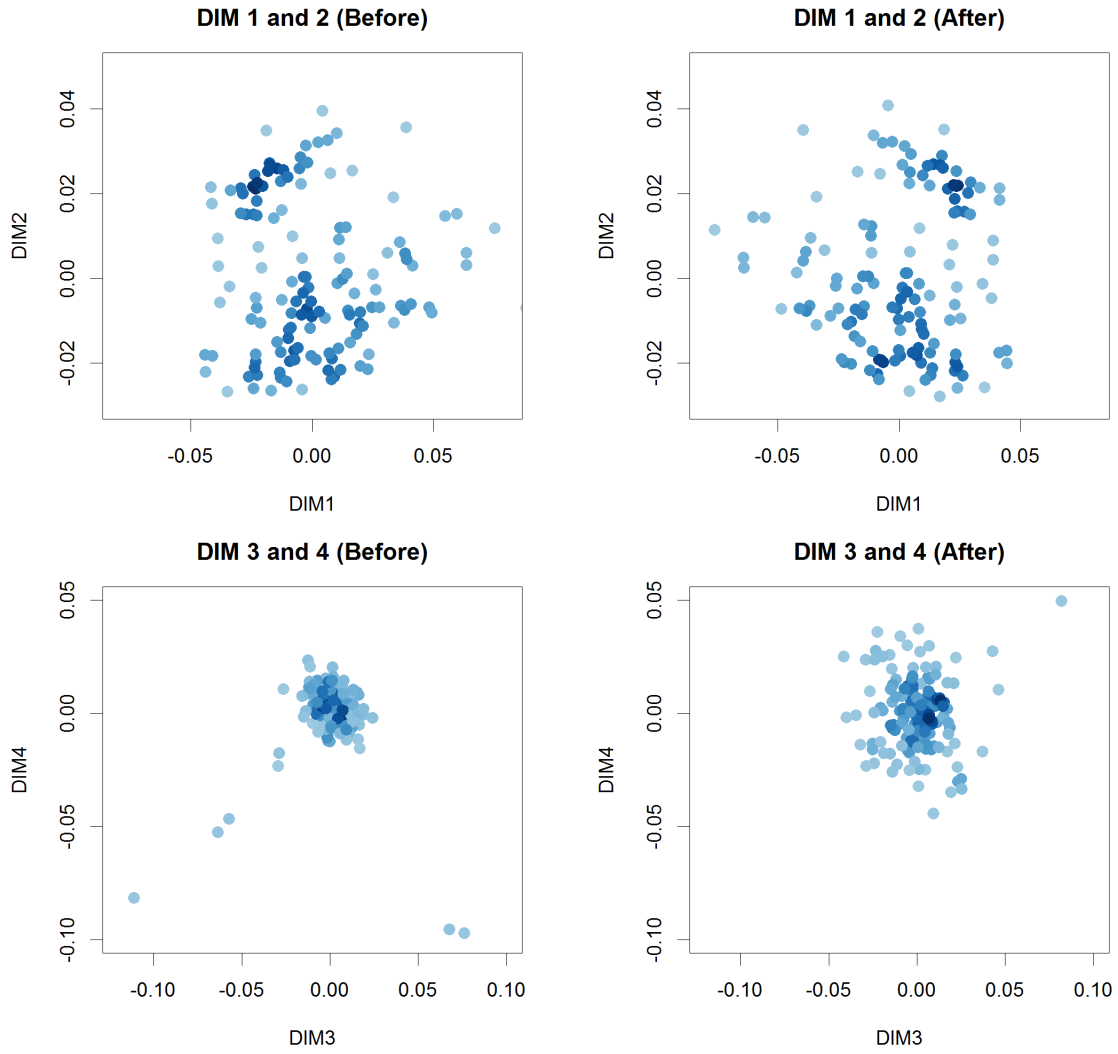

Figure 2.8: MDS plots for top four MDS scores from the genome-wide estimate of IBD sharing before and after removing three singletons from three pairs who have high IBD values.

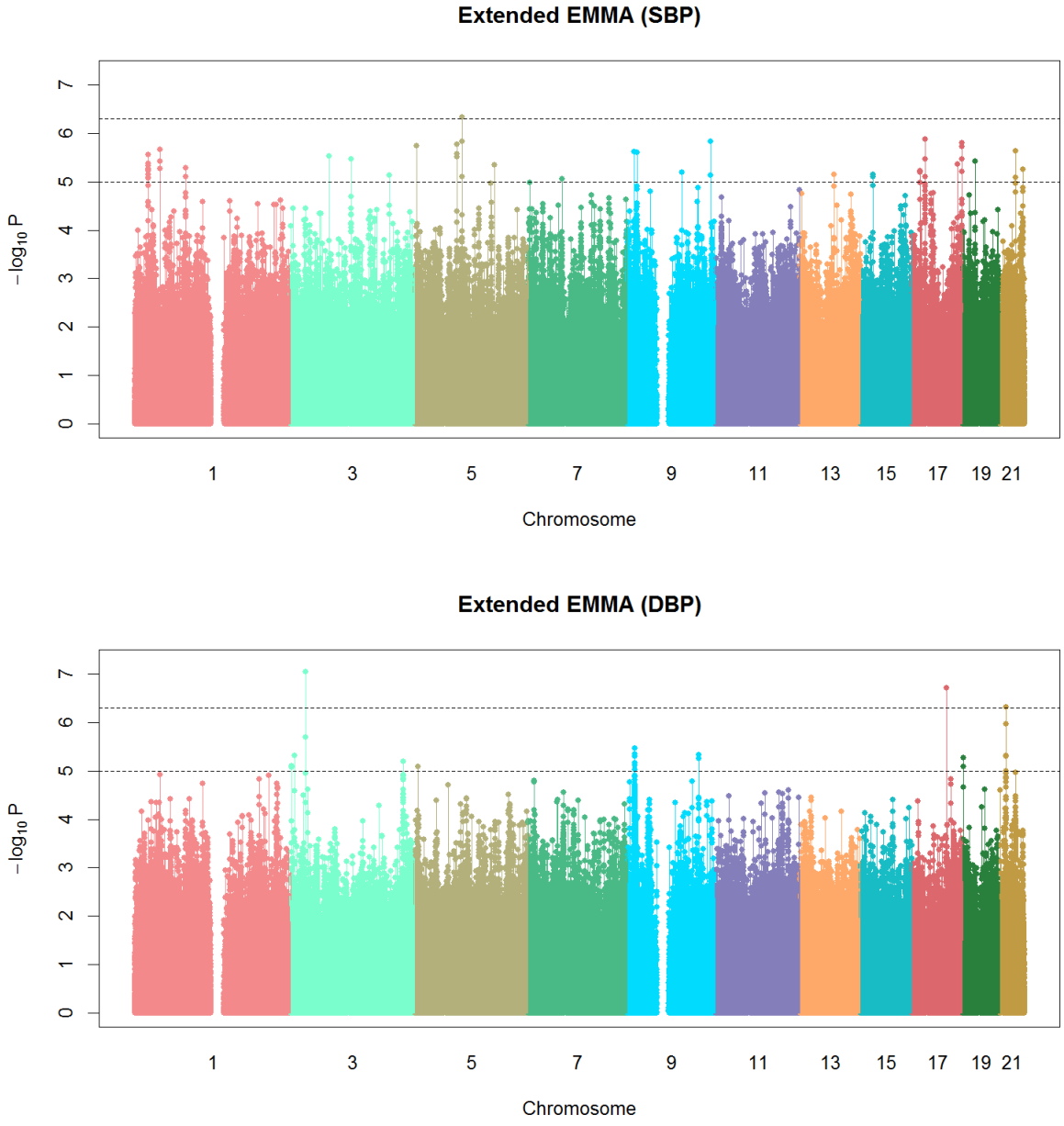

Figure 2.9: Genomewide Manhattan plots of $-log_{10}$(P-value) for association with SBP and DBP measurements from extended EMMA, on the basis of covariance matrix estimated from SAS. Two dashed horizontal lines represent the thresholds for suggestive (P-value=10^{-5}) and significant (P-value=5×10^{-7}) genomewide association.

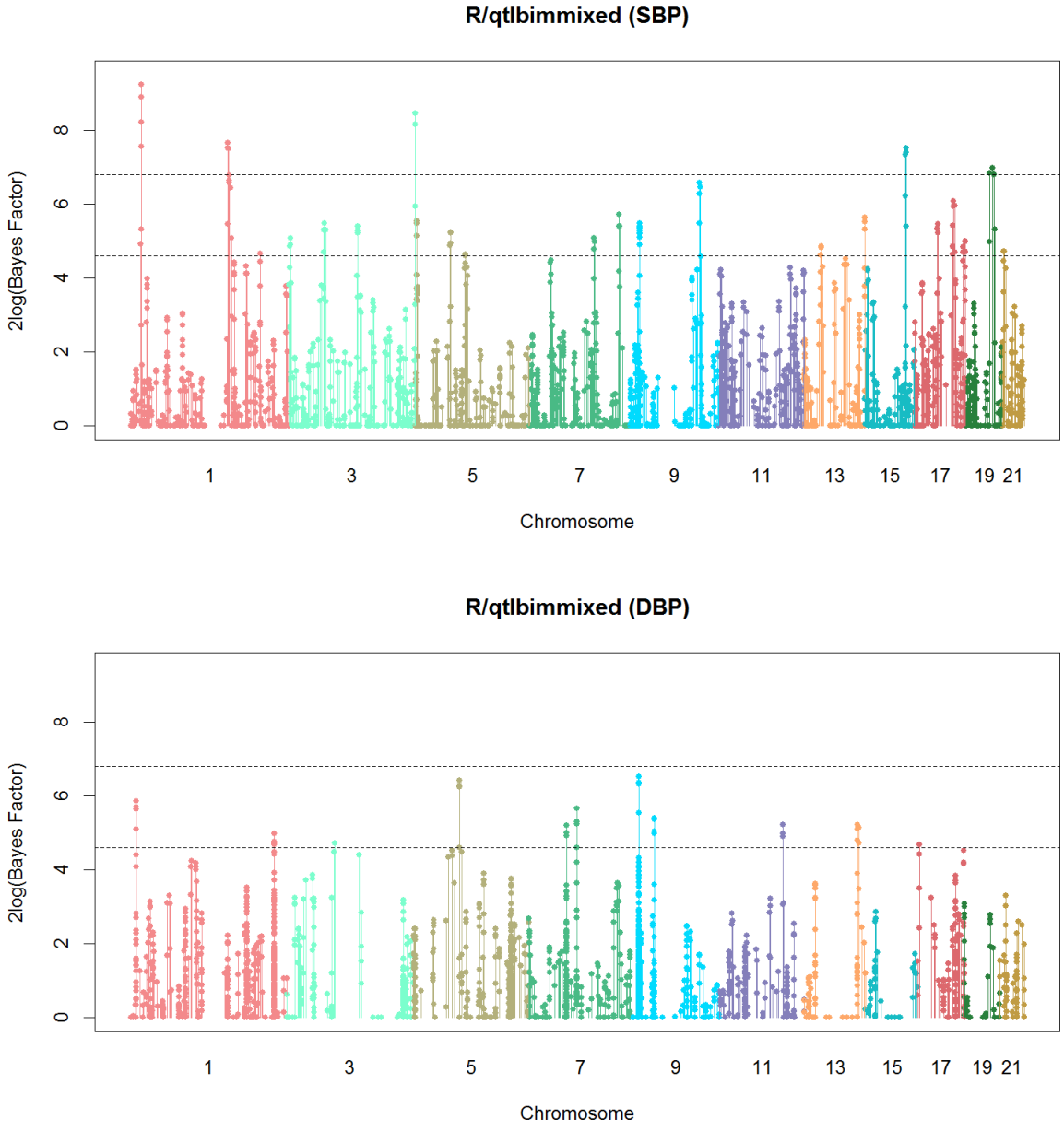

Figure 2.10: Genomewide Manhattan plots of $2log(BF)$ for all combined effects with SBP and DBP measurements from R/qtlbimmixed. Two dashed horizontal lines represent the genomewide thresholds for moderate (BF=10) strong (BF=30) genomewide associations.

Table 2.1: Posterior means, medians, standard deviations and 95% HPD intervals of the parameters for random errors and random effects in the simulation study.

Setup	Par	True	Mean	Med	SD	95% HPD
1	σ^2	1	0.97	0.96	0.06	(0.86,1.1)
	δ_1	1	1.14	1.13	0.15	(0.88,1.44)
	δ_2	1.2	1.24	1.25	0.17	(0.9,1.59)
	δ_3	0.8	0.73	0.73	0.15	(0.44,1.02)
	ψ_{21}	0.6	0.79	0.77	0.24	(0.38,1.31)
	ψ_{31}	0.4	0.55	0.53	0.28	(0.07,1.18)
	ψ_{32}	0.6	0.64	0.59	0.36	(0.03,1.49)
2	σ^2	1	1.06	1.05	0.06	(0.93,1.19)
	δ_1	1	0.88	0.88	0.14	(0.63,1.18)
	δ_2	1.2	1.11	1.12	0.2	(0.72,1.5)
	δ_3	0.8	0.7	0.7	0.18	(0.36,1.04)
	ψ_{21}	0.6	0.9	0.84	0.36	(0.3,1.78)
	ψ_{31}	0.4	0.3	0.28	0.32	(-0.26,0.99)
	ψ_{32}	0.6	0.47	0.44	0.41	(-0.24,1.38)
3	σ^2	1	1.05	1.05	0.07	(0.93,1.19)
	δ_1	1	1.06	1.07	0.15	(0.76,1.34)
	δ_2	1.2	1.23	1.23	0.16	(0.9,1.54)
	δ_3	0.8	0.76	0.76	0.17	(0.45,1.09)
	ψ_{21}	0.6	0.63	0.61	0.23	(0.26,1.17)
	ψ_{31}	0.4	0.4	0.39	0.27	(-0.1,0.96)
	ψ_{32}	0.6	0.6	0.54	0.37	(0,1.4)
4	σ^2	1	0.97	0.97	0.06	(0.86,1.09)
	δ_1	1	1.01	1.01	0.15	(0.71,1.3)
	δ_2	1.2	1.32	1.32	0.16	(0.96,1.61)
	δ_3	0.8	0.86	0.85	0.17	(0.53,1.18)
	ψ_{21}	0.6	0.45	0.42	0.23	(0.1,0.98)
	ψ_{31}	0.4	0.53	0.5	0.3	(-0.01,1.26)
	ψ_{32}	0.6	0.61	0.58	0.29	(0.13,1.25)

Table 2.2: Average DIC, average simplifed BPIC scores and percentage of selection of the right number of true grids for Bayesian mixed effects model with different number of true grid points.

True k	k	Avg DIC	#Sel (%)	Avg Sim BPIC	#Sel (%)	Avg P_D
2	2	3291.68	97	3316.61	99	24.93
	3	3296.2	3	3324.34	1	28.14
	4	3300.92	0	3332.71	0	31.78
3	2	3354.26	0	3378.92	1	24.67
	3	3329.95	91	3358.03	97	28.07
	4	3336.84	9	3368.24	2	31.4
4	2	3360.12	0	3384.67	2	24.54
	3	3347.82	6	3375.16	17	27.34
	4	3336.96	94	3367.84	81	30.88

Table 2.3: Genomewide association results for SBP,DBP-associated SNPs with P-value $< 5 * 10^{-7}$ sorted by P-value via extended EMMA.

BP	Chr	Position	Minor	Major	Counts	MAF	P (HWE)	P (EMMA)
SBP	5	75506197	G	C	31/68/40	0.47	0.87	$4.67 * 10^{-7}$
DBP	3	23715851	C	T	0/20/119	0.07	1.00	$9.00 * 10^{-8}$
	17	54834217	T	C	11/66/62	0.32	0.33	$1.98 * 10^{-7}$
	21	18744081	C	A	33/62/44	0.46	0.23	$4.95 * 10^{-7}$

Table 2.4: Genomewide association results for SBP,DBP-associated SNPs with $2log(BF) > 6.8$ sorted by $2log(BF)$ of all combined effects via R/qtlbimmixed.

BP	Chr	Position	Minor	Major	Counts	MAF	P (HWE)	P (EMMA)	$2log(BF)$
SBP	1	17876090	A	G	1/39/98	0.15	0.31	$2.08 * 10^{-4}$	9.25
	1	17249395	A	G	4/71/62	0.29	0.00	$1.03 * 10^{-3}$	9.25
	1	17939373	T	C	3/68/67	0.27	0.00	$1.53 * 10^{-3}$	8.91
	3	197469358	T	C	2/32/105	0.13	1.00	$1.64 * 10^{-3}$	8.47
	1	17048228	G	A	0/24/112	0.09	0.60	$1.66 * 10^{-3}$	8.24
	3	197252834	A	G	6/52/81	0.23	0.64	$9.22 * 10^{-4}$	8.17
	1	153023956	C	T	1/14/124	0.06	0.37	$1.59 * 10^{-3}$	7.66
	1	17963920	T	C	1/28/110	0.11	1.00	$1.68 * 10^{-3}$	7.57
	1	152300819	T	C	1/14/124	0.06	0.37	$1.54 * 10^{-3}$	7.53
	15	87675666	C	A	0/17/122	0.06	1.00	$1.23 * 10^{-4}$	7.52
	1	153186966	T	A	0/20/119	0.07	1.00	$8.84 * 10^{-4}$	7.51
	15	87968635	T	G	0/19/120	0.07	1.00	$9.99 * 10^{-4}$	7.41
	15	87444856	A	G	1/19/119	0.08	0.56	$2.68 * 10^{-4}$	7.34
	19	41642807	G	C	2/27/110	0.11	0.68	$7.50 * 10^{-4}$	6.99
	19	37607570	G	A	23/86/28	0.48	0.00	$7.74 * 10^{-4}$	6.86
	19	43979439	C	T	21/77/40	0.43	0.12	$1.54 * 10^{-3}$	6.81

CHAPTER 3

GAUSSIAN PROCESS BASED NONPARAMETRIC BAYESIAN QTL MAPPING FOR LONGITUDINAL TRAITS

3.1 Introduction

Recently, there has been a growing interest on mapping time-dependent genetic factors through measuring traits repeatedly over time. In the previous chapter, we extended the Bayesian multiple QTL mapping method with a composite model space framework [Yi, 2004; Yi et al., 2005, 2007] for longitudinal data where a new grid-based covariance estimation method has been proposed to flexibly model the covariance structure. The proposed method effectively identify main, two-way gene-gene interaction of the data and gene by time or gene by environment effects. However, the method only allows for pairwise interactions among genes and time/environmental covariates and may miss genes with higher-order interactions. It is not feasible to include all potential effects and their high-order interactions in this parametric model since this can lead to a dramatic increase in the number of parameters and the search space. To overcome this difficulty, we alternatively consider a nonparametric method to search for QTL without restricting to pairwise interactions among genes and nongenetic factors.

For univariate phenotypes, a nonparametric Bayesian variable selection method with Gaussian process prior has recently been developed [Zou et al., 2010], where both

genetic and nongenetic effects are modeled nonparametrically. These methods were implemented via hybrid Monte Carlo method and Gaussian process prior [Neal, 1996, 1997; Rasmussen and Williams, 2006] on the unknown functions for genetic and nongenetic factors. Rather than modeling each main and interaction term explicitly, this Bayesian method measures the importance of each QTL, regardless whether it functions through main, epistatic, or interactions among genes and environment effect non-explicitly. The importance of each genetic factor and each nongenetic factor included in the function is estimated by a single hyperparameter, which enters the covariance function and captures the main and interaction effects associated with each factor. The task is fulfilled by a Bayesian variable selection method through a set of the latent indicator variables and gamma mixture priors on both genetic and nongenetic effects.

In this chapter, we extend the nonparametric Bayesian variable selection method with Gaussian process (GP) prior of Zou et al. [2010] to longitudinal traits. For modeling the covariance structure, we again use the grid-based method presented in the previous chapter. The usefulness of the proposed approach will be evaluated by simulation studies and a real mouse data analysis.

3.2 Nonparametric GP Model for Longitudinal Data

3.2.1 GP-based Nonparametric Bayesian Model

Suppose there are n subjects under study, with subject i having traits measured at n_i time points ($i = 1, ..., n$). For the ith individual, let $\boldsymbol{x}_{gi} = \{x_{gik}\}_{k=1}^{p}$ denote a $p \times 1$ vector of genotypes where x_{gik} is the kth marker genotype ($k = 1, .., p$), and $\boldsymbol{x}_{ti} = \{x_{tijm}\}_{j=1\ m=1}^{n_i\ q}$ denote a $q \times n_i$ matrix of nongenetic covariates including time where x_{tijm} is the mth

nongenetic covariate ($m = 1, ..., q$) measured at time j. We consider the following GP-based nonparametric Bayesian model:

$$\boldsymbol{y}_i = \boldsymbol{\eta}(\boldsymbol{x}_{gi}, \boldsymbol{x}_{ti}) + \boldsymbol{p}_i\boldsymbol{\nu}_i + \boldsymbol{e}_i \quad i = 1, ..., n, \tag{3.1}$$

where $\boldsymbol{y}_i = (y_{i1}, ..., y_{in_i})^T$ is an $n_i \times 1$ vector of phenotypes or traits where y_{ij} is the phenotype value of subject i at time j; $\boldsymbol{\eta}$ is a unknown function which will be used to flexibly model genetic effects and nongenetic effects; $\boldsymbol{p}_i$ is an $n_i \times k$ incidence matrix which maps each observed time point to its two nearest adjacent grid time points using linear interpolation; $\boldsymbol{\nu}_i$ is a $k \times 1$ vector of random variables for a fixed number of grid time points with $\boldsymbol{\nu}_i \sim N_k(0, \boldsymbol{D})$ where $\boldsymbol{D}$ is a $k \times k$ matrix; $\boldsymbol{e}_i$ is an $n_i \times 1$ vector of random errors with $\boldsymbol{e}_i \sim N_{n_i}(0, \sigma_e^2 \boldsymbol{I}_{n_i})$.

For Bayesian estimation of the nonparametric model (3.1), we conduct the factorization of the covariance matrix, $\boldsymbol{D}$, via the modified Cholesky decomposition of Chen and Dunson [2003]. Let $\boldsymbol{L}$ denote the $k \times k$ lower triangular Cholesky decomposition matrix which have nonnegative diagonal elements, such that $\boldsymbol{D} = \boldsymbol{L}\boldsymbol{L}^T$. Let $\boldsymbol{L} = \boldsymbol{\Delta}\boldsymbol{\Psi}$ where $\boldsymbol{\Delta} = diag(\delta_1, ..., \delta_k)$ and $\boldsymbol{\Psi}$ is a $k \times k$ matrix with the (l, m) element denoted by ψ_{lm}. To make $\boldsymbol{\Delta}$ and $\boldsymbol{\Psi}$ identifiable, we assume the following conditions:

$$\delta_l \geq 0, \;\; \psi_{ll} = 1 \;\; and \;\; \psi_{lm} = 0, \;\; for \;\; l = 1, ..., k; \;\; m = l + 1, ..., k. \tag{3.2}$$

These conditions make $\boldsymbol{\Delta}$ a nonnegative $k \times k$ diagonal matrix and $\boldsymbol{\Psi}$ a lower triangular matrix with all its diagonal elements being 1. This results in the following decomposition of $\boldsymbol{D}$,

$$\boldsymbol{D} = \boldsymbol{\Delta}\boldsymbol{\Psi}\boldsymbol{\Psi}^T\boldsymbol{\Delta}. \tag{3.3}$$

Based on the modified Cholesky decomposition of $\boldsymbol{D}$, we reparameterize our model

(3.1),

$$\boldsymbol{y}_i = \boldsymbol{\eta}(\boldsymbol{x}_{gi}, \boldsymbol{x}_{ti}) + \boldsymbol{p}_i \boldsymbol{\Delta}\boldsymbol{\Psi}\boldsymbol{b}_i + \boldsymbol{e}_i \quad i = 1, ..., n, \tag{3.4}$$

where $\boldsymbol{b}_i = (b_{i1}, ...b_{ik})^T$ such that $b_{ij} \sim N(0,1)$ and $b_{ij} \perp b_{ij'}$ $(j \neq j')$ $j = 1, ..., k$. For the later use, we define $\boldsymbol{v}_i = \boldsymbol{p}_i \boldsymbol{\Delta}\boldsymbol{\Psi} = (\boldsymbol{v}_{i1}^T, ..., \boldsymbol{v}_{in_i}^T)^T$.

3.2.2 Prior Specifications

The above GP-based nonparametric Bayesian model contains a unknown function $\boldsymbol{\eta}$ and a set of random effects. For random effects, we adopt the priors in Chen and Dunson [2003]. Specifically, we impose independent half normal priors on the diagonal elements of $\boldsymbol{\Delta}$ and normal priors on the lower triangular elements of $\boldsymbol{\Psi}$. For the unknown function $\boldsymbol{\eta}$, we extend the Gaussian process prior to contain both genetic and time-varying nongenetic covariates. Below we give details.

Prior on $\boldsymbol{\eta}$

Let $\boldsymbol{\eta}_i = \boldsymbol{\eta}(\boldsymbol{x}_{gi}, \boldsymbol{x}_{ti})$ denote an $n_i \times 1$ vector of unknown functions of $\boldsymbol{x}_{gi}$, $\boldsymbol{x}_{ti}$, and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_n)^T$ denote an $N \times 1$ vector where $N = \Sigma_{i=1}^n n_i$. In order to estimate $\boldsymbol{\eta}$, we assume $\boldsymbol{\eta}$ has a Gaussian process prior which is a stochastic process such that each finite dimensional distribution is a multivariate normal. Any Gaussian process can be specified by its mean function and covariance kernel. For model (3.1), we set the prior of $\boldsymbol{\eta}$ as a Gaussian process prior with mean 0 and covariance matrix Σ_N as follows.

$$\boldsymbol{\eta} \sim N_N(0, \boldsymbol{\Sigma}_N), \ \textit{where} \ \boldsymbol{\Sigma}_N = [\Sigma_{iji'j'}]_{N \times N} \tag{3.5}$$

$$\Sigma_{iji'j'} = cov(\eta_{ij}, \eta_{i'j'}) = \xi^2 exp\{-\sum_{k=1}^{p} \rho_{gk}^2 (x_{ik} - x_{i'k})^2 - \sum_{m=1}^{q} \rho_{tm}^2 (x_{tijm} - x_{ti'j'm})^2\}.$$

where ξ^2, ρ_{gk}^2 and ρ_{tm}^2 are hyperparameters. Hyperparameter ξ^2 defines the vertical

scale of variation and affects the magnitude of the exponential part. Hyperparameters ρ_{gk}^2 and ρ_{tm}^2 determine the relevance of the various input covariates such as genetic and nongenetic covariates. Large values of ρ_{gk}^2 and ρ_{tm}^2 indicate that variables $\tilde{\boldsymbol{x}}_{gk}$ and $\tilde{\boldsymbol{x}}_{tm}$ where $\tilde{\boldsymbol{x}}_{gk} = (x_{g1k}, ..., x_{gnk})^T$ and $\tilde{\boldsymbol{x}}_{tm} = (x_{t11k}, ..., x_{tnn_nk})^T$ are of high importance to the phenotype.

Priors on ρ_{gk}^2, ρ_{tm}^2, γ_{gk}, γ_{tm}, ξ^2 and σ_e^2

Let $\tau_{gk} = 1/\rho_{gk}^2$ and $\tau_{tm} = 1/\rho_{tm}^2$. We conduct Bayesian variable selection by imposing Gamma mixture prior on the parameters τ_{gk} and τ_{tm} [Zou et al., 2010]. We introduce the latent variables γ_{gk} and γ_{tm} to indicate which factors (genetic and nongenetic effects) are relevant (γ_{gk}, γ_{tm}=1), or irrelevant (γ_{gk}, γ_{tm}=0) to the phenotype. Specifically, the Gamma mixture priors for the parameters related to the genetic covariates are given by

$$P(\gamma_{gk} = 1) = 1 - P(\gamma_{gk} = 0) = p_{gk}, \quad p_{gk} \sim Be(p_{gk}|a_{g\gamma}, b_{g\gamma}), \tag{3.6}$$

$$\tau_{gk}|\gamma_{gk} \sim (1-\gamma_{gk})Ga(\tau_{gk}|\frac{\alpha_{g0}}{2}, \frac{\alpha_{g0}}{2\mu_{g0}}) + \gamma_{gk}Ga(\tau_{gk}|\frac{\alpha_{g1}}{2}, \frac{\alpha_{g1}}{2\mu_{g1}}),$$

where $Ga(\tau|a,b)$ is the Gamma density $\tau^{a-1}exp(-b\tau)b^a/\Gamma(a)$ and $Be(p|a,b)$ is the Beta density $p^{a-1}(1-p)^{b-1}/B(a,b)$. Similarly, the Gamma mixture priors for the parameters related to the nongenetic covariates are

$$P(\gamma_{tm} = 1) = 1 - P(\gamma_{tm} = 0) = p_{tm}, \quad p_{tm} \sim Be(p_{tm}|a_{t\gamma}, b_{t\gamma}), \tag{3.7}$$

$$\tau_{tm}|\gamma_{tm} \sim (1-\gamma_{tm})Ga(\tau_{tm}|\frac{\alpha_{t0}}{2}, \frac{\alpha_{t0}}{2\mu_{t0}}) + \gamma_{tm}Ga(\tau_{tm}|\frac{\alpha_{t1}}{2}, \frac{\alpha_{t1}}{2\mu_{t1}}),$$

respectively. Here, α_{g0}, α_{g1}, α_{t0} and α_{t1} are positive shape parameters, and μ_{g0}, μ_{g1}, μ_{t0} and μ_{t1} are the means of the two Gamma distributions in (3.6) and (3.7), respectively.

Let $\mu_{g0} = c_g^2\mu_{g1}$ and $\mu_{t0} = c_t^2\mu_{t1}$. If we set c_g^2 (or c_t^2) to a large value, μ_{g0} (or μ_{t0}) is large. In this case, when $\gamma_{gk} = 0$ (or $\gamma_{tm} = 0$), we let τ_{gk} (or τ_{tm}) be large and thus the corresponding variable is irrelevant. When $\gamma_{gk} = 1$ (or $\gamma_{tm} = 1$), we let τ_{gk} (or τ_{tm}) take on a small value, indicating the corresponding variable is important.

Let $\tau_\xi = 1/\xi^2$, $\tau_e = 1/\sigma_e^2$. We let the prior distributions of the two parameters be Gamma distributions with the following densities:

$$\tau_\xi \sim Ga(\frac{\alpha_\xi}{2}, \frac{\alpha_\xi}{2\mu_\xi}),\ \ P(\tau_\xi) = \frac{\tau_\xi^{\frac{\alpha_\xi}{2}-1}}{\Gamma(\frac{\alpha_\xi}{2})}(\frac{\alpha_\xi}{2\mu_\xi})^{\frac{\alpha_\xi}{2}} exp(-\frac{\alpha_\xi}{2\mu_\xi}\tau_\xi), \tag{3.8}$$

$$\tau_e \sim Ga(\frac{\alpha_e}{2}, \frac{\alpha_e}{2\mu_e}),\ \ P(\tau_e) = \frac{\tau_e^{\frac{\alpha_e}{2}-1}}{\Gamma(\frac{\alpha_e}{2})}(\frac{\alpha_e}{2\mu_e})^{\frac{\alpha_\xi}{2}} exp(-\frac{\alpha_e}{2\mu_e}\tau_e). \tag{3.9}$$

Here, α_ξ, α_e are positive shape parameters and μ_ξ, μ_e are the means of τ_ξ and τ_e, respectively.

Priors on $\boldsymbol{b}$, $\boldsymbol{\delta}$, $\boldsymbol{\psi}$

In model (3.4), we let the prior of b_{ij} follow a standard normal distribution, such that the joint prior distribution of the latent variable $\boldsymbol{b} = (\boldsymbol{b}_1^T, ..., \boldsymbol{b}_n^T)^T$ is $P(\boldsymbol{b}) \overset{d}{=} N(0, I)$. In order to choose priors that facilitate the posterior computation, we consider conjugate prior distributions for $\boldsymbol{\Delta}$ and $\boldsymbol{\Psi}$ and assume that $P(\boldsymbol{\delta}, \boldsymbol{\psi}) = P(\boldsymbol{\delta})P(\boldsymbol{\psi})$. Let $\boldsymbol{\delta} = (\delta_1, ..., \delta_k)^T$, $\boldsymbol{\psi} = (\psi_{ml} : m = 2, ..., k; l = 1, ..., m-1)^T$. Let $N^+(\mu, \sigma^2)$ denote the positive normal density of mean μ and variance σ^2 that is truncated at zero. The prior distribution for $\boldsymbol{\delta}$ is $P(\boldsymbol{\delta}) = \Pi_{l=1}^k P(\delta_l) = \Pi_{l=1}^k N^+(m_{l0}, s_{l0}^2)$ where m_{l0} and s_{l0}^2 are hyperparameters to be specified. The prior distribution for $\boldsymbol{\psi}$ is given by $P(\boldsymbol{\psi}) \overset{d}{=} N(\boldsymbol{\psi}_0, \boldsymbol{R}_0)$ where $\boldsymbol{\psi}_0$, $\boldsymbol{R}_0$ are pre-specified hyperparameters.

3.2.3 Posterior Calculation and MCMC Algorithm

We define $\boldsymbol{\theta}$ as the vector of all unknown parameters, such that $\boldsymbol{\theta} = (\boldsymbol{b}, \boldsymbol{\tau}, \tau_\xi, \tau_e, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\psi})$ where $\boldsymbol{\tau} = (\tau_{g1}, ..., \tau_{gp}, \tau_{t1}, ..., \tau_{tq})^T$ and $\boldsymbol{\gamma} = (\gamma_{g1}, ..., \gamma_{gp}, \gamma_{t1}, ..., \gamma_{tq})^T$.
Letting $\boldsymbol{y} = (\boldsymbol{y}_1^T, ..., \boldsymbol{y}_n^T)^T$, the joint posterior distribution of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ is given by

$$P(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta}) P(\boldsymbol{\eta}|\boldsymbol{\Sigma}_N) P(\boldsymbol{\theta}). \tag{3.10}$$

In order to draw random samples from $P(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{y})$, we use two different MCMC algorithms. For hyperparameters $\boldsymbol{\tau}$, τ_ξ and τ_e, we will use hybrid Monte Carlo method [Neal, 1997] because one cannot sample from their full conditional posteriors. The hybrid Monte Carlo method is a family of MCMC methods which merges the Metropolis-Hastings algorithm with sampling techniques based on dynamic systems in physics [Duane et al., 1987]. For parameters $\boldsymbol{b}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$ and $\boldsymbol{\psi}$, we will use either Gibbs Sampling or Metropolis-Hastings method. In this section, we describe the hybrid Monte Carlo method and conditional posteriors of $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$, $\boldsymbol{b}$, $\boldsymbol{\delta}$ and $\boldsymbol{\psi}$.

Posterior calculation via Hybrid Monte Carlo Method

One crucial problem in working with the joint posterior distribution in equation (3.10) occurs due to the discrete nature of marker data [Zou et al., 2010]. If markers of two individuals are identical (or similar), the covariance matrix of $\boldsymbol{\eta}$ will be (nearly) singular. The joint posterior marginalized with respect to $\boldsymbol{\eta}$ is considered as $P(\boldsymbol{\theta}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\theta}) P(\boldsymbol{\theta})$. The marginal likelihood of $\boldsymbol{y}|\boldsymbol{\theta}$ can be expressed by

$$\boldsymbol{y}|\boldsymbol{\theta} \sim N_N(0, \boldsymbol{\Sigma}) \text{ where } \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_N + \boldsymbol{p}(I_n \otimes \boldsymbol{D})\boldsymbol{p}^T + \frac{1}{\tau_e}\boldsymbol{I}_N, \tag{3.11}$$

where $\boldsymbol{p} = diag(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. The inference based on the joint posterior marginalized with respect to $\boldsymbol{\eta}$ provides clearly superior results to the joint posterior of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ [Zou

et al., 2010]. Letting $\boldsymbol{\nu} = (log(\tau_{g1}), ..., log(\tau_{gp}), , log(\tau_{t1}), ..., log(\tau_{tq}), log(\tau_{\xi}) \ , log(\tau_e))$, the joint posterior distribution marginalized with respect to $\boldsymbol{\eta}$ can be written by

$$P(\boldsymbol{\nu}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\nu}) \propto (2\pi)^{-\frac{N}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}exp(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y})P(\boldsymbol{\nu}). \tag{3.12}$$

The potential energy of the system is defined as

$$\mathcal{E}(\boldsymbol{\nu}) = -logP(\boldsymbol{\nu}|\boldsymbol{y}) \propto \frac{N}{2}log(2\pi) + \frac{1}{2}log|\boldsymbol{\Sigma}| + \frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} - logP(\boldsymbol{\nu}), \tag{3.13}$$

$$\frac{\partial\mathcal{E}(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}_i} = \frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\boldsymbol{\nu}_i}) - \frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\boldsymbol{\nu}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{y} - \frac{1}{P(\boldsymbol{\nu})}\frac{\partial P(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}_i}. \tag{3.14}$$

The kinetic energy of the system is defined as

$$\mathcal{K}(\boldsymbol{\phi}) = \frac{1}{2}\sum_{i=1}^{p+q+2}\phi_i{}^2, \tag{3.15}$$

where $\boldsymbol{\phi}$ is a momentum variable which has $p+q+2$ real-valued components, $\boldsymbol{\phi}_i$, in one-to-one correspondence with the components of $\boldsymbol{\nu}$. The total energy $\mathcal{H}$ of the system which is called "Hamiltonian" function is the sum of the kinetic energy $\mathcal{K}$ and the potential energy $\mathcal{E}$, such that $\mathcal{H}(\boldsymbol{\phi}, \boldsymbol{\nu}) = \mathcal{K}(\boldsymbol{\phi}) + \mathcal{E}(\boldsymbol{\nu})$. The dynamical system simulated through a virtual time t is governed by the following Hamilton's differential equations:

$$\frac{d\nu_i}{dt} = \frac{\partial\mathcal{H}}{\partial\phi_i} = \phi_i, \qquad \frac{d\phi_i}{dt} = -\frac{\partial\mathcal{H}}{\partial\nu_i} = -\frac{\partial\mathcal{E}}{\partial\nu_i}, \tag{3.16}$$

where ν_i is the ith element of $\boldsymbol{\nu}$. Since the partial derivative of $\mathcal{E}$ with respect to ν_i is complicated, the above equation cannot be simulated exactly. We use the leapfrog

steps to approximate the dynamic system via the following equations:

$$\begin{aligned} \phi_i(t+\frac{\epsilon}{2}) &= \phi_i(t) - \frac{\epsilon}{2}\frac{\partial \mathcal{E}}{\partial \nu_i}(\boldsymbol{\nu}(t)), \\ \nu_i(t+\epsilon) &= \nu_i(t) + \epsilon\phi_i(t+\frac{\epsilon}{2}), \\ \phi_i(t+\epsilon) &= \phi_i(t+\frac{\epsilon}{2}) - \frac{\epsilon}{2}\frac{\partial \mathcal{E}}{\partial \nu_i}(\boldsymbol{\nu}(t+\epsilon)), \end{aligned} \tag{3.17}$$

where ϵ is the step size for discretizing the dynamic system. The step sizes ϵ are set to the same value for all hyperparameters and are chosen to scale as $\epsilon \propto N^{-1/2}$ since the magnitude of the gradients under the posterior are expected to be scale roughly as $N^{1/2}$ when the prior is vague. Rasmussen [1996] found that $\epsilon = 0.5N^{-1/2}$ performs reasonably well. In summary, one iteration of the Hybrid Monte Carlo sampling is as follows:

1. Starting from $(i-1)$th sample $(\boldsymbol{\nu}^{i-1},\boldsymbol{\phi}^{i-1})$, perform one leap frog step using equations (4.13) with step size ϵ, resulting in the proposed value $(\boldsymbol{\nu}^*,\boldsymbol{\phi}^*)$.

2. With the acceptance rate, $\min(1, exp[\mathcal{H}(\boldsymbol{\nu}^{i-1},\boldsymbol{\phi}^{i-1}) - \mathcal{H}(\boldsymbol{\nu}^*,\boldsymbol{\phi}^*)])$, accept the proposed value as $(\boldsymbol{\nu}^i,\boldsymbol{\phi}^i) := (\boldsymbol{\nu}^*,\boldsymbol{\phi}^*)$; otherwise retain the previous values with negative momenta as $(\boldsymbol{\nu}^i,\boldsymbol{\phi}^i) := (\boldsymbol{\nu}^{i-1},-\boldsymbol{\phi}^{i-1})$.

3. Update the total energy of the system by perturbing the momenta according to $\phi_i := \alpha\phi_i + p_i\sqrt{1-\alpha^2}$ for all i, where p_i are randomly sampled from a standard normal distribution and α is set to 0.95 to ensure a reasonable level of perturbation [Rasmussen, 1996].

Since $\boldsymbol{\nu}$ and $\boldsymbol{\phi}$ are independent of each other, the Gibbs sampling of the momenta (step 3) allows the Hybrid Monte Carlo to explore regions with different values of $\mathcal{H}$. Finally, we can use the sequence $\{\nu_i | i = 1, ..., N\}$ as the samples generated from the posterior distribution $P(\boldsymbol{\nu}|\boldsymbol{y})$.

Conditional Posterior of $\boldsymbol{\gamma}$

Let $\boldsymbol{\theta}_{-z}$ be the remaining subvector of $\boldsymbol{\theta}$ after removing a subset of parameters, z, from $\boldsymbol{\theta}$. The full conditional distribution of γ_{gk} and γ_{tm} are given by

$$\begin{aligned} P(\gamma_{gk}=1|\boldsymbol{\theta}_{-\gamma_{gk}}) &= \frac{P(\tau_{gk}|\gamma_{gk}=1)P(\gamma_{gk}=1)}{P(\tau_{gk}|\gamma_{gk}=0)P(\gamma_{gk}=0)+P(\tau_{gk}|\gamma_{gk}=1)P(\gamma_{gk}=1)} \\ &= \frac{a_{g\gamma}(\frac{\alpha_{g1}}{2\mu_{g1}})^{\frac{\alpha_{g1}}{2}}\tau_{gk}^{\frac{\alpha_{g1}}{2}-1}exp(-\frac{\alpha_{g1}}{2\mu_{g1}}\tau_{gk})/\Gamma(\frac{\alpha_{g1}}{2})}{b_{g\gamma}(\frac{\alpha_{g0}}{2\mu_{g0}})^{\frac{\alpha_{g0}}{2}}\tau_{gk}^{\frac{\alpha_{g0}}{2}-1}exp(-\frac{\alpha_{g0}}{2\mu_{g0}}\tau_{gk})/\Gamma(\frac{\alpha_{g0}}{2})+a_{g\gamma}(\frac{\alpha_{g1}}{2\mu_{g1}})^{\frac{\alpha_{g1}}{2}}\tau_{gk}^{\frac{\alpha_{g1}}{2}-1}exp(-\frac{\alpha_{g1}}{2\mu_{g1}}\tau_{gk})/\Gamma(\frac{\alpha_{g1}}{2})}, \end{aligned} \tag{3.18}$$

$$\begin{aligned} P(\gamma_{tm}=1|\boldsymbol{\theta}_{-\gamma_{tm}}) &= \frac{P(\tau_{tm}|\gamma_{tm}=1)P(\gamma_{tm}=1)}{P(\tau_{tm}|\gamma_{tm}=0)P(\gamma_{tm}=0)+P(\tau_{tm}|\gamma_{tm}=1)P(\gamma_{tm}=1)} \\ &= \frac{a_{t\gamma}(\frac{\alpha_{t1}}{2\mu_{t1}})^{\frac{\alpha_{t1}}{2}}\tau_{tm}^{\frac{\alpha_{t1}}{2}-1}exp(-\frac{\alpha_{t1}}{2\mu_{t1}}\tau_{tm})/\Gamma(\frac{\alpha_{t1}}{2})}{b_{t\gamma}(\frac{\alpha_{t0}}{2\mu_{t0}})^{\frac{\alpha_{t0}}{2}}\tau_{tm}^{\frac{\alpha_{t0}}{2}-1}exp(-\frac{\alpha_{t0}}{2\mu_{t0}}\tau_{tm})/\Gamma(\frac{\alpha_{t0}}{2})+a_{t\gamma}(\frac{\alpha_{t1}}{2\mu_{t1}})^{\frac{\alpha_{t1}}{2}}\tau_{tm}^{\frac{\alpha_{t1}}{2}-1}exp(-\frac{\alpha_{t1}}{2\mu_{t1}}\tau_{tm})/\Gamma(\frac{\alpha_{t1}}{2})}. \end{aligned} \tag{3.19}$$

We sample $\boldsymbol{\gamma}$ directly from their conditional posterior distributions using Metropolis-Hastings algorithm.

Conditional Posterior of $\boldsymbol{\eta}$

In order to sample parameters related to the random effects, we first sample $\boldsymbol{\eta}$ from its conditional distribution. The full conditional distribution of $\boldsymbol{\eta}$ is given by

$$\begin{aligned} P(\boldsymbol{\eta}|\boldsymbol{\theta},\boldsymbol{y}) &\propto exp\{-\frac{\tau_e}{2}(\boldsymbol{y}-\boldsymbol{\eta}-\boldsymbol{vb})^T(\boldsymbol{y}-\boldsymbol{\eta}-\boldsymbol{vb})\}exp(-\frac{1}{2}\boldsymbol{\eta}^T\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\eta}) \\ &\propto exp\{-\frac{1}{2}(\boldsymbol{\eta}-\boldsymbol{\eta}^*)^T(\tau_e\boldsymbol{I}_N+\boldsymbol{\Sigma}_N^{-1})(\boldsymbol{\eta}-\boldsymbol{\eta}^*)\} \\ &\quad \text{, where } \boldsymbol{\eta}^*=\tau_e(\tau_e\boldsymbol{I}_N+\boldsymbol{\Sigma}_N^{-1})^{-1}(\boldsymbol{y}-\boldsymbol{vb}). \end{aligned} \tag{3.20}$$

Since $\boldsymbol{\eta}|\boldsymbol{\theta},\boldsymbol{y} \sim N_N(\boldsymbol{\eta}^*,\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^*)$, where $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^* = (\tau_e\boldsymbol{I}_N+\boldsymbol{\Sigma}_N^{-1})^{-1}$ and $\boldsymbol{\eta}^* = \tau_e\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^*(\boldsymbol{y}-\boldsymbol{vb})$, we sample $\boldsymbol{\eta}$ using Gibbs sampling scheme.

Conditional Posterior of $\boldsymbol{b}$

The full conditional distribution of $\boldsymbol{b}$ is given by

$$\begin{aligned} P(\boldsymbol{b}|\boldsymbol{\eta},\boldsymbol{\theta}_{-\boldsymbol{b}},\boldsymbol{y}) &\propto exp\{-\frac{\tau_e}{2}(\boldsymbol{y}-\boldsymbol{\eta}-\boldsymbol{v}\boldsymbol{b})^T(\boldsymbol{y}-\boldsymbol{\eta}-\boldsymbol{v}\boldsymbol{b})\}exp(-\frac{1}{2}\boldsymbol{b}^T\boldsymbol{b}) \\ &\propto exp\{-\frac{1}{2}(\boldsymbol{b}-\boldsymbol{b}^*)^T(\tau_e\boldsymbol{v}^T\boldsymbol{v}+\boldsymbol{I}_{nk})(\boldsymbol{b}-\boldsymbol{b}^*)\} \\ &\quad \text{where } \boldsymbol{b}^* = \tau_e(\tau_e\boldsymbol{v}^T\boldsymbol{v}+\boldsymbol{I}_{nk})^{-1}\boldsymbol{v}^T(\boldsymbol{y}-\boldsymbol{\eta}). \end{aligned} \tag{3.21}$$

Since $\boldsymbol{b}|\boldsymbol{\eta},\boldsymbol{\theta}_{-\boldsymbol{b}},\boldsymbol{y} \sim N_{nk}(\boldsymbol{b}^*,\boldsymbol{\Sigma}_b^*)$, where $\boldsymbol{\Sigma}_b^* = (\tau_e\boldsymbol{v}^T\boldsymbol{v}+\boldsymbol{I}_{nk})^{-1}$ and $\boldsymbol{b}^* = \tau_e\boldsymbol{\Sigma}_b^*\boldsymbol{v}^T(\boldsymbol{y}-\boldsymbol{\eta})$, we sample $\boldsymbol{b}$ using Gibbs sampling scheme.

Conditional Posterior of $\boldsymbol{\delta}$

In order to obtain the full conditional distribution of $\boldsymbol{\delta}$, we rewrite the model (3.4) as

$$y_{ij} = \eta_{ij} + \sum_{l=1}^{k}\delta_l(p_{ijl}(b_{il} + \sum_{m=1}^{l-1} b_{im}\psi_{lm})) + e_{ij}, \tag{3.22}$$

and define the $k \times 1$ vector $\boldsymbol{t}_{ij} = (t_{ij1},...,t_{ijk})^T = (p_{ijl}(b_{il} + \sum_{m=1}^{l-1} b_{im}\psi_{lm}) : l = 1,...,k)^T$ and $\xi_{ijl} = y_{ij} - \eta_{ij} - \Sigma_{m\neq l}t_{ijm}\delta_m$. The full conditional distribution of $\boldsymbol{\delta}$ is given by

$$\begin{aligned} P(\boldsymbol{\delta}|\boldsymbol{\eta},\boldsymbol{\theta}_{-\boldsymbol{\delta}},\boldsymbol{y}) &\propto exp\{-\frac{\tau_e}{2}(\boldsymbol{y}-\boldsymbol{\eta}-\boldsymbol{t}\boldsymbol{\delta})^T(\boldsymbol{y}-\boldsymbol{\eta}-\boldsymbol{t}\boldsymbol{\delta})\} \\ &\quad \times\prod_{l=1}^{k}\{exp(-\frac{1}{2s_{l0}^2}(\delta_l-m_{l0})^2)I(\delta_l>0)\} \ \textit{where } \boldsymbol{t} = (\boldsymbol{t}_{11},...,\boldsymbol{t}_{nn_n})^T, \end{aligned} \tag{3.23}$$

$$\begin{aligned} P(\delta_l|\boldsymbol{\eta},\boldsymbol{\theta}_{-\delta_l},\boldsymbol{y}) &\propto exp\{-\frac{\tau_e}{2}(\boldsymbol{\xi}_l-\boldsymbol{t}_l\delta_l)^T(\boldsymbol{\xi}_l-\boldsymbol{t}_l\delta_l)\}\{exp(-\frac{1}{2s_{l0}^2}(\delta_l-m_{l0})^2)I(\delta_l>0)\} \\ &\quad \text{where } \boldsymbol{\xi}_l = (\xi_{11l},...,\xi_{nn_nl})^T \ \textit{and} \ \boldsymbol{t}_l = (t_{11l},...,t_{nn_nl})^T \\ &\propto \{exp(-\frac{1}{2\sigma_l^{*2}}(\delta_l-\delta_l^*)^2)I(\delta_l>0)\} \\ &\quad \text{where } \sigma_l^{*2} = (\tau_e\boldsymbol{t}_l^T\boldsymbol{t}_l + s_{l0}^{-2})^{-1}, \ \delta_l^* = \sigma_l^{*2}(\tau_e\boldsymbol{t}_l^T\boldsymbol{\xi}_l + s_{l0}^{-2}m_{l0}). \end{aligned} \tag{3.24}$$

Since $\delta_l|\boldsymbol{\eta}, \boldsymbol{\theta}_{-\delta_l}, \boldsymbol{y} \sim N^+(\delta_l^*, \sigma_l^{*2})$ where $\sigma_l^{*2} = (\tau_e \boldsymbol{t}_l^T \boldsymbol{t}_l + s_{l0}^{-2})^{-1}$, $\delta_l^* = \sigma_l^{*2}(\tau_e \boldsymbol{t}_l^T \boldsymbol{\xi}_l + s_{l0}^{-2} m_{l0})$, we sample δ_l using Gibbs sampling scheme.

Conditional Posterior of $\boldsymbol{\psi}$

In order to obtain the full conditional distribution of $\boldsymbol{\psi}$, we rewrite the model (3.4) as

$$y_{ij} = \eta_{ij} + \sum_{l=1}^{k} b_{il}(\delta_l p_{ijl} + \sum_{m=l+1}^{k} \delta_m p_{ijm} \psi_{ml}) + e_{ij}, \tag{3.25}$$

and define the $k(k-1)/2 \times 1$ vector $\boldsymbol{u}_{ij} = (b_{il}\delta_m p_{ijm} : l = 1, .., k, m = l+1, ..., k)^T$. The full conditional distribution of $\boldsymbol{\psi}$ is given by

$$\begin{aligned} P(\boldsymbol{\psi}|\boldsymbol{y}, \boldsymbol{\theta}_{-\boldsymbol{\psi}}) &\propto exp\{-\frac{\tau_e}{2}(\boldsymbol{y}-\boldsymbol{\eta}-\boldsymbol{u}\boldsymbol{\psi})^T(\boldsymbol{y}-\boldsymbol{\eta}-\boldsymbol{u}\boldsymbol{\psi})\} exp(-\frac{1}{2}(\boldsymbol{\psi}-\boldsymbol{\psi}_0)^T \boldsymbol{R}_0^{-1}(\boldsymbol{\psi}-\boldsymbol{\psi}_0)) \\ &\quad \text{, where } \boldsymbol{u} = (\boldsymbol{u}_{11}, ..., \boldsymbol{u}_{nn_n})^T, \\ &\propto exp\{-\frac{1}{2}(\boldsymbol{\psi}-\boldsymbol{\psi}^*)^T \boldsymbol{\Sigma}_\psi^{*\,-1}(\boldsymbol{\psi}-\boldsymbol{\psi}^*)\} \\ &\quad \text{, where } \boldsymbol{\Sigma}_\psi^* = (\tau_e \boldsymbol{u}^T\boldsymbol{u} + \boldsymbol{R}_0^{-1})^{-1} \text{ and } \boldsymbol{\psi}^* = \boldsymbol{\Sigma}_\psi^*(\tau_e \boldsymbol{u}^T(\boldsymbol{y}-\boldsymbol{\eta}) + \boldsymbol{R}_0^{-1}\boldsymbol{\psi}_0). \end{aligned} \tag{3.26}$$

Since $\boldsymbol{\psi}|\boldsymbol{y}, \boldsymbol{\theta}_{-\boldsymbol{\psi}} \sim N(\boldsymbol{\psi}^*, \boldsymbol{\Sigma}_\psi^*) I(\boldsymbol{\psi} \in \boldsymbol{\Re}_{\boldsymbol{\delta}})$, where $\boldsymbol{\Sigma}_\psi^* = (\tau_e \boldsymbol{u}^T\boldsymbol{u} + \boldsymbol{R}_0^{-1})^{-1}$ and $\boldsymbol{\psi}^* = \boldsymbol{\Sigma}_\psi^*(\tau_e \boldsymbol{u}^T(\boldsymbol{y}-\boldsymbol{\eta}) + \boldsymbol{R}_0^{-1}\boldsymbol{\psi}_0)$, we sample $\boldsymbol{\psi}$ using Gibbs sampling scheme.

Choice of the Number of Grid Points

The critical issue with the proposed nonparametric GP model is how to efficiently choose the number of grid points, k. To select the optimal number of grid points, we evaluate the goodness of the predictive distributions of our GP model as presented in the previous chapter. Spiegelhalter et al. [2002] proposed the deviance information criterion (DIC) as DIC = $-2E_{\boldsymbol{\eta},\boldsymbol{\theta}|\boldsymbol{y}}\ \{logP(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta})\} + P_D$. The second term of DIC, P_D is the effective number of parameters, which is defined as $P_D = -2E_{\boldsymbol{\eta},\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta})\} + 2logP(\boldsymbol{y}|\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\theta}})$ where $\tilde{\boldsymbol{\eta}}$ and $\tilde{\boldsymbol{\theta}}$ are the posterior means of $\boldsymbol{\eta}$

and $\boldsymbol{\theta}$. Since $P(\boldsymbol{y}_i|\boldsymbol{\eta}_i, \boldsymbol{\theta}) \stackrel{d}{=} N(\boldsymbol{\eta}_i, \boldsymbol{p_i D p_i}^T + \sigma_e^2 \boldsymbol{I}_{n_i})$ in the model (3.1), DIC is easy to compute using MCMC samples. As stated by Robert and Titterington [2002], the same observed data are used twice to construct P_D, and thus the predictive distribution chosen by DIC overfits the observed data. To overcome this overfitting problem of DIC, Ando [2007] developed the following Bayesian predictive information criterion (BPIC) as BPIC = $-2E_{\boldsymbol{\eta},\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta})\} + 2n\hat{b}$ where $\hat{b}$ is the bias of the posterior mean of the expected loglikelihood. Under a certain mild regularity condition, the bias term is given approximately by $n\hat{b} \approx P_D$ [Ando, 2011], so that the simplified BPIC = $2E_{\boldsymbol{\eta},\boldsymbol{\theta}|\boldsymbol{y}}\{logP(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta})\} + 2P_D$. Note that the penalty term of the simplified BPIC is twice of that of original DIC. In order to choose the optimal number of grid points for our nonparametric GP model, we first compute both DIC and simplified BPIC with several pre-selected numbers of grid points. And then we select the number of grid points that gives the minimum DIC or simplified BPIC scores.

Implementation in gpmixed

The proposed methods have been implemented in the package "gpmixed" which is built on top of the C code (called "original gp" from here on) developed by Zou et al. [2010] for univariate trait mapping. The MCMC algorithm and data manipulation procedure were modified for longitudinal data. For the choice of the optimal number of grid points, the gpmixed provides both DIC and simplified BPIC scores.

3.3 Simulation Study and Real Data Analysis

3.3.1 Simulation I

In this section, we simulated data for evaluating our GP-based nonparametric Bayesian variable selection method for longitudinal data. We simulated a backcross population with 200 individuals and a single chromosome with 151 evenly spaced markers at 5cM intervals. Each individual has three to seven measures and the total number of observations is 1000. In order to investigate the ability of the proposed method, we consider four different setups. We first simulate a set of four QTL with only one four-way interaction plus one time effect (Setup 1). The four simulated QTL are located at markers 31, 61, 91 and 121, respectively. The simulated function $\boldsymbol{\eta}$ equals $\boldsymbol{\eta}(x_{i1},...,x_{i151},\boldsymbol{t}_i) = x_{i31}x_{i61}x_{i91}x_{i121} + \boldsymbol{t}_i$ where the x_{ik} $(k = 31, 61, 91, 121)$ are the genotypes of the four simulated QTL and $\boldsymbol{t}_i = (t_{i1},...,t_{in_i})^T$ is the vector of times of individual i. The time covariates $\boldsymbol{t}_i$ are randomly generated from the uniform distribution in $[0, 1]$. Next, we simulated the data sets containing QTL that have one three-way gene-gene interaction and one two-way gene-time interaction (Setup 2), or three QTL with main effect and gene-time interaction (Setup 3). The simulated function $\boldsymbol{\eta}$ for Setup 2 equals $\boldsymbol{\eta}(x_{i1},...,x_{i151},\boldsymbol{t}_i) = x_{i31}x_{i61}x_{i91} + x_{i121}\boldsymbol{t}_i$ and, for Setup 3, $\boldsymbol{\eta}(x_{i1},...,x_{i151},\boldsymbol{t}_i) = 0.5(x_{i31} + x_{i61} + x_{i91} + x_{i121}\boldsymbol{t}_i)$. Last, we simulated data containing four QTL with only main effects (Setup 4). The simulated function equals $\boldsymbol{\eta}(x_{i1},...,x_{i151},\boldsymbol{t}_i) = 0.5(x_{i31}+x_{i61}+x_{i91}+x_{i121}+\boldsymbol{t}_i)$. For all setups, we consider three grid time points at $(0, 0.5, 1)$ and set $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3) = (1, 1.2, 0.8)$ and $\boldsymbol{\psi} = (\psi_{21}, \psi_{31}, \psi_{32}) = (0.6, 0.4, 0.6)$, which results in $\boldsymbol{\nu}_i \sim N(0, \boldsymbol{D})$ with $diag(\boldsymbol{D}) = (1, 1.96, 0.97)$ and the lower triangle elements are $(d_{21}, d_{31}, d_{32}) = (0.72, 0.32, 0.81)$. We set $\sigma_e^2 = 1$.

For the analysis, we choose hyperparameters $\alpha_{x0} = \alpha_{t0} = \alpha_{x1} = \alpha_{t1} = 1$, $\alpha_\xi = \alpha_e = 0.5$, $c_x = c_t = 100$, and $\mu_\xi = \mu_e = 400$. We also set $a_{x\gamma} = a_{t\gamma} = 0.05$ and $b_{x\gamma} = b_{t\gamma} = 0.95$, so that

the prior probabilities that each variable (QTL or nongenetic covariate) is relevant or irrelevant to the phenotype are 0.05 and 0.95, respectively. The prior distributions for the elements of $\boldsymbol{\delta}$ are chosen to be independent $N^{+}(0, 30)$ and the prior distributions for the elements of $\boldsymbol{\psi}$ are independent $N(0, 0.5)$. We have a relatively large variance for the prior of $\boldsymbol{\delta}$ and a somewhat diffused variance for the prior of $\boldsymbol{\psi}$.

The upper left panel of Figure 3.1 displays the posterior mean estimates of the latent variable γ_{gk} and γ_{tm} for Setup 1 from gpmixed with all time points. All four QTL and time effect were detected using the criteria that the average marginal posterior probability of inclusion is larger than 0.5. In order to compare the ability of gpmixed with previously-proposed original gp [Zou et al., 2010], we conduct the analysis with a subset of the data with only one randomly selected measurement of each subject. The results are shown in the middle upper panel of Figure 3.1. All four QTL are identified though the signals are much smaller. The right upper panel of Figure 3.1 summaries estimated marginal posterior probability of each marker/gene for Setup 1 with all time point from the R/qtlbimmixed we proposed in the previous chapter on the full data. We calculated marginal posterior probability of inclusion based on Bayes factor (BF) of R/qtlbimmixed. No markers are detected because R/qtlbimmixed is only capable of detecting main and pairwise interactions.

The lower panels of Figure 3.1 summarize the results for Setup 2. Both gpmixed and the original gp detect the four QTL and time effect based on the criteria that the marginal posterior probability is larger than 0.5, while R/qtlbimmixed identifies only the QTL that interacts with the time covariate and misses the other 3 QTL with the three-way interaction. Setup 3 includes one two-way interaction and all methods find the four QTL but the marginal posterior inclusion probabilities from gpmixed are generally larger than the original gp in Figure 3.2. Setup 4 contains only main effects, and all QTL and time effects were identified by the three methods in Figure 3.2.

To further compare the methods, we generated receiver operating characteristic (ROC) curves by varying the cut-offs imposed on the posterior mean and BF, respectively. For each setup, we conduct 100 simulations with uniformly generated QTL positions which are restricted to be at least 10cM apart. First, we define true QTL intervals as the ones containing one true QTL and its two flanking markers. (if the true QTL is located at one of the two ends, only one flaking marker is included.) And then remaining genome is divided into non-overlapping 10 cM intervals. For a given cut-off on the posterior mean or the BF, a significant interval is defined as an interval containing at least one marker whose posterior mean or BF exceeds the cut-off value. An interval is a true positive if it is a significant true QTL interval, otherwise a false positive. We then define: True positive rate = (# of significant, true intervals)/(# of true intervals); False positive rate = (# of significant, false intervals)/(# of false intervals). The ROC curves up to a false positive rate of 0.2 are shown in Figure 3.3 for all four setups. For Setup 1, our new gpmixed on the full data outperforms R/qtlbimmixed. The new method slightly better performs the original gp method when the later is applied to the subset of the data. For Setup 2, the performance of our gpmixed is similar to the original gp on the subset of the data while both methods outperform R/qtlbimmixed. For Setups 3 and 4, our new gpmixed has higher true positive rates than the other methods.

In Figure 3.4, we display the trace plots of σ^2, δ_1, δ_2, δ_3, ψ_{21}, ψ_{31} and ψ_{32} for each setup. The black lines represent the values of the random draws for all parameters at each iteration and gray lines represents the true values of the parameters. It represents that all chains move around the true values for every parameter, indicating good convergence. Figure 3.5 presents the marginal posterior and prior densities of the parameters for random errors and random effects. It shows that the distributions of posterior samples look approximately normal around true value. Figure 3.6 displays

95% HPD intervals of σ^2, δ_1, δ_2, δ_3, ψ_{21}, ψ_{31} and ψ_{32} for each setup. The blue dots represent the posterior means and blue lines represent 95% HPD intervals. Most of 95% HPD interval contains true values. Table 3.1 summaries the posterior estimates of all parameters for random errors and random effects in the simulation study. The posterior means and medians are close to the true simulated values and all the 95% HPD intervals contain the true values, representing good performance of our algorithm.

3.3.2 Simulation II

We conduct another simulations and compute DIC [Spiegelhalter et al., 2002] and simplified BPIC [Ando, 2007, 2011] to access the number of grid points. The simulation settings are mostly the same as in simulation I except that the true number of grid points now varies from 2 to 4 (i.e., true $k = 2, 3, 4$). We simulate 100 data sets containing four QTL with main effects for each setup. The four simulated QTL are randomly located with at least 10cM apart. The simulated setup equals $\boldsymbol{y}_i = 0.5 \cdot (\boldsymbol{x}_{k_1 i} + \boldsymbol{x}_{k_2 i} + \boldsymbol{x}_{k_3 i} + \boldsymbol{x}_{k_4 i} + \boldsymbol{t}_i) + \boldsymbol{p}_i \boldsymbol{\nu}_i + \boldsymbol{e}_i$, where the $\boldsymbol{x}_{k_j i}$ (j = 1,2,3,4) are the genotypes of the four simulated QTL and $\boldsymbol{t}_i = (t_{i1}, ..., t_{in_i})^T$ are the times at which the ith individual has its data collected. We set $(\delta_1, \delta_2, \delta_3, \delta_4) = (1, 1.2, 0.8, 0.7)$ and $(\psi_{21}, \psi_{31}, \psi_{32}, \psi_{41}, \psi_{42}, \psi_{43}) = (0.6, 0.4, 0.6, 0.2, 0.4, 0.6)$. We evaluate DIC and simplified BPIC to select the number of grid points. Table 3.2 shows the average DIC, average simplified BPIC scores and the percentages of selecting true grid setup for nonparametric Bayesian model with different number of true grid points. All average DIC and average simplified BPIC scores achieved the minimum scores with true grid setup and the percentages of selection of true grid setup are 89%, 96% and 87% for 2, 3 and 4 grid setups using DIC, 98%, 94% and 79% using simplified BPIC in table 3.2.

3.3.3 Real Data Analysis

To further evaluate our GP-based nonparametric Bayesian model, we analyze a real mouse data on plasma HDL cholesterol regulation in backcross progeny involving NZB/ BlNJ and SM/J inbred strains [Pitman et al., 2002]. To identify QTL involved in plasma HDL cholesterol concentrations, SM females were mated to NZB males to produce F1 hybrids and F1 females were backcrossed with NZB males to generate 89 female backcross progeny. For all experiments, all female mice were fed a standard diet until 6-8 weeks of age (0 week time point), and then fed the high-fat diet for 18 weeks. Plasma HDL cholesterol (HDL-C) levels were measured at weeks 0,4,8 and 18. 53 of 89 backcross mice were genotyped at 79 markers. These include 3-5 markers per chromosome except for additional markers typed on those chromosomes with suggestive QTL. Interval mapping were performed by Pitman et al. [2002]. They reported the NZB alleles on chromosome 5(D5Mit370) and 18(D18Mit34) are associated with higher HDL-C concentration in the standard diet-fed mice while the NZB alleles on chromosome 5(D5Mit239) and 19(D19Mit71) are associated in the high-fat diet-fed mice. For our analysis, we consider HDL-C concentration measured at 0,4,8 and 18 weeks. We treat all 79 markers as putative QTL locations. We treat time as a fixed covariate. For all analyses, we run 1×10^4 iterations after discarding the first 1000 burn-ins. The MCMC chain was thinned by one in twenty, yielding 5×10^3 MCMC samples for the posterior analysis. The genomewide profile of inclusion probability for HDL-C is presented in Figure 3.7. This result shows a suggestive evidence of QTL activity on chromosome 5 (D5Mit10-D5MIT239) and 19 (D19MIT71) in Pitman et al. [2002]. No gene-time interactions are detected but time effect is very clear from Figure 3.7.

3.4 Discussion

We have extended the nonparametric Bayesian variable selection method with a Gaussian process prior to longitudinal traits. For modeling the dependence structure of the repeated measurements, we have used a grid-based covariance estimation method to accurately approximate the covariance structure. We have employed a modified Cholesky decomposition which provides an unconstrained reparameterization of any covariance matrix. To draw MCMC samples, both hybrid Monte Carlo and Gibbs sampling methods are used. For the hyperparameters related to QTL, we have adopted a hybrid Monte Carlo method because one cannot simply sample from their full conditional posteriors. For parameters related to the random effects, we have used either Gibbs Sampling or Metropolis-Hastings method.

The performance of the proposed method is evaluated by simulations. For data with higher-order time by QTL interactions, the proposed gpmixed on the full data performs much better than the R/qtlbimmixed and slightly better than the original gp on the subset of the data. For the main effect or pairwise interactions, all methods work reasonably well although our gpmixed has higher true positive rates than the other methods. As expected, the proposed gpmixed that utilizes the full data is more powerful than their corresponding univariate analysis method that only use the subset of the data. Furthermore, the gpmixed is more powerful than the original gp because only gpmixed consider within-subject correlation via the proposed grid-based covariance estimation method.

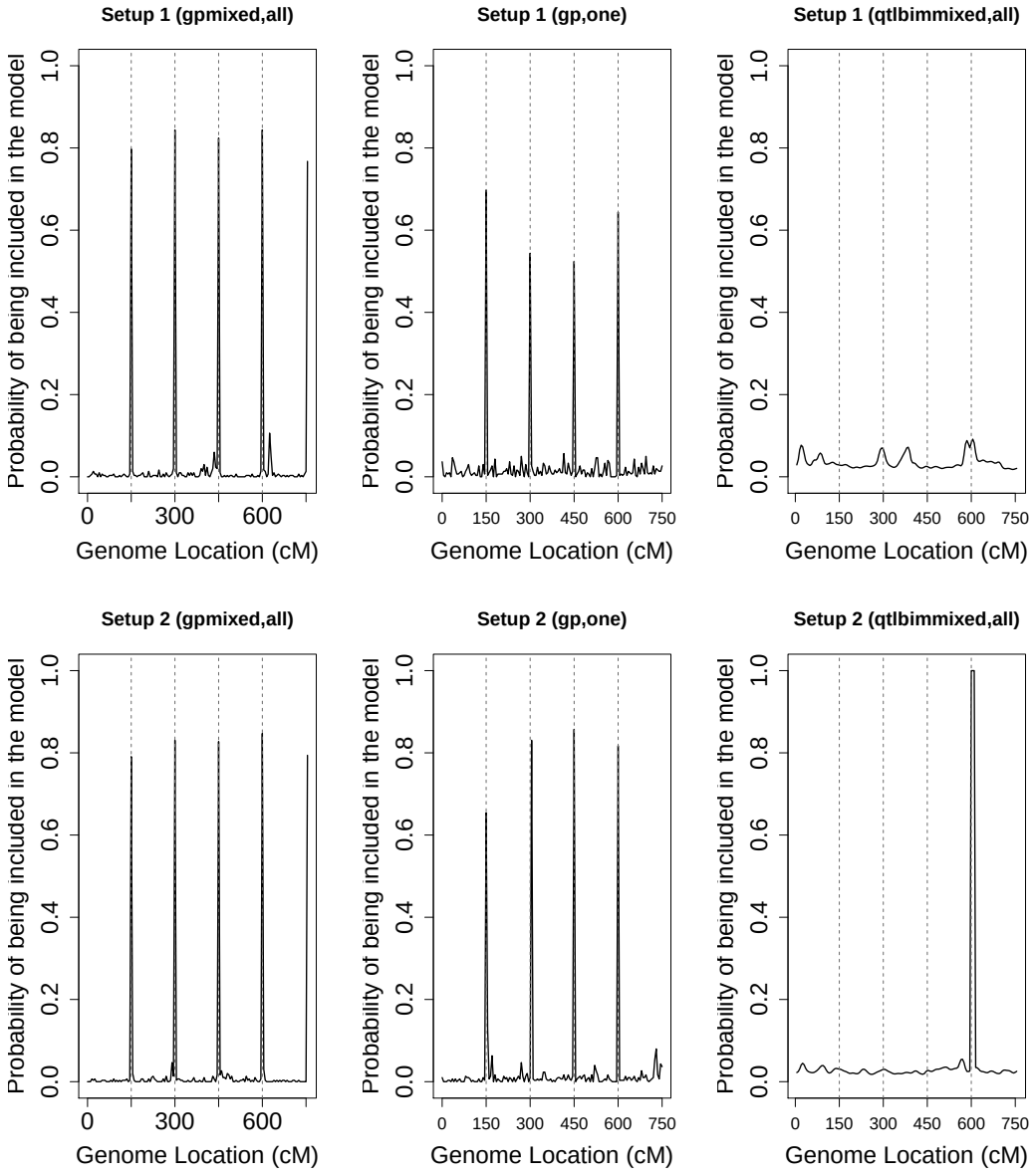

Figure 3.1: Posterior mean estimates of the latent variable γ_{gk} and γ_{tm} from gpmixed with all time points, original gp with one randomly selected time point and R/qtlbimmixed with all time points for Setups 1 and 2.

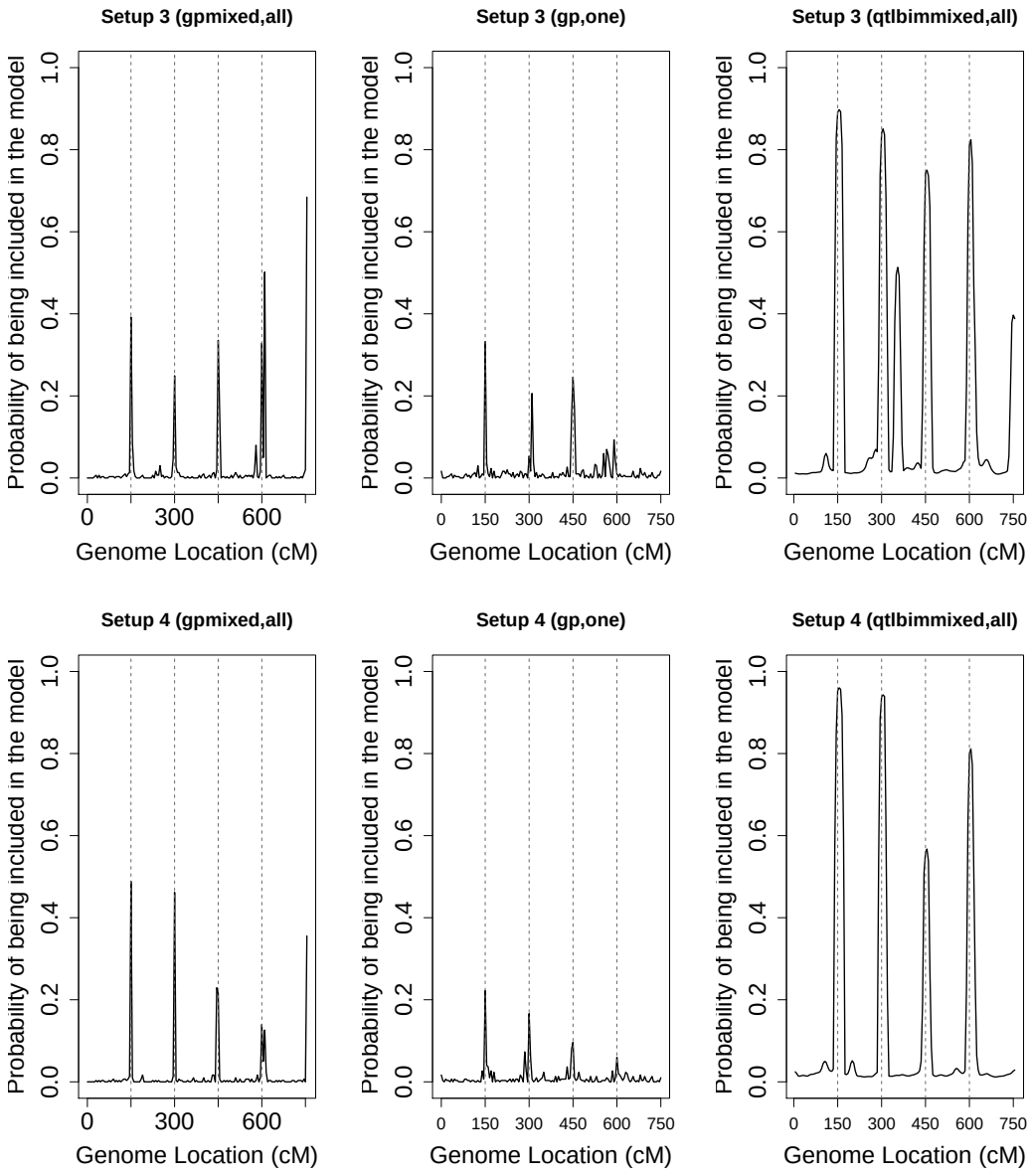

Figure 3.2: Posterior mean estimates of the latent variable γ_{gk} and γ_{tm} from gpmixed with all time points, original gp with one randomly selected time point and R/qtlbimmixed with all time points for for Setups 3 and 4.

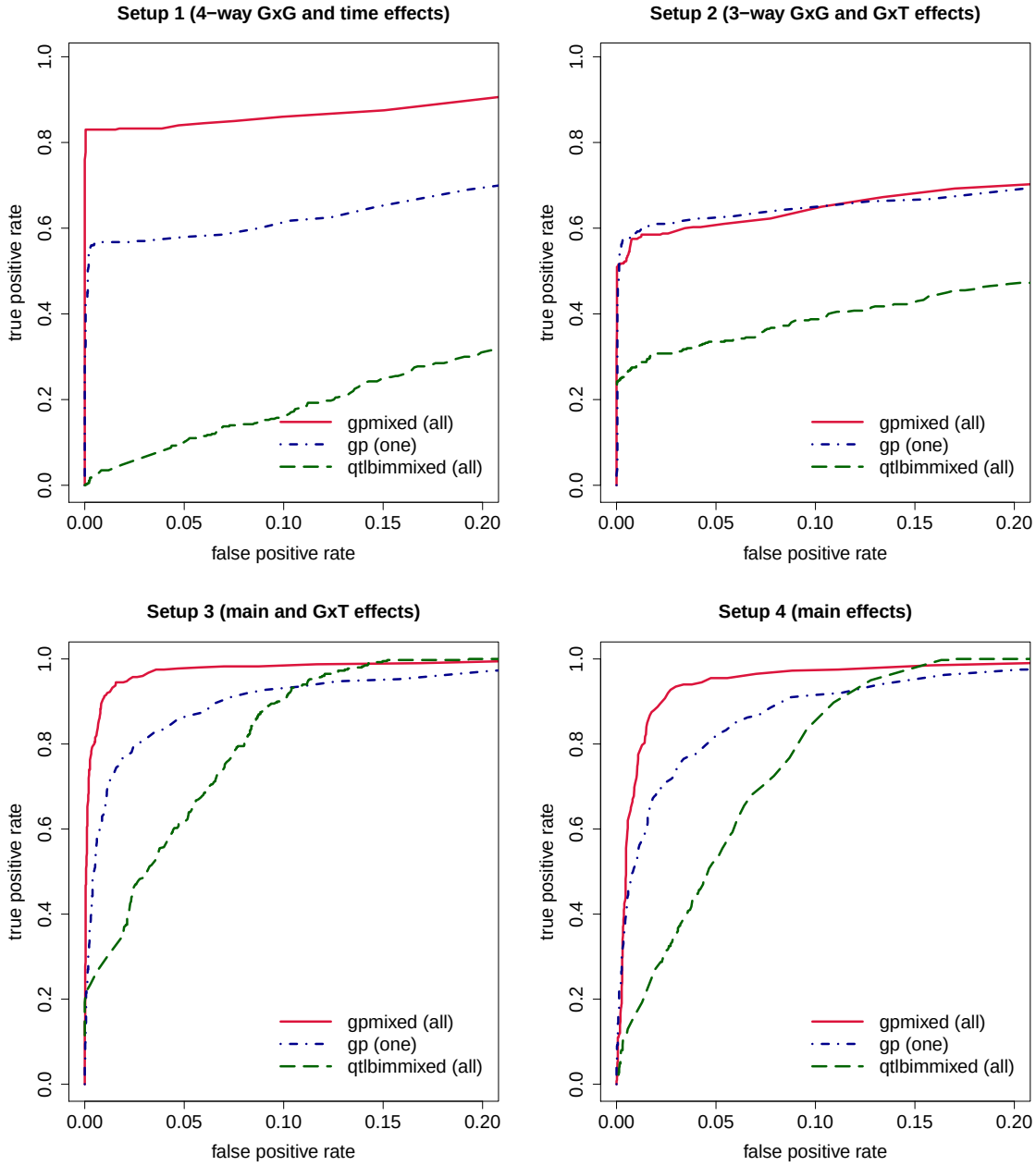

Figure 3.3: Estimated ROC curves for Setups 1,2,3 and 4: solid line (red) - proposed gpmixed on all data; dot-dashed line (blue) - original gp on one randomly selected time point data; long-dashed lines (green) - R/qtlbimmixed on all data.

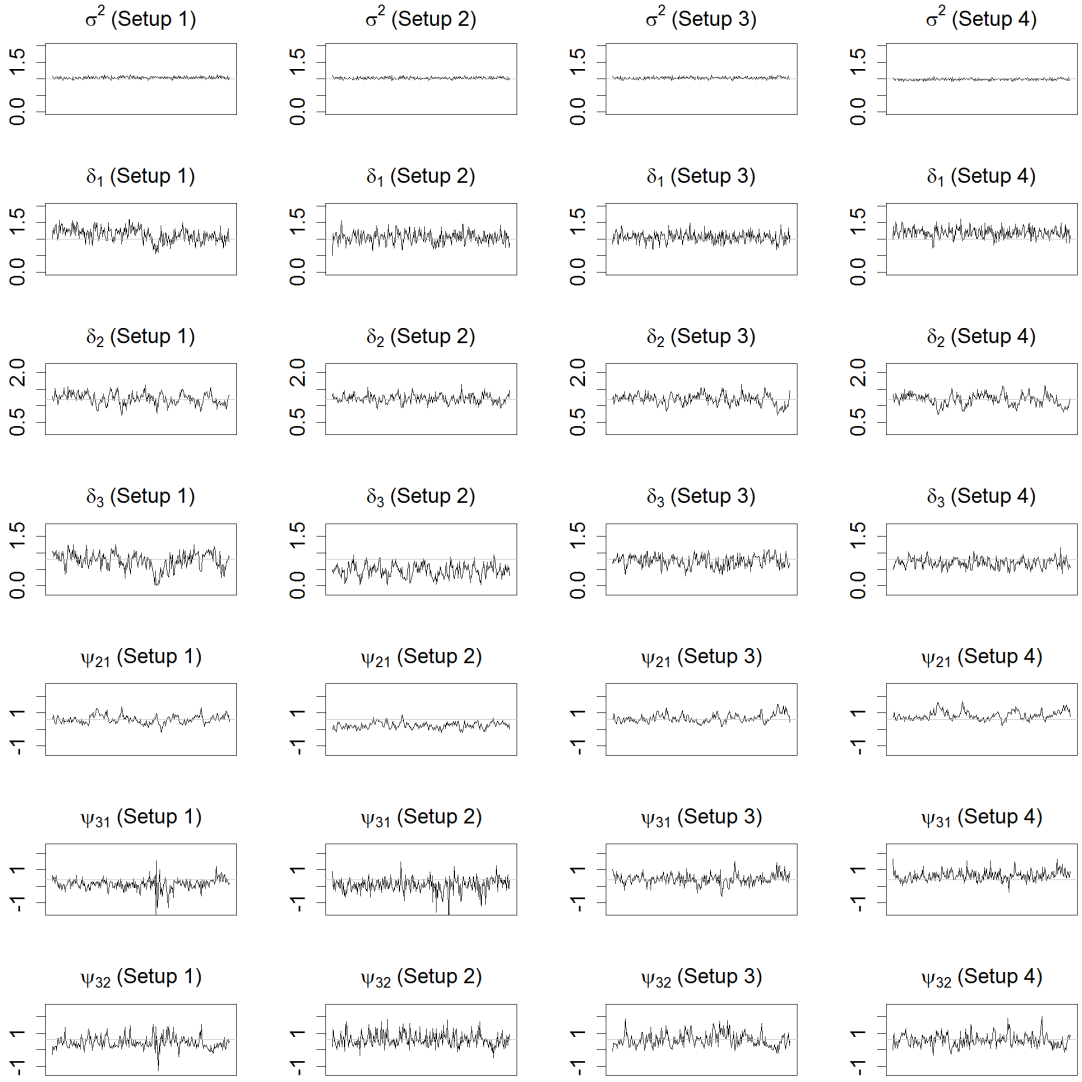

Figure 3.4: Trace plots of σ^2, δ_1, δ_2, δ_3, ψ_{21}, ψ_{31} and ψ_{32} for Setups 1,2,3 and 4 in the simulation study. The black lines represent the values of the draws for all parameters at each iteration and gray lines represents the true values of the parameters.

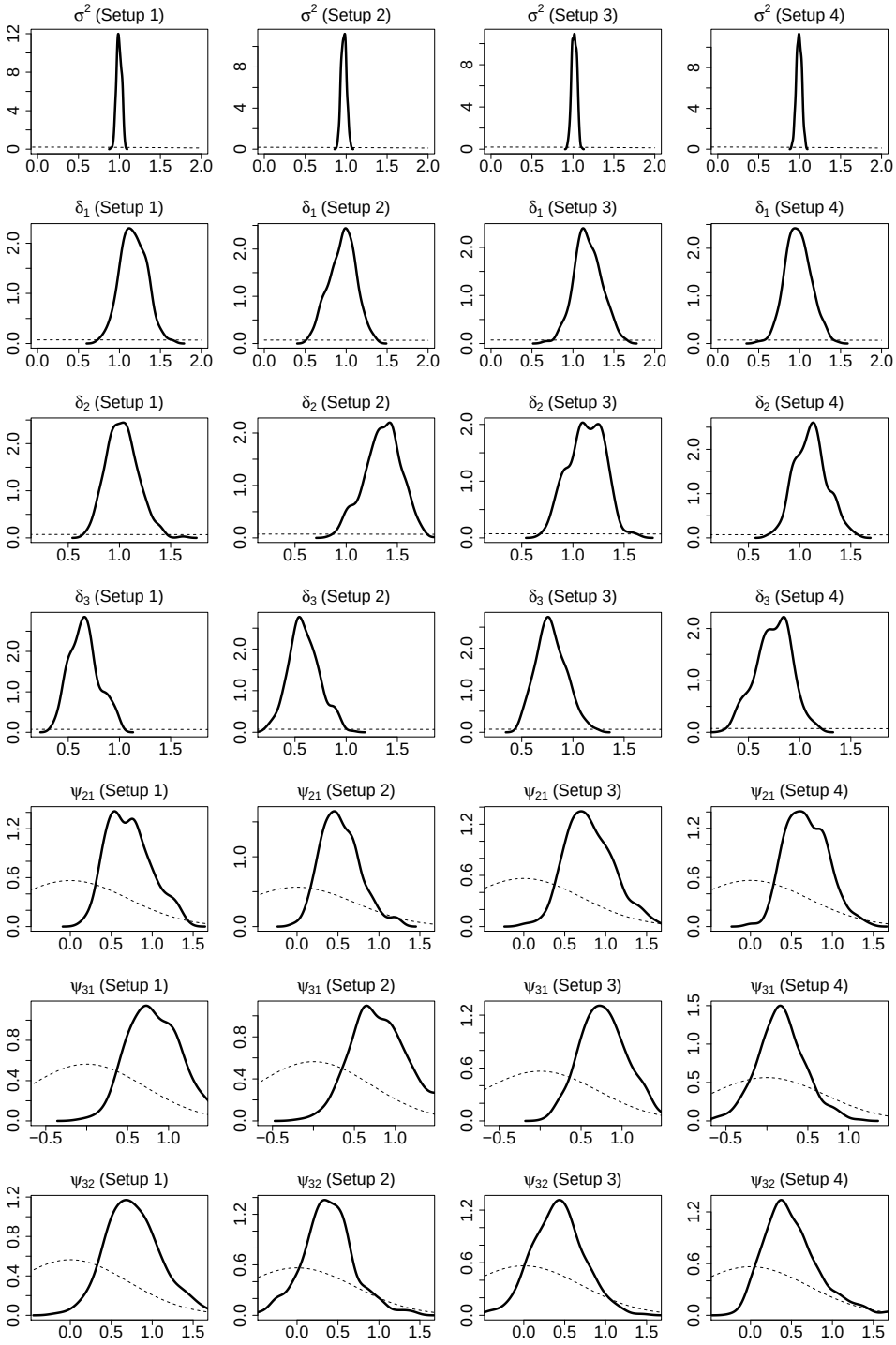

Figure 3.5: Posterior (solid line) and prior (dashed line) densities of the parameters for random errors and random effects for Setups 1,2,3 and 4.

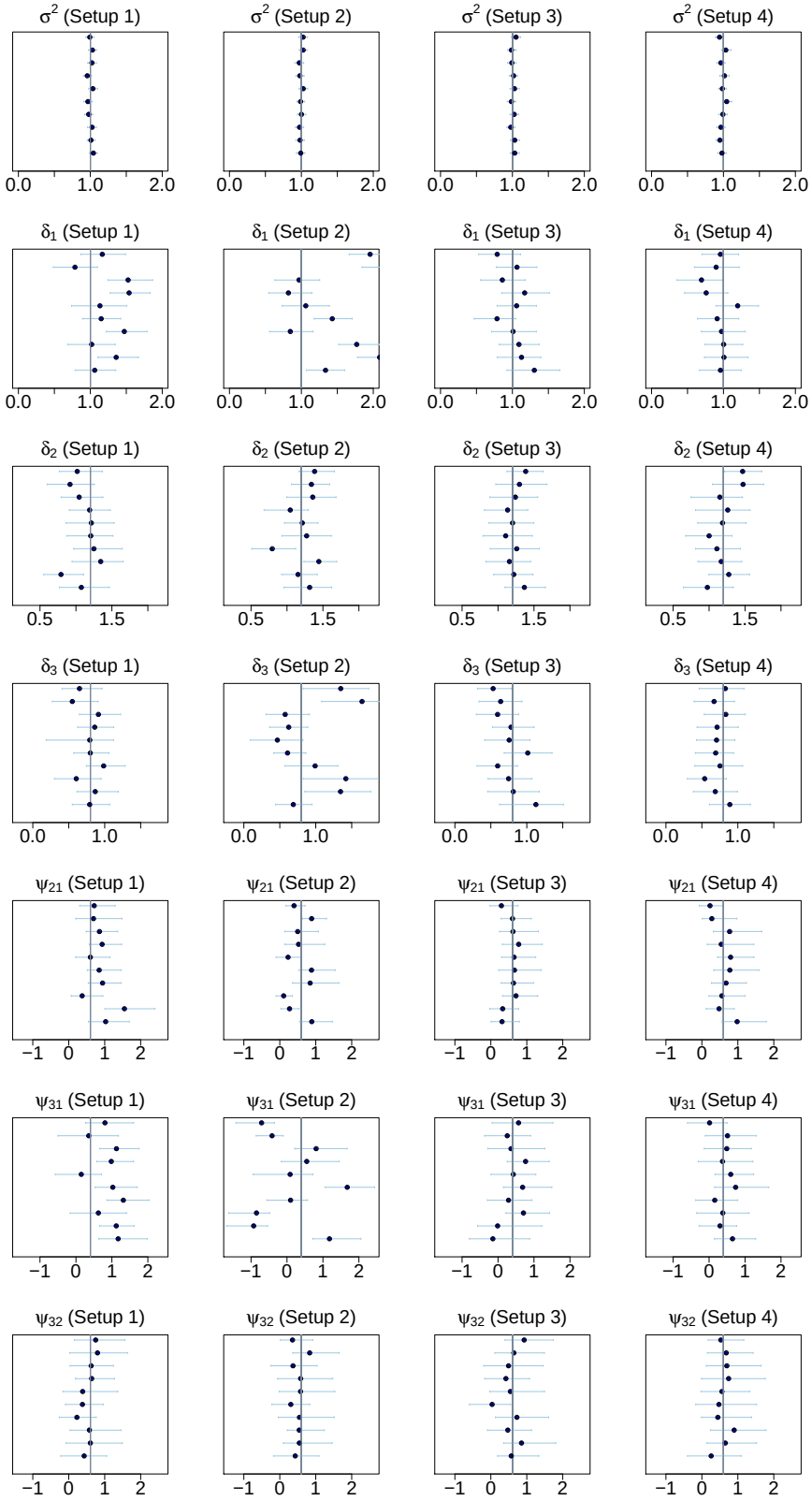

Figure 3.6: 95% HPD intervals of σ^2, δ_1, δ_2, δ_3, ψ_{21}, ψ_{31} and ψ_{32} for Setups 1,2,3 and 4. The blue dots represent the posterior means and blue lines represent HPD intervals.

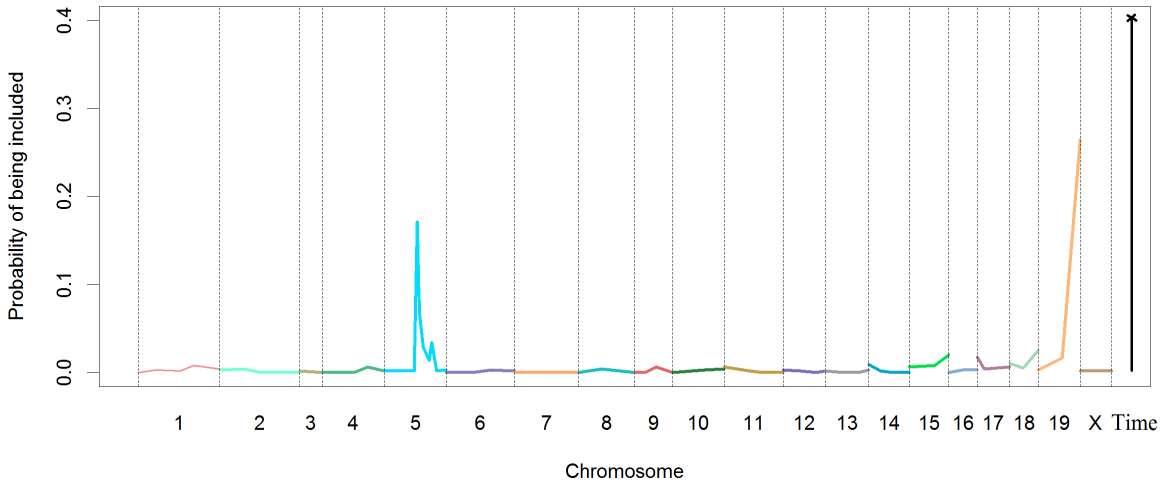

Figure 3.7: Genomewide profile of probability of being included in the model for plasma HDL cholesterol concentration in backcross progeny involving NZB/BINJ and SM/J inbred strains. It visualizes each chromosome in different color and time effect in black.

Table 3.1: Posterior means, medians, standard deviations and 95% HPD intervals of the parameters for random errors and random effects in the simulation study.

Setup	Para	True	Mean	Med	SD	95% HPD
1	σ^2	1	1	0.99	0.03	(0.94,1.05)
	δ_1	1	1.17	1.17	0.16	(0.87,1.49)
	δ_2	1.2	1.03	1.02	0.15	(0.77,1.37)
	δ_3	0.8	0.66	0.65	0.14	(0.41,0.96)
	ψ_{21}	0.6	0.72	0.71	0.27	(0.31,1.29)
	ψ_{21}	0.6	0.9	0.84	0.36	(0.3,1.78)
	ψ_{31}	0.4	0.85	0.81	0.35	(0.27,1.6)
	ψ_{32}	0.6	0.79	0.75	0.36	(0.16,1.56)
2	σ^2	1	0.97	0.97	0.03	(0.91,1.03)
	δ_1	1	0.95	0.97	0.16	(0.63,1.25)
	δ_2	1.2	1.36	1.36	0.18	(1,1.68)
	δ_3	0.8	0.59	0.57	0.15	(0.31,0.91)
	ψ_{21}	0.6	0.53	0.5	0.24	(0.14,1.08)
	ψ_{31}	0.4	0.85	0.81	0.38	(0.22,1.67)
	ψ_{32}	0.6	0.39	0.37	0.32	(-0.25,1.04)
3	σ^2	1	1.02	1.02	0.03	(0.96,1.07)
	δ_1	1	1.18	1.17	0.17	(0.85,1.51)
	δ_2	1.2	1.13	1.13	0.17	(0.81,1.41)
	δ_3	0.8	0.79	0.78	0.15	(0.53,1.1)
	ψ_{21}	0.6	0.8	0.77	0.29	(0.32,1.43)
	ψ_{31}	0.4	0.79	0.77	0.31	(0.26,1.42)
	ψ_{32}	0.6	0.43	0.42	0.32	(-0.17,1.07)
4	σ^2	1	0.99	0.99	0.03	(0.93,1.05)
	δ_1	1	0.98	0.97	0.16	(0.69,1.3)
	δ_2	1.2	1.11	1.11	0.16	(0.81,1.44)
	δ_3	0.8	0.74	0.75	0.17	(0.4,1.07)
	ψ_{21}	0.6	0.68	0.67	0.25	(0.27,1.24)
	ψ_{31}	0.4	0.18	0.16	0.29	(-0.37,0.8)
	ψ_{32}	0.6	0.51	0.45	0.36	(-0.01,1.38)

Table 3.2: Average DIC, average simplifed BPIC scores and percentage of selection of the right number of true grids for nonparametric Bayesian model with different number of true grid points.

True k	k	Avg DIC	#Sel (%)	Avg Sim BPIC	#Sel (%)	Avg P_D
2	2	3297.35	89	3327.69	98	30.49
	3	3300.92	11	3334.55	2	33.78
	4	3306.67	0	3344.17	0	35.25
3	2	3360.15	2	3389.81	4	29.66
	3	3336.31	96	3370.16	94	33.85
	4	3348.48	2	3386.21	2	37.74
4	2	3372.17	0	3403.83	6	31.66
	3	3358.41	13	3393.24	15	34.83
	4	3349.54	87	3389.36	79	39.82

CHAPTER 4

NONPARAMETRIC GAUSSIAN PROCESS MODEL FOR JOINT SNP-SET ANALYSIS

4.1 Introduction

Although genomewide association studies (GWAS) have successfully identified thousands of novel common variants [Gibson, 2012], these variants typically explain only a small fraction of overall heritability, which motivates the investigation of the so-called 'missing heritability'. One of the most important sources of 'missing heritability' is believed to be related to rare variants. Recent advances in whole genome genotyping and next generation sequencing technologies have led to the identification of rare variants (i.e. minor allele frequency (MAF) $< \sim 5\%$) in addition to common variants (i.e. MAF $> \sim 5\%$). As the statistical power to detect an association between a single rare variant and a complex trait is extremely low, rare variant analysis typically seeks to effectively combine multiple rare variants locally. Alternative strategies, such as regional SNP-set analysis have overcome some of the limitation of the standard single SNP analysis.

Several statistical methods have been proposed to jointly analyze multiple rare variants. The cohort allelic sum test (CAST) [Morgenthaler and Thilly, 2007] collapsed rare variants within a genomic region into a single binary variable to indicate whether the subject carries at least one copy of rare variants, and Li and Leal [2008] extended CAST by considering several subgroups instead of two subgroups. The weighted sum

test (WST) [Madsen and Browning, 2009] summarized multiple rare variants by weighting them based on their frequency in the unaffected individuals. Price et al. [2010] incorporated computational predictions of functional importance of each variant when summarizing multiple rare variants. Variance components tests such as C-alpha [Neale et al., 2011] and sequencing kernel association test (SKAT) [Wu et al., 2011] aggregated individual variant test statistics in a genomic region. Recently, Lee et al. [2012] and Derkach et al. [2013] proposed unified tests that combine collapsing and variance component tests. As a Bayesian approach, Yi and Zhi [2011] introduced a novel Bayesian hierarchical generalized linear model for analyzing multiple rare variants.

The most existing methods have been developed to assess only one group of rare or common variants at a time. Since complex traits are likely associated with many genes and environmental factors, it may be more powerful to simultaneously consider multiple groups of rare, common variants and covariates. Yi et al. [2011] introduced a Bayesian hierarchical generalized linear models (GLM) for simultaneously analyzing multiple groups of rare and/or common variants, in which variants within a gene are divided into multiple groups on the basis of their allele frequencies and their functions, and the group effects are jointly estimated. However, this method only allows for main effects and may miss variants with no marginal effects. It also only consider one region/gene at a time. Again, it is not feasible or intuitive to model all potential effects, including interactions parametrically. To get over these problems, we alternatively consider a nonparametric methods to search for sets of rare and common variants within a gene or across multiple genes which are associated with complex traits.

For such a purpose, in this chapter, we develop a Bayesian regional SNP-set analysis which extends the nonparametric Bayesian variable selection method with GP prior [Zou et al., 2010]. This method simultaneously models multiple groups of rare and/or common SNP variants. Instead of assigning a hyperparameter to each SNP, we assign a

common hyperparameter to all SNPs within a SNP-set to measure the cumulative effect of all SNPs. The usefulness of the proposed approach will be evaluated by simulation studies. Finally, we conclude the chapter with comments on future directions.

4.2 Nonparametric GP Model for Multiple Groups of Variants

4.2.1 GP-based Nonparametric Bayesian Model

Suppose that there are n unrelated individuals with a continuous phenotype and rare and/or common SNP variants in multiple genomic regions. Genomic regions can be defined, for example, by genes or by moving windows across the genome. The n observed response variables are denoted by $\boldsymbol{y} = (y_1, ..., y_n)^T$. For the ith individual, the genetic variants are divided into p groups, $\boldsymbol{x}_{gi} = (\boldsymbol{x}_{gi1}, ...\boldsymbol{x}_{gip})^T$, where the jth group $\boldsymbol{x}_{gij}$ contains m_i genetic variants. We have q nongenetic factors, $\boldsymbol{x}_{si} = (x_{si1}, ..., x_{siq})^T$ which are included in the following model as

$$y_i = \boldsymbol{\eta}(\boldsymbol{x}_{gi}, \boldsymbol{x}_{si}) + e_i,\ i = 1, ..., n, \tag{4.1}$$

where $\boldsymbol{\eta}$ is an unknown function of the p groups of genetic variants and the q nongenetic covariates; e_i is a random error with $N(0, \sigma_e^2)$.

To estimate $\boldsymbol{\eta}$, we assume that $\boldsymbol{\eta}$ follows a Gaussian process prior with $\boldsymbol{\eta} \sim \mathcal{GP}(0, \boldsymbol{\Sigma}_n)$. To jointly model multiple groups of genetic variants and other nongenetic covariates, we propose the novel covariance of $\boldsymbol{\eta}$ whose ii' $(i \neq i')$ element is expressed as

$$\Sigma_{ii'} = \xi^2 exp\{-\sum_{j=1}^{p}\sum_{k=1}^{m_j} \rho_{gj}^2 w_{jk}(x_{gijk} - x_{gi'jk})^2 - \sum_{l=1}^{q} \rho_{sl}^2 (x_{sil} - x_{si'l})^2\}, \tag{4.2}$$

where ξ^2, ρ_{gj}^2 and ρ_{sl}^2 are hyperparameters. Hyperparameter ξ^2 defines the vertical scale of variation and affects the magnitude of the exponential part. Hyperparameters

ρ^2_{gj} and ρ^2_{sl} determine the relevance of the various input covariates such as groups of genetic variants and nongenetic covariates. The common hyperparameter ρ^2_{gj} represents the association between the phenotype and all variants in the jth group. The common hyperparameter ρ^2_{gj} can be regarded as the cumulative importance of the m_j individual variants, hence referred as the group effect of the jth group. As seen in the equation (4.2), for the kth variant in jth group, we can weigh the contribution of each variant in a group differently by incorporating w_{jk} into the model. The weight w_{jk} can be set according to our prior knowledges or our mapping goals. For examples, (1) If all w_{jk} are set to 1, the method is the simple sum; (2) If $w_{jk} = sd(x_{gijk})$ where $sd(x_{gijk})$ is the estimated standard deviation of x_{gijk}, the model is comparable to the weighted-sum method [Madsen and Browning, 2009]; (3) If w_{jk} is set according to a prior probability of being functional for each variant, this is similar to Price et al. [2010]'s approach. These weights allow us to model the relative importance of each variants. Since the number of genetic variants may vary across groups, we also need to consider m_j, the number of genetic variants in group j. For example, we may let the weight be $w_{jk} = 1/m_j$. This weight enables the proposed model not to penalize groups with a large number of variants and thus we will use it in our analysis.

Let $\tau_{gj} = 1/\rho^2_{gj}$ and $\tau_{sl} = 1/\rho^2_{sl}$. As in the previous chapter, we conduct Bayesian variable selection by imposing Gamma mixture priors on the parameters τ_{gj} and τ_{sl} [Zou et al., 2010]. We introduce the latent variables γ_{gj} and γ_{sl} to indicate which factors (variant groups and nongenetic factors) are relevant (γ_{gj}, γ_{sl}=1), or irrelevant (γ_{gj}, γ_{sl}=0) to the phenotype. Specifically, the Gamma mixture priors for the parameters related to the variant groups are given by

$$P(\gamma_{gj} = 1) = 1 - P(\gamma_{gj} = 0) = p_{gj}, \quad p_{gj} \sim Be(p_{gj}|a_{g\gamma}, b_{g\gamma}), \tag{4.3}$$

$$\tau_{gj}|\gamma_{gj} \sim (1-\gamma_{gj})Ga(\tau_{gj}|\frac{\alpha_{g0}}{2}, \frac{\alpha_{g0}}{2\mu_{g0}}) + \gamma_{gj}Ga(\tau_{gj}|\frac{\alpha_{g1}}{2}, \frac{\alpha_{g1}}{2\mu_{g1}}),$$

where $Ga(\tau|a,b)$ is the Gamma density $\tau^{a-1}exp(-b\tau)b^a/\Gamma(a)$ and $Be(p|a,b)$ is the Beta density $p^{a-1}(1-p)^{b-1}/B(a,b)$. Similarly, the Gamma mixture priors for the parameters related to the nongenetic covariates are

$$P(\gamma_{sl}=1)=1-P(\gamma_{sl}=0)=p_{sl},\quad p_{sl}\sim Be(p_{sl}|a_{s\gamma},b_{s\gamma}), \tag{4.4}$$

$$\tau_{sl}|\gamma_{sl}\sim(1-\gamma_{sl})Ga(\tau_{sl}|\frac{\alpha_{s0}}{2},\frac{\alpha_{s0}}{2\mu_{s0}})+\gamma_{sl}Ga(\tau_{sl}|\frac{\alpha_{s1}}{2},\frac{\alpha_{s1}}{2\mu_{s1}}),$$

respectively. Here, α_{g0}, α_{g1}, α_{s0} and α_{s1} are positive shape parameters, and μ_{g0}, μ_{g1}, μ_{s0} and μ_{s1} are the means of the two Gamma distributions in (4.3) and (4.4), respectively. We set $\mu_{g0}=c_g^2\mu_{g1}$ and $\mu_{s0}=c_s^2\mu_{s1}$. A large c_g^2 (or c_s^2) implies a large μ_{g0} (or μ_{s0}). For $\gamma_{gj}=0$ (or $\gamma_{sj}=0$), a large τ_{gj} (or τ_{sj}) implies the corresponding variable is irrelevant. For $\gamma_{gj}=1$ (or $\gamma_{sj}=1$), a small τ_{gj} (or τ_{sj}) indicates the corresponding variable is important.

Letting $\tau_\xi=1/\xi^2$, $\tau_e=1/\sigma_e^2$, we set the prior distributions of the two parameters to be Gamma distributions with the following densities:

$$\tau_\xi\sim Ga(\frac{\alpha_\xi}{2},\frac{\alpha_\xi}{2\mu_\xi}),\ \ P(\tau_\xi)=\frac{\tau_\xi^{\frac{\alpha_\xi}{2}-1}}{\Gamma(\frac{\alpha_\xi}{2})}(\frac{\alpha_\xi}{2\mu_\xi})^{\frac{\alpha_\xi}{2}}exp(-\frac{\alpha_\xi}{2\mu_\xi}\tau_\xi), \tag{4.5}$$

$$\tau_e\sim Ga(\frac{\alpha_e}{2},\frac{\alpha_e}{2\mu_e}),\ \ P(\tau_e)=\frac{\tau_e^{\frac{\alpha_e}{2}-1}}{\Gamma(\frac{\alpha_e}{2})}(\frac{\alpha_e}{2\mu_e})^{\frac{\alpha_e}{2}}exp(-\frac{\alpha_e}{2\mu_e}\tau_e). \tag{4.6}$$

Here, α_ξ,α_e are positive shape parameters and μ_ξ,μ_e are the means of τ_ξ and τ_e, respectively.

4.2.2 Posterior Computation and Hybrid MCMC

We define $\boldsymbol{\theta} = (\boldsymbol{\tau}, \boldsymbol{\gamma}, \tau_\xi, \tau_e)$ as the vector of all unknown parameters where $\boldsymbol{\tau} = (\tau_{g1}, ..., \tau_{gp}, \tau_{s1}, ..., \tau_{sq})^T$ and $\boldsymbol{\gamma} = (\gamma_{g1}, ..., \gamma_{gp}, \gamma_{s1}, ..., \gamma_{sq})^T$. The joint posterior distribution of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ conditional on the phenotype $\boldsymbol{y}$ is given by

$$P(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta}) P(\boldsymbol{\eta}|\boldsymbol{\Sigma}_n) P(\boldsymbol{\theta}). \tag{4.7}$$

In order to sample from the posterior distribution $P(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{y})$, we adopt a hybrid Monte Carlo method [Neal, 1997]. The hybrid Monte Carlo method is a family of MCMC methods on the basis of the concept of dynamic systems in physics. Recently, it has been shown in various applications to converge significantly faster than the Metropolis-Hastings algorithm since the dynamic method avoids the random walk behavior [Rasmussen, 1996; Duane et al., 1987]. This section provides a brief description of Hybrid Monte Carlo method based on the implementation of Rasmussen [1996].

The likelihood function of $\boldsymbol{y}$ conditional on $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ is given by $\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta} \sim N_n(\boldsymbol{\eta}, \frac{1}{\tau_e}\boldsymbol{I}_n)$. The marginal likelihood of $\boldsymbol{y}|\boldsymbol{\theta}$ is the integral of the likelihood function of $\boldsymbol{y}$ times the prior as $P(\boldsymbol{y}|\boldsymbol{\theta}) = \int P(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta}) P(\boldsymbol{\eta}) d\boldsymbol{\eta}$. Since $\boldsymbol{\eta} \sim N_n(\boldsymbol{0}, \boldsymbol{\Sigma}_n)$, we have $\boldsymbol{y}|\boldsymbol{\theta} \sim N_n(0, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_n + \frac{1}{\tau_e}\boldsymbol{I}_n$. Since $\boldsymbol{\Sigma}$ is nonsingular even if $\boldsymbol{\Sigma}_n$ is singular and the inference based on the marginalized posterior provides clearly superior results to the joint posterior of both unknown parameters and $\boldsymbol{\eta}$ [Zou et al., 2010], we only consider the marginalized posterior from here on.

We set the log-transformed hyperparameters to $\boldsymbol{\nu} = (log(\tau_{g1}), ..., log(\tau_{gp}), log(\tau_{s1}), ..., log(\tau_{sq}), log(\tau_\xi), log(\tau_e))$ and the joint posterior distribution marginalized with respect to $\boldsymbol{\eta}$ can be written as

$$P(\boldsymbol{\nu}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\theta}) P(\boldsymbol{\nu}) \propto (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp(-\frac{1}{2}\boldsymbol{y}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{y}) P(\boldsymbol{\nu}). \tag{4.8}$$

The hybrid Monte Carlo method creates a virtual dynamic system by augmenting the log-transformed hyperparameters $\boldsymbol{\nu}$ with momentum variables $\boldsymbol{\phi}$. It samples from the distribution of the combined system $P(\boldsymbol{\nu}, \boldsymbol{\phi}) \propto exp(-\mathcal{E} - \mathcal{K})$, where $\mathcal{E}$ is the "potential energy" of the parameters and $\mathcal{K}$ is the "kinetic energy" of momenta. The potential energy of the system is defined as

$$\mathcal{E}(\boldsymbol{\nu}) = -logP(\boldsymbol{\nu}|\boldsymbol{y}) \propto \frac{n}{2}log(2\pi) + \frac{1}{2}log|\boldsymbol{\Sigma}| + \frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} - logP(\boldsymbol{\nu}), \tag{4.9}$$

$$\frac{\partial\mathcal{E}(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}_i} = \frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\boldsymbol{\nu}_i}) - \frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\boldsymbol{\nu}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{y} - \frac{1}{P(\boldsymbol{\nu})}\frac{\partial P(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}_i}. \tag{4.10}$$

The kinetic energy of the system is defined as

$$\mathcal{K}(\boldsymbol{\phi}) = \frac{1}{2}\sum_{i=1}^{p+q+2}{\phi_i}^2, \tag{4.11}$$

where $\boldsymbol{\phi}$ is momentum variable which has $p+q+2$ real-valued components, $\boldsymbol{\phi}_i$, in one-to-one correspondence with the components of $\boldsymbol{\nu}$. The total energy $\mathcal{H}$ of the system which is called "Hamiltonian" function is the sum of the kinetic energy $\mathcal{K}$ and the potential energy $\mathcal{E}$, such that $\mathcal{H}(\boldsymbol{\phi}, \boldsymbol{\nu}) = \mathcal{K}(\boldsymbol{\phi}) + \mathcal{E}(\boldsymbol{\nu})$. The dynamical system evolved through virtual time t is governed by the following Hamilton's differential equations:

$$\frac{d\nu_i}{dt} = \frac{\partial\mathcal{H}}{\partial\phi_i} = \phi_i, \qquad \frac{d\phi_i}{dt} = -\frac{\partial\mathcal{H}}{\partial\nu_i} = -\frac{\partial\mathcal{E}}{\partial\nu_i}, \tag{4.12}$$

where ν_i is the ith elements of $\boldsymbol{\nu}$. Since the partial derivative of $\mathcal{E}$ with respect to ν_i is a complicated function, the above equation cannot be simulated exactly. We use the

leapfrog steps to approximate the dynamic system using the following equations:

$$\begin{aligned} \phi_i(t+\frac{\epsilon}{2}) &= \phi_i(t) - \frac{\epsilon}{2}\frac{\partial \mathcal{E}}{\partial \nu_i}(\boldsymbol{\nu}(t)), \\ \nu_i(t+\epsilon) &= \nu_i(t) + \epsilon\phi_i(t+\frac{\epsilon}{2}), \\ \phi_i(t+\epsilon) &= \phi_i(t+\frac{\epsilon}{2}) - \frac{\epsilon}{2}\frac{\partial \mathcal{E}}{\partial \nu_i}(\boldsymbol{\nu}(t+\epsilon)), \end{aligned} \tag{4.13}$$

where ϵ is the step size for discretizing the dynamic system. The step sizes ϵ are set to the same value for all hyperparameters and are chosen to scale as $\epsilon \propto n^{-1/2}$ since the magnitude of the gradients under the posterior are expected to be scale roughly as $n^{1/2}$ when the prior is vague. Rasmussen [1996] found that $\epsilon = 0.5n^{-1/2}$ performs reasonably well. In summary, one iteration of the Hybrid Monte Carlo sampling is as follows:

1. Starting from $(i-1)$th sample $(\boldsymbol{\nu}^{i-1},\boldsymbol{\phi}^{i-1})$, perform one leap frog step using equations (4.13) with step size ϵ, resulting in the proposed value $(\boldsymbol{\nu}^*,\boldsymbol{\phi}^*)$.

2. With the acceptance rate, $\min(1, exp[\mathcal{H}(\boldsymbol{\nu}^{i-1},\boldsymbol{\phi}^{i-1}) - \mathcal{H}(\boldsymbol{\nu}^*,\boldsymbol{\phi}^*)])$, accept the proposed value as $(\boldsymbol{\nu}^i,\boldsymbol{\phi}^i) := (\boldsymbol{\nu}^*,\boldsymbol{\phi}^*)$; otherwise retain the previous values with negative momenta as $(\boldsymbol{\nu}^i,\boldsymbol{\phi}^i) := (\boldsymbol{\nu}^{i-1},-\boldsymbol{\phi}^{i-1})$.

3. Update the total energy of the system by perturbing the momenta according to $\phi_i := \alpha\phi_i + p_i\sqrt{1-\alpha^2}$ for all i, where p_i are randomly sampled from a standard normal distribution and α is set to 0.95 to ensure a reasonable level of perturbation [Rasmussen, 1996].

Since $\boldsymbol{\nu}$ and $\boldsymbol{\phi}$ are independent of each other, the Gibbs sampling of the momenta (step 3) allows the Hybrid Monte Carlo to explore regions with different values of $\mathcal{H}$. Finally, we can use the sequence $\{\nu_i | i = 1, ..., n\}$ as the samples generated from the posterior distribution $P(\boldsymbol{\nu}|\boldsymbol{y})$.

In order to sample the γ_{gj}, γ_{sl}, we derive the full conditional distribution as follows:

$$P(\gamma_{gj}=1|\boldsymbol{\theta}_{-\gamma_{gj}}) = \frac{P(\tau_{gj}|\gamma_{gj}=1)P(\gamma_{gj}=1)}{P(\tau_{gj}|\gamma_{gj}=0)P(\gamma_{gj}=0)+P(\tau_{gj}|\gamma_{gj}=1)P(\gamma_{gj}=1)} \tag{4.14}$$
$$= \frac{a_{g\gamma}(\frac{\alpha_{g1}}{2\mu_{g1}})^{\frac{\alpha_{g1}}{2}}\tau_{gj}^{\frac{\alpha_{g1}}{2}-1}exp(-\frac{\alpha_{g1}}{2\mu_{g1}}\tau_{gj})/\Gamma(\frac{\alpha_{g1}}{2})}{b_{g\gamma}(\frac{\alpha_{g0}}{2\mu_{g0}})^{\frac{\alpha_{g0}}{2}}\tau_{gj}^{\frac{\alpha_{g0}}{2}-1}exp(-\frac{\alpha_{g0}}{2\mu_{g0}}\tau_{gj})/\Gamma(\frac{\alpha_{g0}}{2})+a_{g\gamma}(\frac{\alpha_{g1}}{2\mu_{g1}})^{\frac{\alpha_{g1}}{2}}\tau_{gk}^{\frac{\alpha_{g1}}{2}-1}exp(-\frac{\alpha_{g1}}{2\mu_{g1}}\tau_{gj})/\Gamma(\frac{\alpha_{g1}}{2})},$$

$$P(\gamma_{sl}=1|\boldsymbol{\theta}_{-\gamma_{sl}}) = \frac{P(\tau_{sl}|\gamma_{sl}=1)P(\gamma_{sl}=1)}{P(\tau_{sl}|\gamma_{sl}=0)P(\gamma_{sl}=0)+P(\tau_{sl}|\gamma_{sl}=1)P(\gamma_{sl}=1)} \tag{4.15}$$
$$= \frac{a_{s\gamma}(\frac{\alpha_{s1}}{2\mu_{s1}})^{\frac{\alpha_{s1}}{2}}\tau_{sl}^{\frac{\alpha_{s1}}{2}-1}exp(-\frac{\alpha_{s1}}{2\mu_{s1}}\tau_{sl})/\Gamma(\frac{\alpha_{s1}}{2})}{b_{s\gamma}(\frac{\alpha_{s0}}{2\mu_{s0}})^{\frac{\alpha_{s0}}{2}}\tau_{sl}^{\frac{\alpha_{s0}}{2}-1}exp(-\frac{\alpha_{s0}}{2\mu_{s0}}\tau_{sl})/\Gamma(\frac{\alpha_{s0}}{2})+a_{s\gamma}(\frac{\alpha_{s1}}{2\mu_{s1}})^{\frac{\alpha_{s1}}{2}}\tau_{sl}^{\frac{\alpha_{s1}}{2}-1}exp(-\frac{\alpha_{s1}}{2\mu_{s1}}\tau_{sl})/\Gamma(\frac{\alpha_{s1}}{2})},$$

where $\boldsymbol{\theta}_{-z}$ is the remaining subvector of $\boldsymbol{\theta}$ after removing parameter subset z from $\boldsymbol{\theta}$. We sample $\boldsymbol{\gamma}$ directly from their conditional posterior distributions using Metropolis-Hastings algorithm.

4.3 Simulation Study

In this section, we present simulation studies to evaluate the proposed GP model for multiple groups of rare and/or common SNPs. We considered 300 individuals (n = 300), each with total of 50 groups of SNPs. To simulate SNPs with realistic linkage disequilibrium (LD), we used the calibrated coalescent models [Schaffner et al., 2005] which can simulate one or more populations retrospectively. In our simulation, we restricted our samples to European population. SNPs in each group span a 1Mb region. Each SNP has two states: 0 and 1 which represent the major and the minor alleles, respectively. Three types of SNPs are simulated:(1) rare SNPs: all SNPs are rare (i.e. $0.01 \leq \text{MAF} < 0.05$); (2) common SNPs: all SNPs are common (i.e. $\text{MAF} \geq 0.05$); (3) mixed SNPs: SNPs can be rare or common.

From the 50 simulated groups, we selected the 10th, 20th, 30th and 40th groups as

the causal groups. In each causal group, there are three causal SNPs. We simulated responses under four settings. For individual i, the phenotype value was first generated as $y_i = c_{g1}(x_{g10,i} + x_{g20,i} + x_{g30,i} + x_{g40,i}) + e_i$, where $x_{gk,i}$ $(k = 10, 20, 30, 40)$ are the sum of the three causal variants in the kth simulated group, c_{g1} is used to control over all genetic effects of the causal variants and e_i was generated from $N(0, 1)$ (Setup 1). Next, we simulated data where causal variants have two gene-gene interactions (Setup 2), or a three-way interaction (Setup 3). Specifically, the simulated phenotype of individual i, y_i for Setup 2 equals $c_{g2}(x_{g10,i}x_{g20,i} + x_{g30,i}x_{g40,i}) + e_i$ and y_i for Setup 3 equals $c_{g3}(x_{g10,i}x_{g20,i}x_{g30,i} + x_{g40,i}) + e_i$. Last, we simulated data containing one four-way gene-gene interaction (Setup 4) where $y_i = c_{g4}(x_{g10,i}x_{g20,i}x_{g30,i}x_{g40,i}) + e_i$. We varied c_{gj} $(j = 1, 2, 3, 4)$ in our simulation to ensure that the proposed method has adequate power. After removing all causal SNPs from the downstream analysis, each individual has 1000 SNPs and each group either has the same number of SNPs (20 SNPs) or varying number of SNPs (10,15,20,25 or 30 SNPs) in them. For the analysis, we chose hyperparameters $\alpha_{x0} = \alpha_{s0} = \alpha_{x1} = \alpha_{s1} = 1$, $\alpha_\xi = \alpha_e = 0.5$, $c_x = c_t = 100$, and $\mu_\xi = \mu_e = 400$. We also set $a_{x\gamma} = a_{t\gamma} = 0.05$ and $b_{x\gamma} = b_{t\gamma} = 0.95$, so that the prior probabilities that each variable is relevant or irrelevant to the phenotype are 0.05 and 0.95, respectively.

Figure 4.1 displays the posterior mean estimates of the latent variable γ_{gk} for Setup 1 and Setup 2 using the proposed GP model and the original GP model of Zou et al. [2010] with common variants. The results for Setup 3 and Setup 4 are shown in Figure 4.2. For all the setups, the proposed GP model performs better than the original GP model in general. A similar conclusion holds when all the genetic variants are rare (Figures 4.3 and 4.4). The results with both common and rare variants for all Setups are presented in Figures 4.5 and 4.6. The proposed GP model identified the true causal variants clearer than the original GP model.

To further compare the proposed method with the original GP model, we generated

receiver operating characteristic (ROC) curves by varying the cut-offs imposed on the inclusion probability with 100 simulations for each setup. For the original GP model, a group is defined as significant if at least one variant in the group exceeds the given cut-off value. We define a significant group as a true positive finding if it contains any causal variants. Otherwise, it is a false positive. We let true positive rate = (# of significant, true groups)/(# of true groups) and false positive rate = (# of significant, false groups)/(# of false groups). The estimated ROC curves with common variants are shown in Figures 4.7 and 4.8 for all four setups. The proposed GP model clearly outperforms the original GP model for all situations. The estimated ROC curves with rare variants are displayed in Figures 4.9 and 4.10. For Setup 1 and 2, the proposed GP model performs better than the original GP model, especially when the false positive rate is low. For setup 3 and 4, the proposed GP model performs much better than the original GP model. The estimated ROC curves with both common and rare variants are presented in Figures 4.11 and 4.12. For all cases, the proposed GP model clearly outperforms the original GP model.

4.4 Discussion

We have proposed a nonparametric Bayesian variable selection method with Gaussian process priors for simultaneously analyzing multiple groups of rare and common variants and nongenetic covariates. Since complex traits are likely influenced by complex interactions among genes and nongenetic factors, the joint analysis of multiple groups of genetic variants can potentially improve mapping power. However, most existing methods only allow for main effects and often ignore higher interactions among genetic variants and environmental factors. Our proposed novel nonparametric GP model measures the importance of each group regardless of whether it functions through its main effect or epistasis effects with other groups or environmental factors. We modified the

covariance of unknown function in the GP prior by introducing a common hyperparameter for all the variants in each group, though each variant can be weighted differently. The common hyperparameter represents the association between the phenotype and the genetic variants in the group. The proposed GP model outperforms the original GP model for both common and rare variants. In addition to the advantages described above, our method enhances the speed of computation since the proposed GP model dramatically reduces the number of hyperparameters in the original GP model.

As seen in the covariance of $\boldsymbol{\eta}$, each genetic variant can be modified by a pre-specified weight. Weights can be set according to our prior knowledge on the genetic variant from other GWAS studies or based on the functional importance of the variants such as whether they are in exon or intron regions. Weights can also be used to deal with the number of genetic variants, m_j in each group which may varies across groups. In our analysis, we set $w_{jk} = 1/m_j$. This weight enables the proposed model not to penalize groups with a large number of variants. Furthermore, we can assign different weights to common and rare variants based on MAF of each variant. How to choose weights efficiently is important and will be carefully evaluated as a future research. Finally, we are currently extending nonparametric Gaussian process model for SNP-set analysis from univariate traits to longitudinal traits. The grid-based approach proposed in the previous chapter will be applied to the method and approximate the covariance matrix of each subject for SNP-set analysis.

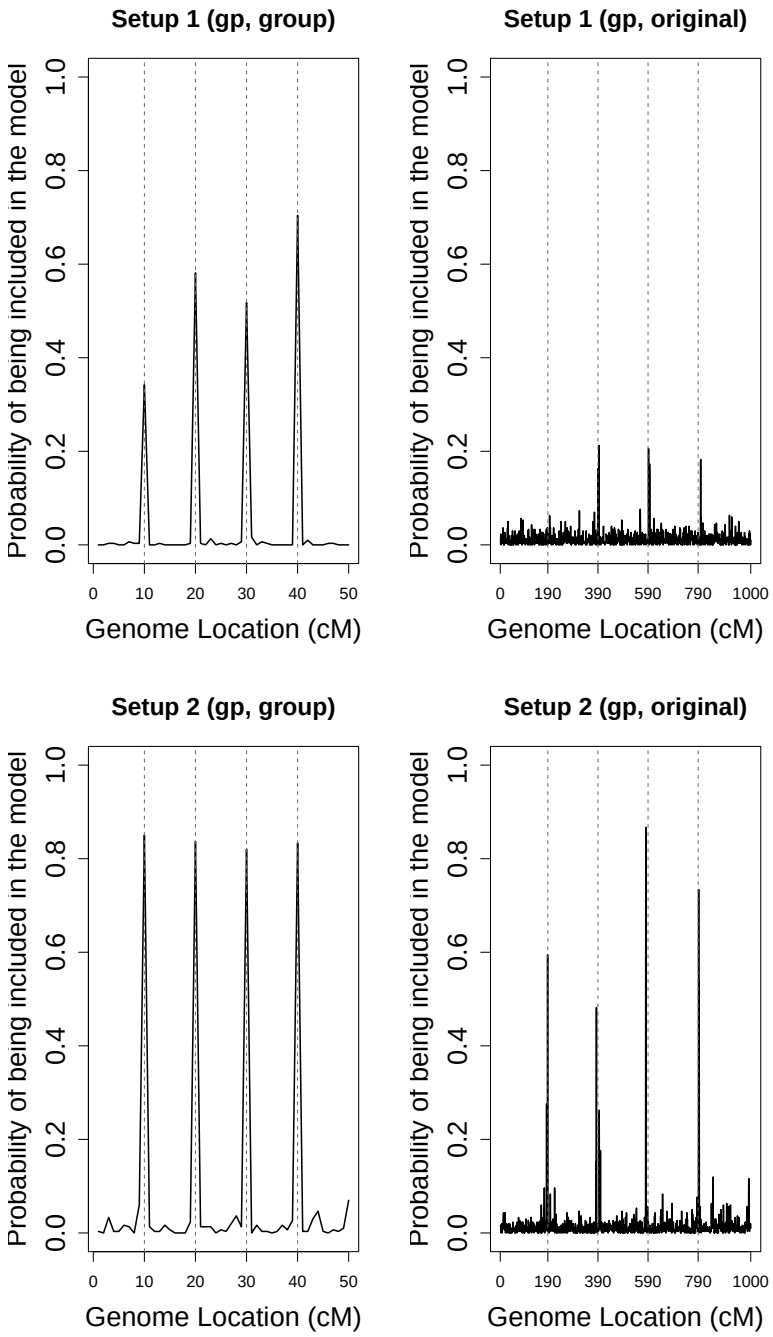

Figure 4.1: Posterior mean estimates of the latent variable γ_{gk} from the proposed GP model and the original GP model for Setups 1 and 2 with common variants ($c_{g1} = 0.30$ and $c_{g2} = 0.20$).

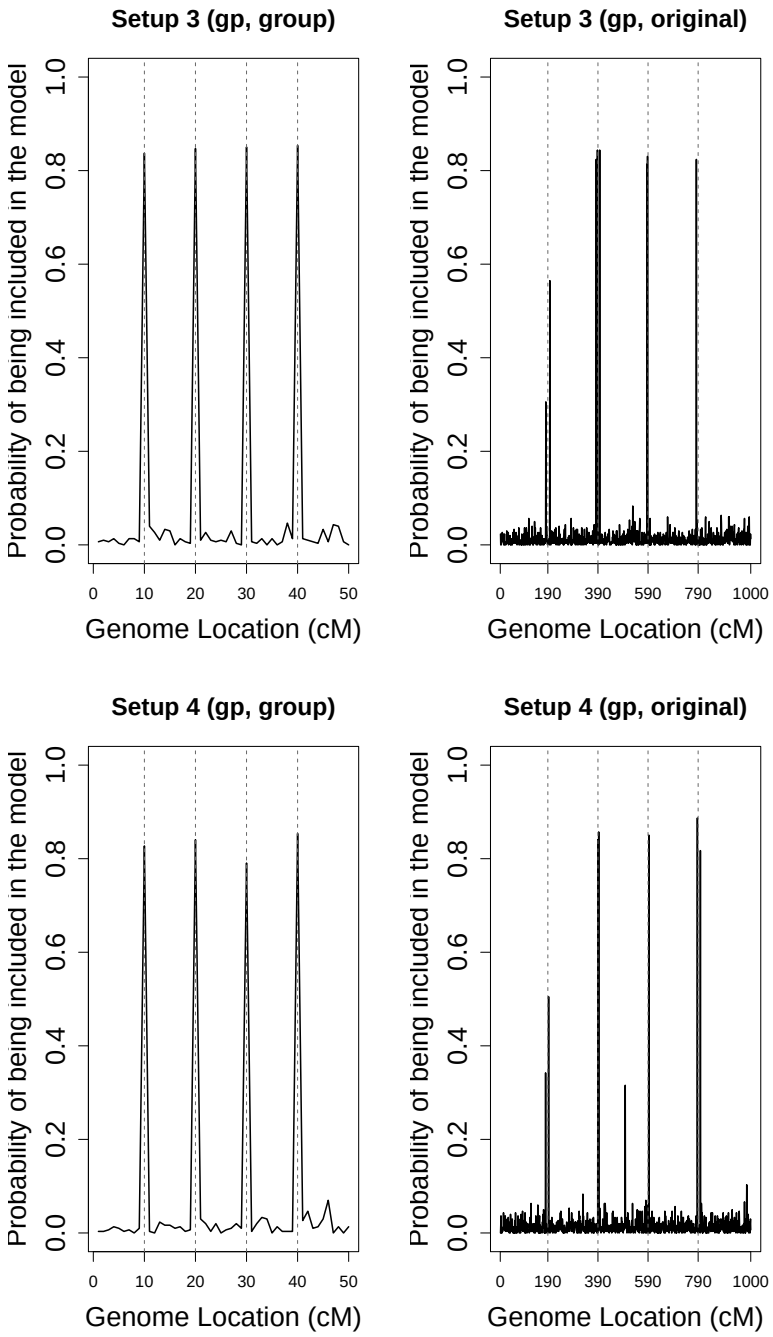

Figure 4.2: Posterior mean estimates of the latent variable γ_{gk} from the proposed GP model and the original GP model for Setups 3 and 4 with common variants ($c_{g3} = 0.20$ and $c_{g4} = 0.20$).

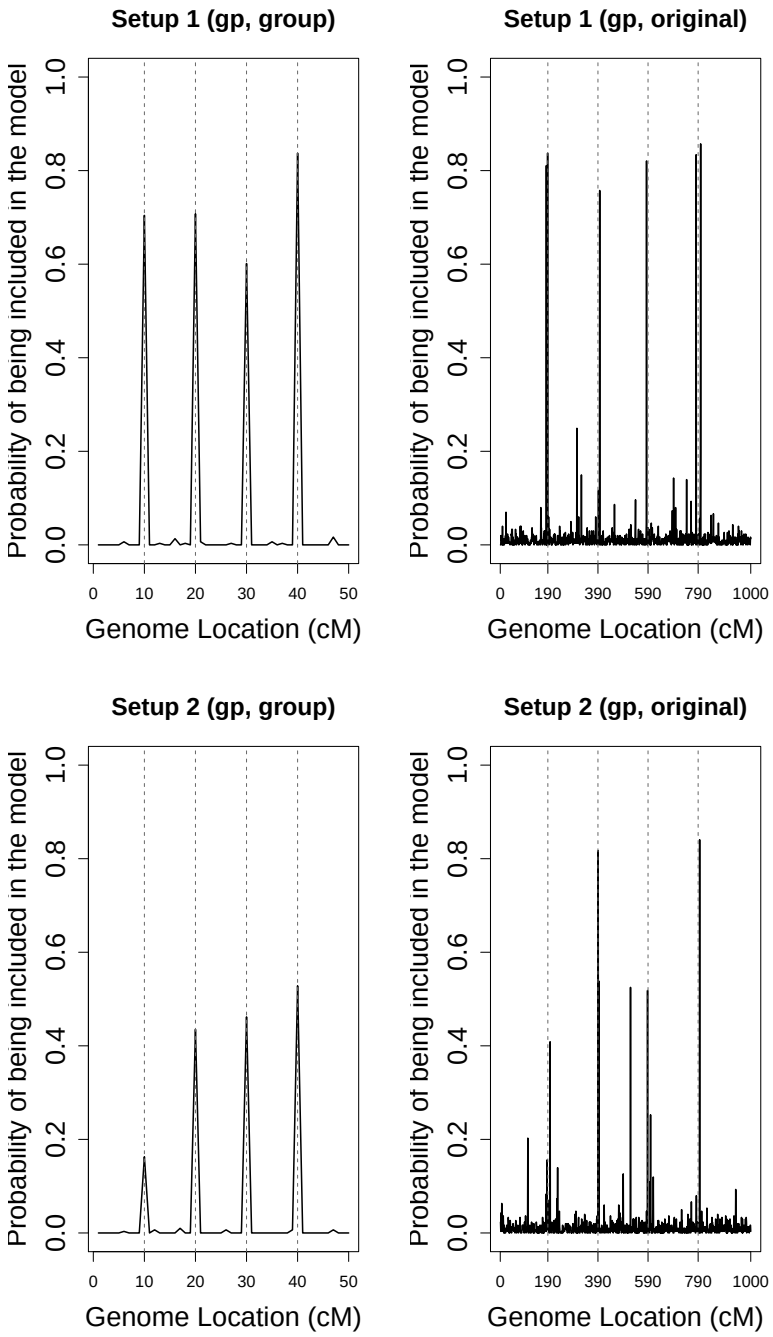

Figure 4.3: Posterior mean estimates of the latent variable γ_{gk} from the proposed GP model and the original GP model for Setups 1 and 2 with rare variants ($c_{g1} = 2.50$ and $c_{g2} = 1.00$).

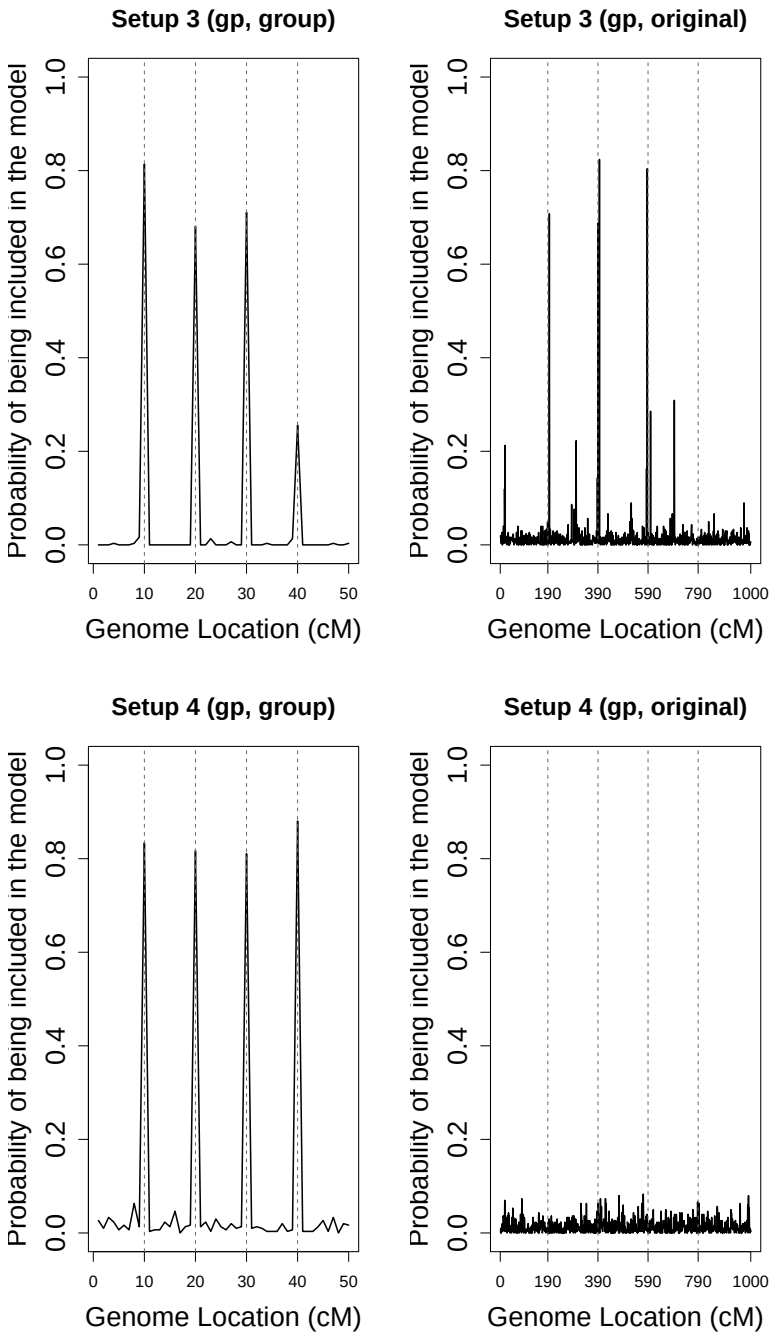

Figure 4.4: Posterior mean estimates of the latent variable γ_{gk} from the proposed GP model and the original GP model for Setups 3 and 4 with rare variants (c_{g3} = 2.00 and c_{g4} = 1.00).

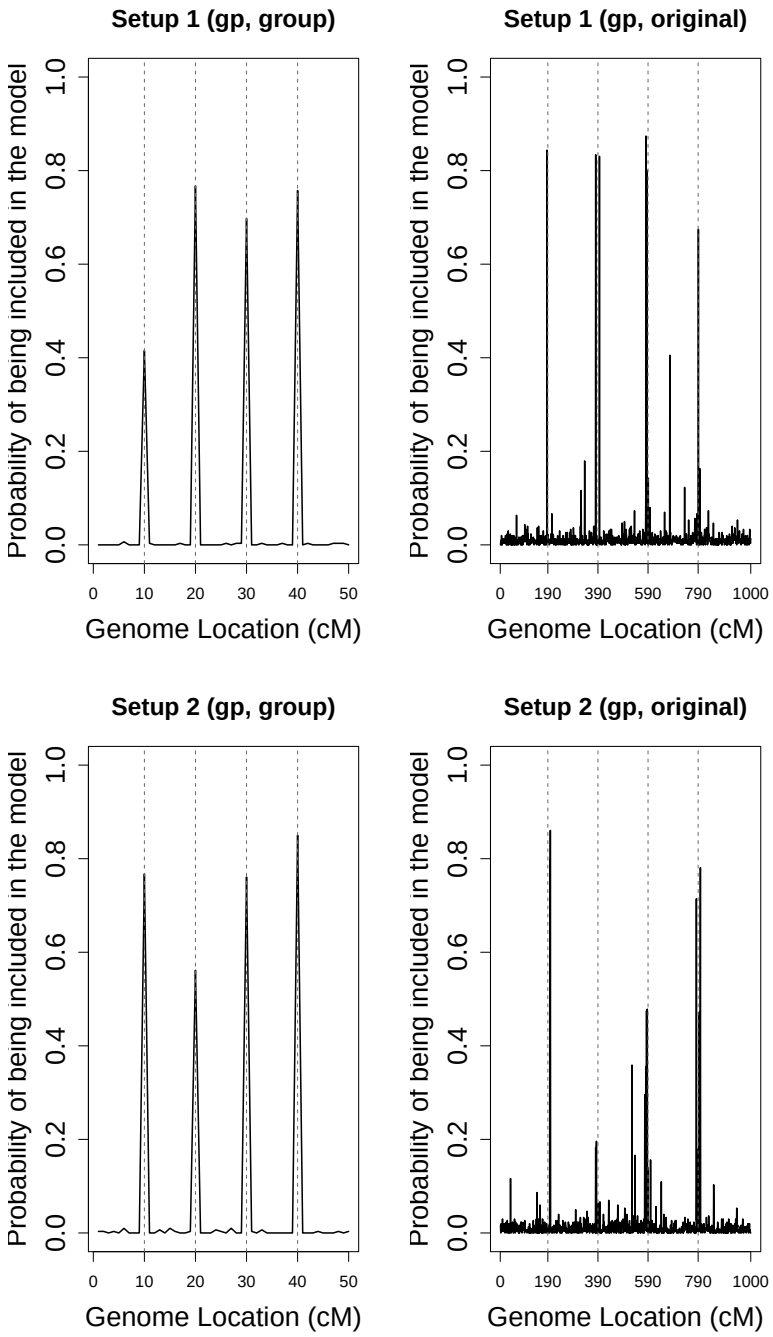

Figure 4.5: Posterior mean estimates of the latent variable γ_{gk} from the proposed GP model and the original GP model for Setups 1 and 2 with both common and rare variants ($c_{g1} = 2.50$ and $c_{g2} = 1.00$).

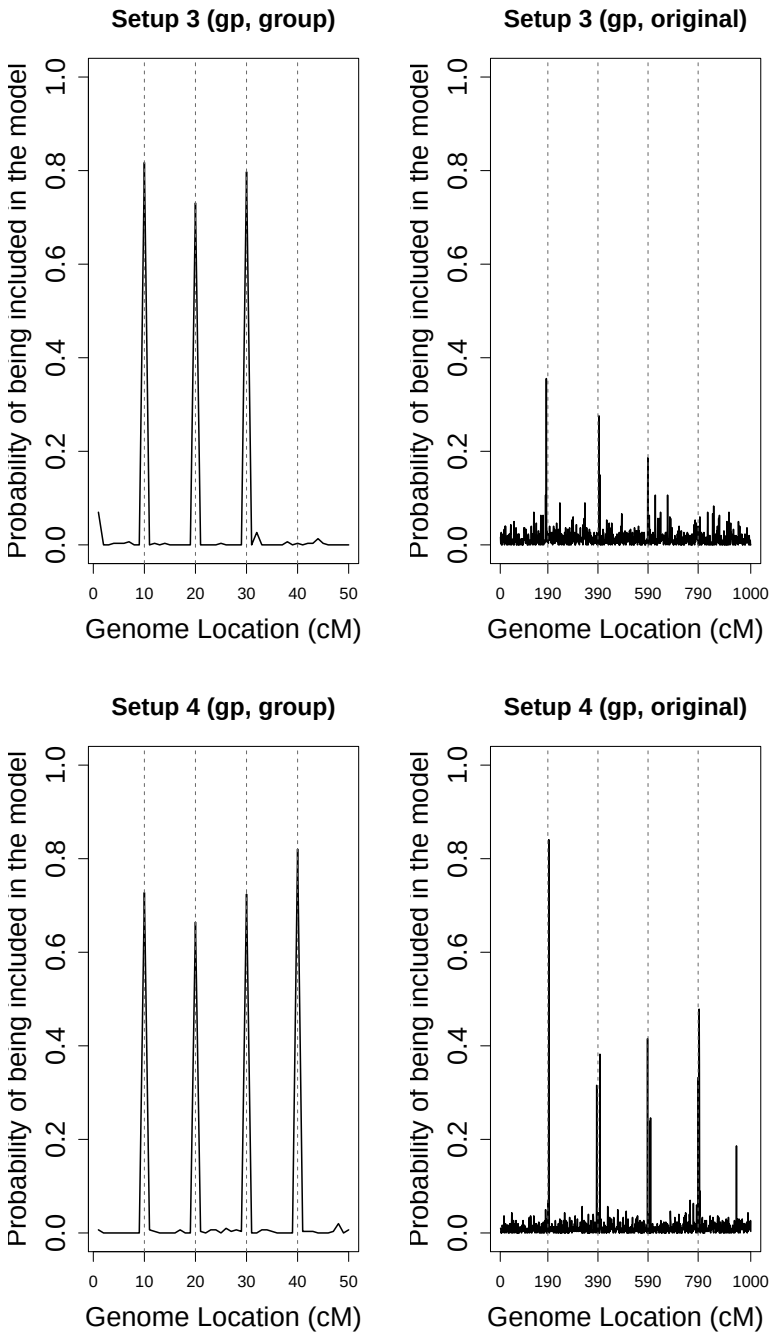

Figure 4.6: Posterior mean estimates of the latent variable γ_{gk} from the proposed GP model and the original GP model for Setups 3 and 4 with both common and rare variants ($c_{g3} = 2.00$ and $c_{g4} = 1.00$).

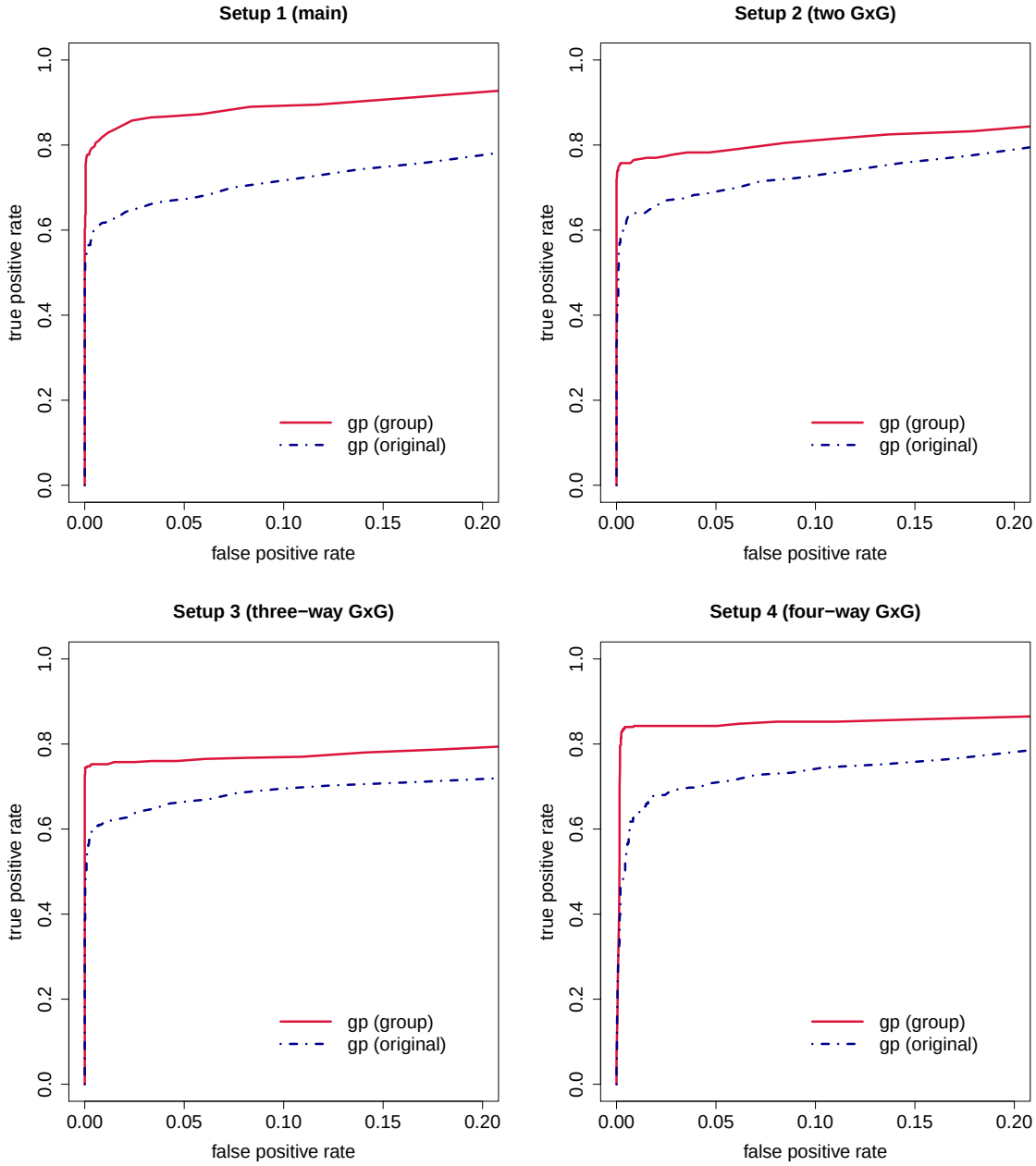

Figure 4.7: Estimated ROC curves for Setups 1,2,3 and 4 where each group has twenty common variants: solid line (red) - proposed GP model; dot-dashed line (blue) - original GP model.

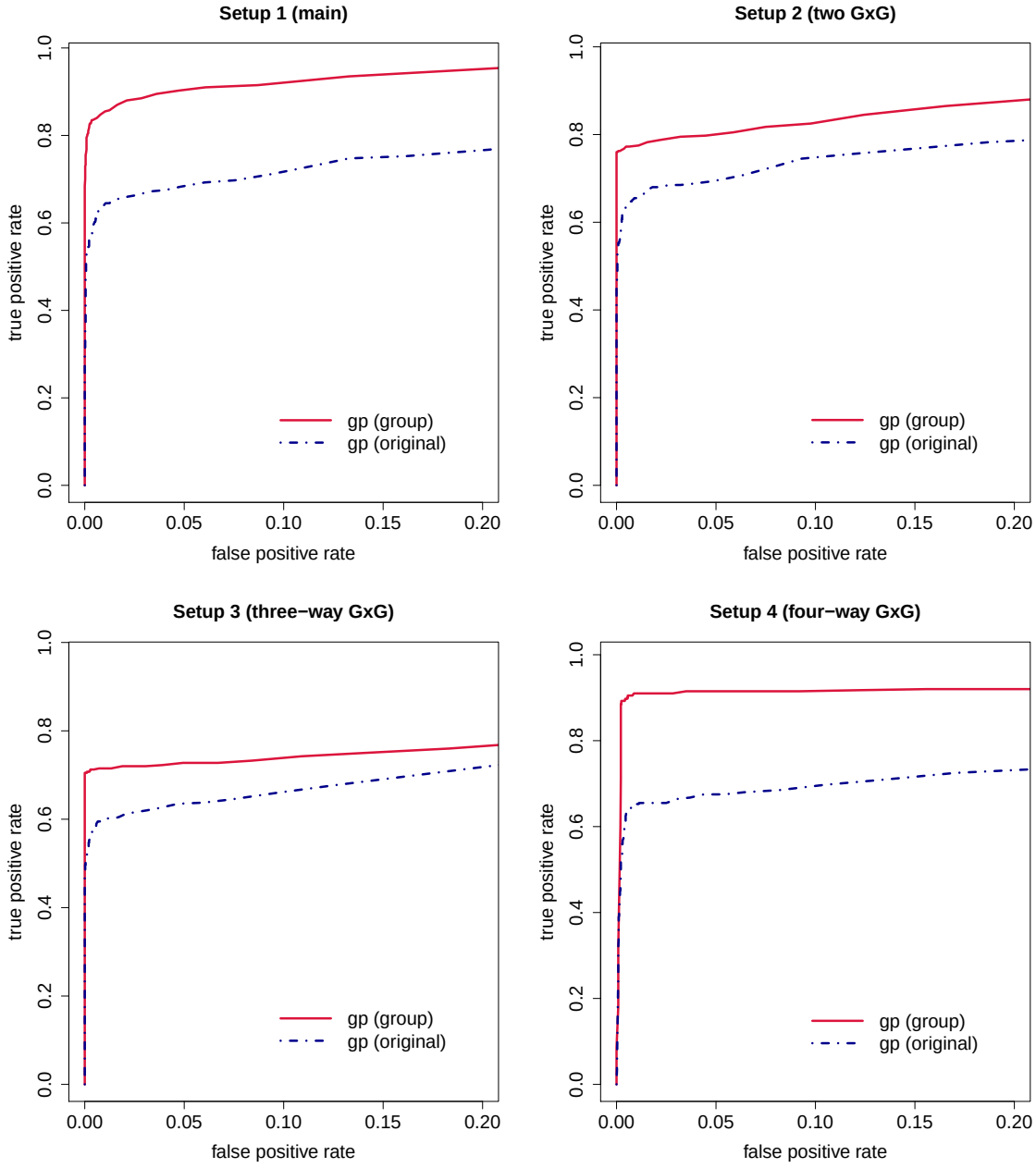

Figure 4.8: Estimated ROC curves for Setups 1,2,3 and 4 where each group has different number of common variants (10, 15, 20, 25 or 30): solid line (red) - proposed GP model; dot-dashed line (blue) - original GP model.

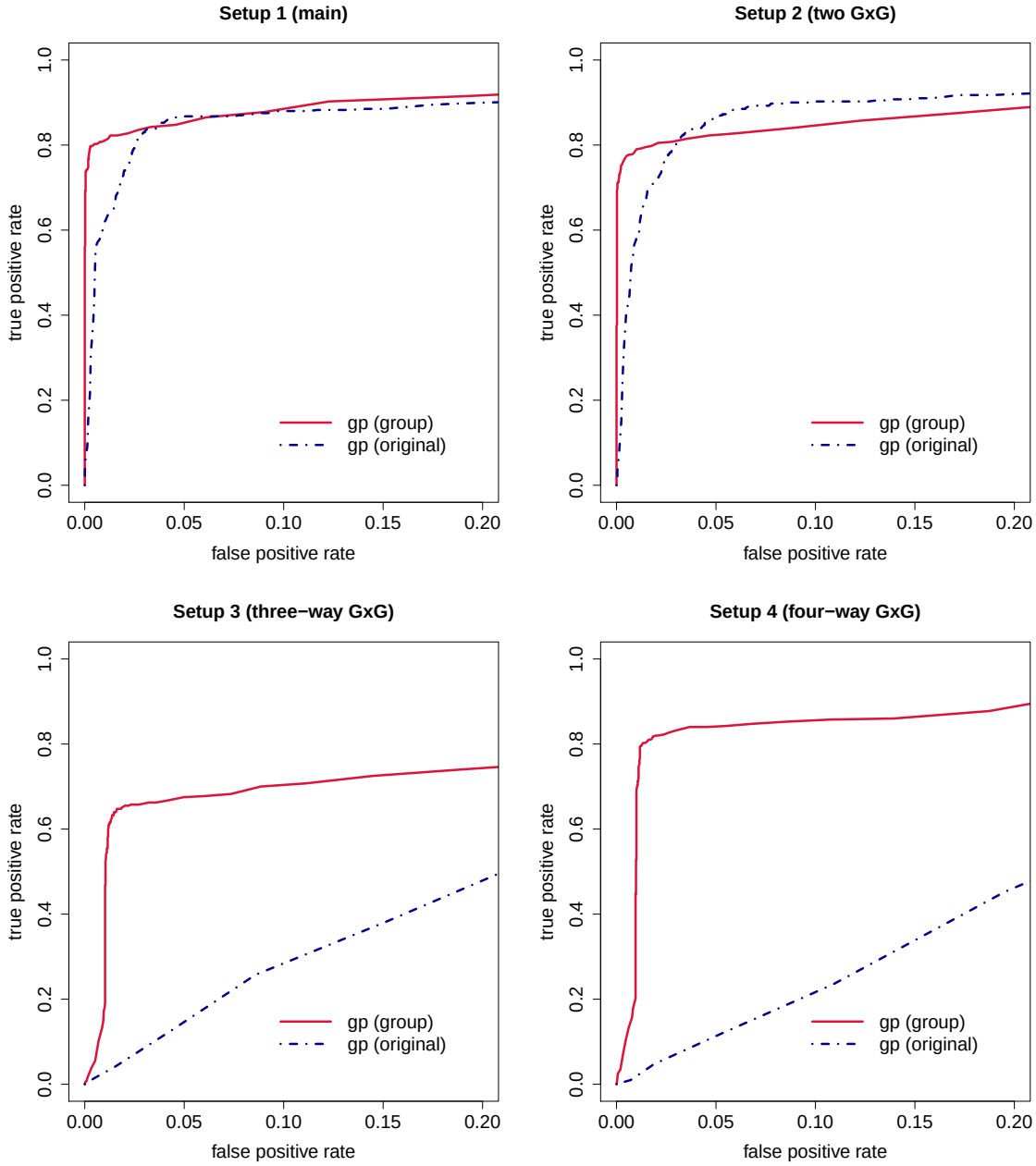

Figure 4.9: Estimated ROC curves for Setups 1,2,3 and 4 where each group has twenty rare variants: solid line (red) - proposed GP model; dot-dashed line (blue) - original GP model.

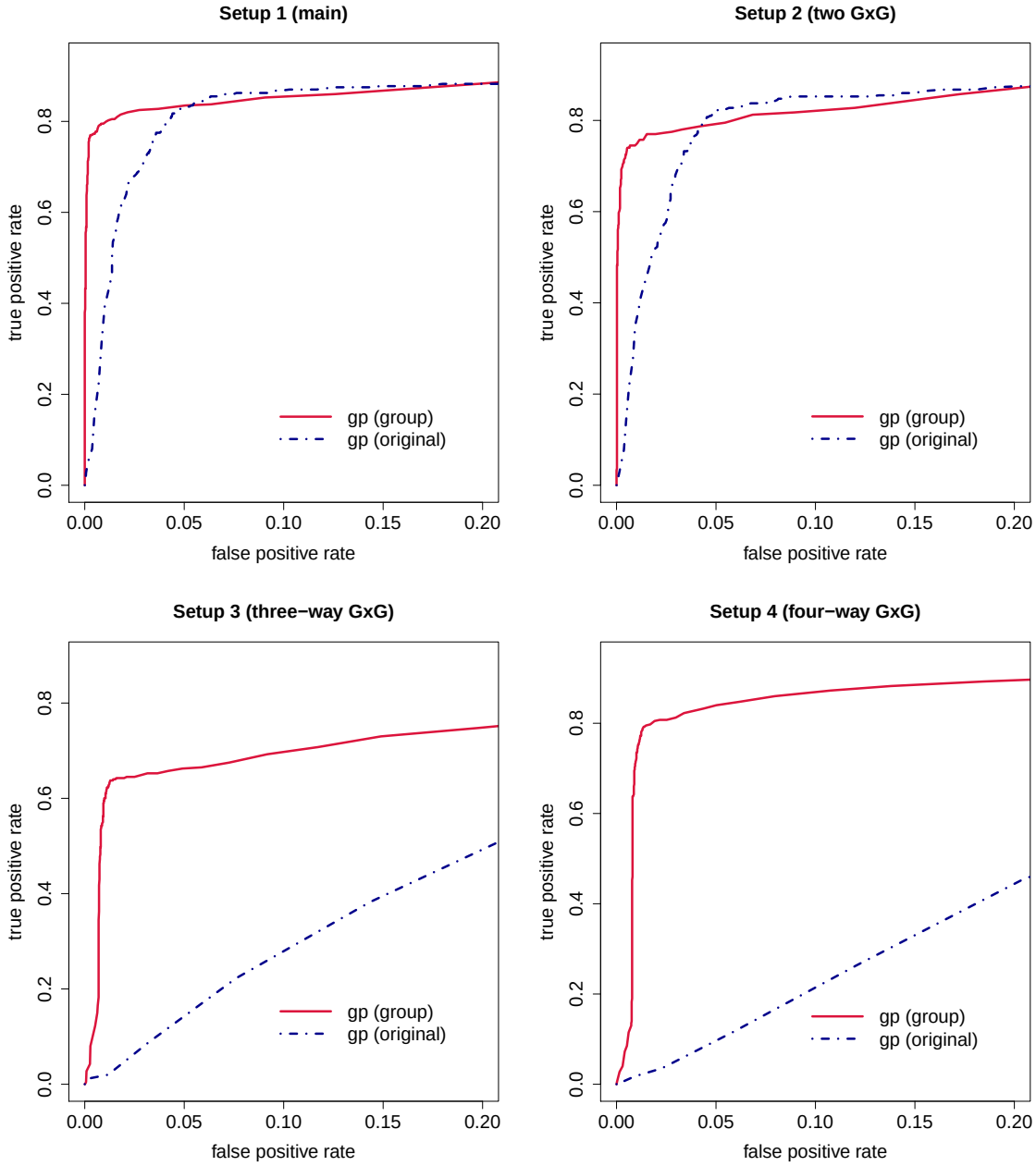

Figure 4.10: Estimated ROC curves for Setups 1,2,3 and 4 where each group has different number of rare variants (10, 15, 20, 25 or 30): solid line (red) - proposed GP model; dot-dashed line (blue) - original GP model.

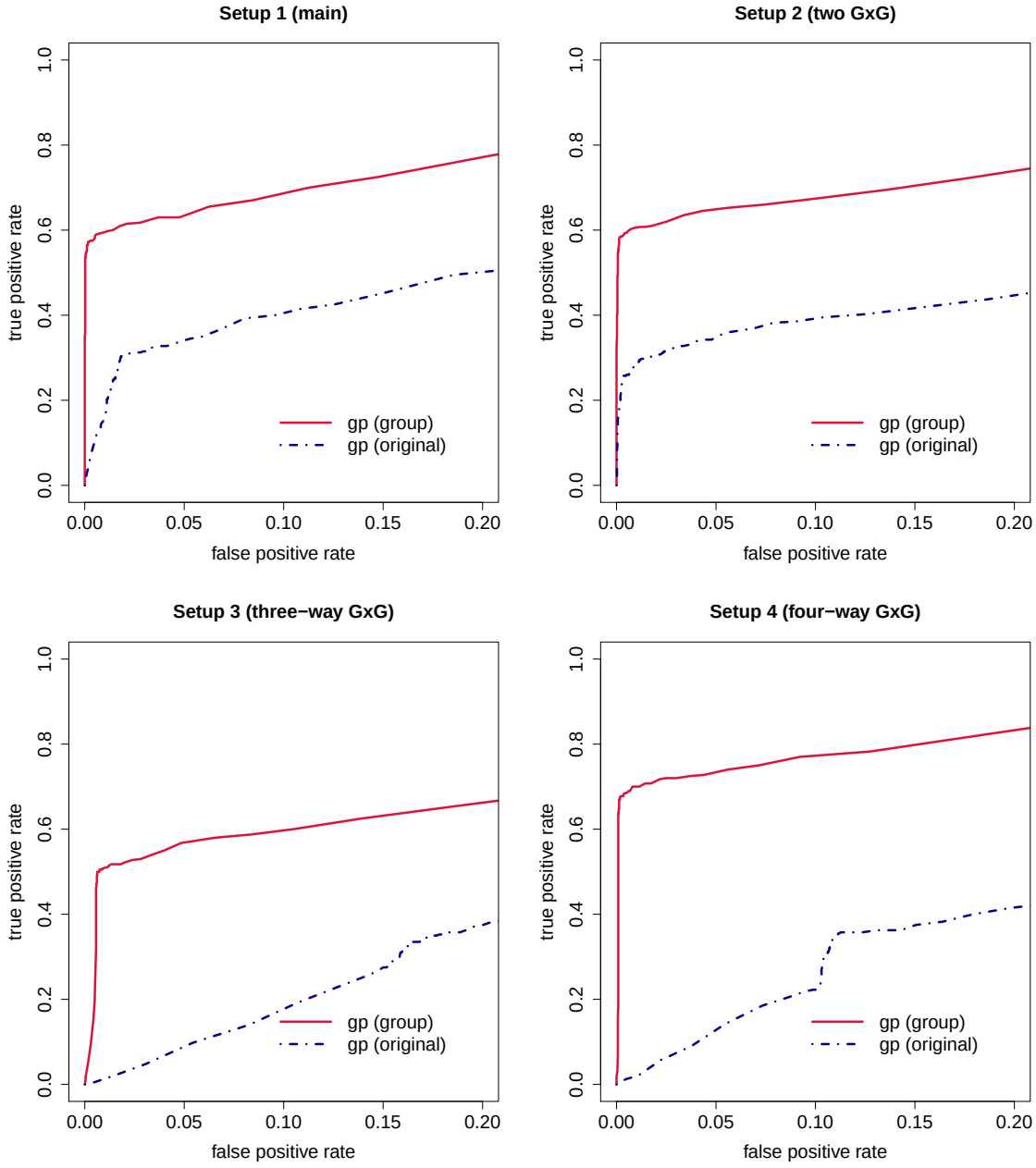

Figure 4.11: Estimated ROC curves for Setups 1,2,3 and 4 where each group has twenty common and rare variants. All causal variants are rare: solid line (red) - proposed GP model; dot-dashed line (blue) - original GP model.

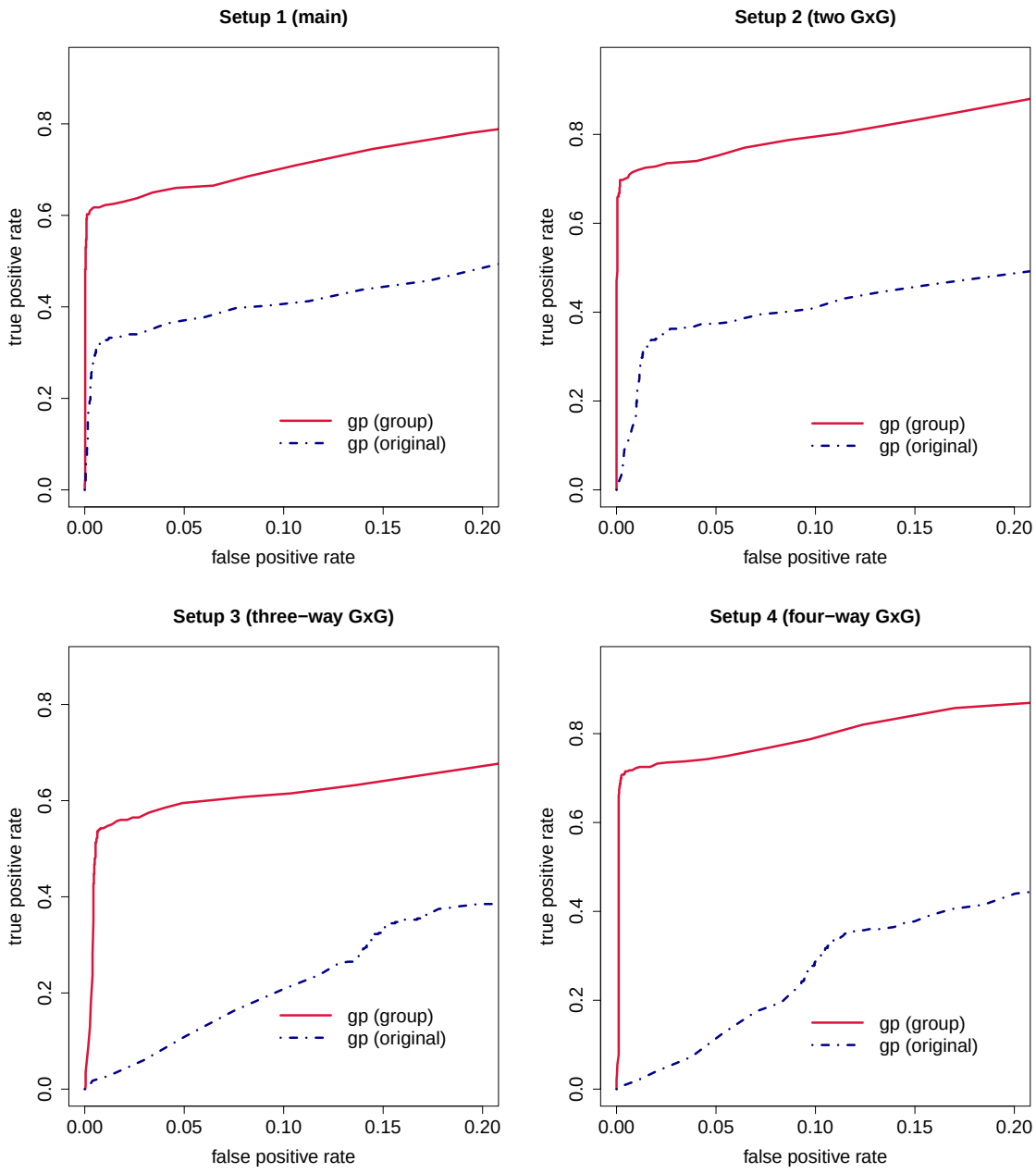

Figure 4.12: Estimated ROC curves for Setups 1,2,3 and 4 where each group has different number of common and rare variants (10, 15, 20, 25 or 30). All causal variants are rare: solid line (red) - proposed GP model; dot-dashed line (blue) - original GP model.

BIBLIOGRAPHY

Almasy, L., Dyer, T. D., Peralta, J. M., Jun, G., Fuchsberger, C., Almeida, M. A., Kent, J. W. J. r., Fowler, S., Duggirala, R., and Blangero, J. Data for genetic analysis workshop 18: Human whole genome sequence, blood pressure, and simulated phenotypes in extended pedigrees.

Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical bayesian and empirical bayes models. *Biometrika*, 94(2):443–458.

Ando, T. (2011). Predictive bayesian model selection. *American Journal of Mathematical and Management Sciences*, 31:13–38.

Banerjee, S., Yandell, B., and Yi, N. (2008). Bayesian quantitative trait loci mapping for multiple traits. *Genetics*, 179(4):2275–2289.

Barnard, J., McCulloch, R., and Meng, X. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4):1281–1312.

Berger, J. and Pericchi, L. (1996). The intrinsic bayes factor for linear models. *Bayesian Statistics*, 5:25–44.

Boik, R. (2002). Spectral models for covariance matrices. *Biometrika*, 89(1):159–182.

Broman, K., Wu, H., Sen, Ś., and Churchill, G. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19(7):889–890.

Bungartz, H.-J. and Griebel, M. (2004). *Sparse Grids. Acta Numerica*, volume 13. Cambridge University Press.

Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D. P., McCarthy, M. I., Ouwehand, W. H., Samani, N. J., et al. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447:661–678.

Carlin, B. and Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 473–484.

Chen, M. and Shao, Q. (1997). On monte carlo methods for estimating ratios of normalizing constants. *The Annals of Statistics*, 25(4):1563–1594.

Chen, M. and Shao, Q. (1998). Monte carlo methods for bayesian analysis of constrained parameter problems. *Biometrika*, 85(1):73–87.

Chen, Z. and Dunson, D. (2003). Random effects selection in linear mixed models. *Biometrics*, 59(4):762–769.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90(432):1313–1321.

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, 96(453):270–281.

Chung, W. and Zou, F. (2013). Mixed effects models for GAW18 longitudinal blood pressure data. *BMC proceedings*. Accepted for publication.

Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, 138(3):963–971.

Cockerham, C. (1954). An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics*, 39(6):859.

Crow, J., Kimura, M., et al. (1970). An introduction to population genetics theory. *An introduction to population genetics theory.*

Daniels, M. and Kass, R. (1999). Nonconjugate bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, 94(448):1254–1263.

Dellaportas, P., Forster, J., and Ntzoufras, I. (2002). On bayesian model and variable selection using MCMC. *Statistics and Computing*, 12(1):27–36.

Derkach, A., Lawless, J. F., and Sun, L. (2013). Robust and powerful tests for rare variants using fisher's method to combine evidence of association from two or more complementary tests. *Genetic Epidemiology*, 37(1):110–121.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics letters B*, 195(2):216–222.

Dupuis, J. and Siegmund, D. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics*, 151(1):373–386.

Fisher, R. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52:399–433.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian data analysis*. CRC press.

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

George, E. and McCulloch, R. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Geweke, J. et al. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Federal Reserve Bank of Minneapolis, Research Department.

Gibson, G. (2012). Rare and common variants: twenty arguments. *Nature Reviews Genetics*, 13(2):135–145.

Gilks, W., Wang, C., Yvonnet, B., and Coursaget, P. (1993). Random-effects models for longitudinal data using Gibbs sampling. *Biometrics*, pages 441–453.

Godsill, S. (2001). On the relationship between Markov chain Monte Carlo methods for model uncertainty. *Journal of Computational and Graphical Statistics*, 10(2):230–248.

Gong, Y. and Zou, F. (2012). Varying coefficient models for mapping quantitative trait loci using recombinant inbred intercrosses. *Genetics*, 190(2):475–486.

Goodnight, C. (2001). Quantitative trait loci and gene interaction: the quantitative genetics of metapopulations. *Heredity*, 84(5):587–598.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732.

Haley, C., Knott, S., et al. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 69(4):315.

Han, C. and Carlin, B. (2001). Markov chain Monte Carlo methods for computing bayes factors. *Journal of the American Statistical Association*, 96(455):1122–1132.

Hegland, M. (2007). Approximate maximum a posteriori with gaussian process priors. *Constructive Approximation*, 26(2):205–224.

Jiang, C. and Zeng, Z. (1997). Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica*, 101(1):47–58.

Kang, H., Zaitlen, N., Wade, C., Kirby, A., Heckerman, D., Daly, M., and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics*, 178(3):1709–1723.

Kang, H. M., Sul, J. H., Zaitlen, N. A., Kong, S., Freimer, N. B., Sabatti, C., Eskin, E., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature genetics*, 42:348–354.

Kao, C. and Zeng, Z. (1997). General formulas for obtaining the mles and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the em algorithm. *Biometrics*, pages 653–665.

Kao, C. and Zeng, Z. (2002). Modeling epistasis of quantitative trait loci using cockerham's model. *Genetics*, 160(3):1243–1261.

Kao, C., Zeng, Z., and Teasdale, R. (1999). Multiple interval mapping for quantitative trait loci. *Genetics*, 152(3):1203–1216.

Kass, R. and Natarajan, R. (2006). A default conjugate prior for variance components in generalized linear mixed models (comment on article by browne and draper). *Bayesian Analysis*, 1(3):535–542.

Kearsey, M., Pooni, H., et al. (1998). *The genetical analysis of quantitative traits.* Stanley Thornes (Publishers) Ltd.

Lagarias, J., Reeds, J., Wright, M., and Wright, P. (1998). Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147.

Laird, N. and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.

Lander, E. and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using rflp linkage maps. *Genetics*, 121(1):185–199.

Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., Christiani, D. C., Wurfel, M. M., and Lin, X. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *The American Journal of Human Genetics*.

Leonard, T. and Hsu, J. (1992). Bayesian inference for a covariance matrix. *The Annals of Statistics*, 20(4):1669–1696.

Levy, D., Ehret, G. B., Rice, K., Verwoert, G. C., Launer, L. J., Dehghan, A., Glazer, N. L., Morrison, A. C., Johnson, A. D., Aspelund, T., et al. (2009). Genome-wide association study of blood pressure and hypertension. *Nature genetics*, 41:677–687.

Li, B. and Leal, S. M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *The American Journal of Human Genetics*, 83(3):311–321.

Ma, C., Casella, G., and Wu, R. (2002). Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics*, 161(4):1751–1762.

MacKay, D. (1992). Bayesian interpolation. *Neural computation*, 4(3):415–447.

MacKay, D. (1998). Introduction to gaussian processes. *NATO ASI Series F Computer and Systems Sciences*, 168:133–166.

Madsen, B. E. and Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS genetics*, 5(2):e1000384.

Mather, K. (1967). Complementary and duplicate gene interactions in biometrical genetics. *Heredity*, 22(1):97.

Mather, K., Jinks, J., et al. (1977). *Introduction to biometrical genetics*. Cambridge Univ Press.

Morgenthaler, S. and Thilly, W. G. (2007). A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 615(1):28–56.

Neal, R. (1996). *Bayesian learning for neural networks*, volume 118. Springer Verlag.

Neal, R. (1997). Monte carlo implementation of gaussian process models for bayesian regression and classification. *Arxiv preprint physics/9701026*.

Neale, B. M., Rivas, M. A., Voight, B. F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S. M., Roeder, K., and Daly, M. J. (2011). Testing for an unusual distribution of rare variants. *PLoS genetics*, 7(3):e1001322.

Newton, M. and Raftery, A. (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 3–48.

Padmanabhan, S., Melander, O., Johnson, T., Di Blasio, A. M., Lee, W. K., Gentilini, D., Hastie, C. E., Menni, C., Monti, M. C., Delles, C., et al. (2010). Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. *PLoS genetics*, 6:e1001177.

Pitman, W. A., Korstanje, R., Churchill, G. A., Nicodeme, E., Albers, J. J., Cheung, M. C., Staton, M. A., Sampson, S. S., Harris, S., and Paigen, B. (2002). Quantitative trait locus mapping of genes that regulate hdl cholesterol in SM/J and NZB/B1NJ inbred mice. *Physiological genomics*, 9(2):93–102.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3):677–690.

Pourahmadi, M. (2007). Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance–correlation parameters. *Biometrika*, 94(4):1006–1013.

Pourahmadi, M. (2011). Covariance estimation: The glm and regularization perspectives. *Statistical Science*, 26(3):369–387.

Price, A. L., Kryukov, G. V., de Bakker, P. I., Purcell, S. M., Staples, J., Wei, L.-J., and Sunyaev, S. R. (2010). Pooled association tests for rare variants in exon-resequencing studies. *The American Journal of Human Genetics*, 86(6):832–838.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., Maller, J., Sklar, P., De Bakker, P., Daly, M., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3):559–575.

Rasmussen, C. (1996). Evaluation of gaussian processes and other methods for non-linear regression.

Rasmussen, C. and Williams, C. (2006). *Gaussian processes for machine learning*, volume 1. MIT press Cambridge, MA.

Reifsnyder, P., Churchill, G., and Leiter, E. (2000). Maternal environment and genotype interact to establish diabesity in mice. *Genome research*, 10(10):1568–1578.

Robert, C. P. and Titterington, D. M. (2002). Discussion of a paper by d.j. spiegelhalter et al. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):621–622.

Roberts, S., Hegland, M., and Altas, I. (2003). Approximation of a thin plate spline smoother using continuous piecewise polynomial functions. *SIAM journal on numerical analysis*, 41(1):208–234.

Satagopan, J., Yandell, B., Newton, M., and Osborn, T. (1996). A bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics*, 144(2):805–816.

Schaffner, S. F., Foo, C., Gabriel, S., Reich, D., Daly, M. J., and Altshuler, D. (2005). Calibrating a coalescent simulation of human genome sequence variation. *Genome research*, 15(11):1576–1583.

Schott, J. R. (2005). Matrix analysis for statistics.

Smith, A. and Roberts, G. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 3–23.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics*, 89(1):82–93.

Wu, W., Li, W., Tang, D., Lu, H., and Worland, A. (1999). Time-related mapping of quantitative trait loci underlying tiller number in rice. *Genetics*, 151(1):297–303.

Wu, W., Zhou, Y., Li, W., Mao, D., and Chen, Q. (2002). Mapping of quantitative trait loci based on growth models. *TAG Theoretical and Applied Genetics*, 105(6):1043–1049.

Yandell, B., Mehta, T., Banerjee, S., Shriner, D., Venkataraman, R., Moon, J., Neely, W., Wu, H., Von Smith, R., and Yi, N. (2007). R/qtlbim: Qtl with bayesian interval mapping in experimental crosses. *Bioinformatics*, 23(5):641–643.

Yang, R. and Berger, J. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, 22(3):1195–1211.

Yang, R., Tian, Q., and Xu, S. (2006). Mapping quantitative trait loci for longitudinal traits in line crosses. *Genetics*, 173(4):2339–2356.

Yap, J., Fan, J., and Wu, R. (2009). Nonparametric modeling of longitudinal covariance structure in functional mapping of quantitative trait loci. *Biometrics*, 65(4):1068–1077.

Yi, N. (2004). A unified Markov chain Monte Carlo framework for mapping multiple quantitative trait loci. *Genetics*, 167(2):967–975.

Yi, N., George, V., and Allison, D. B. (2003). Stochastic search variable selection for identifying multiple quantitative trait loci. *Genetics*, 164(3):1129–1138.

Yi, N., Liu, N., Zhi, D., and Li, J. (2011). Hierarchical generalized linear models for multiple groups of rare and common variants: jointly estimating group and individual-variant effects. *PLoS genetics*, 7(12):e1002382.

Yi, N., Shriner, D., Banerjee, S., Mehta, T., Pomp, D., and Yandell, B. (2007). An efficient bayesian model selection approach for interacting quantitative trait loci models with many effects. *Genetics*, 176(3):1865–1877.

Yi, N., Xu, S., et al. (2002). Mapping quantitative trait loci with epistatic effects. *Genetical Research*, 79(2):185–198.

Yi, N., Yandell, B., Churchill, G., Allison, D., Eisen, E., and Pomp, D. (2005). Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics*, 170(3):1333–1344.

Yi, N. and Zhi, D. (2011). Bayesian analysis of rare variants in genetic association studies. *Genetic epidemiology*, 35(1):57–69.

Zeger, S. and Karim, M. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American statistical association*, 86(413):79–86.

Zeng, Z. (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the National Academy of Sciences*, 90(23):10972–10976.

Zou, F., Huang, H., Lee, S., and Hoeschele, I. (2010). Nonparametric bayesian variable selection with applications to multiple quantitative trait loci mapping with epistasis and gene–environment interaction. *Genetics*, 186(1):385–394.