

Bayesian model comparison in genetic association analysis: linear mixed modeling and SNP set testing

XIAOQUAN WEN

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA
xwen@umich.edu

SUMMARY

We consider the problems of hypothesis testing and model comparison under a flexible Bayesian linear regression model whose formulation is closely connected with the linear mixed effect model and the parametric models for Single Nucleotide Polymorphism (SNP) set analysis in genetic association studies. We derive a class of analytic approximate Bayes factors and illustrate their connections with a variety of frequentist test statistics, including the Wald statistic and the variance component score statistic. Taking advantage of Bayesian model averaging and hierarchical modeling, we demonstrate some distinct advantages and flexibilities in the approaches utilizing the derived Bayes factors in the context of genetic association studies. We demonstrate our proposed methods using real or simulated numerical examples in applications of single SNP association testing, multi-locus fine-mapping and SNP set association testing.

Keywords: Bayes factor; Genetic association; Linear mixed model; Model comparison; SNP set analysis.

1. INTRODUCTION

In the past decades, genetic association studies have taken a prominent position in uncovering the role of genetic variants in disease etiology. Most recently, two related statistical approaches have become especially important in the analysis of genetic association data: the use of linear mixed models (LMMs) to control for confounding factors and account for polygenic effects and the application of Single Nucleotide Polymorphism (SNP) set analysis for regions of (rare) genetic variants. As demonstrated by many authors ([Kang and others, 2010](#); [Segura and others, 2012](#); [Zhou and Stephens, 2012](#); [Zhou and others, 2013](#)), LMMs effectively thwart the identification of false positive associations caused by relatedness or population structures (e.g. cryptic relatedness) in the samples while at the same time increase the power of detecting genuine genetic association signals. SNP set testing ([Madsen and Browning, 2009](#); [Wu and others, 2011](#); [Lee and others, 2012](#)) is emerging as a method of choice in detecting associations of rare genetic variants, which may be critical in explaining the phenomenon of “missing heritability”. Recent studies have also shown the necessity of jointly applying both approaches when analyzing the genetic association of rare variants to control for population stratification or using pedigree data.

Currently, the majority of the methodological work employing LMM and/or SNP set analysis in genetic association studies has focused on reporting p -values for hypothesis testing. In this paper, we discuss a Bayesian alternative to address both topics within the model comparison framework in which hypothesis testing is regarded as a special case. We first show that both problems can be naturally formulated by

a unified Bayesian parametric model, and we then derive a class of analytic approximate Bayes factors for use as our primary statistical device for model comparison. We establish the connections between the approximate Bayes factors and various commonly applied frequentist test statistics in a similar fashion, as reported by Wakefield (2009), Wen (2014), and Wen and Stephens (2014).

Despite its similarities in performance to the frequentist approaches in traditional hypothesis testing settings, the Bayesian approach exhibits great convenience and flexibility in dealing with complicated practical settings within and beyond hypothesis testing. One of the most significant advantages of the Bayesian comparison method is its acceptance of explicitly modeling various alternative scenarios (which are not necessarily nested) and the fluidity with which it combines the evidence from the data via Bayesian model averaging. Beyond single unit (i.e. either an SNP or an SNP set) association testing, we show that the Bayesian model comparison approach can be straightforwardly extended to a joint analysis of multiple association signals, especially when dealing with linkage disequilibrium (LD) among SNPs commonly present in the genetic data. We illustrate a highly efficient multi-locus fine-mapping approach that is facilitated by our results based on approximate Bayes factors.

2. MODEL AND NOTATIONS

We consider a general form of the LMM,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{G}\boldsymbol{\beta} + \mathbf{u} + \mathbf{e}, \quad \mathbf{e} \sim N(\mathbf{0}, \tau^{-1}\mathbf{I}), \quad (2.1)$$

where \mathbf{y} is an n -vector of quantitative response measurements, \mathbf{X} is an $n \times q$ matrix of covariate variables to be controlled as *fixed effects*, and their coefficients are encoded in the q -vector $\boldsymbol{\alpha}$. \mathbf{G} is an $n \times p$ matrix of covariates whose effect, represented by the p -vector $\boldsymbol{\beta}$, is of primary interest for inference. Finally, the n -vectors \mathbf{u} and \mathbf{e} represent the *random effects* and the i.i.d residual errors, respectively. In the general LMM inference framework, the random effects vector, \mathbf{u} , is assumed to be drawn from a multivariate normal (MVN) distribution, i.e.

$$\mathbf{u} \sim N(\mathbf{0}, \lambda\tau^{-1}\mathbf{K}), \quad (2.2)$$

where the $n \times n$ matrix \mathbf{K} is assumed known (while the variance component parameter λ is typically unknown). In typical genetic applications, \mathbf{G} represents the genotypes of p candidate SNPs, \mathbf{X} includes intercept term and factors like age, sex that need to be controlled for, and \mathbf{u} usually represents the random effects due to cryptic genetic relatedness or population structure. The ultimate goal is to make inference of the genetic effect $\boldsymbol{\beta}$.

We now present a Bayesian counterpart of the LMM, the likelihood part of which is identical to (2.1). From the Bayesian perspective, it is natural to regard the “random effect” assumption (2.2) as a standard MVN prior on \mathbf{u} . For *controlled* “fixed” effect coefficient $\boldsymbol{\alpha}$, we assume the MVN prior:

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \boldsymbol{\Psi}), \quad (2.3)$$

where $\boldsymbol{\Psi}$ is a diagonal matrix. When performing inference, we take the limit $\boldsymbol{\Psi}^{-1} \rightarrow \mathbf{0}$, which essentially assigns independent flat priors to each fixed effect coefficient. A flat prior might be interpreted as an assumption that the *a priori* effects of $\boldsymbol{\alpha}$ are extremely large. This assumption intuitively leads to a conservative inference on $\boldsymbol{\beta}$. However, for variables that must be controlled for, such conservative assumptions are most likely welcome.

We also assign an MVN prior for the parameter of interest, $\boldsymbol{\beta}$, such that

$$\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{W}). \quad (2.4)$$

The variance–covariance matrix \mathbf{W} fully characterizes a distinct candidate model in our model comparison framework. The choice of \mathbf{W} is context-dependent and has critical implications on the inference results. In practice, we recommend modeling the effect size on the unit-free scales of signal-noise ratios (Wen, 2014; Wen and Stephens, 2014) by assigning an MVN prior on the *standardized effect*, i.e. $\sqrt{\tau} \boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{\Phi})$, which induces a prior variance matrix on the original scale of $\boldsymbol{\beta}$ as $\mathbf{W} = \tau^{-1} \boldsymbol{\Phi}$. (Note that the prior on the random effect \mathbf{u} is formulated in the same scale.)

Finally, we assume a general joint prior distribution, $p(\lambda, \tau)$, for the variance component parameters. As we will show later, the actual functional form of $p(\lambda, \tau)$ has little impact on our asymptotic approximations of Bayes factors. To emphasize the connection with the frequentist linear mixed effect model, we will henceforth call the above Bayesian linear regression model the Bayesian linear mixed effect model (BLMM).

3. MODEL COMPARISON IN THE BLMM

We derive Bayes factors for the BLMM in order to perform Bayesian model comparisons. More specifically, we consider a space of candidate BLMMs that only differ in their specifications of \mathbf{W} . We denote H_0 as the trivial null model, in which $\boldsymbol{\beta} \equiv \mathbf{0}$ (or equivalently $\mathbf{W} = \mathbf{0}$), and we define a null-based Bayes factor for an alternative model characterized by its prior variance on $\boldsymbol{\beta}$ as

$$\text{BF}(\mathbf{W}) = \lim_{\Psi^{-1} \rightarrow 0} \frac{P(\mathbf{y} | \mathbf{X}, \mathbf{G}, \mathbf{K}, \mathbf{W})}{P(\mathbf{y} | \mathbf{X}, \mathbf{G}, \mathbf{K}, H_0)} = \lim_{\Psi^{-1} \rightarrow 0} \frac{P(\mathbf{y} | \mathbf{X}, \mathbf{G}, \mathbf{K}, \mathbf{W})}{P(\mathbf{y} | \mathbf{X}, \mathbf{G}, \mathbf{K}, \mathbf{W} = \mathbf{0})}. \quad (3.1)$$

To present our results regarding the Bayes factors, we begin by introducing several necessary additional notations. We denote $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, $\hat{\lambda}$, and $\hat{\tau}$ as the MLEs of the full LMM model (2.1) by treating $\boldsymbol{\beta}$ as a fixed effect parameter. In addition, we denote $\hat{\mathbf{V}} = \text{Var}(\hat{\boldsymbol{\beta}})$. Correspondingly, we use $\tilde{\boldsymbol{\alpha}}$, $\tilde{\tau}$, and $\tilde{\lambda}$ to represent the MLEs of the null model, where $\boldsymbol{\beta}$ is restricted to 0. Furthermore, provided that parameter λ is *known*, we note that $\hat{\boldsymbol{\beta}}$ can be analytically computed as a function of λ (Appendix A.1 of supplementary material available at *Biostatistics* online), which we denote by $\hat{\boldsymbol{\beta}}(\lambda)$. Accordingly, we use $\hat{\mathbf{V}}(\lambda, \tau)$ to represent the corresponding variance of $\hat{\boldsymbol{\beta}}(\lambda)$ (specifically, when $\lambda = \hat{\lambda}$, $\hat{\boldsymbol{\beta}}(\hat{\lambda}) = \hat{\boldsymbol{\beta}}$, and $\hat{\mathbf{V}}(\hat{\lambda}, \hat{\tau}) = \hat{\mathbf{V}}$). Finally, we consider a class of general estimators of λ , denoted by $\check{\lambda}$, for which a tuning parameter $\kappa \in [0, 1]$ is built-in. The statistical details of this class of estimators are explained in Appendix A.2 of supplementary material available at *Biostatistics* online. Most importantly, it follows that $\check{\lambda}(\kappa = 0) = \tilde{\lambda}$, and $\check{\lambda}(\kappa = 1) = \hat{\lambda}$ for the two extreme κ values. Deriving from $\check{\lambda}$, we establish a corresponding estimator of τ , denoted by $\check{\tau}(\kappa)$, which can be analytically expressed in terms of $\check{\lambda}$ (see Appendix A.2 of supplementary material available at *Biostatistics* online) and also shares a similar property such that $\check{\tau}(\kappa = 0) = \tilde{\tau}$, and $\check{\tau}(\kappa = 1) = \hat{\tau}$. Finally, we use the notations $\check{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\check{\lambda})$, $\check{\mathbf{V}} = \hat{\mathbf{V}}(\check{\lambda}, \check{\tau})$. In the case that \mathbf{W} is specified as a function of λ and/or τ , we denote $\check{\mathbf{W}} = \mathbf{W}(\check{\lambda}, \check{\tau})$. With these additional notations, we show that the desired Bayes factor can be approximated analytically. We summarize the main result in Proposition 1, whose formal proof is given in Appendix A.2 of supplementary material available at *Biostatistics* online.

PROPOSITION 1 Under the BLMM, the Bayes factor can be approximated by

$$\text{ABF}(\mathbf{W}, \kappa) = |\mathbf{I} + \check{\mathbf{V}}^{-1} \check{\mathbf{W}}|^{-1/2} \cdot \exp \left(\frac{1}{2} \check{\boldsymbol{\beta}}' \check{\mathbf{V}}^{-1} [\check{\mathbf{W}}(\mathbf{I} + \check{\mathbf{V}}^{-1} \check{\mathbf{W}})^{-1}] \check{\mathbf{V}}^{-1} \check{\boldsymbol{\beta}} \right). \quad (3.2)$$

It follows that

$$\text{BF}(\mathbf{W}) = \text{ABF}(\mathbf{W}, \kappa) \cdot \left(1 + O \left(\frac{1}{n} \right) \right) \quad \text{for any } \kappa \in [0, 1].$$

REMARK 1 The approximate Bayes factors in the BLMM share the same functional form as the ABFs discussed in Wen (2014) and enjoy some of the computational properties discussed therein. In particular, the computation of the ABF is robust to the potential collinearity presented in the data matrix \mathbf{G} . Furthermore, \mathbf{W} is allowed to be rank-deficient.

REMARK 2 For single SNP analysis, i.e. $p = 1$, both $\check{\mathbf{W}}$ and $\check{\mathbf{V}}$ degenerate to scalars (which we denote by $\check{\omega}$ and \check{v} , respectively). The expression of (3.2) is reduced to

$$\text{ABF}(\omega, \kappa) = \sqrt{\frac{\check{v}}{\check{v} + \check{\omega}}} \exp\left(\frac{1}{2} \frac{\check{\omega}}{\check{v} + \check{\omega}} \frac{\check{\beta}^2}{\check{v}}\right), \quad (3.3)$$

which has the same functional form as the ABF discussed in Wakefield (2009).

Although all suitable κ values yield the same asymptotic error bound, they have practical implications on the approximation accuracy for finite samples. Our numerical experiments (Appendix B of supplementary material available at *Biostatistics* online) indicate that with sample size around hundreds, the ABFs with $\kappa = 0$ and $\kappa = 1$ both become quite accurate.

3.1 Connection with frequentist test statistics

3.1.1 Connection with fixed effect test statistics. Consider a specific class of prior, $\mathbf{W} = c\mathbf{V}$, for which the ABF can be simplified to

$$\text{ABF}(\mathbf{W} = c\mathbf{V}, \kappa) = \left(\sqrt{\frac{1}{c+1}}\right)^p \exp\left(\frac{1}{2} \frac{c}{c+1} \check{\boldsymbol{\beta}}' \check{\mathbf{V}}^{-1} \check{\boldsymbol{\beta}}\right).$$

Consequently, the ABF becomes a monotonic transformation of the quadratic form $\check{\boldsymbol{\beta}}' \check{\mathbf{V}}^{-1} \check{\boldsymbol{\beta}}$. We note that, in the following two special cases, the quadratic form corresponds to some popular frequentist statistics to test $\boldsymbol{\beta}$ as a fixed effect. Particularly, when $\kappa = 1$, the quadratic form becomes the (multivariate) Wald statistic $\hat{\boldsymbol{\beta}}' \hat{\mathbf{V}}^{-1} \hat{\boldsymbol{\beta}}$; as κ is set to 0, it coincides with the Rao's score statistic (Appendix C.1 of supplementary material available at *Biostatistics* online).

The monotonic correspondence between the ABF and these two popular frequentist test statistics indicates that, under the prior specified, the ABF ranks candidate models (or SNP associations in single SNP analysis) exactly the same way as both the Wald statistic (for $\kappa = 1$) and the score statistic (for $\kappa = 0$). Furthermore, applying the strategy of Bayes/non-Bayes compromise (Good, 1992; Servin and Stephens, 2007) by treating the ABF as a regular test statistic, it becomes obvious that the ABF possesses a p -value identical to that of the corresponding Wald or score statistic, depending on the κ values. Wakefield (2009) first named the prior specification of the kind $\mathbf{W} = c\mathbf{V}$ as the *implicit p -value prior*. In the special case of single SNP association testing and assuming Hardy–Weinberg equilibrium, it follows that $\mathbf{V} \propto 1/nf(1-f)$ (where f represents the allele frequency of a target SNP). As a consequence, the implicit p -value prior essentially assumes a larger *a priori* effect for SNPs that are less informative (either due to a smaller sample size or minor allele frequency). Although, from the Bayesian point of view, there seems to be a lack of proper justification for such prior assumptions (Wakefield, 2009; Wen and Stephens, 2014), we often note that the overall effect of the implicit p -value prior on the final inference may be negligible in practice, especially when the sample size is large (see Section 5.1.1 for illustration).

3.1.2 *Connection with the variance component score statistic.* In SNP set analysis, it has become common practice to construct a variance component score test for the genetic effect β (Wu and others, 2011; Lee and others, 2012; Schifano and others, 2012). That is, for a set of p SNPs, the genetic effects are assumed to be *random* and follow the distribution $\beta \sim N(\mathbf{0}, \gamma \mathbf{M})$, where the matrix \mathbf{M} is pre-defined. To test $H_0 : \gamma = 0$ vs. $H_1 : \gamma \neq 0$, the score statistic is given by $T_{\text{score}} = \tilde{r}^2 (\mathbf{y} - \mathbf{X}\tilde{\alpha})' \tilde{\Sigma}^{-1} \mathbf{G} \mathbf{M} \mathbf{G}' \tilde{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\tilde{\alpha})$, where $\tilde{\Sigma} = \mathbf{I} + \lambda \tilde{\mathbf{K}}$. In the special case that the random effect \mathbf{u} is ignored (i.e. $\lambda = 0$, $\tilde{\Sigma} = \mathbf{I}$), T_{score} is reduced to the form of the original SKAT statistic (Wu and others, 2011). By re-parameterizing $\mathbf{W} = \gamma \mathbf{M}$, we show that $\text{ABF}(\kappa = 0)$ can be represented as a function of T_{score} (Appendix C.2 of supplementary material available at *Biostatistics* online). In particular, as $\gamma \rightarrow 0$, it follows that

$$\text{ABF}(\mathbf{W} = \gamma \mathbf{M}, \kappa = 0) \approx \exp\left(\frac{\gamma}{2} T_{\text{score}}\right).$$

That is, $\text{ABF}(\kappa = 0)$ becomes monotonic to the variance component score statistic. Interestingly, the condition $\gamma \rightarrow 0$ represents a *local alternative* scenario (i.e. β only slight deviates from $\mathbf{0}$), for which score tests are known to be most powerful.

4. GENETIC ASSOCIATION ANALYSIS WITH BAYES FACTORS

4.1 Bayesian hypothesis testing

Bayes factors present two major advantages in the hypothesis testing of genetic association signals: namely, the convenience of Bayesian model averaging and the flexibility of utilizing useful prior information. Before we delve into the details of the advantages of Bayesian models in hypothesis testing, it is worth noting that the practical usage of Bayesian model comparison in hypothesis testing is limited, mostly due to the difficulty involved in determining significance thresholds based on Bayes factors. Traditionally, this issue has been addressed by treating a Bayes factor as a regular test statistic and deriving its p -value accordingly (Good, 1992; Servin and Stephens, 2007). Because the null distribution of a Bayes factor is generally non-trivial, most practical implementations rely on permutation procedures. Recently, Wen (2013) proposed a robust Bayesian false discovery rate (FDR) control procedure that directly uses the Bayes factors as inputs. This procedure ensures FDR control, even under the mis-specification of alternative models, a property resembling the behavior of p -value based procedures under similar circumstances. Most importantly, this procedure is highly computationally efficient and generally does not require extensive permutations.

4.1.1 *Model averaging.* In hypothesis testing, there often exist multiple alternative scenarios, and a single parametric model (or its corresponding test statistic) can hardly accommodate all cases. For example, in SNP set testing of rare-variant genetic associations, there exist two primary types of competing approaches that target different alternative scenarios. The first type, represented by the burden tests (Madsen and Browning, 2009), collapses the genetic variants in a region to form a single characteristic genetic unit, with respect to which the association test is then performed. This approach is ideal for a particular alternative scenario in which most of the variants considered are either *consistently* deleterious or *consistently* protective. The second type of the approach, represented by the C-alpha (Neale and others, 2011) and SKAT tests, targets a complementary scenario in which the variants included in the SNP set can have bi-directional effects on the phenotype of interest. In practice, because the true alternative model is never known *a priori*, it remains a challenge to reconcile/combine the results from the two distinct approaches into the frequentist testing paradigm. Bayesian model averaging provides a principled way to naturally address this issue. Suppose that there are k possible alternative models in consideration, and

for each model i , a Bayes factor BF_i can be computed and a prior probability/weight, π_i is assigned. An overall Bayes factor then can be computed by $\overline{\text{BF}} = \sum_{i=1}^k \pi_i \text{BF}_i$, which summarizes the overall evidence from the data compared with the null model while accounting for the uncertainty of the true alternative scenario.

In the context of SNP set analysis, [Lee and others \(2012\)](#) showed that the alternative scenarios considered in the burden and SKAT tests can both be represented in the LMM framework with different specification of random effect \mathbf{W} matrix. In brief, let the column vector $\mathbf{w} = (w_1, \dots, w_p)$ denote the marginal prior effect sizes for p SNPs in a set. The burden test assumes $\mathbf{W} = \mathbf{W}_b = (\sqrt{\mathbf{w}}) (\sqrt{\mathbf{w}})'$, whereas the SKAT model assumes $\mathbf{W} = \mathbf{W}_s = \text{diag}(\mathbf{w})$. Given these results and within the framework of BLMM, we can straightforwardly average the evidence over the two competing alternative models by computing an overall Bayes factor, $\overline{\text{BF}}(\pi) = \pi \cdot \text{BF}(\mathbf{W}_b) + (1 - \pi) \cdot \text{BF}(\mathbf{W}_s)$, where the probability π denotes the relative prior frequency of the burden model. Without prior preference over the two alternatives, a natural “objective” choice is to set $\pi = 0.5$.

[Lee and others \(2012\)](#) provided an alternative interpretation by connecting the two models. They considered a class of \mathbf{W} matrices indexed by a non-negative correlation coefficient ρ : namely,

$$\mathbf{W}_\rho = \text{diag}(\sqrt{\mathbf{w}})[(1 - \rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}'] \text{diag}(\sqrt{\mathbf{w}}) = (1 - \rho)\mathbf{W}_s + \rho\mathbf{W}_b, \quad (4.1)$$

which we will refer to as the SKAT-O prior. It should be noted that the prior distribution for $\boldsymbol{\beta}$ assumed by Bayesian model averaging is essentially a normal mixture, which itself is not necessarily normal and hence differs from the SKAT-O prior. Nevertheless, the SKAT-O prior can be viewed as a normal approximation of this mixture distribution (to the first two moments).

4.1.2 Informative prior. The explicit specification of the prior distribution on $\boldsymbol{\beta}$ for alternative models is seemingly a distinct feature of Bayesian hypothesis testing. However, as we have shown, even the most commonly applied frequentist test statistics can be viewed as resulting from some implicit Bayesian priors. Therefore, it is only natural to regard the prior specification of $\boldsymbol{\beta}$ as an integrative component in alternative modeling. This fact should encourage practitioners to explicitly formulate appropriate informative priors in Bayesian hypothesis testing: if the prior does capture some essence of reality, it improves the overall statistical power; even if the prior is mis-specified, testing with Bayes factors using the procedures, such as either the Bayes/non-Bayes compromise or the robust Bayesian FDR control, only results in a reduction in power but no inflation of type I error.

For SNP set analysis, it has become common practice to pre-define some “weight” for each individual participating SNP in both the burden and SKAT types of approaches (i.e. the aforementioned \mathbf{w} vector). Most commonly, these priors are set up to prioritize genetic variants with low allele frequencies. When performing genetic association analysis, it is now becoming increasingly popular to incorporate genomic annotation and/or pathway information. In all of these examples, BLMM provides a convenient way to formally integrate the prior information into the hypothesis testing.

Finally, we note that there exist practical settings, especially in the studies of genome-wide scale, in which the information of the desired priors can be sufficiently “learned” from data facilitated by the Bayes factors. Take, for example, the problem of SNP set analysis with two competing alternatives, and consider inferring the weights of the burden and the SKAT models (π) from the data. Hypothetically, if (i) many SNP sets are investigated (in a single or multiple studies) and (ii) a sufficient amount of modest to strong signals are presented in the data, it should be intuitive that π can be accurately estimated by pooling the information across all SNP sets. More specifically, for each SNP set, we can augment a latent indicator to represent the true generative model of the observed data. Subsequently, a straightforward EM algorithm (where the complete data likelihood can be evaluated via Bayes factors) can be used to estimate π .

4.2 Bayesian variable selection in the BLMM

Beyond hypothesis testing, many practical problems in genetic association studies can be tackled using model comparison/selection techniques via Bayes factors. Here, we consider the problem of multi-locus fine-mapping analysis. In practice, the fine-mapping analysis usually focuses on relatively small genomic regions flagged by SNP association signals, with the aim of identifying multiple potential signals and narrowing down the candidate causal variants within a region while accounting for LD.

Consider a region of p candidate variants whose genetic effects are jointly modeled by the p -vector $\boldsymbol{\beta}$. Ultimately, we are interested in making an inference on the binary vector $\boldsymbol{\xi} := (\mathbf{1}(\beta_1 \neq 0), \dots, \mathbf{1}(\beta_p \neq 0))$. Under the BLMM, we assume the following spike-and-slab prior for variable selection, namely,

$$\boldsymbol{\beta} \mid \boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{W}) \text{ with } \mathbf{W} = \phi^2 \text{diag}(\boldsymbol{\xi}) \quad \text{and} \quad \Pr(\boldsymbol{\xi}) = \prod_{i=1}^p p_1^{\xi_i} (1 - p_1)^{(1-\xi_i)}, \quad (4.2)$$

where the parameter p_1 denotes the prior inclusion probability of an SNP and the parameter ϕ^2 represents the prior genetic effect size of each SNP. The posterior distribution of $\boldsymbol{\xi}$ can be computed by

$$\Pr(\boldsymbol{\xi} \mid \phi^2, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{G}) \propto \Pr(\boldsymbol{\xi}) \cdot P(\mathbf{y} \mid \boldsymbol{\xi}, \phi^2, \mathbf{X}, \mathbf{Z}, \mathbf{G}) \propto \Pr(\boldsymbol{\xi}) \cdot \text{BF}(\mathbf{W}), \quad (4.3)$$

where the Bayes factor can be further approximated by $\text{ABF}(\mathbf{W}, \kappa)$. It is then conceptually straightforward to design an MCMC algorithm to perform Bayesian variable selection. We note that, in the case of setting $\kappa = 0$, there are substantial computational savings in the proposed MCMC computation. We give the detailed description and explanation of the MCMC algorithm in Appendix D of supplementary material available at *Biostatistics* online.

5. NUMERICAL ILLUSTRATION

5.1 Application of BLMM to an *A. thaliana* data set

In this example, we apply the BLMM to study the genetic associations between the genotypes of an inbred *A. thaliana* line and the quantitative phenotype of sodium concentration in the leaves using the data described in [Baxter and others \(2010\)](#). The data set consists of 336 inbred individuals, and each individual is genotyped at 214K SNP positions genome-wide. The data set was previously analyzed by [Segura and others \(2012\)](#) under the LMM setting. We conduct an additional quantile normalization step for the original phenotype measurements to prevent the influence of potential outliers.

5.1.1 Single SNP association analysis. We first perform single SNP association tests using the approximate Bayes factors of the BLMM and compare the results with the analyses based on p -values. To specify the alternative models in the BLMM, we consider a natural exchangeable prior on the standardized effect scale, i.e. $\sqrt{\tau}\boldsymbol{\beta} \sim N(0, \phi^2)$. Unlike the implicit p -value prior, this prior does not assume a relationship between the genetic effect size and the features of a target SNP. Furthermore, instead of fixing a single ϕ value, we assume that ϕ is uniformly drawn from the set $L := \{\phi : 0.1, 0.2, 0.4, 0.8, 1.6\}$, where the various levels of ϕ values cover a range of small, modest to large potential effect sizes. The use of multiple ϕ values forms a mixture normal prior, which is helpful for describing a longer-tailed distribution of effect sizes ([Servin and Stephens, 2007](#); [Wen, 2014](#)). The range of the ϕ values is selected following the suggestion of [Stephens and Balding \(2009\)](#). We use the software package GEMMA ([Zhou and Stephens, 2012](#)) to estimate the kinship matrix, \mathbf{K} , for the random effect, and obtain the MLEs, $\hat{\beta}(\hat{\lambda})$, $\hat{\beta}(\hat{\lambda})$, along with their standard errors for all the SNPs. Applying (3.3), we then compute the approximate Bayes factors at $\kappa = 1$

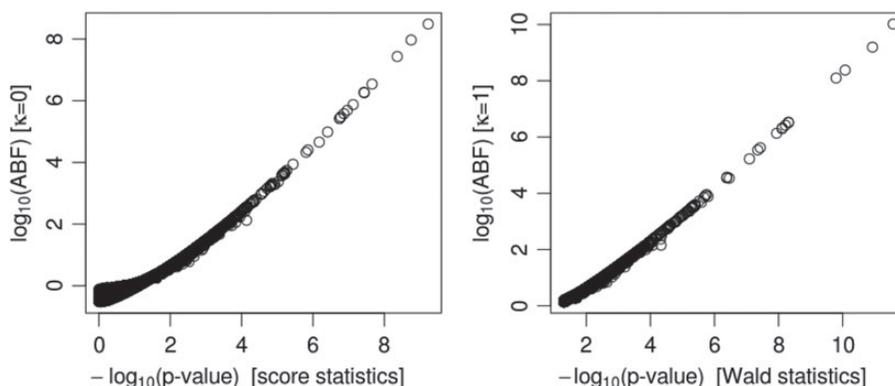


Fig. 1. Direct comparison of the ABFs and p -values on the log scale. The plot shows that the rankings of the association signals based on the Bayes factor and the p -value are largely in agreement, especially for SNPs showing modest to strong signs of association.

and $\kappa = 0$ for each ϕ_i value. Finally, we compute an overall Bayes factor by averaging over all the prior effect size models, i.e. $\text{BF} = (1/|L|) \sum_i \text{BF}(\phi_i)$.

We first investigate the ranking of the association signals by the ABFs under the natural Bayesian prior and the p -values based on the score and Wald test statistics. To this end, we compute the Spearman's rank correlation coefficient (ρ) of the $\log_{10}(\text{ABF})$ and $-\log_{10}(p\text{-value})$. The overall rank correlation (from all 214K association tests) between $-\log_{10}(p\text{-value})$ based on the score statistic and $\log_{10}[\text{ABF}(\kappa = 0)]$ is 0.817. However, we note that the majority of the discordance in ranking comes from the unlikely association signals (see Figure 1), which are generally not of interest. Focusing on the subset of 10 913 SNPs with $p < 0.05$, the rank correlation becomes nearly perfect ($\rho = 0.995$). Similarly, the $-\log_{10}(p\text{-value})$ based on the Wald statistic has an overall rank correlation of 0.821 with $\log_{10}[\text{ABF}(\kappa = 1)]$, and for the subset of 11 379 SNPs with corresponding $p < 0.05$, $\rho = 0.996$. The direct comparison between the approximate Bayes factors and corresponding p -values is shown in Figure 1.

As an illustration, we further apply the Bayesian and the frequentist FDR control procedures for the Bayes factors and p -values to determine the significance cut-offs, ignoring correlations among the tests. Ultimately, both the Benjamini–Hochberg and the Storey procedures using the score statistic p -values select 17 significant SNPs (denoted by set S_p). In comparison, the standard Bonferroni procedure selects 12 SNP (denoted by set S_b). The Bayesian FDR control procedure (i.e. the EBF procedure, described in Wen, 2013) based on $\text{ABF}(\kappa = 0)$ selects 14 significant SNPs (denoted by set S_{BF}). Importantly, we note that $S_b \subset S_{BF} \subset S_p$. The results from the $\text{ABF}(\kappa = 1)$ and Wald statistic p -values are nearly identical.

Based on this result, we conclude that, under this particular GWAS setting with a very modest sample size, there is no obvious practical difference in applying the Bayes factors and the p -values in single SNP hypothesis testing. We view this result as a numerical validation of our theoretical results discussed in Section 3.1.

5.1.2 Fine-mapping analysis. Following Segura and others (2012), we further perform a multi-locus fine-mapping analysis of a 200 kb genomic region centered around the top single SNP association signal at chr4:6392280, where 508 SNPs are included. Using the MCMC algorithm described in Section 4.2, we assign the prior inclusion probability $p_1 = \frac{1}{508}$ for each candidate SNP, which conservatively sets the prior expected number of signals in the region to 1. Conditional on a SNP having a non-zero effect (i.e. $\mathbf{1}(\beta_i) \neq 0$), we use the same normal mixture prior for the effect size β_i described in the single SNP

Table 1. Top 5 associated SNPs according to their marginal inclusion probabilities in the Bayesian fine-mapping analysis

SNP	Posterior inclusion prob.	Marginal $\log_{10}(\text{ABF})$
chr4:6414956	0.795	4.98
chr4:6392280	0.741	7.96
chr4:6420777	0.528	6.03
chr4:6455695	0.451	5.30
chr4:6391204	0.405	7.92

The last column shows the values of $\log_{10} \text{ABF}(\kappa = 0)$ from the single SNP association testing. Only SNP chr4:6392280 and SNP chr4:6391204 show a very modest LD, whereas all of the other pairs of SNPs are very weak in LD.

association analysis. We obtain the posterior samples from 300 000 MCMC repeats after 150 000 burn-in steps, and the convergence of the MCMC algorithm is diagnosed using the procedure described in [Brooks and others \(2003\)](#).

The analysis based on the posterior samples clearly indicates that there are multiple independent association signals residing in this relatively small genomic region. There is zero probability mass on those posterior models containing fewer than 3 SNPs; the probabilities for the posterior models having 3, 4, 5, and 6 independent signals are 0.175, 0.452, 0.350, and 0.023, respectively. Inspecting individual SNPs, we summarize the top 5 associated SNPs according to their posterior inclusion probabilities in Table 1. The correlations among the top 5 SNPs are very modest. Thus, far, our result has been largely consistent with what is reported in [Segura and others \(2012\)](#), in which a stepwise variable selection scheme with a Bayesian Information Criterion-like model selection criteria is employed. Nevertheless, we note a great deal of uncertainty within the individual models from our analysis. The details of the top 10 models ranked by their posterior probabilities are shown in Table 2. The *maximum a posteriori* model only has a probability of 0.05, and all of the top models have similar complexities and very comparable likelihoods. In addition, we find that 61% of the posterior models contain both of the top two SNPs, and 32% of the posterior models contain a combination of the top 3 SNPs. One may naturally suspect that the uncertainty in relative large models (i.e. with more SNPs included) is partially due to the stringent p_1 prior. To this end, we modify the prior distribution to $\log_{10} p_1 \sim \text{Uniform}[-2.71, -1.40]$ (the two end points correspond to p_1 equaling $\frac{1}{508}$ and $\frac{20}{508}$, respectively), but the results do not qualitatively change. Biologically, it might be the case that the true causal variants are not directly genotyped and the observed signals are only partially correlated with them. It is then worth following up with dense genotyping experiments or genotype imputations. Statistically, it seems evident that, in this particular case, reporting a single “best” model from the variable selection procedure yields an over-simplified picture and can be misleading for the follow-up analysis.

5.2 Simulation study of SNP set analysis

In this section, we perform simulation studies to illustrate the effectiveness of the proposed Bayesian model comparison approach in SNP set analysis. In each simulated data set, we generate 5000 phenotype-SNP set pairs that mimics the data structure from genome-wide investigation of expression quantitative trait loci. We randomly select 3500 SNP sets and simulate their phenotypes from a null model. For the remaining SNP sets, we use two types of alternative models described in [Lee and others \(2012\)](#) to generate their phenotypes: one model assumes consistent directional effects of rare variants, whereas the other allows inconsistent directional effects. We use π to denote the relative frequency of the sign-consistent models in all the alternative models, and vary this parameter in different simulation sets. We give a detailed account of the simulation schemes in Appendix E.1 of supplementary material available at *Biostatistics* online.

Table 2. *Top 10 posterior models in the Bayesian fine-mapping analysis*

Model	Posterior Prob.	$\log_{10}(\text{ABF})$
chr4:6392280 + chr4:6394774 + chr4:6414956 + chr4:6421034	0.052	19.55
chr4:6392280 + chr4:6414956 + chr4:6420777 + chr4:6455695	0.039	18.89
chr4:6391204 + chr4:6392280 + chr4:6414956 + chr4:6420777	0.032	18.47
chr4:6380552 + chr4:6391204 + chr4:6414956 + chr4:6455695	0.028	18.67
chr4:6391204 + chr4:6414956 + chr4:6420777 + chr4:6455695	0.026	18.72
chr4:6392280 + chr4:6414956 + chr4:6418442 + chr4:6420777	0.024	18.68
chr4:6391286 + chr4:6392280 + chr4:6414956 + chr4:6420777	0.022	18.35
chr4:6392280 + chr4:6414956 + chr4:6418442	0.018	16.75
chr4:6380552 + chr4:6391204 + chr4:6392280 + chr4:6420777	0.017	18.19
chr4:6380552 + chr4:6392280 + chr4:6394774 + chr4:6414956 + chr4:6421034	0.016	21.25

The models are ranked according to their posterior probabilities (second column). The last column shows the values of \log_{10} ABF ($\kappa = 0$) of the corresponding models. Our prior specification encourages sparse models: complicated models with more predictors are penalized more severely by the prior inclusion probability. The most important feature of these results is that there is not a unique simple model that is clearly better than the others.

Table 3. *Realized FDR and power in simulation studies of SNP set analysis*

Setting (π)	FDR			Power		
	SKAT-O	Bayesian-D	Bayesian-E	SKAT-O	Bayesian-D	Bayesian-E
0.20	0.024	0.027	0.023	0.768	0.741	0.821
0.40	0.046	0.030	0.028	0.791	0.773	0.828
0.50	0.051	0.045	0.041	0.836	0.825	0.869
0.60	0.049	0.046	0.045	0.909	0.908	0.919
0.80	0.050	0.049	0.048	0.933	0.943	0.948

The first column (setting) indicates the percentage of the SNP sets with sign-consistent effects among all of the non-null SNP sets in the simulated data. For the SKAT-O procedure, the resulting p -values are further processed by the Storey procedure for FDR controls. “Bayesian-D” indicates the Bayesian testing procedure with the default uniform weights. “Bayesian-E” indicates the Bayesian procedure that estimates π from the data. The FDR control for the Bayes factors is performed using the EBF procedure described in [Wen \(2013\)](#).

We analyzed the simulated data sets using the proposed Bayesian model comparison approach and the SKAT-O method implemented in the R package SKAT (version 0.95) to examine their controls of FDR and powers. For both approaches, we again follow the previous work ([Wu and others, 2011](#); [Lee and others, 2012](#)) and assign the marginal weight for each SNP as a function of their allele frequencies. In Bayesian analysis, we apply two strategies in choosing the prior weights for Bayes factor computation. The first strategy assumes an “objective” uniform prior setting $\pi = 0.5$, and the second strategy *estimates* π and the distribution of genetic effect sizes from the data by pooling information across all phenotype-SNP set pairs using a hierarchical model. The details of the analysis procedure are provided in Appendix E.2 of supplementary material available at *Biostatistics* online.

We summarize the simulation results in Table 3. The FDRs in all the methods are well controlled. The performance of the Bayesian procedure with the default uniform prior weights is very similar to that of the SKAT-O, and the Bayesian procedure based on informative priors achieves the best power in all settings defined by different true π values. These results are well expected because the Bayesian method with estimated weights has the unique advantage of effectively borrowing information across genes through the use of Bayes factors and hierarchical modeling. In addition, we want to emphasize that all of the Bayesian models assumed in the analysis are indeed very “wrong” comparing to the true data-generating model;

nevertheless, the robust Bayesian FDR control procedure using Bayes factors ensures the targeted FDR level.

Going beyond SNP set testing targeting rare-variant associations, in Appendix F of supplementary material available at *Biostatistics* online, we further demonstrate that our Bayesian model averaging framework can be conveniently extended to integrate models for detecting *common variant associations* into SNP set testing. We envision that this approach will have a profound impact in studies of expression trait quantitative loci at genome-wide scale.

6. DISCUSSION

In this paper, we have presented a unified Bayesian framework to perform model comparisons in the contexts of an LMM and SNP set analysis. Although our statistical results are presented exclusively for the quantitative response variables, it is possible to extend them to the generalized LMMs context to incorporate binary outcomes and count data using a quadratic approximation of the corresponding log-likelihood functions.

Primarily based on the results of the approximate Bayes factors, we have demonstrated an efficient Bayesian sparse variable selection algorithm to perform multi-locus association analysis using the BLMM. Recently, [Zhou and others \(2013\)](#) also proposed an elegant Bayesian solution for multiple SNP association analysis under the LMM model on the genome-wide scale. It should be noted that their method also has a primary focus on estimating the heritability, whereas our method is designed for fine-mapping analysis. In addition, by treating SNP sets as selection units, our approach can be straightforwardly extended to the identification of multiple associated genes/SNP sets, which may be attractive for biological pathway analysis. Previous studies ([Guan and Stephens, 2011](#); [Wen, 2014](#)) have shown that Bayesian methods generally hold advantages over penalized regression approaches in variable selection problems with correlated covariates (e.g. SNPs in LD) and/or non-i.i.d. residual error structures. More importantly, as we have demonstrated, there can be great uncertainty regarding any single “best fitting” model. As a practical consequence, reporting a single “best” model but ignoring appropriate uncertainty assessments could hinder follow-up scientific investigation.

Finally, we want to note that Bayesian model comparison approaches have been successfully assessed in other areas of genetic association studies, e.g. meta-analysis ([Wen and Stephens, 2014](#)), association mapping of multiple-traits, and detecting gene–environment interactions ([Flutre and others, 2013](#); [Wen and Stephens, 2014](#)). Our results can be conveniently integrated into those existing tools, and their usages can be naturally extended to incorporate LMM and SNP set analysis.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at <http://biostatistics.oxfordjournals.org>.

ACKNOWLEDGMENTS

We thank Seunggeun Lee and Xiang Zhou for helpful discussions. *Conflict of Interest*: None declared.

FUNDING

This work was supported by NIH grants R01-MH101825 and R01-HG007022.

REFERENCES

- BAXTER, I., BRAZELTON, J. N., YU, D., HUANG, Y. S., LAHNER, B., YAKUBOVA, E., LI, Y., BERGELSON, J., BOREVITZ, J. O., NORDBERG, M., VITEK, O. AND SALT, D. E. (2010). A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter *athtk1*; 1. *PLoS Genetics* **6**(11), e1001193.

- BROOKS, S. P., GIUDICI, P. AND PHILIPPE, A. (2003). Nonparametric convergence assessment for MCMC model selection. *Journal of Computational and Graphical Statistics* **12**(1), 1–22.
- FLUTRE, T., WEN, X., PRITCHARD, J. AND STEPHENS, M. (2013). A statistical framework for joint eqtl analysis in multiple tissues. *PLoS Genetics* **9**(5), e1003486.
- GOOD, I. J. (1992). The Bayes/non-Bayes compromise: a brief review. *Journal of the American Statistical Association* **87**(419), 597–606.
- GUAN, Y. AND STEPHENS, M. (2011). Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics* **5**(3), 1780–1815.
- KANG, H. M., SUL, J. H., SERVICE, S. K., ZAITLEN, N. A., KONG, S. Y., FREIMER, N. B., SABATTI, C. AND ESKIN, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* **42**(4), 348–354.
- LEE, S., WU, M. C. AND LIN, X. (2012). Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13**(4), 762–775.
- 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.
- NEALE, B. M., RIVAS, M. A. AND VOIGHT, B. F. (2011). Testing for an unusual distribution of rare variants. *PLoS Genetics* **7**(3), e1001322.
- SCHIFANO, E. D., EPSTEIN, M. P., BIELAK, L. F., JHUN, M. A., KARDIA, S. L., PEYSER, P. A. AND LIN, X. (2012). SNP set association analysis for familial data. *Genetic Epidemiology* **36**(8), 797–810.
- SEGURA, V., VILHJÁLMSSON, B. J., PLATT, A., KORTE, A., SEREN, Ü., LONG, Q. AND NORDBORG, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44**(7), 825–830.
- SERVIN, B. AND STEPHENS, M. (2007). Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genetics* **3**(7), e114.
- STEPHENS, M. AND BALDING, D. J. (2009). Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics* **10**(10), 681–690.
- WAKEFIELD, J. (2009). Bayes factors for genome-wide association studies: comparison with p-values. *Genetic Epidemiology* **33**(1), 79–86.
- WEN, X. (2013). Robust Bayesian FDR control with Bayes factors. arXiv preprint arXiv:1311.3981.
- WEN, X. (2014). Bayesian model selection in complex linear systems, as illustrated in genetic association studies. *Biometrics* **70**(1), 73–83.
- WEN, X. AND STEPHENS, M. (2014). Bayesian methods for genetic association analysis with heterogeneous subgroups: from meta-analyses to gene-environment interactions. *Annals of Applied Statistics* **8**(1), 176–203.
- 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.
- ZHOU, X., CARBONETTO, P. AND STEPHENS, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genetics* **9**(2), e1003264.
- ZHOU, X. AND STEPHENS, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**(7), 821–824.