Bayes Factor for Linear Mixed Model in Genetic Association Studies

Yongtao Guan and Daniel Levy
National Heart, Lung, and Blood Institute

May 24, 2024

Abstract

Motivation: Bayes factor has advantages over p-value as test statistics for association, particularly when comparing multiple alternative models. A software package to compute Bayes factor for linear mixed model is lacking.

Results: We transformed the standard linear mixed model as Bayesian linear regression, substituting the random effect by fixed effects with eigenvectors as covariates whose prior effect sizes are proportional to their corresponding eigenvalues. Using conjugate normal inverse gamma priors on regression parameters, Bayes factors can be computed in a closed form. We then showed that the transformed Bayesian linear regression produced identical estimates to those of the best linear unbiased prediction (BLUP), providing a new derivation to a known connection between BLUP and Bayesian estimates.

Availability and implementation: Methods described in this note are implemented in the software IDUL as two new functionalities: computing Bayes factors and residuals for the linear mixed model. IDUL and its source code are freely available at `https://github.com/haplotype/idul`.

1 Introduction

The linear mixed model (LMM) has emerged as an effective tool in controlling for population structure and relatedness in a study sample to reduce false positives in genetic association studies (Kang et al., 2008; Zhou and Stephens, 2012; Jiang et al.,

2019). Consider a standard LMM in genetic association literature

$$\begin{aligned} \mathbf{y} &= \mathbf{W}\mathbf{a} + \mathbf{x}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e} \\ \mathbf{u} &\sim \mathrm{MVN}_n(0, \tau^{-1}\, \eta\mathbf{K}) \\ \mathbf{e} &\sim \mathrm{MVN}_n(0, \tau^{-1}\, \mathbf{I}_n), \end{aligned} \tag{1}$$

where $\mathbf{W}$ contains conventional covariates such as age and sex, including a column of 1; $\mathbf{x}$ contains genetic variant(s) to be tested for association; and $\mathbf{u}$ is the random effect with $\mathbf{Z}$ as its loading matrix and $\mathbf{K}$, whose (i, j)-th entry is twice the kinship between samples i and j, as its covariance matrix (both $\mathbf{Z}$ and $\mathbf{K}$ are known). MVN_n denotes an n-dimensional multivariate normal distribution, $\mathbf{I}_n$ is n-dimensional identity matrix. In genetic association studies, the random effect $\mathbf{Z}\mathbf{u}$ is a nuisance term that absorbs part of the phenotype $\mathbf{y}$ that is attributable to population stratification and relatedness. Here η is a hyperparameter, measuring the ratio between two dispersion terms (random effect $\mathbf{u}$ and random noise $\mathbf{e}$).

In classical literature, η is known (Henderson, 1975). In GWAS applications, however, η is an important parameter to be estimated, either using maximimum likelihood (ML) or restriced maximum likelihood (REML) approaches (Zhou and Stephens, 2012; Guan and Levy, 2024a). Conditioning on η, estimates of $\hat{\mathbf{a}}$ and $\hat{\beta}$ can be obtained and either the likelihood ratio test or Wald test can be used to compute a p-value to test against the following null model

$$\begin{aligned} \mathbf{y} &= \mathbf{W}\mathbf{a} + \mathbf{Z}\mathbf{u} + \mathbf{e} \\ \mathbf{u} &\sim \mathrm{MVN}_n(0, \tau^{-1}\, \eta\mathbf{K}) \\ \mathbf{e} &\sim \mathrm{MVN}_n(0, \tau^{-1}\, \mathbf{I}_n), \end{aligned} \tag{2}$$

where the hyperparameter η was estimated from the alternative model (1). IDUL algorithm is an efficient and asymptotically exact algorithm to estimate η (Guan and Levy, 2024a), and IDUL software can compute likelihood ratio test p-value using ML estimate of η or Wald test p-value using REML estimate of η.

Our main objective here is to compute Bayes factors for the linear mixed model. In order to do that, we transform the linear mixed model to a Bayesian linear regression with a specific prior. Taking an empirical Bayesian approach to the hyperparameter, plugging in ML estimates of η under the alternative model, and using normal inverse gamma priors on regression parameters, we can compute Bayes factors analytically and efficiently. The transformed Bayesian linear regression model can also be used for prediction, and we show that the Bayesian prediction is identical to that of best linear unbiased prediction (BLUP) (Robinson, 1991) .

2 Results

2.1 Bayesian linear regression

Define $\mathbf{G} = \mathbf{ZKZ}^t$ as genetic relatedness matrix. Let $\mathbf{Q}$ and $\mathbf{D}$ be eigen decomposition such that $\mathbf{G} = \mathbf{QDQ}^t$, where $\mathbf{D} = diag(d_1, \ldots, d_n)$ with $d_1 \geq d_2 \geq \cdots \geq d_n$ and $\mathbf{QQ}^t = \mathbf{I}_n$. Model (2) can be rewritten as

$$\begin{aligned} \mathbf{y} &= \mathbf{Wa} + \mathbf{Q}\gamma + \mathbf{e} \\ \gamma &\sim MVN_n(0, \tau^{-1}\eta\mathbf{D}) \\ \mathbf{e} &\sim \mathrm{MVN}_n(0, \tau^{-1}\,\mathbf{I}_n), \end{aligned} \tag{3}$$

where the prior on γ has been specified. Model (3) is equivalent to model (2). To see this, $E(\mathbf{Zuu}^t\mathbf{Z}^t) = \eta\mathbf{ZKZ}^t = \eta\mathbf{G} = \eta\mathbf{QDQ}^t = E(\mathbf{Q}\gamma\gamma^t\mathbf{Q}^t)$. Similarly, Model (1) can be rewritten as

$$\begin{aligned} \mathbf{y} &= \mathbf{Wa} + \mathbf{x}\beta + \mathbf{Q}\gamma + \mathbf{e} \\ \gamma &\sim MVN_n(0, \tau^{-1}\eta\mathbf{D}) \\ \mathbf{e} &\sim \mathrm{MVN}_n(0, \tau^{-1}\,\mathbf{I}_n). \end{aligned} \tag{4}$$

Note the priors on γ are the same between the null and the alternative models.

2.2 Priors and Bayes factors

Let $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$. By specifying a Gamma prior on τ, we have the standard normal-inverse-gamma on a linear model (c.f Servin and Stephens, 2007; Zhou and Guan, 2018).

$$\begin{aligned} \mathbf{a} &\sim MVN(0, \tau^{-1}V_0) \\ \beta &\sim N(0, \tau^{-1}\sigma^2) \\ \tau &\sim \Gamma(\kappa_1/2, \kappa_2/2) \end{aligned} \tag{5}$$

For model (2), after integrating out $\mathbf{a}, \gamma$ and τ and letting $\kappa_1, \kappa_2 \to 0$, we have

$$\begin{aligned} p(\mathbf{y}|M_0) = &\frac{(2\pi)^{-n/2}\Gamma(n/2)}{|\mathbf{X}^t\mathbf{X} + V_{\mathbf{X}}^{-1}|^{1/2} \cdot |V_{\mathbf{X}}|^{1/2}} \\ &\cdot \left(\frac{\mathbf{y}^t\mathbf{y} - \mathbf{y}^t\mathbf{X}\left(\mathbf{X}^t\mathbf{X} + V_{\mathbf{X}}^{-1}\right)^{-1}\mathbf{X}^t\mathbf{y}}{2} \right)^{-n/2} \end{aligned} \tag{6}$$

where $V_{\mathbf{X}} = \mathrm{diag}(\eta\mathbf{D}, V_0)$. Similarly, for model (1), we have

$$p(\mathbf{y}|M_1) = \frac{(2\pi)^{-n/2}\Gamma(n/2)}{|(\mathbf{X},\mathbf{x})^t(\mathbf{X},\mathbf{x}) + V_{(\mathbf{X},\mathbf{x})}^{-1}|^{1/2} \cdot |V_{(\mathbf{X},\mathbf{x})}|^{1/2}} \cdot \left(\frac{\mathbf{y}^t\mathbf{y} - \mathbf{y}^t(\mathbf{X},\mathbf{x}) \left((\mathbf{X},\mathbf{x})^t(\mathbf{X},\mathbf{x}) + V_{(\mathbf{X},\mathbf{x})}^{-1} \right)^{-1} (\mathbf{X},\mathbf{x})^t\mathbf{y}}{2} \right)^{-n/2}, \quad (7)$$

where $V_{(\mathbf{X},\mathbf{x})} = \mathrm{diag}(\eta\mathbf{D}, V_0, \sigma^2)$. The Bayes factor $p(\mathbf{y}|M_1)/p(\mathbf{y}|M_0)$, after letting $V_0 \to \mathbf{0}$, is

$$BF(\eta,\sigma) = \frac{|\mathbf{X}^t\mathbf{X} + V_{\mathbf{X}}^{-1}|^{1/2}}{|(\mathbf{X},\mathbf{x})^t(\mathbf{X},\mathbf{x}) + V_{(\mathbf{X},\mathbf{x})}^{-1}|^{1/2}} \frac{1}{\sigma} \cdot \left(\frac{\mathbf{y}^t\mathbf{y} - \mathbf{y}^t(\mathbf{X},\mathbf{x}) \left((\mathbf{X},\mathbf{x})^t(\mathbf{X},\mathbf{x}) + V_{(\mathbf{X},\mathbf{x})}^{-1} \right)^{-1} (\mathbf{X},\mathbf{x})^t\mathbf{y}}{\mathbf{y}^t\mathbf{y} - \mathbf{y}^t\mathbf{X} \left(\mathbf{X}^t\mathbf{X} + V_{\mathbf{X}}^{-1} \right)^{-1} \mathbf{X}^t\mathbf{y}} \right)^{-n/2}. \quad (8)$$

Note that here we used improper priors, but both the posterior and Bayes factor are proper (Servin and Stephens, 2007).

2.3 Numerical examples

Figure 1 shows comparisons between Bayes factors computed using Equation (8) and p-values from the likelihood ratio test calculated using IDUL software (Guan and Levy, 2024a), stratified according to minor allele frequencies (MAF) of the SNPs. The data, consisting 5757 whole genome sequenced samples with high level of relatedness and 79 protein assay phenotypes, were described previously (Guan and Levy, 2024b). We make the following observations: 1) In general, large Bayes factors correspond to small p-values. 2) Different Bayes factors might have the same p-values, and different p-values might have the same Bayes factors. 3) Left panel shows that, for SNPs with the same p-values, those having larger MAF tend to have larger Bayes factors than those having smaller MAF. This is because Bayes factors penalize SNPs of small MAF for their lower information contents through the prior effect size (Guan and Stephens, 2011). 4) Previously it has been demonstrated that asymptotically $2\log BF = \lambda\chi^2 + \log(1-\lambda)$, where $\lambda \in (0,1)$ increases with the SNP MAF and is also a function of prior effect size (Zhou and Guan, 2018). Right panel reflects the mathematical insight between Bayes factors and their corresponding χ^2-values.

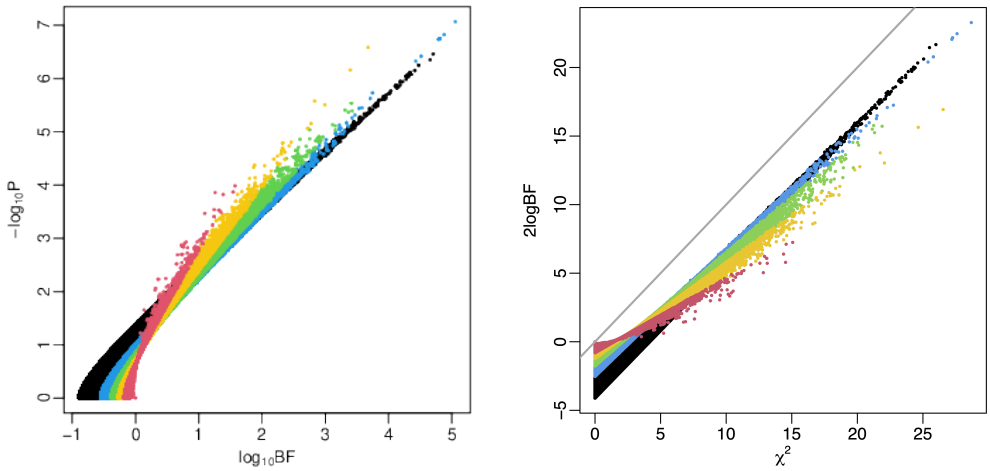

Figure 1: Comparison between p-values and Bayes factors for linear mixed model. The left panel shows $\log_{10} BF$ versus $-\log_{10} P$. The right panel shows χ^2 versus $2 \log BF$, where χ^2 values are quantiles of the p-values. The gray line is $x = y$. The color is based on MAF f, with black for $f \in (0.05, 0.5]$, blue for $f \in (0.02, 0.05]$, green for $f \in (0.01, 0.02]$, orange for $f \in (0.005, 0.01]$, and red for $f \in (0, 0.005]$.

2.4 Predictions and residuals

From model (3) we estimate parameters of $\hat{\mathbf{a}}$ and $\hat{\gamma}$ by $\left(\mathbf{X}^t\mathbf{X} + V_{\mathbf{X}}^{-1}\right)^{-1}\mathbf{X}^t\mathbf{y}$ (Methods). For a new sample, we have a vector $\mathbf{w}$ for covariates and a vector $\mathbf{g}$ for genetic relatedness between the new sample with n existing samples. Mapping $\mathbf{g}$ onto $\mathbf{Q}$ to get the loading $\mathbf{q} = \mathbf{Q}^t\mathbf{G}^{-1}\mathbf{g}$. (For an existing sample j, $\mathbf{q}$ is the j-th row of the matrix $\mathbf{Q}$.) Then the Bayesian prediction for the new sample is $\hat{y} = \mathbf{w}^t\hat{\mathbf{a}} + \mathbf{q}^t\hat{\gamma}$.

We now show that the best linear unbiased prediction (BLUP) is identical to Bayesian prediction. The Bayesian interpretation of BLUP is well known (Gianola and Fernando, 1986; Robinson, 1991). Here we offer a new derivation that is perhaps more natural from the perspective of genetic association studies. For BLUP one first estimates $\tilde{\mathbf{a}} = (\mathbf{W}^t\mathbf{H}^{-1}\mathbf{W})^{-1}\mathbf{W}^t\mathbf{H}^{-1}\mathbf{y}$, where $\mathbf{H} = \eta\mathbf{G} + \mathbf{I}_n$, and then one estimates $\tilde{\mathbf{u}} = \eta\mathbf{K}\mathbf{Z}^t\mathbf{H}^{-1}(\mathbf{y} - \mathbf{W}\tilde{\mathbf{a}})$. It is sufficient to show that $\hat{\mathbf{a}} = \tilde{\mathbf{a}}$ and $\mathbf{Q}\hat{\gamma} = \mathbf{Z}\tilde{\mathbf{u}}$. With reference to (14) and the fact that $\mathbf{Q}\mathbf{F}\mathbf{Q}^t = \mathbf{I}_n - \mathbf{H}^{-1}$, we have

$$\begin{aligned}\hat{\mathbf{a}} &= -\mathbf{M}\mathbf{W}^t\mathbf{Q}\mathbf{F}\mathbf{Q}^t\mathbf{y} + \mathbf{M}\mathbf{W}^t\mathbf{y} = -\mathbf{M}\mathbf{W}^t(\mathbf{I}_n - \mathbf{H}^{-1})\mathbf{y} + \mathbf{M}\mathbf{W}^t\mathbf{y} \\ &= \mathbf{M}\mathbf{W}^t\mathbf{H}^{-1}\mathbf{y} = (\mathbf{W}^t\mathbf{H}^{-1}\mathbf{W})^{-1}\mathbf{W}^t\mathbf{H}^{-1}\mathbf{y} = \tilde{\mathbf{a}},\end{aligned} \tag{9}$$

and

$$
\begin{aligned}
\mathbf{Q}\hat{\gamma} &= \mathbf{QFQ}^t\mathbf{y} + \mathbf{QFQ}^t\mathbf{WMW}^t\mathbf{QFQ}^t\mathbf{y} - \mathbf{QFQ}^t\mathbf{WMW}^t\mathbf{y} \\
&= \mathbf{QFQ}^t\mathbf{y} + \mathbf{QFQ}^t\mathbf{WMW}^t(\mathbf{I}_n - \mathbf{H}^{-1})\mathbf{y} - \mathbf{QFQ}^t\mathbf{WMW}^t\mathbf{y} \\
&= \mathbf{QFQ}^t\mathbf{y} - \mathbf{QFQ}^t\mathbf{WMW}^t\mathbf{H}^{-1}\mathbf{y} \\
&= (\mathbf{I}_n - \mathbf{H}^{-1})(\mathbf{y} - \mathbf{WMW}^t\mathbf{H}^{-1}\mathbf{y}) \\
&= (\mathbf{H} - \mathbf{I}_n)\mathbf{H}^{-1}(\mathbf{y} - \mathbf{W}\tilde{\mathbf{a}}) \\
&= \eta\mathbf{ZKZ}^t\mathbf{H}^{-1}(\mathbf{y} - \mathbf{W}\tilde{\mathbf{a}}) = \mathbf{Z}\tilde{u}.
\end{aligned}
\tag{10}
$$

Therefore, BLUP is identical to Bayesian prediction under a specific prior. For the existing samples, we can compute residual $\mathbf{r} = \mathbf{y} - \mathbf{W}\hat{\mathbf{a}} - \mathbf{Q}\hat{\gamma} = \mathbf{H}^{-1}(\mathbf{y} - \mathbf{W}\hat{\mathbf{a}})$. Residuals are useful in removing spurious correlations due to population structure and (cryptic) relatedness for subsequent analysis, such as studying gene-gene interaction and gene networks using gene expression values.

3 Discussion

It is a common practice to control for top principal components (PCs, the columns of the $\mathbf{Q}$ matrix) in genetic association study to control for population stratification (Price et al., 2006). It is self evident that if a putative causal SNP is correlated with both phenotype and top PCs, this practice of controlling for top PCs will reduce power. Our derivation suggests that when controlling for top PCs, one should put prior on their effect sizes $\eta\mathbf{D}_*$, where $\mathbf{D}_*$ is the a submatrix of $\mathbf{D}$ corresponding to the top PCs. Intuitively, the prior will keep some variations of PCs in the phenotypes for the putative causal SNPs, and a priori, the eigenvalues are the right amount of variation to be controlled for. Estimating η can be done efficiently. If estimating η is a hurdle for a practitioner, however, one can perhaps use $h/(1-h)$ instead, where h is the heritability of the phenotype.

In classical literature dealing with linear mixed model (Henderson, 1975; Robinson, 1991), the covariance structure of the error term is the more general $\mathbf{R}$ instead of our diagonal $\mathbf{I}_n$. We can left multiplying $\mathbf{R}^{-\frac{1}{2}}$ (where $\mathbf{R}^-$ is a generalized inverse) to convert the classical form to the form in Equation (1). In fact, we can rotate the system further and compute the (empirical) Bayes factor based on a rotated simple linear regression. Multiplying both side of Equation (3) by $\mathbf{Q}^t$ we get $\mathbf{y_Q} \sim MVN(\mathbf{W_Q}, \eta\mathbf{D} + \mathbf{I}_n)$, where $\mathbf{y_Q} = \mathbf{Q}^t\mathbf{y}$ and $\mathbf{W_Q} = \mathbf{Q}^t\mathbf{W}$. Let $\Lambda = \eta\mathbf{D} + \mathbf{I}_n$

and scale $\mathbf{y_Q}$ and $\mathbf{W_Q}$ by left multiplying $\Lambda^{-\frac{1}{2}}$, we get

$$\begin{aligned} \mathbf{y}_\Lambda &= \mathbf{W}_\Lambda \mathbf{a} + \mathbf{e} \\ \mathbf{e} &\sim \mathrm{MVN}_n(0, \tau^{-1}\, \mathbf{I}_n), \end{aligned} \tag{11}$$

where $\mathbf{y}_\Lambda = \Lambda^{-\frac{1}{2}}\mathbf{Q}^t\mathbf{y}$ and $\mathbf{W}_\Lambda = \Lambda^{-\frac{1}{2}}\mathbf{Q}^t\mathbf{W}$. Similarly Equation (4) becomes

$$\begin{aligned} \mathbf{y}_\Lambda &= \mathbf{W}_\Lambda \mathbf{a} + \mathbf{x}_\Lambda \beta + \mathbf{e} \\ \mathbf{e} &\sim \mathrm{MVN}_n(0, \tau^{-1}\, \mathbf{I}_n), \end{aligned} \tag{12}$$

where $\mathbf{x}_\Lambda = \Lambda^{-\frac{1}{2}}\mathbf{Q}^t\mathbf{x}$. Then using priors specified in (5), we can compute Bayes factors (Servin and Stephens, 2007). This rotation approach was briefly mentioned by (Wen, 2015). Bayes factor from this approach in equivalent to Equation (8), but this approach obscures insights on Bayesian prediction and its equivalence to BLUP.

We took an empirical Bayesian approach and plugged in ML estimates of the hyperparameter. Of course, we can take the full Bayesian approach by coming up with a prior for η. One sensible choice is to define $h = \eta/(1+\eta)$ and assign a uniform prior $\mathrm{Unif}(0,1)$ on h. Then η has a probability density function $f(\eta) = 1/(1+\eta)^2$, which is heavy tailed (Guan and Stephens, 2011). This full Bayesian approach is advantageous in principle, as it accounts for variation of η, but the computation is difficult, as this prior cannot be integrated out analytically, and either sampling or numerical integration are required for computation. For example, Wen (2015) took a full Bayesian approach assigning priors on η and τ, and elected to compute approximate Bayes factors for linear mixed model. On the other hand, by plugging in ML estimate $\hat{\eta}$, the subsequent computation can be done analytically. And this empirical Bayes approach is considered approximation to the full Bayesian treatment (Carlin and Louis, 2002).

Finally, Bayes factor is more desirable than p-values in applications that require comparing alternative models. One such example is to compare paternal effect and maternal effect of the same SNP. Due to sampling variation, the paternal allele frequencies can deviate from the maternal allele frequencies by a few standard deviation. When this happens, two p-values are no longer comparable due to different power. But their Bayes factors are comparable. Methods developed here are implemented in a software package that is freely available at `https://github.com/haplotype/idul`.

4 Methods

Bayes factor (8) can be evaluated efficiently. This computation was documented previously (Guan and Levy, 2024b), and we reproduced it here for context clarity. As $\mathbf{X}^t\mathbf{X} + V_{\mathbf{X}}^{-1} = \begin{pmatrix} I_n + \frac{1}{\eta\mathbf{D}} & \mathbf{Q}^t\mathbf{W} \\ \mathbf{W}^t\mathbf{Q} & \mathbf{W}^t\mathbf{W} \end{pmatrix}$ and denote $\mathbf{F} = (\mathbf{I}_n + \frac{1}{\eta\mathbf{D}})^{-1}$, $\mathbf{M} = (\mathbf{W}^t\mathbf{W} - \mathbf{W}^t\mathbf{Q}F\mathbf{Q}^t\mathbf{W})^{-1}$, we compute its determinant using the identity $\left|\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right| = |A| \cdot |D - CA^{-1}B|$ to get

$$|\mathbf{X}^t\mathbf{X} + V_{\mathbf{X}}^{-1}| = |\mathbf{I}_n + \frac{1}{\eta\mathbf{D}}| \cdot |\mathbf{W}^t\mathbf{W} - \mathbf{W}^t\mathbf{Q}\mathbf{F}\mathbf{Q}^t\mathbf{W}|;$$

Using the identities $\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}, & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1}, & (D - CA^{-1}B)^{-1} \end{pmatrix}$, we get

$$(\mathbf{X}^t\mathbf{X} + V_{\mathbf{X}}^{-1})^{-1} = \begin{pmatrix} \mathbf{F} + \mathbf{F}\mathbf{Q}^t\mathbf{W}\mathbf{M}\mathbf{W}^t\mathbf{Q}\mathbf{F}, & -\mathbf{F}\mathbf{Q}^t\mathbf{W}\mathbf{M} \\ -\mathbf{M}\mathbf{W}^t\mathbf{Q}\mathbf{F}, & \mathbf{M} \end{pmatrix}. \tag{13}$$

Note that for large matrices the computation only involves an inexpensive matrix multiplying vector, without expensive matrix multiplication and matrix inversion. For small matrices such $\mathbf{W}^t\mathbf{W}$ and M, the multiplication and inversion are inexpensive. The only expensive calculation is the eigendecomposition to obtain Q and D, which only needs to be done once. Similarly, both $|(\mathbf{X}, \mathbf{x})^t(\mathbf{X}, \mathbf{x}) + V_{(\mathbf{X},\mathbf{x})}^{-1}|$ and $((\mathbf{X}, \mathbf{x})^t(\mathbf{X}, \mathbf{x}) + V_{(\mathbf{X},\mathbf{x})}^{-1})^{-1}$ can be efficiently computed. Finally,

$$\begin{aligned} \begin{pmatrix} \gamma \\ \mathbf{a} \end{pmatrix} &= (\mathbf{X}^t\mathbf{X} + V_{\mathbf{X}}^{-1})^{-1} \cdot \begin{pmatrix} \mathbf{Q}^t\mathbf{y} \\ \mathbf{W}^t\mathbf{y} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{F} + \mathbf{F}\mathbf{Q}^t\mathbf{W}\mathbf{M}\mathbf{W}^t\mathbf{Q}\mathbf{F}, & -\mathbf{F}\mathbf{Q}^t\mathbf{W}\mathbf{M} \\ -\mathbf{M}\mathbf{W}^t\mathbf{Q}\mathbf{F}, & \mathbf{M} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{Q}^t\mathbf{y} \\ \mathbf{W}^t\mathbf{y} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{F}\mathbf{Q}^t\mathbf{y} + \mathbf{F}\mathbf{Q}^t\mathbf{W}\mathbf{M}\mathbf{W}^t\mathbf{Q}\mathbf{F}\mathbf{Q}^t\mathbf{y} - \mathbf{F}\mathbf{Q}^t\mathbf{W}\mathbf{M}\mathbf{W}^t\mathbf{y} \\ -\mathbf{M}\mathbf{W}^t\mathbf{Q}\mathbf{F}\mathbf{Q}^t\mathbf{y} + \mathbf{M}\mathbf{W}^t\mathbf{y} \end{pmatrix} \end{aligned} \tag{14}$$

References

Carlin, B. P. and T. A. Louis (2002). Empirical bayes: Past, present, and future. In A. E. Raftery, M. A. Tanner, and M. T. Wells (Eds.), *Statistics in the 21st Century*, pp. 312–318. Chapman & Hall.

Gianola, D. and R. L. Fernando (1986, 07). Bayesian Methods in Animal Breeding Theory. *Journal of Animal Science 63*(1), 217–244.

Guan, Y. and D. Levy (2024a). Asymptotically exact fit for linear mixed model. *bioRxiv*.

Guan, Y. and D. Levy (2024b, 02). Estimation of inbreeding and kinship coefficients via latent identity-by-descent states. *Bioinformatics 40*(2), btae082.

Guan, Y. and M. Stephens (2011). Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics 5*(3), 1780 – 1815.

Henderson, C. R. (1975). Best linear unbaised estimation and prediction under a selection model. *Biometrics 31*, 423–447.

Jiang, L., Z. Zheng, T. Qi, K. E. Kemper, N. R. Wray, P. M. Visscher, and J. Yang (2019). A resource-efficient tool for mixed model association analysis of large-scale data. *Nature Genetics 51*(12), 1749–1755.

Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly, and E. Eskin (2008, 03). Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics 178*(3), 1709–1723.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006, August). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet. 38*(8), 904–909.

Robinson, G. (1991). That blup is a good thing: the estimation of random effects (with disucssioins). *Statistical Science 6*(1), 15–51.

Servin, B. and M. Stephens (2007, 07). Imputation-based analysis of association studies: Candidate regions and quantitative traits. *PLOS Genetics 3*(7), 1–13.

Wen, X. (2015, 03). Bayesian model comparison in genetic association analysis: linear mixed modeling and SNP set testing. *Biostatistics 16*(4), 701–712.

Zhou, Q. and Y. Guan (2018). On the null distribution of bayes factors in linear regression. *Journal of the American Statistical Association 113*(523), 1362–1371.

Zhou, X. and M. Stephens (2012, June). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet 44*(7), 821–824.