Bayesian Linear Regression Models

Palle Duun Rohde, Izel Fourie Sørensen & Peter Sørensen

2023-08-24

Contents

1 Introduction

Bayesian linear regression models have been proposed as a unified framework for gene mapping, prediction of genetic predispostion, estimation of genetic parameters and effect size distribution (Moser et al. 2015).

Bayesian linear regression models attempts to account for the underlying genetic architecture of the trait. This is achieved by using many linked markers covering the entire genome to jointly estimate marker effects, and by allowing the genetic signal to be heterogeneous distributed over the genome (i.e. some regions have stronger genetic signal than others). This may in some situations allow a more accurate estimate of the true underlying genetic signal leading to more accurate predictions.

Bayesian linear regression models can also be used to map genetic variants associated with phenotypes and to estimate the total variance explained by the genetic markers. Because they fit all markers simultaneously and account for linkage disequilibrium between markers, they should have greater power to detect true associations, find less false negatives and give unbiased estimates of the larger marker effects. They can also provide information about the genetic architecture of the trait from the hyper-parameters of the distribution of marker effects.

Bayesian linear regression models fit all markers simultaneously and their effects as drawn from a prior distribution that attempts to match the true distribution of marker effects as closely as possible. However

the true distribution of effect sizes is unknown but a mixture of normal distributions can approximate a wide variety of distributions by varying the mixing proportions. Erbe et al. used this prior and included one component of the mixture with zero variance. A similar model was proposed by Zhou et al. but with a mixture of two normal distributions, one with a small variance and one with a larger variance.

In the following the statistical model, prior distributions of model parameters, algorithms for estimation of model parameters, extensions for handling multiple marker sets and multiple traits for the Bayesian linear regression models is presented. We will present two alternative BLR models, BLR model for marker effects and BLR model for individual effects.

2 Bayesion Linear Regression Model for Marker Effects

2.1 Statistical model and model parameters

In the multiple regression model the phenotype is related to the set of genetic markers:

$$y = Xb + e \tag{1}$$

where y is the phenotype, X a matrix of SNP genotypes, where values are standardised to give the ijth element as: $x_{ij} = (x_{ij} - 2p_j)/\sqrt{2p_j(1-p_j)}$, with x_{ij} the number of copies of the effect allele (e.g. 0, 1 or 2) for the ith individual at the jth SNP and p_j the allele frequency of the effect allele. b are the genetic effects for each SNP, and e the residual error. The dimensions of y, X, b and e are dependent upon the number of traits, k, the number of SNP markers, m, and the number of individuals, n. The residuals, e, are a priori assumed to be independently and identically distributed multivariate normal with null mean and covariance matrix $I\sigma_e^2$.

2.2 Estimation of parameters using Bayesian methods

In the Bayesian multiple regression model the posterior density of the model parameters $(b,\sigma_b^2,\sigma_e^2)$ depend on the likelihood of the data given the parameters and a prior probability for the model parameters:

$$p(b, \sigma_b^2, \sigma_e^2|y) \propto p(y|b, \sigma_b^2, \sigma_e^2)p(b|\sigma_b^2)p(\sigma_b^2)p(\sigma_e^2) \tag{2}$$

The prior density of marker effects, $p(b|\sigma_b^2)$, defines whether the model will induce variable selection and shrinkage or shrinkage only. Also, the choice of prior will define the extent and type of shrinkage induced. Ideally the choice of prior for the marker effect should reflect the genetic architecture of the trait, and will vary (perhaps a lot) across traits. Most complex traits and diseases are likely highly polygenic, with hundreds to thousands of causal variants, most frequently of small effect. So, the prior distribution must include many small and few large effects. Furthermore marker effects are a priory assumed to be uncorrelated (but markers can be in strong linkage disequilibrium and therefore a high posterior correlation). Many priors for marker effects have been proposed. These priors come more from practical (ease of computation) than from biological reasons. Each prior originates a method or family of methods, and we will describe some of them next, as well as their implications.

Prior marker variance Bayes N

In the Bayes N approach the prior the marker effect, b, follows a priori a normal distribution with a variance σ_b^2 which is constant across markers:

$$p(b) = \prod_i p(b_i) \tag{3}$$

where

$$p(b_i) = N\left(0, \sigma_b^2\right) \tag{4}$$

In a normal distribution most effects are concentrated around 0, whereas few effects will be large. Therefore, the prior assumption of normality precludes few markers of having very large effects – unless there is a lot of information to compensate for this prior information.

Prior marker variance Bayes A

In the Bayes A approach it is assumed that a priori we have some information on the marker variance. For instance, this can be σ_b^2. Thus, we may attach some importance to this value and use it as prior information for $\sigma_{b_i}^2$. A natural way of doing this is using an inverted chi-squared distribution with with v_b degrees of freedom and scale parameter $S_b^2 = v_b \sigma_b^2$

$$p(b_i | \sigma_{b_i}^2) = N\left(0, \sigma_{b_i}^2\right) \tag{5}$$

In the second stage, we postulate a prior distribution for the variance themselves:

$$p(\sigma_{b_i}^2 | v_b, S_b^2) = S_b^2 \chi_{v_b}^{-1} \tag{6}$$

The value of σ_b^2 should be set as $\sigma_b^2 = \frac{v_b - 2}{v_b} \frac{\sigma_g^2}{2 \sum_i p_i (1-p_i)}$ because the variance of a t distribution is $\frac{v_b}{v_b - 2}$. It can be shown that this corresponds to a prior on the marker effects corresponding to a scaled t distribution (Gianola et al. 2009):

$$p(b_i | \sigma_b^2, v_b) = \sigma_b t\left(0, v_b\right) \tag{7}$$

which has the property of having "fat tails". This means that large marker effects are more likely a priori compared to a normal distribution.

Prior marker variance Bayes C

In the Bayes C approach the marker effects, b, are a priori assumed to be sampled from a mixture with a point mass at zero and univariate normal distribution conditional on common marker effect variance σ_b^2. This reflect a very common thought was that there were not many causal loci. This can be implemented by introducing additional variables δ_i which explain if the i-th marker has an effect or not. In turn, these variables δ have a prior distribution called Bernouilli with a probability π of being 0. Therefore the hierarchy of priors is:

$$p(b_j | \delta_i, \sigma_{b_i}^2, \pi) = \begin{cases} 0 & \text{with probability } \pi, \\ \sim N(0, \sigma_{b_i}^2) & \text{with probability } 1 - \pi, \end{cases} \tag{8}$$

$$p(\sigma_{b_i}^2 | v_b, S_b^2) = S_b^2 \chi_{v_b}^{-1} \tag{9}$$

where $S_b^2 = \sigma_b^2 v_b$ with $\sigma_b^2 = \frac{\sigma_g^2}{(1-\pi)2 \sum_i p_i (1-p_i)}$ because the variance of a t distribution is $\frac{v_b}{v_b - 2}$.

Prior marker variance Bayes R

In the Bayes R approach the marker effects, b, are a priori assumed to be sampled from a mixture with a point mass at zero and univariate normal distributions conditional on common marker effect variance σ_b^2, and

variance scaling factors, γ:

$$b_j|\pi,\sigma_b^2 = \begin{cases} 0 & \text{with probability } \pi_1, \\ \sim N(0,\gamma_2\sigma_b^2) & \text{with probability } \pi_2, \\ \vdots & \\ \sim N(0,\gamma_C\sigma_b^2) & \text{with probability } 1 - \sum_{c=1}^{C-1}\pi_c, \end{cases} \tag{10}$$

where $\pi = (\pi_1,\pi_2,....,\pi_C)$ is a vector prior probabilities and $\gamma = (\gamma_1,\gamma_2,.....,\gamma_C)$ is a vector of variance scaling factors for each of C marker variance classes. The γ coefficients are prespecified and constrain how the common marker effect variance σ_b^2 scales within each mixture distribution. Typically $\gamma = (0,0.01,0.1,1.0)$. and $\pi = (0.95,0.02,0.02,0.01)$.

The prior distribution for the marker variance σ_b^2 is assumed to be an inverse Chi-square prior distribution,$\chi^{-1}(S_b,\nu_b)$.

The proportion of markers in each mixture class π follows a Direchlet $(C,c+\alpha)$ distribution, where c is a vector of length C that contains the counts of the number of variants in each variance class and $\alpha = (1,1,1,1)'$.

Using the concept of data augmentation, an indicator variable $d = (d_1,d_2,..,d_{m-1},d_m)$, is introduced, where d_j indicates whether the j'th marker effect is zero or nonzero.

2.2.1 Estimation of model parameters

Bayesian linear regression methods use an iterative algorithm for estimating joint marker effects. Estimation of the joint marker effects depend on additional model parameters such as a probability of being causal (π), an overall marker variance (σ_b^2), and residual variance (σ_e^2). Estimation of model parameters can be done using MCMC techniques by sampling from fully conditional posterior distributions.

To illustrate consider the following MCMC algorithm used to obtain estimates of parameters in BayesR models. The multiple linear regression model is parameterized in terms of $\theta = (\sigma_e^2,\sigma_\beta^2,\gamma,\pi,\ b,\ d)$. The full conditional sampling distributions for these parameters are presented below.

The joint posterior for all parameter in the multiple regression model can be written as :

$$f\left(\sigma_e^2,\sigma_\beta^2,\gamma,\pi,\ b,\ d|y\right) \propto \left(y|\sigma_e^2,\sigma_\beta^2,\gamma,\pi,\ b,\ d\right) f\left(b|d,\gamma,\sigma_\beta^2\right) f\left(d|\pi\right) f\left(\sigma_e^2\right) f\left(\sigma_\beta^2\right) f\left(\pi\right) \tag{11}$$

The parameters d_j and b_j are sampled jointly from their joint full conditional distributions, which can be written as the product of the full conditional distribution of b_j given d_j and the marginal full conditional distribution of d_j:

$$f_j\left(d_j,b_j|\theta_{-j},\widetilde{y}\right) = f_j\left(b_j|\theta_{-j},d_j,\widetilde{y}\right) f_j\left(d_j,|\theta_{-j},\widetilde{y}\right) \tag{12}$$

Where θ_{-j} is all parameters except d_j and b_j, and $\widetilde{y} = y - X_{-j}b_{-j}$ is the phenotype adjusted for all marker effects except for j'th marker (b_j).

The full conditional distribution for b_j can be be written as:

$$f_j\left(b_j|\theta_{-j},d_j,\widetilde{y}\right) \propto N\left(C_j^{-1}r_j,C_j^{-1}\right) \tag{13}$$

where $C_j^{-1} = \frac{\sigma_e^2}{x_j'x_j+\frac{\sigma_e^2}{\sigma_c^2}}$ and $r_j = \frac{x_j'\widetilde{y}}{\sigma_e^2}$

The marginal full conditional probability for the indicator variable of d_j is:

$$f_j\left(d_j = c | \theta_{-j}, \widetilde{y}\right) = \frac{f_j\left(\widetilde{y}|d_j = c, \theta_{-j}\right) f\left(d_j | \pi_c\right)}{\sum_{k=1}^{C} f_j\left(\widetilde{y}|d_j = k, \theta_{-j}\right) f\left(d_j | \pi_k\right)} \tag{14}$$

where $f_j\left(\widetilde{y}|d_j = c, \theta_{-j}\right) = \left(\sigma_c^2\right)^{-0.5} \left(C_j^{-1}\right)^{0.5} \exp\left[\frac{1}{2}C_j^{-1}r_j^2\right]$

From these expression probabilities for the categorical distribution for an arbitrary number of mixture components can be computed. Given these probabilities sample from a categorical distribution, which determines which class the variant will be sampled from. Conditional on the marker variance class sample the effect from the relevant normal distribution or give it a zero effect. To sample from the categorical distribution: 1) create a vector of cumulative probabilities calculated from above ordered by category, and 2) accept the lowest c such that the cumulative probability > u, where u is sample from a uniform distribution $U\left(0,1\right)$.

The proportion of markers in each mixture class π follows a *Direchlet* $\left(C, c + \alpha\right)$ distribution, where **c** is a vector of length C that contains the counts of the number of variants in each variance class and $\alpha = \left(1, 1, .., 1\right)$.

The full conditional posterior distribution for σ_β^2 is a scaled inverse chi-squared distribution with $v_\beta = \widetilde{\nu}_\beta + q$ degrees of freedom and scale parameter $S_\beta^2 = \frac{\widetilde{\nu}_\beta \widetilde{S}_\beta^2 + \sum_{j=1}^{q} \frac{b_j^2}{\gamma_{d_j}}}{\widetilde{\nu}_\beta + q}$ where $\widetilde{\nu}_\beta$ and $\widetilde{S}_\beta^2$ are the prior degrees of freedom and scale parameters.

The full conditional posterior distribution for σ_e^2 is a scaled inverse chi-squared distribution with $v_e = \widetilde{\nu}_e + n$ degrees of freedom and scale parameter $S_\beta^2 = \frac{\widetilde{\nu}_e \widetilde{S}_e^2 + sse}{\widetilde{\nu}_e + n}$ where $\widetilde{\nu}_e$ and $\widetilde{S}_e^2$ are the prior degrees of freedom and scale parameters and $sse = y'y - b'r^* - b'X'y$.

2.3 Extensions to summary statistics

The key parameter of interest in the multiple regression model are the marker effects. These can be obtained by solving an equation system similar to:

$$b = \left(X'X + I\frac{\sigma_e^2}{\sigma_b^2}\right)^{-1} X'y \tag{15}$$

In order to solve this equation system individual level data (genotypes X and phenotypes y) is required. If these are not available, it is possible to reconstruct $X'y$ and $X'X$ from a LD correlation matrix B (from a population matched LD reference panel) and summary statistics:

$$X'X = D^{0.5}BD^{0.5} \tag{16}$$
$$X'y = Db_m \tag{17}$$

where $D_i = \frac{1}{\sigma_{b_i}^2 + b_i^2/n_i}$ if the genotypes have been centered to mean 0 or $D_i = n_i$ if the genotypes have been centered to mean 0 and scaled to unit variance, and $b_m = D^{-1}X'y$ is the marginal marker effects obtained from a standard GWAS. The construction of a LD correlation matrix, B, is shown in session 5.

The summary statistics methods used require the construction of a reference LD correlation matrix. Typically this is done through the use of a fixed 1–10-Mb window approach, as in GCTA-SBLUP or LDpred, which sets LD correlation values outside this window to zero. Zhu and Stephens detail the reasons for using the shrinkage estimator of the LD matrix, which shrinks the off-diagonal entries of the LD correlation matrix

towards zero and is required for the implementation based on summary statistics. Experimentation with different types of sparse LD correlation matrices led to the conclusion that the shrinkage estimator was the most stable for SBayesR implementation. Briefly, each element of the reference LD correlation matrix Bij is shrunk by the factor $exp(-\varrho_{ij}/2m)$, where m is taken to be the sample size used to generate the genetic map, (ϱ_{ij}) is an estimate of the population-scaled recombination rate between SNPs i and j taken as $\varrho_{ij} = 4N_ec_{ij}$, for Ne the effective population size and c_{ij} the genetic distance between sites i and j in centimorgans as stated in Li and Stephens. LD matrix entries are set to zero if $exp(-\varrho_{ij}/2m)$ is less than a user-chosen cutoff.

Genetic distance between sites is derived from the genetic map files containing interpolated map positions for the CEU population generated from the 1000G OMNI arrays (Data availability). The calculation of the shrunk LD matrix requires the effective population sample size, which we set to be 11,400 (as in Zhu and Stephens), the sample size of the genetic map reference, which corresponds to the 183 individuals from the CEU cohort of the 1000G and the hard threshold on the shrinkage value, which we set to 10^{-3}. This threshold gave a good balance between computational efficiency and accuracy with, on average, each SNP having a window width of 10.6 Mb (SD=5.6Mb) across the autosomes. The shrunk LD matrix is stored in a sparse matrix format (ignoring matrix elements equal to 0) for efficient SBayesR computation. Currently, the LD matrix construction can only be performed with PLINK hard-call genotypes.

2.4   Fine mapping approaches based on BLR models

BLR models used for fine-mapping have been specialized in order to focus on the markers that have the largest chance of being causal. The challenge of BLR models is determining which markers have non-zero effect sizes on a trait. The BLR models are fitted using all markers simultaneously, and fine mapping approaches must therefore consider sets of markers in a genomic region rather than individual markers, because the effect of a causative mutation may be distributed across multiple markers. Although markers with non-zero effect sizes may be referred to as causal, it is important to realize that statistical methods alone cannot determine causality.

For BLR models, several association statistics have been developed to identify which markers or windows of markers can be considered as explaining a substantial or significant proportion of the genetic variance.

The genomic region used for defining the marker set can be defined by non-overlapping or sliding windows across the genome. The genomic region used in the fine mapping procedure may be determined by a fixed number of markers (e.g. 1000 markers) or by physical (e.g. 1 Mb) or genetic maps (e.g. 1 cM).

In general a BLR model for fine-mapping can be represented by an indicator variable for each marker (or marker set), with values of 1 for causal and 0 for not, and by organizing these indicators for all markers (or marker set) of interest in vector d. For m markers, there are 2^m possible d vectors (hence, 2^m possible models), ranging from all values of d equal to 0 for no markers causal to all values equal to 1 for all markers causal. There are several ways to specify the prior probability for a model, such as assuming that variants are independent and equally likely to be causal or assuming a fixed number of causal variants out of the total variants.

- To what extend mapping power and false discoveries of the BLR models depends on the choice of prior require further investigation.

- To what extend mapping power and false discoveries of the BLR models depends on the choice of association statistics require further investigation.

- To what extend mapping power and false discoveries of the BLR models depends on the choice of genomic region require further investigation.

2.5 Extensions to Multiple Traits and Multiple Components

Bayesian linear regression models have been extended to multiple trait analyses. A multiple trait analysis is a natural choice in quests for understanding and dissecting genetic correlations between traits using genetic markers, e.g., evaluating whether pleiotropy or linkage disequilibrium are at the root of between-trait associations. Furthermore pleiotropy among traits can be utilised to increase accuracy of genomic predictions.

Multiple trait Bayesian linear regression models can implemented under an restrictive assumption that a locus simultaneously affects all the traits or none of them (Jia and Jannick 2012). This approach has the advantage of simplicity, but the disadvantage that many effects might need to be estimated for loci that have no effect on a trait, and this may erode the accuracy of prediction. This should not be a problem for asymptotically large datasets, as in that case the fitted locus effects should converge to zero for those traits not influenced by that locus. Furthermore, this assumption is not biologically meaningful, especially in multiple trait analyses involving many traits. This assumption of genetic architecture is violated if some loci have no effect on at least one of the traits while having an effect on the remaining traits.

Therefore a general multiple trait Bayesian linear regression model based on the Bayes C prior have been proposed (Cheng et al. 2018). In this model a locus is allowed to affect any combination of traits, e.g., in a 2-trait analysis, the "restrictive" model only allows two situations, whereas in the general model allow all $2^2 = 4$ situations. This model is particularly interesting because it provides insight into whether markers affect all, some, or none of the traits addressed.For example, the proportion of markers in each of the (0,0), (1,0), (0,1) and (1,1) categories, where (0,0) means "no effect," and (1,1) denotes "effect" on both traits. This model becomes computational intensive for a multiple trait analysis of a large number of traits. To partly reduce the computational complexicity it can be implemented such that the multi-variate prior distributions used for marker effects assume the same correlation for all markers.

A multiple trait Bayesian linear regression model based on the Bayes R prior (Kemper et al. 2022) which can be considered a hybrid model (compared to MT-BLR-STRICT and MT-BLR-FLEX). In this model the prior for the marker effects is a mixture of normal distributions (i.e. Bayes R prior). However these distributions were assumed to be independent between traits but with a specified prior proportion of markers having no effect on any trait. Thus the "learning" between traits is primarily from the marker-trait indicator variable which is a slightly more flexible approach.

The above mentioned MT-BLR can also be implemented using the Bayes N and Bayes A priors (Fernando and Gianola).

In summary MT-BLR models provide a flexible approach for the detection of pleiotropic loci which can be utilised to increase accuracy of genomic predictions. However, the impact of prior assumptions on marker effects across traits requires further investigation. Also, an important limitation of all methods that attempt to use genotype–phenotype associations to detect pleiotropic loci is that they cannot differentiate the presence of a pleiotropic locus from the presence of two closely linked single-trait loci, depending on the extent of LD in the region. Furthermore issues with inferences about genetic (co) variance based on BLR models can misrepresent the true genetic parameters if the causal loci are not genotyped because of incomplete LD between markers and causal loci and among causal loci (de los Campos et al., Gianola et al.).

2.5.1 Posterior sampling distributions for multiple trait BLR model based on the Bayes C prior

A multiple trait BLR (MT-BLR) model estimate joint marker effect accounting for LD and borrow information across traits based on the following equation (two-trait example):

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \left(\begin{bmatrix} X_1^{'} X_1 & 0 \\ 0 & X_2^{'} X_2 \end{bmatrix} + I \otimes B^{-1} E \right)^{-1} \begin{bmatrix} X_1^{'} y_1 \\ X_2^{'} y_2 \end{bmatrix} \tag{18}$$

- b_1 and b_2 are the marker effects for trait 1 and 2

- y_1 and y_2 are the vectors of phenotypes for trait 1 and 2

- X_1 and X_2 are the matrices of genotypes for trait 1 and 2

- $X_1' X_1$ and $X_2' X_2$ are the LD t-erms which may differ across ancestries

- B is the genetic co-variance matrix between traits

$$B = \begin{bmatrix} \sigma_{b_1}^2 & \sigma_{b_{12}}^2 \\ \sigma_{b_{21}}^2 & \sigma_{b_2}^2 \end{bmatrix} \tag{19}$$

- E is the residual co-variance matrix between traits

$$E = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}}^2 \\ \sigma_{e_{21}}^2 & \sigma_{e_2}^2 \end{bmatrix} \tag{20}$$

- the shrinkage term, $B^{-1}E$, that depends on both prior distribution of the marker effects (e.g. Bayes C) and the model selection procedure (e.g. marker-trait configuration)

The covariance matrix **B** for the marker effect is a priori assumed to follow an inverse Wishart distribution $IW(\widetilde{S}_b^2, \widetilde{\nu}_b)$ where $\widetilde{\nu}_b$ and $\widetilde{S}_b^2$ (t by t matrix) are the prior degrees of freedom and scale parameters.

The full conditional distribution for **B** the covariance matrix for the jth marker is also an inverse Wishart distribution of the form:

$IW(\widetilde{S}_b^2 + bb', \widetilde{\nu}_b + m)$, where **b** is a matrix (m x t) of sampled marker effects

or

$IW(\widetilde{S}_\beta^2 + b_j b_j', \widetilde{\nu}_\beta + 1)$, where b_j is a vector (t x 1) of sampled marker effects.

Difference between the two sampling distributions is that the former is the same for all markers whereas the latter is more flexible as it allows the markers to their own covariance matrix.

The covariance matrix **E** for the residual effects is a priori assumed to follow an inverse Wishart distribution $IW(\widetilde{S}_e^2, \widetilde{\nu}_e)$ where $\widetilde{\nu}_e$ and $\widetilde{S}_e^2$ (t by t matrix) are the prior degrees of freedom and scale parameters.

The full conditional distributions for the covariance matrix **E** for residuals is also an inverse Wishart distribution of the form:

$IW(\widetilde{S}_e^2 + ee'$, $\widetilde{\nu}_e + m)$, where **e** is a matrix (n x t) of residual effects

The full conditional distribution for b_j can be be written as:

$$f_j\left(b_j | \theta_{-j}, d_j, \widetilde{y}\right) \propto N\left(C_j^{-1} r_j, C_j^{-1}\right) \tag{21}$$

where $C = D_j' E^{-1} D_j x_j' x_j + G^{-1}$ and $r_j = x_j' \widetilde{y} E^{-1} D_j = \widetilde{b}_j x_j' x_j E^{-1} D_j$

where D_j is a diagonal matrix whose kth diagonal entry is an indicator variable indicating whether the marker jth effect for trait k is zero or nonzero.

The marginal full conditional probability for the indicator variable of d_j is:

$$f_j\left(d_j = c | \theta_{-j}, \widetilde{y}\right) = \frac{f_j\left(\widetilde{y} | d_j = c, \theta_{-j}\right) f\left(d_j | \pi_c\right)}{\sum_{k=1}^{C} f_j\left(\widetilde{y} | d_j = k, \theta_{-j}\right) f\left(d_j | \pi_k\right)} \tag{22}$$

where $f_j \left(\widetilde{y} | d_j = c, \theta_{-j} \right) = |C_j^{-1}|^{0.5} \exp \left[\frac{1}{2} r_j' C_j^{-1} r_j \right]$

In the most general case, any marker effect might be zero for any possible combination of t traits resulting in 2^t possible combinations of d_j: For example, in a t=2 trait model, there are $2^t = 4$ combinations for d_j: (0,0), (0,1), (1,0), (1,1). Suppose, in general, we use numerical labels 1,2,...,l for the 2^t possible outcomes for d_j then the prior for d_j is a categorical distribution.

3   Bayesion Linear Regression Model for Individual Effects

3.1   Statistical model and model parameters

We begin with a description of the general two-trait model since the simpler models are special cases. The general two-trait model partitions the total phenotypic variance into a component explained by genetic marker effects, g, and a component due to residual effects that are not associated with markers, e. Additionally, the component due to genetic markers g is subdivided into contributions from C marker sets g_i, $i = 1, \ldots, C$.

The model for trait 1 is assumed to be

$$y_1 = \mu_1 + \sum_{i=1}^{C} g_{1i} + e_1, \tag{23}$$

and similar for trait 2,

$$y_2 = \mu_2 + \sum_{i=1}^{C} g_{2i} + e_2, \tag{24}$$

In these expressions, the μ's are scalar means for each trait, the $g's$ are contributions to genetic effects from each of C marker sets that can be associated with genetic marker information (genomic effects or genomic values), the $e's$ represent a residual component that cannot be captured by regression on markers.

The genomic value for marker set k is defined as the sum of the effects of all the markers in marker set k,

$$\begin{aligned} g_1 k &= \sum_{i=1}^{p_k} w_{ijk} b_{1ik} \\ &= w_{jk}' b_{1k}, \end{aligned} \tag{25}$$

with an equivalent expression for trait 2. In (25), p_k is the number of markers in marker set k, the scalar w_{ijk} is the observed (centered and scaled) label for marker i in individual j for marker set k, and b_{1ik} is the effect of marker i for marker set k of trait 1.

The n by 1 vectors of genomic effects for trait 1 and trait 2 for marker set k are $g_{1k} = W_k b_{1k}$ and $g_{2k} = W_k b_{2k}$, respectively, where $W_k = \{w_{ijk}\}$ is the observed n by p_k matrix of marker genotypes of marker set k. The joint distribution of vectors b_{1k} and b_{2k} is assumed to be

$$\left[\begin{array}{c} b_{1k} \\ b_{2k} \end{array} \right] \sim N \left(\left[\begin{array}{c} 0 \\ 0 \end{array} \right], \left[\begin{array}{cc} I\sigma_{b_{1k}}^2 & I\sigma_{b_{1k}b_{2k}} \\ I\sigma_{b_{1k}b_{2k}} & I\sigma_{b_{2k}}^2 \end{array} \right] \right) \tag{26}$$

where the $I's$ represent p_k by p_k identity matrices, $\sigma_{b_{1k}}^2 \left(\sigma_{b_{2k}}^2 \right)$ is the prior variance of marker effects for trait 1 (trait 2) in marker set k, and $\sigma_{b_{1k}b_{2k}}$ is their prior covariance.

Due to the centering the rank of W_k is $n-1$. It follows from these assumptions and from standard properties of the multivariate normal distribution that the model for the joint distribution of genomic values in trait 1 and 2 is the singular multinormal (SN) distribution,

$$\begin{bmatrix} g_{1k} \\ g_{2k} \end{bmatrix} \sim SN \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G_k \sigma_{g_{1k}}^2 & G_k \sigma_{g_{1k}g_{fk}} \\ G_k \sigma_{g_{1k}g_{2k}} & G_k \sigma_{g_{2k}}^2 \end{bmatrix} \right), \quad k = 1, \ldots, C. \tag{27}$$

Above, $\sigma_{g_{1k}}^2$ is the genomic variance in trait 1 for marker set k, $\sigma_{g_{2k}}^2$ is the genomic variance in trait 2 for marker set k, $\sigma_{g_{1k}g_{2k}}$ is the genomic covariance between trait 1 and 2 for marker set k, $G_k = \frac{1}{p_k} W_k W_k'$ is the genomic relationship matrix of rank $n-1$ for marker set k, and SN denotes the singular normal distribution (details in the Supplementary Methods). A more compact notation for (27) is

$$g_k = (g_{1k}', g_{2k}')' \sim SN \left(0, G_k \otimes V g_k \right) \tag{28}$$

where

$$V_{g_k} = \begin{bmatrix} \sigma_{g_{1k}}^2 & \sigma_{g_{1k}g_{2k}} \\ \sigma_{g_{1k}g_{2k}} & \sigma_{g_{2k}}^2 \end{bmatrix}. \tag{29}$$

Similarly, collecting the $e's$ in vectors with n elements for each trait joint distribution is assumed to be

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I\sigma_{e_1}^2 & I\sigma_{e_1 e_2} \\ I\sigma_{e_1 e_2} & I\sigma_{e_2}^2 \end{bmatrix} \right), \tag{30}$$

The identity matrix I is of order n by n in (30). The phenotypes for trait 1 and 2, y_1 and y_2, each with n elements, are conditionally normally distributed, given g, and e, and take the form

$$y_1 | \mu_1, g_1 \quad \sim \quad N \left(1\mu_1 + \sum_{i=1}^{C} g_{1i}, I\sigma_{e_1}^2 \right), \tag{31}$$

$$y_2 | \mu_2, g_2 \quad \sim \quad N \left(1\mu_2 + \sum_{i=1}^{C} g_{2i}, I\sigma_{e_2}^2 \right), \tag{32}$$

where σ_{e1}^2 (σ_{e2}^2) is the residual variance for the trait 1 (trait 2).

3.2 Extensions to multiple components

The model specified quantify the relative contribution from each of the marker sets by fitting separate variance components for each of those marker sets. Here we describe the basis for this partitioning. The point of departure is to assume that the $2n_t$ by 1 vector of genomic effects $g = (g_1', g_2')'$ is equal to the sum of the vectors of genomic effects belonging to C marker set components. Then

$$g = \sum_{i=1}^{C} g_i, Var(g|W) = \sum_{i=1}^{C} Var(g_i|W_i) = \sum_{i=1}^{C} \frac{1}{p_i} W_i W_i' \otimes V_{g_i} \tag{33}$$

where g_i is the $2n$ by 1 vector of genomic values for trait 1 and 2, associated with component i, $i = 1, \ldots, C$, p_i is the number of marker genotypes in component i, and V_{g_i} is the 2 by 2 genomic covariance matrix for marker set i. This holds under the model assumption that the components of the vector of marker effects $b = (b_1', \ldots, b_i', \ldots b_C')'$ are realisations from $b_i \sim N(0, I \otimes V_{bi})$, with $Cov(b_i, b_j') = 0$, $i, j = 1, \ldots, C; i \neq j$. Then $Cov(g_i, g_j'|W) = Cov(W_i b_i, b_j' W_j') = 0$. In these expressions, $b_i = (b_{1i}', b_{2i}')'$ is the vector of marker effects of marker set component i, and

$$V_{bi} = \begin{bmatrix} \sigma_{b_1 i}^2 & \sigma_{b_{1i} b_{2i}} \\ \sigma_{b_{1i} b_{2i}} & \sigma_{b_2 i}^2 \end{bmatrix}. \tag{34}$$

Two simple two-trait models are implemented that partition the total genomic variance into multiple component defined by marker sets. Details of the prior distributions of the parameters of the Bayesian models and of the Markov chain Monte Carlo algorithm can be found in the Supplementary Methods.

3.2.1 Inferences

For single trait analysis, inferences are reported in terms of ratios for the ith marker sets defined as:

$$h_{g_i}^2 = \frac{\sigma_{g_i}^2}{\sigma_g^2}, \quad i = 1, \ldots, C. \tag{35}$$

with $\sigma_g^2 = Var\left(\sum_{i=1}^C g_i\right)$. When the elements of W are scaled so that the average of the diagonal elements of WW' is equal to 1, this ratio quantifies the proportion of total genomic variance, captured by genetic marker information associated with marker set i. The ratio involving marker effects from all the marker sets for trait 1 is

$$h_{g_1}^2 = \frac{\sigma_{g_1}^2}{\sigma_{g_1}^2 + \sigma_{e_1}^2} \tag{36}$$

with a similar expression for trait 2.

We also report within marker set genomic correlations

$$r_{g_i} = \frac{\sigma_{g_{i_1} g_{i_2}}}{\sigma_{g_{i_1}} \sigma_{g_{i_2}}}, \quad i = 1, \ldots, C, \tag{37}$$

and the total genomic correlation between traits, defined as

$$r_{g_i} = \frac{Cov\left(\sum_{i=1}^C g_{1_i}, \sum_{i=1}^C g_{2_i}\right)}{\sqrt{\sigma_{g_1}^2 \sigma_{g_2}^2}} \tag{38}$$

The absence of symmetry of posterior distributions is well captured by the Bayesian McMC methods. In a situation like this, summarizing results in the form of posterior means and standard deviations would be misleading. In the presence of asymmetry, posterior means are poor indicators of points of high probability mass. We therefore summarize inferences in terms of posterior modes and posterior intervals.

3.3 Estimation of model parameters

The models are implemented using empirical Bayesian methods. The models were implemented using a Bayesian, Markov chain Monte Carlo approach, whereby the hyperparameters of the dispersion parameters of the Bayesian model were estimated by maximum likelihood, and conditional on these, the model was fitted using Markov chain Monte Carlo. To illustrate, consider determining a value of the scale parameter of an inverse chi square distribution, which is the prior assumed for the variance components. This is achieved by equating the mode of the inverse chi square prior density, for a fixed value of the degrees of freedom, to the maximum likelihood estimate and solving for the scale parameter. The decision to huse this approach rather than classical likelihood is based on the need to obtain measures of uncertainty that account as much

as possible for the limited amount of information in the data without resorting to large sample theory, which is inherent in classical likelihood. This is relevant if the data are of limited size and results in marginal posterior distributions that are expected to be asymmetric, particularly in the case of the more complex, general two-trait model.

A number of technical details are presented in this section. These include a formal derivation of the probability model for the variance between traits and details of a spectral decomposition that plays an important computational role in the Markov chain Monte Carlo strategy implemented, as well as a full description of the Bayesian models.

3.3.1 The spectral decomposition

The Bayesian implementation of the model is facilitated making use of the following spectral decomposition performed within each chromosome segment (we drop subscripts that refer to marker sets to avoid clotting the notation). Let

$$g_1 = Wb_1 \quad g_2 = Wb_2 \tag{39}$$

where W is of order $n \times p$ and g_m is of order $n \times 1$. The spectral decomposition of WW' (for each marker set) is

$$WW' = U\Delta U' = \sum_{i=1}^{n_t} \lambda_i U_i U_i', \tag{40}$$

where $U = [U_1, U_2, \ldots, U_{n_t}]$, of order $n \times n$ is the matrix of eigenvectors of WW', U_j is the jth column (dimension $n \times 1$), and Δ is a diagonal matrix with elements equal to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{n_t}$ associated to the n_t eigenvectors. Since WW' is positive semidefinite the eigenvalues are $\lambda_i \geq 0$, $i = 1, 2, \ldots, n$. The eigen vectors satisfy $U'U = UU' = I$.

Define the $n \times n$ matrix $G = \frac{1}{p}WW' = \frac{1}{p}WW'$ and write this as

$$
\begin{aligned}
G &= \tfrac{1}{p}WW' \\
&= \tfrac{1}{p}U\Delta U' \\
&= UDU'
\end{aligned}
$$

where

$$D = \frac{1}{p}\Delta.$$

Matrix G (peculiar to each marker set) is singular and the diagonal matrix D (also peculiar to each marker set) contains only $n - 1$ positive eigenvalues.

3.3.2 The case of singular G

Due to the singularity of G, $[g|W, \sigma_g^2]$ does not follow a multivariate normal distribution but a singular normal distribution instead. The singular normal density is

$$p\left(g_1|W, \sigma_{g1}^2\right) = \frac{1}{(2\pi)^{\frac{n-1}{2}} \left(\lambda_1 \sigma_{g_1}^2 \ldots \lambda_{n-1}\sigma_{g_1}^2\right)^{\frac{1}{2}}} \exp\left(-\frac{g_1' G^- g_1}{2\sigma_{g_1}^2}\right) \tag{41}$$

where the $n - 1$ $\lambda's$ are the non-zero eigenvalues of G and G^- is any generalised inverse of G. One choice choice of generalised inverse of G is

$$G^- = UD^-U' \tag{42}$$

where U and D are of order n_t by n_t

$$D^- = \frac{1}{p} \begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \cdots & \cdots & 0 \\ \vdots & \vdots & \frac{1}{\lambda_{n-1}} & \cdots & 0 \\ \vdots & \vdots & \cdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \frac{1}{p} \begin{bmatrix} D_1^{-1} & 0 \\ 0' & 0 \end{bmatrix}$$

Above, $D_1 = diag\,(\lambda_i)_{i=1}^{n-1}$, a diagonal matrix of dimension $n-1$ by $n-1$ that contains the nonzero eigenvalues λ_i. The remaining elements of D^- are all equal to zero.

3.3.3   A probabilistically equivalent reparameterisation of the random regression model

We describe a reparameterisation of the original two-trait model that simplifies the Markov chain Monte Carlo computations.

For each marker set (omitting the subscripts) define the row vector random variable $\alpha' = (\alpha_1', \alpha_2')$ of dimension 1 by $2n$, with vectors α_1 and α_2, each of dimension n by 1 associated with trait 1 and 2, with distribution

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \sim SN \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} D\sigma_{g_1}^2 & D\sigma_{g_1 g_2} \\ D\sigma_{g_1 g_2} & D\sigma_{g_2}^2 \end{bmatrix} \right) \tag{43}$$

where D is a diagonal matrix (associated with a particular marker set) of dimension $n \times n$, with has eigenvalues λ_i, $i = 1, \ldots, n$ as diagonal elements, of which the first $n-1$ are positive and the rest are equal to zero. We define for each marker set

$$V_g = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_1 g_2} \\ \sigma_{g_1 g_2} & \sigma_{g_1}^2 \end{bmatrix}. \tag{44}$$

In a particular marker set, the random variables

$$\begin{pmatrix} U\alpha_1 \\ U\alpha_2 \end{pmatrix} \sim SN \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} UDU'\sigma_{g_1}^2 & UDU'\sigma_{g_1 g_2} \\ UDU'\sigma_{g_1 g_2} & UDU'\sigma_{g_2}^2 \end{bmatrix} \right) \tag{45}$$

and

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \sim SN \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{p}WW'\sigma_{g_1}^2 & \frac{1}{p}WW'\sigma_{g_1 g_2} \\ \frac{1}{p}WW'\sigma_{g_1 g_2} & \frac{1}{p}WW'\sigma_{g_2}^2 \end{bmatrix} \right) \tag{46}$$

have the same distribution since $UDU' = \frac{1}{p}WW'$. Here p is the number of genetic markers in the marker set. Therefore the structure can be written as

$$y_1|\mu_1, \sum_{i=1}^{C} \alpha_{1i} \quad \sim \quad N \left(1\mu_1 + \sum_{i=1}^{C} U_i\alpha_{1i}, I\sigma_{1e}^2 \right), \tag{47}$$

$$y_2|\mu_2, \sum_{i=1}^{C} \alpha_{2i} \quad \sim \quad N \left(1\mu_2 + \sum_{i=1}^{C} U_i\alpha_{2i}, I\sigma_{2e}^2 \right) \quad i = 1, \ldots, C. \tag{48}$$

3.3.4 Prior distributions

The prior distribution of the $\mu's$ is $N\left(0, 10^5\right)$. The prior distributions of V_g and V_h are scaled inverted Wishart with degrees of freedom set equal to 2.5 and scale parameters P_g and P_h, respectively. The degrees of freedom generate a proper distribution with overdispersed values. The prior distributions of $\sigma_{r_m}^2$ and $\sigma_{r_f}^2$ are scale inverted chi square densities. The degrees of freedom of these densities are set equal to 1.0 which leads to vague prior information. For example, when the scale is equal to 0.1, the modal value of the prior distribution is 0.03 and the prior probability that the variance component is smaller than this value is 56. The prior probability that the variance component is between 0.03 and 0.3 is 29.

The scale parameters of all these distributions are estimated using maximum likelihood. This involved obtaining maximum likelihood estimates of the two-trait model (for Models H and G, each sex considered as one trait), and then equating these estimates to the mode of the relevant prior distribution, written as a function of the scale parameter. In the case of the general two-trait model, single-trait likelihoods were fitted instead. The off-diagonal elements of the scale parameter of inverse Wishart distributions were set equal to zero.

3.3.5 Fully conditional posterior distributions

The fully conditional posterior distributions of a parameter θ is denoted $[\theta|All, z]$ where All denotes all the parameters of the model except θ. Here we sketch the form of these densities.

3.3.5.1 Updating the $\alpha's$ with subscripts 1 for trait 1, 2 for trait 2 and k for marker set, from the bivariate normal distribution

$$\left[\begin{array}{c}\alpha_{ki1}\\\alpha_{ki2}\end{array}\middle|All, y\right] \sim N\left(\left[\begin{array}{c}\widehat{\alpha}_{ki1}\\\widehat{\alpha}_{ki2}\end{array}\right], \left[I + \frac{\sigma_e^2}{\lambda_{ki}}\left(\begin{array}{cc}\sigma_{g_1}^2 & \sigma_{g_1 g_2}\\\sigma_{g_1 g_2} & \sigma_{g_2}^2\end{array}\right)_k^{-1}\right]^{-1}\sigma_e^2\right), \quad i = 1, 2, \ldots, n-1,$$

where

$$\left(I + \lambda_{ki}^{-1} V_{g_k}^{-1}\sigma_e^2\right)\widehat{\alpha}_{is} = \left[\begin{array}{c}U_{ki}'\left(y_1 - 1\mu_1 - \sum_{i\neq k} U_i \alpha h_{i1}\right)\\U_{ki}'\left(y_2 - 1\mu_2 - \sum_{i\neq k} U_i \alpha_{i2}\right)\end{array}\right], \qquad \alpha_{is}' = (\alpha_{i1}, \alpha_{i2}).$$

The strategy updates jointly the $\alpha's$ for both traits from a given marker set, conditional on the $\alpha's$ from the remaining marker sets.

3.3.5.2 Updating μ_1 and μ_2 The prior distribution of both scalars μ_1 and μ_2 are the normal process $N\left(0, 10^5\right)$. The fully conditional for μ_1 is proportional to

$$[\mu_1|All, x] \sim N\left(\widehat{\mu}_1, \left(1'1 + k_{\mu_1}\right)^{-1}\sigma_e^2\right)$$

where

$$\left(1'1 + k_{\mu_1}\right)\widehat{\mu}_1 = 1'\left(y_m - U\alpha_m - T\gamma_1\right), \qquad k_{\mu_1} = \frac{\sigma_e^2}{10^5}.$$

A similar expression holds for μ_2.

3.3.5.3   Updating V_g   For each of the marker sets, the update involves drawing samples from scaled inverse Wishart distributions. The fully conditional of the 2×2 matrix V_g is proportional to

$$[\alpha_m, \alpha_f | V_g, D] \, [V_g | \nu_g, P_g] \,. \tag{49}$$

The second term is inverse Wishart the prior distribution of V_g, $IW\,(\nu_g, P_g)$ with density

$$p\,(V_g | \nu_g, P_g) \propto |Vg|^{-\frac{1}{2}(v_g+3)} \exp\left[-\frac{1}{2}tr\left(V_g^{-1}P_g\right)\right] \tag{50}$$

where the hyperparameters ν_g and P_g are the degrees of freedom and the scale, respectively. The modal value of this distribution is given by $P_g/(\nu_g + p + 1)$, where in our case, $p = 2$. On defining (see **?**, page 574)

$$S_g = \left[\begin{array}{cc} \alpha_1' D^{-1}\alpha_1 & \alpha_1' D^{-1}\alpha_2 \\ \alpha_2' D^{-1}\alpha_1 & \alpha_2' D^{-1}\alpha_2 \end{array} \right],$$

the density of the fully conditional posterior distribution of V_g is

$$\begin{aligned} p\,(V_g | All, y) \quad &\propto \quad |V_g|^{-\frac{k}{2}} |Vg|^{-\frac{1}{2}(v_g+3)} \exp\left[-\frac{1}{2}tr\left(V_g^{-1}P_g\right)\right] \exp\left[-\frac{1}{2}tr\left(V_g^{-1}S_g\right)\right] \\ &= \quad |Vg|^{-\frac{1}{2}(k+v_g+3)} \exp\left[-\frac{1}{2}tr\left[V_g^{-1}\left(S_g + P_g\right)\right]\right] \end{aligned} \tag{51}$$

where $k = n - 1$, which is in the form of an inverse Wishart distribution of dimension 2, $k + v_g$ degrees of freedom and scale matrix $(S_g + P_g)$.

3.3.5.4   Updating V_e   The update here is similar and it involves again drawing samples from scaled inverse Wishart distributions. The fully conditional posterior distribution of the 2×2 matrix V_e defined in **(??)** is proportional to

$$[e_1, e_2 | V_e, E] \, [V_e | \nu_e, P_e] \tag{52}$$

where the second term is the prior distribution of V_e with hyperparameters ν_e and P_e of the form

$$p\,(V_e | \nu_e, P_e) \propto |V_e|^{-\frac{1}{2}(v_e+3)} \exp\left[-\frac{1}{2}tr\left(V_e^{-1}P_e\right)\right] \tag{53}$$

symbolised $IW\,(\nu_e, P_e)$.

The density is

$$p\,(V_e | All, y) \propto |V_e|^{-\frac{1}{2}(n+v_e+3)} \exp\left[-\frac{1}{2}tr\left[V_e^{-1}\left(S_e + P_e\right)\right]\right],$$

an inverse Wishart distribution of dimension 2, $n + v_e$ degrees of freedom and scale matrix $(S_e + P_e)$, where

$$S_h = \left[\begin{array}{cc} \gamma_m' E^{-1}\gamma_m & \gamma_m' E^{-1}\gamma_f \\ \gamma_f' E^{-1}\gamma_m & \gamma_f' E^{-1}\gamma_f \end{array} \right].$$

References