
Bayesian multiple logistic regression for meta-analyses of GWAS – Supplemental Methods

Saikat Banerjee¹, Lingyao Zeng², Heribert Schunkert² and Johannes Söding^{1,*}

¹Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany

²German Heart Centre, 80636 Munich, Germany

*Corresponding author: soeding@mpibpc.mpg.de

1 Method details

1.1 Correction of population substructure

To correct for non-causal correlations of genotype with disease via population substructure, we performed a singular value decomposition (SVD) of the genotype matrix $\mathbf{X}' = \sum_s \lambda_s \mathbf{u}_s \mathbf{v}_s^T$. As discussed in the main text, \mathbf{X}' is the centered and normalized genotype matrix with N rows for each patient, and $I + 1$ columns for I SNPs and extra column with a value of 1 for the bias term v_0 . Here, λ_s are the singular values sorted in decreasing order, $\mathbf{v}_s \in \mathbb{R}^{I+1}$ the singular vectors or principal components and $\mathbf{u}_s \in \mathbb{R}^N$ are the *loadings* of the singular vectors. We store the loadings multiplied by the singular values of the first S (e.g. $S = 3 \dots 10$) principal components in a vector of covariates, $\mathbf{x}_n'' = (\lambda_1 u_{1,n}, \dots, \lambda_S u_{S,n})^T$ for each patient. We consider these as covariates \mathbf{X}'' and append them to the genotype matrix \mathbf{X}' . The total data matrix after the concatenation has rows $\mathbf{x}^T = (\mathbf{x}', \mathbf{x}'')^T \in \mathbb{R}^{I+S+1}$, and the covariates are treated as a separate locus. The number of loadings used for the population substructure correction can be controlled by using the `-pca` option in B-LORE. i

1.2 Notes on quasi-Laplace approximation

The quasi-Laplace approximation, as discussed in the main text, is a key concept in the methodology presented here. We approximate the product of the likelihood of \mathbf{v} and a *regularizer* with a Gaussian distribution.

$$p(\boldsymbol{\phi} \mid \mathbf{X}, \mathbf{v}) \mathcal{N}(\mathbf{v} \mid \tilde{\boldsymbol{\mu}}, \text{diag}(\tilde{\boldsymbol{\sigma}}^2)) \propto \mathcal{N}(\mathbf{v} \mid \tilde{\mathbf{v}}, \tilde{\boldsymbol{\Lambda}}^{-1}) \quad (1)$$

The logarithm of the left-hand side, which we call regularised log likelihood,

$$\text{LL}_{\text{reg}}(\mathbf{v} \mid \boldsymbol{\theta}, \mathbf{z}) = \log p(\boldsymbol{\phi} \mid \mathbf{X}, \mathbf{v}) + \log \mathcal{N}(\mathbf{v} \mid \tilde{\boldsymbol{\mu}}, \text{diag}(\tilde{\boldsymbol{\sigma}}^2)) \quad (2)$$

is a quadratic function with respect to any of the v_i , $\log \mathcal{N}(\mathbf{v} \mid \tilde{\boldsymbol{\mu}}, \text{diag}(\tilde{\boldsymbol{\sigma}}^2))$, plus the sum of N concave functions. The Hessians of these concave functions must all have negative (or zero) diagonal elements and therefore their sum will grow roughly proportionally with the number of patients N . In contrast to the second derivatives, the third and higher partial derivatives will take both positive and negative signs. If the number of diseased and control patients is roughly equal, $p(\phi_n \mid \mathbf{x}_n, \mathbf{v})$ will lie mostly around $(1/N) \sum_n \mathbb{I}(\phi_n = 1) \approx 0.5$, and therefore $\mathbf{v}^T \mathbf{x}_n$ will be roughly as often positive as negative. Consequently, the third partial derivatives will tend to be close to zero and have no preferred sign. The same is true for the higher derivatives. Their signs fluctuate around 0 for all patients n , hence the magnitudes of the third and higher derivatives will grow only as \sqrt{N} . The larger N is, the more the second derivatives will dominate over the higher derivatives and the better will the log likelihood be approximated by a quadratic function, or in other words, by the logarithm of a multivariate Gaussian.

If we are dealing with a GWA study with strongly differing numbers N_{dis} and N_{ctr} of disease and control patients, a simple reweighting of the control group patients can ensure that $p(\phi_n \mid \mathbf{x}_n, \mathbf{v})$

will lie around 0.5. We would multiply the contributions of the control patients to the log likelihood with weights $N_{\text{dis}}/N_{\text{ctr}}$.

1.3 Integration of the marginal likelihood

We optimize the hyperparameters by maximizing the marginal likelihood function, which is obtained by integrating over the effect sizes \mathbf{v} :

$$\begin{aligned}
mL(\theta) &= p(\phi \mid \mathbf{X}, \theta) \\
&= \int p(\phi \mid \mathbf{X}, \mathbf{v}) p(\mathbf{v} \mid \theta) d\mathbf{v} \\
&= \sum_{\mathbf{z}} p(\mathbf{z} \mid \theta) \int p(\phi \mid \mathbf{X}, \mathbf{v}) \mathcal{N}(\mathbf{v} \mid \mu_{\mathbf{z}}, \text{diag}(\sigma_{\mathbf{z}}^2)) d\mathbf{v} \\
&= \sum_{\mathbf{z}} p(\mathbf{z} \mid \theta) \int p(\phi \mid \mathbf{X}, \mathbf{v}) \mathcal{N}(\mathbf{v} \mid \tilde{\mu}, \text{diag}(\tilde{\sigma}^2)) \frac{\mathcal{N}(\mathbf{v} \mid \mu_{\mathbf{z}}, \text{diag}(\sigma_{\mathbf{z}}^2))}{\mathcal{N}(\mathbf{v} \mid \tilde{\mu}, \text{diag}(\tilde{\sigma}^2))} d\mathbf{v} \\
&\propto \sum_{\mathbf{z}} p(\mathbf{z} \mid \theta) \int \mathcal{N}(\mathbf{v} \mid \tilde{\mathbf{v}}, \tilde{\Lambda}^{-1}) \frac{\mathcal{N}(\mathbf{v} \mid \mu_{\mathbf{z}}, \text{diag}(\sigma_{\mathbf{z}}^2))}{\mathcal{N}(\mathbf{v} \mid \tilde{\mu}, \text{diag}(\tilde{\sigma}^2))} d\mathbf{v} \tag{3}
\end{aligned}$$

where we have used the quasi-Laplace approximation of Eq. (1) in the last step. We can analytically evaluate the integral in Eq. (3) because the logarithm of the integrand is now a quadratic function of \mathbf{v} . We make use of the following equality, which we prove in Appendix A:

$$\int \mathcal{N}(\mathbf{v} \mid \mu_1, \Lambda_1^{-1}) \frac{\mathcal{N}(\mathbf{v} \mid \mu_2, \Lambda_2^{-1})}{\mathcal{N}(\mathbf{v} \mid \mu_3, \Lambda_3^{-1})} d\mathbf{v} = \frac{\mathcal{N}(\mathbf{0} \mid \mu_1, \Lambda_1^{-1}) \mathcal{N}(\mathbf{0} \mid \mu_2, \Lambda_2^{-1})}{\mathcal{N}(\mathbf{0} \mid \mu_{123}, \Lambda_{123}^{-1}) \mathcal{N}(\mathbf{0} \mid \mu_3, \Lambda_3^{-1})} \tag{4}$$

where

$$\Lambda_{123} := \Lambda_1 + \Lambda_2 - \Lambda_3 \tag{5}$$

$$\mu_{123} := \Lambda_{123}^{-1} (\Lambda_1 \mu_1 + \Lambda_2 \mu_2 - \Lambda_3 \mu_3) \tag{6}$$

Identifying $\Lambda_1 = \tilde{\Lambda}$, $\Lambda_2 = \text{diag}(\lambda_{\mathbf{z}})$, $\Lambda_3 = \text{diag}(\tilde{\lambda})$, $\mu_1 = \tilde{\mathbf{v}}$, $\mu_2 = \mu_{\mathbf{z}}$, $\mu_3 = \tilde{\mu}$ and defining

$$\Lambda_{\mathbf{z}} := \tilde{\Lambda} + \text{diag}(\lambda_{\mathbf{z}}) - \text{diag}(\tilde{\lambda}) = \Lambda_{123} \tag{7}$$

$$\mathbf{v}_{\mathbf{z}} := \Lambda_{\mathbf{z}}^{-1} (\tilde{\Lambda} \tilde{\mathbf{v}} + \text{diag}(\lambda_{\mathbf{z}}) \mu_{\mathbf{z}} - \text{diag}(\tilde{\lambda}) \tilde{\mu}) = \mu_{123} \tag{8}$$

we obtain

$$mL(\theta) \approx \sum_{\mathbf{z}} p(\mathbf{z} \mid \theta) \frac{\mathcal{N}(\mathbf{0} \mid \tilde{\mathbf{v}}, \tilde{\Lambda}^{-1}) \mathcal{N}(\mathbf{0} \mid \mu_{\mathbf{z}}, \text{diag}(\lambda_{\mathbf{z}}^{-1}))}{\mathcal{N}(\mathbf{0} \mid \mathbf{v}_{\mathbf{z}}, \Lambda_{\mathbf{z}}^{-1}) \mathcal{N}(\mathbf{0} \mid \tilde{\mu}, \text{diag}(\tilde{\lambda}^{-1}))} \tag{9}$$

The upper left and lower right Gaussians do not depend on θ or \mathbf{z} and can be pulled into a constant factor D' , yielding

$$mL(\theta) = D' \sum_{\mathbf{z}} p(\mathbf{z} \mid \theta) \frac{|\text{diag}(\lambda_{\mathbf{z}})|^{\frac{1}{2}}}{|\Lambda_{\mathbf{z}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \mu_{\mathbf{z}}^{\top} \text{diag}(\lambda_{\mathbf{z}}) \mu_{\mathbf{z}} + \frac{1}{2} \mathbf{v}_{\mathbf{z}}^{\top} \Lambda_{\mathbf{z}} \mathbf{v}_{\mathbf{z}}\right) \tag{10}$$

1.4 Estimation of the regularization parameters $\tilde{\mu}$ and $\tilde{\sigma}$

We use an iterative estimation procedure to find optimum regularisation parameters $\tilde{\mu}$ and $\tilde{\sigma}$. For the quasi-Laplace approximation to hold well, we need to find regularisation parameters that make the regulariser almost as good a prior as the full two-component mixtures prior. This can be achieved by optimising (with respect to the regularisation parameters) a simplified marginal likelihood in which the regulariser replaces the full prior.

We start by setting the regularisation parameters to $\tilde{\mu} = \mathbf{0}$ and $\tilde{\sigma} = \sigma_{\text{reg}} \mathbf{1} = 0.011$, corresponding to a coupling of a causal SNP with $v_i = 0.2$ distributed over 20 strongly correlated SNPs with coupling $v_i = 0.01$ each. We then find improved estimates of $\tilde{\mu}$ and $\tilde{\sigma}$ by setting $\mu_{\mathbf{z}} := \mu \mathbf{1}$ and $\lambda_{\mathbf{z}} := \sigma^{-2} \mathbf{1}$ – which corresponds to only one single causality configuration $\mathbf{z} = \mathbf{1}$ in which all SNPs are causal – and maximising the likelihood in Eq. (10) with respect to μ and σ . We iterate a few times using the new, updated estimates of μ and σ to update the regulariser parameters $\tilde{\mu}$ and $\tilde{\sigma}$ and reestimating new, improved values for μ and σ .

1.5 Summary statistics: Optimization of the regularised likelihood

As discussed in the main text, we intend to learn the mode $\tilde{\mathbf{v}}$ and precision $\tilde{\Lambda}$ of the regularized likelihood given by Eq. (1). Once we know the regularization parameters $\tilde{\mu}$ and $\tilde{\sigma}$, we maximize the regularized log likelihood of Eq. (2) with respect to \mathbf{v} using the gradient-based optimisation method L-BFGS. The partial derivatives are:

$$\frac{\partial}{\partial v_i} \text{LL}_{\text{reg}}(\mathbf{v} \mid \boldsymbol{\theta}, \mathbf{z}) = \sum_{n=1}^N (\phi_n - p_n) x_{ni} - \tilde{\lambda}_i v_i \quad (11)$$

where $p_n = p(\phi_n = 1 \mid \mathbf{x}_n, \mathbf{v})$. The solution of this optimisation, $\tilde{\mathbf{v}}$, is the mean and mode of the Gaussian. The term $-\tilde{\lambda}_i v_i$ pulls v_i towards zero. It prevents the maximum likelihood solutions v_i^* to assume large values in the absence of strong evidence for a SNP-disease coupling, *e.g.* when two SNPs are in near-perfect linkage disequilibrium and therefore very highly anti-correlated.

To find the precision matrix, we note that it should be equal to the negative Hessian matrix $-\mathbf{H}$ of the regularised log likelihood at $\tilde{\mathbf{v}}$. Using the derivative of the logistic function $\text{lf}(x) = 1/(1 + e^{-x})$ given by,

$$\frac{d \text{lf}(x)}{dx} = \text{lf}(x)(1 - \text{lf}(x)), \quad (12)$$

we obtain the matrix elements of the Hessian \mathbf{H} ,

$$\begin{aligned} \frac{\partial^2}{\partial v_i \partial v_j} \text{LL}_{\text{reg}}(\tilde{\mathbf{v}}) &= - \sum_{n=1}^N p(\phi_n = 1 \mid \mathbf{x}_n, \tilde{\mathbf{v}}) p(\phi_n = 0 \mid \mathbf{x}_n, \tilde{\mathbf{v}}) x_{ni} x_{nj} - \tilde{\lambda}_i \delta_{ij} \\ &= - \sum_{n=1}^N \tilde{p}_n (1 - \tilde{p}_n) x_{ni} x_{nj} - \tilde{\lambda}_i \delta_{ij} \end{aligned} \quad (13)$$

where $\tilde{p}_n = p(\phi_n = 1 \mid \mathbf{x}_n, \tilde{\mathbf{v}})$. The equation shows that strongly correlated SNPs will have high coupling coefficients in the Hessian matrix. The precision matrix of the Gaussian distribution in our quasi-Laplace approximation is therefore,

$$\tilde{\Lambda} = \sum_{n=1}^N \tilde{p}_n (1 - \tilde{p}_n) \mathbf{x}_n \mathbf{x}_n^T + \text{diag}(\tilde{\lambda}) \quad (14)$$

which is the sum of two matrices, one proportional to a weighted sample covariance matrix of the x_1, \dots, x_I and the precision matrix of the regularisation prior.

1.6 Factorization over loci

If the covariance matrix of the genotype $\mathbf{X}^T \mathbf{X}$ is block-diagonal, as discussed in the main text, then it is obvious from Eq. (14) that $\tilde{\Lambda}$ is also block-diagonal. This allows us to factorise the marginal likelihood in equation Eq. (10), writing \mathbf{z}_l for the binary configuration vector corresponding to the I_l SNPs of locus l and $\mu_{\mathbf{z},l}$, $\lambda_{\mathbf{z},l}$, $\Lambda_{\mathbf{z},l}$, $\mathbf{v}_{\mathbf{z},l}$ for the subvectors and submatrices corresponding to only the SNPs of locus l :

$$mL(\boldsymbol{\theta}) = D' \sum_{\mathbf{z}} \prod_{l=1}^{L+1} p(\mathbf{z}_l \mid \boldsymbol{\theta}) \frac{|\text{diag}(\lambda_{\mathbf{z},l})|^{\frac{1}{2}}}{|\Lambda_{\mathbf{z},l}|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} \mu_{\mathbf{z},l}^T \text{diag}(\lambda_{\mathbf{z},l}) \mu_{\mathbf{z},l} + \frac{1}{2} \mathbf{v}_{\mathbf{z},l}^T \Lambda_{\mathbf{z},l} \mathbf{v}_{\mathbf{z},l} \right) \quad (15)$$

For notational brevity, we defined a virtual locus $l = L + 1$ for the confounding variables, $\mathbf{x}_{L+1} = \mathbf{x}''$. We allow only one configuration $\mathbf{z}_{L+1} := \mathbf{1}$ and define $p(\mathbf{z}_{L+1} | \boldsymbol{\theta}) := 1$, $\boldsymbol{\mu}_{\mathbf{z}_{L+1}} := \mathbf{0}$ and $\boldsymbol{\Lambda}_{\mathbf{z}_{L+1}} := \boldsymbol{\Lambda}''$. Hence this virtual locus is treated exactly the same as the real loci, except that the sum over \mathbf{z} runs only over real loci while $\mathbf{z}_{L+1} := \mathbf{1}$.

Finally, we denote all possible configurations in locus l as \mathbf{z}_l , which allows us to write

$$mL(\boldsymbol{\theta}) = p(\boldsymbol{\phi} | \mathbf{X}, \boldsymbol{\theta}) = D' \prod_{l=1}^{L+1} \sum_{\mathbf{z}_l} p(\mathbf{z}_l | \boldsymbol{\theta}) F_l(\mathbf{z}_l, \boldsymbol{\theta}) \quad (16)$$

with

$$F_l(\mathbf{z}_l, \boldsymbol{\theta}) = \frac{|\text{diag}(\boldsymbol{\Lambda}_{\mathbf{z}_l})|^{\frac{1}{2}}}{|\boldsymbol{\Lambda}_{\mathbf{z}_l}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \boldsymbol{\mu}_{\mathbf{z}_l}^\top \text{diag}(\boldsymbol{\Lambda}_{\mathbf{z}_l}) \boldsymbol{\mu}_{\mathbf{z}_l} + \frac{1}{2} \mathbf{v}_{\mathbf{z}_l}^\top \boldsymbol{\Lambda}_{\mathbf{z}_l} \mathbf{v}_{\mathbf{z}_l}\right) \quad (17)$$

1.7 Inferences

1.7.1 Finemapping

It follows from the derivation of the marginal likelihood in Eq. (16) that

$$p(\boldsymbol{\phi}, \mathbf{z} | \mathbf{X}, \boldsymbol{\theta}) = D' \prod_{l=1}^{L+1} p(\mathbf{z}_l | \boldsymbol{\theta}) F_l(\mathbf{z}_l, \boldsymbol{\theta}) \quad (18)$$

Using the definition of the conditional probability, we conclude

$$\begin{aligned} p(\mathbf{z} | \boldsymbol{\phi}, \mathbf{X}, \boldsymbol{\theta}) &= \frac{p(\boldsymbol{\phi}, \mathbf{z} | \mathbf{X}, \boldsymbol{\theta})}{\sum_{\mathbf{z}'} p(\boldsymbol{\phi}, \mathbf{z}' | \mathbf{X}, \boldsymbol{\theta})} \\ &= \frac{\prod_{l=1}^{L+1} p(\mathbf{z}_l | \boldsymbol{\theta}) F_l(\mathbf{z}_l, \boldsymbol{\theta})}{\prod_{l=1}^{L+1} \sum_{\mathbf{z}'} p(\mathbf{z}_l | \boldsymbol{\theta}) F_l(\mathbf{z}_l, \boldsymbol{\theta})} \\ &= \prod_{l=1}^{L+1} \frac{p(\mathbf{z}_l | \boldsymbol{\theta}) F_l(\mathbf{z}_l, \boldsymbol{\theta})}{\sum_{\mathbf{z}'} p(\mathbf{z}_l | \boldsymbol{\theta}) F_l(\mathbf{z}_l, \boldsymbol{\theta})} \end{aligned} \quad (19)$$

Therefore, the posterior probability for SNP i to be causative is

$$p(z_i = 1 | \boldsymbol{\phi}, \mathbf{X}, \boldsymbol{\theta}) = \sum_{\mathbf{z}: z_i=1} p(\mathbf{z} | \boldsymbol{\phi}, \mathbf{X}, \boldsymbol{\theta}) \quad (20)$$

1.7.2 Posterior probability for causality of loci

The probability for a locus to be causally associated with the disease is equal to the probability of the locus harbouring at least one causally associated SNP, which is equal to 1 minus the probability of not containing a single causal SNP:

$$\begin{aligned} \text{Pr}_{\text{causal}} &= p(\text{locus is causal} | \boldsymbol{\phi}, \mathbf{X}, \boldsymbol{\theta}) \\ &= 1 - p(\mathbf{z} = \mathbf{0} | \boldsymbol{\phi}, \mathbf{X}, \boldsymbol{\theta}) \\ &= 1 - \prod_{l=1}^{L+1} \frac{p(\mathbf{z}_l = \mathbf{0} | \boldsymbol{\theta}) F_l(\mathbf{z}_l = \mathbf{0}, \boldsymbol{\theta})}{\sum_{\mathbf{z}'} p(\mathbf{z}_l | \boldsymbol{\theta}) F_l(\mathbf{z}_l, \boldsymbol{\theta})} \end{aligned} \quad (21)$$

2 Efficient computation

2.1 Setting σ_{bg} to zero

We will show here that in the limit of $\sigma_{\text{bg}} \rightarrow 0$, Eqs. (16) and (17) remain valid if we replace the vectors with their subvectors containing only those components i for which $z_i = 1$: $\mu_{\mathbf{z}_l} \rightarrow \hat{\mu}_{\mathbf{z}_l} \in \mathbb{R}^{\|\mathbf{z}_l\|_1}$, $\lambda_{\mathbf{z}_l} \rightarrow \hat{\lambda}_{\mathbf{z}_l}$, $\mathbf{v}_{\mathbf{z}_l} \rightarrow \hat{\mathbf{v}}_{\mathbf{z}_l}$, and analogously, $\Lambda_{\mathbf{z}_l} \rightarrow \hat{\Lambda}_{\mathbf{z}_l} \in \mathbb{R}^{\|\mathbf{z}_l\|_1 \times \|\mathbf{z}_l\|_1}$.

To prove this, let us first prove that when $\sigma_{\text{bg}} \rightarrow 0$ then $\frac{|\text{diag}(\lambda_{\mathbf{z}_l})|}{|\Lambda_{\mathbf{z}_l}|} \rightarrow \frac{|\text{diag}(\hat{\lambda}_{\mathbf{z}_l})|}{|\hat{\Lambda}_{\mathbf{z}_l}|}$. The i^{th} element of the vector $\lambda_{\mathbf{z}_l}$ will be $\lambda_{\mathbf{z}_l, i} = \sigma_{\text{bg}}^{-2} \rightarrow \infty$ for all i with $z_i = 0$. From Eq. (7), we note that the diagonal entries in $\Lambda_{\mathbf{z}_l}$ that go to infinity will dominate the sum over all permutations in the determinant as $\lambda_{\mathbf{z}_l, i} \rightarrow \infty$ for all i with $z_i = 0$. Hence, we can separate them out: $|\Lambda_{\mathbf{z}_l}| \rightarrow |\hat{\Lambda}_{\mathbf{z}_l}| \prod_{i: z_i=0} \sigma_{\text{bg}}^{-2}$. The same can be done with $|\text{diag}(\lambda_{\mathbf{z}_l})| \rightarrow |\text{diag}(\hat{\lambda}_{\mathbf{z}_l})| \prod_{i: z_i=0} \sigma_{\text{bg}}^{-2}$, and dividing both terms proves the first part.

Second, we need to prove that

$$\begin{aligned} \mu_{\mathbf{z}_l}^\top \text{diag}(\lambda_{\mathbf{z}_l}) \mu_{\mathbf{z}_l} &\rightarrow \hat{\mu}_{\mathbf{z}_l}^\top \text{diag}(\hat{\lambda}_{\mathbf{z}_l}) \hat{\mu}_{\mathbf{z}_l} \\ \mathbf{v}_{\mathbf{z}_l}^\top \Lambda_{\mathbf{z}_l} \mathbf{v}_{\mathbf{z}_l} &\rightarrow \hat{\mathbf{v}}_{\mathbf{z}_l}^\top \hat{\Lambda}_{\mathbf{z}_l} \hat{\mathbf{v}}_{\mathbf{z}_l} \end{aligned} \quad (22)$$

The first line follows from the fact that for all i with $z_i = 0$, we have $\mu_{\mathbf{z}_l, i} = 0$ and hence those components do not contribute anything on the left hand-side. To prove the second line, we use Eqs. (7) and (8) to write

$$\begin{aligned} \mathbf{v}_{\mathbf{z}_l}^\top \Lambda_{\mathbf{z}_l} \mathbf{v}_{\mathbf{z}_l} &= \left(\tilde{\mathbf{v}}_l^\top \tilde{\Lambda}_l + \text{diag}(\lambda_{\mathbf{z}_l}) \mu_{\mathbf{z}_l} - \text{diag}(\tilde{\lambda}_l) \tilde{\mu}_l \right) \times \\ &\quad \left(\tilde{\Lambda}_l + \text{diag}(\lambda_{\mathbf{z}_l}) - \text{diag}(\tilde{\lambda}_l) \right)^{-1} \left(\tilde{\Lambda}_l \tilde{\mathbf{v}}_l + \text{diag}(\lambda_{\mathbf{z}_l}) \mu_{\mathbf{z}_l} - \text{diag}(\tilde{\lambda}_l) \tilde{\mu}_l \right) \end{aligned} \quad (23)$$

The inverted matrix in the center of the right hand-side will force contributions from all components with $z_i = 0$ to zero. What remains are all other components, which is just the expression on the right hand-side of the second line.

As a consequence of setting σ_{bg} to zero, the complexity of a single gradient computation is reduced by an order of I^2 leading to a huge speed up.

2.2 Branch and bound algorithm to restrict the sum over \mathbf{z} to non-negligible terms

The sums over \mathbf{z}_l in Eq. (16) run over 2^{I_l} terms. In the following we propose a method that omits terms in the sum over \mathbf{z}_l that do not stand a chance of contributing significantly to it. We denote the total number of significant configurations chosen for locus l as C_l . For any given configuration \mathbf{z}_l , the number of causal SNPs is simply the norm of the \mathbf{z} -state, given by $\|\mathbf{z}_l\| = k$. We progressively increase the allowed number of causal SNPs k upto an allowed maximum of k_{max} . At every step of $\|\mathbf{z}_l\| = k$, the total number of possible configurations is given by C_k , out of which we select only significant configurations C'_k to be appended to C_l .

We define the unnormalised posterior probabilities according to equation Eq. (19) as

$$\tilde{p}(\mathbf{z}_l \mid \phi, \mathbf{X}, \theta) := p(\mathbf{z}_l \mid \theta) F_l(\mathbf{z}_l, \theta) \quad (24)$$

The algorithm is initialized with $k = 0$ and $k = 1$:

$$\begin{aligned} S_0 &= \tilde{p}(\mathbf{z}_l = \mathbf{0} \mid \phi, \mathbf{X}, \theta) \\ S_1 &= \sum_{\|\mathbf{z}_l\|=1} \tilde{p}(\mathbf{z}_l \mid \phi, \mathbf{X}, \theta) \\ C_1 &= C'_1 = \{\mathbf{z} \in \{0, 1\}^{I_l} : \|\mathbf{z}_l\| = 1\} \end{aligned} \quad (25)$$

At every step with $k > 1$ we use all the possible configurations \mathcal{C}_k to compute

$$S_k = \sum_{\mathbf{z}_l \in \mathcal{C}_k} \tilde{p}(\mathbf{z}_l | \boldsymbol{\phi}, \mathbf{X}, \boldsymbol{\theta}) \quad (26)$$

To determine the significant configurations \mathcal{C}'_k , we sort all the \mathcal{C}_k configurations by decreasing posterior probability and keep appending to \mathcal{C}'_k until

$$\sum_{\mathbf{z}_l \in \mathcal{C}'_k} \tilde{p}(\mathbf{z}_l | \boldsymbol{\phi}, \mathbf{X}, \boldsymbol{\theta}) \geq 0.98 S_k \quad (27)$$

We combine these \mathcal{C}'_k configurations with all the $I_l - k$ SNPs into new \mathcal{C}_{k+1} configurations. We again select \mathcal{C}'_{k+1} configurations out of all the available \mathcal{C}_{k+1} configurations, and this continues iteratively. We stop the iteration when $S_{k+1} < 0.02 S_k$, because the terms with even higher $\|\mathbf{z}_l\|$ will not contribute significantly to the total sum over all configurations. The posterior probabilities are then approximated as

$$p(\mathbf{z}_l | \boldsymbol{\phi}, \mathbf{X}, \boldsymbol{\theta}) = \frac{\tilde{p}(\mathbf{z}_l | \boldsymbol{\phi}, \mathbf{X}, \boldsymbol{\theta})}{\sum_{k=0}^{k_{\max}} S_k} \quad (28)$$

Since risk loci will usually only contain a few causal SNPs, this procedure should normally stop without needing to compute millions of terms.

Appendix A

We show that the following result holds for any three D -dimensional Gaussian distributions $\mathcal{N}(\mathbf{v} | \boldsymbol{\mu}_1, \boldsymbol{\Lambda}_1^{-1})$, $\mathcal{N}(\mathbf{v} | \boldsymbol{\mu}_2, \boldsymbol{\Lambda}_2^{-1})$, and $\mathcal{N}(\mathbf{v} | \boldsymbol{\mu}_3, \boldsymbol{\Lambda}_3^{-1})$,

$$\int \mathcal{N}(\mathbf{v} | \boldsymbol{\mu}_1, \boldsymbol{\Lambda}_1^{-1}) \frac{\mathcal{N}(\mathbf{v} | \boldsymbol{\mu}_2, \boldsymbol{\Lambda}_2^{-1})}{\mathcal{N}(\mathbf{v} | \boldsymbol{\mu}_3, \boldsymbol{\Lambda}_3^{-1})} d\mathbf{v} = \frac{\mathcal{N}(\mathbf{0} | \boldsymbol{\mu}_1, \boldsymbol{\Lambda}_1^{-1}) \mathcal{N}(\mathbf{0} | \boldsymbol{\mu}_2, \boldsymbol{\Lambda}_2^{-1})}{\mathcal{N}(\mathbf{0} | \boldsymbol{\mu}_{123}, \boldsymbol{\Lambda}_{123}^{-1}) \mathcal{N}(\mathbf{0} | \boldsymbol{\mu}_3, \boldsymbol{\Lambda}_3^{-1})} \quad (29)$$

where

$$\boldsymbol{\Lambda}_{123} := \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_3 \quad (30)$$

$$\boldsymbol{\mu}_{123} := \boldsymbol{\Lambda}_{123}^{-1} (\boldsymbol{\Lambda}_1 \boldsymbol{\mu}_1 + \boldsymbol{\Lambda}_2 \boldsymbol{\mu}_2 - \boldsymbol{\Lambda}_3 \boldsymbol{\mu}_3) \quad (31)$$

We start by writing out the Gaussian functions explicitly:

$$\begin{aligned} \int \mathcal{N}(\mathbf{v} | \boldsymbol{\mu}_1, \boldsymbol{\Lambda}_1^{-1}) \frac{\mathcal{N}(\mathbf{v} | \boldsymbol{\mu}_2, \boldsymbol{\Lambda}_2^{-1})}{\mathcal{N}(\mathbf{v} | \boldsymbol{\mu}_3, \boldsymbol{\Lambda}_3^{-1})} d\mathbf{v} &= \frac{|\boldsymbol{\Lambda}_1|^{\frac{1}{2}} |\boldsymbol{\Lambda}_2|^{\frac{1}{2}}}{|\boldsymbol{\Lambda}_3|^{\frac{1}{2}} (2\pi)^{\frac{D}{2}}} \\ &\times \int \exp\left(-\frac{1}{2} [(\mathbf{v} - \boldsymbol{\mu}_1)^T \boldsymbol{\Lambda}_1 (\mathbf{v} - \boldsymbol{\mu}_1) + (\mathbf{v} - \boldsymbol{\mu}_2)^T \boldsymbol{\Lambda}_2 (\mathbf{v} - \boldsymbol{\mu}_2) - (\mathbf{v} - \boldsymbol{\mu}_3)^T \boldsymbol{\Lambda}_3 (\mathbf{v} - \boldsymbol{\mu}_3)]\right) d\mathbf{v} \end{aligned} \quad (32)$$

To perform the integral, we need to write the terms within the exp function as a Gaussian. For that purpose, we need to sort them into quadratic, linear and constant terms in \mathbf{v} ,

$$\begin{aligned} &(\mathbf{v} - \boldsymbol{\mu}_1)^T \boldsymbol{\Lambda}_1 (\mathbf{v} - \boldsymbol{\mu}_1) + (\mathbf{v} - \boldsymbol{\mu}_2)^T \boldsymbol{\Lambda}_2 (\mathbf{v} - \boldsymbol{\mu}_2) - (\mathbf{v} - \boldsymbol{\mu}_3)^T \boldsymbol{\Lambda}_3 (\mathbf{v} - \boldsymbol{\mu}_3) \\ &= \mathbf{v}^T (\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_3) \mathbf{v} - 2 \left(\boldsymbol{\mu}_1^T \boldsymbol{\Lambda}_1 + \boldsymbol{\mu}_2^T \boldsymbol{\Lambda}_2 - \boldsymbol{\mu}_3^T \boldsymbol{\Lambda}_3 \right) \mathbf{v} + \boldsymbol{\mu}_1^T \boldsymbol{\Lambda}_1 \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^T \boldsymbol{\Lambda}_2 \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3^T \boldsymbol{\Lambda}_3 \boldsymbol{\mu}_3 \\ &= \mathbf{v}^T \boldsymbol{\Lambda}_{123} \mathbf{v} - 2 \boldsymbol{\mu}_{123}^T \boldsymbol{\Lambda}_{123} \mathbf{v} + \boldsymbol{\mu}_1^T \boldsymbol{\Lambda}_1 \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^T \boldsymbol{\Lambda}_2 \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3^T \boldsymbol{\Lambda}_3 \boldsymbol{\mu}_3 \\ &= \left(\mathbf{v}^T - \boldsymbol{\mu}_{123}^T \right) \boldsymbol{\Lambda}_{123} (\mathbf{v} - \boldsymbol{\mu}_{123}) - \boldsymbol{\mu}_{123}^T \boldsymbol{\Lambda}_{123} \boldsymbol{\mu}_{123} + \boldsymbol{\mu}_1^T \boldsymbol{\Lambda}_1 \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^T \boldsymbol{\Lambda}_2 \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3^T \boldsymbol{\Lambda}_3 \boldsymbol{\mu}_3 \end{aligned}$$

We can insert this expression into Eq. (32) and perform the integration over the Gaussian function,

$$\begin{aligned}
& \int \mathcal{N}(\mathbf{v} \mid \boldsymbol{\mu}_1, \boldsymbol{\Lambda}_1^{-1}) \frac{\mathcal{N}(\mathbf{v} \mid \boldsymbol{\mu}_2, \boldsymbol{\Lambda}_2^{-1})}{\mathcal{N}(\mathbf{v} \mid \boldsymbol{\mu}_3, \boldsymbol{\Lambda}_3^{-1})} d\mathbf{v} \\
&= \frac{|\boldsymbol{\Lambda}_1|^{\frac{1}{2}} |\boldsymbol{\Lambda}_2|^{\frac{1}{2}}}{|\boldsymbol{\Lambda}_3|^{\frac{1}{2}} (2\pi)^{\frac{D}{2}}} \times \int \exp\left(-\frac{1}{2} (\mathbf{v}^\top - \boldsymbol{\mu}_{123}^\top) \boldsymbol{\Lambda}_{123} (\mathbf{v} - \boldsymbol{\mu}_{123})\right) d\mathbf{v} \\
&\quad \times \exp\left(-\frac{1}{2} [\boldsymbol{\mu}_1^\top \boldsymbol{\Lambda}_1 \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^\top \boldsymbol{\Lambda}_2 \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3^\top \boldsymbol{\Lambda}_3 \boldsymbol{\mu}_3 - \boldsymbol{\mu}_{123}^\top \boldsymbol{\Lambda}_{123} \boldsymbol{\mu}_{123}]\right) \\
&= \left(\frac{|\boldsymbol{\Lambda}_1| |\boldsymbol{\Lambda}_2|}{|\boldsymbol{\Lambda}_{123}| |\boldsymbol{\Lambda}_3|}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2} [\boldsymbol{\mu}_1^\top \boldsymbol{\Lambda}_1 \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^\top \boldsymbol{\Lambda}_2 \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3^\top \boldsymbol{\Lambda}_3 \boldsymbol{\mu}_3 - \boldsymbol{\mu}_{123}^\top \boldsymbol{\Lambda}_{123} \boldsymbol{\mu}_{123}]\right) \\
&= \frac{\mathcal{N}(\mathbf{0} \mid \boldsymbol{\mu}_1, \boldsymbol{\Lambda}_1^{-1}) \mathcal{N}(\mathbf{0} \mid \boldsymbol{\mu}_2, \boldsymbol{\Lambda}_2^{-1})}{\mathcal{N}(\mathbf{0} \mid \boldsymbol{\mu}_{123}, \boldsymbol{\Lambda}_{123}^{-1}) \mathcal{N}(\mathbf{0} \mid \boldsymbol{\mu}_3, \boldsymbol{\Lambda}_3^{-1})} \tag{33}
\end{aligned}$$

This proves the proposition in Eq. (29).