

Appendix S1

Manuel A. Rivas and Matti Pirinen

January 28, 2018

1 MRP model comparison for association testing

We consider the multivariate linear regression model

$$\underset{(N \times K)}{\mathbf{Y}} = \underset{(N \times K)}{\mathbf{\Psi}} + \underset{(N \times M)(M \times K)}{\mathbf{X}} \underset{(M \times K)}{\mathbf{B}} + \underset{(N \times K)}{\mathbf{E}},$$

where the matrices $\mathbf{Y} = [y_{ik}]$, $\mathbf{X} = [x_{im}]$, $\mathbf{B} = [\beta_{mk}]$ and $\mathbf{E} = [e_{ik}]$ describe the phenotype values (y_{ik}), copies of minor allele (x_{im}), variant-phenotype effects (β_{mk}), and residual errors (e_{ik}), for individual i , phenotype k , and variant m . We assume that each phenotype has been transformed to a standard normal distribution and that the columns of \mathbf{X} have been centered, which means that the estimate for the intercept term $\mathbf{\Psi}$ is 0 and independent of the estimate of \mathbf{B} . We use vectorized notation where the rows of \mathbf{B} form vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_M)^\top$ of length MK .

We define the MRP model comparison as a Bayes factor (BF) between the alternative model, where at least one variant affects at least one phenotype, and the null model, where all variant-phenotype effects are zero. BF is the ratio of the marginal likelihoods for these two models:

$$\text{BF} = \frac{\int_{\boldsymbol{\beta}} p(\text{Data}|\boldsymbol{\beta}) p(\boldsymbol{\beta}|\text{ALT}) d\boldsymbol{\beta}}{\int_{\boldsymbol{\beta}} p(\text{Data}|\boldsymbol{\beta}) p(\boldsymbol{\beta}|\text{NULL}) d\boldsymbol{\beta}},$$

where Data can correspond either to the effect size estimates $\hat{\boldsymbol{\beta}}$ and the estimated variance-covariance matrix of $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{V}}_{\boldsymbol{\beta}}$, or to the original phenotypes and genotypes, $\underset{(N \times K)}{\mathbf{Y}}$ and $\underset{(N \times M)}{\mathbf{X}}$, and any other covariates that we want to regress out from the phenotypes.

The prior distribution for the null model, $p(\boldsymbol{\beta}|\text{NULL})$, is simply the point mass at $\boldsymbol{\beta} = 0$. In section 2 we show how we approximate the likelihood function for $\boldsymbol{\beta}$, $p(\text{Data}|\boldsymbol{\beta})$, in section 3 we define the prior distribution $p(\boldsymbol{\beta}|\text{ALT})$ for the alternative model, and finally, in section 4, we compute the BF.

2 Likelihood function

A maximum likelihood estimator of \mathbf{B} is given by the ordinary least-squares method

$$\hat{\mathbf{B}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y},$$

that in vectorized form is denoted by $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_M)^\top$. An estimator of the variance-covariance of $\hat{\boldsymbol{\beta}}$ is given by

$$\hat{\mathbf{V}}_{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \otimes \hat{\mathbf{V}}_{\mathbf{Y}},$$

where $\hat{\mathbf{V}}_{\mathbf{Y}}$ is the estimated residual variance-covariance matrix of \mathbf{Y} given \mathbf{X} .

Following Band et al. [1], we approximate the likelihood function of β by a multivariate normal distribution with mean $\hat{\beta}$ and variance-covariance matrix \hat{V}_β . Note that by approximating \hat{V}_Y by the trait correlation matrix, this likelihood approximation does not require access to the individual level data X and Y but only to the summary data of effect sizes $\hat{\beta}$, LD-matrix $X^T X$ and a trait correlation estimate.

3 Prior of β in the alternative model

We construct the prior distribution $p(\beta|\text{ALT})$ for the alternative model in three steps allowing user to specify correlations between effects of different variants on different traits across different studies.

In a single study, the prior density for β incorporates the expected correlation of genetic effects among a group of variants (\mathbf{R}_{var}) and among a group of phenotypes (\mathbf{R}_{phen}). In addition, we incorporate an expected spread of the effect size of each variant by scaling \mathbf{R}_{var} as

$$\mathbf{S}_{\text{var}} = \Delta(\sigma_m) \mathbf{R}_{\text{var}} \Delta(\sigma_m),$$

where $\Delta(\sigma_m)$ is a diagonal matrix with entries σ_m determining the spread of the effect size distribution for each variant $m \leq M$. Thus, we can model settings where, e.g., protein-truncating variants have larger effect sizes ($\sigma = 1$) than missense variants ($\sigma = 0.1$). Note that when $\sigma_m = 1$ for all m then $\mathbf{S}_{\text{var}} = \mathbf{R}_{\text{var}}$.

All in all, our prior density for β under alternative model is

$$\beta|\text{ALT} \sim \mathcal{N}(\mathbf{0}, \mathbf{U}), \text{ where } \mathbf{U} = \mathbf{S}_{\text{var}} \otimes \mathbf{R}_{\text{phen}}.$$

When we have data from multiple studies we allow for possible differences in genetic effects across ethnicities or populations extending the Approximate Bayes Factors of Band et al. [1] and the summary statistics approach of RAREMETAL [2] from

univariate to multivariate phenotypes. Let $\hat{\beta} = (\hat{\beta}_{s,m,k}) =$

$(\hat{\beta}_{1,1,1}, \hat{\beta}_{1,1,2}, \dots, \hat{\beta}_{1,1,K}, \hat{\beta}_{1,2,1}, \dots, \hat{\beta}_{1,2,K}, \dots, \hat{\beta}_{1,M,1}, \hat{\beta}_{1,M,2}, \dots, \hat{\beta}_{1,M,K})$, where S is the number of studies, M is the number of variants, and K is the number of phenotypes. As with a single study, we incorporate the expected correlation of genetic effects between a pair of variants and a single phenotype using the matrix \mathbf{S}_{var} , between a variant and a pair of phenotypes using the matrix \mathbf{R}_{phen} , and we introduce the matrix $\mathbf{R}_{\text{study}}$ to specify prior on the similarity in effect sizes across the studies. Thus, the prior is

$$\beta \sim \mathcal{N}(\mathbf{0}, \mathbf{U}), \text{ where } \mathbf{U} = \mathbf{R}_{\text{study}} \otimes (\mathbf{S}_{\text{var}} \otimes \mathbf{R}_{\text{phen}}).$$

It is straightforward to include a non-zero vector μ as a prior mean of genetic effects, in which case the prior is

$$\beta \sim \mathcal{N}(\mu, \mathbf{U}).$$

We use this, for example, when screening for protective rare variants that have a pre-specified beneficial profile on a set of risk factors.

4 BF_{MRP}

The Bayes Factor is the ratio of the marginal likelihoods between the alternative and the null model. The marginal likelihood for the alternative model is

$$\int_{\beta} p(\text{Data}|\beta) p(\beta|\text{ALT}) d\beta = c \times \mathcal{N}(\hat{\beta}; \mu, \hat{V}_\beta + \mathbf{U})$$

and the marginal likelihood for the null model is

$$\int_{\boldsymbol{\beta}} p(\text{Data}|\boldsymbol{\beta}) p(\boldsymbol{\beta}|\text{NULL}) d\boldsymbol{\beta} = c \times \mathcal{N}(\hat{\boldsymbol{\beta}}; 0, \hat{\mathbf{V}}_{\boldsymbol{\beta}}).$$

The Bayes Factor (derived below in section 4.1) is given by

$$\text{BF}_{\text{MRP}} = \frac{\det(\hat{\mathbf{V}}_{\boldsymbol{\beta}} + \mathbf{U})^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})^{\top} (\hat{\mathbf{V}}_{\boldsymbol{\beta}} + \mathbf{U})^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})\right]}{\det(\hat{\mathbf{V}}_{\boldsymbol{\beta}})^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\hat{\boldsymbol{\beta}}^{\top} \hat{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \hat{\boldsymbol{\beta}}\right]}.$$

When $\boldsymbol{\mu} = 0$, BF_{MRP} is an increasing function of the following quadratic form

$$Q(\hat{\boldsymbol{\beta}}; \hat{\mathbf{V}}_{\boldsymbol{\beta}}, \mathbf{U}) = \hat{\boldsymbol{\beta}}^{\top} (\hat{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} - (\hat{\mathbf{V}}_{\boldsymbol{\beta}} + \mathbf{U})^{-1}) \hat{\boldsymbol{\beta}}. \quad (1)$$

Furthermore, this quadratic form is the only part of the BF_{MRP} that depends on $\hat{\boldsymbol{\beta}}$. Thus, by deriving a distribution of $Q(\hat{\boldsymbol{\beta}}; \hat{\mathbf{V}}_{\boldsymbol{\beta}}, \mathbf{U})$ under the null model we can compute a p-value when BF_{MRP} is used as a test statistic. According to basic properties of quadratic forms of Gaussian variables, $Q(\hat{\boldsymbol{\beta}}; \hat{\mathbf{V}}_{\boldsymbol{\beta}}, \mathbf{U}) \sim \sum_{i=1}^n d_i \chi_i^2$, where χ_i^2 are an independent sample from χ_1^2 distribution (chi-square with one degree of freedom), and d_i are the eigenvalues of matrix $\mathbf{I} - (\hat{\mathbf{V}}_{\boldsymbol{\beta}} + \mathbf{U})^{-1} \hat{\mathbf{V}}_{\boldsymbol{\beta}}$. The distribution function for a mixture of chi-squares can be numerically evaluated by the R-package ‘CompQuadForm’ [3].

4.1 MRP Bayes Factor derivation

To compute the Bayes Factor

$$\text{BF}_{\text{MRP}} = \frac{\det(\hat{\mathbf{V}}_{\boldsymbol{\beta}} + \mathbf{U})^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})^{\top} (\hat{\mathbf{V}}_{\boldsymbol{\beta}} + \mathbf{U})^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})\right]}{\det(\hat{\mathbf{V}}_{\boldsymbol{\beta}})^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\hat{\boldsymbol{\beta}}^{\top} \hat{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \hat{\boldsymbol{\beta}}\right]},$$

we first consider the term inside the exponential function:

$$\mathcal{E}(\hat{\boldsymbol{\beta}}, \boldsymbol{\mu}, \hat{\mathbf{V}}_{\boldsymbol{\beta}}, \mathbf{U}) = \frac{1}{2} \hat{\boldsymbol{\beta}}^{\top} \hat{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \hat{\boldsymbol{\beta}} - \frac{1}{2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})^{\top} (\hat{\mathbf{V}}_{\boldsymbol{\beta}} + \mathbf{U})^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}).$$

Since $\hat{\mathbf{V}}_{\boldsymbol{\beta}}$ and \mathbf{U} are typically defined through Kronecker products of smaller matrices, their inverses are easier to compute than the inverse of their sum. Hence we use Woodbury matrix identity to write

$$\mathcal{E}(\hat{\boldsymbol{\beta}}, \boldsymbol{\mu}, \hat{\mathbf{V}}_{\boldsymbol{\beta}}, \mathbf{U}) = \frac{1}{2} \hat{\boldsymbol{\beta}}^{\top} \hat{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \hat{\boldsymbol{\beta}} - \frac{1}{2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})^{\top} \left(\hat{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} - \hat{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} (\mathbf{U}^{-1} + \hat{\mathbf{V}}_{\boldsymbol{\beta}}^{-1})^{-1} \hat{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \right) (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}).$$

To simplify the determinant calculation we write

$$\det(\hat{\mathbf{V}}_{\boldsymbol{\beta}} + \mathbf{U}) = \det(\hat{\mathbf{V}}_{\boldsymbol{\beta}}) \det(\mathbf{I} + \hat{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \mathbf{U}).$$

The logarithm of the Bayes Factor is then

$$\log(\text{BF}_{\text{MRP}}) = -\frac{1}{2} \log(\det(\mathbf{I} + \hat{\mathbf{V}}_{\boldsymbol{\beta}}^{-1} \mathbf{U})) + \mathcal{E}(\hat{\boldsymbol{\beta}}, \boldsymbol{\mu}, \hat{\mathbf{V}}_{\boldsymbol{\beta}}, \mathbf{U}).$$

If studies do not share individuals, \hat{V}_β is a block-diagonal matrix

$$\hat{V}_\beta = \left[\begin{array}{c|c|c|c} \hat{V}_\beta^1 & 0 & \cdots & 0 \\ \hline 0 & \hat{V}_\beta^2 & \cdots & 0 \\ \hline \vdots & & \ddots & \vdots \\ \hline 0 & 0 & \cdots & \hat{V}_\beta^s \end{array} \right].$$

If studies share individuals, e.g., controls, we take the approach of Cichonska et al. [4] to use summary level data to estimate the correlation structure of the non-diagonal blocks caused by overlapping individuals.

References

- [1] Band G, Le QS, Jostins L, Pirinen M, Kivinen K, Jallow M, et al. Imputation-based meta-analysis of severe malaria in three African populations. *PLoS genetics*. 2013;9(5):e1003509.
- [2] Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, et al. Meta-analysis of gene-level tests for rare variant association. *Nature Genetics*. 2014;46(2):200–204.
- [3] Duchesne P, de Micheaux PL. Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. *Computational Statistics and Data Analysis*. 2010;54:858–862.
- [4] Cichonska A, Rousu J, Marttinen P, Kangas AJ, Soininen P, Lehtimäki T, et al. metaCCA: summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. *Bioinformatics*. 2016;32(13):1981–1989.