Genome-wide scan for linear mixed models

Danilo Horta and Francesco P. Casale

April 5, 2017

1 Least squares

We apply the method of least squares to estimate the fixed-effect sizes.

1.1 Single block

Let

$$\mathbf{y} \sim \mathcal{N}(\mathrm{F}\boldsymbol{\alpha};\, \mathrm{K}) \tag{1}$$

be the marginal likelihood of a LMM with a single covariate block F. The maximum likelihood estimation of the fixed effects is

$$\hat{\boldsymbol{\alpha}} = (\mathrm{F}^\intercal \mathrm{K}^\dagger \mathrm{F})^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \mathbf{y}.$$

In practice, the method of least squares can be used to solve the above equation without explicitly finding pseudoinverses.

1.2 Double block

Let

$$\mathbf{y} \sim \mathcal{N}(\mathrm{F}\boldsymbol{\alpha} + \mathrm{G}\boldsymbol{\beta};\, \mathrm{K})$$

be the marginal likelihood of a LMM with two covariate blocks F and G. We want to solve

$$\begin{bmatrix}\hat{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\beta}}\end{bmatrix} = (\mathrm{X}^\intercal \mathrm{X})^\dagger \mathrm{X}^\intercal {\mathrm{K}^{\frac{1}{2}}}^\dagger \mathbf{y}$$

for

$$\mathrm{L} = {\mathrm{K}^{\frac{1}{2}}}^\dagger \mathrm{F},\ \ \mathrm{R} = {\mathrm{K}^{\frac{1}{2}}}^\dagger \mathrm{G}\ \text{ and }\ \mathrm{X} = \begin{bmatrix}\mathrm{L} & \mathrm{R}\end{bmatrix}.$$

We have

$$(\mathrm{X}^\intercal \mathrm{X})^\dagger = \begin{bmatrix} (\mathrm{L}^\intercal \mathrm{L})^\dagger + (\mathrm{L}^\intercal \mathrm{L})^\dagger (\mathrm{L}^\intercal \mathrm{R}) \mathrm{W}^\dagger (\mathrm{L}^\intercal \mathrm{R})^\intercal (\mathrm{L}^\intercal \mathrm{L})^\dagger & -(\mathrm{L}^\intercal \mathrm{L})^\dagger (\mathrm{L}^\intercal \mathrm{R}) \mathrm{W}^\dagger \\ -\mathrm{W}^\dagger (\mathrm{L}^\intercal \mathrm{R})^\intercal (\mathrm{L}^\intercal \mathrm{L})^\dagger & \mathrm{W}^\dagger \end{bmatrix},$$

from Eq. 3 of [1], where $\mathrm{W} = \mathrm{R}^\intercal \mathrm{R} - \mathrm{R}^\intercal \mathrm{L} (\mathrm{L}^\intercal \mathrm{L})^\dagger \mathrm{L}^\intercal \mathrm{R}$.

Defining $\mathrm{A} = \mathrm{L}^\intercal \mathrm{L}$ and $\mathrm{B} = \mathrm{L}^\intercal \mathrm{R}$ leads us to

$$(\mathrm{X}^\intercal \mathrm{X})^\dagger \mathrm{X}^\intercal = \begin{bmatrix} \mathrm{A}^\dagger \mathrm{L}^\intercal + \mathrm{A}^\dagger \mathrm{B} \mathrm{W}^\dagger \mathrm{B}^\intercal \mathrm{A}^\dagger \mathrm{L}^\intercal - \mathrm{A}^\dagger \mathrm{B} \mathrm{W}^\dagger \mathrm{R}^\intercal \\ -\mathrm{W}^\dagger \mathrm{B}^\intercal \mathrm{A}^\dagger \mathrm{L}^\intercal + \mathrm{W}^\dagger \mathrm{R}^\intercal \end{bmatrix}.$$

Finally,

$$(\mathrm{X}^\intercal \mathrm{X})^\dagger \mathrm{X}^\intercal {\mathrm{K}^{\frac{1}{2}}}^\dagger \mathbf{y} = \begin{bmatrix} \mathrm{A}^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \mathbf{y} + \mathrm{A}^\dagger \mathrm{B} \mathrm{W}^\dagger \mathrm{B}^\intercal \mathrm{A}^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \mathbf{y} - \mathrm{A}^\dagger \mathrm{B} \mathrm{W}^\dagger \mathrm{G}^\intercal \mathrm{K}^\dagger \mathbf{y} \\ \mathrm{W}^\dagger \mathrm{G}^\intercal \mathrm{K}^\dagger \mathbf{y} - \mathrm{W}^\dagger \mathrm{B}^\intercal \mathrm{A}^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \mathbf{y} \end{bmatrix}.$$

A robust implementation of the above equation has to: (i) associate matrix multiplications in such a way that a sequence of $\mathrm{K}^\dagger \ldots \mathrm{K}^\dagger$ is avoided; and (ii) handle low-rank matrices W. A better association of matrix multiplications is given by

$$(\mathrm{X}^\intercal \mathrm{X})^\dagger \mathrm{X}^\intercal {\mathrm{K}^{\frac{1}{2}}}^\dagger \mathbf{y} = \begin{bmatrix} \mathrm{A}^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \mathbf{y} + \mathrm{A}^\dagger \mathrm{B} \mathrm{W}^\dagger (\mathrm{B}^\intercal \mathrm{A}^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \mathbf{y} - \mathrm{G}^\intercal \mathrm{K}^\dagger \mathbf{y}) \\ \mathrm{W}^\dagger (\mathrm{G}^\intercal \mathrm{K}^\dagger \mathbf{y} - \mathrm{B}^\intercal \mathrm{A}^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \mathbf{y}) \end{bmatrix}.$$

$\mathrm{W}^\dagger$ can be found via economic SVD decomposition.

1.3 Batch scan

Given a $n \times p$ covariate block G, we want to quickly infer the maximum likelihood estimations of $\boldsymbol{\alpha}_i$ and β_i, for $i \in \{1,\ldots,p\}$ denoting the G columns. We define a diagonal matrix

$$\mathrm{W} = \mathrm{dotd}(\mathrm{G}^\intercal, \mathrm{K}^\dagger \mathrm{G}) - \mathrm{dotd}(\mathrm{G}^\intercal, \mathrm{K}^\dagger \mathrm{F} (\mathrm{L}^\intercal \mathrm{L})^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \mathrm{G}),$$

where $\mathrm{dotd}(\cdot,\cdot)$ is a function that returns the diagonal elements of a matrix multiplication with asymptotically lower computational cost and memory use.

Clearly,

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_p\end{bmatrix} = \mathrm{W}^\dagger (\mathrm{G}^\intercal \mathrm{K}^\dagger \mathbf{y} - \mathrm{B}^\intercal \mathrm{A}^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \mathbf{y}).$$

We know that

$$\hat{\boldsymbol{\alpha}}_i = \mathrm{A}^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger (\mathbf{y} + (\mathrm{G}_i \mathrm{W}_i {\mathrm{G}_i}^\intercal)(\mathrm{K}^\dagger \mathrm{F} \mathrm{A}^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \mathbf{y} - \mathrm{K}^\dagger \mathbf{y})),$$

where G_i is the i-th column of G. For one-shot computation we do

$$\hat{\boldsymbol{\alpha}} = \mathrm{A}^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \Big(\mathbf{y} \oplus \mathrm{G} \otimes \big((\mathrm{W}\mathrm{G}^\intercal)(\mathrm{K}^\dagger \mathrm{F} \mathrm{A}^\dagger \mathrm{F}^\intercal \mathrm{K}^\dagger \mathbf{y} - \mathrm{K}^\dagger \mathbf{y})\big)^\intercal\Big),$$

where $\oplus$ and $\otimes$ are element-wise summation and multiplication with broadcasting.

2 Marginal likelihood

Replace K by $\sigma^2\mathrm{K}$ in Eq. (1). Log of the marginal likelihood is given by

$$\mathcal{L}(\boldsymbol{\beta},\sigma^2) = -\frac{1}{2}\big(\log(2\pi) - \mathrm{logdet}|\sigma^2 \mathrm{K}| - (\mathbf{y} - \mathrm{F}\boldsymbol{\beta})^\intercal (\sigma^2 \mathrm{K})^\dagger (\mathbf{y} - \mathrm{F}\boldsymbol{\beta})\big).$$

The maximum likelihood estimation and the restricted likelihood estimation of σ^2 are give by

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathrm{F}\hat{\boldsymbol{\beta}})^\intercal \mathrm{K}^\dagger (\mathbf{y} - \mathrm{F}\hat{\boldsymbol{\beta}})}{k}$$

for $k = n$ and $k = n - p$, respectively.

3 Likelihood ratio test

Suppose that the null hyphotesis is given by $\mathcal{H}_0 : \boldsymbol{\beta} \in \Theta_0$ and the alternative one is given by $\mathcal{H}_1 : \boldsymbol{\beta} \in \Theta_1$, where Θ_0 is a subspace of Θ_1, and let d be the difference between the Θ_1 and Θ_0 dimensions. Give that the null hyphotesis is the true one, it follows from classical likelihood theory that

$$\tilde{\chi}^2_d \sim -2\ln\left(\frac{\mathcal{L}(\hat{\boldsymbol{\beta}}_0, \hat{\sigma}_0^2)}{\mathcal{L}(\hat{\boldsymbol{\beta}}_1, \hat{\sigma}_1^2)}\right),$$

asymptotically, for maximum likelihood parameter estimation[1] subject to $\hat{\boldsymbol{\beta}}_0 \in \Theta_0$ and $\hat{\boldsymbol{\beta}}_1 \in \Theta_1$ [2].

References

[1] Charles A Rohde. "Generalized inverses of partitioned matrices". In: *Journal of the Society for Industrial and Applied Mathematics* 13.4 (1965), pp. 1033–1035.

[2] Geert Verbeke and Geert Molenberghs. *Linear mixed models for longitudinal data*. Springer Science & Business Media, 2009.

[1] This is not true for rescricted maximum likelihood estimation.