Method of Moments and MM Algorithm

Yida Wu

2023-08-25

Contents

Linear mixed models (LMMs) have emerged as a key tool for heritability estimation where the parameters of the LMMs, i.e. the variance components, are related to the heritability attributable to the SNPs analyzed.

1 Linear Mixed Model

The linear mixed model builds upon a linear relationship from $\mathbf{y}$ to $\mathbf{X}$ and $\mathbf{Z}$ by

$$\mathbf{y} = \mathbf{Z}\omega + \mathbf{X}\beta + \mathbf{e}. \tag{1}$$

- $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{y}$ is centered so that $\sum_n y_n = 0$;
- $\mathbf{X} \in \mathbb{R}^{n\times p}$, each column of $\mathbf{X}$ is centered and scaled so that $\sum_n x_{n,p} = 0$ and $\sum_n x_{n,p}^2 = \frac{1}{p}$;
- $\mathbf{Z}$ is a $n \times c$ matrix of covariates;
- $\omega \in \mathbb{R}^p$ is the vector of fixed effects;
- β is the vector of random effects with $\beta \sim \mathcal{N}\left(0, \sigma_\beta^2 \mathbf{I}_p\right)$;
- $\mathbf{e} \sim \mathcal{N}\left(0, \sigma_e^2 \mathbf{I}_n\right)$ is the independent noise term.

Note that the linear mixed model (1) can be re-written as:

$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{Z}\omega, \Sigma\right), \tag{2}$$

where $\Sigma = \sigma_\beta^2 \mathbf{K} + \sigma_e^2 \mathbf{I}_n$ and $\mathbf{K} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$. The main target is to estimate the set of unknown parameters $= \left\{\omega, \sigma_\beta^2, \sigma_e^2\right\}$. We will derive and implement two methods (MoM and MM) in this project.

2 Method-of-Moments

2.1 Derivation

The **principle** of the Method-of-Moments (MoM) is to obtain estimates of the model parameters such that the theoretical moments match the sample moments.

First, Equation (1) is transformed by multiplying by the projection matrix $\mathbf{V} = \mathbf{I}_n - \mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T$ (Note that $\mathbf{V}^{\mathrm{T}} = \mathbf{V}$ and $\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{V}$) :

$$\mathbf{V}\mathbf{y} = \mathbf{V}\mathbf{X}\beta + \mathbf{V}\mathbf{e}. \tag{3}$$

From Equation (2), the first theoretical moment and the second theoretical moment can be derived. $\mathbb{E}\left[\mathbf{V}\mathbf{y}\right] = \mathbf{0}$ while the population covariance of the vector $\mathbf{V}\mathbf{y}$ is:

$$\begin{aligned}
\mathrm{Cov}\left(\mathbf{V}\mathbf{y}\right) &= \mathbb{E}\left[\mathbf{V}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{V}\right] - \mathbb{E}\left[\mathbf{V}\mathbf{y}\right]\mathbb{E}\left[\mathbf{V}\mathbf{y}\right]^{\mathrm{T}} && (4)\\
&= \sigma_\beta^2 \mathbf{V}\mathbf{K}\mathbf{V} + \sigma_e^2 \mathbf{V}. && (1)
\end{aligned}$$

Next, the MoM estimator is obtained by solving the following ordinary least squares (OLS) problem:

$$\left(\widehat{\sigma_\beta^2}, \widehat{\sigma_e^2}\right) = \mathrm{argmin}_{\sigma_\beta^2, \sigma_e^2} \left\|(\mathbf{V}\mathbf{y})(\mathbf{V}\mathbf{y})^{\mathrm{T}} - \left(\sigma_\beta^2 \mathbf{V}\mathbf{K}\mathbf{V} + \sigma_e^2 \mathbf{V}\right)\right\|_F^2, \tag{5}$$

Due to the fact that $\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}\left(\mathbf{A}\mathbf{A}^T\right)}$, the OLS problem can be re-written as:

$$\left(\widehat{\sigma_\beta^2}, \widehat{\sigma_e^2}\right) = \mathrm{argmin}_{\sigma_\beta^2, \sigma_e^2} \, \mathrm{tr}\left[\left((\mathbf{V}\mathbf{y})(\mathbf{V}\mathbf{y})^{\mathrm{T}} - \left(\sigma_\beta^2 \mathbf{V}\mathbf{K}\mathbf{V} + \sigma_e^2 \mathbf{V}\right)\right)\left((\mathbf{V}\mathbf{y})(\mathbf{V}\mathbf{y})^{\mathrm{T}} - \left(\sigma_\beta^2 \mathbf{V}\mathbf{K}\mathbf{V} + \sigma_e^2 \mathbf{V}\right)\right)^{\mathrm{T}}\right]. \tag{6}$$

Then, the MoM estimator satisfies the normal equations:

$$\mathbf{A}\hat{\theta} = \mathbf{b}, \tag{7}$$

where

$$\mathbf{A} = \left[\begin{array}{ll} \mathrm{tr}\left(\mathbf{V}\mathbf{K}\mathbf{V}\mathbf{K}\right) & \mathrm{tr}\left(\mathbf{V}\mathbf{K}\right) \\ \mathrm{tr}\left(\mathbf{V}\mathbf{K}\right) & n-c \end{array}\right],$$

$$\hat{\theta} = \left[\begin{array}{c} \widehat{\sigma_\beta^2} \\ \widehat{\sigma_e^2} \end{array}\right],$$

$$\mathbf{b} = \left[\begin{array}{c} \mathbf{y}^{\mathrm{T}}\mathbf{VKVy} \\ \mathbf{y}^{\mathrm{T}}\mathbf{Vy} \end{array} \right].$$

Hence, the MoM estimates of θ is $\hat{\theta} = \mathbf{A}^{-1}\mathbf{b}$. Once the $\hat{\theta}$ is obtained, estimating the vector of fixed effects ω is a standard general least-squares problem, that is:

$$\hat{\omega} = \left(\mathbf{Z}^{\mathrm{T}}\hat{\Sigma}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}(\hat{\Sigma})^{-1}\mathbf{y}, \tag{2}$$

where $\hat{\Sigma} = \hat{\sigma_\beta^2}\mathbf{K} + \hat{\sigma_e^2}\mathbf{I}_n$.

2.2 Modification 1: Sandwich Estimator

From Equation (7), the covariance matrix of $\hat{\theta}$ can be given by the sandwich estimator: $\operatorname{Cov}\left(\hat{\theta}\right) = \mathbf{A}^{-1}\operatorname{Cov}\left(\mathbf{b}\right)\mathbf{A}^{-1}$, where

$$\operatorname{Cov}(\mathbf{b}) = \operatorname{Cov}\left(\left[\begin{array}{c} \mathbf{y}^{\mathrm{T}}\mathbf{VKVy} \\ \mathbf{y}^{\mathrm{T}}\mathbf{Vy} \end{array} \right]\right) = \left[\begin{array}{cc} \operatorname{Var}\left(\mathbf{y}^{\mathrm{T}}\mathbf{VKVy}\right) & \operatorname{Cov}\left(\mathbf{y}^{\mathrm{T}}\mathbf{VKVy}, \mathbf{y}^{\mathrm{T}}\mathbf{Vy}\right) \\ \operatorname{Cov}\left(\mathbf{y}^{\mathrm{T}}\mathbf{Vy}, \mathbf{y}^{\mathrm{T}}\mathbf{VKVy}\right) & \operatorname{Var}\left(\mathbf{y}^{\mathrm{T}}\mathbf{Vy}\right) \end{array} \right], \tag{8}$$

Using the Lemma 1, the elements of $\operatorname{Cov}(\mathbf{b})$ are calculated by $\operatorname{Var}\left(\mathbf{y}^{\mathrm{T}}\mathbf{VKVy}\right) = 2\operatorname{tr}\left[(\mathbf{VKV}\Sigma)^2\right]$, $\operatorname{Var}\left(\mathbf{y}^{\mathrm{T}}\mathbf{Vy}\right) = 2\operatorname{tr}\left[(\mathbf{V}\Sigma)^2\right]$, $\operatorname{Cov}\left(\mathbf{y}^{\mathrm{T}}\mathbf{VKVy}, \mathbf{y}^{\mathrm{T}}\mathbf{Vy}\right) = 2\operatorname{tr}(\mathbf{VKV}\Sigma\mathbf{V}\Sigma)$.

Since $\hat{\theta} - \theta_0$ is asymptotically normal and θ_0 is the true value of θ, that is:

$$\operatorname{Cov}\left(\hat{\theta}\right)^{-1/2}\left(\hat{\theta} - \theta_0\right) \to_d \mathcal{N}\left(\mathbf{0}, \mathbf{I}_2\right), \tag{3}$$

Then when $\hat{\theta} = \theta_0$, the rejection region is:

$$\left(\hat{\theta} - \theta_0\right)^{\mathrm{T}}\operatorname{Cov}\left(\hat{\theta}\right)^{-1}\left(\hat{\theta} - \theta_0\right) > \chi^2_{2,\alpha}. \tag{4}$$

2.3 Modification 2: Delta Method

Denote $\hat{h}^2 = g\left(\hat{\theta}\right) = \frac{\hat{\sigma_\beta^2}}{\hat{\sigma_\beta^2} + \hat{\sigma_e^2}}$, and the gradient matrix can be computed:

$$\nabla g\left(\theta\right) = \left(\frac{\hat{\sigma_e^2}}{\left(\hat{\sigma_\beta^2} + \hat{\sigma_e^2}\right)^2}, \frac{-\hat{\sigma_\beta^2}}{\left(\hat{\sigma_\beta^2} + \hat{\sigma_e^2}\right)^2}\right)^{\mathrm{T}}. \tag{9}$$

Then, using the delta method, the variance of $\hat{h}^2$ is:

$$\operatorname{Cov}\left(\hat{h}^2\right) = \nabla^{\mathrm{T}}g\left(\theta\right)\operatorname{Cov}\left(\hat{\theta}\right)\nabla g\left(\theta\right). \tag{10}$$

After the variance of $\hat{h}^2$ is obtained, using the delta theorem, we know that $g(\theta)-g(\theta_0)$ is also asymptotically normal, that is:

$$\operatorname{Cov}\left(\hat{h}^2\right)^{-1/2}\left(g\left(\theta\right)-g\left(\theta_0\right)\right)\rightarrow_d \mathcal{N}\left(0,1\right), \tag{5}$$

and when $g\left(\theta\right)=g\left(\theta_0\right)$, the rejection region is:

$$\left(g\left(\theta\right)-g\left(\theta_0\right)\right)^{\mathrm{T}}\operatorname{Cov}\left(\hat{h}^2\right)^{-1}\left(g\left(\theta\right)-g\left(\theta_0\right)\right)>\chi^2_{1,\alpha}. \tag{6}$$

2.4 Application

2.5 Input the data

2.6 MoM

```
## [1] 1
```

```
## [1] 0.9896469
```

3 MM Algorithm

Unlike the MoM, the minorization-maximization (MM) algorithm consider the likelihood of the variance components model directly. The log-likelihood function $\mathcal{L}(\mathbf{y}\mid\omega,\sigma^2_\beta,\sigma^2_e;\mathbf{Z},\mathbf{K})$ is given as:

$$\mathcal{L}(\mathbf{y}\mid\omega,\sigma^2_\beta,\sigma^2_e;\mathbf{Z},\mathbf{K})=-\frac{1}{2}\log\det\Sigma-\frac{1}{2}(\mathbf{y}-\mathbf{Z}\omega)^{\mathrm{T}}\Sigma(\mathbf{y}-\mathbf{Z}\omega), \tag{7}$$

where $\Sigma=\sigma^2_\beta\mathbf{K}+\sigma^2_e\mathbf{I}_n$. The MM algorithm is utilized to maximizing the log-likelihood function and such an algorithm follow from the inequalities:

$$f\left(\theta^{(t+1)}\right)\geq g\left(\theta^{(t+1)}\mid\theta^{(t)}\right)\geq g\left(\theta^{(t)}\mid\theta^{(t)}\right)=f\left(\theta^{(t)}\right). \tag{8}$$

Therefore, the key step of MM algorithm is to identify the surrogate function $g\left(\theta|\theta^{(t)}\right)$ by using proper inequalities.

The strategy for maximizing the log-likelihood is to alternate updating the fixed effects ω and the variance components $\theta=\left(\sigma^2_\beta,\sigma^2_e\right)$. Updating ω is a standard general least-squares problem with solution

$$\omega^{(t+1)}=\left(\mathbf{Z}^{\mathrm{T}}\Sigma^{-(t)}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\Sigma^{-(t)}\mathbf{y}, \tag{9}$$

where $\Sigma^{-(t)}=\sigma^{2(t)}_\beta\mathbf{K}+\sigma^{2(t)}_e\mathbf{I}_n$.

Then, Updating θ given $\omega^{(t)}$ depends on two minorizations.

First, using the **joint convexity** of $\Sigma^{(t)}\Sigma^{-1}\Sigma^{(t)}$,

$$-(\mathbf{y}-\mathbf{Z}\omega)^{\mathrm{T}}\Sigma(\mathbf{y}-\mathbf{Z}\omega) \geq -(\mathbf{y}-\mathbf{Z}\omega)^{\mathrm{T}}\Sigma^{-(t)}\left(\frac{\sigma_\beta^{4(t)}}{\sigma_\beta^2}\mathbf{K}+\frac{\sigma_e^{4(t)}}{\sigma_e^2}\right)(\mathbf{y}-\mathbf{Z}\omega). \tag{10}$$

Second, using the **supporting hyperplane**,

$$-\log\det\Sigma \geq -\log\det\Sigma^{(t)} - \operatorname{tr}\left[\Sigma^{-(t)}\left(\Sigma-\Sigma^{(t)}\right)\right]. \tag{11}$$

Combining of the minorizations gives the overall minorization:

$$
\begin{aligned}
& g\left(\theta|\theta^{(t)}\right) && (12)\\
&= -\frac{1}{2}\operatorname{tr}\left(\Sigma^{-(t)}\Sigma\right) - \frac{1}{2}(\mathbf{y}-\mathbf{Z}\omega^{(t)})^{\mathrm{T}}\Sigma^{-(t)}\left(\frac{\sigma_\beta^{4(t)}}{\sigma_\beta^2}\mathbf{K}+\frac{\sigma_e^{4(t)}}{\sigma_e^2}\right)(\mathbf{y}-\mathbf{Z}\omega^{(t)}) + c^{(t)} && (13)\\
&= -\frac{\sigma_\beta^2}{2}\operatorname{tr}\left(\Sigma^{-(t)}\mathbf{K}\right) - \frac{1}{2}\frac{\sigma_\beta^{4(t)}}{\sigma_\beta^2}(\mathbf{y}-\mathbf{Z}\omega^{(t)})^{\mathrm{T}}\Sigma^{-(t)}\mathbf{K}\Sigma^{-(t)}(\mathbf{y}-\mathbf{Z}\omega^{(t)}) && (14)\\
&\quad -\frac{\sigma_e^2}{2}\operatorname{tr}\left(\Sigma^{-(t)}\right) - \frac{1}{2}\frac{\sigma_e^{4(t)}}{\sigma_e^2}(\mathbf{y}-\mathbf{Z}\omega^{(t)})^{\mathrm{T}}\Sigma^{-2(t)}(\mathbf{y}-\mathbf{Z}\omega^{(t)}) + c^{(t)}, && (15)
\end{aligned}
$$

where $c^{(t)}$ is an irrelevant constant. By setting that $\frac{\partial g\left(\theta|\theta^{(t)}\right)}{\partial\sigma_\beta^2}=0$ and $\frac{\partial g\left(\theta|\theta^{(t)}\right)}{\partial\sigma_e^2}=0$, the updates of θ are given as follows:

$$\sigma_\beta^{2(t+1)} = \sigma_\beta^{2(t)}\sqrt{\frac{(\mathbf{y}-\mathbf{Z}\omega^{(t)})^{\mathrm{T}}\Sigma^{-(t)}\mathbf{K}\Sigma^{(t)}(\mathbf{y}-\mathbf{Z}\omega^{(t)})}{\operatorname{tr}\left(\Sigma^{-(t)}\mathbf{K}\right)}}, \tag{16}$$

$$\sigma_e^{2(t+1)} = \sigma_e^{2(t)}\sqrt{\frac{(\mathbf{y}-\mathbf{Z}\omega^{(t)})^{\mathrm{T}}\Sigma^{-2(t)}(\mathbf{y}-\mathbf{Z}\omega^{(t)})}{\operatorname{tr}\left(\Sigma^{-(t)}\right)}}. \tag{17}$$

3.1 The covariance matrix of θ

The covariance matrix of $\hat{\theta}$ can be calculated from the inverse of Fisher Information Matrix (FIM). Hence, the first step is to obtain FIM, that is

$$\mathrm{FIM} = -\mathbb{E}\left[\frac{\partial^2\mathcal{L}}{\partial\theta^2}\right]. \tag{18}$$

The first derivatives are:

$$\frac{\partial\mathcal{L}}{\partial\sigma_\beta^2} = \frac{1}{2}\operatorname{tr}\left[-\Sigma^{-1}\mathbf{K} + (\mathbf{y}-\mathbf{Z}\omega)^{\mathrm{T}}\Sigma^{-1}\mathbf{K}\Sigma^{-1}(\mathbf{y}-\mathbf{Z}\omega)\right], \tag{19}$$

$$\frac{\partial\mathcal{L}}{\partial\sigma_e^2} = \frac{1}{2}\operatorname{tr}\left[-\Sigma^{-1} + (\mathbf{y}-\mathbf{Z}\omega)^{\mathrm{T}}\Sigma^{-2}(\mathbf{y}-\mathbf{Z}\omega)\right]. \tag{20}$$

And the second derivatives are:

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial\left(\sigma_\beta^2\right)^2} &= \frac{1}{2}\operatorname{tr}\left[\left(\Sigma^{-1}\mathbf{K}\right)^2 - 2\left(\Sigma^{-1}\mathbf{K}\right)^2\Sigma^{-1}(\mathbf{y}-\mathbf{Z}\omega)(\mathbf{y}-\mathbf{Z}\omega)^{\mathrm{T}}\right],\\
\frac{\partial^2 \mathcal{L}}{\partial(\sigma_e^2)^2} &= \frac{1}{2}\operatorname{tr}\left[\Sigma^{-2} - 2\Sigma^{-3}(\mathbf{y}-\mathbf{Z}\omega)(\mathbf{y}-\mathbf{Z}\omega)^{\mathrm{T}}\right],\\
\frac{\partial^2 \mathcal{L}}{\partial\sigma_\beta^2\partial\sigma_e^2} &= \frac{1}{2}\operatorname{tr}\left[\Sigma^{-1}\mathbf{K}\Sigma^{-1} - \left(\Sigma^{-1}\mathbf{K}\Sigma^{-2} + \Sigma^{-2}\mathbf{K}\Sigma^{-1}\right)(\mathbf{y}-\mathbf{Z}\omega)(\mathbf{y}-\mathbf{Z}\omega)^{\mathrm{T}}\right].
\end{aligned}
$$

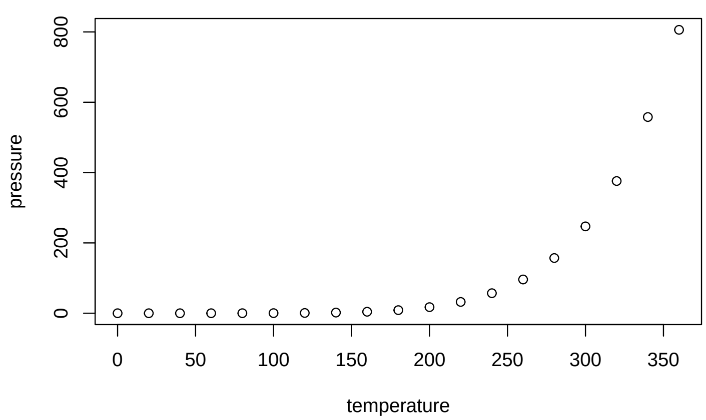
Since $\mathbb{E}\left[(\mathbf{y}-\mathbf{Z}\omega)(\mathbf{y}-\mathbf{Z}\omega)^{\mathrm{T}}\right] = \Sigma$, the FIM is:

$$
\begin{aligned}
\text{FIM} &= -\,\mathbb{E}\left[\frac{\partial^2\mathcal{L}}{\partial\theta^2}\right] && (21)\\
&= \frac{1}{2}\begin{bmatrix} \operatorname{tr}\left[\left(\Sigma^{-1}\mathbf{K}\right)^2\right] & \operatorname{tr}\left(\Sigma^{-2}\mathbf{K}\right) \\ \operatorname{tr}\left(\Sigma^{-2}\mathbf{K}\right) & \operatorname{tr}\left[\Sigma^{-2}\right] \end{bmatrix}. && (22)
\end{aligned}
$$

Therefore, the covariance matrix of $\hat{\theta}$ is the inverse of FIM.

3.2 Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.