Details of Derivation: MoM and MM

Yida Wu

2023-08-29

Contents

1 Linear Mixed Model

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

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

- $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{y}$ 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;
- $\boldsymbol{\omega} \in \mathbb{R}^c$ is the vector of fixed effects;
- $\boldsymbol{\beta}$ is the vector of random effects with $\boldsymbol{\beta} \sim \mathcal{N}\left(0, \sigma_\beta^2 \mathbf{I}_p\right)$;
- $\mathbf{e} \sim \mathcal{N}\left(0, \sigma_e^2 \mathbf{I}_n\right)$ is the independent noise term.

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

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

where $\boldsymbol{\Sigma} = \sigma_\beta^2 \mathbf{K} + \sigma_e^2 \mathbf{I}_n$ and $\mathbf{K} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$. The first target is to estimate the set of unknown parameters $\boldsymbol{\Theta} = \left\{\boldsymbol{\omega}, \sigma_\beta^2, \sigma_e^2\right\}$. 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 $\mathbf{V} = \mathbf{I}_n - \mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}$ (Note that $\mathbf{V}^{\mathrm{T}} = \mathbf{V}$ and $\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{V}$) :

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

From Equation (3), 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:

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

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

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

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

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

Then, the MoM estimator satisfies the normal equations:

$$\mathbf{A}\hat{\boldsymbol{\sigma}}^2 = \mathbf{b}, \tag{7}$$

where

$$\mathbf{A} = \left[\begin{array}{ll} \operatorname{tr}\left(\mathbf{V}\mathbf{K}\mathbf{V}\mathbf{K}\right) & \operatorname{tr}\left(\mathbf{V}\mathbf{K}\right) \\ \operatorname{tr}\left(\mathbf{V}\mathbf{K}\right) & n-c \end{array}\right], \hat{\boldsymbol{\sigma}}^2 = \left[\begin{array}{l} \hat{\sigma}_\beta^2 \\ \hat{\sigma}_e^2 \end{array}\right], \mathbf{b} = \left[\begin{array}{l} \mathbf{y}^{\mathrm{T}}\mathbf{V}\mathbf{K}\mathbf{V}\mathbf{y} \\ \mathbf{y}^{\mathrm{T}}\mathbf{V}\mathbf{y} \end{array}\right].$$

Hence, the MoM estimates of $\boldsymbol{\sigma}^2$ is $\hat{\boldsymbol{\sigma}}^2 = \mathbf{A}^{-1}\mathbf{b}$. Once the $\boldsymbol{\sigma}^2$ is obtained, estimating the vector of fixed effects $\boldsymbol{\omega}$ is a **standard general least-squares problem**, that is:

$$\hat{\boldsymbol{\omega}} = \left(\mathbf{Z}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}(\hat{\boldsymbol{\Sigma}})^{-1}\mathbf{y}, \tag{8}$$

where $\hat{\boldsymbol{\Sigma}} = \hat{\sigma}_\beta^2\mathbf{K} + \hat{\sigma}_e^2\mathbf{I}_n$.

The one remaining quantity that we need to compute efficiently is $\operatorname{tr}(\mathbf{V}\mathbf{K}\mathbf{V}\mathbf{K})$. Since $\mathbf{V}\mathbf{K}$ is a symmetric matrix, we can use $\|\mathbf{V}\mathbf{K}\|_F^2$ to replace $\operatorname{tr}(\mathbf{V}\mathbf{K}\mathbf{V}\mathbf{K})$, which can compute efficiently. Another method is using a randomized estimator $L_B = \frac{1}{B}\sum_b z_b^{\mathrm{T}}\mathbf{V}\mathbf{K}\mathbf{V}\mathbf{K}z_b$, which is the unbiased estimator of the trace of $\mathbf{V}\mathbf{K}\mathbf{V}\mathbf{K}$ given B random vectors, $z_1, \ldots, z_B$, drawn independently from a standard normal distribution.

2.2 Modification 1: Sandwich Estimator

From Equation (7), the covariance matrix of $\hat{\boldsymbol{\sigma}}^2$ can be given by **the sandwich estimator**: $\operatorname{Cov}\left(\hat{\boldsymbol{\sigma}}^2\right) = \mathbf{A}^{-1}\operatorname{Cov}(\mathbf{b})\,\mathbf{A}^{-1}$, where

$$\operatorname{Cov}(\mathbf{b}) = \operatorname{Cov}\left(\begin{bmatrix} \mathbf{y}^{\mathrm{T}}\mathbf{VKVy} \\ \mathbf{y}^{\mathrm{T}}\mathbf{Vy} \end{bmatrix}\right) = \begin{bmatrix} \operatorname{Var}\left(\mathbf{y}^{\mathrm{T}}\mathbf{VKVy}\right) & \operatorname{Cov}\left(\mathbf{y}^{\mathrm{T}}\mathbf{VKVy}, \mathbf{y}^{\mathrm{T}}\mathbf{Vy}\right) \\ \operatorname{Cov}\left(\mathbf{y}^{\mathrm{T}}\mathbf{Vy}, \mathbf{y}^{\mathrm{T}}\mathbf{VKVy}\right) & \operatorname{Var}\left(\mathbf{y}^{\mathrm{T}}\mathbf{Vy}\right) \end{bmatrix}, \tag{9}$$

Using the Lemma 1 in [1], the elements of $\operatorname{Cov}(\mathbf{b})$ are calculated by

$$\begin{aligned} \operatorname{Var}\left(\mathbf{y}^{\mathrm{T}}\mathbf{VKVy}\right) &= 2\operatorname{tr}\left[(\mathbf{VKV\Sigma})^2\right], \\ \operatorname{Var}\left(\mathbf{y}^{\mathrm{T}}\mathbf{Vy}\right) &= 2\operatorname{tr}\left[(\mathbf{V\Sigma})^2\right], \\ \operatorname{Cov}\left(\mathbf{y}^{\mathrm{T}}\mathbf{VKVy}, \mathbf{y}^{\mathrm{T}}\mathbf{Vy}\right) &= 2\operatorname{tr}(\mathbf{VKV\Sigma V\Sigma}). \end{aligned}$$

Since $\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}_0^2$ is asymptotically normal and $\boldsymbol{\sigma}_0^2$ is the true value of $\boldsymbol{\sigma}^2$, that is:

$$\operatorname{Cov}\left(\hat{\boldsymbol{\sigma}}^2\right)^{-1/2}\left(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}_0^2\right) \rightarrow_d \mathcal{N}\left(\mathbf{0}, \mathbf{I}_2\right), \tag{10}$$

Then when $\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}_0^2$, the rejection region is:

$$\left(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}_0^2\right)^{\mathrm{T}}\operatorname{Cov}\left(\hat{\boldsymbol{\sigma}}^2\right)^{-1}\left(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}_0^2\right) > \chi^2_{2,\alpha}. \tag{11}$$

2.3 Modification 2: Delta Method

Denote $\hat{h}^2 = g\left(\hat{\boldsymbol{\sigma}}^2\right) = \frac{\hat{\sigma}_\beta^2}{\hat{\sigma}_\beta^2 + \hat{\sigma}_e^2}$, and the gradient matrix can be computed:

$$\nabla g\left(\hat{\boldsymbol{\sigma}}^2\right) = \left(\frac{\hat{\sigma}_e^2}{\left(\hat{\sigma}_\beta^2 + \hat{\sigma}_e^2\right)^2}, \frac{-\hat{\sigma}_\beta^2}{\left(\hat{\sigma}_\beta^2 + \hat{\sigma}_e^2\right)^2}\right)^{\mathrm{T}}. \tag{12}$$

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

$$\operatorname{Var}\left(\hat{h}^2\right) = \nabla^{\mathrm{T}} g\left(\hat{\boldsymbol{\sigma}}^2\right)\operatorname{Cov}\left(\hat{\boldsymbol{\sigma}}^2\right)\nabla g\left(\hat{\boldsymbol{\sigma}}^2\right). \tag{13}$$

After the variance of $\hat{h}^2$ is obtained, using the **delta theorem**, we know that $g(\hat{\boldsymbol{\sigma}}^2) - g(\boldsymbol{\sigma}_0^2)$ is also **asymptotically normal**, that is:

$$\operatorname{Cov}\left(\hat{h}^2\right)^{-1/2}\left(g\left(\hat{\boldsymbol{\sigma}}^2\right) - g\left(\boldsymbol{\sigma}_0^2\right)\right) \rightarrow_d \mathcal{N}(0, 1), \tag{14}$$

and when $g\left(\hat{\boldsymbol{\sigma}}^2\right) = g\left(\boldsymbol{\sigma}_0^2\right)$, the rejection region is:

$$\left(g\left(\hat{\boldsymbol{\sigma}}^2\right) - g\left(\boldsymbol{\sigma}_0^2\right)\right)^{\mathrm{T}}\operatorname{Cov}\left(\hat{h}^2\right)^{-1}\left(g\left(\hat{\boldsymbol{\sigma}}^2\right) - g\left(\boldsymbol{\sigma}_0^2\right)\right) > \chi^2_{1,\alpha}. \tag{15}$$

3 MM Algorithm

3.1 Derivation

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

$$\mathcal{L}(\mathbf{y} \mid \boldsymbol{\omega}, \sigma_\beta^2, \sigma_e^2; \mathbf{Z}, \mathbf{K}) = -\frac{1}{2} \log\det \boldsymbol{\Sigma} - \frac{1}{2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}), \tag{16}$$

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

$$f\left(\boldsymbol{\theta}^{(t+1)}\right) \geq g\left(\boldsymbol{\theta}^{(t+1)} \mid \boldsymbol{\theta}^{(t)}\right) \geq g\left(\boldsymbol{\theta}^{(t)} \mid \boldsymbol{\theta}^{(t)}\right) = f\left(\boldsymbol{\theta}^{(t)}\right), \tag{17}$$

where $f(\boldsymbol{\theta}) \geq g\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\right)$ and the equality holds true if and only if $\boldsymbol{\theta}$ equals $\boldsymbol{\theta}^{(t)}$. Therefore, the key step of MM algorithm is to identify the surrogate function $g\left(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)}\right)$ by using proper inequalities.

The strategy for maximizing the log-likelihood is to alternate updating the fixed effects $\boldsymbol{\omega}$ and the variance components $\boldsymbol{\sigma}^2 = \left(\sigma_\beta^2, \sigma_e^2\right)$. Updating $\boldsymbol{\omega}$ is a standard general least-squares problem with solution

$$\boldsymbol{\omega}^{(t+1)} = \left(\mathbf{Z}^{\mathrm{T}} \boldsymbol{\Sigma}^{-(t)} \mathbf{Z}\right)^{-1} \mathbf{Z}^{\mathrm{T}} \boldsymbol{\Sigma}^{-(t)} \mathbf{y}, \tag{18}$$

where $\boldsymbol{\Sigma}^{-(t)} = \sigma_\beta^{2(t)} \mathbf{K} + \sigma_e^{2(t)} \mathbf{I}_n$.

Then, Updating $\boldsymbol{\sigma}^2$ given $\boldsymbol{\omega}^{(t)}$ depends on two minorizations.

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

$$-(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}) \geq -(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^{\mathrm{T}} \boldsymbol{\Sigma}^{-(t)} \left(\frac{\sigma_\beta^{4(t)}}{\sigma_\beta^2} \mathbf{K} + \frac{\sigma_e^{4(t)}}{\sigma_e^2} \right) \boldsymbol{\Sigma}^{-(t)} (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}). \tag{19}$$

Second, using the **supporting hyperplane**,

$$-\log\det \boldsymbol{\Sigma} \geq -\log\det \boldsymbol{\Sigma}^{(t)} - \operatorname{tr}\left[\boldsymbol{\Sigma}^{-(t)} \left(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^{(t)}\right)\right]. \tag{20}$$

Combining of the minorizations gives the overall minorization:

$$\begin{aligned}
& g\left(\boldsymbol{\sigma}^2 \mid \boldsymbol{\sigma}^{2(t)}\right) \\
= & -\frac{1}{2} \operatorname{tr}\left(\boldsymbol{\Sigma}^{-(t)} \boldsymbol{\Sigma}\right) - \frac{1}{2} (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^{\mathrm{T}} \boldsymbol{\Sigma}^{-(t)} \left(\frac{\sigma_\beta^{4(t)}}{\sigma_\beta^2} \mathbf{K} + \frac{\sigma_e^{4(t)}}{\sigma_e^2} \right) \boldsymbol{\Sigma}^{-(t)} (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)}) + c^{(t)} \\
= & -\frac{\sigma_\beta^2}{2} \operatorname{tr}\left(\boldsymbol{\Sigma}^{-(t)} \mathbf{K}\right) - \frac{1}{2} \frac{\sigma_\beta^{4(t)}}{\sigma_\beta^2} (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^{\mathrm{T}} \boldsymbol{\Sigma}^{-(t)} \mathbf{K} \boldsymbol{\Sigma}^{-(t)} (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)}) \\
& -\frac{\sigma_e^2}{2} \operatorname{tr}\left(\boldsymbol{\Sigma}^{-(t)}\right) - \frac{1}{2} \frac{\sigma_e^{4(t)}}{\sigma_e^2} (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^{\mathrm{T}} \boldsymbol{\Sigma}^{-2(t)} (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)}) + c^{(t)},
\end{aligned} \tag{21}$$

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

$$\sigma_\beta^{2(t+1)} = \sigma_\beta^{2(t)} \sqrt{\frac{(\mathbf{y}-\mathbf{Z}\boldsymbol{\omega}^{(t)})^{\mathrm{T}}\boldsymbol{\Sigma}^{-(t)}\mathbf{K}\boldsymbol{\Sigma}^{(-t)}(\mathbf{y}-\mathbf{Z}\boldsymbol{\omega}^{(t)})}{\operatorname{tr}\left(\boldsymbol{\Sigma}^{-(t)}\mathbf{K}\right)}}, \tag{22}$$

$$\sigma_e^{2(t+1)} = \sigma_e^{2(t)} \sqrt{\frac{(\mathbf{y}-\mathbf{Z}\boldsymbol{\omega}^{(t)})^{\mathrm{T}}\boldsymbol{\Sigma}^{-2(t)}(\mathbf{y}-\mathbf{Z}\boldsymbol{\omega}^{(t)})}{\operatorname{tr}\left(\boldsymbol{\Sigma}^{-(t)}\right)}}. \tag{23}$$

Since the major computational cost this algorithm is inversion of the covariance matrix $\boldsymbol{\Sigma}$ at each iteration. **Eigenvalue decomposition** can be conduct to boost the efficiency. Using the fact that $\mathbf{K}$ is a real symmetric matrix, we can get $\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$. Since $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$, the covariance matrix $\boldsymbol{\Sigma}$ is given as:

$$\begin{aligned}\boldsymbol{\Sigma} =& \sigma_\beta^2\mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}} + \sigma_e^2\mathbf{I}\\ =& \sigma_\beta^2\mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}} + \sigma_e^2\mathbf{U}\mathbf{U}^{\mathrm{T}}\\ =& \mathbf{U}\left(\sigma_\beta^2\mathbf{D} + \sigma_e^2\mathbf{I}\right)\mathbf{U}^{\mathrm{T}}.\end{aligned} \tag{24}$$

Hence, the inversion of the covariance matrix $\boldsymbol{\Sigma}$ is:

$$\boldsymbol{\Sigma}^{-1} = \mathbf{U}\left(\sigma_\beta^2\mathbf{D} + \sigma_e^2\mathbf{I}\right)^{-1}\mathbf{U}^{\mathrm{T}}. \tag{25}$$

When $\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$, the responses $\mathbf{y}$ should be transformed into $\mathbf{U}^{\mathrm{T}}\mathbf{y}$ and $\mathbf{Z}$ should be transformed into $\mathbf{U}^{\mathrm{T}}\mathbf{Z}$.

3.2 The inverse of Fisher Information Matrix

The covariance matrix of $\hat{\boldsymbol{\sigma}}^2$ can be calculated from the inverse of Fisher Information Matrix (FIM). Hence, the first step is to obtain FIM, that is

$$\mathrm{FIM} = -\mathbb{E}\left[\frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\theta}^2}\right]. \tag{26}$$

The **first derivatives** are:

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

And **the second derivatives** are:

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

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

$$
\begin{aligned}
\mathrm{FIM} &= -\,\mathbb{E}\left[\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta}^2}\right] \\
&= \frac{1}{2}\begin{bmatrix} \operatorname{tr}\left[\left(\boldsymbol{\Sigma}^{-1}\mathbf{K}\right)^2\right] & \operatorname{tr}\left(\boldsymbol{\Sigma}^{-2}\mathbf{K}\right) \\ \operatorname{tr}\left(\boldsymbol{\Sigma}^{-2}\mathbf{K}\right) & \operatorname{tr}\left[\boldsymbol{\Sigma}^{-2}\right] \end{bmatrix}.
\end{aligned}
\tag{27}
$$

Therefore, the covariance matrix of $\hat{\boldsymbol{\sigma}}^2$ is the inverse of FIM.

References

[1] Wu, Y., & Sankararaman, S. (2018). A scalable estimator of SNP heritability for biobank-scale data. *Bioinformatics*, 34(13), i187-i194.

[2] Zhou, H., Hu, L., Zhou, J., & Lange, K. (2018). MM algorithms for variance components models. *Journal of Computational and Graphical Statistics*, 28(2), 350-361.

[3] Freedman, D. A. (2006). On the so-called ”Huber sandwich estimator” and ”robust standard errors”. *The American Statistician*, 60(4), 299-302.

[4] Zhou, C. Lecture Notes on Asymptotic Statistics.

[5] Wu, H. MM Algorithm.