Vignette of Variance Components Model

CAI Mingxuan

2019/2/25

This package contains three approaches for estimating variance components of linear model [1]: the parameter expanded EM (PX-EM) algorithm [2, 3], Minorization-Maximization (MM) algotirhm [4] and the Method of Moments (MoM) [5].

Variance components model

Suppose we have dataset $\{\mathbf{X}, \mathbf{Z}, \mathbf{y}\}$ where $\mathbf{X} \in \mathbb{R}^{n \times p}$ is the design matrix whose columns are normalized with mean 0 and variance $1/p$, $\mathbf{y} \in \mathbb{R}^n$ is the response vector and $\mathbf{Z} \in \mathbb{R}^{n \times c}$ is the covariate matrix of fixed effects. The linear mixed model links $\mathbf{y}$ with $\mathbf{X}$ and $\mathbf{Z}$:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\omega} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad \boldsymbol{\beta} \sim \mathcal{N}(0, \sigma_\beta^2 \mathbf{I}_p), \quad \mathbf{e} \sim \mathcal{N}(0, \sigma_e^2 \mathbf{I}_n), \tag{1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is the random effect, $\boldsymbol{\omega} \in \mathbb{R}^c$ is the fixed effect and σ_β^2 and σ_e^2 are model parameters. This linear mixed model can also be re-written as a variance components model:

$$\mathbf{y} = \mathcal{N}(\mathbf{Z}\boldsymbol{\omega}, \sigma_\beta^2 \mathbf{K} + \sigma_e^2 \mathbf{I}_n), \tag{2}$$

where $\mathbf{K} = \mathbf{X}\mathbf{X}^T$. Since $\mathbf{X}$ has been normalized with mean 0 and variance $1/p$, $\text{tr}(K) = \text{tr}(\mathbf{I}_n) = n$. The goal is to estimate the variance components $\boldsymbol{\theta} = \{\sigma_\beta^2, \sigma_e^2\}$ [1]. This package provides three approaches to estimate the parameters: PX-EM algorithm, MM algorithm and the method of moments. The first two are based on maximum likelihood (MLE) approach and the third one adopts the moment matching approach.

PX-EM algorithm

The PX-EM algorithm is an extension of classical EM algorithm with faster speed [2, 3]. We first consider the parameter expanded version of (1):

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\omega} + \delta\mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

where $\delta \in \mathbb{R}^1$ is the expanded parameter. The complete-data log-likelihood is given as

$$
\begin{aligned}
\mathcal{L} &= \log \Pr(\mathbf{y}, \boldsymbol{\beta} | \boldsymbol{\theta}; \mathbf{Z}, \mathbf{X}) \\
&= -\frac{n}{2} \log(2\pi\sigma_e^2) - \frac{1}{2\sigma_e^2} ||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega} - \delta\mathbf{X}\boldsymbol{\beta}||^2 \\
&\quad - \frac{p}{2} \log(2\pi\sigma_\beta^2) - \frac{1}{2\sigma_\beta^2} ||\boldsymbol{\beta}||^2,
\end{aligned}
\tag{3}
$$

from which we can easily recognize that the terms involving $\boldsymbol{\beta}$ are of a quadratic form:

$$\boldsymbol{\beta}^T (-\frac{\delta^2}{2\sigma_e^2} \mathbf{X}^T\mathbf{X} - \frac{1}{2\sigma_\beta^2} \mathbf{I}_p)\boldsymbol{\beta} + \frac{\delta}{\sigma_e^2} (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T \mathbf{X}\boldsymbol{\beta} + \text{Constant}.$$

Therefore, the posterior distribution of $\boldsymbol{\beta}$ is Gaussian $\mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma}^{-1} = \frac{\delta^2}{\sigma_e^2}\mathbf{X}^T\mathbf{X} + \frac{1}{\sigma_\beta^2}\mathbf{I}_p,$$

$$\boldsymbol{\mu} = (\frac{\delta^2}{\sigma_e^2}\mathbf{X}^T\mathbf{X} + \frac{1}{\sigma_\beta^2}\mathbf{I}_p)^{-1}\frac{\delta}{\sigma_e^2}\mathbf{X}^T(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}).$$

Now in the E-step, we evaluate the Q-function by taking the expectation of the complete-data log-likelihood with respect to the posterior $\mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Specifically, the quadratic terms involving $\hat{\boldsymbol{\beta}}$ are evaluated as following:

$$
\begin{aligned}
\mathbb{E}[||\tilde{\mathbf{y}} - \delta\mathbf{X}\boldsymbol{\beta}||^2] &= \mathbb{E}[\tilde{\mathbf{y}}^T\tilde{\mathbf{y}} - 2\delta\tilde{\mathbf{y}}^T\mathbf{X}\boldsymbol{\beta} + \delta^2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}] \\
&= \tilde{\mathbf{y}}^T\tilde{\mathbf{y}} - 2\delta\tilde{\mathbf{y}}^T\mathbf{X}\boldsymbol{\mu} + \delta^2\boldsymbol{\mu}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\mu} + \delta^2\mathrm{tr}(\mathbf{X}^T\mathbf{X}\boldsymbol{\Sigma}), \\
\mathbb{E}[||\boldsymbol{\beta}||^2] &= \boldsymbol{\mu}^T\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{\Sigma}),
\end{aligned}
$$

where $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{Z}\boldsymbol{\omega}$. Then the Q-function given the current parameter estimtes $\boldsymbol{\theta}_{old}$ is obtained as:

$$
\begin{aligned}
\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}_{old}) = &-\frac{n}{2}\log(2\pi\sigma_e^2) - \frac{p}{2}\log(2\pi\sigma_\beta^2) \\
&- \frac{1}{2\sigma_e^2}||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega} - \delta\mathbf{X}\boldsymbol{\mu}||^2 - \frac{1}{2\sigma_\beta^2}\boldsymbol{\mu}^T\boldsymbol{\mu} \\
&- \mathrm{tr}\left(\left(\frac{\delta^2}{2\sigma_e^2}\mathbf{X}^T\mathbf{X} + \frac{1}{2\sigma_\beta^2}\mathbf{I}_p\right)\boldsymbol{\Sigma}\right).
\end{aligned}
\tag{4}
$$

It the M-step, the new estimates of parameter $\boldsymbol{\theta}$ is obtained by setting the derivative of Q-function to be zero. The resulting updates are given as follows:

$$
\begin{aligned}
\delta &= \frac{(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T\mathbf{X}\boldsymbol{\mu}}{\boldsymbol{\mu}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\mu} + \mathrm{tr}(\mathbf{X}^T\mathbf{X}\boldsymbol{\Sigma})}, \\
\boldsymbol{\omega} &= (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T(\mathbf{y} - \delta\mathbf{X}\boldsymbol{\mu}), \\
\sigma_e^2 &= \frac{1}{n}[||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega} - \delta\mathbf{X}\boldsymbol{\mu}||^2 + \delta^2\mathrm{tr}(\mathbf{X}^T\mathbf{X}\boldsymbol{\Sigma})], \\
\sigma_\beta^2 &= \frac{1}{p}[\boldsymbol{\mu}^T\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{\Sigma})].
\end{aligned}
$$

To check the convergence of PX-EM algorithm, we evaluate the lower bound after each E-step, when the incomplete-data log-likelihood is exactly equal to the lower bound (i.e. the bound is tight).

This PX-EM algorithm is summarized in Algorithm 1. After convergence, the posterior mean and variance of $\mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be evaluated given the obtained parameter estimates $\hat{\boldsymbol{\theta}} = \{\hat{\sigma}_e^2, \hat{\sigma}_\beta^2\}$ and $\hat{\boldsymbol{\omega}}$:

$$
\begin{aligned}
\boldsymbol{\Sigma}^{-1} &= \frac{1}{\hat{\sigma}_e^2}\mathbf{X}^T\mathbf{X} + \frac{1}{\hat{\sigma}_\beta^2}\mathbf{I}_p, \\
\boldsymbol{\mu} &= \left(\frac{1}{\hat{\sigma}_e^2}\mathbf{X}^T\mathbf{X} + \frac{1}{\hat{\sigma}_\beta^2}\mathbf{I}_p\right)^{-1}\frac{1}{\hat{\sigma}_e^2}\mathbf{X}^T(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\omega}}).
\end{aligned}
\tag{5}
$$

Algorithm 1 PX-EM algorithm for model (1)

Initialization: Parameters are initialized by setting $\boldsymbol{\omega} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}$, $\sigma_e^2 = \sigma_\beta^2 = \text{Var}(y - \mathbf{Z}\boldsymbol{\omega})/2$.

repeat

    E-step: At the t-th iteration, evaluate the posterior $\mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ given the current parameter estimates $\boldsymbol{\theta}^{(t)} = \{(\sigma_e^{(t)})^2, (\sigma_\beta^{(t)})^2\}$, $\boldsymbol{\omega}^{(t)}$ and set $\delta^{(t)} = 1$:

$$\boldsymbol{\Sigma}^{-1} = \frac{(\delta^{(t)})^2}{(\sigma_e^{(t)})^2}\mathbf{X}^T\mathbf{X} + \frac{1}{(\sigma_\beta^{(t)})^2}\mathbf{I}_p,$$

$$\boldsymbol{\mu} = \left(\frac{(\delta^{(t)})^2}{(\sigma_e^{(t)})^2}\mathbf{X}^T\mathbf{X} + \frac{1}{(\sigma_\beta^{(t)})^2}\mathbf{I}_p\right)^{-1}\frac{(\delta^{(t)})^2}{(\sigma_e^{(t)})^2}\mathbf{X}^T(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)}),$$

$$ELBO^{(t)} = \mathcal{Q}(\boldsymbol{\theta}^{(t)}) + \frac{1}{2}\log|2\pi\boldsymbol{\Sigma}|, \text{ where } \mathcal{Q} \text{ is defined in Equation(4)}.$$

    M-step: Update the model parameters by

$$\delta^{(t+1)} = \frac{(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^T\mathbf{X}\boldsymbol{\mu}}{\boldsymbol{\mu}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\mu} + \text{tr}(\mathbf{X}^T\mathbf{X}\boldsymbol{\Sigma})},$$

$$\boldsymbol{\omega}^{(t+1)} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T(\mathbf{y} - \delta\mathbf{X}\boldsymbol{\mu}),$$

$$(\sigma_e^{(t+1)})^2 = \frac{1}{n_r}[||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)} - \delta\mathbf{X}\boldsymbol{\mu}||^2 + \delta^2\text{tr}(\mathbf{X}^T\mathbf{X}\boldsymbol{\Sigma})],$$

$$(\sigma_\beta^{(t+1)})^2 = \frac{1}{p}[\boldsymbol{\mu}^T\boldsymbol{\mu} + \text{tr}(\boldsymbol{\Sigma})].$$

    Reduction-step: Rescale $(\sigma_\beta^{(t+1)})^2 = (\delta^{(t+1)})^2(\sigma_\beta^{(t+1)})^2$ and reset $\delta^{(t+1)} = 1$.

until the incomplete-data log-likelihood ($ELBO^{(t)}$) stop increasing or maximum iteration reached

---

In practice. We can avoid frequently inverting the $p \times p$ matrix $\mathbf{\Sigma}$ by conducting a single eigen-dedomposition on $\mathbf{X}\mathbf{X}^T$ or $\mathbf{X}^T\mathbf{X}$, depending on the relative sizes of p and n. If $n \geq p$, we conduct the eigen-decomposition $\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{Q}\mathbf{V}^T$ before the iteration, where $\mathbf{Q} \in \mathbb{R}^{p \times p}$ is a diagonal matrix of eigenvalues q_j and $\mathbf{V} \in \mathbb{R}^{p \times p}$ is a matrix whose columns are corresponding eigenvectors of q_j. The resulting algorithm is shown in Algorithm 2.

Algorithm 2 PX-EM algorithm for model (1) when $n \geq p$

Initialization: $\boldsymbol{\omega} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}$, $\sigma_e^2 = \sigma_\beta^2 = \text{Var}(y - \mathbf{Z}\boldsymbol{\omega})/2$; conduct eigen-deomposition $\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{Q}\mathbf{V}^T$.

repeat
   E-step: At the t-th iteration, evaluate the posterior $\mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\mu}, \mathbf{\Sigma})$ given the current parameter estimates $\boldsymbol{\theta}^{(t)} = \{(\sigma_e^{(t)})^2, (\sigma_\beta^{(t)})^2\}$, $\boldsymbol{\omega}^{(t)}$ and set $\delta^{(t)} = 1$:

$$\tilde{q}_j = q_j/(\sigma_e^{(t)})^2 + 1/(\sigma_\beta^{(t)})^2, \ \text{diag}(\tilde{\mathbf{Q}}) = \tilde{q} = [\tilde{q}_1, ...\tilde{q}_p]$$

$$\boldsymbol{\mu} = \frac{1}{(\sigma_e^{(t)})^2}\mathbf{V}[\mathbf{V}^T\mathbf{X}^T(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)}) \odot 1/\tilde{q}]$$

$$ELBO^{(t)} = -\frac{n}{2}\log(2\pi(\sigma_e^{(t)})^2) - \frac{p}{2}\log(2\pi(\sigma_\beta^{(t)})^2) - \frac{1}{2(\sigma_e^{(t)})^2}||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)} - \delta\mathbf{X}\boldsymbol{\mu}||^2$$

$$- \frac{1}{2(\sigma_\beta^{(t)})^2}\boldsymbol{\mu}^T\boldsymbol{\mu} - \frac{1}{2}\sum_j^p \log \tilde{q}_j.$$

   M-step: Update the model parameters by

$$\delta^{(t+1)} = \frac{(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^T\mathbf{X}\boldsymbol{\mu}}{\boldsymbol{\mu}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\mu} + \sum_j^p q_j/\tilde{q}_j},$$

$$\boldsymbol{\omega}^{(t+1)} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T(\mathbf{y} - \delta\mathbf{X}\boldsymbol{\mu}),$$

$$(\sigma_e^{(t+1)})^2 = \frac{1}{n_r}[||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)} - \delta\mathbf{X}\boldsymbol{\mu}||^2 + \delta^2\sum_j^p q_j/\tilde{q}_j],$$

$$(\sigma_\beta^{(t+1)})^2 = \frac{1}{p}[\boldsymbol{\mu}^T\boldsymbol{\mu} + \sum_j^p 1/\tilde{q}_j].$$

   Reduction-step: Rescale $(\sigma_\beta^{(t+1)})^2 = (\delta^{(t+1)})^2(\sigma_\beta^{(t+1)})^2$ and reset $\delta^{(t+1)} = 1$.
   until the incomplete-data log-likelihood ($ELBO^{(t)}$) stop increasing or maximum iteration reached

---

If $p > n$, we conduct the eigen-decomposition $\mathbf{X}\mathbf{X}^T = \mathbf{U}\mathbf{D}\mathbf{U}^T$ before the iteration, where $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a diagonal matrix of eigenvalues d_i and $\mathbf{U} \in \mathbb{R}^{n \times n}$ is a matrix whose columns are corresponding eigenvectors of d_i. The resulting algorithm is shown in Algorithm 3.

MM algorithm

Unlike the PX-EM algorithm, the MM algorithm maximize the incomplete-data log-likelihood by considering the variance components model (2) [4]. The incomplete-data log-likelihood is given as

$$\log \Pr(\mathbf{y}|\boldsymbol{\theta}; \mathbf{Z}, \mathbf{K}) = -\frac{1}{2}\log|\mathbf{\Omega}| - \frac{1}{2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T\mathbf{\Omega}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}), \tag{6}$$

where $\mathbf{\Omega} = \sigma_\beta^2\mathbf{K} + \sigma_e^2\mathbf{I}_n$. The MM algorithm updates $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$ alternatively by iteratively maximizing the

---

Algorithm 3 PX-EM algorithm for model (1) when $p > n$

Initialization: $\boldsymbol{\omega} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}$, $\sigma_e^2 = \sigma_\beta^2 = \text{Var}(y - \mathbf{Z}\boldsymbol{\omega})/2$; conduct eigen-deomposition $\mathbf{X}\mathbf{X}^T = \mathbf{U}\mathbf{D}\mathbf{U}^T$.

repeat

   E-step: At the t-th iteration, evaluate the posterior $\mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ given the current parameter estimates $\boldsymbol{\theta}^{(t)} = \{(\sigma_e^{(t)})^2, (\sigma_\beta^{(t)})^2\}$, $\boldsymbol{\omega}^{(t)}$ and set $\delta^{(t)} = 1$:

$$\tilde{d}_i = d_i/(\sigma_e^{(t)})^2 + 1/(\sigma_\beta^{(t)})^2, \ \text{diag}(\tilde{\mathbf{D}}) = \tilde{d} = [\tilde{d}_i, ...\tilde{d}_n]$$

$$\boldsymbol{\mu} = \frac{1}{(\sigma_e^{(t)})^2}\mathbf{X}^T\mathbf{U}[\mathbf{U}^T(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)}) \odot 1/\tilde{d}]$$

$$ELBO^{(t)} = -\frac{n}{2}\log(2\pi(\sigma_e^{(t)})^2) - \frac{p}{2}\log(2\pi(\sigma_\beta^{(t)})^2) - \frac{1}{2(\sigma_e^{(t)})^2}||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)} - \delta\mathbf{X}\boldsymbol{\mu}||^2$$

$$- \frac{1}{2(\sigma_\beta^{(t)})^2}\boldsymbol{\mu}^T\boldsymbol{\mu} - \frac{1}{2}\sum_i^n \log\tilde{d}_i + \frac{p-n}{2}\log(\sigma_\beta^{(t)})^2.$$

   M-step: Update the model parameters by

$$\delta^{(t+1)} = \frac{(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^T\mathbf{X}\boldsymbol{\mu}}{\boldsymbol{\mu}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\mu} + \sum_i^n d_i/\tilde{d}_i},$$

$$\boldsymbol{\omega}^{(t+1)} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T(\mathbf{y} - \delta\mathbf{X}\boldsymbol{\mu}),$$

$$(\sigma_e^{(t+1)})^2 = \frac{1}{n_r}[||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)} - \delta\mathbf{X}\boldsymbol{\mu}||^2 + \delta^2\sum_i^n d_i/\tilde{d}_i],$$

$$(\sigma_\beta^{(t+1)})^2 = \frac{1}{p}[\boldsymbol{\mu}^T\boldsymbol{\mu} + \sum_i^n 1/\tilde{d}_i + (n - p)(\sigma_\beta^{(t)})^2].$$

   Reduction-step: Rescale $(\sigma_\beta^{(t+1)})^2 = (\delta^{(t+1)})^2(\sigma_\beta^{(t+1)})^2$ and reset $\delta^{(t+1)} = 1$.

until the incomplete-data log-likelihood ($ELBO^{(t)}$) stop increasing or maximum iteration reached

---

lower bound of incomplete-data log-likelihood. Given $\boldsymbol{\theta}^{(t)}$, the updata of $\boldsymbol{\omega}$ is simply a weighted least square problem

$$\boldsymbol{\omega}^{(t)} = (\mathbf{Z}^T \boldsymbol{\Omega}^{(t)} \mathbf{Z})^{-1} \mathbf{Z}^T (\boldsymbol{\Omega}^{(t)})^{-1} \mathbf{y}. \tag{7}$$

The update of $\boldsymbol{\theta}$ given $\boldsymbol{\omega}$ depends on two minorizations. First,

$$- (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T (\boldsymbol{\Omega}^{(t)})^{-1} (\frac{(\sigma_\beta^{(t)})^4}{\sigma_\beta^2} \mathbf{K} + \frac{(\sigma_e^{(t)})^4}{\sigma_e^2})(\boldsymbol{\Omega}^{(t)})^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}) \leq -\frac{1}{2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T \boldsymbol{\Omega}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}), \tag{8}$$

which separates the variance components in the quadratic term of the likelihood (6). Second, the convexity of function $-\log |\boldsymbol{\Omega}|$ implies that

$$- \log |\boldsymbol{\Omega}^{(t)}| - \text{tr}((\boldsymbol{\Omega}^{(t)})^{-1}(\boldsymbol{\Omega} - \boldsymbol{\Omega}^{(t)})) \leq -\frac{1}{2} \log |\boldsymbol{\Omega}|, \tag{9}$$

which separates the variance components in the log determinant of the likelihood (6). Combining (8) and (9), the overall minorization if given as

$$\begin{aligned}
&\mathcal{G}(\boldsymbol{\theta}|\boldsymbol{\theta}_{old}) \\
&= - \log |\boldsymbol{\Omega}^{(t)}| - \frac{1}{2}\text{tr}((\boldsymbol{\Omega}^{(t)})^{-1}\boldsymbol{\Omega}) - \frac{1}{2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^T (\boldsymbol{\Omega}^{(t)})^{-1}(\frac{(\sigma_\beta^{(t)})^4}{\sigma_\beta^2}\mathbf{K} + \frac{(\sigma_e^{(t)})^4}{\sigma_e^2})(\boldsymbol{\Omega}^{(t)})^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)}) \\
&= - \log |\boldsymbol{\Omega}^{(t)}| - \frac{\sigma_\beta^2}{2}\text{tr}((\boldsymbol{\Omega}^{(t)})^{-1}\mathbf{K}) - \frac{1}{2}\frac{(\sigma_\beta^{(t)})^4}{\sigma_\beta^2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^T(\boldsymbol{\Omega}^{(t)})^{-1}\mathbf{K}(\boldsymbol{\Omega}^{(t)})^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)}) \\
&\quad - \frac{\sigma_e^2}{2}\text{tr}((\boldsymbol{\Omega}^{(t)})^{-1}) - \frac{1}{2}\frac{(\sigma_e^{(t)})^4}{\sigma_e^2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^T(\boldsymbol{\Omega}^{(t)})^{-2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)}).
\end{aligned} \tag{10}$$

By setting the derivative of $\mathcal{G}$-function to be zero, the resulting updates are given as follows:

$$\begin{aligned}
(\sigma_\beta^{(t+1)})^2 &= (\sigma_\beta^{(t)})^2 \sqrt{\frac{(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^T(\boldsymbol{\Omega}^{(t)})^{-1}\mathbf{K}(\boldsymbol{\Omega}^{(t)})^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})}{\text{tr}((\boldsymbol{\Omega}^{(t)})^{-1}\mathbf{K})}} \\
(\sigma_e^{(t+1)})^2 &= (\sigma_e^{(t)})^2 \sqrt{\frac{(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^T(\boldsymbol{\Omega}^{(t)})^{-2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})}{\text{tr}((\boldsymbol{\Omega}^{(t)})^{-1})}}.
\end{aligned}$$

The convergence of MM algorithm is checked by evaluating the log-likelihood at each iteration. The resulting algorithm is summarized in Algorithm 4.

To avoid frequently inverting $\boldsymbol{\Omega}$ in the iterations, we can conduct eigen-decomposition on $\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ before the algorithm, where $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a diagonal matrix of eigenvalues d_i and $\mathbf{U} \in \mathbb{R}^{n \times n}$ is a matrix whose columns are corresponding eigenvectors of d_i. The resulting procedure is summarized in Algorithm 5.

Standard errors of variance components for MLE methods

For MLE methods including MM and PX-EM, the covariance matrix of variance components estimates are calculated from inverse of Fisher Information Matrix (FIM). $FIM = -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta}^2}\right] = -\mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} - \frac{1}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})\right]$. The first derivatives are:

Algorithm 4 MM algorithm for model (2)

Initialization: Parameters are initialized by setting $\boldsymbol{\omega} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}$, $\sigma_e^2 = \sigma_\beta^2 = \mathrm{Var}(y - \mathbf{Z}\boldsymbol{\omega})/2$.
repeat

$$\boldsymbol{\Omega}^{(t)} = (\sigma_\beta^{(t)})^2\mathbf{K} + (\sigma_e^{(t+1)})^2\mathbf{I}_n,$$

$$\boldsymbol{\omega}^{(t)} = (\mathbf{Z}^T\boldsymbol{\Omega}^{(t)}\mathbf{Z})^{-1}\mathbf{Z}^T(\boldsymbol{\Omega}^{(t)})^{-1}\mathbf{y},$$

evaluate $\mathcal{L}^{(t)}(\boldsymbol{\Omega}^{(t)}, \boldsymbol{\omega}^{(t)})$ from Equation (6),

$$(\sigma_\beta^{(t+1)})^2 = (\sigma_\beta^{(t)})^2\sqrt{\frac{(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^T(\boldsymbol{\Omega}^{(t)})^{-1}\mathbf{K}(\boldsymbol{\Omega}^{(t)})^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})}{\mathrm{tr}((\boldsymbol{\Omega}^{(t)})^{-1}\mathbf{K})}},$$

$$(\sigma_e^{(t+1)})^2 = (\sigma_e^{(t)})^2\sqrt{\frac{(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})^T(\boldsymbol{\Omega}^{(t)})^{-2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}^{(t)})}{\mathrm{tr}((\boldsymbol{\Omega}^{(t)})^{-1})}}.$$

until the incomplete-data log-likelihood ($\mathcal{L}^{(t)}$) stop increasing or maximum iteration reached

---

Algorithm 5 Efficient MM algorithm for model (2)

Initialization: $\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^T$, $\bar{\mathbf{Z}} = \mathbf{U}^T\mathbf{Z}$, $\bar{\mathbf{y}} = \mathbf{U}^T\mathbf{y}$, $\boldsymbol{\omega} = (\bar{\mathbf{Z}}^T\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}^T\bar{\mathbf{y}}$, $\sigma_e^2 = \sigma_\beta^2 = \mathrm{Var}(y - \mathbf{Z}\boldsymbol{\omega})/2$.
repeat

$$\tilde{d}_i = d_i/(\sigma_e^{(t)})^2 + 1/(\sigma_\beta^{(t)})^2, \ \mathrm{diag}(\tilde{\mathbf{D}}) = \tilde{d} = [\tilde{d}_i, ...\tilde{d}_n]$$

$$\boldsymbol{\omega}^{(t)} = (\bar{\mathbf{Z}}^T\tilde{\mathbf{D}}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}^T(\bar{\mathbf{y}} \odot \tilde{d}),$$

$$\mathcal{L}^{(t)}(\boldsymbol{\Omega}^{(t)}, \boldsymbol{\omega}^{(t)}) = -\frac{1}{2}\sum_i^n \log\tilde{d}_i - \frac{n}{2}\log\sigma_\beta^2 - \frac{n}{2}\log\sigma_e^2 - \frac{n}{2}\log 2\pi - \frac{1}{2}\sum_i^n[(\bar{y}_i - \bar{\mathbf{Z}}_i^T\boldsymbol{\omega}^{(t)})^2/\tilde{d}_i]$$

$$(\sigma_\beta^{(t+1)})^2 = \frac{\sigma_\beta^{(t)}}{\sigma_e^{(t)}}\sqrt{\frac{\sum_i^n[(\bar{y}_i - \bar{\mathbf{Z}}_i^T\boldsymbol{\omega}^{(t)})^2 d_i/\tilde{d}_i^2]}{\sum_i^n d_i/\tilde{d}_i}},$$

$$(\sigma_e^{(t+1)})^2 = \frac{\sigma_e^{(t)}}{\sigma_\beta^{(t)}}\sqrt{\frac{\sum_i^n[(\bar{y}_i - \bar{\mathbf{Z}}_i^T\boldsymbol{\omega}^{(t)})^2/\tilde{d}_i^2]}{\sum_i^n 1/\tilde{d}_i}}.$$

until the incomplete-data log-likelihood ($\mathcal{L}^{(t)}$) stop increasing or maximum iteration reached

---

$$\frac{\partial \mathcal{L}}{\partial \sigma_g^2} = \frac{1}{2}\text{tr}\left[-\mathbf{\Omega}^{-1}\mathbf{K} + (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T\mathbf{\Omega}^{-1}\mathbf{K}\mathbf{\Omega}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})\right],$$

$$\frac{\partial \mathcal{L}}{\partial \sigma_e^2} = \frac{1}{2}\text{tr}\left[-\mathbf{\Omega}^{-1} + (\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T\mathbf{\Omega}^{-2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})\right];$$

and the second derivatives are given as

$$\frac{\partial^2 \mathcal{L}}{\partial (\sigma_g^2)^2} = \frac{1}{2}\text{tr}\left[(\mathbf{\Omega}^{-1}\mathbf{K})^2 - 2(\mathbf{\Omega}^{-1}\mathbf{K})^2\mathbf{\Omega}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T\right],$$

$$\frac{\partial^2 \mathcal{L}}{\partial (\sigma_e^2)^2} = \frac{1}{2}\text{tr}\left[\mathbf{\Omega}^{-2} - 2\mathbf{\Omega}^{-3}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T\right],$$

$$\frac{\partial^2 \mathcal{L}}{\partial \sigma_g^2 \partial \sigma_e^2} = \frac{1}{2}\text{tr}\left[\mathbf{\Omega}^{-1}\mathbf{K}\mathbf{\Omega}^{-1} - (\mathbf{\Omega}^{-1}\mathbf{K}\mathbf{\Omega}^{-2} + \mathbf{\Omega}^{-2}\mathbf{K}\mathbf{\Omega}^{-1})(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T\right].$$

Since the the only random variable is $\mathbf{y}$, and $\mathbb{E}[(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T] = \mathbf{\Omega}$, the FIM is

$$\begin{aligned}
FIM &= -\frac{1}{2}\begin{bmatrix} \text{tr}[(\mathbf{\Omega}^{-1}\mathbf{K})^2] - 2\text{tr}[(\mathbf{\Omega}^{-1}\mathbf{K})^2] & \text{tr}(\mathbf{\Omega}^{-1}\mathbf{K}\mathbf{\Omega}^{-1}) - 2\text{tr}(\mathbf{\Omega}^{-1}\mathbf{K}\mathbf{\Omega}^{-1}) \\ \cdot & \text{tr}[\mathbf{\Omega}^{-2}] - 2\text{tr}[\mathbf{\Omega}^{-2}] \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} \text{tr}[(\mathbf{\Omega}^{-1}\mathbf{K})^2] & \text{tr}(\mathbf{\Omega}^{-1}\mathbf{K}\mathbf{\Omega}^{-1}) \\ \cdot & \text{tr}[\mathbf{\Omega}^{-2}] \end{bmatrix}.
\end{aligned}$$

Inverting the FIM leads to the covariance matrix of $\hat{\boldsymbol{\theta}}$.

When handling the FIM, we can again make use of the pre-calculated eigenvectors and eigenvalues to avoid inverting $\mathbf{\Omega}$. For MM algorithm and PX-EM algorithm with $p \geq n$ case, we can evaluate $\mathbf{\Omega}^{-1}$ using the identity $\mathbf{\Omega}^{-1} = \mathbf{U}\tilde{\mathbf{D}}\mathbf{U}^T$ with $\mathbf{U}$, $\tilde{\mathbf{D}}$ from Algorithm 3 and 5. For PX-EM algorithm with $n > p$, we first define $\mathbf{\Lambda}^{-1} = (\sigma_e^2\mathbf{I}_p + \sigma_\beta^2\mathbf{X}^T\mathbf{X})^{-1}$. Then using the matrix inverse lemma, we have $\mathbf{X}^T\mathbf{\Omega}^{-1} = \mathbf{\Lambda}^{-1}\mathbf{X}^T$. Therefore, we can express the FIM using the dual form:

$$\begin{aligned}
\text{tr}[(\mathbf{\Omega}^{-1}\mathbf{K})^2] &= \text{tr}(\mathbf{\Omega}^{-1}\mathbf{X}\mathbf{X}^T\mathbf{\Omega}^{-1}\mathbf{X}\mathbf{X}^T) = \text{tr}(\mathbf{\Lambda}^{-1}\mathbf{X}^T\mathbf{X}\mathbf{\Lambda}^{-1}\mathbf{X}^T\mathbf{X}) \\
\text{tr}[\mathbf{\Omega}^{-1}\mathbf{K}\mathbf{\Omega}^{-1}] &= \text{tr}[\mathbf{\Omega}^{-1}\mathbf{X}\mathbf{X}^T\mathbf{\Omega}^{-1}] = \text{tr}(\mathbf{\Lambda}^{-2}\mathbf{X}^T\mathbf{X}) \\
\text{tr}[\mathbf{\Omega}^{-2}] &= n(\frac{1}{\sigma_e^2})^2 - 2\frac{\sigma_\beta^2}{(\sigma_e^2)^2}\text{tr}[\mathbf{\Lambda}^{-1}\mathbf{X}^T\mathbf{X}] + (\frac{\sigma_\beta^2}{\sigma_e^2})^2\text{tr}[\mathbf{\Lambda}^{-1}\mathbf{X}^T\mathbf{X}\mathbf{\Lambda}^{-1}\mathbf{X}^T\mathbf{X}],
\end{aligned}$$

where $\mathbf{\Lambda}^{-1} = \mathbf{V}\tilde{\mathbf{Q}}\mathbf{V}^T$ with $\mathbf{V}$ and $\tilde{\mathbf{Q}}$ from Algorithm 2.

Method of Moments

While the MM algorithm and PX-EM algorithm adopts the MLE, MoM estimator is obtained by first multiplying Equation (2) by the projection matrix $\mathbf{M} = \mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$ and then solving the following ordinary least squares (OLS) problem [5]:

$$argmin_{\sigma_\beta^2, \sigma_e^2}||(\mathbf{M}\mathbf{y})(\mathbf{M}\mathbf{y})^T - (\sigma_\beta^2\mathbf{M}\mathbf{K}\mathbf{M} + \sigma_e^2\mathbf{M})||_F^2. \tag{11}$$

Using the fact that $||\mathbf{A}||_F = \sqrt{\text{tr}(\mathbf{A}\mathbf{A}^T)}$, the OLS problem in (11) can be re-written as

$$argmin_{\sigma_\beta^2, \sigma_e^2} \mathrm{tr}[((\mathbf{My})(\mathbf{My})^T - (\sigma_\beta^2 \mathbf{MKM} + \sigma_e^2 \mathbf{M}))((\mathbf{My})(\mathbf{My})^T - (\sigma_\beta^2 \mathbf{MKM} + \sigma_e^2 \mathbf{M}))^T],$$

which leads to the normal equation

$$\mathbf{S}\boldsymbol{\theta} = \mathbf{q}, \tag{12}$$

$$\text{with } \mathbf{S} = \begin{bmatrix} \mathrm{tr}(\mathbf{MKMK}) & \mathrm{tr}(\mathbf{MK}) \\ \mathrm{tr}(\mathbf{MK}) & n-c \end{bmatrix}, \ \boldsymbol{\theta} = \begin{bmatrix} \sigma_\beta^2 \\ \sigma_e^2 \end{bmatrix}, \ \mathbf{q} = \begin{bmatrix} \mathbf{y}^T \mathbf{MKMy} \\ \mathbf{y}^T \mathbf{My} \end{bmatrix}.$$

The MoM estimates of $\boldsymbol{\theta}$ is then given by $\hat{\boldsymbol{\theta}} = \mathbf{S}^{-1}\mathbf{q}$. The covariance matrix of MoM estimators are given by the sanwich estimator: $\boldsymbol{\Sigma_\theta} = \mathbb{E}\left[\frac{\partial B}{\partial \boldsymbol{\theta}}\right]^{-1} \mathrm{Cov}(B) \mathbb{E}\left[\frac{\partial B}{\partial \boldsymbol{\theta}}\right]^{-1}$, where B is the normal equation $\mathbf{q} - \mathbf{S}\boldsymbol{\theta}$. Specifically,

$$\mathbb{E}\left[\frac{\partial B}{\partial \theta}\right]^{-1} = \mathbf{S}^{-1}, \tag{13}$$

and

$$\mathrm{Cov}(B) = \mathrm{Cov}\left(\begin{bmatrix} \mathbf{y}^T \mathbf{MKMy} \\ \mathbf{y}^T \mathbf{My} \end{bmatrix}\right) = \begin{bmatrix} \mathrm{Var}(\mathbf{y}^T \mathbf{MKMy}) & \mathrm{Cov}(\mathbf{y}^T \mathbf{MKMy}, \mathbf{y}^T \mathbf{My}) \\ \mathrm{Cov}(\mathbf{y}^T \mathbf{MKMy}, \mathbf{y}^T \mathbf{MKMy}) & \mathrm{Var}(\mathbf{y}^T \mathbf{My}) \end{bmatrix}, \tag{14}$$

where the elements are calculated by $\mathrm{Var}(\mathbf{y}^T \mathbf{MKMy}) = 2\mathrm{tr}([\mathbf{MKM\Omega}]^2)$, $\mathrm{Var}(\mathbf{y}^T \mathbf{My}) = 2\mathrm{tr}([\mathbf{M\Omega}]^2)$, $\mathrm{Cov}(\mathbf{y}^T \mathbf{MKMy}, \mathbf{y}^T \mathbf{My}) = 2\mathrm{tr}(\mathbf{MKM\Omega M\Omega})$.

Example

```r
library(VCM)
n <- 1000
d <- 1000

sb2 <- 0.1
se2 <- 1
X <- matrix(rnorm(n*d),n,d)
X <- scale(X)/sqrt(d)

w  <- c(rnorm(d,0,sqrt(sb2)))
y0 <- X%*%w

y  <- y0 + sqrt(se2)*rnorm(n)

fit_PXEM <- linRegPXEM(X=X,y=y,tol = 1e-6,maxIter =500,verbose=F)
fit_MM <- linRegMM(X=X,y=y,tol=1e-6,maxIter = 500,verbose=F)
fit_MoM <- linReg_MoM(X=X,y=y)

c(fit_PXEM$se2,fit_PXEM$sb2)

## [1] 0.9444521 0.1722404
c(fit_MM$se2,fit_MM$sb2)

## [1] 0.9443750 0.1723421
```

```
c(fit_MoM$se2,fit_MoM$sb2)
```

```
## [1] 0.9674995 0.1486726
```

References

[1] Jiang, J., Li, C., Paul, D., Yang, C., & Zhao, H. (2016). On high-dimensional misspecified mixed model analysis in genome-wide association study. *The Annals of Statistics*, 44(5), 2127-2160.

[2] Liu, C., Rubin, D. B., & Wu, Y. N. (1998). Parameter expansion to accelerate EM: the PX-EM algorithm. *Biometrika*, 85(4), 755-770.

[3] Foulley, J. L., & Van Dyk, D. A. (2000). The PX-EM algorithm for fast stable fitting of Henderson's mixed model. *Genetics Selection Evolution*, 32(2), 143.

[4] Zhou, H., Hu, L., Zhou, J., & Lange, K. (2018). MM algorithms for variance components models. *Journal of Computational and Graphical Statistics*, (just-accepted), 1-30.

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