PX-EM Algorithm with Three Variance Components

Cai Mingxuan

November 10, 2021

1 Model setting

Consider a three variance components model:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\omega} + \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^n$ is the vector of response, $\mathbf{Z} \in \mathbb{R}^{n\times q}$, $\mathbf{X}_1 \in \mathbb{R}^{n\times p_1}$ and $\mathbf{X}_2 \in \mathbb{R}^{n\times p_2}$ are design matrices, $\boldsymbol{\omega} \in \mathbb{R}^q$ is the vector of fixed effects, $\boldsymbol{\beta}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\beta}_2 \in \mathbb{R}^{p_2}$ are two vectors of random effects and $\mathbf{e} \in \mathbb{R}^n$ is the vector of independent noise. We assume that

$$\boldsymbol{\beta}_1 \sim \mathcal{N}(\mathbf{0}, \sigma_1^2\mathbf{I}_{p_1}), \boldsymbol{\beta}_2 \sim \mathcal{N}(\mathbf{0}, \sigma_2^2\mathbf{I}_{p_2}), \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2\mathbf{I}_n). \tag{2}$$

We will derive a parameter-expanded EM algorithm for this problem.

2 Parameter-expanded EM algorithm

For convenience, denote $\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix}$ and $\boldsymbol{\beta} = \begin{bmatrix}\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2\end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \begin{bmatrix}\sigma_1^2\mathbf{I}_{p_1} & 0 \\ 0 & \sigma_2^2\mathbf{I}_{p_2}\end{bmatrix}$. The complete data log likelihood with the expanded parameter δ is given as

$$\begin{aligned}\mathcal{L} &= -\frac{n}{2}\log(2\pi\sigma_e^2) - \frac{1}{2\sigma_e^2}||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega} - \delta\mathbf{X}\boldsymbol{\beta}||^2 - \frac{p_1}{2}\log(2\pi\sigma_1^2) - \frac{p_2}{2}\log(2\pi\sigma_2^2) - \frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta} \\ &= \boldsymbol{\beta}^T(-\frac{\delta^2}{2\sigma_e^2}\mathbf{X}^T\mathbf{X} - \frac{1}{2}\boldsymbol{\Sigma}^{-1})^{-1}\boldsymbol{\beta} + \frac{\delta}{\sigma_e^2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T\mathbf{X}\boldsymbol{\beta} + \text{constant.}\end{aligned} \tag{3}$$

We can obtain the posterior distribution of $\boldsymbol{\beta}$ from the quadratic form (3):

$$\begin{aligned}&\boldsymbol{\beta}|\mathbf{y}, \mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \tilde{\boldsymbol{\Sigma}}) \\ &\tilde{\boldsymbol{\Sigma}}^{-1} = \frac{\delta^2}{\sigma_e^2}\mathbf{X}^T\mathbf{X} + \boldsymbol{\Sigma}^{-1}, \boldsymbol{\mu} = \frac{\delta}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}\mathbf{X}^T(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega}).\end{aligned} \tag{4}$$

Taking the expectation of $\mathcal{L}$ with respect to the posterior distribution of $\boldsymbol{\beta}$ in (4) involves the following terms:

$$\begin{aligned}\mathbb{E}(||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega} - \mathbf{X}\boldsymbol{\beta}||^2) &= \mathbb{E}(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}} - 2\delta\tilde{\mathbf{y}}^T\mathbf{X}\boldsymbol{\beta} + \delta^2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}) \\ &= \tilde{\mathbf{y}}^T\tilde{\mathbf{y}} - 2\delta\tilde{\mathbf{y}}^T\mathbf{X}\boldsymbol{\mu} + \delta^2\boldsymbol{\mu}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\mu} + \delta^2 tr(\mathbf{X}^T\mathbf{X}\tilde{\boldsymbol{\Sigma}}),\end{aligned} \tag{5}$$

where $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{Z}\boldsymbol{\omega}$ and

$$\begin{aligned}\mathbb{E}(\boldsymbol{\beta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}) &= \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + tr(\tilde{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) \\ &= \frac{1}{\sigma_1^2}\boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 + \frac{1}{\sigma_1^2}tr(\tilde{\boldsymbol{\Sigma}}_1) + \frac{1}{\sigma_2^2}\boldsymbol{\mu}_2^T\boldsymbol{\mu}_2 + \frac{1}{\sigma_2^2}tr(\tilde{\boldsymbol{\Sigma}}_2),\end{aligned} \tag{6}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the sub-vectors of $\boldsymbol{\mu}$ corresponding to $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively, and $\tilde{\boldsymbol{\Sigma}}_1$ and $\tilde{\boldsymbol{\Sigma}}_2$ are sub-matrices corresponding to $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively. With (5) and (6), we can obtain the Q function as the expectation of (3):

$$\begin{aligned}\mathcal{Q} =& -\frac{n}{2}\log(2\pi\sigma_e^2) - \frac{p_1}{2}\log(2\pi\sigma_1^2) - \frac{p_2}{2}\log(2\pi\sigma_2^2) \\ &- \frac{1}{2\sigma_e^2}||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega} - \delta\mathbf{X}\boldsymbol{\mu}||^2 - \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - tr((\frac{\delta^2}{2\sigma_e^2}\mathbf{X}^T\mathbf{X} + \frac{1}{2}\boldsymbol{\Sigma}^{-1})\tilde{\boldsymbol{\Sigma}})\end{aligned} \tag{7}$$

2.1 E-step

We compute $\boldsymbol{\mu}$ and $\tilde{\boldsymbol{\Sigma}}$ in (4) by setting $\delta = 1$.

2.2 M-step

With $\boldsymbol{\mu}$ and $\tilde{\boldsymbol{\Sigma}}$ obtained in the E-step, we estimate the model parameters $\{\delta, \omega, \sigma_e^2, \sigma_1^2, \sigma_2^2\}$ by taking derivative of the Q function (7) with respect to the model parameters and setting them as zero.

$$\begin{aligned}\delta =& \frac{(\mathbf{y} - \mathbf{Z}\boldsymbol{\omega})^T\mathbf{X}\boldsymbol{\mu}}{\boldsymbol{\mu}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\mu} + tr(\mathbf{X}^T\mathbf{X}\tilde{\boldsymbol{\Sigma}})} \\ \boldsymbol{\omega} =& (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T(\mathbf{y} - \delta\mathbf{X}\boldsymbol{\mu}) \\ \sigma_e^2 =& \frac{1}{n}||\mathbf{y} - \mathbf{Z}\boldsymbol{\omega} - \delta\mathbf{X}\boldsymbol{\mu}||^2 + \delta^2 tr(\mathbf{X}^T\mathbf{X}\tilde{\boldsymbol{\Sigma}}) \\ \sigma_1^2 =& \frac{1}{p_1}\boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 + tr(\tilde{\boldsymbol{\Sigma}}_1) \\ \sigma_2^2 =& \frac{1}{p_2}\boldsymbol{\mu}_2^T\boldsymbol{\mu}_2 + tr(\tilde{\boldsymbol{\Sigma}}_2)\end{aligned} \tag{8}$$