An EM algorithm for a linear mixed model

Consider the growth data from Jennrich and Schluchter (1986). Let \( i = 1, \ldots, 27 \) index subject and \( j = 1, \ldots, 4 \) index repeated measure within subject. We will ignore gender and consider a random intercept and slope model:

\[
y_{ij} = \beta_0 + \beta_1 age_{ij} + b_{0i} + b_{1i} age_{ij} + e_{ij},
\]

\( b_{0i} \overset{i.i.d.}{\sim} \text{MVN}(0, G) \) and \( e_{ij} \overset{i.i.d.}{\sim} \text{N}(0, \sigma_e^2) \).

The model we will consider for each subject is

\[
\begin{pmatrix}
y_{i1} \\
y_{i2} \\
y_{i3} \\
y_{i4}
\end{pmatrix} =
\begin{pmatrix}
1 & 8 \\
1 & 10 \\
1 & 12 \\
1 & 14
\end{pmatrix}
\begin{pmatrix}
\beta_0 \\ \\
\beta_1
\end{pmatrix}
+ 
\begin{pmatrix}
b_{0i} \\
b_{1i}
\end{pmatrix}
+ 
\begin{pmatrix}
e_{i1} \\
e_{i2} \\
e_{i3} \\
e_{i4}
\end{pmatrix},
\]
concisely,

\[
y_i = X_i \beta + Z_i b_i + e_i.
\]

Please work through the questions below and then apply your results in the code \texttt{lmm.em.for.Mon.R}. You will need to hand in your finished version of \texttt{lmm.em.for.Mon.R} and answers to questions on this sheet. Depending on your progress, we may take more than one class to do this. You are welcome to work in groups.

1 Specify \( X, \beta, Z, b, \) and \( e \) to write the model for \( y = (y_{i1}^T, \ldots, y_{i27}^T)^T \) in canonical form (\( y = X \beta + Zb + e \)).

2 What are the distributions for \( b \) and \( e \)? Let \( \text{cov}(u) = \tilde{G} \). How does \( G \) relate to \( \tilde{G} \)?

3 If we had estimates of \( \tilde{G} \) and and \( \sigma_e^2 \), what would the MLE of \( \beta \) be?

4 EM Algorithm for ML estimates of \( G \) and \( \sigma_e^2 \):

   The likelihood of \( y \) is \( f(y; \beta, G, \sigma_e^2) = \text{MVN}(X \beta, Z \tilde{G} Z^T + \sigma_e^2 I_{108}) \) pdf. The EM algorithm requires that we define a “latent” \( u \) so that \( f(y; \beta, G, \sigma_e^2) = \int_u f(y, u) du \).

   In the following, we will first let \( b = u \) to estimate \( G \). Then we will let \( e = u \) to estimate \( \sigma_e^2 \).

   (a) Estimating \( G \):

      i. Show that

      \[
      f(y, b) = \begin{pmatrix} y \\ b \end{pmatrix} \sim \text{MVN}
      \begin{pmatrix}
      X \beta \\
      0
      \end{pmatrix},
      \begin{pmatrix}
      Z \tilde{G} Z^T + \sigma_e^2 I_{108} \\
      \tilde{G}
      \end{pmatrix}
      \]

      Note that the requirement \( \int_b f(y, b) db = \text{MVN}(X \beta, Z \tilde{G} Z^T + \sigma_e^2 I_{108}) \) is satisfied by construction.
ii. Find the conditional distributions of \( b | y \) and \( b_i | y \).  

iii. Let \( C \) be a constant that does not involve \( G \). Show that 
\[
E_{b|y} \{ \log(f(y, b) | y) \} = C - \frac{27}{2} \log(|G|) - \frac{1}{2} \sum_{i=1}^{27} \text{tr} \left\{ G^{-1} E_{b_i|y} (b_i b_i^T | y) \right\}.
\]

Hint: \( f(y, b) = f(y|b) f(b) \) and the first term does not involve \( G \).

iv. Show that (4aiii) is maximized by
\[
\hat{G} = \frac{1}{27} \sum_{i=1}^{27} \left\{ E_{b_i|y} (b_i | y) E_{b_i|y} (b_i | y)^T + \text{cov}_{b_i|y} (b_i | y) \right\}.
\]

(b) Estimating \( \sigma_e^2 \):

i. Show that
\[
f(y, e) = \left( \begin{array}{c} y \\ e \end{array} \right) \sim \text{MVN} \left( \begin{array}{c} X \beta \\ Z \hat{G} Z^T + \sigma_e^2 I_{108} \\ \sigma_e^2 I \end{array} \right).
\]

Note that the requirement \( \int e f(y, e) de = \text{MVN}(X \beta, Z \hat{G} Z^T + \sigma_e^2 I_{108}) \) is satisfied by construction.

ii. Find the conditional distribution of \( e | y \).

iii. Let \( C \) be a constant that does not involve \( \sigma_e^2 \). Show that 
\[
E_{e|y} \{ \log(f(y, e) | y) \} = C - \frac{108}{2} \log(\sigma_e^2) - \frac{1}{2\sigma_e^2} E_{e|y} (e^T e | y).
\]

Hint: \( f(y, e) = f(y|e) f(e) \) and the first term does not involve \( \sigma_e^2 \).

iv. Show that (4biii) is maximized by
\[
\hat{\sigma}_e^2 = \frac{1}{108} \left[ E_{e|y} (e^T e | y) E_{e|y} (e | y) + \text{tr} \left\{ \text{cov}_{e|y} (e | y) \right\} \right].
\]

5 EM Algorithm for REML estimates of \( G \) and \( \sigma_e^2 \):

REML finds estimates of \( G \) and \( \sigma_e^2 \) to maximize the likelihood of \( Ky \) where 
\( KnX = 0 \) and \( K \) has rank \( \text{length}(y) - \text{rank}(X) = n - p \). One choice is to use the first \( n - p \) rows of \( I - X(X^T X)^{-1} X^T \). As a result, this only changes steps (i) and (ii) in the methods above, and \( Ky \) replaces \( y \) in all estimates of the variance components ( \( G \) and \( \sigma_e^2 \)). What are the new versions of (i) and (ii)?