Linear Mixed Models

Phi Le

2/28/2021

In this project, we will write a program to find the coefficients of the Linear Mixed Models for repeated measure. In this note, we present the details how we obtain the coefficients from Maximum Likelihood method for latent data.

$$
\begin{aligned}
Y_{ij} &= X_{ij}\beta + Z_{ij}b_i + \epsilon_{ij}, \quad \text{for} j = 1, ..., k_i \\
b_i &\sim N(0, D) \\
\epsilon_i &\sim N(0, \sigma^2 I_{k_i}) \text{ where} I_{k_i} = (1, ..., 1), k_i \text{times} \\
\epsilon_{ij} & \quad \text{are iids and } \ , b_i \perp \epsilon_i
\end{aligned}
$$

Where the Y_{ij} is the observation of individual i at the time j, $X_{ij}\beta$ is the fixed effect and $Z_{ij}b_i$ is the random effect for individual i with unknown unstructured covariance D and ϵ is the unknown error measurement for each individual. Notice that if we write the formula in matrix form then

$$
\begin{aligned}
Y &= X\beta + Zb + \epsilon \\
b &\sim N(0, \mathbf{D}) \\
\epsilon &\sim N(0, \sigma^2 Id)
\end{aligned}
$$

EM algorithm for linear mixed effect models

In this section, we will give details how to get EM algorithm for the model above. We will compute the Q function for E step and then perform M step. Before we work on the details, let me state some facts about linear algebra results which we will need later

Trace properties and some facts: For any matrix A, B, C such that ABC has meaning and a is a number

1. $Trace(ABC) = Trace(BCA)$
2. $a = Trace(a)$
3. $Trace(aA) = aTrace(A)$
4. $A^{-1} = \frac{Adj(A)}{det(A)}$, where $Adj(A)$ is the adjugate of matrix A
5. $Adj(\sigma^2 Id) = \sigma^{2(m-1)} Id$ where Id is a identity matrix with $dim(Id) = m \times m$
6. Derivative of determinant function: $d(det(A)) = \frac{1}{det(A)} Trace(Adj(A)d(A))$, where dA is the differentiation of A
7. Derivative of inverse matrix: $d(A^{-1}) = A^{-1}d(A)A^{-1}$

where A^T is the transpose of matrix A.

Multivariable completing the square: For any vector $y = (y_1, y_2, \cdots, y_k)$

1. $\|y\|_2^2 = yy^T$
2. $\|ay + bx\|_2^2 = a^2\|y\|_2^2 + (ay)^T bx + (bx)^T ay + b^2\|x\|_2^2$, for any vector y, x and scalars a, b.
3. $b^2 xx^T + abxy^T + abx^T y = (bx + ay)(bx + ay)^T + C(y)$, $C(y)$ is a function of y but it is independent on x
4. For any positive definite matrix D, we can write $D = LL^T$ where L is a lower triangle matrix

Loglikelihood of complete data

First, we compute the likelihood function of complete data. Suppose the parameters are $\theta = (\beta, b, D, \sigma)$ and the observed data is y with latent data b.

$$\begin{aligned} f_{complete}(y, b|D, \sigma; \beta) &= f(y|b; \beta, D, \sigma) f(b|D) \\ &= \prod_{i=1}^{N} \frac{1}{(2\pi\sigma)^{k_i/2}} \exp\left\{-\frac{(y_i - X_{ij}\beta - Z_{ij}b_i)(y_i - X_{ij}\beta - Z_{ij}b_i)^T}{2\sigma^2}\right\} \\ &\times \frac{1}{(2\pi)^{k_i/2}\det(D)^{1/2}} \exp\left\{\frac{-b_i D^{-1} b_i^T}{2}\right\} \end{aligned}$$

So, the log likelihood function is

$$log(L) = \sum_{i=1}^{N} -\frac{1}{2}\log(det(\sigma^2 Id)) - \frac{1}{2}\log(det(D)) - \frac{(y_i - X_i\beta - Z_i b_i)Id(y_i - X_i\beta - Z_i b_i)^T}{2\sigma^2} - \frac{b_i D^{-1} b_i^T}{2}$$

The density function of missing (latent) data under condition of observed y

$$\begin{aligned} f_{miss}(b|y; \beta, D, \sigma) &\propto f(y|b) f(b|D) \\ &= \prod_{i=1}^{N} \frac{1}{(2\pi)^{k_i/2} det(\sigma^2 Id_i)^{1/2}} \exp\left\{-\frac{(y_i - X_{ij}\beta - Zb)(y_i - X_{ij}\beta - Z_{ij}b_i)^T}{2\sigma^2}\right\} \\ &\times \frac{1}{(2\pi)^{k_i/2}\det(D)^{1/2}} \exp\left\{-\frac{b_i D^{-1} b_i^T}{2}\right\} \\ &= \prod_{i=1}^{N} \frac{1}{(2\pi)^{k_i} det(\sigma^2 Id_i)^{1/2}} \frac{1}{\det(D)^{1/2}} \exp\left\{-\frac{(y_i - X_{ij}\beta - Z_{ij}b_i)(y_i - X_{ij}\beta - Z_{ij}b_i)^T + \sigma^2 b_i D^{-1} b_i^T}{2\sigma^2}\right\} \end{aligned}$$

We need to do complete the square the numerator of the exp expression to get the density function of the normal distribution in term of b_i variable

$$\begin{aligned} & (y_i - X_{ij}\beta - Z_{ij}b_i)(y_i - X_{ij}\beta - Z_{ij}b_i)^T + \sigma^2 b_i D^{-1} b_i^T \\ = & (Z_{ij}b_i)^T (Z_{ij}b_i) + b_i^T (\sigma^2 D^{-1}) b_i + (Z_{ij}b_i)^T (y_i - X_{ij}\beta) + (y_i - X_{ij}\beta)^T (Z_{ij}b_i) + (y_i - X_{ij}\beta)^T (y_i - X_{ij}\beta) \\ = & b_i^T (Z_{ij}^T Z_{ij} + \sigma^2 D^{-1}) b_i + b_i^T Z_{ij}^T (y_i - X_{ij}\beta) + (y_i - X_{ij}\beta)^T Z_{ij} b_i + (y_i - X_{ij}\beta)^T (y_i - X_{ij}\beta) \\ = & [b_i - \Sigma Z_i^T (y_i - X_{ij}\beta)]^T \Sigma^{-1} [b_i - \Sigma Z_i^T (y_i - X_{ij}\beta)] + C(y_i, Z, D) \end{aligned}$$

where $\Sigma_i = (Z_{ij}^T Z_{ij} + \sigma^2 D^{-1})^{-1}$

Now, back to the density function of the missing information. By replacing the expression in the parentheses by the above expression, then

$$\begin{aligned}
f_{miss}(b|y;\beta,D,\sigma) &\propto f(y|b)f(b|D) \\
&\propto \prod_{i=1}^{N} \frac{1}{\sigma^{k_i/2}} \frac{1}{\det(D)^{1/2}} \exp\left\{-\frac{[b_i - \Omega_i\sigma^{-2}Z_i^T(y_i - X_{ij}\beta)]^T\Omega_i^{-1}[b_i - \Omega_i\sigma^{-2}Z_i^T(y_i - X_{ij}\beta)]}{2}\right\}
\end{aligned}$$

where $\Omega_i = \left(\frac{\Sigma^{-1}}{\sigma^2}\right)^{-1} = \left(\frac{Z_i^T Z_i}{\sigma^2} + D^{-1}\right)^{-1}$

Compute the Q function.

To compute the Q function, we need to compute the $\log(L)$ over the $f_{miss}(b|y;\beta^{(s)}, D^{(s)}, \sigma^{(s)})$. It is enough to compute the integral over the terms having b_i variable. Before we go into the details, by the first fact of Trace function above:

$$b_i D^{(-1)} b_i^T = Trace(b_i D^{(-1)} b_i^T) = Trace(D^{-1} b_i^T b_i)$$

and we also have

$$\begin{aligned}
& (y_i - X_{ij}\beta - Z_{ij}b_i)^T(y_i - X_{ij}\beta - Z_{ij}b_i) \\
= \;& y_i^T y_i - y_i^T X_i\beta - y_i^T Z_i b_i - (X_i\beta)^T y_i + (X_i\beta)^T(X_i\beta) + (X_i\beta)^T Z_i b_i - (Z_i b_i)^T y_i + (Z_i b_i)^T X_i\beta + (Z_i b_i)^T(Z_i b_i) \\
= \;& y_i^T y_i - y_i^T X_i\beta - y_i^T Z_i b_i - (X_i\beta)^T y_i + 2(X_i\beta)^T(X_i\beta) \\
& +(X_i\beta)^T Z_i b_i - (y_i^T(Z_i b_i))^T + ((X_i\beta)^T Z_i b_i)^T + Trace(Z_i^T Z_i b_i b_i^T))
\end{aligned}$$

Thus, to compute the Q function, we only need to find the expectation of b_i and $b_i b_i^T$ which is the mean and variance of the multivariate normal distribution. So,

$$\begin{aligned}
Q(\theta, \theta^{(s)}) &= \sum_{i=1}^{N}\left(-\frac{1}{2}\log(det(\sigma Id)) - \frac{1}{2}\log(\det(D))\right) + \\
&+ \sum_{i=1}^{N}\frac{1}{2\sigma^2}\left(-Trace(Z_i^T Z_i E(b_i b_i^T|y;\theta^{(s)})) - y_i^T y_i + y_i^T X_i\beta\right) \\
&\quad + \sum_{i=1}^{N}\frac{1}{2\sigma^2}\left(y_i^T Z_i E(b_i|y;\theta^{(s)}) + (X_i\beta)^T y_i - 2(X_i\beta)^T(X_i\beta) - (X_i\beta)^T Z_i E(b_i|y;\theta)\right) \\
&\quad + \sum_{i=1}^{N}\frac{1}{2\sigma^2}\left((y_i^T Z_i E(b_i|y;\theta))^T - ((X_i\beta)Z_i E(b_i|y;\theta))^T\right) \\
&\quad - \frac{1}{2}Trace(D^{-1}E(b_i b_i^T|y;\theta^{(s)}))
\end{aligned}$$

where

$$\begin{aligned}
E(b_i|y;\theta^{(s)}) &= \frac{1}{\sigma^2}\Omega_i Z_i^T(y_i - X_{ij}\beta^{(s)}) =: \mu_i^{(s)} \\
E(b_i b_i^T|y;\theta^{(s)}) &= Cov(b_i, b_i^T) + E(b_i|y;\theta^{(s)})E(b_i|y;\theta^{(s)})^T = \Omega_i^{(s)} + \mu^{(s)}(\mu^{(s)})^T =: \mu_{ii}^{(s)}
\end{aligned}$$

Hence, the parameters β, D, σ^2 can be computed as followings:

- For D covariance: The derivative of the Q function with respect to D matrix. Take the derivative of $\log(det(D))$ and derivative of $Trace(D^{-1}E(b_i^T b_i|y;\theta^{(s)}))$ by using the facts in section Trace properties and facts of matrix, then we obtain $D^{(s+1)}$ is the solution of the following

$$\begin{aligned}
\frac{Trace(Adj(D)dD)}{det(D)} - Trace(D^{-1}(dD)D^{-1}E(b_i^T b_i|y;\theta^{(s)})) &= 0 \\
\frac{Trace(Adj(D)dD)}{det(D)} - Trace\left(\frac{Adj(D)(dD)Adj(D)}{(det(D))^2}E(b_i^T b_i|y;\theta^{(s)})\right) &= 0 \\
\frac{1}{det(D)}Trace\left[(Adj(D)dD)\left(\frac{Adj(D)}{det(D)}E(b_i^T b_i|y;\theta^{(s)}) - ID\right)\right] &= 0 \\
\frac{1}{det(D)}Trace\left[(Adj(D)dD)\left(D^{-1}E(b_i^T b_i|y;\theta^{(s)}) - ID\right)\right] &= 0, \quad \text{Since} D^{-1} = \frac{Adj(D)}{det(D)}
\end{aligned}$$

We can see that $D = E(b_i^T b_i|y;\theta^{(s)})$ is a solution of the equation. By using the result of $E(b_i^T b_i|y;\theta^{(s)})$ we found above, then

$$D^{(s+1)} = \frac{1}{N}\sum_{i=1}^{N}\Omega(y_i - X_{ij}\beta^{(s)}) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Z_i^T Z_i}{\sigma^2} + (D^{(s)})^{-1}\right)^{-1}(y_i - X_{ij}\beta^{(s)})$$

- For β: The derivative of the Q function with respect to β is

$$\sum_{i=1}^{N}(X_i)^T(y_i - X_i\beta - Z_i E(b_i|y;\theta^{(s)}) = 0$$

the solution is

$$X^T X\beta = X^T y - X^T Z\Omega(y - X\beta^{(s)})$$

or

$$\beta^{(s+1)} = (X^T X)^{-1}X^T\left(y - Z\mu)\right)$$

- For σ: Taking derivative of Q function with respect to matrix σI_k by using the same method that we applied for finding D and $d(\sigma^2 ID) = (2\sigma ID)$, we get $\sigma^{(s+1)}$ is the solution of the following equation

$$-\frac{\sum_{i=1}^{N}k_i}{\sigma} + \sum_{i=1}^{N}\frac{2}{\sigma^3}\left(\|y_i - X_i\beta\|^2 + Trace(Z_i^T Z_i E(b_i^T b_i|y;\theta^{(s)})) - 2(y_i - X_i\beta)Z_i E(b_i|y;\theta^{(s)})\right) = 0$$

or we have

$$(\sigma^{(s+1)})^2 = \frac{1}{\sum_{i=1}^{N}k_i}\left[\|y - X\beta\|^2 + \sum_{i=1}^{N}\left\{Trace(Z_i^T Z_i\Omega_i^{(s)} + \mu_i^{(s)}(\mu_i^{(s)})^T) - 2(y_i - X_i\beta)Z_i\Omega_i^{(s)}(y_i - X_i\beta^{(s)})\right\}\right]$$