General Linear Model for Longitudinal Data

MLE estimation for β, Σ:

The likelihood function is:

$$L = \prod_{i=1}^{N} (2\pi)^{-n/2} |\Sigma|^{-1/2} e^{-1/2(Y_i - X_i\beta)^T \Sigma^{-1} (Y_i - X_i\beta)}$$

$$l = \log L = -\frac{Nn}{2}\log 2\pi - \frac{N}{2}\log|\Sigma| - 1/2 \sum_{i=1}^{N} (Y_i - X_i\beta)^T \Sigma^{-1} (Y_i - X_i\beta)$$

In order to obtain MLE, we can take the derivative with respective to these two parameters and set them to be zero:

$$\frac{\partial l}{\partial \beta} = -1/2 \sum_{i=1}^{N} (2\beta^T X_i^T \Sigma^{-1} X_i - 2X_i^T \Sigma^{-1} Y_i) = 0$$

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i^T \Sigma^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i^T \Sigma^{-1} Y_i\right)$$

$$\frac{\partial l}{\partial \Sigma^{-1}} = \frac{N}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{N} (Y_i - X_i\beta)(Y_i - X_i\beta)^T = 0$$

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} (Y_i - X_i\beta)(Y_i - X_i\beta)^T$$

MLE Algorithm:

Start with some initial values of Σ (for example, $\Sigma = \mathrm{I}$)

For t in 1:T

Estimate $\hat{\beta}_{(t)}$ based on $\hat{\Sigma}_{(t-1)}$

Estimate $\hat{\Sigma}_{(t)}$ based on $\hat{\beta}_{(t)}$

Until convergence

Properties of MLE:

Under regulatory conditions, $\hat{\beta}_{MLE}$ and $\hat{\Sigma}_{MLE}$ are consistent, and

$$\sqrt{N}\left(\hat{\beta}_{MLE} - \beta\right) \to N(0, N\Gamma)$$

Where $\Gamma = -\mathrm{E}\left(\frac{\partial^2 \log L(\beta,\Sigma)}{\partial \beta^2}\right)^{-1}$. There, the asymptotic variance of $\sqrt{N}\left(\hat{\beta}_{MLE} - \beta\right)$ is $\mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{N} X_i^T {\hat{\Sigma}_{MLE}}^{-1} X_i\right)^{-1}$, which is consistently estimated by

$$\hat{V}\left(\sqrt{N}\left(\hat{\beta}_{MLE} - \beta\right)\right) = \left(\frac{1}{N}\sum_{i=1}^{N} X_i^T {\hat{\Sigma}_{MLE}}^{-1} X_i\right)^{-1}$$

So that $\hat{V}\left(\hat{\beta}_{MLE}\right) = \left(\sum_{i=1}^{N} X_i^T {\hat{\Sigma}_{MLE}}^{-1} X_i\right)^{-1}$

REML estimation:

$$L_r = \int_\beta L \, d\beta$$

$$= \int_\beta \prod_{i=1}^{N} (2\pi)^{-n/2} |\Sigma|^{-1/2} e^{-1/2(Y_i - X_i\beta)^T \Sigma^{-1} (Y_i - X_i\beta)} \, d\beta$$

$$= (2\pi)^{-nN/2} |\Sigma|^{-N/2} \int_\beta e^{-1/2 \sum_{i=1} (Y_i - X_i\beta)^T \Sigma^{-1} (Y_i - X_i\beta)} d\beta$$

$$= (2\pi)^{-nN/2} |\Sigma|^{-N/2} \int_\beta e^{-1/2 \sum_{i=1} (Y_i^T \Sigma^{-1} Y_i - 2\beta^T X_i^T \Sigma^{-1} Y_i + \beta^T X_i^T \Sigma^{-1} X_i \beta)} d\beta$$

$$= C \int_\beta e^{-1/2(\beta^T (\sum_{i=1} X_i^T \Sigma^{-1} X_i)\beta - 2\beta^T (\sum_{i=1} X_i^T \Sigma^{-1} Y_i) + \sum_i Y_i^T \Sigma^{-1} Y_i)} d\beta$$

$$= C \int_\beta e^{-1/2((\beta - \hat{\beta})^T (\sum_{i=1} X_i^T \Sigma^{-1} X_i)(\beta - \hat{\beta}) + \sum_i Y_i^T \Sigma^{-1} Y_i - \hat{\beta}^T (\sum_{i=1} X_i^T \Sigma^{-1} X_i)\hat{\beta})} d\beta$$

$$= C (2\pi)^{p/2} \left| \left(\sum_{i=1} X_i^T \Sigma^{-1} X_i \right) \right|^{-1/2} e^{-1/2(\sum_i Y_i^T \Sigma^{-1} Y_i - \hat{\beta}^T (\sum_{i=1} X_i^T \Sigma^{-1} X_i)\hat{\beta})}$$

$$= (2\pi)^{-(nN-p)/2} |\Sigma|^{-N/2} \left| \left(\sum_{i=1} X_i^T \Sigma^{-1} X_i \right) \right|^{-1/2} e^{-1/2(\sum_i (Y_i - X_i\hat{\beta})^T \Sigma^{-1} (Y_i - X_i\hat{\beta}))}$$

So in some sense, $L_r \propto L * \left| V(\hat{\beta}) \right|^{1/2}$

$$\Sigma_{REML} = 1/N \sum_i (Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})^T + 1/N \sum_{i=1}^{N} X_i V(\hat{\beta}) X_i^T$$

Where $V(\hat{\beta}) = \left(\sum_{i=1}^{N} X_i^T \Sigma_{REML}^{-1} X_i \right)^{-1}$

For REML estimation, we need to iterate the estimation procedure, and is more complicated than the MLE estimation

In certain settings, $\widehat{\Sigma}_{REML}$ is unbiased for Σ (unlike $\widehat{\Sigma}_{MLE}$). This is not always the case. However, REML often has smaller bias and MSE than MLE estimates.

Finally, $\hat{\beta}_{REML} = \left(\sum_{i=1}^{N} X_i^T \widehat{\Sigma}_{\text{REML}}^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i^T \widehat{\Sigma}_{\text{REML}}^{-1} Y_i\right)$