EM Algorithm

LMEM: $y_j = X_j \beta + Z_j a_j + \epsilon_j \qquad j = 1 \dots m$

$$a_1 \dots a_m \sim \text{iid } N_q(0, \psi)$$
$$\epsilon_1 \dots \epsilon_m \sim \text{iid } N_n(0, \sigma^2 I)$$

Unknown parameters: $(\underbrace{\beta}_{\text{fixed}}, \underbrace{\psi, \sigma^2}_{\text{varcomp}})$

Likelihood:

$$p(y_1 \dots y_m \mid \beta, \psi, \sigma^2) = \prod_{j=1}^{m} p(y_j \mid \beta, \psi, \sigma^2)$$

$$p(y_j \mid \beta, \psi, \sigma^2) = \int p(y_j \mid \beta, \sigma^2, a_j)\, p(a_j \mid \psi)\, da_j$$

Example: $y_{ij} = \overset{\text{"}\mu\text{"}}{\beta} + a_j + \epsilon_{ij} \qquad a_1 \dots a_m \sim \text{iid } N(0, \overset{\text{"}\tau^2\text{"}}{\psi})$

$$y_j = 1\beta + I a_j + \epsilon_j$$

$$p(y_j \mid \beta, \psi^2, \sigma^2): \quad y_j \mid \beta, \psi^2, \sigma^2 \sim N\left(1\beta, \begin{bmatrix} \sigma^2 + \psi^2 & \psi^2 \\ & \ddots \end{bmatrix}\right)$$

$$p(y_j \mid \beta, \sigma^2, a_j): \quad y_j \mid \beta, \sigma^2, a_j \sim N(1(\beta + a_j), \sigma^2 I)$$

More generally

$$\underset{\sim}{y}_j \mid \beta, \Psi, \sigma^2 \sim N_{n_j}(X_j\beta,\ \overset{\text{hard}}{\swarrow}\ Z_j\Psi Z_j' + \sigma^2 I)$$

$$\underset{\sim}{y}_j \mid \beta, a_j, \sigma^2 \sim N_{n_j}(X_j\beta + Z_j \underset{\sim}{a}_j,\ \sigma^2 I)$$

$\nwarrow$ easy

$$\underset{\sim}{y}_j - Z_j \underset{\sim}{a}_j \mid \beta, a_j, \sigma^2 \sim N_{n_j}(X_j\beta,\ \sigma^2 I)$$

$\nwarrow$ very easy

<u>Idea</u>: Suppose $\underset{\sim}{a}_1 \ldots \underset{\sim}{a}_m$ were known.
How would we estimate β, σ^2?

$$\begin{pmatrix} y_1 - Z_1 a_1 \\ y_2 - Z_2 a_2 \\ \vdots \\ y_m - Z_m a_m \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} \beta + \sigma^2 \begin{pmatrix} \underset{\sim}{e}_1 \\ \underset{\sim}{e}_2 \\ \vdots \\ e_m \end{pmatrix}$$

$$\tilde{y} = X\beta + \sigma^2 e$$

This is just an ordinary linear reg. model

Could use $\hat{\beta} = (X'X)^{-1}X'\tilde{y}$, $\hat{\sigma}^2 = \frac{\|\tilde{y} - X\hat{\beta}\|^2}{N}$

Also, if $\underset{\sim}{a}_1 \ldots \underset{\sim}{a}_n$ were known, we could estimate Ψ:

$$\underset{\sim}{a}_1 \ldots \underset{\sim}{a}_n \sim \text{iid } N_q(0, \Psi)$$

$$\hat{\Psi}_{MLE} = \sum_{j=1}^{m} a_j a_j^T / m$$ (~ sample cov. matrix of $a_1 \ldots a_n$)

Complete Data Log Likelihood

Suppose $\underset{\sim}{a}_1 \ldots \underset{\sim}{a}_n$ were observed.
We could estimate β, σ^2, Ψ w complete:

$$p(\underset{\sim}{y}_1 \ldots \underset{\sim}{y}_m, \underset{\sim}{a}_1 \ldots \underset{\sim}{a}_m \mid \beta, \sigma^2, \Psi)$$

$$= \prod_{j=1}^{m} p(\underset{\sim}{y}_j, \underset{\sim}{a}_j \mid \beta, \sigma^2, \Psi)$$

$$p(\underset{\sim}{y}_j, \underset{\sim}{a}_j \mid \beta, \sigma^2, \Psi) = p(\underset{\sim}{y}_j \mid \underset{\sim}{a}_j, \beta, \sigma^2 \Psi)\, p(\underset{\sim}{a}_j \mid \ldots)$$

$$= p(\underset{\sim}{y}_j \mid \underset{\sim}{a}_j, \beta, \sigma^2)\, p(\underset{\sim}{a}_j \mid \Psi)$$

4

So the lik is

$$p(y_1 \ldots y_m\, a_1 \ldots a_m \mid \sigma^2, \Psi, \beta) = \prod p(y_j \mid a_j \ldots)\, p(a_j \mid \ldots)$$

$$= \left\{ \prod_{j=1}^{m} p(y_j \mid \beta, \sigma^2, a_j) \right\} \times \left\{ \prod_{j=1}^{m} p(a_j \mid \Psi) \right\}$$

$$p(\tilde{y} \mid \beta, \sigma^2, a_1 \ldots a_m) \times \qquad p(a_1 \ldots a_m \mid \Psi)$$

$$\tilde{y}_j = y_j - Z_j a_j$$

$$-2 \log p(y_1 \ldots y_m, a_1 \ldots a_m \mid \sigma^2, \Psi, \beta)$$

$$= \underbrace{nm \log \sigma^2 + \frac{(\tilde{y} - X\beta)^T (\tilde{y} - X\beta)}{\sigma^2}}_{\text{dep only on } \beta,\, \sigma^2} + \underbrace{m \log |\Psi| + \operatorname{tr}(\Psi^{-1} A^T A)}_{\text{only dep on } \Psi}$$

opt is $(\hat{\beta}, \hat{\sigma}^2) = \left((X^T X)^{-1} X \tilde{y},\ \frac{\|\tilde{y} - X\hat{\beta}\|^2}{nm} \right)$

opt is $\hat{\Psi} = A^T A / m$ $= \operatorname{cov}(a_1 \ldots a_m)$

Problem: We don't know $\underset{\sim}{a}_1 \ldots \underset{\sim}{a}_m$.

Just know $\underset{\sim}{a}_1 \ldots \underset{\sim}{a}_m \sim_{\text{IID}} N(0, \Psi)$

Idea: replace CDLL with its expectation over $a_1 \ldots a_m$.

EM algorithm: given starting values $(\beta_0, \sigma_0^2, \Psi_0)$ iterate the following

① E-step

Compute Expectation of CDLL given $\underset{\sim}{y}_1 \ldots \underset{\sim}{y}_m$, $\beta_{(0)}, \sigma_{(0)}^2, \Psi_{(0)}$

$$f(\beta, \sigma^2, \Psi) = E\left[n m \log \sigma^2 + \frac{(\tilde{y} - X\beta)^T (y - X\beta)}{\sigma^2} + m \log |\Psi| + \operatorname{tr}(\Psi^{-1} A^T A) \,\middle|\, \beta_{(0)}, \sigma_{(0)}^2, \Psi_{(0)}, y_1 \ldots y_m \right]$$

② M step

optimize f in β, σ^2, Ψ : $(\beta_{(1)}, \sigma_{(1)}^2, \Psi_{(1)}) = \underset{\beta, \sigma^2, \Psi}{\arg\min} f(\beta, \sigma^2, \Psi)$