1. Repeated-measures data

A variance component model:

\[ y_i = X_i \beta + Z_i b_i + e_i, \quad i = 1, 2, \ldots, m; \]

where \( y_i \) is a vector with length \( n_i \) and denotes the \( n_i \) measurements observed on subject \( i \); \( X_i \) and \( Z_i \) are \( n_i \times p \) and \( n_i \times q \) design matrices, \( \beta \) is the fixed effect, \( b_i \sim N(0, D) \) is the random effect, \( e_i \sim N(0, \sigma^2 I_{n_i}) \) are the measurement errors.

An obvious choice for the complete-data is \( \{y_i, b_i, e_i, i = 1, \ldots, m \} \). and the complete data sufficient statistics are

\[
\sum_{i=1}^{m} b_i b_i^T, \quad \sum_{i=1}^{m} e_i^T e_i.
\]

Let \( V_i = Z_i D Z_i^T + \sigma^2 I_{n_i} \), suppose the current estimate is \( \theta^{(t)} = (\mu^{(t)}, D^{(t)}, \sigma^{(t)}) \). In the E-step,

\[
E(b_i \mid y_i, \theta^{(t)}) = b_i^{(t)} = (Z_i^T Z_i + \sigma^{(t)} D^{(k)} - 1)^{-1} Z_i^T (y_i - X_i \beta^{(t)}),
\]

\[
E(b_i b_i^T \mid y_i, \theta^{(t)}) = b_i^{(t)} b_i^{(t)T} + \text{var}(b_i \mid y_i, \theta^{(t)})
\]

\[
= b_i^{(t)} b_i^{(t)T} + \sigma^{(t)} \left( \sigma^{(t)} - 2 Z_i^T Z_i + D^{(k)} \right)^{-1}.
\]

\[
E(e_i^T e_i \mid y_i, \theta^{(t)}) = E\{(y_i - X_i \beta^{(t)} - Z_i b_i)^T (y_i - X_i \beta^{(t)} - Z_i b_i) \mid y_i, \theta^{(t)}\}
\]
CM-step 1

\[ D^{(k+1)} = \frac{1}{m} \sum_{i=1}^{m} E(b_ib_i^T \mid y_i, \theta^{(t)}) \]
\[ \sigma^{(t+1)^2} = \frac{1}{\sum_{i=1}^{m} n_i} \sum_{i=1}^{m} E(e_i^T e_i \mid y_i, \theta^{(t)}) \]

CM-step 2

\[ \beta^{(k+1)} = \left( \sum_{i=1}^{m} X_i^T V_i^{(t+1)^{-1}} X_i \right)^{-1} \left( \sum_{i=1}^{m} X_i V_i^{(t+1)^{-1}} y_i \right) \]
2. Nonignorable missing-data models

In general, there are three models in terms of different factorization of the full likelihood function of hypothetic complete data $Y$ and the missing-data mechanism $M$.

1. Selection models

$$f(M_i, y_i; \theta, \psi) = f(y_i; \theta) f(M_i | y_i; \psi).$$

2. Pattern-mixture models

$$f(M_i, y_i; \theta, \psi) = f(M_i; \psi) f(y_i | M_i; \theta)$$

3. Pattern-set mixture models

$$f(M_i^{(1)}, M_i^{(2)}, y_i; \theta, \psi, \phi) = f(M_i^{(1)}; \psi) f(y_i | M_i^{(1)}, \theta) f(M_i^{(2)} | M_i^{(1)}, y_i; \phi).$$

Remarks

(1) Selection models are more popular, have better interpretation but the missing-data mechanisms usually have to be specified.

(2) Parameters of pattern-mixture models are not fully identified and require prior information or parameter restrictions.

(3) Pattern-set mixture models are extension of the pattern-mixture models and its applications are quite limited so far.

(4) The likelihood inference is based on $f(Y_{obs}, M)$. Calculation can be done by EM algorithm.

Biostat 2065: Analysis of Incomplete Data, 2005
3. Examples

Example 15.3. Grouped exponential samples.
The hypothetic complete data are a random sample from $\text{Exp}(\theta)$; $y_i$ is observed for $i = 1, 2, \ldots, r$, and the remaining $(n - r)$ values are not observed but grouped into $J$ categories. If subject $i$ belongs to the $j$th category, then $y_i \in (a_j, b_j)$, where $a_j$ and $b_j$ are known constants.

Example 15.7. Bivariate normal stochastic censoring model. Suppose $Y_i$ is incompletely observed, $Y_2$ is never observed, $p$ covariates $X$ are fully observed, and for case $i$, $f(Y \mid X, \theta)$ is specified by

$$
\begin{pmatrix}
  y_{i1} \\
  y_{i2}
\end{pmatrix}
\sim_{\text{ind}} N_2
\begin{bmatrix}
  \left( \begin{array}{c}
  x_i \beta_1 \\
  x_i \beta_2
  \end{array} \right), \\
  \begin{pmatrix}
  \sigma_1^2 & \rho \sigma_1 \\
  \rho \sigma_1 & 1
  \end{pmatrix}
\end{bmatrix},
$$

The covariates $x_i$ are vectors with length $(p + 1)$. The missing-data indicators $M_{i1} = 1$ if and only if $y_{i2} \leq 0$; $M_{i2} \equiv 1$. Therefore:

$$
pr(M_{i1} = 1 \mid y_{i1}, x_i) = pr(y_{i2} \leq 0 \mid y_{i1}, x_i) = 1 - \Phi\left\{\frac{\mu_{i2} + \rho \sigma_1^{-1}(y_{i1} - \mu_{i1})}{\sqrt{1 - \rho^2}}\right\},
$$

where $\mu_{i1} = x_i \beta_1, \mu_{i2} = x_i \beta_2$.

The complete-data sufficient statistics are $\{\sum_i y_{i1} x_{ij}, \sum_i y_{i2} x_{ij}, \sum_i y_{i1} y_{i2}, \sum_i y_{i1}^2, \sum_i y_{i2}^2\}$.
4. EM algorithm for the Type II Tobit model

E step:

\[
E(y_{i2} | y_{i2} \leq 0) = \mu_{i2} - \lambda(-\mu_{i2}),
\]
\[
E(y_{i1} | y_{i2} \leq 0) = \mu_{i1} - \rho \sigma_1 \lambda(-\mu_{i2}),
\]
\[
E(y_{i2}^2 | y_{i2} \leq 0) = 1 + \mu_{i2}^2 - \mu_{i2} \lambda(-\mu_{i2}),
\]
\[
E(y_{i1}^2 | y_{i2} \leq 0) = \mu_{i1}^2 + \sigma_1^2 - \rho \sigma_1 \lambda(-\mu_{i2})(2\mu_{i1} - \rho \sigma_1 \mu_{i2}),
\]
\[
E(y_{i1}y_{i2} | y_{i2} \leq 0) = \mu_{i1} \{\mu_{i2} - \lambda(-\mu_{i2})\} + \rho \sigma_1,
\]
\[
E(y_{i2} | y_{i1}, y_{i2} > 0) = \mu_{i2.1} + \sqrt{1 - \rho^2} \lambda(\mu_{i2.1}/\sqrt{1 - \rho^2}),
\]
\[
E(y_{i2}^2 | y_{i1}, y_{i2} > 0) = 1 - \rho^2 + \mu_{i2.1}^2 + \mu_{i2.1} \sqrt{1 - \rho^2} \lambda(\mu_{i2.1}/\sqrt{1 - \rho^2}),
\]

where \(\lambda(t) = \phi(t)/\Phi(t)\), \(\mu_{i2.1} = \mu_{i2} + \rho \sigma_1^{-1}(y_{i1} - \mu_{i1})\) is the conditional mean of \((Y_2 | Y_1)\) corresponding to subject \(i\).

M step

(A). Regress \(Y_2\) on \(X\), yielding estimates of \(\beta_2\), say \(\hat{\beta}_2\).
(B). Regress \(Y_1\) on \(X\) and \(Y_2\), yielding regression parameters of \((Y_1 | X, Y_2)\): coefficients \(\hat{\delta}\) for \(Y_2\), \(\hat{\beta}_1^*\) for \(X\), and residual variance \(\hat{\sigma}_{11.2}\).
(C). Let \(\hat{\beta}_1 = \hat{\beta}_1^* + \hat{\delta} \hat{\beta}_2\), \(\hat{\sigma}^2_1 = \hat{\sigma}_{11.2} + \hat{\delta}^2\), and \(\hat{\rho} = \hat{\delta}/\hat{\sigma}_1\).

Biostat 2065: Analysis of Incomplete Data, 2005