Mixed Effects Models for Censored Data

Brought to you by the letters E and M!

Leigh Fisher

May 1, 2012
Remember the problem

- HIV Therapeutic drug trial originally analyzed by Paxton et al. (1997)
- Outcome of interest: Viral load measurements
- Viral load is measured in RNA copies per mL of blood
- At the time, lower detection limit was 500 RNA copies/ml
- In Paxton’s data, 38% of observations were censored

In some cases censoring for a given individual was much higher.

Hughes proposed a Monte Carlo EM Algorithm for mixed effects models with censoring.
The Model

We model the complete data for the $i^{th}$ individual as

$$Y_i = X_i \beta + Z_i b_i + e_i$$

Where

- $e_i$ is a vector of random errors independent of $b_i$
- $b_i \sim N(0, G(\alpha))$
- $e_{ij} \sim N(0, \sigma^2)$

We want to estimate $\theta = (\beta, \alpha, \sigma^2)$. 
In a perfect world...

We would observe \((Y_i, b_i)\) and would estimate \(\beta\) using weighted least squares (WLS)

\[
\hat{\beta} = \left( \sum_{i=1}^{m} X_i^T \Sigma^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{m} X_i^T \Sigma^{-1} Y_i \right)
\]

And we would estimate the covariance components with:

\[
\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^{m} e_i^T e_i / \sum_{i=1}^{m} n_i = t_1 / \sum_{i=1}^{m} n_i
\]

\[
\hat{G} = \frac{1}{m} \sum_{i=1}^{m} b_i b_i^T = t_2 / m
\]
In the real world, we didn’t see $b_i$

We estimate what we don’t observe:

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

$$\hat{b}_i = \hat{G} Z_i^T \hat{\Sigma}_i^{-1} (Y_i - X_i \hat{\beta})$$

$$\hat{e}_i = Y_i - X_i \hat{\beta} - Z_i \hat{b}_i = (I - Z_i \hat{G} Z_i^T \hat{\Sigma}_i^{-1})(Y_i - X_i \hat{\beta})$$

And use them to estimate the variance components:

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^{m} \hat{e}_i^T \hat{e}_i / \sum_{i=1}^{m} n_i = \hat{t}_1 / \sum_{i=1}^{m} n_i$$

$$\hat{G} = \frac{1}{m} \sum_{i=1}^{m} \hat{b}_i \hat{b}_i^T = \hat{t}_2 / m$$

That is (basically) what the EM Algorithm does.
Estimating the $b_i$'s

With our knowledge of normal distributions we can write:

\[
\begin{bmatrix}
Y_i \\
b_i
\end{bmatrix} \sim \text{ind } N \left( \begin{bmatrix}
X_i \beta \\
0
\end{bmatrix}, \begin{bmatrix}
\Sigma_i(\alpha) & Z_iG(\alpha) \\
G(\alpha)Z_i^T & G(\alpha)
\end{bmatrix} \right)
\]

Where

\[
\Sigma_i(\alpha) = Z_iGZ_i^T + \sigma^2 I
\]

Marginally,

\[
Y_i \sim \text{ind } N \left( X_i \beta, \Sigma_i(\alpha) \right)
\]

And conditionally,

\[
b_i | Y_i \sim \text{ind } N \left( GZ_i^T \Sigma_i^{-1} (Y_i - X_i \beta), (\sigma^2 Z_i^T Z_i + G^{-1})^{-1} \right)
\]
E is for Expectation

Call $\theta = (\beta, \alpha, \sigma^2)$ and let $\hat{\theta}^{(k)}$ be the estimate at step $k$.

In the E-step, we take our current estimate, $\hat{\theta}^{(k)}$ to compute $\hat{t}_1$ and $\hat{t}_2$

$$
\hat{t}_1^{(k)} = E \left[ \sum_{i=1}^{m} e_i^T e_i | Y_i, \hat{\theta}^{(k)} \right] \\
= \sum_{i=1}^{m} \hat{e}_i^T \hat{e}_i + \text{trace} \left( \text{Var} \left[ e_i | Y_i, \hat{\theta}^{(k)} \right] \right) \\
= \sum_{i=1}^{m} \left[ \hat{e}_i^T \hat{e}_i + \hat{\sigma}^2(k) \left( n_i - \hat{\sigma}^2(k) \text{tr} \left( \hat{\Sigma}_i^{(k)} \right) \right) \right]
$$

Where $\hat{\Sigma}_i^{(k)} = Z_i G(\hat{\alpha}^{(k)}) Z_i^T + \hat{\sigma}^2(k) I$
And we can similarly compute

$$
\hat{t}_2^{(k)} = E \left[ \sum_{i=1}^{m} b_i b_i^T | Y_i, \hat{\theta}^{(k)} \right]
$$

$$
= \sum_{i=1}^{m} \hat{b}_i \hat{b}_i^T + \text{Var} \left[ b_i | Y_i, \hat{\theta}^{(k)} \right]
$$

$$
= \sum_{i=1}^{m} \left[ \hat{b}_i \hat{b}_i^T + G(\hat{\alpha}^{(k)}) - G(\hat{\alpha}^{(k)})Z_i^T \hat{\Sigma}_i^{(k)} Z_i G(\hat{\alpha}^{(k)}) \right]
$$
M is for Maximize

We use the expectations from the E-step to compute the $k + 1$ estimates:

\[
\hat{\sigma}^2(k+1) = \frac{\hat{\tau}_1^{(k)}}{\sum_{i=1}^m n_i}
\]

\[
\hat{G}^{(k+1)} = \frac{\hat{\tau}_2^{(k)}}{m}
\]

Iterate between the E and M steps until convergence.
Let $d_l$ and $d_u$ be the lower and upper detection limits. Define $c_{ij}$ to be an indicator variable which not only indicates censoring and the direction of the censoring:

$$c_{ij} = \begin{cases} 
-1 & \text{if } Y_{ij} \leq d_l \\
0 & \text{if } d_l \leq Y_{ij} \leq d_u \\
1 & \text{if } Y_{ij} \geq d_u 
\end{cases}$$

So we observe $(Q_i, C_i)$ where

$$Q_{ij} \geq Y_{ij} \quad \text{if } c_{ij} = -1$$
$$Q_{ij} = Y_{ij} \quad \text{if } c_{ij} = 0$$
$$Q_{ij} \leq Y_{ij} \quad \text{if } c_{ij} = 1$$
What do we do with \((Q_i, C_i)\)?

If the observation is uncensored, ie \(c_{ij} = 0\), we use the observed data in our computations (duh!)

If the observation is censored, we use conditional expectations in our computations

If \(c_{ij} = -1\), use \(E[Y_{ij} | Y_{ij} \leq d_l, \theta]\)

If \(c_{ij} = 1\), use \(E[Y_{ij} | Y_{ij} \geq d_u, \theta]\)
In the E-Step, we will need to compute the following conditional expectations:

$$
\hat{\beta}^{(k)} = \left( \sum_{i=1}^{m} X_i^T \hat{\Sigma}^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{m} X_i^T \hat{\Sigma}^{-1} E[Y_i|Q_i, C_i, \hat{\theta}^{(k)}] \right)
$$

$$
\hat{t}_1^{(k)} = \sum_{i=1}^{m} E\left[ e_i^T e_i | Q_i, C_i, \hat{\theta}^{(k)} \right]
$$

$$
\hat{t}_2^{(k)} = \sum_{i=1}^{m} E\left[ b_i b_i^T | Q_i, C_i, \hat{\theta}^{(k)} \right]
$$
Consider computing for the $i$th individual

$$E \left[ e_i^T e_i | Q_i, C_i, \hat{\theta}^{(k)} \right] = E \left[ E[ e_i^T e_i | Y_i, \hat{\theta}^{(k)} ] | Q_i, C_i, \hat{\theta}^{(k)} \right]$$

$$= E \left[ \hat{e}_i^T \hat{e}_i + \hat{\sigma}^2(k) n_i - \hat{\sigma}^4(k) \text{tr} \left( \hat{\Sigma}^{(k)}_i \right) | Q_i, C_i, \hat{\theta}^{(k)} \right]$$

If $e_i$ has length 1 - this is a function of the variance of a truncated normal, which we can compute (ie use R or look up the formula on Wikipedia).

Two dimensional case can be done with some bivariate truncated normals.

If there are 10 observations, and 6 of them are censored, we are considering a multivariate normal distribution where some of the observations are truncated...

Which is why we use Gibbs!
And now for the easy part

The M-Step is to compute:

\[
\hat{\beta}^{(k+1)} = \left( \sum_{i=1}^{m} X_i^T \hat{\Sigma}^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{m} X_i^T \hat{\Sigma}^{-1} E[Y_i|Q_i, C_i, \hat{\theta}^{(k)}] \right)
\]

\[
\hat{\sigma}^{2(k+1)} = \sum_{i=1}^{m} E\left[ e_i^T e_i | Q_i, C_i, \hat{\theta}^{(k)} \right] / \sum_{i=1}^{m} n_i
\]

\[
\hat{G}^{(k+1)} = \frac{1}{m} \sum_{i=1}^{m} E\left[ b_i b_i^T | Q_i, C_i, \hat{\theta}^{(k)} \right]
\]
MCEM Algorithm

The basic idea

- E-step: Use Monte Carlo methods to evaluate the estimated expectations of the log-likelihood functions with respect to the conditional distributions under the current estimates of the parameters
- M-step: Find new estimates of the parameters that maximize the expectations

Repeat until convergence
Next Steps

Next up:

- Start simulating
- Clean up the math
- Thoughts? Ideas?