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

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i$$

Where

- $\mathbf{e}_i$ is a vector of random errors independent of $\mathbf{b}_i$
- $\mathbf{b}_i \sim N(0, \mathbf{G}(\boldsymbol{\alpha}))$
- $e_{ij} \sim N(0, \sigma^2)$

We want to estimate $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2)$.

In a perfect world...

We would observe $(\mathbf{Y}_i, \mathbf{b}_i)$ and would estimate $\boldsymbol{\beta}$ using weighted least squares (WLS)

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m} \mathbf{X}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{m} \mathbf{X}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{Y}_i \right)$$

And we would estimate the covariance components with:

$$\begin{aligned} \hat{\sigma}^2 &= \sum_{i=1}^{m} \mathbf{e}_i^T \mathbf{e}_i \Big/ \sum_{i=1}^{m} n_i = t_1 \Big/ \sum_{i=1}^{m} n_i \\ \hat{\mathbf{G}} &= \frac{1}{m} \sum_{i=1}^{m} \mathbf{b}_i \mathbf{b}_i^T = \mathbf{t}_2 / m \end{aligned}$$

In the real world, we didn't see $\mathbf{b}_i$

We estimate what we don't observe:

$$\begin{aligned}
\hat{\beta} &= \left(\sum_{i=1}^{m} \mathbf{X}_i^T \hat{\mathbf{\Sigma}}^{-1} \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{m} \mathbf{X}_i^T \hat{\mathbf{\Sigma}}^{-1} \mathbf{Y}_i\right) \\
\hat{\mathbf{b}}_i &= \hat{\mathbf{G}}\mathbf{Z}_i^T \hat{\mathbf{\Sigma}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}) \\
\hat{\mathbf{e}}_i &= \mathbf{Y}_i - \mathbf{X}_i \hat{\beta} - \mathbf{Z}_i \hat{\mathbf{b}}_i = (\mathbf{I} - \mathbf{Z}_i \hat{\mathbf{G}} \mathbf{Z}_i^T \hat{\mathbf{\Sigma}}_i^{-1})(\mathbf{Y}_i - \mathbf{X}_i \hat{\beta})
\end{aligned}$$

And use them to estimate the variance compontents:

$$\begin{aligned}
\hat{\sigma}^2 &= \sum_{i=1}^{m} \hat{\mathbf{e}}_i^T \hat{\mathbf{e}}_i \Big/ \sum_{i=1}^{m} n_i = \hat{t}_1 \Big/ \sum_{i=1}^{m} n_i \\
\hat{\mathbf{G}} &= \frac{1}{m} \sum_{i=1}^{m} \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^T = \hat{\mathbf{t}}_2 / m
\end{aligned}$$

That is (basically) what the EM Algorithm does.

Estimating the $\mathbf{b}_i$'s

With our knowledge of normal distributions we can write:

$$\begin{bmatrix} \mathbf{Y}_i \\ \mathbf{b}_i \end{bmatrix} \overset{\text{ind}}{\sim} N\left(\begin{bmatrix} \mathbf{X}_i\boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_i(\boldsymbol{\alpha}) & \mathbf{Z}_i\mathbf{G}(\boldsymbol{\alpha}) \\ \mathbf{G}(\boldsymbol{\alpha})\mathbf{Z}_i^T & \mathbf{G}(\boldsymbol{\alpha}) \end{bmatrix} \right)$$

Where

$$\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}) = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^T + \sigma^2\mathbf{I}$$

Marginally,

$$\mathbf{Y}_i \overset{\text{ind}}{\sim} N\Big(\mathbf{X}_i\boldsymbol{\beta},\ \boldsymbol{\Sigma}_i(\boldsymbol{\alpha})\Big)$$

And conditionally,

$$\mathbf{b}_i|\mathbf{Y}_i \overset{\text{ind}}{\sim} N\Big(\mathbf{G}\mathbf{Z}_i^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}),\ (\sigma^2\mathbf{Z}_i^T\mathbf{Z}_i + \mathbf{G}^{-1})^{-1}\Big)$$

E is for Expectation

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

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

$$\begin{aligned}
\hat{t}_1^{(k)} &= \mathrm{E}\left[\sum_{i=1}^{m} \mathbf{e}_i^T \mathbf{e}_i | \mathbf{Y}_i, \hat{\boldsymbol{\theta}}^{(k)}\right] \\
&= \sum_{i=1}^{m} \hat{\mathbf{e}}_i^T \hat{\mathbf{e}}_i + \mathrm{trace}\Big(\mathrm{Var}\left[\mathbf{e}_i | \mathbf{Y}_i, \hat{\boldsymbol{\theta}}^{(k)}\right]\Big) \\
&= \sum_{i=1}^{m} \left[\hat{\mathbf{e}}_i^T \hat{\mathbf{e}}_i + \hat{\sigma}^{2(k)}\Big(n_i - \hat{\sigma}^{2(k)} \mathrm{tr}\Big(\hat{\boldsymbol{\Sigma}}_i^{(k)}\Big)\Big)\right]
\end{aligned}$$

Where $\hat{\boldsymbol{\Sigma}}_i^{(k)} = \mathbf{Z}_i \mathbf{G}(\hat{\alpha}^{(k)}) \mathbf{Z}_i^T + \hat{\sigma}^{2(k)} \mathbf{I}$

E is for Expectation

And we can similarly compute

$$
\begin{aligned}
\hat{\mathbf{t}}_2^{(k)} &= \mathrm{E}\left[\sum_{i=1}^{m} \mathbf{b}_i \mathbf{b}_i^T | \mathbf{Y}_i, \hat{\theta}^{(k)}\right] \\
&= \sum_{i=1}^{m} \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^T + \mathrm{Var}\left[\mathbf{b}_i | \mathbf{Y}_i, \hat{\theta}^{(k)}\right] \\
&= \sum_{i=1}^{m} \left[\hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^T + \mathbf{G}(\hat{\alpha}^{(k)}) - \mathbf{G}(\hat{\alpha}^{(k)}) \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{(k)} \mathbf{Z}_i \mathbf{G}(\hat{\alpha}^{(k)})\right]
\end{aligned}
$$

M is for Maximize

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

$$
\begin{aligned}
\hat{\sigma}^{2(k+1)} &= \hat{t}_1^{(k)} \Big/ \sum_{i=1}^{m} n_i \\
\hat{\mathbf{G}}^{(k+1)} &= \hat{\mathbf{t}}_2^{(k)} / m
\end{aligned}
$$

Iterate between the E and M steps until convergence.

The Observed Data

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:

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

So we observe $(\mathbf{Q}_i, \mathbf{C}_i)$ where

$$\begin{aligned} Q_{ij} \geq Y_{ij} & \quad \text{if} \quad c_{ij} = -1 \\ Q_{ij} = Y_{ij} & \quad \text{if} \quad c_{ij} = 0 \\ Q_{ij} \leq Y_{ij} & \quad \text{if} \quad c_{ij} = 1 \end{aligned}$$

What do we do with $(\mathbf{Q}_i, \mathbf{C}_i)$?

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 $\mathrm{E}[Y_{ij} | Y_{ij} \leq d_l, \boldsymbol{\theta}]$
If $c_{ij} = 1$, use $\mathrm{E}[Y_{ij} | Y_{ij} \geq d_u, \boldsymbol{\theta}]$

E as in cEnsored Expectation

In the E-Step, we will need to compute the following conditional expectations

$$\hat{\beta}^{(k)} = \left(\sum_{i=1}^{m} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{m} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}^{-1} \mathrm{E}[\mathbf{Y}_i | \mathbf{Q}_i, \mathbf{C}_i, \hat{\theta}^{(k)}]\right)$$

$$\hat{t}_1^{(k)} = \sum_{i=1}^{m} \mathrm{E}\left[\mathbf{e}_i^T \mathbf{e}_i | \mathbf{Q}_i, \mathbf{C}_i, \hat{\theta}^{(k)}\right]$$

$$\hat{\mathbf{t}}_2^{(k)} = \sum_{i=1}^{m} \mathrm{E}\left[\mathbf{b}_i \mathbf{b}_i^T | \mathbf{Q}_i, \mathbf{C}_i, \hat{\theta}^{(k)}\right]$$

E as in Eeeks!

Consider computing for the ith individual

$$\begin{aligned} \mathrm{E}\left[\mathbf{e}_i^T\mathbf{e}_i|\mathbf{Q}_i,\mathbf{C}_i,\hat{\boldsymbol{\theta}}^{(k)}\right] &= \mathrm{E}\left[\mathrm{E}[\mathbf{e}_i^T\mathbf{e}_i|\mathbf{Y}_i,\hat{\boldsymbol{\theta}}^{(k)}]\,\middle|\,\mathbf{Q}_i,\mathbf{C}_i,\hat{\boldsymbol{\theta}}^{(k)}\right] \\ &= \mathrm{E}\left[\hat{\mathbf{e}}_i^T\hat{\mathbf{e}}_i+\hat{\sigma}^{2(k)}n_i-\hat{\sigma}^{4(k)}\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_i^{(k)}\right)\,\middle|\,\mathbf{Q}_i,\mathbf{C}_i,\hat{\boldsymbol{\theta}}^{(k} \end{aligned}$$

If $\mathbf{e}_i$ 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:

$$
\begin{aligned}
\hat{\beta}^{(k+1)} &= \left(\sum_{i=1}^{m} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{m} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}^{-1} \mathrm{E}[\mathbf{Y}_i | \mathbf{Q}_i, \mathbf{C}_i, \hat{\theta}^{(k)}]\right) \\
\hat{\sigma}^{2(k+1)} &= \sum_{i=1}^{m} \mathrm{E}\left[\mathbf{e}_i^T \mathbf{e}_i | \mathbf{Q}_i, \mathbf{C}_i, \hat{\theta}^{(k)}\right] \Big/ \sum_{i=1}^{m} n_i \\
\hat{\mathbf{G}}^{(k+1)} &= \frac{1}{m} \sum_{i=1}^{m} \mathrm{E}\left[\mathbf{b}_i \mathbf{b}_i^T | \mathbf{Q}_i, \mathbf{C}_i, \hat{\theta}^{(k)}\right]
\end{aligned}
$$

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?