Mixed Effects Models for Censored Data with Application to HIV RNA Levels

By James P Hughes

Leigh Fisher

University of Washington, Department of Biostatistics

May 24, 2012

Linear mixed effects models with informative censoring

Motivating Example

- 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

Concern is how to accurately estimate fixed effects and variance components.

As statisticians, we are especially wary of the ad hoc methods of simply replacing the censored observation with the known level of detection (dl) or some function of it.

Hughes presents a Monte Carlo EM algorithm to obtain better estimates.

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 $\mathbf{b}_i$
- $\mathbf{b}_i \sim N(0, \mathbf{G}(\boldsymbol{\alpha}))$
- $e_{ij} \sim N(0, \sigma^2)$

Marginally,

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

Where

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

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

Recall the EM Method (Complete Data)

(The E Step) We first estimate:

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

And for each individual compute

$$\begin{aligned}
\mathrm{E}\left[\mathbf{e}_i^T \mathbf{e}_i | \mathbf{Y}_i, \hat{\theta}^{(k)}\right] &= \hat{\mathbf{e}}_i^T \hat{\mathbf{e}}_i + \hat{\sigma}^{2(k)} \left(n_i - \hat{\sigma}^{2(k)} \mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_i^{-1(k)}\right)\right) \\
\mathrm{E}\left[\mathbf{b}_i \mathbf{b}_i^T | \mathbf{Y}_i, \hat{\theta}^{(k)}\right] &= \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^T + \hat{\mathbf{G}}^{(k)} - \hat{\mathbf{G}}^{(k)} \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1(k)} \mathbf{Z}_i \hat{\mathbf{G}}^{(k)}
\end{aligned}$$

Recall the EM Method (Complete Data)

(The M Step) And use them to estimate the variance components:

$$\hat{\sigma}^{2(k+1)} = \hat{t}_1^{(k)} \Big/ \sum_{i=1}^{m} n_i \quad \text{and} \quad \hat{\mathbf{G}}^{(k+1)} = \hat{\mathbf{t}}_2^{(k)}/m$$

Where

$$\hat{t}_1^{(k)} = \sum_{i=1}^{m} \mathrm{E}\left[\mathbf{e}_i^T \mathbf{e}_i | \mathbf{Y}_i, \hat{\boldsymbol{\theta}}^{(k)}\right] \quad \text{and} \quad \hat{\mathbf{t}}_2^{(k)} = \sum_{i=1}^{m} \mathrm{E}\left[\mathbf{e}_i^T \mathbf{e}_i | \mathbf{Y}_i, \hat{\boldsymbol{\theta}}^{(k)}\right]$$

Then iterate between the E and M step until convergence!

But in our problem we don't have complete data.

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

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

Modified EM for Censored Data

In the E-Step, we compute

$$
\begin{aligned}
\hat{\beta}^{(k)} &= \left(\sum_{i=1}^{m} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{m} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}^{-1} \mathrm{E}[\mathbf{Y}_i | \mathbf{Q}_i, \mathbf{C}_i, \hat{\theta}^{(k)}]\right) \\
\hat{t}_1^{(k)} &= \sum_{i=1}^{m} \mathrm{E}\left[\mathbf{e}_i^T \mathbf{e}_i | \mathbf{Q}_i, \mathbf{C}_i, \hat{\theta}^{(k)}\right] \\
\hat{\mathbf{t}}_2^{(k)} &= \sum_{i=1}^{m} \mathrm{E}\left[\mathbf{b}_i \mathbf{b}_i^T | \mathbf{Q}_i, \mathbf{C}_i, \hat{\theta}^{(k)}\right]
\end{aligned}
$$

Instead of integrating something hard $\mathrm{E}\left[\mathbf{e}_i^T \mathbf{e}_i | \mathbf{Q}_i, \mathbf{C}_i, \hat{\theta}^{(k)}\right]$, we use a Gibbs sampler.

Simple Example

S'pose we observe $(\mathbf{Q}_i, \mathbf{C}_i)$ and $\mathbf{C}_i = (0, -1, 0, 1)^T$ and let $\hat{\boldsymbol{\theta}}^{(k)}$ be our current estimate of $\boldsymbol{\theta}$. We want to use Gibbs to obtain a

$$\mathbf{Y}_i^{(s)} \sim f(\mathbf{Y}_i | \mathbf{Q}_i, \mathbf{C}_i, \hat{\boldsymbol{\theta}}^{(k)})$$

We need a (good) starting value and we know $Y_{i2} \leq d_l$

sample $u \sim \text{Unif}(0,\, d_l)$ and let $Y_{i2}^0 = \Phi(u)$

And then we can do the standard Gibbs procedure.

Introducing MC Gibbs

Sample Y_{i4} from the conditional distribution given Y_{i1}, Y_{i3}, and Y_{12}^0.

$$Y_{i4}^0 \sim f\left(Y_{i4} \,\middle|\, Y_{i1}, Y_{i3}, Y_{i2} = Y_{i2}^0, Y_{i4} \geq d_u, \hat{\boldsymbol{\theta}}^{(k)}\right)$$

Update the conditional distribution of Y_{i2} to gain a new estimate of Y_{i2}

$$Y_{i2}^1 \sim f\left(Y_{i2} \,\middle|\, Y_{i1}, Y_{i3}, Y_{i4} = Y_{i4}^0, Y_{i2} \leq d_l, \hat{\boldsymbol{\theta}}^{(k)}\right)$$

Good thing we assumed normality - these are simply truncated univariate conditional normal distributions.

The E Step

After a brief burn-in, $\mathbf{Y}_i^{(s)} = \left(Y_{i1}, Y_{i2}^{(s)}, Y_{i3}, Y_{i4}^{(s)}\right)^T$ represent random samples from $f(\mathbf{Y}_i|\mathbf{Q}_i, \mathbf{C}_i, \hat{\theta}^{(k)})$. So we just generate N_{mc} samples. For each $\mathbf{Y}_i^{(s)}$, we compute

$$
\begin{aligned}
\hat{t}_{1,s} &= \hat{\mathbf{e}}_{i,s}^T \hat{\mathbf{e}}_{i,s} + \hat{\sigma}^2 \left(n_i - \hat{\sigma}^2 \mathrm{tr}\left(\hat{\mathbf{\Sigma}}_i \right) \right) \\
\hat{\mathbf{t}}_{2,s} &= \hat{\mathbf{b}}_{i,s} \hat{\mathbf{b}}_{i,s}^T + \hat{\mathbf{G}} - \hat{\mathbf{G}} \mathbf{Z}_i^T \hat{\mathbf{\Sigma}}_i \mathbf{Z}_i \hat{\mathbf{G}}
\end{aligned}
$$

Where

$$
\begin{aligned}
\hat{\mathbf{e}}_{i,s} &= (\mathbf{I}_{n_i} - \mathbf{Z}_i \hat{\mathbf{G}} \mathbf{Z}_i^T \hat{\mathbf{\Sigma}}_i^{-1})(\mathbf{Y}_i^{(s)} - \mathbf{X}_i \hat{\beta}) \\
\hat{\mathbf{b}}_{i,s} &= \hat{\mathbf{G}} \mathbf{Z}_i^T \hat{\mathbf{\Sigma}}_i^{-1} (\mathbf{Y}_i^{(s)} - \mathbf{X}_i \hat{\beta})
\end{aligned}
$$

The MCE Step

For each cluster, we estimate the conditional expectations with average over the Gibbs samples. We let

$$
\begin{aligned}
\mathrm{E}\left[\mathbf{Y}_i|\mathbf{Q}_i,\mathbf{C}_i,\hat{\boldsymbol{\theta}}^{(k)}\right] &\approx \frac{1}{N_{mc}}\sum_{s=1}^{N_{mc}}\mathbf{Y}_{i,s} \\
\mathrm{E}\left[\mathbf{e}_i^T\mathbf{e}_i|\mathbf{Q}_i,\mathbf{C}_i,\hat{\boldsymbol{\theta}}^{(k)}\right] &\approx \frac{1}{N_{mc}}\sum_{s=1}^{N_{mc}}\hat{\mathbf{e}}_{i,s}^T\hat{\mathbf{e}}_{i,s}+\hat{\sigma}^2\left(n_i-\hat{\sigma}^2\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_i\right)\right) \\
\mathrm{E}\left[\mathbf{b}_i\mathbf{b}_i^T|\mathbf{Q}_i,\mathbf{C}_i,\hat{\boldsymbol{\theta}}^{(k)}\right] &\approx \frac{1}{N_{mc}}\sum_{s=1}^{N_{mc}}\hat{\mathbf{b}}_{i,s}\hat{\mathbf{b}}_{i,s}^T+\hat{\mathbf{G}}-\hat{\mathbf{G}}\mathbf{Z}_i^T\hat{\boldsymbol{\Sigma}}_i\mathbf{Z}_i\hat{\mathbf{G}}
\end{aligned}
$$

And then use those estimates in the M step.

Modified EM for Censored Data

The M-Step is to compute:

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

Head back to the MCE Step using $\hat{\boldsymbol{\theta}}^{(k+1)}$ and repeat until convergence!

A few details

- Double Gibbs sample size at each iteration that reduces the log-likelihood
- Stop when the change in log-likelihood was less than 0.01%

Aside: ML vs REML

Why do we care about REML estimates?

> We know the MLE for σ^2 is biased downwards in classical linear models; REML standard error estimates are typically less biased than the MLE.

Laird & Ware showed how to obtain both ML and REML estimates using the EM algorithm, so an analogous extension can be made for the case of censored data.

The good news: the MCEM procedures are equivalent, but there are some differences in the E step

I will highlight the key differences.

Aside: ML vs REML

In the E Step, estimate $\hat{\mathbf{b}}_i$ and $\hat{\mathbf{e}}_i$ using the current estimate $\boldsymbol{\theta}^{(k)}$.

ML: $\mathbf{Y}_i - \mathbf{X}_i\hat{\beta} = \left(\mathbf{I}_{n_i} - \mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T\right)\mathbf{Y}_i$

$$\begin{aligned}
\hat{\mathbf{b}}_i^{(k)} &= \hat{\mathbf{G}}\mathbf{Z}_i^T\hat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta}) \\
\hat{\mathbf{e}}_i^{(k)} &= (\mathbf{I} - \mathbf{Z}_i\hat{\mathbf{G}}\mathbf{Z}_i^T\hat{\boldsymbol{\Sigma}}_i^{-1})(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta})
\end{aligned}$$

REML: $\left(\mathbf{I}_i^+ - \mathbf{X}_i\left(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\boldsymbol{\Sigma}^{-1}\right)\mathbf{Y}$

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

Where $\mathbf{I}_i^+$ is such that $\mathbf{I}_i^+\mathbf{Y} = \mathbf{Y}_i$

Aside: ML to REML

And the expectations can be worked out ML:

$$\begin{aligned}
\mathrm{E}\left[\mathbf{e}_i^T\mathbf{e}_i|\mathbf{Y}_i,\hat{\boldsymbol{\theta}}^{(k)}\right] &= \hat{\mathbf{e}}_i^T\hat{\mathbf{e}}_i + \hat{\sigma}^{2(k)}\left(n_i - \hat{\sigma}^{2(k)}\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_i^{-1(k)}\right)\right) \\
\mathrm{E}\left[\mathbf{b}_i\mathbf{b}_i^T|\mathbf{Y}_i,\hat{\boldsymbol{\theta}}^{(k)}\right] &= \hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^T + \hat{\mathbf{G}}^{(k)} - \hat{\mathbf{G}}^{(k)}\mathbf{Z}_i^T\hat{\boldsymbol{\Sigma}}_i^{-1(k)}\mathbf{Z}_i\hat{\mathbf{G}}^{(k)}
\end{aligned}$$

REML:

$$\begin{aligned}
\mathrm{E}\left[\mathbf{e}_i^T\mathbf{e}_i|\mathbf{Y}_i,\hat{\boldsymbol{\theta}}^{(k)}\right] &= \hat{\mathbf{e}}_i^T\hat{\mathbf{e}}_i + \hat{\sigma}^{2(k)}\left(n_i - \hat{\sigma}^{2(k)}\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_i^{-1(k)} - \mathbf{H}_i\right)\right) \\
\mathrm{E}\left[\mathbf{b}_i\mathbf{b}_i^T|\mathbf{Y}_i,\hat{\boldsymbol{\theta}}^{(k)}\right] &= \hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^T + \hat{\mathbf{G}}^{(k)} - \hat{\mathbf{G}}^{(k)}\mathbf{Z}_i^T\left(\hat{\boldsymbol{\Sigma}}_i^{-1(k)} - \mathbf{H}_i\right)\mathbf{Z}_i\hat{\mathbf{G}}^{(k)}
\end{aligned}$$

Where

$$\mathbf{H}_i = \boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\left(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}_i\boldsymbol{\Sigma}_i^{-1}$$

The variables without i's are the 'stacked' matrices and $\boldsymbol{\Sigma}$ is a $mn_i \times mn_i$ block diagonal matrix of $\boldsymbol{\Sigma}_i$'s

Simulations

Simulated, instead

$$\mathbf{Y}_i = \beta_0 \mathbf{1}_{n_i} + \beta_1 \,\text{time} + \beta_2 \mathbf{X}_i + b_{i0} \mathbf{1}_{n_i} + b_{i1} \,\text{time} + e_{ij}$$

Where

- 40 individuals ($m = 40$) with 10 observations each ($n_i = 10$)
- $\mathbf{X}_i$ is binary treatment indicator
- $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G})$
- $e_{ij} \sim N(0, \sigma^2)$
- time is from 0 to 9

Simulations

We specify:

- $\boldsymbol{\beta} = (11, -1, -0.5)^T$
- **G** is diag(1.5, 2.5) & b_{i0} is independent of b_{i1}
- $\sigma^2 = 2$
- Only considering censoring below the level of detection
- Percentage of censored observations: 10-80%
- Gibbs sampling sizes start at 500

Note: Censoring must not be done randomly!

To start, each individual has the same level of censoring, although at different (known) levels.

Simulations

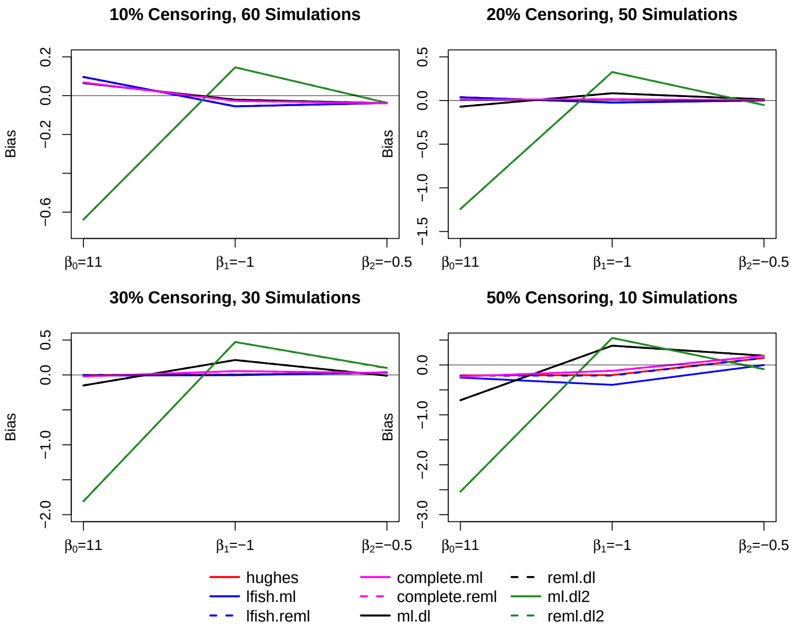
For each dataset, we compare:

- Hughes' FORTRAN program (ML only)
- Fisher's hand-coded ML and REML programs
- ML & REML using the detection limit (dl)
- ML & REML using half the detection limit (dl2)
- ML & REML estimates using the complete data

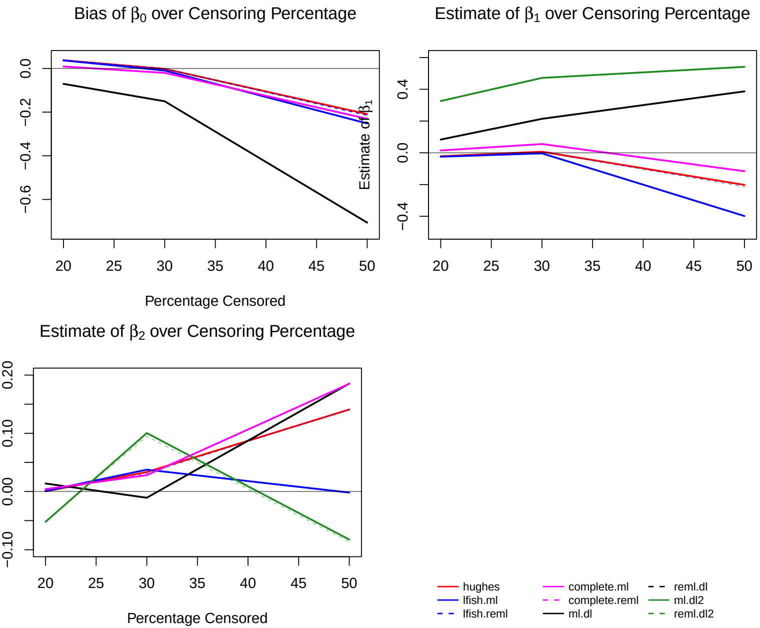
A few differences between Fisher and Hughes

- Max Gibbs sample sizes: 4,000 vs 16,384,000+?
- Max number of EM steps: 60 vs 100

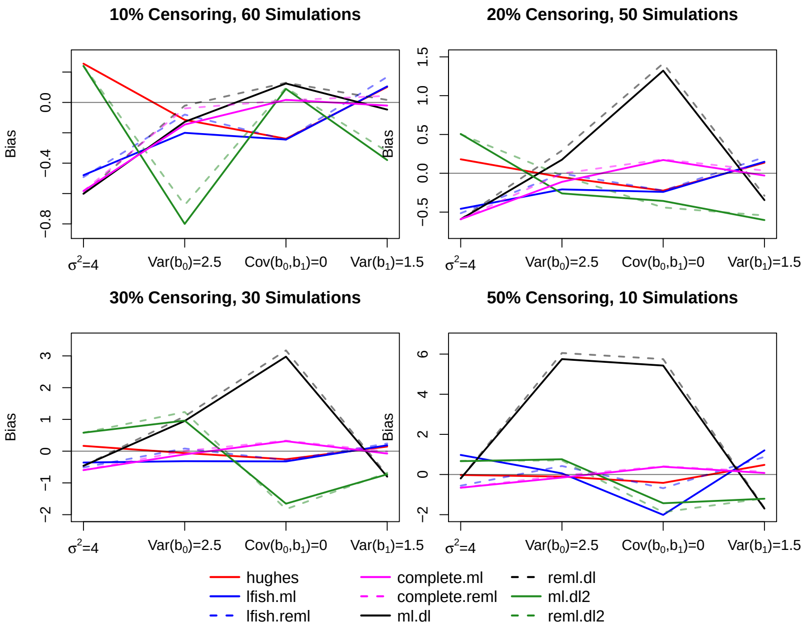
Estimating the β's



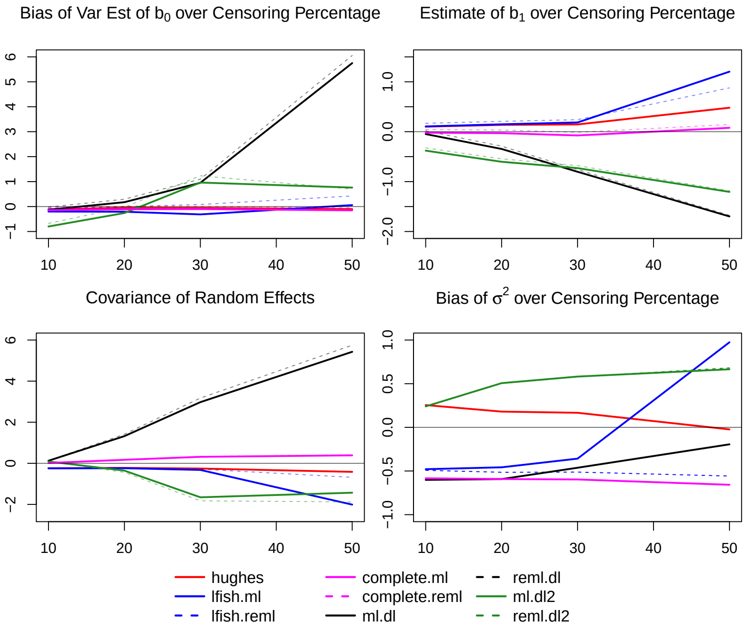
Bias of $\hat{\beta}$'s vs Percentage Censored



Estimating the variance components



Bias in Variance Components vs Percentage Censored



Big Picture

We should think hard about how to obtain good estimates when faced with informative censoring.

Ad hoc methods are almost always a bad idea, especially when there are relatively simple methods that are known to give less biased results.

The MCEM takes the sensible EM algorithm, uses a Gibbs sampler to do difficult integration, and results in much less biased estimates, especially for the variance components.

Any questions?