Introduction and Motivation

Leigh Fisher

April 10, 2012

The Paper

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

James P. Hughes

Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A.
email: hughes@biostat.washington.edu

Published in Biometrics 1999

Censoring

In this context, we are concerned with censoring as a result to detection limits of laboratory outcomes.

This kind of censoring can be nontrivial in many real-life situations.

A Feeble Example

1795706 [RF] © www.visualphotos.com

The 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

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

Previous Solutions to the Problem

Ad hoc procedures:

- Use the detection limit; e.g. use 500 copies/ml
- Use something smaller (or larger) than the detection limit

Narrow range of situations where these will lead to reasonable estimates. Generally, these procedures lead to grossly biased estimates especially as proportion of the data censored increases

Previous Solutions to the Problem

Statistically motivated procedures:

- Paxton et al. [2] used iterative imputation to adjust for censoring - fails to take the correlated structure of the data into account
- Pettitt [3] proposed a similar result, but only reached a tractable solution for the one-way random effects model

Hughes proposes a general approach using Monte Carlo methods and the EM algorithm that can be used with any censoring pattern and an arbitrarily complex design matrix for the random effects.

The Complete Data

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

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\beta_i + \mathbf{e}_i$$

Where

- $\mathbf{Y}_i$ is a vector of n_i outcomes
- $\boldsymbol{\alpha}$ is a vector of fixed effects
- $\boldsymbol{\beta}_i$ is a vector of random effects
- $\mathbf{X}_i$ and $\mathbf{Z}_i$ are the design matrices
- $\mathbf{e}_i$ is a vector of random errors independent of $\boldsymbol{\beta}_i$

For now, we also assume $\boldsymbol{\beta}_i \sim N(0, \Sigma)$ and $e_{ij} \sim N(0, \sigma^2)$.

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

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

Parameter Estimation

When there is no censoring, we know how to obtain LM and REML estimates (see 571).

Following the work of Laird and Ware (1982) [1], Hughes proposes a modified Monte Carlo EM algorithm to accommodate censored data and obtain less biased estimates

EM Algorithm

Let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}, \hat{\Sigma}, \hat{\sigma}^2)$ be our current estimation.

- **E-step:** Compute expectations $\mathrm{E}[\mathbf{Y}|\mathbf{C}, \mathbf{Q}, \hat{\theta}]$, $\mathrm{E}[\boldsymbol{\beta}_i \boldsymbol{\beta}_i^T | \mathbf{C}_i, \mathbf{Q}_i, \hat{\boldsymbol{\theta}}]$, and $\mathrm{E}[\mathbf{e}_i \mathbf{e}_i^T | \mathbf{C}_i, \mathbf{Q}_i, \hat{\boldsymbol{\theta}}]$
- **M-step:** Maximize each parameter with

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

Where's the Monte Carlo?

Where's the Monte Carlo?

Computing these expectations is problematic:

- multi-dimensional integration over the censoring ranges at each iteration (not fun!)
- when a given individual has many censored observations (basically impossible!)

Instead, we repeatedly sample $\mathbf{Y}_i$ from $f(\mathbf{Y}_i|\mathbf{C}_i, \mathbf{Q}_i, \hat{\theta})$ to estimate the expected values. Rejection sampling can be very slow, so instead we use the Gibbs sampler.

The Gibbs Sampler

Generate new values of $\mathbf{Y}_i$ by iteratively sampling from the univariate conditional distributions:

$$f(Y_{ij}|Y_{ik:k\neq j}, \hat{\boldsymbol{\theta}}) = \begin{cases} Y_{ij} < Q_{ij} & \text{if} \quad C_{ij} = -1 \\ Y_{ij} > Q_{ij} & \text{if} \quad C_{ij} = 1 \end{cases}$$

After a burn-in period, the resulting $\mathbf{Y}_i$ will have the desired distribution and the expectations in the EM algorithm can be computed easily.

Last thing to do is adjust the estimated variances of the estimated fixed effects for the information lost due to censoring.

Results

Via simulation, we see that

- *A*d hoc procedures generally produce biased point estimates and underestimate the variance components of the random effects
- Iterative imputation removes bias of the fixed effect estimate, and is unpredictably biased when estimating the variance components, σ^2 and Σ
- The MCEM algorithm produces unbiased estimates for the fixed effects and within-person variance, σ^2 and the bias of the between-person variance is much lower

Also this method is applied to previously analyzed HIV RNA viral load data

Next Steps (for the Paper)

- Utilize a fully Bayesian implementation to relax the assumptions
- Determine if there is a faster way to implement these algorithms - See Vaida & Liu (2009) [4]

Next Steps (for me)

- Understand the details of the Monte Carlo EM algorithm in this setting
- Consider a BUGS implementation
- Consider the faster implementations that have been purposed


Nan M. Laird and James H. Ware.
Random-effects models for longitudinal data.
Biometrics, 38:963–974, 1982.


WB Paxton, RW Coombs, MJ McElrath, MC Keefer, J Hughes, F Sinangil, D Chernoff, L Demeter, B Williams, and L Corey.
Longitudinal analysis of quantitative virologic measures in human immunodeficiency virus-infected subjects with ¿=400 CD4 lymphocytes: Implications for applying measurements to individual patients.
JOURNAL OF INFECTIOUS DISEASES, 175(2):247–254, FEB 1997.


A. N. Pettitt.
Censored observations, repeated measures and mixed effects models: An approach using the EM algorithm and normal errors.
Biometrika, 73:635–643, 1986.


Florin Vaida and Lin Liu.

Fast Implementation for Normal Mixed Effects Models With Censored Response.
JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS, 18(4):797–817, DEC 2009.