Generalized Estimating Equations

Robert Weiss

Biostatistics 411

email: `robweiss@ucla.edu`

`rem.ph.ucla.edu//biostat411`

Analysis of Correlated Data

Current Reading

Week 8: Chapters 21 and 22 in ALA.

Today: Reading: Section 13.2 on estimation of Marginal Models

Week 10: Chapter 13 in Weiss (2005) Modeling Longitudinal Data covers *bivariate longitudinal data*. Available online through the UCLA library at `http://www.library.ucla.edu/libraries/sel/e-books-science-engineering`

- Scroll down to and click on “Springer eBooks (2005-12)”.
- In the search box, enter “Modeling Longitudinal Data” hit enter.
- Click on the first book that appears.
- Scroll down to chapter 13.

Generalized Estimating Equations

- Abbreviated GEE.
- Due to (Zeger and Liang 1986 Biometrics; Liang and Zeger 1986 Biometrika)
- Software easy to write, runs great.
- Does not include a full model specification.
- Has a *working correlation matrix*
- *Fix* step to adjust SEs: Sandwich estimator.
- Fixed effect coefficient estimates are *population-averaged*.
- Fits *marginal* models.

Why GEE?

- There is no convenient or natural specification of the joint multivariate distribution of
$$Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in})'$$
when the responses are discrete.
- Thus GEE is a popular alternative to maximum likelihood (ML) estimation.
- Liang and Zeger (1986) proposed GEE a method based on the concept of 'estimating equations'.
- GEE provides a general approach for analyzing discrete and continuous responses with marginal models.

Overview of marginal models

- The vector of observations from subject i are
$$Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in})'$$
for $(i = 1, \ldots N)$.
- We consider a marginal model that has
  - A mean function, a function of predictors,
  - A known variance function: discrete data the mean determines the variance
  - Possibly a scale parameter for overdispersion.
  - A *working* correlation matrix.

Marginal Model Details

More precisely:

- The marginal mean of the response is
$$\mathsf{E}[Y_{ij}] = \mu_{ij},$$
- The mean depends on explanatory variables X_{ij}
- Through a known link function
$$g(\mu_{ij}) = \eta_{ij} = X_{ij}'\beta.$$

Marginal Models Details II

- The marginal variance of Y_{ij} depends on the marginal mean according to
$$\text{Var}[Y_{ij}] = v(\mu_{ij})\phi$$
- Where the *variance function* $v(\mu_{ij})$ is known
- $\phi = 1$ is fixed for Poisson, Bernoulli.
- Parameter ϕ may have to be estimated (for normal models or overdispersed data).
- Correlation between Y_{ij} and Y_{ik} is a function of extra parameters, α.
- Correlation may depend on means μ_{ij} and μ_{ik}.

Marginal Models

- The essential idea is to generalize the usual univariate likelihood equations by introducing the covariance matrix of the vector of responses, Y_i.
- For linear models, weighted least squares (WLS) or generalized least squares (GLS) can be considered a special case of this 'estimating equations' approach.
- For non-linear models, this approach is called 'generalized estimating equations' or GEE.

Computation Overview

- We haven’t discussed computational algorithms much.
- A computational interlude now.
- Somewhat more complex material.
- If difficult on first review, not to worry, it is not be primary material.
- Do your best; however, this is quite useful material.
- It will be good to hear this for a first time.

Computation Overview II

- Computation can take several hearings to understand.
- The purpose is to understand computation in our models by analogy with simpler models and algorithms.
- Thus I will be talking you through the equations and pointing out some key ideas.

Likelihood Equations, Normal Regression

Consider normal linear regression

$$y_i = x_i'\beta + \epsilon_i,$$
$$\epsilon_i \sim N(0, \sigma^2).$$

The normal density for N observations looks like $(2 * \pi * \sigma^2)^{-N/2}$ times the exponential of

$$\sum_{i=1}^{N} -.5(y_i - x_i'\beta)\sigma^{-2}(y_i - x_i'\beta)$$

This is also the *log likelihood* function.

Likelihood Equations, Normal Regression

We find the *maximum likelihood* estimates by finding the value $\hat{\beta}$ of β that maximizes the likelihood.

$$\ell(\beta) = \sum_{i=1}^{N} -.5(y_i - x_i'\beta)\sigma^{-2}(y_i - x_i'\beta)$$

Maximum of a function $\ell(\beta)$ is found by *differentiating* the function $\ell(\beta)$ with respect to β, setting this equal to zero (which we will do) and then solving (which we mostly won't do).

$$\frac{d\ell(\beta)}{d\beta} = 0$$

Regression Likelihood Equations

The log likelihood looks like

$$\sum_{i=1}^{N} -.5(y_i - x_i'\beta)\sigma^{-2}(y_i - x_i'\beta)$$

Differentiating with respect to β and setting equal to zero gives

$$\sum_{i=1}^{N} x_i'\sigma^{-2}(y_i - x_i'\beta) = 0$$

This is the *likelihood equation*. Solving for β gives

$$\hat{\beta} = \left(\sum_i x_i x_i'\right)^{-1}\left(\sum_i x_i y_i\right).$$

Normal Longitudinal Model

- Generalize to multivariate regression, i.e. longitudinal models.
- Suppose continuous data Y_i, $n \times 1$
- Predictor matrix X_i, n rows by p columns.
- Coefficient vector β, p by 1.

Normal Longitudinal Model

- Model is

$$Y_i = X_i\beta + \epsilon_i$$

- The residual distribution is

$$\epsilon \sim N(0, V_i),$$

- Variance matrix is V_i to distinguish from summation Σ.
- Covariance matrix V_i depends on unknown parameters θ.

Multivariate Likelihood

- The multivariate normal log likelihood is

$$\sum_{i=1}^{N} -.5(Y_i - X_i\beta)' V_i^{-1}(Y_i - X_i\beta)$$

- Looks like a the normal regression likelihood earlier.
- Vector Y_i has replaced scalar y_i;
- Vector $X_i\beta$ has replaced scalar $x_i'\beta$; and
- Matrix V_i^{-1} has replaced scalar σ^{-2}.

Multivariate Observation Likelihood

- Differentiate the log likelihood and set to zero.
- Gives the likelihood equation

$$\sum_{i=1}^{N} X_i' V_i^{-1} (Y_i - X_i \beta) = 0$$

- Similar to linear regression likelihood equation.
- The solution, if we know the V_i is

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' V_i^{-1} Y_i \right)$$

- The *weighted least squares estimator* for $\hat{\beta}$.

Interpreting the Likelihood Equations

$$\sum_{i=1}^{N} X_i' V_i^{-1}(Y_i - X_i\beta) = 0$$

- The residual $(Y_i - X_i\beta)$ is $(Y_i - \mu_i)$, the observation vector Y_i minus its mean $\mu_i = X_i\beta$.
- The V_i^{-1} is the inverse of the covariance matrix.
- The X_i can be thought of as the derivative of μ_i with respect to the regression parameters β, written $d\mu_i/d\beta$.

Multivariate Generalized Linear Models Equations

- Generalize to generalized linear models (GLMs) with correlated data.
- No longer a likelihood equation, hence estimating equation.
- For generalized linear models, hence generalized estimating equations.
- Rather a weighted least squares equation.
- We replace each term in the previous likelihood equation with a generalized linear model equivalent, and include something to adjust for correlations.

MV GLM Estimating Equations

- Data vector Y_i remains as Y_i.
- Mean vector is μ_i rather than $X_i\alpha$.
- Link function and covariates specified later.
- We replace X_i by the derivative
$$D_i = d\mu_i/d\beta.$$
- For GLMs D_i has components that look like $D_1^{-1}X_i$.
- Matrix D_1 is diagonal and has elements $v(\mu_{ij})$ the variance functions of the means.
- Thus D_i is a weighted predictor matrix.
- The overdispersion parameter ϕ is not involved at this stage.

MV GLM Estimating Equations II

- The V_i's are constructed to be similar to the covariance matrix Var$[Y_i]$, but not actually equal.
- The Generalized Estimating Equations are

$$\sum_{i=1}^{N} D_i V_i^{-1}(Y_i - \mu_i(\beta)) = 0$$

- Solve for β giving the GEE estimate we write as

$$\hat{\beta}_{\text{GEE}}.$$

Constructing the Working Covariance Matrix

- We construct a *working* covariance matrix $V_i \approx \text{Var}[Y_i]$.
- Separate V_i into two parts, variance matrix and correlation matrix.
- The variance matrix is a diagonal matrix A_i of variances.
- The diagonal elements of A_i are $\phi v(\mu_{ij})$ are the variances of the observation Y_{ij}.

Constructing the Working Covariance Matrix II

- The correlation matrix $\mathrm{Corr}(Y_i)$ is a function of unknown parameters α.
- The matrix V_i is known as a *working* covariance matrix.
- It is *not* the true underlying correlation matrix $\mathrm{Corr}[Y_i]$.
- We hope V_i is close to Var$[Y_i]$, but do not assume it is exactly correct.

Working Correlation Matrix

- We specify the working correlation matrix $\mathrm{Corr}(Y_i)$ in SAS Proc Genmod as part of our model.
- It is akin but not identical to using the repeated statement in mixed.
- Options are similar to proc mixed repeated statement options valid for balanced equally spaced data: AR, EXCH (or CS), IND, and UN.
- The choice will affect the standard errors (SEs).
- A bad choice will generally inflate SEs, a better choice should reduce the SEs.

Working Correlation Matrix Choice

- UN, Unstructured, is theoretically appropriate, but too many unknown parameters.
- AR, Autoregressive, is for when observations close together in time are more highly correlated than observations far apart in time.
- EXCH or CS, Compound Symmetry: when correlations are constant, no matter how far apart in time.
- IND, Independent, is unlikely to be correct for longitudinal data.

Example: UN didn’t run, other 3 quite similar for CLEAR analysis of count of HIV-/unknown partners.

Variances of the GEE Estimates

- The working correlation is a postulated covariance matrix for $Y_i - \mu_i$.
- Not the final or true or modeled covariance matrix.
- Rather, a *working correlation matrix* used to create estimates $\hat{\beta}_{\text{GEE}}$.

Variances of the GEE Estimates

- *If the model for the data is correct*, then the covariance matrix of the estimates is

$$B^{-1} = \text{Var}\left[\hat{\beta}_{\text{GEE}}\right] = \left(\sum_{i=1}^{N} D_i' V_i^{-1} D_i\right)^{-1}$$

- This is the same as the proc mixed covariance model

$$\left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i\right)^{-1}$$

  but with $D_i = d\mu_i / d\beta$ replacing X_i.

However, . . .

Sandwich Estimator

- Assume working correlation matrix incorrect. Modify covariance matrix B^{-1}
$$\text{Var}[\hat{\beta}_{\text{GEE}}] = B^{-1} M B^{-1}.$$
- Increases the variance estimate!
- Inflates $\text{Var}[\hat{\beta}_{\text{GEE}}]$ to pay for errors in $\mathrm{Corr}[Y_i]$.
- The B^{-1} is the bread; M is the meat in the $B^{-1} M B^{-1}$ sandwich.
- Hence the nickname 'sandwich estimator'.

The Middle

$$M = \sum_{i=1}^{N} D_i' V_i^{-1} \text{Cov}[Y_i] V_i^{-1} D_i$$

- We estimate $D_i = d\mu_i/d\beta$ (the X_i but for GLMs),
- And V_i is the working correlation matrix times the variance function.
- Because we don't know $\text{Cov}[Y_i]$, we replace it with

$$(Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)'$$

- Which has the right expectation.
- $B^{-1}MB^{-1}$ aka *Huber sandwich estimator* or *robust standard errors*.

The Good News

- The estimated regression coefficient $\hat{\beta}_{\text{GEE}}$ is asymptotically correct if the underlying regression mean model is correct.
- Even if assumed correlation model is incorrect!
- Valid standard errors are attained with the sandwich covariance estimator.
- Easy to use methodology.
- Software runs well.
- Produces population estimates directly.

GEE: The Bad News

- The assumptions can correspond to a mathematically impossible model.
- That is, some combinations of correlations and variances are mathematically impossible, no matter how the data is generated.
- Sandwich estimator inflates standard errors
- Inflated SE’s are a serious cost of not modeling the covariance correctly.
- Does not fully specify a statistical model, and covers instead with asymptotic statements.
- Full assumptions are hidden.

GEE: The News

- Does not allow or make individual subject predictions.
- GEE methodology most suited to balanced longitudinal designs where
  - Number of observations N is large
  - Number n of repeated measures is small.
- Not good for highly unbalanced data sets.