Marginal Models and GEE

Biostatistics 653

Applied Statistics III: Longitudinal Analysis

Marginal Models

- Marginal models specify the mean and covariance models separately. We will assume that the marginal distribution for Y_{ij} is given by

$$P\left(Y_{ij}\middle|\theta_{ij}\right) = \exp\{\frac{Y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} - c(Y_{ij}, \phi)\}$$

so that each Y_{ij} has a distribution from the exponential family. We model the marginal expectation of each response, $\mathrm{E}(Y_{ij})$, as a function of covariates of interest.

Marginal Models

We will assume the following

- The marginal expectation of the response, $E(Y_{ij}) = \mu_{ij}$, depends on covariates X_{ij} through the known link function
$$g(\mu_{ij}) = \eta_{ij} = X_{ij}\beta$$
- The marginal variance of Y_{ij} depends on the marginal mean as
$$Var(Y_{ij}) = \nu(\mu_{ij})\phi$$

where $\nu(\mu_{ij})$ is a known variance function and ϕ is a scale parameter we may need to estimate.

- $Cov(Y_{ij}, Y_{ik})$ is a function of the means and perhaps additional parameters that may also need to be estimated.

Examples: Continuous Response

- The marginal mean (a linear regression):

$$\mu_{ij} = \eta_{ij} = X_{ij}\beta$$

- The marginal variance (homogeneous variance):

$$Var(Y_{ij}) = \phi$$

- The covariance function (autoregressive correlation):

$$Cov(Y_{ij}, Y_{ik}) = \alpha^{|k-j|}, 0 \leq \alpha \leq 1$$

Examples: Binary Response

- The marginal mean (logistic regression):

$$logit(\mu_{ij}) = \eta_{ij} = X_{ij}\beta$$

- The marginal variance (based on Bernoulli):

$$Var(Y_{ij}) = \mu_{ij}(1 - \mu_{ij})$$

- The covariance function (unstructured correlation):

$$Cov(Y_{ij}, Y_{ik}) = \alpha_{jk}$$

Examples: Count Response

- The marginal mean (Poisson regression):

$$log(\mu_{ij}) = \eta_{ij} = X_{ij}\beta$$

- The marginal variance (overdispersion or extra-Poisson variance):

$$Var(Y_{ij}) = \mu_{ij}\phi$$

- The covariance function (compound symmetry correlation):

$$Cov(Y_{ij}, Y_{ik}) = \alpha$$

Marginal Models

- In marginal models, parameters have population-averaged interpretations. It is important to note that the type or size of the correlation does not affect the interpretation of β. The regression parameters β describe the effect of covariates on the average responses (or marginal expectations).

Inference for Marginal Models: Maximum Likelihood

- In linear models for normal data, specifying the means and covariance fully determines the likelihood; however, this is not the case for discrete response data. In the absence of such a convenient likelihood function, there is no single unified likelihood-based approach for marginal models. In order to specify the likelihood for multivariate discrete data, additional assumptions about higher-order moments must be made.

- When $Y_i = (Y_{i1}, \cdots, Y_{in})^T$ is binary, the joint distribution is multinomial with 2^n probabilities, and the number of parameters grows very quickly as n increases. This becomes very difficult to handle very quickly.

Inference for Marginal Models: Maximum Likelihood

- Another approach is to relate the joint distribution of Y_i to that of some underlying latent continuous response. Consider a vector of unobserved continuous variables $Z_i = (Z_{i1}, \cdots, Z_{in})^T$ with joint cdf $H_i(.)$. Define Y_{ij} as

$$Y_{ij} = 1 \; when \; Z_{ij} \leq X_{ij}\beta$$

where X_{ij} represents some threshold. The dependencies among the Z_{ij} induce dependencies among the Y_{ij}.

- For example

$$\begin{aligned} &P(Y_{i1} = 1, Y_{i2} = 1, \cdots, Y_{in} = 1) \\ &= P(Z_{i1} \leq X_{i1}\beta, Z_{i2} \leq X_{i2}\beta, \cdots, Z_{in} \leq X_{in}\beta) \\ &= H_i(X_{i1}\beta, X_{i2}\beta, \cdots, X_{in}\beta) \end{aligned}$$

- Similarly

$P(Y_{i1} = 0, Y_{i2} = 1, \cdots, Y_{in} = 1) = P(Z_{i1} > X_{i1}\beta, Z_{i2} \leq X_{i2}\beta, \cdots, Z_{in} \leq X_{in}\beta) = H_i(\infty, X_{i2}\beta, \cdots, X_{in}\beta)$- $H_i(X_{i1}\beta, X_{i2}\beta, \cdots, X_{in}\beta)$

Inference for Marginal Models: Maximum Likelihood

- The choice of H determines the form of marginal distribution obtained. Because there is no single convenient approach to maximum likelihood with binary data, generalized estimating equations have soared in popularity.

Inference for Marginal Models: Generalized Estimating Equations (GEE)

- Liang and Zeger (1986) proposed an alternative to maximum likelihood based on the concept of estimating equations. Their approach is a single general and unified method for analyzing discrete (or continuous) responses using marginal models.
- Their idea was to generalize the usual univariate likelihood equations by adding the covariance matrix of the response vector Y_i. For linear models, generalized/weighted least squares is a special case of their proposed approach (a nice property).

Inference for Marginal Models: Generalized Estimating Equations (GEE)

We assume the following marginal model for $Y_i = (Y_{i1}, \cdots, Y_{in})^T$:

- The marginal expectation of the response, $E\left(Y_{ij}\right) = \mu_{ij}$, depends on covariates X_{ij} through the known link function

$$g\left(\mu_{ij}\right) = \eta_{ij} = X_{ij}\beta$$

- The marginal variance of Y_{ij} depends on the marginal mean as

$$Var\left(Y_{ij}\right) = \nu\left(\mu_{ij}\right)\phi$$

where $\nu\left(\mu_{ij}\right)$ is a known variance function and ϕ is a scale parameter we may need to estimate.

- $Cov(Y_{ij}, Y_{ik})$ is a function of the means and perhaps additional parameters that may also need to be estimated.
- We note that in general, we have not specified the full joint distribution of the data through these assumptions.

Inference for Marginal Models: Generalized Estimating Equations (GEE)

- Based on this model, we define the working covariance matrix

$$V_i = \phi A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}}$$

where $A_i = diag(\nu(\mu_{i1}), \nu(\mu_{i2}), \cdots, \nu(\mu_{in}))$, and $R(\alpha)$ is a working correlation matrix indexed by α.

Inference for Marginal Models: Generalized Estimating Equations (GEE)

- Recall that the generalized/weighted least squares estimate of for the linear model minimizes the objective function

$$\sum_{i=1}^{N} (Y_i - X_i\beta)^T \Sigma_i^{-1} (Y_i - X_i\beta)$$

- It can be shown that if a minimum exists, it must solve

$$\sum_{i=1}^{N} X_i^T \Sigma_i^{-1} (Y_i - \mu_i) = 0$$

where $\mu_i = \mu_i(\beta) = X_i\beta$

Inference for Marginal Models: Generalized Estimating Equations (GEE)

- The GEE estimator of β can be thought of as arising from minimizing the objective function

$$\sum_{i=1}^{N} \left(Y_i - \mu_i(\beta)\right)^T V_i^{-1} (Y_i - \mu_i(\beta))$$

where V_i is treated as known, and μ_i is a vector of mean responses with elements $\mu_{ij} = g^{-1}(X_{ij}^T \beta)$.

- It can be shown that if a minimum exists, it must solve

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

where $D_i = \frac{\partial \mu_i}{\partial \beta} = \begin{pmatrix} \frac{\partial \mu_{i1}}{\partial \beta_1} & \frac{\partial \mu_{i1}}{\partial \beta_2} & \cdots & \frac{\partial \mu_{i1}}{\partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mu_{in}}{\partial \beta_1} & \frac{\partial \mu_{in}}{\partial \beta_2} & \cdots & \frac{\partial \mu_{in}}{\partial \beta_p} \end{pmatrix}$

Inference for Marginal Models: Generalized Estimating Equations (GEE)

- Then, we estimate β as the solution to the generalized estimating equations (GEE)

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

- Because the generalized estimating equations depend both on β and α, ϕ, we use a two-stage estimation procedure:

1. Given current estimates of α, ϕ, we obtain an estimate of β as the solution to the generalized estimating equations.

2. Given a current estimate of β, we estimate α, ϕ based on the standardized residuals

$$r_{ij} = \frac{(Y_{ij} - \hat{\mu}_{ij})}{\nu(\hat{\mu}_{ij})^{\frac{1}{2}}}$$

GEE: More Details on the 2nd Stage

- When we use the working independence assumption, we assume that $R(\alpha) = I$, so that we need not estimate any parameters α.
- However, more generally, we need to estimate α. Method of moments is often also used to estimate α given a fixed $\hat{\beta}^{(t)}$ at iteration t of the iterative algorithm.
- First, consider the balanced case so that $n_i = n$, and no data are missing, with unstructured $R(\alpha)$, so that α has $n(n-1)/2$ parameters (diagonal 1)
- Because $V_i = \phi A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}}$, and the estimates $\hat{\alpha}_{ij}$, the j, jth elements of $\hat{A}_i$ are determined by our variance as $A_i = diag(\nu(\mu_{i1}), \nu(\mu_{i2}), \cdots, \nu(\mu_{in}))$. Then it remains for us to estimate R and ϕ.

GEE: More Details on the 2nd Stage

- Often, the residual-based estimate is used:

$$\hat{\phi} = \frac{1}{Nn}\sum_{i=1}^{N} r_i^T r_i$$

where $r_{ij} = \frac{(Y_{ij} - \widehat{\mu}_{ij}(\widehat{\beta}))}{\nu(\widehat{\mu}_{ij})^{\frac{1}{2}}}$ is the standardized residual.

- Then, to estimate α, just note that under the working covariance model

$$E[(Y_i - \mu_i)(Y_i - \mu_i)^T] = V_i$$

so that we may obtain $\widehat{V}_i = \left(Y_i - \mu_i(\hat{\beta})\right)\left(Y_i - \mu_i\hat{\beta}\right)^T$. This relationship gives us an estimate of the unknowns $\hat{\alpha}$.

GEE: More Details on the 2nd Stage

- In particular, we can formalize an ad hoc estimation of α using a set of estimating equations. Define

$$\rho_{ijk}(\alpha) = Cor\left(Y_{ij}, Y_{ik}\right)$$

$$r_{ijk} = \phi \frac{\left(Y_{ij} - \mu_{ij}\right)\left(Y_{ij} - \mu_{ik}\right)^T}{a_{ij}^{\frac{1}{2}} a_{ik}^{\frac{1}{2}}}$$

- We have $E\left(r_{ijk}\right) = \rho_{ijk}$. We denote r_i, ρ_i as vectors of size $\frac{n_i(n_i-1)}{2}$.
- Then, α estimating equations can be given by

$$\sum_{i=1}^{N} \left(\frac{\partial \rho_i}{\partial \alpha}\right)^T [Var(r_i)]^{-1}(r_i - \rho_i) = 0$$

- Because specifying $Var(r_i)$ requires specification of higher-order moments of Y_{ij} , most people focus on a set of "working independence" estimating equations given by

$$\sum_{i=1}^{N} \left(\frac{\partial \rho_i}{\partial \alpha}\right)^T (r_i - \rho_i) = 0$$

GEE Algorithm Summary

- Therefore, GEE consists of two estimating equations solved separately, but iteratively:
- At iteration t+1, given fixed $\hat{\alpha}^{(t)}$, solve for $\hat{\beta}^{(t+1)}$ as the solution to

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

where $D_i = \frac{\partial \mu_i}{\partial \beta}$.

- Given fixed $\hat{\beta}^{(t+1)}$, solve for $\hat{\alpha}^{(t+1)}$ as the solution to

$$\sum_{i=1}^{N} C_i^T(r_i - \rho_i) = 0$$

where $\mathrm{C_i} = \frac{\partial \rho_i}{\partial \alpha}$

- Iterate until convergence.

Inference for Marginal Models: Generalized Estimating Equations (GEE)

Assuming we use consistent estimators of α, ϕ and β, then the solution to the generalized estimating equations has the following properties:

- $\hat{\beta}$ is a consistent estimate of β
- $\hat{\beta}$ is asymptotically multivariate normally distributed
- $Cov(\hat{\beta}) = F^{-1}GF^{-1}$, where

$$F = \sum_{i=1}^{N} D_i^T V_i^{-1} D_i$$

$$G = \sum_{i=1}^{N} D_i^T V_i^{-1} Cov(Y_i) V_i^{-1} D_i$$

- F and G may be estimated by replacing α, ϕ and β by their estimates, and replacing $Cov(Y_i)$ by $(Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)^T$. That is, we can use the empirical or sandwich variance estimator.

Inference for Marginal Models: Generalized Estimating Equations (GEE)

- The model-based variance estimator is given by F^{-1}. If we correctly specify the variance-covariance model for Y_i, we can use the quasi-likelihood estimator $\hat{\beta}$ with the model-based variance estimator in order to conduct inferences. We would expect to gain some power in this situation. However, if the covariance model is reasonable and the sample size is large enough, the empirical and model-based standard errors should not differ greatly. We note that you want to have around N = 100 subjects to use the empirical covariance estimator with binary data.

Inference for Marginal Models: Generalized Estimating Equations (GEE)

Other attractive properties of GEE estimators include the following:

- $\hat{\beta}$ is almost efficient when compared to MLE in many cases. For example, GEE has the same form as the likelihood equations for multivariate normal model.
- $\hat{\beta}$ is consistent even if $Cov(Y_i)$ has been mis-specified.
- Standard errors for $\hat{\beta}$ may be obtained using the empirical or sandwich variance estimator.
- As long as we model the mean correctly, our inferences using the empirical standard errors are robust, even if our knowledge of the full joint distribution is imperfect.
- The closer you get to the right covariance structure, the better your efficiency will be.