Generalized Estimating Equations (GEE) for Correlated Data

Key reference:

Liang and Zeger, 1986, Biometrika. 73. p13-22

Longitudinal Data (Clustered Data):

subject	time	# of obs
1	x x	n_1
2	x x x	n_2
$\vdots$	$\vdots$	$\vdots$
m	x x	n_m

1. Assume m independent subjects (clusters)
2. For the ith of m subjects ($i = 1, \cdots, m$), there are n_i observations over time.

$$\begin{aligned} \mathbf{Y}_i = (Y_{i1}, \cdots, Y_{in_i})^T &\quad - \quad \text{outcome } n_i \times 1 \\ \mathbf{X}_i = (\mathbf{X}_{i1}, \cdots, \mathbf{X}_{in_i})^T &\quad - \quad \text{covariate matrix } n_i \times p \end{aligned}$$

   where Y_{ij} is the outcome and $\mathbf{X}_{ij}$ is a $p \times 1$covariate vector at the jth time point of the ith subject.
3. Note that $\mathbf{X}_{ij}$ may contain both subject-level covariates and time varying covariates.

Example: Indonesia Infectious Disease Data

- 1200 Indonesian children, each was followed for up to 6 consecutive quarters.
- Outcome=respiratory infection (Y/N).
- Covariates=age, sex, xerophthalmia status, etc.

For each subject:

Y=infection	0	1	1
XERO	0	0	1
sex	1	1	1
age	0	3	9

$$\mathbf{Y}_i = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{X}_i = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 3 \\ 1 & 1 & 1 & 9 \end{pmatrix}$$

Problem

We now have a **dichotomous outcome**!

Question:
How can we extend GLMs to model the relationship between $\mathbf{Y}_i$ and $\mathbf{X}_i$ while accounting for the within-subject correlation?

Challenge:
The likelihood of $\mathbf{Y}_i$ is hard to specify, since the joint likelihood of $(Y_{i1}, \cdots, Y_{in_i})$ is hard (not impossible) to specify except for the normal case.

Objective: Modeling the Mean

If one is only interested in modeling the dependence of the MEAN of Y_{ij} on $\mathbf{X}_{ij}$ while treating the correlation as nuisance parameters, how can we make as fewer assumptions as possible and construct consistent and asymptotically normal regression coefficient estimators?

Answer: Construct unbiased estimating equations or generalized estimating equations (GEEs).

GEEs: Distributional Assumptions

1. Joint distribution of $\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{in_i})^T$ is hard (but not impossible).

2. Only specify the **marginal distribution** of Y_{ij} using QL (or exponential family).

$$\begin{aligned} E(Y_{ij}) &= \mu_{ij} \\ var(Y_{ij}) &= \phi a_{ij}^{-1} v(\mu_{ij}) \end{aligned}$$

$$\Rightarrow \ell(Y_{ij}) = \int_{Y_{ij}}^{\mu_{ij}} \frac{Y_{ij} - u}{\phi a_{ij}^{-1} v(u)} du$$

Mean Model: Independent Data

Recall:

$$g(\mu_i) = \mathbf{X}_i^T \beta$$

Quasi - score:

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

where $\mathbf{D}_i = \frac{\partial \mu_i}{\partial \beta^T}$ is $1 \times p$, $V_i = var(Y_i) = \phi a_i^{-1} v(\mu_i)$ is 1×1, and $Y_i - \mu_i$ is 1×1.

Clustered Data: Generalized Estimating Equations (GEEs)

Assumptions:

1. Marginal mean & variance: $E(Y_{ij}) = \mu_{ij}$, $var(Y_{ij}) = \phi a_{ij}^{-1} v(\mu_{ij})$
2. Mean Model: $g(\mu_{ij}) = \mathbf{X}_{ij}^T \beta$

GEEs:

$$\sum_{i=1}^{m} \mathbf{D_i^T V_i^{-1} (Y_i - \boldsymbol{\mu}_i) = 0}$$

where $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^T}$ is $n_i \times p$, $(\mathbf{Y_i} - \boldsymbol{\mu}_\mathbf{i})$ is $n_i \times 1$, and $\mathbf{V}_i$ is an $n_i \times n_i$ working covariance matrix.

Independent vs. Correlated Setting

Forms for the estimating equations are nearly identical.

- Specify mean model:

$$g(\mu_i) = \mathbf{X}_i^T \beta \text{ vs. } g(\mu_{ij}) = \mathbf{X}_{ij}^T \beta$$

$$\mathbf{D}_i = \frac{\partial \mu_i}{\partial \beta^T} \text{ vs. } \mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \beta^T}$$

- Specify distribution OR mean/variance:

$$var(Y_i) = \phi a_i^{-1} v(\mu_i) \text{ vs. } var(Y_{ij}) = \phi a_{ij}^{-1} v(\mu_{ij})$$

(Quasi-)Score for Independent Data

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

GEE for Correlated Data

$$\sum_{i=1}^{m} \mathbf{D_i^T} \mathbf{V_i^{-1}} (\mathbf{Y_i} - \boldsymbol{\mu}_\mathbf{i}) = \mathbf{0}$$

Working correlation matrix:

Since $\mathbf{V}_i$ is a matrix instead of a scalar, we need to specify non-diagonal elements (diag. are determined by mean/variance model)

$$\mathbf{V}_i = \mathbf{V_{M_i}}^{\frac{1}{2}} \mathbf{R}_i(\alpha) \mathbf{V_M}_i^{\frac{1}{2}},$$

where $\mathbf{V_M}_i = diag\{\phi a_{ij}^{-1} v(\mu_{ij})\}$ is the marginal variance of $\mathbf{Y}_i$, $\mathbf{R_i}(\alpha)$ is a working correlation matrix, and α is a working correlation parameter, which is a nuisance parameter.

Key Results

1. $\hat{\beta}$ is consistent and asymptotically normal given the mean model $g(\mu_{ij}) = \mathbf{X}_{ij}^T\beta$ is correctly specified even when the correlation matrix $\mathbf{R_i(\alpha)}$ is misspecified.

2. If the working correlation $\mathbf{R_i(\alpha)}$ is correctly specified, $\hat{\beta}$ is efficient within the linear estimating function family.

Formal Asymptotic Results

Conditions:

(1) $\hat{\phi}$ and $\hat{\boldsymbol{\alpha}}$ are $\sqrt{m}$-consistent, e.g., a moment estimator, for some ϕ_* and $\boldsymbol{\alpha}_*$

(2) $\frac{\partial \mathbf{U}}{\partial \boldsymbol{\beta}^T} \xrightarrow{\mathcal{P}} \mathbf{A}$ uniformly in an open neighborhood of $\boldsymbol{\beta}$

If (1) and (2) hold, then

(a) $\hat{\boldsymbol{\beta}}$ is consistent.

(b) $\sqrt{m}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically normal with mean 0 and covariance $\boldsymbol{\Sigma}$

Variance

$$
\begin{aligned}
\mathbf{\Sigma} &= lim_{m\to\infty}\left(\frac{1}{m}\sum_{i=1}^{m}\mathbf{D_i^T V_i^{-1} D_i}\right)^{-1} \\
&\times\left\{\frac{1}{m}\sum_{i=1}^{m}\mathbf{D_i^T V_i^{-1}}(\mathbf{Y_i}-\boldsymbol{\mu_i})(\mathbf{Y_i}-\boldsymbol{\mu_i})^{\mathbf{T}}\mathbf{V_i^{-1} D_i}\right\} \\
&\times\left(\frac{1}{m}\sum_{i=1}^{m}\mathbf{D_i^T V_i^{-1} D_i}\right)^{-1}
\end{aligned}
$$

$\widehat{\mathbf{\Sigma}}$ which is consistent for $\mathbf{\Sigma}$ can be obtained by plugging in consistent estimates for parameters.

Variance - One Interpretation

$$\begin{aligned} var(\widehat{\beta}) &= \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} \\ \mathbf{A} &= \mathbf{D_i^T V_i^{-1} D_i} \\ \mathbf{B} &= \mathbf{D_i^T V_i^{-1}}(\mathbf{Y_i} - \boldsymbol{\mu_i})(\mathbf{Y_i} - \boldsymbol{\mu_i})^{\mathbf{T}}\mathbf{V_i^{-1} D_i} \\ &= \mathbf{D_i^T V_i^{-1}} var(\mathbf{Y}_i)\mathbf{V_i^{-1} D_i} \end{aligned}$$

Meat: **B** indicates stability of individual contributions to EE; bigger $\rightarrow$ less information

Bread: **A** tells us how contributions distinguish the true β from other values; bigger $\rightarrow$ more information

When $\mathbf{V}_i$ is Correct

Corollary: If $\mathbf{V}_i = \mathbf{V_M}_i^{\frac{1}{2}} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{V_M}_i^{\frac{1}{2}}$ is correctly specified, i.e., the working correlation matrix $\mathbf{R}_i(\boldsymbol{\alpha})$ is correctly specified, writing $\boldsymbol{\Sigma} = \frac{\mathbf{1}}{\boldsymbol{m}} \boldsymbol{A}^{-\mathbf{1}} \boldsymbol{B} \boldsymbol{A}^{-\mathbf{1}}$, then

$$E[\mathbf{B}] = \mathbf{A}$$

and

$$\boldsymbol{\Sigma} = lim_{m \to \infty} \mathbf{A}^{-1} = lim_{m \to \infty} \left(\frac{1}{m} \sum_{i=1}^{m} \mathbf{D_i^T} \mathbf{V_i^{-1}} \mathbf{D_i}\right)^{-1}.$$

i.e. $\boldsymbol{\Sigma}$ is the variance from the full likelihood and $\widehat{\beta}$ is efficient within the linear estimating function family.

Fisher Scoring for Estimating β

$$\mathbf{U}(\hat{\beta}) = \sum_{i=1}^{m} \mathbf{D_i}(\widehat{\beta})^{\mathbf{T}} \mathbf{V_i}(\widehat{\beta}, \widehat{\alpha})^{-1} (\mathbf{Y_i} - \boldsymbol{\mu}_{\mathbf{i}}(\widehat{\beta})) = \mathbf{0}$$

Procedure:

1. Initialize $\widehat{\beta}^{(0)}$ to some value, often from GLM
2. Calculate $\widehat{\alpha}^{(k)}$ (and $\widehat{\phi}^{(k)}$) using moment-based formulas and residuals from $\widehat{\beta}^{(k)}$.
3. Get update $\widehat{\beta}^{(k+1)}$ via Fisher scoring:

$$\widehat{\beta}^{(k+1)} = \widehat{\beta}^{(k)} + \left[\sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i(\widehat{\alpha}^{(k)})^{-1} \mathbf{D}_i \right]^{-1} \left[\sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i(\widehat{\alpha}^{(k)})^{-1} (\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i^{(k)}) \right]$$

4. Repeat steps 2 and 3 til convergence.

Working Correlation Choices

Independence

$$\rho(\alpha) = 0; \quad \mathbf{R}_i(\boldsymbol{\alpha}) = \mathbf{I}$$

Essentially fit usual GLM to the data, but then correct the naive SEs of $\widehat{\beta}$ using sandwich estimators

Exchangeable

$$\rho(\alpha) = \alpha; \quad \mathbf{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & & \vdots \\ \vdots & & \ddots & \alpha \\ \alpha & \cdots & \alpha & 1 \end{bmatrix}$$

Note that CS structure is similar but places additional constraints on the variance.

Working Correlation Choices (2)

Autoregressive (AR-1)

$$\rho(\alpha)_{j,k} = \alpha^{|j-k|}; \quad \mathbf{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{n-2} \\ \alpha^2 & \alpha & 1 & \cdots & \alpha^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{n-1} & \alpha^{n-2} & \alpha^{n-3} & \cdots & 1 \end{bmatrix}$$

Toeplitz/Banded

$$\mathbf{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_{n-1} \\ \alpha_1 & 1 & \alpha_1 & \alpha_2 & \cdots & \alpha_{n-2} \\ \alpha_2 & \alpha_1 & 1 & \alpha_1 & \cdots & \alpha_{n-3} \\ \alpha_3 & \alpha_2 & \alpha_1 & 1 & \cdots & \alpha_{n-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{n-1} & \alpha_{n-2} & \alpha_{n-3} & \alpha_{n-4} & \cdots & 1 \end{bmatrix}$$

Working Correlation Choices (3)

Unstructured

$$\rho(\alpha)_{j,k} = \alpha_{j,k}; \quad \mathbf{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1,n-1} \\ \alpha_{21} & 1 & \alpha_{23} & \cdots & \alpha_{2,n-2} \\ \alpha_{31} & \alpha_{32} & 1 & \cdots & \alpha_{3,n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{n-1,1} & \alpha_{n-1,2} & \alpha_{n-1,3} & \cdots & 1 \end{bmatrix}$$