# Statistical Methods for Analysis with Missing Data

## Lecture 13: intro to (weighted generalized) estimating equations

Mauricio Sadinle

Department of Biostatistics
UNIVERSITY of WASHINGTON

# Previous Lecture

- Inverse-probability weighting (IPW)
  - Origins in survey sampling (Horvitz-Thompson estimator)
  - Does not require modeling of the full-data distribution
  - Sensitive to misspecification of the propensity score model and to extreme weights
- Augmented IPW
  - Enjoys *double-robustness* property
  - However "*in at least some settings, two wrong models are not better than one*" (Kang and Schafer, 2007)
- We focused on estimation of a mean

# Today's Lecture

Introduction to:

- Estimating equations
- Generalized estimating equations
- Weighted generalized estimating equations

# Outline

# Estimating Equations

- Consider i.i.d. data $\mathbf{Z} = \{Z_i\}_{i=1}^n$
- An *estimating function* $M(Z;\theta)$ is a continuously differentiable function of $Z$ and parameters $\theta$ that satisfies

$$E_{Z|\theta}[M(Z;\theta)] = \mathbf{0}$$

- Given an estimating function, the *estimating equations* are given by

$$\frac{1}{n}\sum_{i=1}^{n} M(Z_i;\theta) = \mathbf{0}$$

- If

$$E_{\mathbf{Z}|\theta}\left[\frac{1}{n}\sum_{i=1}^{n} M(Z_i;\theta)\right] = \mathbf{0},$$

the estimating equation is called *unbiased*

# Estimating Equations

- Consider i.i.d. data $\mathbf{Z} = \{Z_i\}_{i=1}^n$
- An *estimating function* $M(Z;\theta)$ is a continuously differentiable function of $Z$ and parameters $\theta$ that satisfies

$$E_{Z|\theta}[M(Z;\theta)] = \mathbf{0}$$

- Given an estimating function, the *estimating equations* are given by

$$\frac{1}{n}\sum_{i=1}^{n} M(Z_i;\theta) = \mathbf{0}$$

- If

$$E_{\mathbf{Z}|\theta}\left[\frac{1}{n}\sum_{i=1}^{n} M(Z_i;\theta)\right] = \mathbf{0},$$

the estimating equation is called *unbiased*

# Estimating Equations

- Consider i.i.d. data $\mathbf{Z} = \{Z_i\}_{i=1}^n$
- An *estimating function* $M(Z;\theta)$ is a continuously differentiable function of $Z$ and parameters $\theta$ that satisfies

$$E_{Z|\theta}[M(Z;\theta)] = \mathbf{0}$$

- Given an estimating function, the *estimating equations* are given by

$$\frac{1}{n}\sum_{i=1}^{n} M(Z_i;\theta) = \mathbf{0}$$

- If

$$E_{\mathbf{Z}|\theta}\left[\frac{1}{n}\sum_{i=1}^{n} M(Z_i;\theta)\right] = \mathbf{0},$$

the estimating equation is called *unbiased*

# Estimating Equations

- Consider i.i.d. data $\mathbf{Z} = \{Z_i\}_{i=1}^n$
- An *estimating function* $M(Z; \theta)$ is a continuously differentiable function of $Z$ and parameters $\theta$ that satisfies

$$E_{Z|\theta}[M(Z; \theta)] = \mathbf{0}$$

- Given an estimating function, the *estimating equations* are given by

$$\frac{1}{n} \sum_{i=1}^{n} M(Z_i; \theta) = \mathbf{0}$$

- If

$$E_{\mathbf{Z}|\theta} \left[ \frac{1}{n} \sum_{i=1}^{n} M(Z_i; \theta) \right] = \mathbf{0},$$

the estimating equation is called *unbiased*

# $M$-Estimators

- The solution $\hat{\theta}$ to the $p \times 1$ estimating equations

$$\frac{1}{n}\sum_{i=1}^{n} M(Z_i;\theta) = \mathbf{0}$$

  is referred to as an *M-estimator*

# Examples

- Say we are interested in estimating a mean $\mu = E(Y)$
- Take the estimating function as $M(Y;\mu) = Y - \mu$
- The sample mean is the solution to the unbiased estimating equation

$$\sum_{i=1}^{n}(Y_i - \mu) = 0$$

# Examples

- Say we are interested in a regression model

$$E(Y \mid x) = \mu(x; \beta)$$

- With no further assumptions, this is a *semiparametric model*
- With full data $\{(Y_i, X_i)\}_{i=1}^n$, estimation of $\beta$ is done by solving the *least squares estimating equation*

$$\sum_{i=1}^{n} \frac{\partial}{\partial \beta}[\mu(X_i; \beta)][Y_i - \mu(X_i; \beta)] = \mathbf{0}$$

- This is an estimating equation with estimating function

$$\frac{\partial}{\partial \beta}[\mu(X; \beta)][Y - \mu(X; \beta)]$$

- The estimating equation is unbiased if the regression model is correctly specified
- Under the additional assumption of $Y \mid x$ being normal, this corresponds to MLE

# Examples

- Say we are interested in a regression model

$$E(Y \mid x) = \mu(x;\beta)$$

- With no further assumptions, this is a *semiparametric model*
- With full data $\{(Y_i, X_i)\}_{i=1}^n$, estimation of $\beta$ is done by solving the *least squares estimating equation*

$$\sum_{i=1}^{n} \frac{\partial}{\partial \beta}[\mu(X_i;\beta)][Y_i - \mu(X_i;\beta)] = \mathbf{0}$$

- This is an estimating equation with estimating function

$$\frac{\partial}{\partial \beta}[\mu(X;\beta)][Y - \mu(X;\beta)]$$

- The estimating equation is unbiased if the regression model is correctly specified
- Under the additional assumption of $Y \mid x$ being normal, this corresponds to MLE

# Examples

- Say we are interested in a regression model

$$E(Y \mid x) = \mu(x; \beta)$$

- With no further assumptions, this is a *semiparametric model*
- With full data $\{(Y_i, X_i)\}_{i=1}^n$, estimation of $\beta$ is done by solving the *least squares estimating equation*

$$\sum_{i=1}^{n} \frac{\partial}{\partial \beta}[\mu(X_i; \beta)][Y_i - \mu(X_i; \beta)] = \mathbf{0}$$

- This is an estimating equation with estimating function

$$\frac{\partial}{\partial \beta}[\mu(X; \beta)][Y - \mu(X; \beta)]$$

- The estimating equation is unbiased if the regression model is correctly specified
- Under the additional assumption of $Y \mid x$ being normal, this corresponds to MLE

# Examples

- Say we are interested in a regression model

$$E(Y \mid x) = \mu(x; \beta)$$

- With no further assumptions, this is a *semiparametric model*
- With full data $\{(Y_i, X_i)\}_{i=1}^n$, estimation of $\beta$ is done by solving the *least squares estimating equation*

$$\sum_{i=1}^{n} \frac{\partial}{\partial \beta}[\mu(X_i; \beta)][Y_i - \mu(X_i; \beta)] = \mathbf{0}$$

- This is an estimating equation with estimating function

$$\frac{\partial}{\partial \beta}[\mu(X; \beta)][Y - \mu(X; \beta)]$$

- The estimating equation is unbiased if the regression model is correctly specified
- Under the additional assumption of $Y \mid x$ being normal, this corresponds to MLE

# Examples

- Say we are interested in a regression model

$$E(Y \mid x) = \mu(x; \beta)$$

- With no further assumptions, this is a *semiparametric model*
- With full data $\{(Y_i, X_i)\}_{i=1}^n$, estimation of $\beta$ is done by solving the *least squares estimating equation*

$$\sum_{i=1}^{n} \frac{\partial}{\partial \beta}[\mu(X_i; \beta)][Y_i - \mu(X_i; \beta)] = \mathbf{0}$$

- This is an estimating equation with estimating function

$$\frac{\partial}{\partial \beta}[\mu(X; \beta)][Y - \mu(X; \beta)]$$

- The estimating equation is unbiased if the regression model is correctly specified
- Under the additional assumption of $Y \mid x$ being normal, this corresponds to MLE

# Examples

- Say we are interested in a regression model
$$E(Y \mid x) = \mu(x; \beta)$$
- With no further assumptions, this is a *semiparametric model*
- With full data $\{(Y_i, X_i)\}_{i=1}^n$, estimation of $\beta$ is done by solving the *least squares estimating equation*
$$\sum_{i=1}^{n} \frac{\partial}{\partial \beta}[\mu(X_i; \beta)][Y_i - \mu(X_i; \beta)] = \mathbf{0}$$
- This is an estimating equation with estimating function
$$\frac{\partial}{\partial \beta}[\mu(X; \beta)][Y - \mu(X; \beta)]$$
- The estimating equation is unbiased if the regression model is correctly specified
- Under the additional assumption of $Y \mid x$ being normal, this corresponds to MLE

# Examples

- Say we have a parametric model for the study variables $p(z \mid \theta)$
- Taking
$$M(Z;\theta) = \frac{\partial}{\partial\theta} \log p(Z \mid \theta),$$
leads to the usual *score equation*
$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log p(Z_i \mid \theta) = \mathbf{0},$$
that can be solved to obtain the MLE of $\theta$
- The estimating (score) equation is unbiased if the model is correctly specified
- MLEs are $M$-estimators

# Examples

- Say we have a parametric model for the study variables $p(z \mid \theta)$
- Taking

$$M(Z;\theta) = \frac{\partial}{\partial\theta} \log p(Z \mid \theta),$$

leads to the usual *score equation*

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log p(Z_i \mid \theta) = \mathbf{0},$$

that can be solved to obtain the MLE of $\theta$

- The estimating (score) equation is unbiased if the model is correctly specified
- MLEs are $M$-estimators

# Examples

- Say we have a parametric model for the study variables $p(z \mid \theta)$
- Taking
$$M(Z;\theta) = \frac{\partial}{\partial\theta}\log p(Z \mid \theta),$$
leads to the usual *score equation*
$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log p(Z_i \mid \theta) = \mathbf{0},$$
that can be solved to obtain the MLE of $\theta$
- The estimating (score) equation is unbiased if the model is correctly specified
- MLEs are $M$-estimators

# Examples

- Say we have a parametric model for the study variables $p(z \mid \theta)$
- Taking
$$M(Z;\theta) = \frac{\partial}{\partial\theta} \log p(Z \mid \theta),$$
leads to the usual *score equation*
$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log p(Z_i \mid \theta) = \mathbf{0},$$
that can be solved to obtain the MLE of $\theta$
- The estimating (score) equation is unbiased if the model is correctly specified
- MLEs are $M$-estimators

# $M$-Estimators

- Heuristically, we say that with a large sample size the approximate distribution of $\hat{\theta}$ is

$$\hat{\theta} \approx \text{Normal}[\theta_0, n^{-1} U_n^{-1} V_n (U_n^{-1})^T]$$

where

$$U_n = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta^T} M(Z_i; \hat{\theta})$$

and

$$V_n = \frac{1}{n} \sum_{i=1}^{n} M(Z_i; \hat{\theta}) M(Z_i; \hat{\theta})^T$$

# Outline

# Generalized Estimating Equations

*Generalized estimating equations* (GEEs) were introduced by Liang and Zeger (Biometrika 1986)

- Data to be collected at $T$ time points
- $n$ i.i.d. individuals
- $Y_{ij}$: outcome of individual $i$ at time $j$
- $Y_i = (Y_{i1}, \ldots, Y_{iT})$
- $X_i$: exogenous vector of covariates for individual $i$
- We are interested in a model

$$E(Y_i \mid X_i) = \begin{bmatrix} E(Y_{i1} \mid X_i) \\ \vdots \\ E(Y_{iT} \mid X_i) \end{bmatrix} = \begin{bmatrix} \mu_1(X_i;\beta) \\ \vdots \\ \mu_T(X_i;\beta) \end{bmatrix}$$

- Without further specification of the distribution of the data, this is a *semiparametric model*

# Generalized Estimating Equations

*Generalized estimating equations* (GEEs) were introduced by Liang and Zeger (Biometrika 1986)

- Data to be collected at $T$ time points
- $n$ i.i.d. individuals
- $Y_{ij}$: outcome of individual $i$ at time $j$
- $Y_i = (Y_{i1}, \ldots, Y_{iT})$
- $X_i$: exogenous vector of covariates for individual $i$
- We are interested in a model

$$E(Y_i \mid X_i) = \begin{bmatrix} E(Y_{i1} \mid X_i) \\ \vdots \\ E(Y_{iT} \mid X_i) \end{bmatrix} = \begin{bmatrix} \mu_1(X_i; \beta) \\ \vdots \\ \mu_T(X_i; \beta) \end{bmatrix}$$

- Without further specification of the distribution of the data, this is a *semiparametric model*

# Generalized Estimating Equations

*Generalized estimating equations* (GEEs) were introduced by Liang and Zeger (Biometrika 1986)

- Data to be collected at $T$ time points
- $n$ i.i.d. individuals
- $Y_{ij}$: outcome of individual $i$ at time $j$
- $Y_i = (Y_{i1}, \ldots, Y_{iT})$
- $X_i$: exogenous vector of covariates for individual $i$
- We are interested in a model

$$E(Y_i \mid X_i) = \begin{bmatrix} E(Y_{i1} \mid X_i) \\ \vdots \\ E(Y_{iT} \mid X_i) \end{bmatrix} = \begin{bmatrix} \mu_1(X_i; \beta) \\ \vdots \\ \mu_T(X_i; \beta) \end{bmatrix}$$

- Without further specification of the distribution of the data, this is a *semiparametric model*

# Generalized Estimating Equations

*Generalized estimating equations* (GEEs) were introduced by Liang and Zeger (Biometrika 1986)

- Data to be collected at $T$ time points
- $n$ i.i.d. individuals
- $Y_{ij}$: outcome of individual $i$ at time $j$
- $Y_i = (Y_{i1}, \ldots, Y_{iT})$
- $X_i$: exogenous vector of covariates for individual $i$
- We are interested in a model

$$E(Y_i \mid X_i) = \begin{bmatrix} E(Y_{i1} \mid X_i) \\ \vdots \\ E(Y_{iT} \mid X_i) \end{bmatrix} = \begin{bmatrix} \mu_1(X_i; \beta) \\ \vdots \\ \mu_T(X_i; \beta) \end{bmatrix}$$

- Without further specification of the distribution of the data, this is a *semiparametric model*

# Generalized Estimating Equations

*Generalized estimating equations* (GEEs) were introduced by Liang and Zeger (Biometrika 1986)

- Data to be collected at $T$ time points
- $n$ i.i.d. individuals
- $Y_{ij}$: outcome of individual $i$ at time $j$
- $Y_i = (Y_{i1}, \dots, Y_{iT})$
- $X_i$: exogenous vector of covariates for individual $i$
- We are interested in a model

$$E(Y_i \mid X_i) = \begin{bmatrix} E(Y_{i1} \mid X_i) \\ \vdots \\ E(Y_{iT} \mid X_i) \end{bmatrix} = \begin{bmatrix} \mu_1(X_i; \beta) \\ \vdots \\ \mu_T(X_i; \beta) \end{bmatrix}$$

- Without further specification of the distribution of the data, this is a *semiparametric model*

# Generalized Estimating Equations

*Generalized estimating equations* (GEEs) were introduced by Liang and Zeger (Biometrika 1986)

- Data to be collected at $T$ time points
- $n$ i.i.d. individuals
- $Y_{ij}$: outcome of individual $i$ at time $j$
- $Y_i = (Y_{i1}, \ldots, Y_{iT})$
- $X_i$: exogenous vector of covariates for individual $i$
- We are interested in a model

$$E(Y_i \mid X_i) = \begin{bmatrix} E(Y_{i1} \mid X_i) \\ \vdots \\ E(Y_{iT} \mid X_i) \end{bmatrix} = \begin{bmatrix} \mu_1(X_i; \beta) \\ \vdots \\ \mu_T(X_i; \beta) \end{bmatrix}$$

- Without further specification of the distribution of the data, this is a *semiparametric model*

# Generalized Estimating Equations

*Generalized estimating equations* (GEEs) were introduced by Liang and Zeger (Biometrika 1986)

- Data to be collected at $T$ time points
- $n$ i.i.d. individuals
- $Y_{ij}$: outcome of individual $i$ at time $j$
- $Y_i = (Y_{i1}, \ldots, Y_{iT})$
- $X_i$: exogenous vector of covariates for individual $i$
- We are interested in a model

$$E(Y_i \mid X_i) = \begin{bmatrix} E(Y_{i1} \mid X_i) \\ \vdots \\ E(Y_{iT} \mid X_i) \end{bmatrix} = \begin{bmatrix} \mu_1(X_i; \beta) \\ \vdots \\ \mu_T(X_i; \beta) \end{bmatrix}$$

- Without further specification of the distribution of the data, this is a *semiparametric model*

# Generalized Estimating Equations

*Generalized estimating equations* (GEEs) were introduced by Liang and Zeger (Biometrika 1986)

- Data to be collected at $T$ time points
- $n$ i.i.d. individuals
- $Y_{ij}$: outcome of individual $i$ at time $j$
- $Y_i = (Y_{i1}, \ldots, Y_{iT})$
- $X_i$: exogenous vector of covariates for individual $i$
- We are interested in a model

$$E(Y_i \mid X_i) = \begin{bmatrix} E(Y_{i1} \mid X_i) \\ \vdots \\ E(Y_{iT} \mid X_i) \end{bmatrix} = \begin{bmatrix} \mu_1(X_i; \beta) \\ \vdots \\ \mu_T(X_i; \beta) \end{bmatrix}$$

- Without further specification of the distribution of the data, this is a *semiparametric model*

# Generalized Estimating Equations

Liang and Zeger (Biometrika 1986) assumed a *generalized linear model* marginally at each time point

- $p(y_{ij}) = \exp[\{y_{ij}\theta_{ij} - a(\theta_{ij}) + b(y_{ij})\}\phi]$
- $E(y_{ij}) = a'(\theta_{ij})$, $V(y_{ij}) = a''(\theta_{ij})/\phi$
- $\mu_j(X_i; \beta) = a'(\theta_{ij})$

# Generalized Estimating Equations

Liang and Zeger (Biometrika 1986) assumed a *generalized linear model* marginally at each time point

- $p(y_{ij}) = \exp[\{y_{ij}\theta_{ij} - a(\theta_{ij}) + b(y_{ij})\}\phi]$
- $E(y_{ij}) = a'(\theta_{ij})$, $V(y_{ij}) = a''(\theta_{ij})/\phi$
- $\mu_j(X_i; \beta) = a'(\theta_{ij})$

# Generalized Estimating Equations

Liang and Zeger (Biometrika 1986) assumed a *generalized linear model* marginally at each time point

- $p(y_{ij}) = \exp[\{y_{ij}\theta_{ij} - a(\theta_{ij}) + b(y_{ij})\}\phi]$
- $E(y_{ij}) = a'(\theta_{ij}),\ V(y_{ij}) = a''(\theta_{ij})/\phi$
- $\mu_j(X_i; \beta) = a'(\theta_{ij})$

# Generalized Estimating Equations

Liang and Zeger (Biometrika 1986) assumed a *generalized linear model* marginally at each time point

- $p(y_{ij}) = \exp[\{y_{ij}\theta_{ij} - a(\theta_{ij}) + b(y_{ij})\}\phi]$
- $E(y_{ij}) = a'(\theta_{ij})$, $V(y_{ij}) = a''(\theta_{ij})/\phi$
- $\mu_j(X_i; \beta) = a'(\theta_{ij})$

# Generalized Estimating Equations

- The estimate $\hat{\beta}$ is obtained from solving the *generalized estimating equations*

$$\sum_{i=1}^{n} \mathcal{D}_i^T \mathcal{V}_i^{-1} \begin{bmatrix} Y_{i1} - \mu_1(X_i;\beta) \\ \vdots \\ Y_{iT} - \mu_T(X_i;\beta) \end{bmatrix} = \mathbf{0}$$

- $\mathcal{D}_i = \frac{\partial}{\partial \beta^T}[\mu(X_i;\beta)]$ is $T \times p$
- $\mathcal{V}_i$ is a $T \times T$ *working covariance matrix*, specified through the working correlation matrix $R(\alpha)$

$$\mathcal{V}_i = A_i^{1/2} R(\alpha) A_i^{1/2} / \phi$$

where $A_i = \text{diag}\{a''(\theta_{ij})\}$

# Generalized Estimating Equations

- The estimate $\hat{\beta}$ is obtained from solving the *generalized estimating equations*

$$\sum_{i=1}^{n} \mathcal{D}_i^T \mathcal{V}_i^{-1} \begin{bmatrix} Y_{i1} - \mu_1(X_i;\beta) \\ \vdots \\ Y_{iT} - \mu_T(X_i;\beta) \end{bmatrix} = \mathbf{0}$$

- $\mathcal{D}_i = \frac{\partial}{\partial \beta^T}[\mu(X_i;\beta)]$ is $T \times p$
- $\mathcal{V}_i$ is a $T \times T$ *working covariance matrix*, specified through the working correlation matrix $R(\alpha)$

$$\mathcal{V}_i = A_i^{1/2} R(\alpha) A_i^{1/2} / \phi$$

where $A_i = \text{diag}\{a''(\theta_{ij})\}$

# Generalized Estimating Equations

- The estimate $\hat{\beta}$ is obtained from solving the *generalized estimating equations*

$$\sum_{i=1}^{n} \mathcal{D}_i^T \mathcal{V}_i^{-1} \begin{bmatrix} Y_{i1} - \mu_1(X_i;\beta) \\ \vdots \\ Y_{iT} - \mu_T(X_i;\beta) \end{bmatrix} = \mathbf{0}$$

- $\mathcal{D}_i = \frac{\partial}{\partial \beta^T}[\mu(X_i;\beta)]$ is $T \times p$
- $\mathcal{V}_i$ is a $T \times T$ *working covariance matrix*, specified through the working correlation matrix $R(\alpha)$

$$\mathcal{V}_i = A_i^{1/2} R(\alpha) A_i^{1/2} / \phi$$

where $A_i = \text{diag}\{a''(\theta_{ij})\}$

# Generalized Estimating Equations

Examples of correlation matrices $R(\alpha) \equiv R$

- $R_{jj'} = 0$ (independence)
- $R_{jj'} = \alpha$ (exchangeability)
- $R_{jj'} = \alpha^{|j-j'|}$ (autoregressive of order 1 – AR 1)
- $R_{jj'} = \alpha_{jj'}$ (unstructured)

# Generalized Estimating Equations

Examples of correlation matrices $R(\alpha) \equiv R$

- $R_{jj'} = 0$ (independence)
- $R_{jj'} = \alpha$ (exchangeability)
- $R_{jj'} = \alpha^{|j-j'|}$ (autoregressive of order 1 – AR 1)
- $R_{jj'} = \alpha_{jj'}$ (unstructured)

# Generalized Estimating Equations

Examples of correlation matrices $R(\alpha) \equiv R$

- $R_{jj'} = 0$ (independence)
- $R_{jj'} = \alpha$ (exchangeability)
- $R_{jj'} = \alpha^{|j-j'|}$ (autoregressive of order 1 – AR 1)
- $R_{jj'} = \alpha_{jj'}$ (unstructured)

# Generalized Estimating Equations

Examples of correlation matrices $R(\alpha) \equiv R$

- $R_{jj'} = 0$ (independence)
- $R_{jj'} = \alpha$ (exchangeability)
- $R_{jj'} = \alpha^{|j-j'|}$ (autoregressive of order 1 – AR 1)
- $R_{jj'} = \alpha_{jj'}$ (unstructured)

# Generalized Estimating Equations

Examples of correlation matrices $R(\alpha) \equiv R$

- $R_{jj'} = 0$ (independence)
- $R_{jj'} = \alpha$ (exchangeability)
- $R_{jj'} = \alpha^{|j-j'|}$ (autoregressive of order 1 – AR 1)
- $R_{jj'} = \alpha_{jj'}$ (unstructured)

# Generalized Estimating Equations

Comments

- $\mathcal{V}(X;\beta) = V(Y \mid X)$ if $R(\alpha)$ is indeed the true correlation of $Y_1, \ldots, Y_T$
- Parameter estimates from the GEE are consistent even when the correlation structure is misspecified
- Approximately correct specification of $R(\alpha)$ improves the efficiency of the estimator
- Full specification of the joint distribution of the correlated responses is not needed

# Generalized Estimating Equations

Comments

- $\mathcal{V}(X;\beta) = V(Y \mid X)$ if $R(\alpha)$ is indeed the true correlation of $Y_1, \ldots, Y_T$
- Parameter estimates from the GEE are consistent even when the correlation structure is misspecified
- Approximately correct specification of $R(\alpha)$ improves the efficiency of the estimator
- Full specification of the joint distribution of the correlated responses is not needed

# Generalized Estimating Equations

Comments

- $\mathcal{V}(X;\beta) = V(Y \mid X)$ if $R(\alpha)$ is indeed the true correlation of $Y_1, \ldots, Y_T$
- Parameter estimates from the GEE are consistent even when the correlation structure is misspecified
- Approximately correct specification of $R(\alpha)$ improves the efficiency of the estimator
- Full specification of the joint distribution of the correlated responses is not needed

# Generalized Estimating Equations

Comments

- $\mathcal{V}(X;\beta) = V(Y \mid X)$ if $R(\alpha)$ is indeed the true correlation of $Y_1, \ldots, Y_T$
- Parameter estimates from the GEE are consistent even when the correlation structure is misspecified
- Approximately correct specification of $R(\alpha)$ improves the efficiency of the estimator
- Full specification of the joint distribution of the correlated responses is not needed

# Outline

# GEEs and Missing Data

- When we have missing data, we could ask about the validity of the GEE estimator $\hat{\beta}^{ac}$ based on *available cases*, derived from

$$\sum_{i=1}^{n} \mathcal{D}_i^T \mathcal{V}_i^{-1} \mathcal{R}_i \begin{bmatrix} Y_{i1} - \mu_1(X_i; \beta) \\ \vdots \\ Y_{iT} - \mu_T(X_i; \beta) \end{bmatrix} = \mathbf{0}$$

  where $\mathcal{R}_i = \text{diag}(R_{i1}, \ldots, R_{iT})$

- The solution to this system of equations will converge to the solution of

$$E\left\{ \mathcal{D}^T(X) \mathcal{V}^{-1}(X) \mathcal{R} \begin{bmatrix} Y_1 - \mu_1(X; \beta) \\ \vdots \\ Y_T - \mu_T(X; \beta) \end{bmatrix} \right\} = \mathbf{0}$$

# GEEs and Missing Data

- When we have missing data, we could ask about the validity of the GEE estimator $\hat{\beta}^{ac}$ based on *available cases*, derived from

$$\sum_{i=1}^{n} \mathcal{D}_i^T \mathcal{V}_i^{-1} \mathcal{R}_i \begin{bmatrix} Y_{i1} - \mu_1(X_i;\beta) \\ \vdots \\ Y_{iT} - \mu_T(X_i;\beta) \end{bmatrix} = \mathbf{0}$$

where $\mathcal{R}_i = \text{diag}(R_{i1}, \ldots, R_{iT})$

- The solution to this system of equations will converge to the solution of

$$E\left\{ \mathcal{D}^T(X)\mathcal{V}^{-1}(X)\mathcal{R} \begin{bmatrix} Y_1 - \mu_1(X;\beta) \\ \vdots \\ Y_T - \mu_T(X;\beta) \end{bmatrix} \right\} = \mathbf{0}$$

# GEEs and Missing Data

- We can write the left-hand side of the above expression as

$$E_X\left(\mathcal{D}^T(X)\mathcal{V}^{-1}(X)E_{Y|X}\left\{E[\mathcal{R}\mid X, Y]\begin{bmatrix} Y_1-\mu_1(X;\beta) \\ \vdots \\ Y_T-\mu_T(X;\beta)\end{bmatrix}\mid X\right\}\right)$$

- Note that if $R \perp\!\!\!\perp Y \mid X$ then this expression is zero when $\mu(X;\beta)$ is correctly specified and the solution $\hat{\beta}^{ac}$ is a consistent estimator of $\beta$
- Otherwise there's no guarantee that $\hat{\beta}^{ac}$ is consistent

# GEEs and Missing Data

- We can write the left-hand side of the above expression as

$$E_X\left(\mathcal{D}^T(X)\mathcal{V}^{-1}(X)E_{Y|X}\left\{E[\mathcal{R} \mid X, Y]\begin{bmatrix} Y_1-\mu_1(X;\beta) \\ \vdots \\ Y_T-\mu_T(X;\beta)\end{bmatrix} \mid X\right\}\right)$$

- Note that if $R \perp\!\!\!\perp Y \mid X$ then this expression is zero when $\mu(X;\beta)$ is correctly specified and the solution $\hat{\beta}^{ac}$ is a consistent estimator of $\beta$
- Otherwise there's no guarantee that $\hat{\beta}^{ac}$ is consistent

# GEEs and Missing Data

- We can write the left-hand side of the above expression as

$$E_X\left(\mathcal{D}^T(X)\mathcal{V}^{-1}(X)E_{Y|X}\left\{E[\mathcal{R}\mid X,Y]\begin{bmatrix}Y_1-\mu_1(X;\beta)\\ \vdots \\ Y_T-\mu_T(X;\beta)\end{bmatrix}\mid X\right\}\right)$$

- Note that if $R \perp\!\!\!\perp Y \mid X$ then this expression is zero when $\mu(X;\beta)$ is correctly specified and the solution $\hat{\beta}^{ac}$ is a consistent estimator of $\beta$
- Otherwise there's no guarantee that $\hat{\beta}^{ac}$ is consistent

# Observation/Occasion-Specific Weighted GEE

- The *observation- or occasion- specific weighted GEE method* solves the *weighted generalized estimating equations*

$$\sum_{i=1}^{n} \mathcal{D}_i^T \mathcal{V}_i^{-1} \mathcal{W}_i \begin{bmatrix} Y_{i1} - \mu_1(X_i;\beta) \\ \vdots \\ Y_{iT} - \mu_T(X_i;\beta) \end{bmatrix} = \mathbf{0}$$

- $\mathcal{W}_i = \text{diag}\{R_{i1}w_{i1}, \ldots, R_{iT}w_{iT}\}$, with $w_{ij} = p(R_{ij} = 1 \mid X_i, Y_i)^{-1}$
- Note that while under likelihood-based inference we can ignore the response mechanism, here it needs to be explicitly modeled to estimate the weights, even if the missingness mechanism is assumed to be ignorable[1]

[1] See Sun & Tchetgen Tchetgen (JASA 2018) https://doi.org/10.1080/01621459.2016.1256814

# Observation/Occasion-Specific Weighted GEE

- The *observation- or occasion- specific weighted GEE method* solves the *weighted generalized estimating equations*

$$\sum_{i=1}^{n} \mathcal{D}_i^T \mathcal{V}_i^{-1} \mathcal{W}_i \begin{bmatrix} Y_{i1} - \mu_1(X_i; \beta) \\ \vdots \\ Y_{iT} - \mu_T(X_i; \beta) \end{bmatrix} = \mathbf{0}$$

- $\mathcal{W}_i = \text{diag}\{R_{i1} w_{i1}, \ldots, R_{iT} w_{iT}\}$, with $w_{ij} = p(R_{ij} = 1 \mid X_i, Y_i)^{-1}$
- Note that while under likelihood-based inference we can ignore the response mechanism, here it needs to be explicitly modeled to estimate the weights, even if the missingness mechanism is assumed to be ignorable[1]

[1] See Sun & Tchetgen Tchetgen (JASA 2018) https://doi.org/10.1080/01621459.2016.1256814

# Observation/Occasion-Specific Weighted GEE

- The *observation- or occasion- specific weighted GEE method* solves the *weighted generalized estimating equations*

$$\sum_{i=1}^{n} \mathcal{D}_i^T \mathcal{V}_i^{-1} \mathcal{W}_i \begin{bmatrix} Y_{i1} - \mu_1(X_i; \beta) \\ \vdots \\ Y_{iT} - \mu_T(X_i; \beta) \end{bmatrix} = \mathbf{0}$$

- $\mathcal{W}_i = \text{diag}\{R_{i1} w_{i1}, \ldots, R_{iT} w_{iT}\}$, with $w_{ij} = p(R_{ij} = 1 \mid X_i, Y_i)^{-1}$
- Note that while under likelihood-based inference we can ignore the response mechanism, here it needs to be explicitly modeled to estimate the weights, even if the missingness mechanism is assumed to be ignorable[1]

[1] See Sun & Tchetgen Tchetgen (JASA 2018) https://doi.org/10.1080/01621459.2016.1256814

# Missing at Random and Dropout

- It is not too difficult to accommodate dropout into WGEEs under MAR
- Let $D$ be the *dropout* indicator, where $D = j + 1$ indicates that the individual is last seen at time $j$
- Denote the *hazard function* as $\lambda_j(Z) = p(D = j \mid D \geq j, Z)$
- It can be shown that (HW4)

$$p(D = j + 1 \mid Z) = \lambda_{j+1}(Z) \prod_{\ell=1}^{j} [1 - \lambda_\ell(Z)]$$

and

$$p(R_j = 1 \mid Z) = \prod_{\ell=1}^{j} [1 - \lambda_\ell(Z)]$$

# Missing at Random and Dropout

- It is not too difficult to accommodate dropout into WGEEs under MAR
- Let $D$ be the *dropout* indicator, where $D = j + 1$ indicates that the individual is last seen at time $j$
- Denote the *hazard function* as $\lambda_j(Z) = p(D = j \mid D \geq j, Z)$
- It can be shown that (HW4)

$$p(D = j + 1 \mid Z) = \lambda_{j+1}(Z) \prod_{\ell=1}^{j} [1 - \lambda_\ell(Z)]$$

and

$$p(R_j = 1 \mid Z) = \prod_{\ell=1}^{j} [1 - \lambda_\ell(Z)]$$

# Missing at Random and Dropout

- It is not too difficult to accommodate dropout into WGEEs under MAR
- Let $D$ be the *dropout* indicator, where $D = j + 1$ indicates that the individual is last seen at time $j$
- Denote the *hazard function* as $\lambda_j(Z) = p(D = j \mid D \geq j, Z)$
- It can be shown that (HW4)

$$p(D = j + 1 \mid Z) = \lambda_{j+1}(Z) \prod_{\ell=1}^{j} [1 - \lambda_\ell(Z)]$$

and

$$p(R_j = 1 \mid Z) = \prod_{\ell=1}^{j} [1 - \lambda_\ell(Z)]$$

# Missing at Random and Dropout

- It is not too difficult to accommodate dropout into WGEEs under MAR
- Let $D$ be the *dropout* indicator, where $D = j + 1$ indicates that the individual is last seen at time $j$
- Denote the *hazard function* as $\lambda_j(Z) = p(D = j \mid D \geq j, Z)$
- It can be shown that (HW4)

$$p(D = j + 1 \mid Z) = \lambda_{j+1}(Z) \prod_{\ell=1}^{j} [1 - \lambda_\ell(Z)]$$

and

$$p(R_j = 1 \mid Z) = \prod_{\ell=1}^{j} [1 - \lambda_\ell(Z)]$$

# Missing at Random and Dropout

- Let us define the *history* up to time $j$ as

$$H_j = \{X, (Y_1, V_1), ..., (Y_j, V_j)\},$$

- The auxiliary $V$ variables might not be of scientific interest, but might be seen as important to model the dropout process
- The MAR assumption in this case is equivalent to (HW4)

$$\lambda_j(Z) = p(D = j \mid D \geq j, Z) = p(D = j \mid D \geq j, H_{j-1}) = \lambda_j(H_{j-1})$$

- Note that each $\lambda_j(H_{j-1})$ can be estimated form the observed data, for example using a logistic regression (explain, HW4)
- Given estimates $\hat{\lambda}_j(H_{j-1})$, we can estimate the weights of the WGEE above as $w_{ij} = \hat{p}(R_j = 1 \mid H_{j-1})^{-1}$

# Missing at Random and Dropout

- Let us define the *history* up to time $j$ as

$$H_j = \{X, (Y_1, V_1), ..., (Y_j, V_j)\},$$

- The auxiliary $V$ variables might not be of scientific interest, but might be seen as important to model the dropout process
- The MAR assumption in this case is equivalent to (HW4)

$$\lambda_j(Z) = p(D = j \mid D \geq j, Z) = p(D = j \mid D \geq j, H_{j-1}) = \lambda_j(H_{j-1})$$

- Note that each $\lambda_j(H_{j-1})$ can be estimated form the observed data, for example using a logistic regression (explain, HW4)
- Given estimates $\hat{\lambda}_j(H_{j-1})$, we can estimate the weights of the WGEE above as $w_{ij} = \hat{p}(R_j = 1 \mid H_{j-1})^{-1}$

# Missing at Random and Dropout

- Let us define the *history* up to time $j$ as

$$H_j = \{X, (Y_1, V_1), ..., (Y_j, V_j)\},$$

- The auxiliary $V$ variables might not be of scientific interest, but might be seen as important to model the dropout process
- The MAR assumption in this case is equivalent to (HW4)

$$\lambda_j(Z) = p(D = j \mid D \geq j, Z) = p(D = j \mid D \geq j, H_{j-1}) = \lambda_j(H_{j-1})$$

- Note that each $\lambda_j(H_{j-1})$ can be estimated form the observed data, for example using a logistic regression (explain, HW4)
- Given estimates $\hat{\lambda}_j(H_{j-1})$, we can estimate the weights of the WGEE above as $w_{ij} = \hat{p}(R_j = 1 \mid H_{j-1})^{-1}$

# Missing at Random and Dropout

- Let us define the *history* up to time $j$ as

$$H_j = \{X, (Y_1, V_1), ..., (Y_j, V_j)\},$$

- The auxiliary $V$ variables might not be of scientific interest, but might be seen as important to model the dropout process
- The MAR assumption in this case is equivalent to (HW4)

$$\lambda_j(Z) = p(D = j \mid D \geq j, Z) = p(D = j \mid D \geq j, H_{j-1}) = \lambda_j(H_{j-1})$$

- Note that each $\lambda_j(H_{j-1})$ can be estimated form the observed data, for example using a logistic regression (explain, HW4)
- Given estimates $\hat{\lambda}_j(H_{j-1})$, we can estimate the weights of the WGEE above as $w_{ij} = \hat{p}(R_j = 1 \mid H_{j-1})^{-1}$

# Missing at Random and Dropout

- Let us define the *history* up to time $j$ as

$$H_j = \{X, (Y_1, V_1), ..., (Y_j, V_j)\},$$

- The auxiliary $V$ variables might not be of scientific interest, but might be seen as important to model the dropout process
- The MAR assumption in this case is equivalent to (HW4)

$$\lambda_j(Z) = p(D = j \mid D \geq j, Z) = p(D = j \mid D \geq j, H_{j-1}) = \lambda_j(H_{j-1})$$

- Note that each $\lambda_j(H_{j-1})$ can be estimated form the observed data, for example using a logistic regression (explain, HW4)
- Given estimates $\hat{\lambda}_j(H_{j-1})$, we can estimate the weights of the WGEE above as $w_{ij} = \hat{p}(R_j = 1 \mid H_{j-1})^{-1}$

# Subject-Specific Weighted GEE

- Under dropout, another way of implementing WGEEs is via the *subject-specific weighted GEE method*, which solves the *weighted generalized estimating equations*

$$\sum_{i=1}^{n} w_i \mathcal{D}_i^T \mathcal{V}_i^{-1} \mathcal{R}_i \begin{bmatrix} Y_{i1} - \mu_1(X_i;\beta) \\ \vdots \\ Y_{iT} - \mu_T(X_i;\beta) \end{bmatrix} = \mathbf{0}$$

  where $\mathcal{R}_i = \text{diag}(R_{i1}, \ldots, R_{iT})$

- The weight $w_i = p(D_i = d_i \mid H_{d_i-1})^{-1}$ for subject $i$ is the inverse probability of a subject $i$ dropping out at the observed dropout time $d_i$

- The weights can be estimated from an estimate of $p(D_i = d_i \mid H_{d_i-1})$ as explained above

# Subject-Specific Weighted GEE

- Under dropout, another way of implementing WGEEs is via the *subject-specific weighted GEE method*, which solves the *weighted generalized estimating equations*

$$\sum_{i=1}^{n} w_i \mathcal{D}_i^T \mathcal{V}_i^{-1} \mathcal{R}_i \begin{bmatrix} Y_{i1} - \mu_1(X_i;\beta) \\ \vdots \\ Y_{iT} - \mu_T(X_i;\beta) \end{bmatrix} = \mathbf{0}$$

  where $\mathcal{R}_i = \mathrm{diag}(R_{i1}, \ldots, R_{iT})$

- The weight $w_i = p(D_i = d_i \mid H_{d_i - 1})^{-1}$ for subject $i$ is the inverse probability of a subject $i$ dropping out at the observed dropout time $d_i$

- The weights can be estimated from an estimate of $p(D_i = d_i \mid H_{d_i - 1})$ as explained above

# Subject-Specific Weighted GEE

- Under dropout, another way of implementing WGEEs is via the *subject-specific weighted GEE method*, which solves the *weighted generalized estimating equations*

$$\sum_{i=1}^{n} w_i \mathcal{D}_i^T \mathcal{V}_i^{-1} \mathcal{R}_i \begin{bmatrix} Y_{i1} - \mu_1(X_i;\beta) \\ \vdots \\ Y_{iT} - \mu_T(X_i;\beta) \end{bmatrix} = \mathbf{0}$$

  where $\mathcal{R}_i = \text{diag}(R_{i1}, \ldots, R_{iT})$

- The weight $w_i = p(D_i = d_i \mid H_{d_i - 1})^{-1}$ for subject $i$ is the inverse probability of a subject $i$ dropping out at the observed dropout time $d_i$

- The weights can be estimated from an estimate of $p(D_i = d_i \mid H_{d_i - 1})$ as explained above

# Subject-Specific Weighted GEE

Davidian and Tsiatis note

- "*Theoretically, it is not straightforward to deduce if the subject level or occasion level approach is preferred in general on the basis of efficiency*"
- Expensive simulation studies have shown that "*under MAR, the occasion level WGEE is to be preferred on efficiency grounds*" (Preisser, Lohman, and Rathouz, 2002)

# Subject-Specific Weighted GEE

Davidian and Tsiatis note

- "*Theoretically, it is not straightforward to deduce if the subject level or occasion level approach is preferred in general on the basis of efficiency*"
- Expensive simulation studies have shown that "*under MAR, the occasion level WGEE is to be preferred on efficiency grounds*" (Preisser, Lohman, and Rathouz, 2002)

# Outline

# Summary

Main take-aways from today's lecture:

- We only scratched the surface of
  - Estimating equations
  - Generalized estimating equations
  - Weighted generalized estimating equations
  - Monotone nonresponse (dropout)
- Comments
  - No need of a full-data model (semiparametric) but missingness mechanism has to be correctly modeled!
  - General doubly robust, augmented inverse probability weighted estimators not covered here! (only for eatimating mean)
  - How to obtain standard errors?: asymptotic covariance matrix can be obtained using the sandwich technique (these are $M$-estimators), but
    - "*The parameter of interest is estimated jointly with the parameters in the dropout models and working covariance model by solving accompanying estimating equations for these parameters*"
    - Or use the bootstrap!

Next lecture:

- R session 4

# Summary

Main take-aways from today's lecture:

- We only scratched the surface of
  - Estimating equations
  - Generalized estimating equations
  - Weighted generalized estimating equations
  - Monotone nonresponse (dropout)
- Comments
  - No need of a full-data model (semiparametric) but missingness mechanism has to be correctly modeled!
  - General doubly robust, augmented inverse probability weighted estimators not covered here! (only for eatimating mean)
  - How to obtain standard errors?: asymptotic covariance matrix can be obtained using the sandwich technique (these are *M*-estimators), but
    - "*The parameter of interest is estimated jointly with the parameters in the dropout models and working covariance model by solving accompanying estimating equations for these parameters*"
    - Or use the bootstrap!

Next lecture:

- R session 4

# Summary

Main take-aways from today's lecture:

- We only scratched the surface of
  - Estimating equations
  - Generalized estimating equations
  - Weighted generalized estimating equations
  - Monotone nonresponse (dropout)
- Comments
  - No need of a full-data model (semiparametric) but missingness mechanism has to be correctly modeled!
  - General doubly robust, augmented inverse probability weighted estimators not covered here! (only for eatimating mean)
  - How to obtain standard errors?: asymptotic covariance matrix can be obtained using the sandwich technique (these are $M$-estimators), but
    - *"The parameter of interest is estimated jointly with the parameters in the dropout models and working covariance model by solving accompanying estimating equations for these parameters"*
    - Or use the bootstrap!

Next lecture:

- R session 4

# Summary

Main take-aways from today's lecture:

- We only scratched the surface of
  - Estimating equations
  - Generalized estimating equations
  - Weighted generalized estimating equations
  - Monotone nonresponse (dropout)
- Comments
  - No need of a full-data model (semiparametric) but missingness mechanism has to be correctly modeled!
  - General doubly robust, augmented inverse probability weighted estimators not covered here! (only for eatimating mean)
  - How to obtain standard errors?: asymptotic covariance matrix can be obtained using the sandwich technique (these are $M$-estimators), but
    - *"The parameter of interest is estimated jointly with the parameters in the dropout models and working covariance model by solving accompanying estimating equations for these parameters"*
    - Or use the bootstrap!

Next lecture:

- R session 4

# Summary

Main take-aways from today's lecture:

- We only scratched the surface of
  - Estimating equations
  - Generalized estimating equations
  - Weighted generalized estimating equations
  - Monotone nonresponse (dropout)
- Comments
  - No need of a full-data model (semiparametric) but missingness mechanism has to be correctly modeled!
  - General doubly robust, augmented inverse probability weighted estimators not covered here! (only for eatimating mean)
  - How to obtain standard errors?: asymptotic covariance matrix can be obtained using the sandwich technique (these are $M$-estimators), but
    - "*The parameter of interest is estimated jointly with the parameters in the dropout models and working covariance model by solving accompanying estimating equations for these parameters*"
    - Or use the bootstrap!

Next lecture:

- R session 4

# Summary

Main take-aways from today's lecture:

- We only scratched the surface of
  - Estimating equations
  - Generalized estimating equations
  - Weighted generalized estimating equations
  - Monotone nonresponse (dropout)
- Comments
  - No need of a full-data model (semiparametric) but missingness mechanism has to be correctly modeled!
  - General doubly robust, augmented inverse probability weighted estimators not covered here! (only for eatimating mean)
  - How to obtain standard errors?: asymptotic covariance matrix can be obtained using the sandwich technique (these are $M$-estimators), but
    - "*The parameter of interest is estimated jointly with the parameters in the dropout models and working covariance model by solving accompanying estimating equations for these parameters*"
    - Or use the bootstrap!

Next lecture:

- R session 4

# Summary

Main take-aways from today's lecture:

- We only scratched the surface of
  - Estimating equations
  - Generalized estimating equations
  - Weighted generalized estimating equations
  - Monotone nonresponse (dropout)
- Comments
  - No need of a full-data model (semiparametric) but missingness mechanism has to be correctly modeled!
  - General doubly robust, augmented inverse probability weighted estimators not covered here! (only for eatimating mean)
  - How to obtain standard errors?: asymptotic covariance matrix can be obtained using the sandwich technique (these are $M$-estimators), but
    - "*The parameter of interest is estimated jointly with the parameters in the dropout models and working covariance model by solving accompanying estimating equations for these parameters*"
    - Or use the bootstrap!

Next lecture:

- R session 4