Why are GEE and LMM biased with time-varying covariate?

Tianchen Qian

2017.11.09

1 Theoretical Explanation

Consider a two-stage study with m subjects, where the data for i-th subject is $(X_{i1}, Y_{i1}, X_{i2}, Y_{i2})$. Note that in order to be consistent with the literature on GEE and LMM, the time index for Y is the same as for X. (This is different from the index Susan usually uses.)

Also note that we don't consider treatment for now, so this is standard regression setting (no causal inference involved).

1.1 Generalized Estimating Equations (GEE)

Define $\mu_{it} := E[Y_{it} \mid X_{it}]$. For simplicity we assume Y is continuous. Suppose we assume the following mean model: $\mu_{it} = \beta_0 + \beta_1 X_{it}$ (assuming X_{it} is a scalar). The goal of GEE is to estimate β_0 and β_1.

Define $X_i := (X_{i1}, X_{i2})^T, Y_i := (Y_{i1}, Y_{i2})^T, \mu_i := (\mu_{i1}, \mu_{i2})^T$. Define $\beta := (\beta_0, \beta_1)^T$. GEE solves the following estimating equation:

$$\sum_{i=1}^{m} \frac{\partial \mu_i}{\partial \beta^T} W_i (Y_i - \mu_i) = 0. \tag{1}$$

Writing out each entry in (1), it becomes

$$\sum_{i=1}^{m} \begin{bmatrix} \frac{\partial \mu_{i1}}{\partial \beta_{i1}} & \frac{\partial \mu_{i2}}{\partial \beta_{i1}} \\ \frac{\partial \mu_{i1}}{\partial \beta_{i2}} & \frac{\partial \mu_{i2}}{\partial \beta_{i2}} \end{bmatrix} W_i \begin{bmatrix} Y_{i1} - \mu_{i1} \\ Y_{i2} - \mu_{i2} \end{bmatrix} = \sum_{i=1}^{m} \begin{bmatrix} 1 & 1 \\ X_{i1} & X_{i2} \end{bmatrix} W_i \begin{bmatrix} Y_{i1} - \beta_0 - \beta_1 X_{i1} \\ Y_{i2} - \beta_0 - \beta_1 X_{i2} \end{bmatrix} = 0. \tag{2}$$

Here, W_i is the inverse of the working covariance matrix. Examples of W_i are:

- Working independence:
$$W_i^{\text{ind}} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}^{-1}.$$
- Compound symmetry:
$$W_i^{\text{cs}} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}^{-1}.$$

GEE (1) gives unbiased $\hat{\beta}_0$ and $\hat{\beta}_1$ if the left hand side of (1) has expectation zero (in which case (1) is called an unbiased estimating equation). Pepe and Anderson (1994) point out that (1) is unbiased if one of the two conditions is satisfied.

Theorem (Pepe and Anderson (1994) in the context of two-stage study)**.** *If*

i) $E(Y_{i1} \mid X_{i1}, X_{i2}) = E(Y_{i1} \mid X_{i1})$ and $E(Y_{i2} \mid X_{i1}, X_{i2}) = E(Y_{i2} \mid X_{i2})$, or

ii) a working independence correlation structure is used (i.e., $W_i = W_i^{ind}$ is diagonal),

then $E\left\{ \frac{\partial \mu_i}{\partial \beta^T} W_i (Y_i - \mu_i) \right\} = 0$, that is (1) is unbiased.

Proof. Write out the matrix W_i as

$$W_i = \begin{bmatrix} w_{i11} & w_{i12} \\ w_{i21} & w_{i22} \end{bmatrix}.$$

A summand in equation (2) becomes

$$\begin{aligned} &\begin{bmatrix} 1 & 1 \\ X_{i1} & X_{i2} \end{bmatrix} \begin{bmatrix} w_{i11} & w_{i12} \\ w_{i21} & w_{i22} \end{bmatrix} \begin{bmatrix} Y_{i1} - \beta_0 - \beta_1 X_{i1} \\ Y_{i2} - \beta_0 - \beta_1 X_{i2} \end{bmatrix} \\ &= \begin{bmatrix} (w_{i11} + w_{i21})(Y_{i1} - \beta_0 - \beta_1 X_{i1}) + (w_{i12} + w_{i22})(Y_{i2} - \beta_0 - \beta_1 X_{i2}) \\ (w_{i11} X_{i1} + w_{i21} X_{i2})(Y_{i1} - \beta_0 - \beta_1 X_{i1}) + (w_{i12} X_{i1} + w_{i22} X_{i2})(Y_{i2} - \beta_0 - \beta_1 X_{i2}) \end{bmatrix}. \end{aligned} \tag{3}$$

By definition of μ_{it} ($\mu_{it} := E[Y_{it} \mid X_{it}] = \beta_0 + \beta_1 X_{it}$), we have

$$\begin{aligned} E(Y_{i1} - \beta_0 - \beta_1 X_{i1}\} &= 0, \\ E(Y_{i2} - \beta_0 - \beta_1 X_{i2}\} &= 0, \\ E\{X_{i1}(Y_{i1} - \beta_0 - \beta_1 X_{i1})\} &= E[E\{X_{i1}(Y_{i1} - \beta_0 - \beta_1 X_{i1}) \mid X_{i1}\}] = 0, \\ E\{X_{i2}(Y_{i2} - \beta_0 - \beta_1 X_{i2})\} &= E[E\{X_{i2}(Y_{i2} - \beta_0 - \beta_1 X_{i2}) \mid X_{i2}\}] = 0. \end{aligned}$$

Therefore, the expectation of (3) equals

$$\begin{bmatrix} 0 \\ w_{i21} E\{X_{i2}(Y_{i1} - \beta_0 - \beta_1 X_{i1})\} + w_{i12} E\{X_{i1}(Y_{i2} - \beta_0 - \beta_1 X_{i2})\} \end{bmatrix}. \tag{4}$$

Under Condition i), we have

$$\begin{aligned} E\{X_{i2}(Y_{i1} - \beta_0 - \beta_1 X_{i1})\} &= E[E\{X_{i2}(Y_{i1} - \beta_0 - \beta_1 X_{i1}) \mid X_{i1}, X_{i2}\}] \\ &= E[X_{i2} E\{(Y_{i1} - \beta_0 - \beta_1 X_{i1}) \mid X_{i1}, X_{i2}\}] \\ &= E[X_{i2} E\{(Y_{i1} - \beta_0 - \beta_1 X_{i1}) \mid X_{i1}\}] = 0, \end{aligned} \tag{5}$$

and by a similar reasoning

$$E\{X_{i1}(Y_{i2} - \beta_0 - \beta_1 X_{i2})\} = 0. \tag{6}$$

By (5) and (6), we know that (4) equals 0.

Under Condition ii), $w_{i21} = w_{i12} = 0$, hence (4) equals 0.

Therefore, we showed that under either i) or ii), the expectation of (3) equals 0. This finishes the proof. □

Remark. In the presence of time-varying covariate, X_{i2} can depend on Y_{i1}. In fact, X_{i2} can be a function of Y_{i1}. In this case, $E(Y_{i1} \mid X_{i1}, X_{i2}) = E(Y_{i1} \mid X_{i1})$ doesn't hold, and it is likely that $E\{X_{i2}(Y_{i1} - \beta_0 - \beta_1 X_{i1})\} \neq 0$. Therefore, if not using a working independence correlation structure, GEE will produce biased $\hat{\beta}_0$ and $\hat{\beta}_1$.

1.2 Linear Mixed Model (LMM)

For simplicity, we consider a LMM with random intercept. Assume $u_i \sim N(0, \sigma_u^2)$ is a random intercept for subject i. LMM assumes that $Y_{it} = \beta_0 + \beta_1 X_{it} + u_i + \epsilon_{it}$, where $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$. LMM also assumes that u_i and ϵ_{it} are all independent of each other and independent of X_i. Thus $Y_i = (Y_{i1}, Y_{i2})^T$ follows the multivariate normal distribution with the following mean vector and covariance matrix

$$\begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix} \sim MVN \left(\begin{bmatrix} \beta_0 + \beta_1 X_{i1} \\ \beta_0 + \beta_1 X_{i2} \end{bmatrix}, \underbrace{\begin{bmatrix} \sigma_u^2 + \sigma_\epsilon^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_\epsilon^2 \end{bmatrix}}_{\text{denote by } \Sigma} \right). \tag{7}$$

Hence, the likelihood of the data is

$$L = \prod_{i=1}^{m} \prod_{i=1}^{m} (2\pi)^{-1} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} (Y_i - \mu_i)^T \Sigma^{-1} (Y_i - \mu_i) \right\},$$

the log-likelihood is

$$l = -\frac{1}{2} \sum_{i=1}^{m} (Y_i - \mu_i)^T \Sigma_i^{-1} (Y_i - \mu_i) - m \log 2\pi - \frac{1}{2} \sum_{i=1}^{m} \log |\Sigma|,$$

and the score equation for β is

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial \mu_i}{\partial \beta^T} \Sigma^{-1} (Y_i - \mu_i). \tag{8}$$

LMM solves estimating equation (8). Note that (8) is the same as (1) when we let $W_i = \Sigma^{-1}$. Therefore, in this case LMM is the same as GEE with compound symmetric working correlation matrix.

Because LMM is the same as GEE with compound symmetric working correlation matrix, we can use the theorem and remark in Section 1.1 to argue that with time-varying covariate, when $E(Y_{i1} \mid X_{i1}, X_{i2}) = E(Y_{i1} \mid X_{i1})$ doesn't hold, (8) doesn't have expectation zero, and hence LMM will produce biased $\hat\beta_0$ and $\hat\beta_1$.

An alternative explanation: (7) describes the joint distribution of (Y_{i1}, Y_{i2}) conditional on (X_{i1}, X_{i2}), which implicily assumes condition i) in the theorem. This assumption does not make scientific sense in our setting, because our X_{i2} could depend on Y_{i1}.

[Note that in LMM with time-varying covariate, because X_{i2} can depend on Y_{i1}, the assumption that u_i and ϵ_{it} are independent of X_i is also questionable.]

1.3 Why this matters in mobile health

In a mobile health study like a micro-randomized trial, A_{it}, the indicator of treatment for person i at time t is time-varying. Thus if part of X_{it} includes A_{it}, then assumption i) in the theorem is hard to believe (which means we have to use GEE with independence working correlation). In particular, assumption such as $E(Y_{i1} \mid X_{i1}, X_{i2}) = E(Y_{i1} \mid X_{i1})$ typically doesn't hold, because A_{i1} could impact both Y_{i1} and X_{i2}, which makes Y_{i1} dependent on X_{i2} even after adjusting for X_{i1}.

Reference

Pepe, M. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. Communications in Statistics-Simulation and Computation **23**, 939–951.