

4. Application to Longitudinal Data: Part II

4.1. Iterative WLS (I-WLS)

- If V is known, an efficient estimator of β is the best linear unbiased estimator (BLUE):

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y.$$

- On the other hand, if β is known, V may be easily estimated.

Example 4.1 (balanced data). $E(y_i) = X_i\beta$, $\text{Var}(y_i) = V_0$, $1 \leq i \leq m$. If β is known, V_0 can be estimated by

$$\hat{V}_0 = \frac{1}{m} \sum_{i=1}^m (y_i - X_i\beta)(y_i - X_i\beta)'$$

However, both are unknown in practice.

Solution: Iteration between the two steps.

An extension of I-WLS to unbalanced data under a semiparametric model will be discussed in Lecture 5.

4.2. Binary responses & counts

Example 4.2 (Seizure counts).

- Thall & Vail (1990, table 2).
- 59 epileptics randomized to a new drug or a placebo.
- Multivariate response variable consisted of a baseline seizures counts as well as seizure counts during the two-weeks before each of 4 clinic visits.

- Methods of generalized linear mixed models (GLMM) may be used to analyze such longitudinal data.

- Some issues

- a. Computation of the MLE

- b. Robustness (e. g., serial correlation)

4.3. Generalized estimating equations (GEE)

- Estimating functions: Another look at the Gauss-Markov theorem

An estimating function is a function of y , a vector of responses, and θ , a vector of parameters, denoted by $g(y, \theta)$, such that

$$E_{\theta}\{g(y, \theta)\} = 0$$

for every θ .

The equation $g(y, \theta) = 0$ is then called an *estimating equation* (for θ).

First assume that y_1, \dots, y_n are indep. Let $y = (y_i)_{1 \leq i \leq n}$. Suppose that $E(y_i) = \theta$, a scalar. Let \mathcal{G} denote the class of estimating functions of the form

$$g(y, \theta) = \sum_{i=1}^n a_i(\theta)(y_i - \theta),$$

where a_i 's are differentiable with $\sum_i a_i(\theta) \neq 0$.

Then, an extension of the G-M theorem can be stated as follows (Godambe 1991).

G-M Theorem. If $\text{var}(y_i) = \sigma^2$, $1 \leq i \leq n$, $g^* = \sum_{i=1}^n (y_i - \theta)$ is an optimal estimating function within \mathcal{G} and the corresponding estimating equation gives \bar{y} , the sample mean, as the estimator of θ .

- Optimality in what sense?

The optimality was introduced by Godambe (1960).

Idea: (i) $g(y, \theta)$ needs to be as close to zero as possible, if θ is the true parameter. Then, by the unbiasedness, one needs to minimize $\text{var}(g)$. (ii) On the other hand, to distinguish the true θ from a false one, it makes sense to maximize $\partial g / \partial \theta$, or the absolute value of its expected value.

When both (i) and (ii) are put on the same scale, the two criteria can be combined by considering

$$\text{var}(g)/\{E(\partial g/\partial\theta)\}^2.$$

- Another look at Godambe's optimality

Consider a multivariate version of the estimating function. Suppose that y is associated with a vector x of explanatory variables. Suppose that $E(y) = \mu = \mu(x, \theta)$. Let $\dot{\mu} = \partial\mu/\partial\theta'$.

Consider the following class of vector-valued estimating functions $\mathcal{H} = \{G = A(Y - \mu)\}$, where $A = A(A, \theta)$.

Theorem 4.1. Suppose that $V = \text{Var}(y)$ is known. Then, the optimal estimating function within \mathcal{H} is given by $G^* = \dot{\mu}'V^{-1}(Y - \mu)$, that is, with $A = A^* = \dot{\mu}'V^{-1}$.

Here the optimality is in the similar sense that G^* maximizes, in the partial order of nonnegative definite matrices (i. e., $A \leq B$ if $B - A$ is nonnegative definite), the generalized information criterion:

$$\mathcal{I}(G) = \{\mathbf{E}(\dot{G})\}' \{\mathbf{E}(GG')\}^{-1} \{\mathbf{E}(\dot{G})\}.$$

It is easy to see that $\mathcal{I}(G)$ equals to the Fisher information matrix when G is the score function of a likelihood.

Also, $\mathcal{I}(G)$ is the reciprocal of Godambe's criterion, $\text{var}(g)/\{\mathbf{E}(\partial g/\partial \theta)\}^2$, in the univariate case.

- Application to longitudinal data

Suppose that $E(y_i) = \mu_i = g(X_i, \beta)$. Then, the optimal estimating equation is given by

$$\sum_{i=1}^n \dot{\mu}_i' V_i^{-1} (y_i - \mu_i) = 0,$$

where $V_i = \text{Var}(y_i)$. Here it is assumed that V_i , $1 \leq i \leq n$ are known.

In practice, however, the V_i 's are unknown. Liang & Zeger (1986) proposed to use “working covariance matrices” $V_{w,i}$ instead of V_i .

Alternatively, V_i may be replaced by \hat{V}_i , a consistent estimator.

- Asymptotic properties of GEE estimator

Under mild conditions,

- (i) the GEE estimator, with V_i replaced by $V_{w,i}$ ($1 \leq i \leq n$), is consistent;
- (ii) if V_i is replaced by \hat{V}_i , $1 \leq i \leq n$, the resulting estimator is asymptotically as efficient as the GEE estimator with the true V_i 's.

- Iterative estimating equations - Lecture 5.

Example 4.3 (Mothers' stress & children's morbidity).

- Zeger & Liang (1986).
- 167 mothers with infants between ages of 18 months and 5 years.
- Each mother asked to keep diary on whether her child was ill and her own stressness.
- First 9 days of the diaries were used.

4.4. Informative missing data

- The probability that an outcome of the i th subject is missing may depend on X_i as well as the previous outcomes.

It is known that a fully parametric likelihood method still provides valid inference about β , if the model is correctly specified and the prob. of missing at time t does not depend on the current and future data (i. e., at times $s \geq t$; Rubin 1976).

However, with incomplete data, the likelihood-based methods can be sensitive to model misspecification and, even with complete data, computationally challenging.

Furthermore, the GEE method is valid only under the stronger assumption that the data is missing completely at random, i. e., the non-response process is indep. of the data process.

- Modelling the probability of missing

Let $R_{it} = 1$ if subject i is observed at time t , and $R_{it} = 0$ otherwise. Robins, Rotnitzky & Zhao (1995) assumed

$$\begin{aligned} & P(R_{it} = 1 | R_{i\underline{t-1}} = 1, y_{i1}, \dots, y_{i\underline{t-1}}, y_{it}) \\ & = P(R_{it} = 1 | R_{i\underline{t-1}} = 1, y_{i1}, \dots, y_{i\underline{t-1}}). \end{aligned}$$

Furthermore, they proposed a logistic model for the right side of the above equation that depends on a vector of additional parameters, ψ .

They proposed a weighted GEE estimator that satisfies

$$\sum_{i=1}^n D_i(\beta) w_i(\hat{\psi})(y_i - \mu_i) = 0,$$

where $\mu_i = E(y_i|X_i) = g(X_i, \beta)$ under a semi-parametric regression model, $D_i(\beta)$ is similar to the factor $\dot{\mu}'_i V_i^{-1}$ in the GEE, where $V_i = \text{Var}(y_i|X_i)$, and $w_i(\psi)$ is a weight depending on ψ .

The weight is computed under the (logistic) model for the missing process R_{it} .

ψ is estimated by maximizing a partial likelihood.

See Robins, Rotnitzky & Zhao (1995) for detail.