Iterative Estimating Equations

Jiming Jiang

University of California, Davis, CA
March 2005

Joint with Prof. YouGan Wang, National University of Singapore and Yihui Luan, Shangdong University, PRC

1. Introduction

- Motivation: A numerical example

  Hand & Crowder (1996) presented data regarding hip replacements of 30 patients. Each patient was measured four times, once before the operation and three times after, for twelve variables including hematocrit, calcium, etc.

- Features of the data:

(i) The responses are longitudinal and likely to be correlated.

(ii) Most of the patients have at least one missing observation for all variables.

- Generalized estimating equations (GEE)

  Discussed in lecture 4. Recall that the GEE is given by

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

where the mean μ_i depends on the parameter of interest, β, and $\dot{\mu}_i = \partial\mu_i/\partial\beta.$

However, the V_i's are unknown in practice.

2. A semiparametric model

Consider a follow-up study over times $t_1, \ldots, t_b$.

Let $Y_i = (Y_{ij})_{j \in J_i}$ be the vector of responses collected from subject i at the visit times, t_j, $j \in J_i \subset J = \{1, \ldots, b\}$.

Let $X_{ij} = (X_{ijl})_{1 \leq l \leq p}$ represent a vector of explanatory variables associated with Y_{ij}.

Write $X_i = (X_{ij})_{j \in J_i}$.

The following assumptions are made:

(i) (X_i, Y_i), $i = 1, \ldots, n$ are independent.

(ii) $\mathsf{E}(Y_{ij}|X_i) = g_j(X_i, \beta)$, where β is a vector of unknown regression coefficients and $g_j(\cdot, \cdot)$ is a known function.

(iii) $V_i = \mathsf{Var}(Y_i|X_i)$ such that the # of different covariances in V_i, $1 \leq i \leq n$ are bounded.

Assumption (iii) is essential for consistent estimation of the covariances.

- Examples

Example 1 (balanced data). $J_i = J$, $1 \le i \le n$, so that $V_i = V_0$, which is $b \times b$, $1 \le i \le n$.

Example 2. $Y_{ij} = x'_{ij}\beta + u_i + w_{ij} + e_{ij}$, where u_i is a subject-specific random effect; w_{ij} corresponds to a serial correlation; and e_{ij} is a measurement error.

Assume $u_i \sim N(0, \sigma_u^2)$, $w_{ij} \sim$ AR(1):

$$w_{ij} = \phi w_{i\underline{j-1}} + z_{ij},$$

where $\phi \in (0, 1)$, $z_{ij} \sim N(0, \sigma_w^2(1 - \phi^2))$; $e_{ij} \sim N(0, \sigma_e^2)$; and u, w and e uncorrelated.

Then, we have

$$\begin{aligned} v_{ijk} &= \text{cov}(Y_{ij}, Y_{ik}|X_i) \\ &= \sigma_u^2 + \sigma_w^2 \phi^{|j-k|} + \sigma_e^2 \delta_{j,k}, \end{aligned}$$

$1 \le i \le n$, where $\delta_{j,k} = 1$ if $j = k$ and $\delta_{j,k} = 0$ otherwise.

Problem of main interest: How to (efficiently) estimate β?

3. Iterative estimating equations (IEE)

a. If the V_i's are known, β can be (efficiently) estimated by the GEE:

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

b. If β is known, V_i's can be estimated by the method of moments. Let $I(j,k,l)$ denote the set of indexes $1 \le i \le n$ such that $v_{ijk} = v(j,k,l)$, and $n(j,k,l) = |I(j,k,l)|$. Define

$$\begin{aligned}\widehat{v}(j,k,l) &= \frac{1}{n(j,k,l)} \sum_{i \in I(j,k,l)} \{Y_{ij} - g_j(X_i,\beta)\} \\ &\quad \times \{Y_{ik} - g_k(X_i,\beta)\}.\end{aligned}$$

c. If β and V_i's are both unknown, iterate between a and b, starting with the identity V_i's.

4. Linear convergence

- A term from numerical analysis

An iterative algorithm that results in $x^{(m)}$, $m = 1, 2, \ldots$ converges linearly to a limit x^*, if there is $0 < \rho < 1$ such that

$$\sup_{m \geq 1} \{|x^{(m)} - x^*|/\rho^m\} < \infty.$$

- Linear convergence of IEE

Let v denote the vector of different covariances in V_i, $1 \leq i \leq n$ and $R = \mathrm{dim}(v)$. Recall $p = \mathrm{dim}(\beta)$. Denote the IEE sequence by

$$\hat{\beta}^{(m)}, v^{(m)}, m = 1, 2, \ldots$$

If the sequence converges, the limit, denoted by (β^*, v^*), is call the IEE estimator, or IEEE.

Theorem 1. Under mild conditions,

$$\mathrm{P}(\text{IEE converges}) \to 1$$

as $n \to \infty$. Furthermore, $\forall 0 < \eta < 1/(p \vee R)$, we have

$$\mathrm{P}\left[\sup_{m\geq 1}\{|\hat{\beta}^{(m)} - \beta^*|/(p\eta)^{m/2}\} < \infty\right] \to 1,$$

$$\mathrm{P}\left[\sup_{m\geq 1}\{|\hat{v}^{(m)} - v^*|/(R\eta)^{m/2}\} < \infty\right] \to 1$$

as $n \to \infty$.

5. Asymptotic behavior of IEEE

Theorem 2. Under the conditions of Theorem 1, the IEEE β^* and v^* are both consistent.

Theorem 3. Under slightly stronger conditions the IEEE β^* satisfies

$$\sqrt{n}(\beta^* - \beta) \to 0$$

in probability. Therefore, β^* is asymptotically as efficient as the GEE estimator with known (true) V_i's.

6. Special case

Consider the (classic) linear models for longitudinal data analysis, in which X_i is considered fixed such that $\mathrm{E}(Y_i) = X_i\beta$, $i = 1, \ldots, n$. Let $X = (X_i)_{1 \le i \le n}$, $Y = (Y_i)_{1 \le i \le n}$ and $V = \mathrm{diag}(V_1, \ldots, V_n)$.

• In this case, the GEE has an explicit solution known as the BLUE (best linear unbiased estimator):

$$\begin{aligned}\widehat{\beta}_{\text{BLUE}} &= (X'V^{-1}X)^{-1}X'V^{-1}Y \\ &= \left(\sum_{i=1}^{m} X_i'V_i^{-1}X_i\right)^{-1} \sum_{i=1}^{m} X_i'V_i^{-1}Y_i \, ,\end{aligned}$$

which is the solution to the following minimization problem

$$\min_{\beta}\{(Y - X\beta)'W(Y - X\beta)\}$$

with $W = V^{-1}$, the optimal W. So IEE becomes iterative WLS, or I-WLS, starting with the OLS.

7. A simulation study

Consider a special case of Example 2:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + w_{ij} + e_{ij},$$

$i = 1, \ldots, 100$, $j \in J_i = \{1, 3, 5\}$ or $\{2, 4\}$.

- x_{ij}'s are generated from $N(0, 1)$ and then fixed throughout the simulation.

Case 1: $\sigma_u^2 = 1$, $\sigma_w^2 = 9$, $\sigma_e^2 = 1$ and $\phi = 0.9$.

Case 2: $\sigma_u^2 = 9$, $\sigma_w^2 = 25$, $\sigma_e^2 = 1$ and $\phi = 0.99$.

Scenario 1: Normality, parametric covariance.

Scenario 2: No normality, param. covariance (i. e., $u \sim$ *centralized exponential*, w, $e \sim$ normal).

Scenario 3: No normality & param. covariance (i. e., $w \sim$ MA instead of AR).

In all cases $\beta_0 = 0.5$, $\beta_1 = 1.0$.

Results:

Table 1: **Number of steps to converge.**

Case	2	3	4	5	6	7	8	9	10	11
1.1	0	4.2	34.1	47.8	11.6	2.1	0.2	0	0	0
1.2	0	0.6	19.4	45.9	25.5	7.1	1.0	0.3	0.1	0.1
2.1	0.1	3.9	42.1	38.9	12.5	2.1	0.2	0.1	0.1	0
2.2	0	1.0	21.5	45.7	23.3	6.8	1.3	0.4	0	0
3.1	0	5.8	39.9	39.3	12.8	2.0	0.2	0	0	0
3.2	0	1.4	25.2	43.1	22.9	5.9	1.1	0.3	0.1	0

Table 2: **Simulated covariance matrices.**

Case	$\hat{V}_{\mathrm{OLS}}$		$\hat{V}_{\mathrm{I-WLS}}$		$\hat{V}_{\mathrm{MLE}}$		V_{BLUE}	
1.1	9.80	0.11	10.36	0.36	10.00	0.48	9.30	0.11
	0.11	5.51	0.36	2.09	0.48	1.87	0.11	2.16
1.2	36.41	0.61	36.41	-0.10	35.57	-0.08	34.10	0.06
	0.61	17.41	-0.10	1.30	-0.08	1.23	0.06	1.29
2.1	9.87	0.14	9.94	0.07	9.19	0.02	9.30	0.11
	0.14	5.35	0.07	2.25	0.02	2.08	0.11	2.16
2.2	38.74	1.78	38.48	-0.01	34.54	0.15	34.10	0.06
	1.78	15.86	-0.01	1.31	0.15	1.11	0.06	1.29
3.1	7.14	0.44	7.25	0.41	6.77	0.45	7.15	0.23
	0.44	5.34	0.41	4.01	0.45	4.13	0.23	3.74
3.2	25.48	0.26	26.22	0.22	24.73	0.46	25.40	0.63
	0.26	17.55	0.22	10.80	0.46	12.35	0.63	9.61

- Summary of results

(i) In most cases the I-WLS algorithm converges in 4 to 6 steps.

(ii) Both I-WLS and MLE significantly outperform OLS.

(iii) The performance of I-WLS is very close to that of BLUE.

(iv) As σ_u^2, σ_w^2 and ϕ increase, the difference between OLS and I-WLS becomes larger.

(v) In the Scenarios 1 and 2, MLE slightly ourperform I-WLS, while in Scenario 3 I-WLS slightly ourperform MLE.

8. Data analysis

Back to the numerical example discussed at the beginning.

8.1. Analysis of hematocrit data

- I-WLS converged in 7 steps.

- Observations: The coefficients β_1, β_3, β_4, β_5 and β_6 are significant and the rest of the coefficients are insignificant. These are consistent with the findings of Hand and Crowder with the only exception being β_6.

Results

Table 3: **Estimates for hematocrit.** *The first row are I-WLS estimates correspond to, from left to right, intercept, sex, occasions (three), sex by occasion interaction (three), age, and age by sex interaction; the second row are estimated standard errors corresponding to the I-WLS estimates; the third row are the Gaussian maximum likelihood estimates were obtained by Hand and Crowder (1996).*

3.19	0.08	0.65	-0.34	-0.21	0.12	-0.051	-0.051	0.033	-0.001
0.39	0.14	0.06	0.06	0.07	0.06	0.061	0.066	0.058	0.021
3.28	0.21	0.65	-0.34	-0.21	0.12	-0.050	-0.048	0.019	-0.020

8.2 Analysis of calcium data

- I-WLS converged in 13 steps.

- Observations: Except for β_0, β_3 and β_4, all the coefficients are not significant (at 5% level). In particular, there seems to be no difference in terms of sex and age. Also, the recovery of calcium after the operation seems to be a little quicker than that of hematocrit, because β_5 is no longer significant.

Results

Table 4: **Estimates for calcium.** *The first row are I-WLS estimates correspond to, from left to right, intercept, sex, occasions (three), sex by occasion interaction (three), age, and age by sex interaction; the second row are estimated standard errors corresponding to the estimates.*

20.1	0.93	1.32	-1.89	-0.13	0.09	0.17	-0.15	0.19	-0.12
1.3	0.57	0.16	0.13	0.16	0.16	0.13	0.16	0.19	0.09