# Asymptotic variance and small sample correction incorporating availability For estimator_EMEE_alwaysCenterA

Tianchen Qian

2019.12.19

## 1 Estimating equations

Data is $(X_1, A_1, Y_2, \cdots, X_T, A_T, Y_{T+1})$. History $H_t := (\bar{Y}_t, \bar{X}_t, \bar{A}_{t-1}) = (X_1, A_1, Y_2, \ldots X_{t-1}, A_{t-1}, Y_{t-1}, X_t)$. Availability at the $t$-th decision point is $I_t$. The randomization probability is $p_t$ which does not depend on $H_t$.

The marginal excursion effect is defined as

$$\begin{aligned}\text{MEE}(t) &= \log \frac{E\{E(Y_{t+1} \mid H_t, A_t = 1, I_t = 1) \mid I_t = 1\}}{E\{E(Y_{t+1} \mid H_t, A_t = 0, I_t = 1) \mid I_t = 1\}} \\ &= \log \frac{E(Y_{t+1} \mid A_t = 1, I_t = 1)}{E(Y_{t+1} \mid A_t = 0, I_t = 1)}.\end{aligned}$$

(The second equality holds in this setting because $p_t$ does not depend on $H_t$.) We assume that

$$\text{MEE}(t) = f(t)^T \beta^*$$

for some unknown $p$-dimensional $\beta^*$. The estimating equation is

$$\mathbb{P}_n \sum_{t=1}^{T} I_t \left\{ e^{-(A_t - p_t) f(t)^T \beta} Y_{t+1} - e^{g(t)^T \alpha} \right\} \begin{bmatrix} g(t) \\ (A_t - p_t) f(t) \end{bmatrix} = 0,$$

where $g(t)$ is a $q$-dimensional feature vector of $t$ such that $p_t f(t)$ is in the linear span of $g(t)$. So we have $\dim(\alpha) = q$, $\dim(\beta) = p$.

Define $D = \begin{bmatrix} D_1^T \\ D_2^T \\ \vdots \\ D_T^T \end{bmatrix}$ a $T \times (p+q)$ matrix, where $D_t = \begin{bmatrix} g(t) \\ (A_t - p_t) f(t) \end{bmatrix}$ is $(p+q) \times 1$. Define $r(\alpha, \beta) = (r_1, \ldots, r_T)^T$ a $T \times 1$ (column) vector, where $r_t(\alpha, \beta) = e^{-(A_t - p_t) f(t)^T \beta} Y_{t+1} - e^{g(t)^T \alpha}$. Let $I$ be the diagonal matrices with the $t$-th diagonal entry being $I_t$. The above estimating equation can be rewritten as $\mathbb{P}_n D^T I r(\alpha, \beta) = 0$.

Notation convention: For a column vector $f = (f_1, \dots, f_k)^T$, its derivative is defined as

$$\frac{\partial f}{\partial \theta^T} = \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1} & \cdots & \frac{\partial f_1}{\partial \theta_k} \\ \vdots & & \vdots \\ \frac{\partial f_k}{\partial \theta_1} & \cdots & \frac{\partial f_k}{\partial \theta_k} \end{bmatrix},$$

and for a scalar $r$, its derivative is defined as $\frac{\partial r}{\partial \theta} = (\frac{\partial r}{\partial \theta_1}, \dots, \frac{\partial r}{\partial \theta_k})^T$ (column vector), and $\frac{\partial r}{\partial \theta^T} = (\frac{\partial r}{\partial \theta_1}, \dots, \frac{\partial r}{\partial \theta_k})$ (row vector).

## 2 Asymptotic variance

Let $\theta := (\alpha, \beta)$. Suppose $(\alpha^*, \beta^*)$ satisfies the population version of the estimating equation, i.e., $E_{\text{truth}}\{D^T I r(\alpha^*, \beta^*)\} = 0$ ($\beta^*$is the $\beta^*$ in MEE($t$), which is shown in Sec 1.2). Then by Theorem 5.21 in vdV, under regularity conditions we have

$$\sqrt{n}\left(\begin{bmatrix}\hat{\alpha} \\ \hat{\beta}\end{bmatrix} - \begin{bmatrix}\alpha^* \\ \beta^*\end{bmatrix}\right) = -V_{\theta^*}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\theta^*}(\text{Data}_i) + o_P(1), \tag{1}$$

where $\theta^* = (\alpha^*, \beta^*)$, $\psi_\theta = D^T I r(\alpha, \beta)$, $V_{\theta^*} = E_{\text{truth}}\left(\frac{\partial \psi_\theta}{\partial \theta^T} |_{\theta^*}\right)$. This implies that the asymptotic variance of $(\hat{\alpha}, \hat{\beta})$ is

$$\begin{aligned}
\text{Var}\left(\begin{bmatrix}\hat{\alpha} \\ \hat{\beta}\end{bmatrix}\right) &\approx \frac{1}{n} V_{\theta^*}^{-1} E_{\text{truth}}(\psi_{\theta^*}\psi_{\theta^*}^T)(V_{\theta^*}^{-1})^T. \\
&= \frac{1}{n} V_{\theta^*}^{-1} E_{\text{truth}}\{D^T I r(\alpha^*, \beta^*) r(\alpha^*, \beta^*)^T I D\}(V_{\theta^*}^{-1})^T \\
&= \frac{1}{n} M^{-1} \Sigma (M^{-1})^T,
\end{aligned}$$

where

$$\begin{aligned}
M &= V_{\theta^*}, \\
\Sigma &= E_{\text{truth}}\{D^T I r(\alpha^*, \beta^*) r(\alpha^*, \beta^*)^T I D\}.
\end{aligned}$$

An estimator of the **asymptotic variance** is $\frac{1}{n} M_n^{-1} \Sigma_n (M_n^{-1})^T$, where

$$\begin{aligned}
\Sigma_n &= \mathbb{P}_n\{D^T I r(\hat{\theta}) r(\hat{\theta})^T I D\}, \\
M_n &= \mathbb{P}_n D^T I \frac{\partial r(\hat{\theta})}{\partial \theta^T}.
\end{aligned}$$

## 3 Small sample correction with hat matrix

For small sample correction as in Mancl and DeRouen (2001), we replace $\Sigma_n$ in the variance estimator by $\tilde{\Sigma}_n$:

$$\tilde{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n}\left[D_i^T I_i \left\{Id_i - H_{ii}(\hat{\theta})\right\}^{-1} r_i(\hat{\theta})r_i(\hat{\theta})^T \left\{Id_i - H_{ii}(\hat{\theta})\right\}^{-1,T} I_i D_i\right],$$

where $Id_i$ is the identity matrix (the same dimension as $H_{ii}(\hat{\theta})$), and

$$H_{ii}(\theta) = \frac{1}{n}\frac{\partial r_i(\theta)}{\partial \theta^T} M_n^{-1} D_i^T.$$

Derivation of this small sample correction formula is in Section 5 of "2019.02.06 - asymptotic variance and small sample correction incorporating availability (new).pdf".

## 4 Asymptotic variance and small sample correction, computing out each terms in order to program it in R

Some of the sections are copied from "2019.12.18 - derivation with only time in covariates.lyx".

### 4.1 Derivative $\frac{\partial r(\theta)}{\partial \theta^T}$

Recall that $r_t(\alpha,\beta) = e^{-(A_t - p_t)f(t)^T\beta}Y_{t+1} - e^{g(t)^T\alpha}$. We have

$$\begin{aligned}\frac{\partial r_t}{\partial \theta^T} &= \left[\frac{\partial r_t}{\partial \alpha^T}, \frac{\partial r_t}{\partial \beta^T}\right] \\ &= \left[-e^{g(t)^T\alpha}g(t)^T, -e^{-(A_t-p_t)f(t)^T\beta}Y_{t+1}(A_t - p_t)f(t)^T\right].\end{aligned}$$

So

$$\frac{\partial r(\theta)}{\partial \theta^T} = \begin{bmatrix} -e^{g(1)^T\alpha}g(1)^T & -e^{-(A_1-p_1)f(1)^T\beta}Y_2(A_1-p_1)f(1)^T \\ \vdots & \vdots \\ -e^{g(T)^T\alpha}g(T)^T & -e^{-(A_T-p_T)f(T)^T\beta}Y_{T+1}(A_T-p_T)f(T)^T \end{bmatrix}.$$

## 4.2 Analytic calculation of $M_n$

We have

$$
\begin{aligned}
M_n &= \mathbb{P}_n D^T I \frac{\partial r(\hat{\theta})}{\partial \theta^T}. \\
&= \mathbb{P}_n \begin{bmatrix} g(1) & \cdots & g(T) \\ (A_1 - p_1) f(1) & \cdots & (A_T - p_T) f(T) \end{bmatrix}_{(q+p)\times T} \begin{bmatrix} I_1 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & I_T \end{bmatrix}_{T\times T} \\
&\times \begin{bmatrix} -e^{g(1)^T \hat{\alpha}} g(1)^T & -e^{-(A_1 - p_1) f(1)^T \hat{\beta}} Y_2 (A_1 - p_1) f(1)^T \\ \vdots & \vdots \\ -e^{g(T)^T \hat{\alpha}} g(T)^T & -e^{-(A_T - p_T) f(T)^T \hat{\beta}} Y_{T+1} (A_T - p_T) f(T)^T \end{bmatrix}_{T\times(q+p)} \\
&= \mathbb{P}_n \sum_{t=1}^T I_t \begin{bmatrix} g(t) \\ (A_t - p_t) f(t) \end{bmatrix} \left[ -e^{g(t)^T \hat{\alpha}} g(t)^T, -e^{-(A_t - p_t) f(t)^T \hat{\beta}} Y_{t+1} (A_t - p_t) f(t)^T \right] \\
&= -\mathbb{P}_n \sum_{t=1}^T I_t \begin{bmatrix} e^{g(t)^T \hat{\alpha}} g(t) g(t)^T & e^{-(A_t - p_t) f(t)^T \hat{\beta}} Y_{t+1} (A_t - p_t) g(t) f(t)^T \\ e^{g(t)^T \hat{\alpha}} (A_t - p_t) f(t) g(t)^T & e^{-(A_t - p_t) f(t)^T \hat{\beta}} Y_{t+1} (A_t - p_t)^2 f(t) f(t)^T \end{bmatrix}.
\end{aligned}
$$

## 4.3 Analytic calculation of $\Sigma_n$

$$
\begin{aligned}
\Sigma_n &= \mathbb{P}_n \{ D^T I r(\hat{\alpha}, \hat{\beta}) r(\hat{\alpha}, \hat{\beta})^T I D \} \\
&= \mathbb{P}_n \left\{ \left( \sum_{t=1}^T D_t I_t r_t(\hat{\alpha}, \hat{\beta}) \right) \left( \sum_{s=1}^T D_s I_s r_s(\hat{\alpha}, \hat{\beta}) \right)^T \right\} \\
&= \mathbb{P}_n \left\{ \sum_{t=1}^T \sum_{s=1}^T I_t I_s r_t(\hat{\alpha}, \hat{\beta}) r_s(\hat{\alpha}, \hat{\beta}) D_t D_s^T \right\} \\
&= \mathbb{P}_n \left\{ \sum_{t=1}^T \sum_{s=1}^T I_t I_s r_t(\hat{\alpha}, \hat{\beta}) r_s(\hat{\alpha}, \hat{\beta}) \begin{bmatrix} g(t) g(t)^T & (A_s - p_s) g(t) f(s)^T \\ (A_t - p_t) f(t) g(s)^T & (A_t - p_t)(A_s - p_s) f(t) f(s)^T \end{bmatrix} \right\}
\end{aligned}
$$

## 4.4 Analytic calculation of $\tilde{\Sigma}_n$

We have

$$
H_{ii}(\hat{\theta}) = \frac{1}{n} \frac{\partial r_i(\hat{\theta})}{\partial \theta^T} M_n^{-1} D_i^T
$$

where each term are explicitly written out as above.

Hence, we have

$$
\tilde{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n \left[ D_i^T I_i \left\{ I d_i - H_{ii}(\hat{\theta}) \right\}^{-1} r_i(\hat{\theta}) r_i(\hat{\theta})^T \left\{ I d_i - H_{ii}(\hat{\theta}) \right\}^{-1,T} I_i D_i \right].
$$