Log linear estimating equation for primary analysis of Sense2Stop (asymptotic theory)

Tianchen Qian

2020.06.22

1 Notation

- Each minute within a user's 10-hour day is a decision point. Over the course of 10 days $t = 1, \ldots, T$ where $T = 600 \times 10 = 6000$. All the following variables are defined for $t = 1, \ldots, T$.
- Outcome: $Y_t \in \{1, 2, 3\}$, where 1 is "detected stressed and not detected physically active", 2 is "detected physically active", and 3 is "not detected stressed and not detected physically active". We let $K = 3$ be the reference category.
- Availability indicator: I_t. See the Overleaf write up for how this is defined.
- Treatment indicator: A_t.
- Missing data indicator: M_t. $M_t = 1$ if Y_t is observed, and $M_t = 0$ if Y_t is unobserved.
- Time-varying covariates: Z_t.
- Stratification variable for this stratified-MRT: X_t. $X_t = 1$ if $Y_t = 1$; $X_t = 0$ if $Y_t = 3$; when $Y_t = 2$ or missing, X_t is undefined, and in this case $I_t = 0$. X_t is the moderator in the primary analysis.

Temporal ordering of the variables (those in {} are observed at the same time):

$$\ldots, \{Z_t, M_t, Y_t, M_t, X_t, I_t\}, A_t, \{Z_{t+1}, M_{t+1}, Y_{t+1}, M_{t+1}, X_{t+1}, I_{t+1}\}, A_{t+1}, \ldots$$

History H_t is everything observed prior to A_t.

- Randomization probability: $p_t(H_t) = P(A_t = 1 \mid H_t)$.
- The proximal outcome following decision point t is: $Y_{t+1}, Y_{t+2}, \ldots, Y_{t+120}$.
- We use O to denote the data observed for a generic individual.

2 Estimating equation

Note: The estimating equations are the same as what I sent last Friday. I am simply revising the notation so that the form of the variance is simplified.

The treatment effect model is (for each of $k \in \{1, 2\}$ and $x \in \{0, 1\}$)

$$\log \frac{P\left\{Y_{t+m}(\bar{A}_{t-1}, 1, \bar{0}_{m-1}) = k \mid I_t(\bar{A}_{t-1}) = 1, X_t(\bar{A}_{t-1}) = x\right\}}{P\left\{Y_{t+m}(\bar{A}_{t-1}, 0, \bar{0}_{m-1}) = k \mid I_t(\bar{A}_{t-1}) = 1, X_t(\bar{A}_{t-1}) = x\right\}} = \beta_{kx} \quad \text{for all } t \text{ and } m. \tag{1}$$

Let $\beta = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})$ denote the parameter of interest.

Assume regularity conditions so that matrices are invertible and/or have finite entries where relevant.

2.1 Estimating numerator of the weights

Let $p(x; \rho)$ be the numerator of the weights for all decision points with $X_t = x$ and $I_t = 1$. Suppose $\hat{\rho}$ is estimated using an estimating equation $\mathbb{P}_n U_N(O; \rho) = 0$. (Subscript "N" for "Numerator".) Assume that, for a finite value ρ^*, $E[U_N(O; \rho^*)] = 0$ and that

$$\sqrt{n}(\hat{\rho} - \rho^*) = -\left\{E\left[\frac{\partial U_N(O; \rho^*)}{\partial \rho}\right]\right\}^{-1} \sqrt{n}\mathbb{P}_n U_N(O; \rho^*) + o_P(1).$$

The above assumptions hold in the following special case: Suppose $p(x; \rho) = \rho_0 1_{x=0} + \rho_1 1_{x=1}$ and let

$$U_N(O; \rho) = \begin{bmatrix} \sum_{t=1}^T I_t 1_{X_t=0} \{p_t(H_t) - \rho_0\} \\ \sum_{t=1}^T I_t 1_{X_t=1} \{p_t(H_t) - \rho_1\} \end{bmatrix}.$$

This would result in the following:

$$p(x; \hat{\rho}) = \frac{\mathbb{P}_n \sum_{t=1}^T I_t 1_{X_t=x} p_t(H_t)}{\mathbb{P}_n \sum_{t=1}^T I_t 1_{X_t=x}}. \tag{2}$$

Let the marginalization weight be

$$W_t(\rho) = \left\{\frac{p(X_t; \rho)}{p_t(H_t)}\right\}^{A_t} \left\{\frac{1 - p(X_t; \rho)}{1 - p_t(H_t)}\right\}^{1-A_t} \times \prod_{m=1}^{120} \frac{1(A_m = 0)}{1 - p_m(H_j)} \times \frac{1}{p(X_t; \rho)\{1 - p(X_t; \rho)\}}.$$

2.2 First stage estimating quation

The working model on the control part is (for each of $k \in \{1, 2\}$)

$$P\left\{Y_{t+m}(\bar{A}_{t-1}, 0, \bar{0}_{m-1}) = k \mid H_t, I_t = 1\right\} \approx e^{Z_t^T \alpha_k} \quad \text{for all } t \text{ and } m.$$

Suppose $\dim(\alpha_k) = q$. I am currently using different α_k for $k = 1, 2$ to make the estimating equation easier to write. In practice we may want to instead set $\alpha_1 = \alpha_2$ and revise the estimating equation for α to save degrees of freedom and possibly improve power.

Let $\theta = (\alpha_1^T, \alpha_2^T, \beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})^T$. Define the feature matrix

$$D_t(\rho) = \begin{bmatrix} Z_t & 0_{q\times 1} \\ 0_{q\times 1} & Z_t \\ \{A_t - p(X_t;\rho)\}\, X_t & 0 \\ 0 & \{A_t - p(X_t;\rho)\}\, X_t \\ \{A_t - p(X_t;\rho)\}\,(1 - X_t) & 0 \\ 0 & \{A_t - p(X_t;\rho)\}\,(1 - X_t) \end{bmatrix},$$

and define the blipped-down residuals

$$R_{ktm}(\theta) = e^{-A_t X_t \beta_{k1} - A_t(1-X_t)\beta_{k0}} 1_{Y_{t+m}=k} - e^{Z_t^T \alpha_k}.$$

The first-stage estimating function for θ is

$$U_1(O;\theta,\hat{\rho}) = \sum_{t=1}^{T} \sum_{m=1}^{120} I_t W_t(\hat{\rho}) M_{t+m} D_t(\hat{\rho}) \begin{bmatrix} R_{1tm}(\theta) \\ R_{2tm}(\theta) \end{bmatrix}. \tag{3}$$

Suppose $\hat{\theta}$ solves $\mathbb{P}_n U_1(O;\theta,\hat{\rho}) = 0$.

2.3 Second stage estimating quation

Define $\pi_t^{(k)}(\theta) = e^{A_t X_t \beta_{k1} + A_t(1-X_t)\beta_{k0} + Z_t^T \alpha_k}$ for $k = 1, 2$. We construct the 2×2 variance-covariance matrix:

$$\Sigma_t(\theta) = \begin{bmatrix} \pi_t^{(1)}(\theta)\left\{1 - \pi_t^{(1)}(\theta)\right\} & -\pi_t^{(1)}(\theta)\pi_t^{(2)}(\theta) \\ -\pi_t^{(1)}(\theta)\pi_t^{(2)}(\theta) & \pi_t^{(2)}(\theta)\left\{1 - \pi_t^{(2)}(\theta)\right\} \end{bmatrix}.$$

Let $\phi = (\alpha_1^T, \alpha_2^T, \beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})^T$. The second stage estimating equation for ϕ is

$$U_2(O;\phi,\hat{\theta},\hat{\rho}) = \sum_{t=1}^{T} \sum_{m=1}^{120} I_t W_t(\hat{\rho}) M_{t+m} D_t(\hat{\rho}) \Sigma_t(\hat{\theta})^{-1} \begin{bmatrix} R_{1tm}(\phi) \\ R_{2tm}(\phi) \end{bmatrix}. \tag{4}$$

Suppose $\hat{\phi}$ solves $\mathbb{P}_n U_2(O;\phi,\hat{\theta},\hat{\rho}) = 0$. The β_{kx}'s in $\hat{\phi}$ are the final estimators.

3 Asymptotic theory

3.1 Consistency

Under regularity conditions, one can show that as $n \to \infty$,

$$\hat{\theta} \xrightarrow{P} \theta^*,$$
$$\hat{\phi} \xrightarrow{P} \phi^*, \tag{5}$$

and that the β_{kx}'s in θ^* and the β_{kx}'s in ϕ^* are equal, and both equal to the β_{kx}'s in (1). In particular, the final estimators for β_{kx} are consistent.

The proof is straightforward by using Theorem 5.9 of van der Vaart (1998). I omitted the proof here. I can fill this in if needed. Basically it involves calculating the expectation of the estimating functions and showing that they are unbiased (having expectation 0).

3.2 Asymptotic normality

Under regularity conditions, one can show that

$$\sqrt{n}(\hat{\phi} - \phi^*) \xrightarrow{d} N(0, \Sigma_\phi), \tag{6}$$

where asymptotic variance-covariance matrix, Σ_ϕ, equals

$$\Sigma_\phi = \left\{ E\left[\frac{\partial \tilde{U}_2(O; \phi^*, \theta^*, \rho^*)}{\partial \phi^T} \right] \right\}^{-1} E\left\{ \tilde{U}_2(O; \theta^*, \phi^*, \rho^*) \tilde{U}_2(O; \phi^*, \theta^*, \rho^*)^T \right\} \left\{ E\left[\frac{\partial \tilde{U}_2(O; \phi^*, \theta^*, \rho^*)}{\partial \phi^T} \right] \right\}^{-1,T} \tag{7}$$

The asymptotic variance-covariance matrix for the estimated treatment effects, $(\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})^T$, is the lower right 4×4 submatrix of Σ_ϕ. Σ_ϕ can be estimated by plugging in $\hat{\phi}, \hat{\theta}, \hat{\rho}$ and replacing expectations by empirical averages.

Terms in Σ_ϕ are defined in the next subsection.

3.3 Terms in Σ_ϕ

$$\begin{aligned}
\tilde{U}_2(O; \phi, \theta, \rho) := {} & U_2(O; \phi, \theta, \rho) \\
& - \sum_{t=1}^{T} \sum_{m=1}^{120} I_t \frac{\partial W_t(\rho)}{\partial \rho^T} \left\{ E\left[\partial U_N(O; \rho) / \partial \rho^T \right] \right\}^{-1} U_N(O; \rho) M_{t+m} D_t(\rho) \Sigma_t(\theta)^{-1} \begin{bmatrix} R_{1tm}(\phi) \\ R_{2tm}(\phi) \end{bmatrix} \\
& - \sum_{t=1}^{T} \sum_{m=1}^{120} I_t W_t(\rho) M_{t+m} \frac{\partial D_t(\rho)}{\partial \rho^T} \left(\left\{ E\left[\partial U_N(O; \rho) / \partial \rho^T \right] \right\}^{-1} U_N(O; \rho) \right) \Sigma_t(\theta)^{-1} \begin{bmatrix} R_{1tm}(\phi) \\ R_{2tm}(\phi) \end{bmatrix} \\
& - \sum_{t=1}^{T} \sum_{m=1}^{120} I_t W_t(\rho) M_{t+m} D_t(\rho) \frac{\partial \left\{ \Sigma_t(\theta)^{-1} \right\}}{\partial \theta^T} \left(\left\{ E\left[\partial \tilde{U}_1(O; \theta, \rho) / \partial \theta^T \right] \right\}^{-1} \tilde{U}_1(O; \theta, \rho) \right) \begin{bmatrix} R_{1tm}(\phi) \\ R_{2tm}(\phi) \end{bmatrix},
\end{aligned}$$

$$
\begin{aligned}
\tilde{U}_1(O;\theta,\rho) &:= U_1(O;\theta,\rho) \\
&\quad - \sum_{t=1}^{T}\sum_{m=1}^{120} I_t M_{t+m} \frac{\partial W_t(\rho)}{\partial \rho^T} \left\{ E\left[\partial U_N(O;\rho)/\partial \rho^T\right] \right\}^{-1} U_N(O;\rho) D_t(\rho) \begin{bmatrix} R_{1tm}(\theta) \\ R_{2tm}(\theta) \end{bmatrix} \\
&\quad - \sum_{t=1}^{T}\sum_{m=1}^{120} I_t M_{t+m} W_t(\rho) \frac{\partial D_t(\rho)}{\partial \rho^T} \left(\left\{ E\left[\partial U_N(O;\rho)/\partial \rho^T\right] \right\}^{-1} U_N(O;\rho) \right) \begin{bmatrix} R_{1tm}(\theta) \\ R_{2tm}(\theta) \end{bmatrix},
\end{aligned}
$$

$$
\frac{\partial W_t(\rho)}{\partial \rho} = W_t(\rho) \times (-1)^{1-A_t} \left[\frac{p(X_t;\rho)}{\{1-p(X_t;\rho)\}} \right]^{A_t} \frac{\partial \log p(X_t;\rho)}{\partial \rho^T}.
$$

Let $q = \dim(Z_t) = \dim(\alpha_1) = \dim(\alpha_2)$. $\frac{\partial D_t(\rho)}{\partial \rho^T}$ is an operator that maps $\mathbb{R}^{\dim(\rho)} \to \mathbb{R}^{(2q+4)\times 2}$, defined as (for any column vector ξ of length $\dim(\rho)$)

$$
\frac{\partial D_t(\rho)}{\partial \rho^T}(\xi) := \begin{bmatrix}
0_{q\times 1} & 0_{q\times 1} \\
0_{q\times 1} & 0_{q\times 1} \\
-\frac{\partial p(X_t;\rho)}{\partial \rho^T}\xi X_t & 0 \\
0 & -\frac{\partial p(X_t;\rho)}{\partial \rho^T}\xi X_t \\
-\frac{\partial p(X_t;\rho)}{\partial \rho^T}\xi(1-X_t) & 0 \\
0 & -\frac{\partial p(X_t;\rho)}{\partial \rho^T}\xi(1-X_t)
\end{bmatrix}.
$$

$\frac{\partial\left\{\Sigma_t(\theta)^{-1}\right\}}{\partial \theta}$ is an operator that maps $\mathbb{R}^{\dim(\theta)} \to \mathbb{R}^{2\times 2}$, defined as (for any column vector ζ of length $\dim(\theta)$)

$$
\begin{aligned}
\frac{\partial\left\{\Sigma_t(\theta)^{-1}\right\}}{\partial \theta^T}(\zeta) &:= \frac{1}{\left\{1-\pi_t^{(1)}(\theta)-\pi_t^{(2)}(\theta)\right\}^2} \left\{ \frac{\partial \pi_t^{(1)}(\theta)}{\partial \theta^T} + \frac{\partial \pi_t^{(2)}(\theta)}{\partial \theta^T} \right\} \zeta \begin{bmatrix} \frac{1-\pi_t^{(2)}(\theta)}{\pi_t^{(1)}(\theta)} & 1 \\ 1 & \frac{1-\pi_t^{(1)}(\theta)}{\pi_t^{(2)}(\theta)} \end{bmatrix} \\
&\quad + \frac{1}{1-\pi_t^{(1)}(\theta)-\pi_t^{(2)}(\theta)} \begin{bmatrix} -\frac{1-\pi_t^{(2)}(\theta)}{\pi_t^{(1)}(\theta)^2}\frac{\partial \pi_t^{(1)}(\theta)}{\partial \theta^T}\zeta & 0 \\ 0 & -\frac{1-\pi_t^{(1)}(\theta)}{\pi_t^{(2)}(\theta)^2}\frac{\partial \pi_t^{(2)}(\theta)}{\partial \theta^T}\zeta \end{bmatrix}.
\end{aligned}
$$

Also we have

$$
\begin{aligned}
\frac{\partial \pi_t^{(1)}(\theta)}{\partial \theta^T} &= \pi_t^{(1)}(\theta) \begin{bmatrix} Z_t^T & 0_{1\times q} & A_t(1-X_t) & A_t X_t & 0 & 0 \end{bmatrix}, \\
\frac{\partial \pi_t^{(2)}(\theta)}{\partial \theta^T} &= \pi_t^{(2)}(\theta) \begin{bmatrix} 0_{1\times q} & Z_t^T & 0 & 0 & A_t(1-X_t) & A_t X_t \end{bmatrix}.
\end{aligned}
$$