Web-based Supplementary Materials for Adaptive Semi-Supervised Inference for Optimal Treatment Decisions with Electronic Medical Record Data

By Kevin Gunn, Wenbin Lu*, Rui Song
*email: wlu4@ncsu.edu

1 Web Appendix

Proof of Theorem 1. Let $\widehat{C}(\mathbf{x}) = \widehat{Q}^{(np)}(\mathbf{x},1) - \widehat{Q}^{(np)}(\mathbf{x},0)$.

$$\begin{aligned}
\left(\widehat{\boldsymbol{\beta}}_{np} - \boldsymbol{\beta}\right) &= \Lambda_N^{-1}\left[N^{-1}\sum_{j=n+1}^{n+N}\mathbf{X}_j\left\{\widehat{C}(\mathbf{X}_j) - \boldsymbol{\beta}'\widetilde{\mathbf{X}}_j\right\}\right] \\
&= \Lambda_N^{-1}\left[N^{-1}\sum_{j=n+1}^{n+N}\mathbf{X}_j\left\{\widehat{C}(\mathbf{X}_j) - C(\mathbf{X}_j)\right\}\right] + \Lambda_N^{-1}\left[N^{-1}\sum_{j=n+1}^{n+N}\mathbf{X}_j\left\{C(\mathbf{X}_j) - \boldsymbol{\beta}'\widetilde{\mathbf{X}}_j\right\}\right] \\
&= \Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{C}(\mathbf{X}) - C(\mathbf{X})\right\}\right] + O_p(N^{-1/2}).
\end{aligned}$$

The first step follows from the normal equations. The last step is due to the fact $\Lambda_N^{-1}\left[N^{-1}\sum_{j=n+1}^{n+N}\mathbf{X}_j\left\{\widehat{C}(\mathbf{X}_j) - C(\mathbf{X}_j)\right\}\right] = \Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{C}(\mathbf{X}) - C(\mathbf{X})\right\}\right] + o_p(1)$ by standard arguments involving the weak law of large numbers. According to the central limit theorem $N^{-1/2}\left[N^{-1/2}\sum_{j=n+1}^{N}\Lambda_n^{-1}\left\{C(\mathbf{X})_j - \boldsymbol{\beta}'\widetilde{\mathbf{X}}_j\right\}\right] = O_p(N^{-1/2})$. Multiplying both sides by $n^{1/2}$ we have,

$$\begin{aligned}
n^{1/2}\left(\widehat{\boldsymbol{\beta}}_{np} - \boldsymbol{\beta}\right) &= n^{1/2}\Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{C}(\mathbf{x}) - C(\mathbf{X})\right\}\right] + O_p\left((n/N)^{\frac{1}{2}}\right) \\
&= n^{1/2}\Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{Q}^{(np)}(\mathbf{X},1) - Q^{(np)}(\mathbf{X},1)\right\}\right] \\
&- n^{1/2}\Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{Q}^{(np)}(\mathbf{X},0) - Q^{(np)}(\mathbf{X},0)\right\}\right] + O_p\left((n/N)^{\frac{1}{2}}\right).
\end{aligned}$$

Note that $n/N \to 0$ implying $O_p\left((n/N)^{\frac{1}{2}}\right) \equiv o_p(1)$. Next, let $\tau(\mathbf{X}) = \pi(\mathbf{X})f(\mathbf{X})$ and $\widehat{\tau}(\mathbf{X}) = \frac{1}{nh^p}\sum_{i=1}^n A_i W_h(\mathbf{X}_i - \mathbf{X})$, where $W_h(\mathbf{X}_i - \mathbf{X}) = W(\frac{\mathbf{X}_i - \mathbf{X}}{h})$. Let's rewrite $E\left[\mathbf{X}\left\{\widehat{Q}^{(np)}(\mathbf{X},1) - Q^{(np)}(\mathbf{X},1)\right\}\right]$ as,

$$= E\left\{\frac{\frac{1}{nh^p}\sum_{i=1}^n A_i\mathbf{X}W_h(\mathbf{X}_i - \mathbf{X})\left\{Y_i - Q^{(np)}(\mathbf{X},1)\right\}}{\tau(\mathbf{X})}\right\} \tag{1.1}$$

$$+ E\left\{\mathbf{X}\left(\widehat{Q}^{(np)}(\mathbf{X},1) - Q^{(np)}(\mathbf{X},1)\right)\left\{\frac{\tau(\mathbf{X}) - \widehat{\tau}(\mathbf{X})}{\tau(\mathbf{X})}\right\}\right\} = H_{n,1}^{(1)} + H_{n,2}^{(1)}. \tag{1.2}$$

Then $H_{n,1}^{(1)}$ is equivalent to:

$$\begin{aligned}
&= \frac{1}{nh^p}\sum_{i=1}^{n} A_i \int \mathbf{X}\left\{Y_i - Q^{(np)}(\mathbf{X},1)\right\}\frac{W_h(\mathbf{X}-\mathbf{X}_i)}{\tau(\mathbf{X})} f(\mathbf{X}) d\mathbf{X} \\
&= \frac{1}{nh^p}\sum_{i=1}^{n} A_i \int \mathbf{X}\left\{Y_i - Q^{(np)}(\mathbf{X},1)\right\}\frac{W_h(\mathbf{X}-\mathbf{X}_i)}{\pi(\mathbf{X})} d\mathbf{X} \\
&= \frac{1}{n}\sum_{i=1}^{n} A_i \int (\mathbf{X}_i + h\mathbf{t}_i)\left\{Y_i - Q^{(np)}(\mathbf{X}_i + h\mathbf{t}_i, 1)\right\}\frac{W(\mathbf{t}_i)}{\pi(\mathbf{X}_i + h\mathbf{t}_i)} d\mathbf{t}_i
\end{aligned}$$

By *assumptions* (1) and (2), Taylor expansion in $h\mathbf{t}_i$ for sufficiently small h leads to,

$$H_{n,1}^{(1)} = \frac{1}{n}\sum_{i=1}^{n}\frac{A_i}{\pi(\mathbf{X}_i)}\mathbf{X}_i\left\{Y_i - Q^{(np)}(\mathbf{X}_i,1)\right\} + O_p(h^r).$$

Since $n^{1/2}h^r \to 0$ as $n \to \infty$,

$$n^{1/2}\Lambda H_{n,1}^{(1)} = \widetilde{H}_{n,1}^{(1)} = n^{-1/2}\sum_{i=1}^{n}\frac{A_i}{\pi(\mathbf{X}_i)}\Lambda\mathbf{X}_i\left\{Y_i - Q^{(np)}(\mathbf{X}_i,1)\right\} + o_p(1). \tag{1.3}$$

Let $q(\mathbf{X}) = \widehat{Q}^{(np)}(\mathbf{X},1) - Q^{(1)}(\mathbf{X},1)$, $l(\mathbf{X}) = \frac{\tau(\mathbf{X}) - \widehat{\tau}(\mathbf{X})}{\tau(\mathbf{X})} = 1 - \frac{\widehat{\tau}(\mathbf{X})}{\tau(\mathbf{X})}$, and $Q^{(np)}(\mathbf{X},1) = \alpha(\mathbf{X})/\tau(\mathbf{X})$, where $\alpha(\mathbf{X})$ is the numerator in Equation 1.1. It follows that $H_{n,2}^{(1)} = E\{\mathbf{X}q(\mathbf{X})l(\mathbf{X})\}$ and,

$$E\{\mathbf{X}q(\mathbf{X})l(\mathbf{X})\} \le \sup_{\mathbf{x}\in\mathcal{X}}\{\|\mathbf{X}\|\ |q(\mathbf{X})|\ |l(\mathbf{X})|\} = o_p(1). \tag{1.4}$$

Equation 1.4 requires that $\mathbf{X}$ is bounded, $\sqrt{\frac{\ln n}{nh^p}} \to 0$ as $n \to \infty$, as well as *assumptions* (3) and (5). It follows by similar argument of *Lemma B.1* in Newey (1994) combined with Taylor series expansion that $\sup_{\mathbf{x}\in\mathcal{X}}|\widehat{\tau}(\mathbf{X}) - \tau(\mathbf{X})| = \sup_{\mathbf{x}\in\mathcal{X}}|\widehat{\alpha}(\mathbf{X}) - \alpha(\mathbf{X})| = o_p(1)$. Afterwards, it holds that $\sup_{\mathbf{x}\in\mathcal{X}}|q(\mathbf{X})| = o_p(1)$ through similar reasoning given by *Theorem 8* of Hansen (2008). The same technique proves,

$$n^{1/2}\Lambda E\left[\mathbf{X}\left\{\widehat{Q}^{(np)}(\mathbf{X},0) - Q^{(np)}(\mathbf{X},0)\right\}\right] = \widetilde{H}_{n,1}^{(0)} + o_p(1).$$

This leaves us with,

$$n^{1/2}\Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{C}(\mathbf{x}) - C(\mathbf{X})\right\}\right] = \widetilde{H}_{n,1}^{(1)} - \widetilde{H}_{n,1}^{(0)} + o_p(1) = n^{-1/2}\sum_{i=1}^{n}\Psi_{np}(\mathbf{O}_i) + o_p(1).$$

Since $\Psi_{np}(\mathbf{O})$ is the influence function for $\widehat{\boldsymbol{\beta}}_{np}$ with $E\left[\Psi_{np}(\mathbf{O})\right] = 0$ and variance $V_{\boldsymbol{\beta}_{np}} = E\left[\Psi_{np}(\mathbf{O})\Psi'_{np}(\mathbf{O})\right]$, the *CLT* states it will converge to $\mathcal{N}_{p+1}(0, V_{\boldsymbol{\beta}_{np}})$. $\square$

Proof of Theorem 2. The proof of Theorem 2 follows similar logic to the proof of *Theorem 3.2* from Chakrabortty et al. (2018). $\square$

References

Chakrabortty, A., Cai, T., et al. (2018). Efficient and adaptive linear regression in semi-supervised settings. *The Annals of Statistics* **46,** 1541–1572.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24,** 726–748.

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory* **10,** 1–21.