Day 2, Lecture 1

The foundation of TMLE: The constructive proof of TMLE

Overview of Day 2

Day 2: 8 – 9

Introduction to TMLE. The constructive proof of TMLE.

- ▷ The decomposition and the role of Step 1 & Step 2.

Overview of Day 2

Day 2: 8 – 9

Introduction to TMLE. The constructive proof of TMLE.

- ▷ The decomposition and the role of Step 1 & Step 2.

Day 2: 9 – 12

The targeting step. Solving the efficient influence curve equation.

- ▷ Implementation of the targeting step.

Overview of Day 2

Day 2: 8 – 9

Introduction to TMLE. The constructive proof of TMLE.

- ▷ The decomposition and the role of Step 1 & Step 2.

Day 2: 9 – 12

The targeting step. Solving the efficient influence curve equation.

- ▷ Implementation of the targeting step.

Lunch.

Overview of Day 2

Day 2: 8 – 9

Introduction to TMLE. The constructive proof of TMLE.

- ▷ The decomposition and the role of Step 1 & Step 2.

Day 2: 9 – 12

The targeting step. Solving the efficient influence curve equation.

- ▷ Implementation of the targeting step.

Lunch.

Day 2: 13 – 15

Step 2: Super learning. (Thomas).

Introduction to TMLE

1. Data is a random variable O with a probability P_0
2. P_0 belongs to a statistical model $\mathcal{M}$
3. Our target is a parameter $\Psi : \mathcal{M} \to \mathbb{R}$
4. Construct estimator $\hat{P}_n$ for (relevant part of) P_0 and estimate the target parameter by $\hat{\psi}_n = \Psi(\hat{P}_n)$
5. Quantify uncertainty for the estimator $\hat{\psi}_n = \Psi(\hat{P}_n)$

Estimation paradigm

1. Minimal amount of parametric assumptions on P_0 (goal)
2. Go after optimal/efficient estimation of $\psi_0 = \Psi(P_0)$

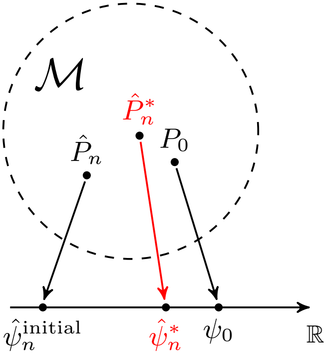
Introduction to TMLE

1. Data is a random variable O with a probability P_0
2. P_0 belongs to a statistical model $\mathcal{M}$
3. Our target is a parameter $\Psi : \mathcal{M} \to \mathbb{R}$
4. Construct estimator $\hat{P}_n$ for (relevant part of) P_0 and estimate the target parameter by $\hat{\psi}_n = \Psi(\hat{P}_n)$
5. Quantify uncertainty for the estimator $\hat{\psi}_n = \Psi(\hat{P}_n)$

Estimation paradigm

1. Minimal amount of parametric assumptions on P_0 (goal)
2. Go after optimal/efficient estimation of $\psi_0 = \Psi(P_0) \Rightarrow$ based the efficient influence function (canonical gradient)

Introduction to TMLE



Tools from semiparametric efficiency theory and empirical process theory tell us how to construct an optimal estimator for a **given target parameter** $\Psi : \mathcal{M} \to \mathbb{R}$

- asymptotic linearity/normality
- asymptotic efficiency

Introduction to TMLE

Recap notation:[1]

- For a function $h : \mathcal{O} \to \mathbb{R}$ and distribution P

$$Ph = \mathbb{E}_P[h(O)] = \int hdP = \int_{\mathcal{O}} h(o)dP(o)$$

  where $\mathcal{O} = \mathbb{R}^d \times \{0,1\} \times \{0,1\}$ is the sample space of $O = (X, A, Y)$.

- For the empirical measure $\mathbb{P}_n$ of the sample $O_1, \ldots, O_n$:

$$\mathbb{P}_n h = \int hd\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} h(O_i);$$

  note: the right-hand-side is really just the empirical average.

- $X_n = o_P(1)$ means that $X_n \xrightarrow{P} 0$; $X_n = o_P(n^{-1/2})$ means that $n^{1/2}X_n \xrightarrow{P} 0$.

[1] van der Vaart, A. W. (2000). Asymptotic statistics (Vol. 3). Cambridge university press.

Introduction to TMLE

Asymptotic linearity

An estimator $\hat{\psi}_n$ is asymptotically linear if,

$$\sqrt{n}\big(\hat{\psi}_n - \Psi(P_0)\big) = \sqrt{n}\,\mathbb{P}_n\phi(P_0) + o_P(1) \qquad (*)$$

with $P_0\phi(P_0) = 0$. We say that $\hat{\psi}_n$ has influence function $O \mapsto \phi(P_0)(O)$.

Introduction to TMLE

Asymptotic linearity

An estimator $\hat{\psi}_n$ is asymptotically linear if,

$$\sqrt{n}\big(\hat{\psi}_n - \Psi(P_0)\big) = \sqrt{n}\,\mathbb{P}_n\phi(P_0) + o_P(1) \qquad (*)$$

with $P_0\phi(P_0) = 0$. We say that $\hat{\psi}_n$ has influence function $O \mapsto \phi(P_0)(O)$.

Use CLT on $(*)$ $\Rightarrow$ asymptotic normality:

$$\hat{\psi}_n \overset{as}{\sim} N(\Psi(P_0), \sigma_0^2/n),$$

where $\sigma_0^2 = P_0\phi(P_0)^2$.

Introduction to TMLE

Asymptotic linearity

An estimator $\hat{\psi}_n$ is asymptotically linear if,

$$\sqrt{n}\big(\hat{\psi}_n - \Psi(P_0)\big) = \sqrt{n}\,\mathbb{P}_n\phi(P_0) + o_P(1) \qquad (*)$$

with $P_0\phi(P_0) = 0$. We say that $\hat{\psi}_n$ has influence function $O \mapsto \phi(P_0)(O)$.

Use CLT on $(*)$ $\Rightarrow$ asymptotic normality:

$$\hat{\psi}_n \overset{as}{\sim} N(\Psi(P_0), \sigma_0^2/n),$$

where $\sigma_0^2 = P_0\phi(P_0)^2$.

Efficient influence function

There exists $\phi^*(P_0)$ such that $P_0\phi(P_0)^2 \geq P_0\phi^*(P_0)^2$.

$\phi^*(P_0)$ is called the efficient influence function.

$\Rightarrow$ estimator with the smallest possible asymptotic variance

Introduction to TMLE

EXAMPLE: Average treatment effect (ATE)

Observed data $O = (X, A, Y) \in \mathbb{R}^d \times \{0,1\} \times \{0,1\} = \mathcal{O}$

- $X \in \mathbb{R}^d$ are covariates
- $A \in \{0,1\}$ is a binary exposure variable (treatment decision)
- $Y \in \{0,1\}$ is a binary outcome variable

$O \sim P_0$ where P_0 assumed to belong to nonparametric model $\mathcal{M}$.

We are interested in estimating the ATE:

$$\Psi(P) = \mathbb{E}_P\big[\mathbb{E}_P[Y \mid A = 1, X] - \mathbb{E}_P[Y \mid A = 0, X]\big].$$

Introduction to TMLE

For the ATE, as we have seen, we can also write the target parameter $\Psi : \mathcal{M} \to \mathbb{R}$ as

$$\Psi(P) = \tilde{\Psi}(f, \mu_X) = \int_{\mathbb{R}^d} \big(f(1,x) - f(0,x)\big) d\mu_X(x) \qquad (*)$$

where

$$f(a,x) = \mathbb{E}[\,Y \mid A = a, X = x]$$

and μ_X is the marginal distribution of X.

I.e., $\hat{\psi}_n = \tilde{\Psi}(\hat{f}_n, \hat{\mu}_n)$.

Introduction to TMLE

An estimator $\hat{\psi}_n$ is asymptotically linear if,

$$\sqrt{n}\big(\hat{\psi}_n - \Psi(P_0)\big) = \sqrt{n}\,\mathbb{P}_n\phi(P_0) + o_P(1). \quad (*)$$

For $P \in \mathcal{M}$, define: $R(P, P_0) := \Psi(P) - \Psi(P_0) + P_0\phi^*(P)$.

Introduction to TMLE

An estimator $\hat{\psi}_n$ is asymptotically linear if,

$$\sqrt{n}\big(\hat{\psi}_n - \Psi(P_0)\big) = \sqrt{n}\,\mathbb{P}_n\phi(P_0) + o_P(1). \quad (*)$$

For $P \in \mathcal{M}$, define: $R(P, P_0) := \Psi(P) - \Psi(P_0) + P_0\phi^*(P)$.

We decompose as follows:

$$\begin{aligned}
\Psi(\hat{P}_n) - \Psi(P_0) &= -P_0\phi^*(\hat{P}_n) + R(\hat{P}_n, P_0) \\
&= -P_0\phi^*(\hat{P}_n) + R(\hat{P}_n, P_0) \\
&\quad + \mathbb{P}_n\phi^*(\hat{P}_n) - \mathbb{P}_n\phi^*(\hat{P}_n) \\
&\quad + (\mathbb{P}_n - P_0)\phi^*(P_0) - (\mathbb{P}_n - P_0)\phi^*(P_0)
\end{aligned}$$

(1)

(2)

(3)

Introduction to TMLE

An estimator $\hat{\psi}_n$ is asymptotically linear if,

$$\sqrt{n}\big(\hat{\psi}_n - \Psi(P_0)\big) = \sqrt{n}\,\mathbb{P}_n\phi(P_0) + o_P(1). \quad (*)$$

For $P \in \mathcal{M}$, define: $R(P, P_0) := \Psi(P) - \Psi(P_0) + P_0\phi^*(P)$.

We decompose as follows:

$$\begin{aligned}
\Psi(\hat{P}_n) - \Psi(P_0) &= -P_0\phi^*(\hat{P}_n) + R(\hat{P}_n, P_0) \\
&= -P_0\phi^*(\hat{P}_n) + R(\hat{P}_n, P_0) \\
&\qquad + \mathbb{P}_n\phi^*(\hat{P}_n) - \mathbb{P}_n\phi^*(\hat{P}_n) \\
&\qquad + (\mathbb{P}_n - P_0)\phi^*(P_0) - (\mathbb{P}_n - P_0)\phi^*(P_0) \\
&= (\mathbb{P}_n - P_0)\phi^*(P_0) \\
&\qquad + (\mathbb{P}_n - P_0)\big(\phi^*(\hat{P}_n) - \phi^*(P_0)\big) && (1) \\
&\qquad + R(\hat{P}_n, P_0) && (2) \\
&\qquad - \mathbb{P}_n\phi^*(\hat{P}_n) && (3)
\end{aligned}$$

Introduction to TMLE

An estimator $\hat{\psi}_n$ is asymptotically linear if,

$$\sqrt{n}\big(\hat{\psi}_n - \Psi(P_0)\big) = \sqrt{n}\,\mathbb{P}_n\phi(P_0) + o_P(1). \quad (*)$$

For $P \in \mathcal{M}$, define: $R(P, P_0) := \Psi(P) - \Psi(P_0) + P_0\phi^*(P)$.

We decompose as follows:

$$\begin{aligned}
\Psi(\hat{P}_n) - \Psi(P_0) &= -P_0\phi^*(\hat{P}_n) + R(\hat{P}_n, P_0) \\
&= -P_0\phi^*(\hat{P}_n) + R(\hat{P}_n, P_0) \\
&\qquad + \mathbb{P}_n\phi^*(\hat{P}_n) - \mathbb{P}_n\phi^*(\hat{P}_n) \\
&\qquad + (\mathbb{P}_n - P_0)\phi^*(P_0) - (\mathbb{P}_n - P_0)\phi^*(P_0) \\
&= \mathbb{P}_n\phi^*(P_0) \\
&\qquad + (\mathbb{P}_n - P_0)\big(\phi^*(\hat{P}_n) - \phi^*(P_0)\big) && (1) \\
&\qquad + R(\hat{P}_n, P_0) && (2) \\
&\qquad - \mathbb{P}_n\phi^*(\hat{P}_n) && (3)
\end{aligned}$$

Introduction to TMLE

An estimator $\hat{\psi}_n$ is asymptotically linear if,

$$\sqrt{n}\big(\hat{\psi}_n - \Psi(P_0)\big) = \sqrt{n}\,\mathbb{P}_n\phi(P_0) + o_P(1). \quad (*)$$

For $P \in \mathcal{M}$, define: $R(P, P_0) := \Psi(P) - \Psi(P_0) + P_0\phi^*(P)$.

We decompose as follows:

$$\begin{aligned}
\Psi(\hat{P}_n) - \Psi(P_0) &= -P_0\phi^*(\hat{P}_n) + R(\hat{P}_n, P_0) \\
&= -P_0\phi^*(\hat{P}_n) + R(\hat{P}_n, P_0) \\
&\qquad + \mathbb{P}_n\phi^*(\hat{P}_n) - \mathbb{P}_n\phi^*(\hat{P}_n) \\
&\qquad + (\mathbb{P}_n - P_0)\phi^*(P_0) - (\mathbb{P}_n - P_0)\phi^*(P_0) \\
&= \mathbb{P}_n\phi^*(P_0) + o_P(n^{-1/2}) \\
&\qquad + (\mathbb{P}_n - P_0)\big(\phi^*(\hat{P}_n) - \phi^*(P_0)\big) && (1) \\
&\qquad + R(\hat{P}_n, P_0) && (2) \\
&\qquad - \mathbb{P}_n\phi^*(\hat{P}_n) && (3)
\end{aligned}$$

Introduction to TMLE

An estimator $\hat{\psi}_n$ is asymptotically linear if,

$$\sqrt{n}\big(\hat{\psi}_n - \Psi(P_0)\big) = \sqrt{n}\,\mathbb{P}_n\phi(P_0) + o_P(1). \quad (*)$$

For $P \in \mathcal{M}$, define: $R(P, P_0) := \Psi(P) - \Psi(P_0) + P_0\phi^*(P)$.

We decompose as follows:

$$\begin{aligned}
\Psi(\hat{P}_n) - \Psi(P_0) &= -P_0\phi^*(\hat{P}_n) + R(\hat{P}_n, P_0) \\
&= -P_0\phi^*(\hat{P}_n) + R(\hat{P}_n, P_0) \\
&\qquad + \mathbb{P}_n\phi^*(\hat{P}_n) - \mathbb{P}_n\phi^*(\hat{P}_n) \\
&\qquad + (\mathbb{P}_n - P_0)\phi^*(P_0) - (\mathbb{P}_n - P_0)\phi^*(P_0) \\
&= \mathbb{P}_n\phi^*(P_0) + o_P(n^{-1/2}) \\
&\qquad + (\mathbb{P}_n - P_0)\big(\phi^*(\hat{P}_n) - \phi^*(P_0)\big) && (1) \\
&\qquad + R(\hat{P}_n, P_0) && (2) \\
&\qquad - \mathbb{P}_n\phi^*(\hat{P}_n) && (3)
\end{aligned}$$

i.e., want (1)–(3) to be $o_P(n^{-1/2})$.

Introduction to TMLE

An estimator $\hat{\psi}_n$ is asymptotically linear if,

$$\sqrt{n}\big(\hat{\psi}_n - \Psi(P_0)\big) = \sqrt{n}\,\mathbb{P}_n\phi(P_0) + o_P(1). \quad (*)$$

$$\Psi(\hat{P}_n) - \Psi(P_0) = \mathbb{P}_n\phi^*(P_0) + o_P(n^{-1/2})$$

$$+ (\mathbb{P}_n - P_0)\big(\phi^*(\hat{P}_n) - \phi^*(P_0)\big) \quad (1)$$

$$+ R(\hat{P}_n, P_0) \quad (2)$$

$$- \mathbb{P}_n\phi^*(\hat{P}_n) \quad (3)$$

- (1) handled by empirical process theory if Donsker.[2]
- (2) depends on the target parameter.
- (3) is called the efficient influence curve equation.

[2] Lemma 19.24 of van der Vaart, A. W. (2000): Asymptotic statistics yields then that $(\mathbb{P}_n - P_0)(\phi^*(\hat{P}_n) - \phi^*(P_0)) = o_P(n^{-1/2})$.

Introduction to TMLE

That is it.

Introduction to TMLE

That is it.

Conditions (asymptotic linearity and efficiency)

(C1) Solve the efficient influence curve equation: $\mathbb{P}_n \phi^*(\hat{P}_n) = o_P(n^{-1/2})$

(C2) Remainder $R(\hat{P}_n, P_0) = o_P(n^{-1/2})$

(C3) Donsker class conditions for $\{\phi^*(P) : P \in \mathcal{M}\}$

Then: $\Psi(\hat{P}_n) \overset{as}{\sim} N(\Psi(P_0), P_0\phi^*(P_0)^2/n)$

Introduction to TMLE

Side note: Usually, we will assume the Donsker class condition (C3)

- this is a way of nonparametrically characterizing the complexity of nuisance parameters.
- classes of functions that are Donsker: Indicator functions, bounded monotone functions, Lipschitz parametric functions, smooth functions, . . .

Donsker classes also include traditional parametric functions.

We will not discuss this further. For a nice intro see Sections 4.2 and 4.3 of Kennedy, E. H. (2016): Semiparametric theory and empirical processes in causal inference.

Introduction to TMLE

$$\Psi(\hat{P}_n) - \Psi(P_0) = \mathbb{P}_n \phi^*(P_0) + o_P(n^{-1/2})$$
$$+ R(\hat{P}_n, P_0)$$
$$- \mathbb{P}_n \phi^*(\hat{P}_n)$$

For **a given target parameter** $\Psi : \mathcal{M} \to \mathbb{R}$, we need to

1. Know the efficient influence curve, so that we can solve the efficient influence curve equation.
2. Analyze the remainder $R(P, P_0) := \Psi(P) - \Psi(P_0) + P_0 \phi^*(P)$.

Introduction to TMLE

EXAMPLE: Average treatment effect (ATE)

Introduction to TMLE

EXAMPLE: Average treatment effect (ATE)

1. The efficient influence function:

$$\begin{aligned}\phi^*(P)(O) &= \tilde{\phi}^*(f,\pi)(O) \\ &= \left(\frac{A}{\pi(A \mid X)} - \frac{1-A}{\pi(A \mid X)}\right)\big(Y - f(A,X)\big) + f(1,X) - f(0,X) - \Psi(P)\end{aligned}$$

Introduction to TMLE

EXAMPLE: Average treatment effect (ATE)

1. The efficient influence function:

$$\begin{aligned}\phi^*(P)(O) &= \tilde{\phi}^*(f,\pi)(O)\\ &= \left(\frac{A}{\pi(A \mid X)} - \frac{1-A}{\pi(A \mid X)}\right)\big(Y - f(A,X)\big) + f(1,X) - f(0,X) - \Psi(P)\end{aligned}$$

2. The remainder:

$$\begin{aligned}R(P,P_0) &= \tilde{R}(f,\pi,f_0,\pi_0)\\ &= \int_{\mathbb{R}^d} \sum_{a=0,1} (2a-1)\frac{\pi_0(a \mid x) - \pi(a \mid x)}{\pi(a \mid x)}\big(f_0(a,x) - f(a,x)\big)d\mu_{0,X}(x)\end{aligned}$$

$$R(P, P_0) := \Psi(P) - \Psi(P_0) + P_0\phi^*(P).$$

Deriving the remainder for the ATE:

$$\begin{aligned}
R(P, P_0) &= \mathbb{E}_P[f(1,X) - f(0,X)] - \mathbb{E}_{P_0}[f_0(1,X) - f_0(0,X)] \\
&+ \mathbb{E}_{P_0}\left[\left(\frac{A}{\pi(A \mid X)} - \frac{1-A}{\pi(A \mid X)}\right)\big(Y - f(A,X)\big)\right] \\
&+ \mathbb{E}_{P_0}\big[f(1,X) - f(0,X)\big] - \Psi(P) \\
&\overset{*}{=} \int_{\mathbb{R}^d} \sum_{a=0,1} (2a-1)\left(\frac{\pi_0(a \mid x)}{\pi(a \mid x)} - 1\right)\big(f_0(a,x) - f(a,x)\big) d\mu_{0,X}(x) \\
&= \int_{\mathbb{R}^d} \sum_{a=0,1} (2a-1)\frac{\pi_0(a \mid x) - \pi(a \mid x)}{\pi(a \mid x)}\big(f_0(a,x) - f(a,x)\big) d\mu_{0,X}(x)
\end{aligned}$$

the equality marked by $*$ is detailed on the next slide.

Introduction to TMLE

We used that:

$$
\begin{aligned}
&\mathbb{E}_{P_0}\left[\left(\frac{A}{\pi(A \mid X)} - \frac{1-A}{\pi(A \mid X)}\right)\big(Y - f(A,X)\big)\right] \\
&= \mathbb{E}_{P_0}\left[\frac{2A-1}{\pi(A \mid X)}\big(Y - f(A,X)\big)\right] \\
&= \mathbb{E}_{P_0}\left[\mathbb{E}_{P_0}\left[\frac{2A-1}{\pi(A \mid X)}\big(Y - f(A,X)\big)\,\middle|\, A, X\right]\right] \\
&= \mathbb{E}_{P_0}\left[\frac{2A-1}{\pi(A \mid X)}\big(f_0(A,X) - f(A,X)\big)\right] \\
&= \int_{\mathbb{R}^d} \sum_{a=0,1} \frac{2a-1}{\pi(a \mid x)}\big(f_0(a,x) - f(a,x)\big)\pi_0(a \mid x)\, d\mu_{0,X}(x) \\
&= \int_{\mathbb{R}^d} \sum_{a=0,1} (2a-1)\frac{\pi_0(a \mid x)}{\pi(a \mid x)}\big(f_0(a,x) - f(a,x)\big)\, d\mu_{0,X}(x)
\end{aligned}
$$

Introduction to TMLE

The remainder determines the asymptotic bias.

For the ATE, the remainder has a really nice structure!

$$
\begin{aligned}
R(P, P_0) &= \tilde{R}(f, \pi, f_0, \pi_0) \\
&= \int_{\mathbb{R}^d} \sum_{a=0,1} (2a-1) \frac{\pi_0(a \mid x) - \pi(a \mid x)}{\pi(a \mid x)} \big(f_0(a,x) - f(a,x) \big) d\mu_{0,X}(x)
\end{aligned}
$$

A "double robust" structure, which has some important implications.

Introduction to TMLE

$$
\begin{aligned}
|R(P, P_0)| &= |\tilde{R}(f, \pi, f_0, \pi_0)| \\
&\leq \sum_{a=0,1} \int_{\mathbb{R}^d} \frac{|\pi_0(a \mid x) - \pi(a \mid x)|}{\pi(a \mid x)} |f_0(a, x) - f(a, x)| d\mu_{0,X}(x)
\end{aligned}
$$

Introduction to TMLE

$$
\begin{aligned}
|R(P, P_0)| &= |\tilde{R}(f, \pi, f_0, \pi_0)| \\
&\leq \sum_{a=0,1} \int_{\mathbb{R}^d} \frac{|\pi_0(a \mid x) - \pi(a \mid x)|}{\pi(a \mid x)} |f_0(a, x) - f(a, x)| d\mu_{0,X}(x) \\
&\overset{*}{\leq} \sum_{a=0,1} \frac{1}{\pi(a \mid x)} \sqrt{\int_{\mathbb{R}^d} \left\{\pi_0(a \mid x) - \pi(a \mid x)\right\}^2 d\mu_{0,X}(x)} \\
&\qquad\qquad \times \sqrt{\int_{\mathbb{R}^d} \left\{f_0(a, x) - f(a, x)\right\}^2 d\mu_{0,X}(x)}
\end{aligned}
$$

$*$ uses Cauchy-Schwarz.

Introduction to TMLE

$$
\begin{aligned}
|R(P, P_0)| &= |\tilde{R}(f, \pi, f_0, \pi_0)| \\
&\leq \sum_{a=0,1} \int_{\mathbb{R}^d} \frac{|\pi_0(a \mid x) - \pi(a \mid x)|}{\pi(a \mid x)} |f_0(a, x) - f(a, x)| d\mu_{0,X}(x) \\
&\overset{*}{\leq} \sum_{a=0,1} \frac{1}{\pi(a \mid x)} \sqrt{\int_{\mathbb{R}^d} \left\{\pi_0(a \mid x) - \pi(a \mid x)\right\}^2 d\mu_{0,X}(x)} \\
&\qquad\qquad \times \sqrt{\int_{\mathbb{R}^d} \left\{f_0(a, x) - f(a, x)\right\}^2 d\mu_{0,X}(x)}
\end{aligned}
$$

Thus, if $\pi(a \mid X) > \delta > 0$ a.s., then:

$$
|\tilde{R}(\hat{f}_n^*, \hat{\pi}_n, f_0, \pi_0)| \leq \sum_{a=0,1} \delta^{-1} \left\|\pi_0(a \mid \cdot) - \hat{\pi}_n(a \mid \cdot)\right\|_{\mu_0} \left\|f_0(a \mid \cdot) - \hat{f}_n(a \mid \cdot)\right\|_{\mu_0}
$$

$*$ uses Cauchy-Schwarz.

Introduction to TMLE

What does this imply for estimation?

Double robustness in consistency

$$|\tilde{R}(\hat{f}_n^*, \hat{\pi}_n, f_0, \pi_0)| \leq \sum_{a=0,1} \delta^{-1} \underbrace{\|\pi_0(a \mid \cdot) - \hat{\pi}_n(a \mid \cdot)\|_{\mu_0}}_{o_P(1), \text{ or}} \underbrace{\|f_0(a \mid \cdot) - \hat{f}_n^*(a \mid \cdot)\|_{\mu_0}}_{o_P(1)}$$

then $\tilde{\Psi}(\hat{f}_n^*) - \tilde{\Psi}(f_0) = o_P(1)$.

Asymptotic linearity (easier to establish due to double robust structure)

$$|\tilde{R}(\hat{f}_n^*, \hat{\pi}_n, f_0, \pi_0)| \leq \sum_{a=0,1} \delta^{-1} \underbrace{\|\pi_0(a \mid \cdot) - \hat{\pi}_n(a \mid \cdot)\|_{\mu_0}}_{=o_P(n^{-1/4})} \underbrace{\|f_0(a \mid \cdot) - \hat{f}_n^*(a \mid \cdot)\|_{\mu_0}}_{=o_P(n^{-1/4})}$$

i.e., $\tilde{R}(\hat{f}_n^*, \hat{\pi}_n, f_0, \pi_0) = o_P(n^{-1/2})$.

I.e., bias is converging at fast enough rate for reliable confidence intervals.

Introduction to TMLE

Side note: Showing the double robustness in consistency ...

Say we have estimators $(\hat{f}_n^*, \hat{\pi}_n)$;

- converging to (f, π)
- solving the efficient influence curve equation.

Per definition, $\tilde{R}(\hat{f}_n^*, \hat{\pi}_n, f_0, \pi_0) = \tilde{\Psi}(\hat{f}_n^*) - \tilde{\Psi}(f_0) + P_0\tilde{\phi}^*(\hat{f}_n^*, \hat{\pi}_n)$.

I.e.,
$$\begin{aligned}\tilde{\Psi}(\hat{f}_n^*) - \tilde{\Psi}(f_0) &= -P_0\tilde{\phi}^*(\hat{f}_n^*, \hat{\pi}_n) + \tilde{R}(\hat{f}_n^*, \hat{\pi}_n, f_0, \pi_0) \\ &= (\mathbb{P}_n - P_0)\phi^*(\hat{f}_n^*, \hat{\pi}_n) + \tilde{R}(\hat{f}_n^*, \hat{\pi}_n, f_0, \pi_0)\end{aligned}$$

The first term is an empirical process term which equals $(\mathbb{P}_n - P_0)\tilde{\phi}^*(f, \pi)$ plus an $o_P(n^{-1/2})$-term.

This then gives
$$\tilde{\Psi}(\hat{f}_n^*) - \tilde{\Psi}(f_0) = \underbrace{(\mathbb{P}_n - P_0)\tilde{\phi}^*(f, \pi)}_{\text{LLN applies}} + \tilde{R}(\hat{f}_n^*, \hat{\pi}_n, f_0, \pi_0) + o_P(n^{-1/2})$$

which yields that $\tilde{\Psi}(\hat{f}_n^*) - \tilde{\Psi}(f_0) = o_P(1)$ if $\tilde{R}(\hat{f}_n^*, \hat{\pi}_n, f_0, \pi_0) = o_P(1)$.

Introduction to TMLE

EXAMPLE: Average treatment effect (ATE)

1. The efficient influence function:

$$\phi^*(P)(O) = \tilde{\phi}^*(f,\pi)(O)$$
$$= \left(\frac{A}{\pi(A \mid X)} - \frac{1-A}{\pi(A \mid X)}\right)\big(Y - f(A,X)\big) + f(1,X) - f(0,X) - \Psi(P)$$

2. The remainder:

$$R(P,P_0) = \tilde{R}(f,\pi,f_0,\pi_0)$$
$$= \int_{\mathbb{R}^d} \sum_{a=0,1} (2a-1)\frac{\pi_0(a \mid x) - \pi(a \mid x)}{\pi(a \mid x)}\big(f_0(a,x) - f(a,x)\big) d\mu_{0,X}(x)$$

Deriving these is done once for a given target parameter $\Psi : \mathcal{M} \to \mathbb{R}$.

Introduction to TMLE

Conditions (asymptotic linearity and efficiency)

(C1) Solve the efficient influence curve equation: $\mathbb{P}_n \phi^*(\hat{P}_n) = o_P(n^{-1/2})$

(C2) Remainder $R(\hat{P}_n, P_0) = o_P(n^{-1/2})$

(C3) Donsker class conditions for $\{\phi^*(P) : P \in \mathcal{M}\}$

Then: $\Psi(\hat{P}_n) \overset{as}{\sim} N(\Psi(P_0), P_0\phi^*(P_0)^2/n)$

Introduction to TMLE

Conditions (asymptotic linearity and efficiency)

(C1) Solve the efficient influence curve equation: $\mathbb{P}_n \phi^*(\hat{P}_n) = o_P(n^{-1/2})$

(C2) Remainder $R(\hat{P}_n, P_0) = o_P(n^{-1/2})$

(C3) Donsker class conditions for $\{\phi^*(P) : P \in \mathcal{M}\}$

Then: $\Psi(\hat{P}_n) \overset{as}{\sim} N(\Psi(P_0), P_0\phi^*(P_0)^2/n)$

TMLE is a two-step procedure:

Step 1 Construct initial estimator $\hat{P}_n$ for P such that $R(\hat{P}_n, P_0) = o_P(n^{-1/2})$.

Step 2 Update the estimator $\hat{P}_n \mapsto \hat{P}_n^*$ such that $\hat{P}_n^*$ solves the efficient influence curve equation.

Introduction to TMLE: Summary so far

- The role of the targeting step (Step 2):
  - Gaining double robustness in consistency.
  - Easier to get rid of second-order remainder.
- The role of the initial estimation step (Step 1):
  - This should be done well enough to get rid of the second-order remainder.

Introduction to TMLE: Overview of today

Day 2: 8 – 9

Introduction to TMLE. The constructive proof of TMLE.

- ▷ The decomposition and the role of Step 1 & Step 2.

Day 2: 9 – 12

The targeting step. Solving the efficient influence curve equation.

- ▷ Implementation of the targeting step.

Lunch.

Day 2: 13 – 15

Step 2: Super learning. (Thomas).