Introduction to targeted maximum likelihood estimation

Laura Fuentes-Vicente[1], Antoine Chambaz[2], Julie Josse[1]

1 Inria PreMeDICaL, Inserm, University of Montpellier, France
2 Paris Cité University, CNRS, MAP5, F-75006 Paris, France

Team seminar
June 10 2026, Montpellier France

Complete data structure and parameter of interest

$\mathbb{D} = (X, Y(0), Y(1), A, Y) \sim \mathbb{P}$

- X: vector of baseline covariates
- $Y(a)$, for $a \in \{0, 1\}$: potential outcome had the patient undergone treatment a (typically unobserved)
- A: binary treatment assignment
- $Y = AY(1) + (1 - A)Y(0)$: factual outcome (observed)

Complete data structure and parameter of interest

$\mathbb{D} = (X, Y(0), Y(1), A, Y) \sim \mathbb{P}$

- X: vector of baseline covariates
- $Y(a)$, for $a \in \{0, 1\}$: potential outcome had the patient undergone treatment a (typically unobserved)
- A: binary treatment assignment
- $Y = AY(1) + (1 - A)Y(0)$: factual outcome (observed)

Let us consider the average treatment effect (ATE) as target parameter

$$\mathbb{P} \mapsto E_{\mathbb{P}}[Y(1) - Y(0)]$$

defined in terms of potential outcomes

Observational data and identifiability

Observational data: n i.i.d. samples $\mathcal{D}_1, \ldots, \mathcal{D}_n \sim P \in \mathcal{M}$
A generic data structure $\mathcal{D} = (X, A, Y)$:

Relevant features:

- $\mu_P(X, a) = E_P[Y|X, A = a]$
- $e_P(X, a) = P(A = a|X)$ (propensity score)

Observational data and identifiability

Observational data: n i.i.d. samples $\mathcal{D}_1, \ldots, \mathcal{D}_n \sim P \in \mathcal{M}$
A generic data structure $\mathcal{D} = (X, A, Y)$:

Relevant features:

- $\mu_P(X, a) = E_P[Y|X, A = a]$
- $e_P(X, a) = P(A = a|X)$ (propensity score)

The causal estimand can be identified under *standard assumptions* as

$$\psi(P) \mapsto E_P\left[\mu_P(X, 1) - \mu_P(X, 0)\right]$$

Observational data and identifiability

Observational data: n i.i.d. samples $\mathcal{D}_1, \ldots, \mathcal{D}_n \sim P \in \mathcal{M}$
A generic data structure $\mathcal{D} = (X, A, Y)$:

Relevant features:

- $\mu_P(X, a) = E_P[Y|X, A = a]$
- $e_P(X, a) = P(A = a|X)$ (propensity score)

The causal estimand can be identified under *standard assumptions* as

$$\psi(P) \mapsto E_P\left[\mu_P(X, 1) - \mu_P(X, 0)\right]$$

Our goal: build a good estimator of $\psi(P)$ using $\mathcal{D}_1, \ldots, \mathcal{D}_n$

Plug-in estimator

Plug-in estimators are **substitution estimators**

Learn the nuisance parameters: build an estimator μ_n of μ_P and define the plug-in estimator of $\psi(P)$ as

$$\psi_n = \frac{1}{n} \sum_{i=1}^{n} \mu_n(X_i, 1) - \mu_n(X_i, 0)$$

Plug-in estimator

Plug-in estimators are **substitution estimators**

Learn the nuisance parameters: build an estimator μ_n of μ_P and define the plug-in estimator of $\psi(P)$ as

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} \mu_n(X_i, 1) - \mu_n(X_i, 0)$$

Using a first-order expansion, the plug-in estimator can be decomposed as:

$$\psi_n = \psi(P) - E_P[\varphi_{\psi,n}(\mathcal{D})] + {\color{red}R(P, P_n)},$$

where:

- $\varphi_{\psi,n}$ estimates the **efficient influence function (EIF)**
  Intuition: A function describing how the estimator behaves under slight perturbations of the data-generating distribution

Plug-in estimator

Plug-in estimators are **substitution estimators**

Learn the nuisance parameters: build an estimator μ_n of μ_P and define the plug-in estimator of $\psi(P)$ as

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} \mu_n(X_i, 1) - \mu_n(X_i, 0)$$

Using a first-order expansion, the plug-in estimator can be decomposed as:

$$\psi_n = \psi(P) - E_P[\varphi_{\psi,n}(\mathcal{D})] + R(P, P_n),$$

where:

- $\varphi_{\psi,n}$ estimates the **efficient influence function (EIF)**
  Intuition: A function describing how the estimator behaves under slight perturbations of the data-generating distribution
- $R(P, P_n)$ is a second order remainder term, typically $o_P(1/\sqrt{n})$

Plug-in bias

Under the assumption that $\varphi_{\psi,n}$ *(a)* is a $L^2(P)$-consistent estimator of $\varphi_\psi(P)$, and *(b)* lives in a P-Donsker class, it holds that

$$\begin{aligned}\psi_n - \psi(P) &= -E_P[\varphi_{\psi,n}(\mathcal{D})] + R(P, P_n) \\ &= -\frac{1}{n}\sum_{i=1}^{n}\varphi_{\psi,n}(\mathcal{D}_i) + \frac{1}{n}\sum_{i=1}^{n}\varphi_\psi(P)(\mathcal{D}_i) + o_P(1/\sqrt{n}) + R(P, P_n)\end{aligned}$$

Asymptotically linear estimator properties

Definition 1

An estimator ψ_n based on $\mathcal{D}_1, \ldots, \mathcal{D}_n$ is **asymptotically linear** with influence function IF if it can be written as

$$\psi_n = \psi(P) + \frac{1}{n}\sum_{i=1}^{n} IF(\mathcal{D}_i) + o_P(1/\sqrt{n})$$

with $IF : \mathcal{D} \to \mathbb{R}$ such that $E_P[IF(\mathcal{D})] = 0, \mathrm{Var}_P(IF(\mathcal{D})) < \infty$.
Asymptotically linear estimators are asymptotically normal:

$$\sqrt{n}(\psi_n - \psi(P)) \underset{n\to\infty}{\overset{P}{\rightsquigarrow}} \mathcal{N}(0, \mathrm{Var}_P(IF(\mathcal{D})))$$

Asymptotically linear estimator properties

Definition 1

An estimator ψ_n based on $\mathcal{D}_1, \dots, \mathcal{D}_n$ is **asymptotically linear** with influence function *IF* if it can be written as

$$\psi_n = \psi(P) + \frac{1}{n}\sum_{i=1}^{n} IF(\mathcal{D}_i) + o_P(1/\sqrt{n})$$

with $IF : \mathcal{D} \rightarrow \mathbb{R}$ such that $E_P[IF(\mathcal{D})] = 0, \mathrm{Var}_P(IF(\mathcal{D})) < \infty$.
Asymptotically linear estimators are asymptotically normal:

$$\sqrt{n}(\psi_n - \psi(P)) \underset{n\rightarrow\infty}{\overset{P}{\rightsquigarrow}} \mathcal{N}(0, \mathrm{Var}_P(IF(\mathcal{D})))$$

$$\psi_n = \psi(P) \boxed{-\frac{1}{n}\sum_{i=1}^{n} \varphi_{\psi,n}(\mathcal{D}_i)} + \frac{1}{n}\sum_{i=1}^{n} \underbrace{\varphi_{\psi}(P)(\mathcal{D}_i)}_{IF(\mathcal{D}_i)} + o_P(1/\sqrt{n}) + R(P, P_n)$$

Plug-in estimator properties

- Plug-in estimators are **substitution estimators**
- The random term , is known and can drive the asymptotic behavior of ψ_n
- This motivates bias-corrected estimators such as the one-step estimator and TMLE

From plug-in bias to one step correction

Intuition: Update ψ_n by directly correcting the first order plug-in bias in the parameter space

From plug-in bias to one step correction

Intuition: Update ψ_n by directly correcting the first order plug-in bias in the parameter space

Let us first decompose the such term:

From plug-in bias to one step correction

Intuition: Update ψ_n by directly correcting the first order plug-in bias in the parameter space

Let us first decompose the such term:

$$\boxed{-\frac{1}{n}\sum_{i=1}^{n}\varphi_{\psi,n}(\mathcal{D}_i)} = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{e_n(X_i,1)}+\frac{1-A_i}{e_n(X_i,0)}\right)\cdot(Y_i-\mu_n(X_i,A_i)) + \mu_n(X_i,1)-\mu_n(X_i,0)-\psi_n$$

$$= \boxed{-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{e_n(X_i,1)}+\frac{1-A_i}{e_n(X_i,0)}\right)\cdot(Y_i-\mu_n(X_i,A_i))}$$

From plug-in bias to one step correction

Intuition: Update ψ_n by directly correcting the first order plug-in bias in the parameter space

Let us first decompose the such term:

$$-\frac{1}{n}\sum_{i=1}^{n}\varphi_{\psi,n}(\mathcal{D}_i) = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{e_n(X_i,1)} + \frac{1-A_i}{e_n(X_i,0)}\right)\cdot(Y_i - \mu_n(X_i,A_i))$$

$$+ \mu_n(X_i,1) - \mu_n(X_i,0) - \psi_n$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{e_n(X_i,1)} + \frac{1-A_i}{e_n(X_i,0)}\right)\cdot(Y_i - \mu_n(X_i,A_i))$$

The one-step estimator is defined as

$$\psi_n^{\mathrm{os}} = \psi_n + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{e_n(X_i,1)} + \frac{1-A_i}{e_n(X_i,0)}\right)\cdot(Y_i - \mu_n(X_i,A_i))$$

One-step estimator

Therefore,

$$\begin{aligned}
\psi_n^{\mathrm{os}} - \psi(P) &= \psi_n - \psi(P) + \boxed{\frac{1}{n}\sum_{i=1}^{n} \varphi_{\psi,n}(D_i)} \\
&= -E_P[\varphi_{\psi,n}(\mathcal{D})] + \color{red}{R(P, P_n)} + \boxed{\frac{1}{n}\sum_{i=1}^{n} \varphi_{\psi,n}(D_i)} \\
&= \frac{1}{n}\sum_{i=1}^{n} \varphi_\psi(P)(\mathcal{D}_i) + o_P(1/\sqrt{n}) + \color{red}{R(P, P_n)}
\end{aligned}$$

One-step estimator properties

The one-step estimator is **not a substitution estimator**

One-step estimator properties

The one-step estimator is **not a substitution estimator**

Asymptotically linear estimator with IF $\varphi_\psi(P)$
By the central limit theorem,

$$\sqrt{n}(\psi_n^{\mathrm{os}} - \psi(P)) \underset{n \to \infty}{\overset{P}{\rightsquigarrow}} \mathcal{N}(0, \mathrm{Var}_P(\varphi_\psi(P)(\mathcal{D})))$$

One-step estimator properties

The one-step estimator is **not a substitution estimator**

Asymptotically linear estimator with IF $\varphi_\psi(P)$
By the central limit theorem,

$$\sqrt{n}(\psi_n^{\mathrm{os}} - \psi(P)) \underset{n\to\infty}{\overset{P}{\rightsquigarrow}} \mathcal{N}(0, \mathrm{Var}_P(\varphi_\psi(P)(\mathcal{D})))$$

Double robust estimator:

$R(P, P_n)^2 \leq \|\mu_n - \mu_P\|_P^2 \times \|(e_n - e_P)/\ell e_P\|_P^2$,
if either $\mu_n = \mu_P$ or $e_P = e_n$, $R(P, P_n) = 0$ giving

$$\psi_n^{\mathrm{os}} = \psi(P) + \frac{1}{n}\sum_{i=1}^{n} \varphi_\psi(P)(\mathcal{D}_i) + o_P(1/\sqrt{n})$$

One-step estimator properties

The one-step estimator is **not a substitution estimator**

Asymptotically linear estimator with IF $\varphi_\psi(P)$
By the central limit theorem,

$$\sqrt{n}(\psi_n^{\mathrm{os}} - \psi(P)) \underset{n\to\infty}{\overset{P}{\rightsquigarrow}} \mathcal{N}(0, \mathrm{Var}_P(\varphi_\psi(P)(\mathcal{D})))$$

Double robust estimator:

$R(P, P_n)^2 \leq \|\mu_n - \mu_P\|_P^2 \times \|(e_n - e_P)/\ell e_P\|_P^2$,
if either $\mu_n = \mu_P$ or $e_P = e_n$, $R(P, P_n) = 0$ giving

$$\psi_n^{\mathrm{os}} = \psi(P) + \frac{1}{n}\sum_{i=1}^{n} \varphi_\psi(P)(\mathcal{D}_i) + o_P(1/\sqrt{n})$$

For binary outcomes, ψ_n^{os} can step outside $[-1, 1]$
(due to propensity score weighting $\cdot/e_n(X, a)$ for $a \in \{0, 1\}$)

Targeted learning motivation

Targeted maximum likelihood estimation (TMLE) performs the correction in the nuisance model space.

Targeted learning motivation

Targeted maximum likelihood estimation (TMLE) performs the correction in the nuisance model space.

How? Update μ_n into a targeted estimator μ_n^*, hence, building an update $\varphi^*_{\psi,n}$ of $\varphi_{\psi,n}$ (which also estimates $\varphi_\psi(P)$) such that

$$\boxed{\frac{1}{n}\sum_{i=1}^{n}\varphi^*_{\psi,n}(\mathcal{D}_i)} \approx 0$$

Targeted maximum likelihood estimation (TMLE)

Targeted maximum likelihood estimation (TMLE) proceeds in two steps:

1. **Learn the nuisance parameters**: build estimators μ_n and e_n of μ_P and e_P

Targeted maximum likelihood estimation (TMLE)

Targeted maximum likelihood estimation (TMLE) proceeds in two steps:

1. **Learn the nuisance parameters**: build estimators μ_n and e_n of μ_P and e_P
2. **Targeting step**: update μ_n into μ_n^* to mitigate the first order plug-in bias

Targeted maximum likelihood estimation (TMLE)

Targeted maximum likelihood estimation (TMLE) proceeds in two steps:

1. **Learn the nuisance parameters**: build estimators μ_n and e_n of μ_P and e_P
2. **Targeting step**: update μ_n into μ_n^* to mitigate the first order plug-in bias

Targeting intuition

Can be viewed as regressing the residuals on the clever covariate (i.e. small fluctuation of μ_n along a direction given by the EIF)

$$Y - \mu_n(X, A) \approx \epsilon H(X, A)$$

where $H(X, A)$ is a function of e_n, defined from $\varphi_{\psi,n}(\mathcal{D}_i)$

Targeted maximum likelihood estimation (TMLE)

Targeted maximum likelihood estimation (TMLE) proceeds in two steps:

1. **Learn the nuisance parameters**: build estimators μ_n and e_n of μ_P and e_P
2. **Targeting step**: update μ_n into μ_n^* to mitigate the first order plug-in bias

Targeting intuition

Can be viewed as regressing the residuals on the clever covariate (i.e. small fluctuation of μ_n along a direction given by the EIF)

$$Y - \mu_n(X, A) \approx \epsilon H(X, A)$$

where $H(X, A)$ is a function of e_n, defined from $\varphi_{\psi,n}(\mathcal{D}_i)$ as

$$\underbrace{\left(\frac{A_i}{e_n(X_i, 1)} - \frac{1 - A_i}{e_n(X_i, 0)}\right)}_{=H(X,A)} (Y_i - \mu_n(X_i, A_i)) + \mu_n(X_i, 1) - \mu_n(X_i, 0) - \psi_n$$

Targeted maximum likelihood estimation (TMLE)

Targeted maximum likelihood estimation (TMLE) proceeds in two steps:

1. **Learn the nuisance parameters**: build estimators μ_n and e_n of μ_P and e_P
2. **Targeting step**: update μ_n into μ_n^* to mitigate the first order plug-in bias

   a. Define a parametric submodel $\mu_{n,\epsilon}$
   (i.e. express Y using μ_n, $H(X, A)$ and $\epsilon \in \mathbb{R}$)

   b. Minimize a measure of discrepancy between Y and $\mu_{n,\epsilon}$
   $\hat{\epsilon} \in \text{argmin}_{\epsilon \in \mathbb{R}} \ell_n(\epsilon)$ to define $\mu_n^* = \mu_{n,\hat{\epsilon}}$

Focus on step 2 (1/2)

2. Update μ_n into μ_n^* to correct plug-in bias **(continuous outcomes)**:
   a. Define a parametric submodel

$$\mu_{n,\epsilon}(X, A) = \mu_n(X, A) + \epsilon H(X, A)$$

Focus on step 2 (1/2)

2. Update μ_n into μ_n^* to correct plug-in bias **(continuous outcomes)**:

   a. Define a parametric submodel

$$\mu_{n,\epsilon}(X, A) = \mu_n(X, A) + \epsilon H(X, A)$$

   b. Define $\mu_n^* = \mu_{n,\hat{\epsilon}}$, where $\hat{\epsilon}$ solves

$$\hat{\epsilon} \in \underset{\epsilon \in \mathbb{R}}{\operatorname{argmin}} -\frac{1}{n}\sum_{i=1}^{n}[Y_i - \mu_{n,\epsilon}(X_i, A_i)]^2$$

which is such that

$$\boxed{\frac{1}{n}\sum_{i=1}^{n}\varphi^*_{\psi,n}(\mathcal{D}_i)} \approx 0$$

Proof

The derivative of $\mu_{n,\epsilon}(X, A) = \mu_n(X, A) + \epsilon H(X, A))$ w.r.t. ϵ:

$$\frac{d}{d\epsilon}\mu_{n,\epsilon}(X, A) = H(X, A)$$

Proof

The derivative of $\mu_{n,\epsilon}(X, A) = \mu_n(X, A) + \epsilon H(X, A))$ w.r.t. ϵ:

$$\frac{d}{d\epsilon}\mu_{n,\epsilon}(X, A) = H(X, A)$$

Therefore, $\hat{\epsilon} \in \operatorname{argmin}_{\epsilon \in \mathbb{R}} \ell_n(\epsilon)$ solves $\frac{d}{d\epsilon}\ell_n(\epsilon) = 0$:

$$\frac{d}{d\epsilon}\ell_n(\epsilon) = \frac{2}{n}\sum_{i=1}^{n}(Y_i - \mu_{n,\epsilon}(X_i, A_i))H(X_i, A_i)$$

giving the desired result

Focus step 2 (2/2)

2. Update μ_n into μ_n^* to correct plug-in bias **(binary outcomes)**:
   a. Define a parametric submodel

$$\mu_{n,\epsilon}(X, A) = \text{expit}(\text{logit}(\mu_n(X, A)) + \epsilon H(X, A))$$

Focus step 2 (2/2)

2. Update μ_n into μ_n^* to correct plug-in bias **(binary outcomes)**:
   a. Define a parametric submodel

$$\mu_{n,\epsilon}(X, A) = \mathrm{expit}(\mathrm{logit}(\mu_n(X, A)) + \epsilon H(X, A))$$

   b. Define $\mu_n^* = \mu_{n,\hat{\epsilon}}$, where $\hat{\epsilon}$ solves

$$\hat{\epsilon} \in \underset{\epsilon \in \mathbb{R}}{\mathrm{argmin}} - \frac{1}{n} \sum_{i=1}^{n} [Y_i \log(\mu_{n,\epsilon}(X_i, A_i)) + (1 - Y_i) \log(1 - \mu_{n,\epsilon}(X_i, A_i))]$$

   such that

$$\boxed{\frac{1}{n} \sum_{i=1}^{n} \varphi^*_{\psi,n}(\mathcal{D}_i) \approx 0}$$

Proof

The derivative of $\mu_{n,\epsilon}(X, A) = \text{expit}(\text{logit}(\mu_n(X, A)) + \epsilon H(X, A))$ w.r.t. ϵ:

$$\frac{d}{d\epsilon}\mu_{n,\epsilon}(X, A) = H(X, A)\mu_{n,\epsilon}(X, A)(1 - \mu_{n,\epsilon}(X, A))$$

Proof

The derivative of $\mu_{n,\epsilon}(X, A) = \text{expit}(\text{logit}(\mu_n(X, A)) + \epsilon H(X, A))$ w.r.t. ϵ:

$$\frac{d}{d\epsilon}\mu_{n,\epsilon}(X, A) = H(X, A)\mu_{n,\epsilon}(X, A)(1 - \mu_{n,\epsilon}(X, A))$$

Therefore, $\hat{\epsilon} \in \text{argmin}_{\epsilon \in \mathbb{R}} \ell_n(\epsilon)$ solves $\frac{d}{d\epsilon}\ell_n(\epsilon) = 0$:

$$\begin{aligned}
\frac{d}{d\epsilon}\ell_n(\epsilon) &= -\frac{1}{n}\sum_{i=1}^{n} Y_i \frac{\frac{d}{d\epsilon}\mu_{n,\epsilon}(X_i, A_i)}{\mu_{n,\epsilon}(X_i, A_i)} - (1 - Y_i)\frac{\frac{d}{d\epsilon}\mu_{n,\epsilon}(X_i, A_i)}{1 - \mu_{n,\epsilon}(X_i, A_i)} \\
&= -\frac{1}{n}\sum_{i=1}^{n} Y_i H(X_i, A_i)(1 - \mu_{n,\epsilon}(X_i, A_i)) - (1 - Y_i)H(X_i, A_i)\mu_{n,\epsilon}(X_i, A_i) \\
&= -\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_{n,\epsilon}(X_i, A_i))H(X_i, A_i)
\end{aligned}$$

which gives the desired result

TMLE estimator

Define the TMLE estimator as

$$\psi_n^* = \frac{1}{n}\sum_{i=1}^{n} \mu_n^*(X_i, 1) - \mu_n^*(X_i, 0)$$

Continuous case:

$$\mu_n^*(X, 1) = \mu_n(X, 1) + \hat{\epsilon} H(X, 1)$$
$$\mu_n^*(X, 0) = \mu_n(X, 0) + \hat{\epsilon} H(X, 0)$$

Binary case:

$$\mu_n^*(X, 1) = \text{expit}(\text{logit}(\mu_n(X, 1)) + \hat{\epsilon} H(X, 1))$$
$$\mu_n^*(X, 0) = \text{expit}(\text{logit}(\mu_n(X, 0)) + \hat{\epsilon} H(X, 0))$$

TMLE estimator

Define the TMLE estimator as

$$\psi_n^* = \frac{1}{n}\sum_{i=1}^{n} \mu_n^*(X_i, 1) - \mu_n^*(X_i, 0)$$

Continuous case:

$$\mu_n^*(X,1) = \mu_n(X,1) + \hat{\epsilon} H(X,1)$$
$$\mu_n^*(X,0) = \mu_n(X,0) + \hat{\epsilon} H(X,0)$$

Binary case:

$$\mu_n^*(X,1) = \text{expit}(\text{logit}(\mu_n(X,1)) + \hat{\epsilon} H(X,1))$$
$$\mu_n^*(X,0) = \text{expit}(\text{logit}(\mu_n(X,0)) + \hat{\epsilon} H(X,0))$$

Therefore,

$$\begin{aligned}\psi_n^* - \psi(P) &= -E_P[\varphi^*_{\psi,n}] + R(P, P_n^*) \\ &= \underbrace{-\frac{1}{n}\sum_{i=1}^{n} \varphi^*_{\psi,n}(\mathcal{D}_i)}_{\approx 0} + \frac{1}{n}\sum_{i=1}^{n} \varphi_\psi(P)(\mathcal{D}_i) + o_P(1/\sqrt{n}) + R(P, P_n^*)\end{aligned}$$

Properties

The TMLE estimator is a **substitution estimator**

Properties

The TMLE estimator is a **substitution estimator**

Asymptotically linear estimator with IF $\varphi_\psi(P)$
By the central limit theorem,

$$\sqrt{n}(\psi_n^* - \psi(P)) \underset{n\to\infty}{\overset{P}{\rightsquigarrow}} \mathcal{N}(0, \mathrm{Var}_P(\varphi_\psi(P)(\mathcal{D})))$$

Properties

The TMLE estimator is a **substitution estimator**

Asymptotically linear estimator with IF $\varphi_\psi(P)$
By the central limit theorem,

$$\sqrt{n}(\psi_n^* - \psi(P)) \underset{n\to\infty}{\overset{P}{\rightsquigarrow}} \mathcal{N}(0, \mathrm{Var}_P(\varphi_\psi(P)(\mathcal{D})))$$

Double robust estimator:

$R(P, P_n)^2 \leq \|\mu_n - \mu_P\|_P^2 \times \|(e_n - e_P)/\ell e_P\|_P^2$,
if either $\mu_n = \mu_P$ or $e_P = e_n$, $R(P, P_n) = 0$ giving

$$\psi_n^* = \psi(P) + \frac{1}{n}\sum_{i=1}^{n} \varphi_\psi(P)(\mathcal{D}_i) + o_P(1/\sqrt{n})$$

Numerical application

Let's switch to R...

Thank you!

TMLE References: Books

1. Rubin, D., & van der Laan, M. (2006). Targeted Maximum Likelihood Learning. The International Journal of Biostatistics, 2(1), 1–38. https://doi.org/10.2202/1557-4679.1043
2. Laan, Mark J. van der, and Sherri Rose. 2011. Targeted Learning. Springer Series in Statistics. Springer, New York. https://doi.org/10.1007/978-1-4419-9782-1.
3. Targeted Learning in Data Science. Springer Series in Statistics. Springer, Cham. https://doi.org/10.1007/978-3-319-65304-4

TMLE References: Practical guides with R code

1. A Ride in Targeted Learning Territory
2. An Illustrated Guide to TMLE

References: R-packages

There are several available R-packages

- tmle
  Gruber, S., & Laan, M. van der. (2012). tmle: An R Package for Targeted Maximum Likelihood Estimation. Journal of Statistical Software, 51(13), 1–35. https://doi.org/10.18637/jss.v051.i13
- tmle3