Doubly Robust Estimation of Individuall Tratment Regime in a Semi-supervised Framework

Xintong Li

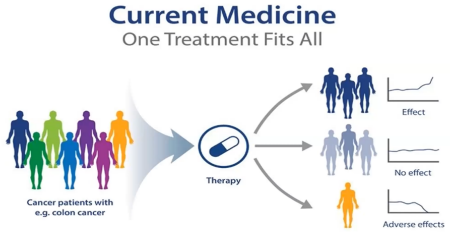
Tsinghua University

Joint work with Mengjiao Peng and Yong Zhou

Outline

Precision Medicine

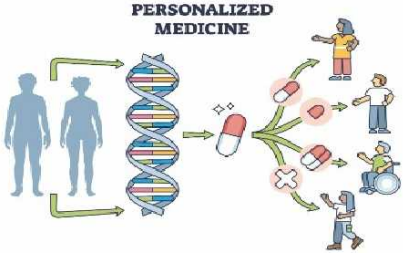
- **Heterogeneity**: different patients respond differently to the same treatment.
  - Positive treatment effects;
  - Side effects.
- **One-size-fits-all** $\rightarrow$ **Precision Medicine**
- **Advantages**:
  - Improve patient adherence;
  - Reduce unnecessary treatments and side effects;
  - Promote recovery;
  - Enhance quality of care and quality of life;
  - Optimize allocation of medical resources;
  - Lower overall healthcare costs;
  - .....



Precision Medicine - Personalized Decision-Making

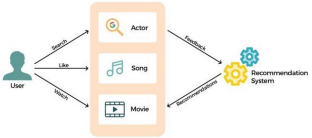
- **Goal**: Find the optimal mapping from individual characteristics $\mathbf{X} \in \mathcal{X}$ to treatments $A \in \mathcal{A}$, i.e. $d^{opt}(\mathbf{X})$, to maximize the expected clinical outcome $E[Y^*(d(\mathbf{X}))]$.
  - $\mathbf{X}$: demographics, clinical features, genetic information, environmental factors, etc.;
  - A: drug choice, dosage, surgery, specific dietary or exercise recommendations, etc.;
  - Y: biomarker levels, survival time, disease progression or remission status, quality of life scores, etc.

- **Applications**:
  - **Disease management**: Recommend the optimal drug dosage based on patient characteristics to optimize treatment efficacy;
  - **Smart health monitoring**: Use wearable devices and biosensors for personalized health management;
  - **Personalized medical intervention**: Combine multimodal data to predict disease risk and enable early intervention.



Personalized Decision-Making - Beyond Precision Medicine

- **Computer Science**: context-aware recommender systems that improve accuracy by incorporating time, location, and social context.



- **Finance**: provide personalized investment advice and wealth management plans based on consumption habits and risk preferences.

- **Public Management**: improve the overall effectiveness of policies through personalized interventions targeting individuals with high social connectivity.

Traditional Methods

- **Q-learning** (Qian and Murphy, 2011; Watkins and Dayan, 1992)

  Define the Q-function: $Q(\mathbf{x}, a) := E[Y|\mathbf{X} = \mathbf{x}, A = a]$, and specify a model $Q(\mathbf{X}, A; \boldsymbol{\beta})$.

$$\widehat{d}^{opt}(\mathbf{X}) = \arg\max_{a \in \mathcal{A}} Q(\mathbf{X}, a; \widehat{\boldsymbol{\beta}}),$$

  where $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} (Y_i - Q(\mathbf{X}_i, A_i; \boldsymbol{\beta}))^2$.

- **A-learning** (Murphy, 2003; Robins, 2004)

  Define the contrast function: $C(\mathbf{X}) = Q(\mathbf{X}, 1) - Q(\mathbf{X}, 0)$, then $d^{opt}(\mathbf{X}) = I(C(\mathbf{X}) \geqslant 0)$.

  Doubly robust A-learning: Let $\nu(\mathbf{X}) = E[Y|\mathbf{X}]$ and $\pi(\mathbf{X}) = E[A|\mathbf{X}]$, with corresponding estimators $\widehat{\nu}(\mathbf{X})$ and $\widehat{\pi}(\mathbf{X})$. Specify a model for the contrast function $C(\mathbf{X}; \boldsymbol{\theta})$, then

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{n} \sum_{i=1}^{n} \{Y_i - \widehat{\nu}(\mathbf{X}_i) - [A_i - \widehat{\pi}(\mathbf{X}_i)] C(\mathbf{X}_i; \boldsymbol{\theta})\}^2.$$

Traditional Methods

- **Direct-search Methds** (Chu et al., 2023; Zhang et al., 2012)
  Denote value function $V(d(\mathbf{X})) := E[Y^*(d(\mathbf{X}))]$, then $d^{opt}(\mathbf{X}) = \underset{d(\mathbf{X})\in\mathcal{D}}{\arg\max}\, V(d(\mathbf{X}))$.
  - IPW-based estimator: $\widehat{V}_{IPW}(d(\mathbf{X})) = P_n\left[\frac{I(A=d(\mathbf{X}))}{\widehat{\pi}(\mathbf{X},A)}Y\right]$.
  - AIPW-based estimator: $\widehat{V}_{AIPW}(d(\mathbf{X})) = P_n\left[\frac{I(A=d(\mathbf{X}))}{\widehat{\pi}(\mathbf{X},A)}Y - \frac{I(A=d(\mathbf{X}))-\widehat{\pi}(\mathbf{X},A)}{\widehat{\pi}(\mathbf{X},A)}\widehat{Q}(\mathbf{X},d(\mathbf{X}))\right]$, where $Q(\mathbf{X},d(\mathbf{X})) = Q(\mathbf{X},1)I(d(\mathbf{X})=1) + Q(\mathbf{X},0)I(d(\mathbf{X})=0)$.

Key Challenges

- **Lack of theoretical guarantees** for parameter estimation of optimal ITRs $\rightarrow$ valid inference is hindered (Zhang et al., 2012).
- **Scarcity of labeled data** $\rightarrow$ large amounts of unlabeled data remain underutilized (Liao et al., 2010).
- **Curse of dimensionality** $\rightarrow$ Imputation-based kernel methods are impractical for multi-dimensional covariates. (Gunn et al., 2024).

Our Contributions

- **Semi-supervised framework:** Propose a **doubly robust direct-search method** that leverages both labeled and unlabeled data to improve estimation efficiency and robustness.
- **Dimension reduction:** Incorporate a **projection-based** technique to address multi-dimensional covariates.
- **Theorical result:** Establish an $n^{-1/3}$ **convergence rate** for parameter estimation, along with its nonstandard asymptotic distribution.
- **Inference:** Develop a **perturbation resampling method** to enable valid statistical inference.

Outline

Methodology

Notations $(\mathbf{X}, A, Y)$

$\mathbf{X} \in \mathcal{X} \subseteq \mathbb{R}^p$: p-dimensional covariates with bounded support $\mathcal{X}$;

$A \in \mathcal{A} = \{0, 1\}$: the binary treatment indicator;

$Y \in \mathcal{Y} \subseteq \mathbb{R}$: the outcome variable, larger values are better.

Observations $\mathcal{L} \cup \mathcal{U}$

$\mathcal{L} = \{(\mathbf{X}_i, A_i, Y_i) : i = 1, 2, \ldots, n\}$: n iid labeled observations;

$\mathcal{U} = \{(\mathbf{X}_i, A_i) : i = n+1, n+2, \ldots, n+N\}$: N iid unlabeled observations.

Methodology

Semi-supervised Assumptions

a. $\mathcal{L} \perp \mathcal{U}$;

b. Observations in $\mathcal{L}$ and $\mathcal{U}$ potentially follow the same distribution;

c. $\frac{n}{N} = \rho_n \to \rho \in [0, \infty)$ as $n, N \to \infty$.

Identifiability Assumptions

Let $Y^*(a)$ be the potential outcome.

a. SUTVA: $Y = Y^*(1)A + Y^*(0)(1 - A)$;

b. Ignoreability: $A \perp \{Y^*(0), Y^*(1)\} \mid \mathbf{X}$;

c. Positivity: $0 < P(A = a|\mathbf{X} = \mathbf{x}) < 1$.

Methodology

For any ITR $d(\mathbf{X}) : \mathcal{X} \to \mathcal{A}$, the potential outcome $Y^*(d(\mathbf{X}))$ can be written by

$$Y^*(d(\mathbf{X})) = Y^*(1)d(\mathbf{X}) + Y^*(0)(1 - d(\mathbf{X})).$$

The optimal ITR is defined as

$$d^{\text{opt}}(\mathbf{X}) = \arg\max_{d \in \mathcal{D}} E\left[Y^*\{d(\mathbf{X})\}\right],$$

where $\mathcal{D}$ is some decision class of interest. We will focus on the **linear decision class**

$$\mathcal{D} = \{d_{\boldsymbol{\beta}}(\mathbf{X}) = I(\boldsymbol{\beta}'\mathbf{X} \geqslant 0) : \boldsymbol{\beta} \in \mathcal{B}\},$$

where $\mathcal{B} = \{\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^p, \|\boldsymbol{\beta}\| = 1\}$, due to its **simple structure** and **good interpretability** (Chu et al., 2023; Fan et al., 2017; Li et al., 2025)..

Methodology

Lemma 1

Denote the conditional average treatment effect (CATE) as $D(\mathbf{X}) = E(Y|\mathbf{X}, A = 1) - E(Y|\mathbf{X}, A = 0)$. *Under identification assumptions, we have*

$$E[\{Y^*(1) - Y^*(0)\}d_{\boldsymbol{\beta}}(\mathbf{X})] = E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})].$$

From Lemma 1 and the definition of optimal ITR, we have

$$d_{\boldsymbol{\beta}}^{\text{opt}} = \arg\max_{d_{\boldsymbol{\beta}} \in \mathcal{D}} E[Y^*(d_{\boldsymbol{\beta}}(\mathbf{X}))] d_{\boldsymbol{\beta}}^{\text{opt}} = \arg\max_{d_{\boldsymbol{\beta}} \in \mathcal{D}} E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})].$$

Define $E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})]$ as the value function and $I(\boldsymbol{\beta}_0'\mathbf{X} \geqslant 0)$ the induced optimal ITR in $\mathcal{D}$, where $\boldsymbol{\beta}_0 = \arg\max_{\boldsymbol{\beta} \in \mathcal{B}} E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})]$.

Methodology

Combining the ideas of the **direct-search method** from Zhang et al. (2012) and **robust A-learning** (Murphy, 2003), we can construct the consistent estimator of the value function as follows:

$$E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})] = E[V(\mathbf{Z},\boldsymbol{\theta})d_{\boldsymbol{\beta}}(\mathbf{X})] := \Delta(\boldsymbol{\beta},\boldsymbol{\theta}),$$

where $\mathbf{Z} = (\mathbf{X}, Y, A)$ and $V(\mathbf{Z},\boldsymbol{\theta}) = \frac{\{Y-\nu(\mathbf{X},\boldsymbol{\theta})\}\{A-\pi(\mathbf{X})\}}{\pi(\mathbf{X})\{1-\pi(\mathbf{X})\}}$:

- satisfying $E[V(\mathbf{Z},\boldsymbol{\theta})|\mathbf{X}] = D(\mathbf{X})$ (Fan et al., 2017);
- $\pi(\mathbf{X}) = P(A=1|\mathbf{X})$: the propensity score(PS) function;
- $\nu(\mathbf{X},\boldsymbol{\theta})$: a model parameterized by $\boldsymbol{\theta}$ for $\nu(\mathbf{X})$ (an arbitrary function of $\mathbf{X}$), assuming $\nu(\mathbf{X}) = \mu_0(\mathbf{X}) = E(Y|\mathbf{X}, A=0)$ w.l.o.g.

Thus we can obtain the **supervised estimator** of the value function that

$$\widehat{\Delta}_{\text{sup}}(\boldsymbol{\beta},\widehat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n} V(\mathbf{Z}_i,\widehat{\boldsymbol{\theta}})I\left(\boldsymbol{\beta}'\mathbf{X}_i \geqslant 0\right),$$

and the optimal ITR parameter β_0 that $\widehat{\boldsymbol{\beta}}_{\text{sup}} = \underset{\boldsymbol{\beta}\in\mathcal{B}}{\arg\max}\, \widehat{\Delta}_{\text{sup}}(\boldsymbol{\beta},\widehat{\boldsymbol{\theta}})$.

Methodology

By the law of iterated expectations, we have $\Delta(\boldsymbol{\beta}, \boldsymbol{\theta}) = E[m(\boldsymbol{\beta}'\mathbf{X}, \boldsymbol{\theta}) d_{\boldsymbol{\beta}}(\mathbf{X})]$ for $d_{\boldsymbol{\beta}}(\mathbf{X}) = I(\boldsymbol{\beta}'\mathbf{X} \geqslant 0)$, where $m(\boldsymbol{\beta}'\mathbf{X}, \boldsymbol{\theta}) = E[V(\mathbf{Z}, \boldsymbol{\theta}) | \boldsymbol{\beta}'\mathbf{X}]$. It can be estimated by

$$\widehat{m}(\boldsymbol{\beta}'\mathbf{X}_j, \boldsymbol{\theta}) = \frac{n^{-1} \sum_{i=1}^{n} K_h(\boldsymbol{\beta}'\mathbf{X}_i - \boldsymbol{\beta}'\mathbf{X}_j) V(\mathbf{Z}_i, \boldsymbol{\theta})}{n^{-1} \sum_{i=1}^{n} K_h(\boldsymbol{\beta}'\mathbf{X}_i - \boldsymbol{\beta}'\mathbf{X}_j)}, \tag{1}$$

where $K_h(u - v) = \frac{1}{h} K\left(\frac{u-v}{h}\right)$ with K being some kernel function and $h > 0$ being the bandwidth.

Methodology

Since $\Delta(\boldsymbol{\beta}, \boldsymbol{\theta}) = \lambda E[V(\mathbf{Z}, \boldsymbol{\theta}) d_{\boldsymbol{\beta}}(\mathbf{X})] + (1-\lambda) E[m(\boldsymbol{\beta}'\mathbf{X}, \boldsymbol{\theta}) d_{\boldsymbol{\beta}}(\mathbf{X})]$ for any weight $\lambda \in [0, 1]$, we construct the **semi-supervised estimator** of the value function that

$$\widehat{\Delta}_{\lambda}(\boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}) = \frac{\lambda}{n} \sum_{i=1}^{n} V(\mathbf{Z}_i, \widehat{\boldsymbol{\theta}}) I(\boldsymbol{\beta}'\mathbf{X}_i \geqslant 0) + \frac{1-\lambda}{N} \sum_{j=n+1}^{n+N} \widehat{m}(\boldsymbol{\beta}'\mathbf{X}_j, \widehat{\boldsymbol{\theta}}) I(\boldsymbol{\beta}'\mathbf{X}_j \geqslant 0),$$

and $\boldsymbol{\beta}_0$ that $\widehat{\boldsymbol{\beta}}_{\lambda} = \arg\max_{\boldsymbol{\beta} \in \mathcal{B}} \widehat{\Delta}_{\lambda}(\boldsymbol{\beta}, \widehat{\boldsymbol{\theta}})$ *.

Furthermore, we propose the **pooled estimator** of the value function that

$$\widehat{\Delta}_{\text{pl}}(\boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}) = \frac{1}{n+N} \sum_{j=1}^{n+N} \widehat{m}(\boldsymbol{\beta}'\mathbf{X}_j, \widehat{\boldsymbol{\theta}}) I\left(\boldsymbol{\beta}'\mathbf{X}_j \geqslant 0\right),$$

and $\boldsymbol{\beta}_0$ that $\widehat{\boldsymbol{\beta}}_{\text{pl}} = \arg\max_{\boldsymbol{\beta} \in \mathcal{B}} \widehat{\Delta}_{\text{pl}}(\boldsymbol{\beta}, \widehat{\boldsymbol{\theta}})$.

*Theoretical results establish the optimal weight $\lambda = \rho^2/(1+\rho^2)$, which we recommend for applications.

Double Robustness

- When $\pi(\mathbf{X})$ is unknown, we develop a doubly robust estimating method for the class of monotonic increasing index model $D(\mathbf{X})$, which satisfing that $D(\mathbf{X}) \geqslant 0$ for $\boldsymbol{\beta}'\mathbf{X} \geqslant 0$ and $D(\mathbf{X}) < 0$ for $\boldsymbol{\beta}'\mathbf{X} < 0$.

- Denote that $V(\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = \frac{\{Y - \nu(\mathbf{X}, \boldsymbol{\theta})\}\{A - \pi(\mathbf{X}, \boldsymbol{\alpha})\}}{\pi(\mathbf{X}, \boldsymbol{\alpha})\{1 - \pi(\mathbf{X}, \boldsymbol{\alpha})\}}$ and $m(\boldsymbol{\beta}'\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = E[V(\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) | \boldsymbol{\beta}'\mathbf{X}]$, where $\pi(\mathbf{X}, \boldsymbol{\alpha})$ is a parametric working model such as a logistic regression model posited for $\pi(\mathbf{X})$.

Double Robustness

- When $\pi(\mathbf{X}, \boldsymbol{\alpha})$ is correctly specified, it is similar to the case discussed before.
- When $\nu(\mathbf{X}, \boldsymbol{\theta})$ is correctly specified, we have

$$m(\boldsymbol{\beta}'\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = E\left[\frac{AD(\mathbf{X})\{A - \pi(\mathbf{X}, \boldsymbol{\alpha})\}}{\pi(\mathbf{X}, \boldsymbol{\alpha})\{1 - \pi(\mathbf{X}, \boldsymbol{\alpha})\}}\middle|\boldsymbol{\beta}'\mathbf{X}\right] = E\left[D(\mathbf{X})\frac{\pi(\mathbf{X})}{\pi(\mathbf{X}, \boldsymbol{\alpha})}\middle|\boldsymbol{\beta}'\mathbf{X}\right].$$

When $D(\mathbf{X})$ is a monotonic increasing index model, for any positive function $g(\cdot)$, we have

$$d_{\boldsymbol{\beta}}^{\text{opt}} = \arg\max_{d_{\boldsymbol{\beta}} \in \mathcal{D}} E[D(\mathbf{X}) d_{\boldsymbol{\beta}}(\mathbf{X})] = \arg\max_{d_{\boldsymbol{\beta}} \in \mathcal{D}} E[D(\mathbf{X}) g(\mathbf{X}) d_{\boldsymbol{\beta}}(\mathbf{X})].$$

Due to the positivity assumption, $\frac{\pi(\mathbf{X})}{\pi(\mathbf{X}, \boldsymbol{\alpha})}$ is always a positive function, hence

$$d_{\boldsymbol{\beta}}^{\text{opt}} = \arg\max_{d_{\boldsymbol{\beta}} \in \mathcal{D}} E[D(\mathbf{X}) d_{\boldsymbol{\beta}}(\mathbf{X})] = \arg\max_{d_{\boldsymbol{\beta}} \in \mathcal{D}} E[m(\boldsymbol{\beta}'\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\alpha}) d_{\boldsymbol{\beta}}(\mathbf{X})].$$

Doubly Robust Estimators

- **DR supervised estimator**

$$\widehat{\boldsymbol{\beta}}_{sup}^{DR} = \arg\max_{\boldsymbol{\beta}} \left\{ \frac{1}{n} \sum_{i=1}^{n} V(\mathbf{Z}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) I\left(\boldsymbol{\beta}' \mathbf{X}_i \geqslant 0\right) \right\}.$$

- **DR semi-supervised estimator**

$$\widehat{\boldsymbol{\beta}}_{\lambda}^{DR} = \arg\max_{\boldsymbol{\beta}} \left\{ \frac{\lambda}{n} \sum_{i=1}^{n} V_i(\mathbf{Z}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) I(\boldsymbol{\beta}' \mathbf{X}_i \geqslant 0) + \frac{1-\lambda}{N} \sum_{i=n+1}^{n+N} \widehat{m}(\boldsymbol{\beta}' \mathbf{X}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) I(\boldsymbol{\beta}' \mathbf{X}_i \geqslant 0) \right\}.$$

- **DR pooled estimator**

$$\widehat{\boldsymbol{\beta}}_{pl}^{DR} = \arg\max_{\boldsymbol{\beta}} \left\{ \frac{1}{n+N} \sum_{i=1}^{n+N} \widehat{m}(\boldsymbol{\beta}' \mathbf{X}_i, \widehat{\theta}, \widehat{\boldsymbol{\alpha}}) I\left(\boldsymbol{\beta}' \mathbf{X}_i \geqslant 0\right) \right\}.$$

Outline

Asymptotic properties

Theorem 1 (Consistency)

Under some regularity conditions, when the propensity score is known, as $n, N \to \infty$, $\frac{n}{N} = \rho_n \to \rho \in [0, \infty)$, $\lambda \in [0, 1]$, we have:

a1. $\widehat{\boldsymbol{\beta}}_{sup} \xrightarrow{p} \boldsymbol{\beta}_0$;

a2. $\widehat{\boldsymbol{\beta}}_{pl} \xrightarrow{p} \boldsymbol{\beta}_0$;

a3. $\widehat{\boldsymbol{\beta}}_{\lambda} \xrightarrow{p} \boldsymbol{\beta}_0$;

Note that double robust estimators $\widehat{\boldsymbol{\beta}}_{sup}^{DR}$, $\widehat{\boldsymbol{\beta}}_{\lambda}^{DR}$ and $\widehat{\boldsymbol{\beta}}_{pl}^{DR}$ also have similar consistency.

Asymptotic properties

Theorem 2 (Asymptotic distributions)

Under some regularity conditions, when the propensity score is known, as $n, N \to \infty$, $\frac{n}{N} = \rho_n \to \rho \in [0, \infty)$, $\lambda \in [0, 1]$, we have:

b1. $n^{\frac{1}{3}}(\widehat{\boldsymbol{\beta}}_{sup} - \boldsymbol{\beta}_0) \xrightarrow{d} \arg\max_t Z(t)$, where $Z(t) = G(t) - \frac{1}{2}t'Vt$ and $-V$ is the second derivative matrix of $E[V(\boldsymbol{\theta}_0)I(\boldsymbol{\beta}'\mathbf{X} \geqslant 0)]$ w.r.t. $\boldsymbol{\beta}$ at $\boldsymbol{\beta}_0$.

b2. $n^{\frac{1}{3}}(\widehat{\boldsymbol{\beta}}_{\lambda} - \boldsymbol{\beta}_0) \xrightarrow{d} \arg\max_t Z_{\lambda}(t)$, where $Z_{\lambda}(t) = G_{\lambda}(t) - \frac{1}{2}t'Vt$.

b3. $n^{\frac{1}{3}}(\widehat{\boldsymbol{\beta}}_{pl} - \boldsymbol{\beta}_0) \xrightarrow{d} \arg\max_t Z_{pl}(t)$, where $Z_{pl}(t) = G_{pl}(t) - \frac{1}{2}t'Vt$.

Here $G(t)$, $G_{\lambda}(t)$ and $G_{pl}(t)$ are all mean-zero Gaussian processes with continuous sample paths and covariance kernel function $Cov(\cdot, \cdot)$, $[\lambda^2 + (1-\lambda)^2\rho^2]Cov(\cdot, \cdot)$ and $(\frac{\rho}{1+\rho})^2 Cov(\cdot, \cdot)$ respectively.

Similar properties hold for double robust estimators.

Asymptotic properties

Denote the corvariance of $\widehat{\boldsymbol{\beta}}_{bm}$, $\widehat{\boldsymbol{\beta}}_{\lambda}$ and $\widehat{\boldsymbol{\beta}}_{pl}$ are Σ_{bm}, Σ_{λ} and Σ_{pl} respectively. We can derive from theorem 2 that Σ_{λ} is minimized when the weight $\lambda = \frac{\rho^2}{1+\rho^2}$.
With this choice, a comparison of Σ_{sup}, Σ_{λ}, and Σ_{pl} reveals that

$$\Sigma_{\text{sup}} \geqslant \Sigma_{\lambda} \geqslant \Sigma_{\text{pl}},$$

since $1 \geqslant \frac{\rho^2}{1+\rho^2} \geqslant \frac{\rho^2}{(1+\rho)^2}$ holds for all $\rho \in [0, \infty)$.

Variance estimation

Propose a resampling method by perturbing the value function repeatedly to estimate the variance. Take the procedure for $\widehat{\beta}_\lambda$ as an example.

1. Generate iid perturbation ξ_i from $Beta(\sqrt{2}-1, 1)$ for $i = 1, \ldots, n+N$.
2. Perturb the value function. Let $\widehat{\boldsymbol{\theta}}^b = \arg\min_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^n \xi_i(1-A_i)[Y_i - \nu(\mathbf{X}_i, \boldsymbol{\theta})]^2$ and $\widehat{m}^b(\boldsymbol{\beta}'\mathbf{X}_j, \boldsymbol{\theta}) = \frac{\sum_{i=1}^n \xi_i K_h(\boldsymbol{\beta}'\mathbf{X}_i - \boldsymbol{\beta}'\mathbf{X}_j) V(\mathbf{Z}_i, \boldsymbol{\theta})}{\sum_{i=1}^n \xi_i K_h(\boldsymbol{\beta}'\mathbf{X}_i - \boldsymbol{\beta}'\mathbf{X}_j)}$, then for linear decision $d_{\boldsymbol{\beta}}(\mathbf{X}) = I(\boldsymbol{\beta}'\mathbf{X} \geqslant 0)$, we perturb the value function by

$$\widehat{\Delta}_\lambda^b\left(\boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}^b\right) = \frac{\lambda}{n}\sum_{i=1}^n \xi_i V\left(\mathbf{Z}_i, \widehat{\boldsymbol{\theta}}^b\right) d_{\boldsymbol{\beta}}(\mathbf{X}_i) + \frac{1-\lambda}{N}\sum_{j=n+1}^{n+N} \xi_j \widehat{m}^b\left(\boldsymbol{\beta}'\mathbf{X}_j, \widehat{\boldsymbol{\theta}}^b\right) d_{\boldsymbol{\beta}}(\mathbf{X}_j).$$

Variance estimation

3. Re-estimate β. We use the iterative algorithm to obtain the new estimator that

$$\widehat{\boldsymbol{\beta}}_{\lambda}^{b}=\underset{\boldsymbol{\beta}\in\mathcal{B}}{\arg\max}\,\widehat{\Delta}_{\lambda}^{b}\left(\boldsymbol{\beta},\widehat{\boldsymbol{\theta}}^{b}\right).$$

4. Estimate the variance. Repeat the above steps for B times and compute the empirical variance matrix $\widehat{\Sigma}_{\lambda}$ of $\{\widehat{\boldsymbol{\beta}}_{\lambda}^{b}, b=1,\ldots,B\}$ to estimate the population variance Σ_{λ}.

The above variance estimation procedure ensures that $n^{\frac{1}{3}}\left(\widehat{\boldsymbol{\beta}}_{\lambda}-\boldsymbol{\beta}_{0}\right)$ and $n^{\frac{1}{3}}\left(\widehat{\boldsymbol{\beta}}_{\lambda}^{b}-\widehat{\boldsymbol{\beta}}_{\lambda}\right)$ have the same asymptotic distribution, so we denote the empirical variance of $\left\{\widehat{\boldsymbol{\beta}}_{\lambda}^{b}:b=1,\ldots,B\right\}$ as an estimator of the population asymptotic variance.

Outline

Simulation setting

Consider a class of monotonic index models with four different cases: $Y = \nu(\mathbf{X}) + AD(\mathbf{X}) + \epsilon$, where $\mathbf{X} = (X_1, X_2, \ldots, X_6)'$ with $(X_1, X_2, X_3, X_4) \sim N_4(0, I_4)$, $X_5 \sim Bernoulli(0.5)$, $X_6 \sim Uniform(0, 1)$, $A \sim Bernoulli\{\pi(\mathbf{X})\}$ and $\epsilon \sim N\left(0, 0.5^2\right)$.

(a) case 1, $\nu(\mathbf{X}) = 1 + \boldsymbol{\gamma}_1'\mathbf{X}$ and $D(\mathbf{X}) = 2\boldsymbol{\beta}_0'\mathbf{X}$;

(b) case 2, $\nu(\mathbf{X}) = 1 + \boldsymbol{\gamma}_1'\mathbf{X}$ and $D(\mathbf{X}) = \exp\left(0.5\boldsymbol{\beta}_0'\mathbf{X}\right) - 1$;

(c) case 3, $\nu(\mathbf{X}) = 1 + \sin\left(\boldsymbol{\gamma}_1'\mathbf{X}\right) + 0.5\left(\boldsymbol{\gamma}_2'\mathbf{X}\right)^2$ and $D(\mathbf{X}) = 10\boldsymbol{\beta}_0'\mathbf{X}$;

(d) case 4, $\nu(\mathbf{X}) = 1 + X_1X_2 + 0.5X_3^2$ and $D(\mathbf{X}) = 10\boldsymbol{\beta}_0'\mathbf{X}$.

(c) case 5, $\nu(\mathbf{X}) = 1 + \sin\left(\boldsymbol{\gamma}_1'\mathbf{X}\right) + 0.5\left(\boldsymbol{\gamma}_2'\mathbf{X}\right)^2$ and $D(\mathbf{X}) = 2\left(\boldsymbol{\beta}_0'\mathbf{X}\right)^3$;

(d) case 6, $\nu(\mathbf{X}) = 1 + X_1X_2 + 0.5X_3^2$ and $D(\mathbf{X}) = 2\left(\boldsymbol{\beta}_0'\mathbf{X}\right)^3$.

We set $\boldsymbol{\beta}_0 = (1, -1, 2, 1, 2, 1)'$, $\boldsymbol{\gamma}_1 = (1, -1, 1, 1, -1, 1)'$ and $\boldsymbol{\gamma}_2 = (1, 0, -1, 0, 1, -1)'$.

The optimal treatment regime is given by $d^{\text{opt}}(\mathbf{X}) = I\left(\boldsymbol{\beta}_0'\mathbf{X} \geqslant 0\right)$.

Simulation setting

- We employ the Gaussian kernel with a bandwidth of $h_n = 0.5n^{-1/3}$ across all the cases.
- We conduct 1000 simulation runs for each case, with each run consisting of 200 resamplings for variance estimation.
- The optimization in the proposed methods is done by the optim function in R with the default method 'Nelder-Mead' for searching the maximizer.
- When the propensity score is known, we set $\pi(\mathbf{X}) = 0.5$. When the propensity score is unknown, we utilize the logistic regression model to estimate it. The true model is $\pi(\mathbf{X}) = s(-0.5 + X_1^2 + X_2^2)$ with $s(x) = \frac{1}{1+e^{-x}}$ under the misspecified setting.

Table: Results under case 1 with known propensity score

Method	N	Statistics	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$
sup		Bias	0.000	0.013	- 0.009	- 0.001	- 0.006	- 0.004
		SE	0.051	0.049	0.047	0.049	0.061	0.115
		SD	0.048	0.046	0.047	0.049	0.060	0.111
		CP(%)	97.5	96.4	97.1	97.1	96.8	92.7
SS	200	Bias	0.005	0.021	- 0.005	0.005	- 0.002	0.007
		SE	0.040	0.034	0.034	0.040	0.036	0.037
		SD	0.035	0.032	0.032	0.035	0.034	0.037
		CP(%)	97.9	94.3	97.4	97.8	98.8	99.6
		Eff	**1.644**	**1.808**	**1.886**	**1.512**	**2.896**	**9.744**
	500	Bias	0.004	0.016	- 0.006	0.006	- 0.006	0.010
		SE	0.028	0.024	0.029	0.028	0.033	0.037
		SD	0.029	0.026	0.029	0.029	0.032	0.037
		CP(%)	97.5	95.8	96.7	98.2	99.0	97.7
		Eff	**3.237**	**3.498**	**2.618**	**3.086**	**3.440**	**9.429**
pl	200	Bias	0.004	0.017	- 0.005	0.005	- 0.006	0.013
		SE	0.020	0.019	0.023	0.021	0.024	0.037
		SD	0.024	0.023	0.027	0.024	0.028	0.039
		CP(%)	98.9	95.0	96.8	99.2	99.6	97.6
		Eff	**6.252**	**5.013**	**4.226**	**5.477**	**6.435**	**9.192**
	500	Bias	0.003	0.014	- 0.004	0.005	- 0.007	0.012
		SE	0.015	0.013	0.019	0.015	0.018	0.033
		SD	0.019	0.018	0.023	0.019	0.023	0.036
		CP(%)	99.2	94.8	95.5	99.2	99.5	95.4
		Eff	**11.543**	**8.900**	**6.244**	**9.878**	**10.831**	**11.369**

AIDS Clinical Trials Group Protocol 175 (ACTG 175)

- The ACTG 175 contains 2139 observed patients.
- Our study focuses on patients randomly assigned to two treatments: zidovudine (coded A=0) and other therapies (coded A=1).
- The response variable Y is the CD4 T cell count at 96 ± 5 weeks, which value is missing for 797 patients.
- 11 covariates are considered including four continuous variables, age, weight, CD4 T cell count at baseline and CD8 T cell count at baseline, and seven binary variables, haemophilia, homosexual activity, history of intravenous drug use, race, gender, antiretroviral history, and symptomatic status. We standardized the continuous variables before estimation.

Real Data Analysis: ACTG 175

Table: Estimated parameters of optimal ITR for ACTG 175 study

Mehods	sup		SS		pl	
Predictors	Est	SD	Est	SD	Est	SD
intercept	0.024	0.032	0.026	0.049	0.138	0.104
hemo	-0.839	0.359	-0.821	0.069	-0.809	0.077
homo	-0.211	0.284	-0.237	0.074	-0.207	0.037
drugs	-0.171	0.290	-0.096	0.103	-0.173	0.037
race	0.270	0.217	0.327	0.077	0.285	0.076
gender	-0.024	0.322	-0.022	0.084	-0.010	0.056
str2	0.189	0.184	0.208	0.059	0.213	0.033
symptom	-0.190	0.206	0.087	0.097	-0.220	0.031
age	0.098	0.115	0.110	0.051	0.092	0.041
weight	0.060	0.107	0.068	0.046	0.048	0.044
cd40	-0.219	0.132	-0.251	0.060	-0.229	0.020
cd80	0.125	0.107	0.145	0.053	0.124	0.030

CI of estimated paramters for ACTG 175 study

Method	sup		SS		pl	
CI	95% CI	90% CI	95% CI	90% CI	95% CI	90% CI
intercept	(-0.038, 0.097)	(-0.025, 0.075)	(-0.057, 0.158)	(-0.031, 0.135)	(-0.177, 0.204)	(-0.127, 0.197)
hemo	(-0.949, 0.323)	(-0.925, 0.101)	**(-0.854, -0.590)**	**(-0.839, -0.631)**	**(-0.859, -0.611)**	**(-0.852, -0.650)**
homo	(-0.658, 0.422)	(-0.586, 0.333)	**(-0.338, -0.049)**	**(-0.324, -0.082)**	**(-0.271, -0.126)**	**(-0.257, -0.133)**
drugs	(-0.647, 0.449)	(-0.605, 0.348)	(-0.290, 0.103)	(-0.283, 0.058)	**(-0.219, -0.082)**	**(-0.218, -0.099)**
race	(-0.217, 0.626)	(-0.146, 0.564)	**(0.194, 0.508)**	**(0.218, 0.463)**	**(0.240, 0.495)**	**(0.245, 0.476)**
gender	(-0.650, 0.505)	(-0.589, 0.474)	(-0.284, 0.082)	(-0.236, 0.049)	(-0.146, 0.065)	(-0.119, 0.058)
str2	(-0.175, 0.528)	(-0.126, 0.483)	**(0.141, 0.378)**	**(0.160, 0.354)**	**(0.176, 0.304)**	**(0.184, 0.281)**
symptom	(-0.529, 0.284)	(-0.450, 0.217)	(-0.368, 0.022)	**(-0.330, -0.032)**	**(-0.249, -0.122)**	**(-0.242, -0.141)**
age	(-0.147, 0.321)	(-0.117, 0.280)	**(0.055, 0.263)**	**(0.087, 0.253)**	**(0.072, 0.209)**	**(0.074, 0.195)**
weight	(-0.164, 0.270)	(-0.122, 0.208)	**(0.011, 0.208)**	**(0.032, 0.187)**	**(0.027, 0.175)**	**(0.028, 0.150)**
cd40	(-0.420, 0.082)	(-0.367, 0.066)	**(-0.383, -0.146)**	**(-0.364, -0.158)**	**(-0.284, -0.193)**	**(-0.273, -0.204)**
cd80	(-0.135, 0.313)	(-0.084, 0.251)	**(0.072, 0.290)**	**(0.102, 0.276)**	**(0.089, 0.206)**	**(0.096, 0.193)**

Real Data Analysis: ACTG 175

Table: Treatment recommendation for ACTG 175 study

ITR	sup	SS	pl
ZDV+ddI	457	545	621
ZDV+ddC	589	501	425

- The SS and pl estimators suggest the ZDV+ddI regime for a larger number of patients.
- In medical research, treatment with ZDV+ddI has demonstrated a more pronounced efficacy in improving patient outcomes and slowing the progression of disease in individuals with HIV/AIDS compared to the treatment with ZDV+ddC (Darbyshire et al., 1996; Hammer et al., 1996; Mauss et al., 1996)

Outline

Future work

- Consider the dimension of covariates p increasing alongside the sample size n.
- Consider the cases involving censoring.
- Consider the heterogeneity between the $\mathcal{L}$ and $\mathcal{U}$.

Reference I

Chu, J., Lu, W., and Yang, S. (2023). Targeted optimal treatment regime learning using summary statistics. *Biometrika*, 110(4):913–931.

Darbyshire, J., Foulkes, M., Peto, R., Duncan, W., Babiker, A., Collins, R., Hughes, M., Peto, T. E., Walker, S. A., and Group, C. H. (1996). Zidovudine (azt) versus azt plus didanosine (ddi) versus azt plus zalcitabine (ddc) in hiv infected adults. *Cochrane database of systematic reviews*, 2010(3).

Fan, C., Lu, W., Song, R., and Zhou, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(5):1565–1582.

Gunn, K., Lu, W., and Song, R. (2024). Adaptive semi-supervised inference for optimal treatment decisions with electronic medical record data. In *Statistics in Precision Health: Theory, Methods and Applications*, pages 229–246. Springer.

Reference II

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., et al. (1996). A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, 335(15):1081–1090.

Li, C., Zeng, D., and Zhu, W. (2025). A robust covariate-balancing method for learning optimal individualized treatment regimes. *Biometrika*, 112(1):asae036.

Liao, K. P., Cai, T., Gainer, V., Goryachev, S., Zeng-treitler, Q., Raychaudhuri, S., Szolovits, P., Churchill, S., Murphy, S., Kohane, I., et al. (2010). Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis care & research*, 62(8):1120–1127.

Mauss, S., Adams, O., Willers, R., and Jablonowski, H. (1996). Combination therapy with zdv+ ddi versus zdv+ ddc in patients with progression of hiv-infection under treatment with zdv. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 11(5):469–477.

Reference III

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(2):331–355.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180–1210.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In Lin, D. Y. and Heagerty, P. J., editors, *Proceedings of the Second Seattle Symposium in Biostatistics*, volume 179 of *Lecture Notes in Statistics*, pages 189–326, New York. Springer.

Watkins, C. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3):279–292.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018.

Thank you for listening.

More details can be referred to “Li, X., Peng, M., and Zhou, Y. (2025). Doubly Robust Estimation of Individuall Tratment Regime in a Semi-supervised Framework. Statistica Sinca (Accepted). doi:10.5705/ss.202025.0168"