Doubly-Robust Estimation of Functional Outcomes

Dominic DiSanto

Junwei Lu Reading Group - Spring 2025

March 6, 2025

Outline

Papers of Focus

- Primarily
  - Doubly-Robust Functional Average Treatment Effect Estimation
    - Lorenzo Testa, Tobia Boschi, Francesca Chiaromonte, Edward H. Kennedy, Matthew Reimherr
- Secondarily
  - Double robust estimation of functional outcomes with data missing at random
    - Xijia Liu, Kreske Ecker, Lina Schelin, Xavier de Luna
- Other related results:
  - Causal Inference on Distribution Functions
    - Zhenhua Lin, Dehan Kong, Linbo Wang
  - One-Step Estimation of Differentiable Hilbert-Valued Parameters
    - Alex Luedtke, Incheoul Chung

Typical Causal Set-Up

- Consider observing iid data $(Y_i, A_i, \mathbf{X}_i)_{i\in[n]}$
  - $Y_i \in \mathbb{R}, A_i \in \{0,1\}, \mathbf{X}_i \in \mathbb{R}^d$
  - Interest lies in contrasts of potential outcomes $Y(1), Y(0)$
- Make "causal" assumptions to translate functions of unobservable potential outcomes to observable data:
  - Unconfoundedness, SUTVA, Positivity
- Construct estimands of interest
  - Most commonly ATE: $\tau = \mathbb{E}[Y(1) - Y(0)]$ or related variant (CATE, ATT)
- AIPW estimator
  $\hat{\tau}_{\text{AIPW}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_i^{(1)} - \hat{\mu}_i^{(0)} + \frac{A_i(Y_i - \hat{\mu}_i)}{\hat{\pi}(X_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{(1-A_i)(Y_i - \hat{\mu}_i)}{1-\hat{\pi}(X_i)}$

Functional Outcomes

- In most settings, Y_i is some scalar (binary/continuous outcome, count of recurrent events, time-to-event)
- Recent results for functional outcomes, say $\mathcal{Y}_i$
  - $\mathcal{Y} \in C(\mathcal{T})$ for some domain, e.g. $\mathcal{T} = [0, T]$
  - e.g. disease trajectories, recurrent temporal data, etc.
- Now define relevant estimands that are themselves random functions:
  - $\text{ATE}_f := \mathbb{E}[\mathcal{Y}(1) - \mathcal{Y}(0)]$, the mean function of a stochastic process
  - Where as before, our estimand was $\text{ATE} : \mathscr{P} \mapsto \mathbb{R}$, we now have $\text{ATE}_f : \mathscr{P} \mapsto C(\mathcal{T})$

Functional Outcomes

- Fundamentally, we still want to construct an estimator that is
  - Doubly-robust (model and ideally rate)
  - Has asymptotic properties that permit inference
- For random functions, we can asymptotcally approximate as Gaussian Processes
- Then conduct inference using existing FDA methods for simultaneous confidence bands

Goals/Strategies

- Review extensions of the "typical" causal/doubly-robust estimation procedure to functional outcomes
  - e.g. disease trajectories, recurrent temporal data, etc.
- Observe a seemingly common tool-kit for constructing an estimator, performing inference
  - Build a "classical" doubly-robust estimator
  - (1) Finite-dimensional asymptotic normality and (2) tightness of measure
  - Approximate as a Gaussian process, construct confidence bands

Outline

Set-Up

- Observe data $\mathcal{D} = (\mathcal{D}_i)_{i\in[n]} = (\mathcal{Y}_i, \mathbf{X}_i, A_i)_{i\in[n]}$
- $\text{ATE}_f := \mathbb{E}[\mathcal{Y}(1) - \mathcal{Y}(0)] = \mathbb{E}[\gamma^{(1)}(\mathcal{D}) - \gamma^{0}(\mathcal{D})]$
- For

$$\gamma^{(a)}(\mathcal{D}) = \mu^{(a)}(\mathbf{X}) + \frac{\mathbb{1}\{A = a\}\left(\mathscr{Y}^{(a)} - \mu^{(a)}(\mathbf{X})\right)}{\pi^{(a)}(\mathbf{X})}$$

where

$$\mu^{(a)} := \mathbb{E}[\mathcal{Y}^{(a)} \mid \mathbf{X}, A = a]$$
$$\pi^{(a)} := P(A = a \mid \mathbf{X})$$

The $\hat{\beta}_{\mathrm{DR-FoS}}$ Estimator

$$\mathbb{E}\left[\gamma^{(1)}(\mathcal{D})\right] = \mathbb{E}\left[\mu^{(1)}(X) + \frac{A\left(\mathscr{Y} - \mu^{(1)}(X)\right)}{\pi^{(1)}(X)}\right]$$

$$= \underbrace{\mathbb{E}[\mathbb{E}[\mathcal{Y} \mid X, A=1]]}_{=\mathbb{E}\left[\mathscr{Y}^{(1)}\right]} + \underbrace{\mathbb{E}\left[\mathbb{E}\left[\left.\frac{A\left(\mathscr{Y} - \mu^{(1)}(X)\right)}{\pi^{(1)}(X)}\right| X\right]\right]}_{=0}$$

We can then estimate ATE_f as (with some background cross-fitting of nuisance functions)

$$\hat{\beta}_{\mathrm{DR-FoS}} := \frac{1}{n}\sum_{i=1}^{n} \hat{\gamma}^{(1)}(\mathcal{D}) - \hat{\gamma}^{(0)}(\mathcal{D})$$

(Here we use the authors' notation, where $\beta \equiv \mathrm{ATE}$, and "DR-FOS' stands for "Doubly-Robust Function-on-Scalar')

Robustness of $\hat{\beta}_{\mathrm{DR-FoS}}$

- Model and rate-robustness come from similar analysis as in the scalar setting
  - Analyze the von mises expansion as three terms (1) CLT term, (2) Empirical Process, and (3) Remainder$=o_{\mathbb{P}}(1)$
- Note on rate-robustness:
  - The rate-robustness proof is shown *for finite-dimensional projections of* $\hat{\beta}_{\mathrm{DR-FoS}}$

Outline

Main Result

Theorem

Theorem 3.9 in Testa 2025 Under the causal identification and inference assumptions (next slide),

$$\sqrt{n}(\hat{\beta}_{\mathrm{DR-FoS}} - \beta) \rightsquigarrow \mathrm{GP}(0, \Sigma)$$

for $\Sigma(s,t) = \mathbb{E}[\varphi(\mathcal{D},s)\varphi(\mathcal{D};t)]$ *where* $\varphi(\mathcal{D}) = \gamma^{(1)}(\mathcal{D}) - \gamma^{(0)}(\mathcal{D}) - \beta$ *is the influence function, and* $\varphi(\mathcal{D};s)$ *the influence function defined at time* s

- Show asymptotic normality of finite-dimensional projection of $\hat{\beta}_{\mathrm{DR-FoS}}$
- Show tightness of measure

Assumptions

Assumption 3.4 (Inference). Let the number of cross-fitting folds be fixed at J, and assume that:

a. For each $j \in \{1,\ldots,J\}$, and for every one-dimensional projection $\varphi(\mathscr{D};t)$, one has

$$\hat{\varphi}^{[-j]}(\mathscr{D};t) \overset{\mathbb{L}^2}{\to} \varphi(\mathscr{D};t)\,.$$

b. For every one-dimensional projection $\varphi(\mathscr{D};t)$, one has

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{J} R_2^{[j]} = o_{\mathbb{P}}(1),$$

where $R_2^{[j]} = \hat{\beta}^{[-j]}_{\texttt{DR-FoS}}(t) - \beta(t) + \int \hat{\varphi}^{[-j]}(\mathscr{D};t)\, d\mathbb{P}$.

c. Given $\xi > 0$, $\hat{\pi}^{(a)}$ is bounded away from ξ and $1-\xi$ with probability 1.

d. For any $\delta > 0$, the functional outcome satisfies

$$\mathbb{E}\left[\sup_{|s-t|\le\delta} |\mathscr{Y}(s) - \mathscr{Y}(t)|\right] \le L\delta \qquad (9)$$

for some constant L.

e. For any $\delta > 0$ and for $a \in \{0,1\}$, the *estimated* regression function satisfies

$$\mathbb{E}\left[\sup_{|s-t|\le\delta} \left|\hat{\mu}^{(a)}(s) - \hat{\mu}^{(a)}(t)\right|\right] \le L^{(a)}\delta$$

for some constant $L^{(a)}$.

Finite-Dimensional Aymptotic Normality

- Take the von Mises expansion and study the (1) CLT, (2) Empirical Process, and (3) remainder terms

Lemma 3.7 (Asymptotic Normality of finite dimensional projections). *Let $k \in \mathbb{N}$ and $t_1, \ldots, t_k \in \mathcal{T}$ be fixed. Under Assumptions 2.1 (identifiability) and 3.4 (inference), one has*

$$\sqrt{n}\left(\left(\hat{\beta}_{\text{DR-FoS}}(t_1), \ldots, \hat{\beta}_{\text{DR-FoS}}(t_k)\right)^T - \left(\beta(t_1), \ldots, \beta(t_k)\right)^T\right) \rightsquigarrow \mathcal{N}\left(0, \Sigma_{t_1,\ldots,t_k}\right), \tag{14}$$

where $v_{t_1,\ldots,t_k}(\mathcal{D}) = \left(\varphi(\mathcal{D}; t_1), \ldots, \varphi(\mathcal{D}; t_k)\right)^T$ *and* $\Sigma_{t_1,\ldots,t_k} = \mathbb{E}\left[v_{t_1,\ldots,t_k}(\mathcal{D})\, v_{t_1,\ldots,t_k}(\mathcal{D})^T\right]$.

Proof Sketch

Take the von mises expansion,

$$\sqrt{n}\left(\hat{B}_{t_1,\ldots,t_k} - B_{t_1,\ldots,t_k}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \nu_{t_1,\ldots,t_k}(\mathscr{D}_i) + \sqrt{n}(\mathbb{P}_n - \mathbb{P})\left(\hat{\nu}_{t_1,\ldots,t_k}(\mathscr{D}) - \nu_{t_1,\ldots,t_k}(\mathscr{D})\right) + \sqrt{n}R_2,$$

where $B_{t_1,\ldots,t_k} := [\hat{\beta}_{\mathrm{DR-FoS}}(t_1, \ldots, \hat{\beta}_{\mathrm{DR-FoS}}(t_k))]^T$

- $\sqrt{n}\sum_i \nu_{t_1,\cdots,t_k}(\mathcal{D}_i)$ is mean 0, finite variance (CLT term)
- $\sqrt{n}(\mathbb{P}_n - \mathbb{P})\left(\hat{\nu}_{t_1,\cdots,t_k}(\mathcal{D}_i) - \nu_{t_1,\cdots,t_k}(\mathcal{D})\right)$ is the Empirical Process term (controlled via cross-fitting to be $o_{\mathbb{P}}(1)$)
- The remainder term is $o_{\mathbb{P}}(1)$ by assumption

Proof Sketch

The remainder term R_2 above is the source of the so-called "strong" or "rate" double-robustness.

$$
\begin{aligned}
R_2^q &= \int \left| \frac{1}{\pi^{(1)}(\mathcal{D})} - \frac{1}{\hat{\pi}^{(1)}(\mathcal{D})} \right| \left| \mu^{(1)}(\mathcal{D}) - \hat{\mu}^{(1)}(\mathcal{D}) \right| \pi^{(1)} d\mathbb{P} \\
&\quad - \int \left| \frac{1}{1 - \pi^{(1)}(\mathcal{D})} - \frac{1}{1 - \hat{\pi}^{(1)}(\mathcal{D})} \right| \left| \mu^{(0)}(\mathcal{D}) - \hat{\mu}^{(0)}(\mathcal{D}) \right| (1 - \pi^{(1)}) d\mathbb{P} \\
&\leq \frac{1}{\varepsilon} \int |\pi^{(1)}(\mathcal{D}) - \hat{\pi}^{(1)}(\mathcal{D})||\mu^{(1)}(\mathcal{D}) - \hat{\mu}^{(1)}(\mathcal{D})| d\mathbb{P} \\
&\quad - \frac{1}{\varepsilon} \int |\pi^{(1)}(\mathcal{D}) - \hat{\pi}^{(1)}(\mathcal{D})||\mu^{(0)}(\mathcal{D}) - \hat{\mu}^{(0)}(\mathcal{D})| d\mathbb{P}
\end{aligned}
$$

with ε coming from the positivity assumption (and all $\mu, \hat{\mu}$ implicitly indexed at a single time point $t_q, q \in [1, \dots, k]$).

Extension to Infinite Dimensional Result

- Remains only to show "tightness" of measure
  - $\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}[w(\hat{\beta}_{\mathrm{DR-FoS}} \geq \varepsilon; \delta)] = 0$
  - Here w is the modulus of continuity $w(f; \delta) = \sup_{|t-s|} \leq \delta |f(t) - f(s)|$

$$
\begin{aligned}
w(\hat{\beta}_{\texttt{DR-FoS}}; \delta) &= \sup_{|s-t| \leq \delta} |\hat{\beta}_{\texttt{DR-FoS}}(s) - \hat{\beta}_{\texttt{DR-FoS}}(t)| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \left(1 + \frac{A_i}{\hat{\pi}^{(1)}(X_i)}\right) \sup_{|s-t| \leq \delta} \left|\hat{\mu}^{(1)}(X_i; s) - \hat{\mu}^{(1)}(X_i; t)\right| \\
&+ \frac{1}{n} \sum_{i=1}^{n} \left(1 + \frac{1 - A_i}{1 - \hat{\pi}^{(1)}(X_i)}\right) \sup_{|s-t| \leq \delta} \left|\hat{\mu}^{(0)}(X_i; s) - \hat{\mu}^{(0)}(X_i; t)\right| \\
&+ \frac{1}{n} \sum_{i=1}^{n} \frac{A_i - \hat{\pi}^{(1)}(X_i)}{\hat{\pi}^{(1)}(X_i)(1 - \hat{\pi}^{(1)}(X_i))} \sup_{|s-t| \leq \delta} |\mathcal{Y}_i(s) - \mathcal{Y}_i(t)| \\
&= M(\hat{\beta}_{\texttt{DR-FoS}}; \delta).
\end{aligned}
$$

Summary

- "Typical" causal set-up extends to function estimation, translated to function estimands
- Retain similar notions of double-robustness
    - Noting rate-double robustness is asymptotic in n for finite-dimensional object of $\hat{\mu}^{(a)}$
- Proof strategy is roughly similar to existing tool-kit for semi-parametric problems
    - Study influence functions/von Mises expansion of finite-dimensional object
    - Show tightness of measure under causal and functional assumptions

Outline

Set-Up and Data Model

- Observe $(\mathcal{Y}_i, \mathbf{X}_i, Z_i)_{i\in[n]}$
  - Z_i is an indicator of whether $\mathcal{Y}_i$ is observed
- Estimand $\mathbb{E}[\mathcal{Y}]$ when $\mathcal{Y}$ is only partially observed
  - Paper is not explicit about estimand but inferring from the structure of their estimator
- Assume a linear functional model $\mathcal{Y}_i = \mathbf{X}_i^T\beta + \varepsilon_i$ for deterministic functions $\beta = [\beta_1, \dots, \beta_p]^T$, $\beta_j \in L^2([0,1], \mathbb{R})$
- Assume a logistic model on Z, $\mathbb{P}(Z_i = 1) = \mathrm{expit}(\mathbf{X}_i^T\gamma)$
- Estimate $\hat{\gamma}$ as the MLE, $\hat{\beta}$ is the functional OLS estimator $\hat{\beta}(t) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathcal{Y}(t)$

Estimator & Convergence

Define the estimator:

$$\hat{\mu}_{DR} = \sum_i \mathbf{X}_i^T \hat{\beta} + \sum_i \frac{Z_i}{g(\mathbf{X}_i^T \hat{\gamma})}(\mathcal{Y}_i - \mathbf{X}_i^T \hat{\beta})$$

Theorem 2.2. *Assume that at least one of the working models* (1) *and* (2) *is correctly spec*
i.e., $E(\mathcal{Y}_i \mid \boldsymbol{x}_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$ *or* $\Pr[Z_i = 1|\boldsymbol{x}_i] = \tau(\boldsymbol{x}_i^\top \boldsymbol{\gamma})$. *Then, the functional DR estimator* $\sqrt{n}(\hat{\mu}_{DR}$ -
is asymptotically distributed as a Gaussian process with zero mean function. Further, if both wo
models are correctly specified, the covariance function simplifies and we have:

$$\sqrt{n}(\hat{\mu}_{DR} - \mu_y) \xrightarrow{d} \mathcal{GP}(0, \boldsymbol{\beta}(s)^\top \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta}(t) + \mathbb{E}[\tau^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\gamma})]\sigma_\epsilon(s,t)).$$

Proof Sketch

- Show asymptotic normality of a finite dimensional object
  - Here the proof is somewhat "brute-foced" by an M-estimation argument
- Show tightness of measure
  - Here, the results are quite technical/measure-theoretic
  - Loosely, uniform tightness is proved via Prohorov's theorem

Tightness Lemmas

Lemma B.1. *With some probability space* $(\Omega, \mathcal{F}, \mathcal{P})$, *let* $\{X_n\}_{n\geq 1}$ *be a sequence of independent random elements in Hilbert space* $(\mathcal{H}, \mathcal{B}_{\mathcal{H}})$ *with mean* 0 *and* $\mathbb{E}\left(\|X_i\|^2\right) < \infty$ *for all* $i = 1, \ldots, n$, *and* $\xi_n = n^{-1/2}\sum_{i=1}^n X_i$. *The sequence of probability measures implied by* ξ_n, $\left\{\mathcal{P} \circ \xi_n^{-1}\right\}_{n\geq 1}$, *is uniformly tight.*

Lemma B.2. *With some probability space* $(\Omega, \mathcal{F}, \mathcal{P})$, *let* $(X_n, Y_n)_{n\geq 1}$ *be a sequence of paired random elements taking values from* $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ *and* $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$ *respectively. Let* $\mathbb{P}_n$ *be the joint probability measures over* $\sigma(\mathcal{B}_{\mathcal{X}} \times \mathcal{B}_{\mathcal{Y}})$ *which is the smallest* σ*-algebra making projections to* $\mathcal{X}$ *and* $\mathcal{Y}$ *all measurable, i.e.* $\mathcal{P} \circ X_n^{-1}(E) = \mathbb{P}_n(E \times \mathcal{Y})\ \forall E \in \mathcal{B}_{\mathcal{X}}$. *Assume* $\forall y \in \mathcal{Y}$, $\exists$ σ*-algebra* $\mathcal{B}_{\mathcal{X}|y}$ *and* $\mathcal{P} \circ (X_n|y)^{-1}$ *such that the joint probability measure can be represented by disintegration, i.e.* $\mathbb{P}_n(E \times F) = \int_F \mathcal{P} \circ (X_n|y)^{-1}(E)d\mathcal{P} \circ Y_n^{-1}$. *If* $\mathcal{P} \circ (X_n|y)^{-1}$ *is uniformly tight, then the probability measure implied by* X_n *is uniformly tight. Further, if* $\mathcal{P} \circ Y_n^{-1}$ *is also uniformly tight, then the joint probability measure is uniformly tight.*

Notes on Liu(2024)

- Similar (global) proof structure (standard machinery to prove convergence to Gaussian processes)
- Explicit proofs of tightness (compared to Testa (2025)'s assumptions)

Goals/Strategies

- Review extensions of the "typical" causal/doubly-robust estimation procedure to functional outcomes
  - e.g. disease trajectories, recurrent temporal data, etc.
- Observe a seemingly common tool-kit for constructing an estimator, performing inference
  - Build a "classical" doubly-robust estimator
  - (1) Finite-dimensional asymptotic normality and (2) tightness of measure
  - Approximate as a Gaussian process, construct confidence bands