Multiply Robust Estimation of Treatment Effect for Time-to-event Outcome under Dependent Left Truncation

Yuyao Wang[1], Andrew Ying and Ronghui Xu[12]

[1]Department of Mathematics, [2]Herbert Wertheim School of Public Health and Halicioglu Data Science Institute, UC San Diego

UC San Diego Mathematics

Introduction

- In prospective cohort studies, only subjects with event times (T^*) greater than enrollment times (Q^*) are included. (e.g., pregnancy studies, aging studies, etc.)
- **Double biases**:
  1) Selection bias from left truncation:
     - Subjects with early event times tend not to be captured;
  2) Confounding bias from non-randomized treatment.
- Conventional methods leveraging covariates information such as IPW or regression-based methods can be used, but they are sensitive to model misspecification.
- Z^*: measured covariates; A^*: binary treatment.
- Observe (Q, T, A, Z) only if $Q^* < T^*$.
- Variables with '*' denote the variables in the data if there were no left truncation, and variables without '*' denote the ones in the observed data.

Review of Wang et al. (2022)

- Derived the efficient influence curve (EIC) for $\theta = \mathbb{E}\{\nu(T^*)\}$ and obtained an doubly robust estimating function from the EIC:
$$U(\theta; F, G) = \frac{\nu(T) - \theta}{G(T|Z)} - \int_0^\infty m_\nu(v, Z; \theta, F) \cdot \frac{F(v|Z)}{1 - F(v|Z)} \cdot \frac{d\bar{M}_Q(v; G)}{G(v|Z)},$$
where $F(t|z) = \mathbb{P}(T^* \le t|Z^* = z)$ and $G(t|z) = \mathbb{P}(Q^* \le t|Z^* = z)$, $m_\nu(v, z; \theta, F) = \mathbb{E}\{\nu(T^*) - \theta \mid T^* < v, Z^* = z\}$.
- **Double robustness**: $\mathbb{E}\{U(\theta_0; F, G)\} = 0$ if either $F = F_0$ or $G = G_0$, where F_0, G_0 denote the truth.
- **Model double robustness**: The estimator is consistent and asymptotically normal (CAN) if both $\hat{F}$ and $\hat{G}$ are asymptotically linear and one of them is consistent; it achieves the semiparametric efficiency bound if both $\hat{F}$ and $\hat{G}$ are consistent.
- **Rate double robustness**: The estimator is CAN and achieves the semiparametric efficiency bound if both $\hat{F}$ and $\hat{G}$ are consistent and the cross integral product of the two estimation error rates is faster than root-n.
- Provided technical conditions for the asymptotic properties that appear to not have been carefully examined in the literature for time-to-event data.

Extension of Wang et al. (2022)

For any unbiased estimating function $u^*(T^*, A^*, Z^*; \theta)$ for θ in the truncation-free data satisfying $\mathbb{E}\{u^*(T^*, A^*, Z^*; \theta)\} = 0$, consider the $\text{AIPW}_{(F,G)}$ operator for left truncation:
$$U(\theta; F, G) = V\{u^*(T^*, A^*, Z^*; \theta); F, G\}$$
$$= \frac{u^*(T, A, Z; \theta)}{G(T|A, Z)} - \int_0^\infty m(v, A, Z; \theta, u^*, F) \cdot \frac{F(v|A, Z)}{1 - F(v|A, Z)} \cdot \frac{d\bar{M}_Q(v; G)}{G(v|A, Z)}, \quad (1)$$
where $F(t|a, z) = \mathbb{P}(T^* \le t|A^* = a, Z^* = z)$, $G(t|a, z) = \mathbb{P}(Q^* \le t|A^* = a, Z^* = z)$, $m(v, a, z; \theta, u^*, F) = \mathbb{E}\{u^*(T^*, A^*, Z^*; \theta) \mid T^* < v, A^* = a, Z^* = z\}$.
- **Double robustness**: $\mathbb{E}[U(\theta_0; F, G)] = 0$ if either $F = F_0$ or $G = G_0$.

Assumptions

- Consistency: $T^* = A^*T^*(1) + (1 - A^*)T^*(0)$, $Q^* = A^*Q^*(1) + (1 - A^*)Q^*(0)$
- No unmeasured confounding: $A^* \perp\!\!\!\perp (T^*(a), Q^*(a)) \mid Z^*$.
- Strict positivity: $0 < \delta \le \mathbb{P}(A^* = 1|Z^*) \le 1 - \delta$.
- Conditional quasi-independence: $Q^*(a)$ and $T^*(a)$ are conditional independent given Z^* on the observed data region, i.e., on the region of $\{q < t\}$.
- Overlap assumption for F and G.

Doubly robust estimation for propensity score

Supposed we assume a parametric model for the propensity score $\pi(z; \gamma) = \mathbb{P}(A^* = 1|Z^* = z; \gamma)$, and denote the corresponding estimating function in the truncation-free data as $u_A^*(A^*, Z^*; \gamma)$.
- Apply (1) to $u_A^*(A^*, Z^*; \gamma)$ $\Rightarrow$ doubly robust estimating function for γ:
$$U_A(\gamma; F, G) = u_A^*(A, Z; \gamma) \cdot W(F, G),$$
where
$$W(F, G) = \frac{1}{G(T|A, Z)} - \int_0^\infty \frac{F(v|A, Z)}{1 - F(v|A, Z)} \cdot \frac{d\bar{M}_Q(v; G)}{G(v|A, Z)}.$$
- Construct the estimator $\hat{\gamma}$ by solving $\sum_{i=1}^n u_A^*(A_i, Z_i; \gamma) \cdot W_i(\hat{F}, \hat{G}) = 0$.
- Obtain estimators with model double robustness under asymptotic linearity and rate double robustness from cross-fitting.

Multiply robust estimation for treatment effect

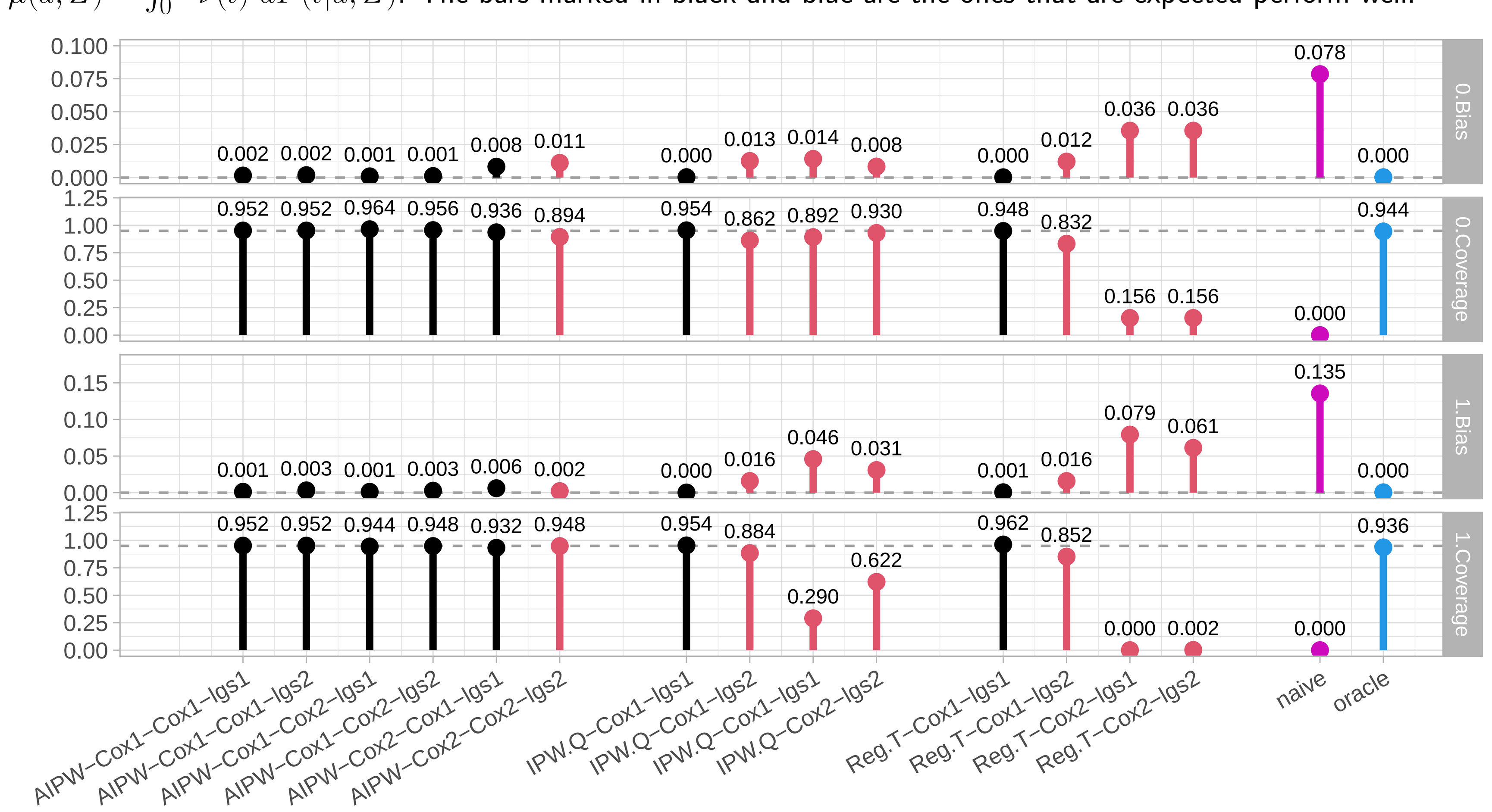
- Estimand: $\theta_a = \mathbb{E}[\nu\{T^*(a)\}]$, for $a = 0, 1$.
- Denote $\mu(a, z) = \mathbb{E}\{\nu(T^*)|A^* = a, Z^* = z\}$.
- Consider the $\text{AIPW}_{(\pi,\mu)}$ estimating functions for θ_a if there were no left truncation:
$$u_a^*(T^*, A^*, Z^*; \theta_a, \pi, \mu) = \frac{(A^*)^a(1 - A^*)^{1-a}\{\nu(T^*) - \theta\}}{\pi(Z^*)^a\{1 - \pi(Z^*)\}^{1-a}} + \frac{(-1)^a\{A^* - \pi(Z^*)\}\{\mu(a, Z^*) - \theta\}}{\pi(Z^*)^a\{1 - \pi(Z^*)\}^{1-a}}.$$
- Apply (1) to $u_a^*(T^*, A^*, Z^*; \theta_a, \pi, \mu)$:
$$U_a(\theta_a; F, G, \pi, \mu)$$
$$= \frac{A^a(1 - A)^{1-a}}{\pi(Z)^a\{1 - \pi(Z)\}^{1-a}} \left\{ \frac{\nu(T) - \theta}{G(T|A, Z)} - \int_0^\infty \frac{\int_0^v \{\nu(t) - \theta\} dF(t|A, Z)}{1 - F(v|A, Z)} \cdot \frac{d\bar{M}_Q(v; G)}{G(v|A, Z)} \right\}$$
$$+ \frac{(-1)^a\{A - \pi(Z)\}\{\mu(a, Z) - \theta\}}{\pi(Z)^a\{1 - \pi(Z)\}^{1-a}} \cdot W(F, G).$$
- Construct the estimator $\hat{\theta}_a$ by solving $\sum_{i=1}^n U_a(\theta_a; \hat{F}, \hat{G}, \hat{\pi}, \hat{\mu}) = 0$.

Multiple Robustness

$\mathbb{E}\{U_a(\theta_a; F, G, \pi, \mu)\} = 0$ if the following two conditions are true: (i) either $F = F_0$ or $G = G_0$; (ii) either $\pi = \pi_0$ or $\mu = \mu_0$.

F	G	π	μ
✓		✓	
✓			✓
	✓	✓	
	✓		✓

Simulation for $\mathbb{P}\{T^*(0) > 3\} = 0.6796$ (top 2 chunks) and $\mathbb{P}\{T^*(1) > 3\} = 0.5629$ (bottom 2 chunks) from 500 simulated data sets each with sample size 2000; the truncation rate is 22.8%; the treated rate is 50.0%. We take $\hat{\mu}(a, Z) = \int_0^\infty \nu(t)\, d\hat{F}(t|a, Z)$. The bars marked in black and blue are the ones that are expected perform well.



Doubly robust estimation under marginal Cox model

- Under randomization, consider the marginal Cox model:
$$\lambda^*(t|A^*) = \lambda_0(t)e^{\beta A^*}$$
- Denote $\Lambda(t) = \int_0^t \lambda_0(u)du$.
- Consider the estimating functions for (β, Λ) if there were no left truncation:
$$D_1^*(\beta, \Lambda, t) = dM^*(t; \beta, \Lambda), \ \forall t \ge 0; \qquad D_2^*(\beta, \Lambda) = \int_{\tau_1}^{\tau_2} A\, dM^*(t; \beta, \Lambda),$$
where $M^*(t; \beta, \Lambda) = N^*(t) - \int_0^t Y^*(u)e^{\beta A^*}d\Lambda(u)$, $N^*(t) = I(T^* \le t)$, $Y^*(t) = I(T^* \ge t)$.
- Apply (1) to D_1^* and D_2^*:
$$D_1(\beta, \Lambda, t; F, G) = V\{D_1^*(\beta, \Lambda, t); F, G\}, \quad \forall t \ge 0,$$
$$D_2(\beta, \Lambda; F, G) = V\{D_2^*(\beta, \Lambda); F, G\}.$$

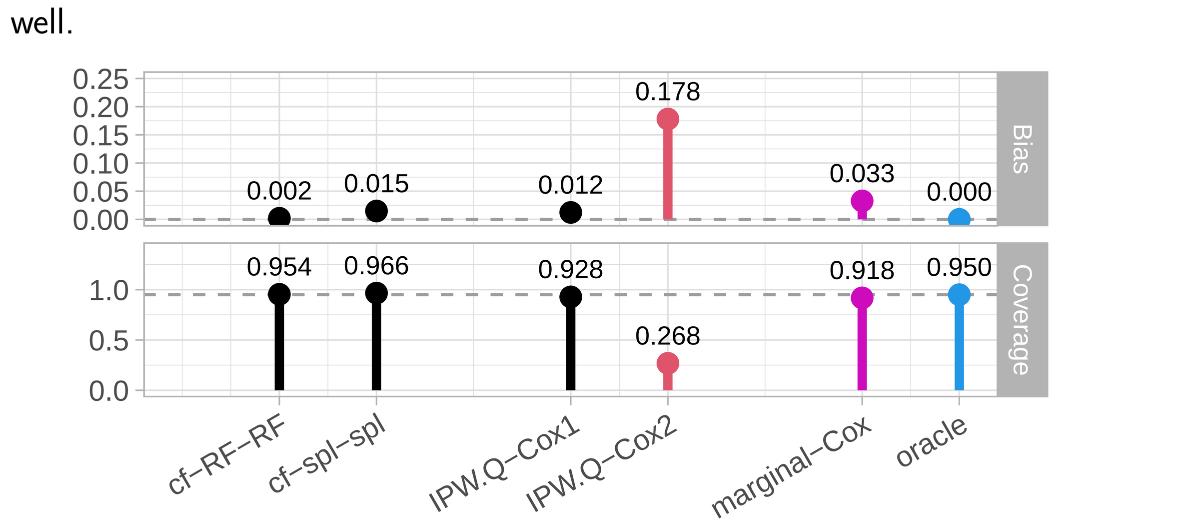
***K*-fold cross-fitting algorithm**:

1. Split the data into K folds of equal size with the index sets $\mathcal{I}_1, ..., \mathcal{I}_K$.
2. For each fold k:
   - Estimate the nuisance parameters F and G using the out-of-k-fold data. $\Rightarrow$ $\hat{F}_{-k}$, $\hat{G}_{-k}$.
   - Consider the following estimating equations for (β, Λ):
$$\frac{1}{|\mathcal{I}_k|}\sum_{i \in \mathcal{I}_k} D_{1i}(\beta, \Lambda, t; \hat{F}_{-k}, \hat{G}_{-k}) = 0, \quad (2)$$
$$\frac{1}{|\mathcal{I}_k|}\sum_{i \in \mathcal{I}_k} D_{2i}(\beta, \Lambda; \hat{F}_{-k}, \hat{G}_{-k}) = 0. \quad (3)$$
   - First solve for $\Lambda(t)$ from (2) and then plug the estimate into (3)
     $\Rightarrow$ An estimating function $U_k(\beta, \hat{F}_{-k}, \hat{G}_{-k})$ for β.
3. Obtain the estimator $\hat{\beta}_{cf}$ by solving $\sum_{k=1}^K U_k(\beta, \hat{F}_{-k}, \hat{G}_{-k}) = 0$.

Rate Double Robustness

The estimator $\hat{\beta}_{cf}$ is consistent and asymptotically normal if both $\hat{F}$ and $\hat{G}$ are consistent and the cross integral product of the two estimation error rates is faster than root-n.

Simulation for β from 500 simulated data sets each with sample size 1000; the truncation rate is 20.9%; the truth $\beta_0 = 0.3$. The ones marked in black and blue are the ones that are expected to perform well.



Discussion

- When the treatment is not randomized, multiply robust estimators for the hazard ratio of the marginal structural Cox model can be developed by applying the $\text{AIPW}_{(F,G)}$ in (1) to the AIPW estimating equations for (β, Λ) developed in Rava (2021).
- Right censoring can be handled using IPCW or AIPCW.

- References: Wang et al. (2022), arXiv:2208.06836.
  Rava (2021), PhD thesis, UCSD.

- Email: yuw079@ucsd.edu (Y. Wang); aying9339@gmail.com (A. Ying); rxu@ucsd.edu (R. Xu)