Multiply Robust Prediction Sets for Missing Data

Arun Kumar Kuchibhotla

Carnegie Mellon University

6 August, 2026

Collaborators



Eric Tchetgen Tchetgen



Manit Paul

Table of contents

1. Introduction to Missing Data
2. Introduction to (Conformal) Prediction
3. Prediction Sets for Missing Data
4. General Framework
5. Conclusions

Outline

Introduction to Missing Data

Introduction to (Conformal) Prediction

Prediction Sets for Missing Data

General Framework

Conclusions

Missing Data

- ⋆ Suppose we have an outcome Y of interest on a population and we are interested in estimating $\mathbb{E}[Y]$.
- ⋆ In an idealized sampling setting, we draw n units/subjects IID from the population and observe their outcomes as $Y_1, \ldots, Y_n$ and report the estimator

$$\widehat{\theta}_n := \frac{1}{n} \sum_{i=1}^{n} Y_i.$$

- ⋆ Of course, reality does not care, and often we end up with an observation sequence $Y_1^*, \ldots, Y_n^*$ where $Y_i^* = Y_i$ (observed) for some i and $Y_i^* = \star$ (unobserved) for the remaining i.
- ⋆ Estimation of $\mathbb{E}[Y]$ now becomes challenging, or even impossible without some assumptions on the missingness mechanism. For example, if all Y_i's above some threshold are left unobserved, then under no additional assumptions, $\mathbb{E}[Y]$ is not identifiable.

Missing Data

- ⋆ Missing data is a special case of coarsened data where only some "function" of idealized observation is revealed to the analyst.
- ⋆ In many cases, to aid with identification and estimation of quantities like $\mathbb{E}[Y]$, additional auxiliary information in the form of covariates is available.

Setting	**Ideal**	**Realized**
classical missing data	(X, Y)	(X, T, TY)
Causal inference	$(X, Y(0), Y(1))$	$(X, A, Y(A))$
Monotone missing	(X, Z, Y)	$(X, R_1, R_1Z, R_1R_2, R_1R_2Y)$
Non-monotone missing	(X, Z, Y)	$(X, R_1, R_1Z, R_2, R_2Y)$

Missing Data

- ⋆ Because the reality does not match the ideal sample we want, we often require some identifying assumption that expresses $\mathbb{E}[Y]$ in terms of observed data distribution.
- ⋆ If $\widetilde{O}_1, \ldots, \widetilde{O}_n$ represents the ideal sample and $O_1, \ldots, O_n$ represents the observed sample, we often have

$$O_i = \mathcal{C}_i(\widetilde{O}_i),$$

for some coarsening function $\mathcal{C}_i(\cdot)$ that has information about what parts of $\widetilde{O}_i$ is missing/coarsened.

- ⋆ The identifying assumption provides

$$\mathbb{E}[Y] = \mathbb{E}[g(\widetilde{O})] = \mathbb{E}[f(O; \eta^*)],$$

for some function $f(\cdot;\cdot)$ that is completely known and η^* is some nuisance component that is identified from the observed data distribution.

Efficiency and Multiple Robustness

- The identifying assumption provides

$$\mathbb{E}[Y] = \mathbb{E}[f(O; \eta^*)], \tag{1}$$

  for some known function f with an unknown nuisance η^*.

- In general, there are infinitely many such $f(\cdot;\cdot)$ and the efficient influence function for $\mathbb{E}[Y]$ yields the one that leads to minimal variance.

- Additionally, in many cases, double or multiple robustness occurs.

- Irrespective of efficiency, (1) allows us to estimate $\mathbb{E}[Y]$ as

$$\widehat{\theta}_n := \frac{1}{n}\sum_{i=1}^{n} f(O_i; \widehat{\eta}).$$

  There are multiple estimators depending on whether $\widehat{\eta}$ is computed on a separate sample or whether cross-fitting is employed and so on.

Illustration of Multiplicity and Multiple Robustness

- ⋆ Consider the classical missing data setting: ideal data is $(X_i, Y_i), 1 \leq i \leq n$, observed data is $(X_i, T_i, T_i Y_i)$.
- ⋆ Here

$$T_i = \begin{cases} 1, & \text{if } Y_i \text{ is observed,} \\ 0, & \text{if } Y_i \text{ is unobserved.} \end{cases}$$

- ⋆ One of the most classical identifying assumptions is Missing-at-random (MAR): $T_i \perp Y_i | X_i$.
- ⋆ Under MAR,

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}\left[\frac{TY}{p^*(X)}\right] \quad \text{IPW} \\ &= \mathbb{E}[\mu^*(X)] \quad \text{OR} \\ &= \mathbb{E}\left[\frac{T}{p^*(X)}(Y - \mu^*(X)) + \mu^*(X)\right], \quad \text{AIPW} \end{aligned}$$

where $p^*(X) = \mathbb{P}(T = 1|X)$ and $\mu^*(X) = \mathbb{E}[Y|X] = \mathbb{E}[Y|X, T = 1]$.

Illustration of Multiplicity and Multiple Robustness

⋆ Under MAR,

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}\left[\frac{TY}{p^*(X)}\right] \quad \text{IPW} \\ &= \mathbb{E}[\mu^*(X)] \quad \text{OR} \\ &= \mathbb{E}\left[\frac{T}{p^*(X)}(Y-\mu^*(X))+\mu^*(X)\right], \quad \text{AIPW}\end{aligned}$$

where $p^*(X) = \mathbb{P}(T=1|X)$ and $\mu^*(X) = \mathbb{E}[Y|X] = \mathbb{E}[Y|X,T=1]$.

⋆ Only AIPW satisfies double robustness:

$$|\mathbb{E}[Y] - \text{AIPW}(p,\mu)| \;=\; \left|\mathbb{E}\left[\left(\frac{p^*(X)}{p(X)}-1\right)(\mu(X)-\mu^*(X))\right]\right|.$$

We only need one of p, μ to match of p^*, μ^*.

Outline

Introduction to Missing Data

Introduction to (Conformal) Prediction

Prediction Sets for Missing Data

General Framework

Conclusions

The Problem

⋆ Given IID observations $Y_1, \ldots, Y_n$ from some distribution P, the goal is to construct a prediction set $\widehat{C}_{n,\alpha}$ such that

$$\mathbb{P}_P(Y \in \widehat{C}_{n,\alpha}) \geq 1 - \alpha.$$

⋆ If one has covariate information available apriori, then one might want

$$\mathbb{P}_P(Y \in \widehat{C}_{n,\alpha}(X)) \geq 1 - \alpha.$$

⋆ There are three conditional versions that one might care about:

$$\begin{aligned} \mathbb{P}_P(Y \in \widehat{C}_{n,\alpha}(X)|\text{Data}) &\geq 1 - \alpha, \\ \mathbb{P}_P(Y \in \widehat{C}_{n,\alpha}(X)|X) &\geq 1 - \alpha, \\ \mathbb{P}_P(Y \in \widehat{C}_{n,\alpha}(X)|X, \text{Data}) &\geq 1 - \alpha. \end{aligned}$$

The Basics

- Suppose $W_1, \ldots, W_m$ are IID real-valued random variables from CDF $F(\cdot)$. Assume continuity for now. Set $\mathcal{D}_m = \{W_1, \ldots, W_m\}$
- If

$$W_{(1)} \leq W_{(2)} \leq \cdots \leq W_{(m)},$$

are the order statistics, then

$$F(W_{(k)}) \stackrel{d}{=} U_{(k)},$$

the k-th uniform order statistic from m uniform observations.

- Two observations:

$$\mathbb{P}(W \leq W_{(k)} | \mathcal{D}_n) = F(W_{(k)}) \quad \text{and} \quad U_{(k)} \sim \text{Beta}(k, m+1-k).$$

- Therefore,

$$\mathbb{P}(W \leq W_{(k)}) = \frac{k}{m+1} \quad \Rightarrow \quad (-\infty, W_{(k)}] \text{ is a valid prediction interval.}$$

Non-conformity Score

⋆ What made conformal prediction such a powerful framework is the idea of using a non-conformity score to apply this basic prediction interval to a much broader setting.

⋆ If the observed data is $(X_{-n}, Y_{-n}), \ldots, (X_0, Y_0)$, $(X_1, Y_1), \ldots, (X_n.Y_n)$, then using the first half for $\widehat{\mu}(\cdot)$, we can define, e.g.,

$$R(x, y) = |y - \widehat{\mu}(x)|,$$

some notion of how well (x, y) fits with our understanding/interpretation of underlying DGP.

⋆ Then defining $W_i = |Y_i - \widehat{\mu}(X_i)|, 1 \leq i \leq n$, we get

$$\mathbb{P}(Y \in [\widehat{\mu}(X) \pm q_k]) = \frac{k}{n+1},$$

where q_k is the k-th order statistic of $W_i, 1 \leq i \leq n$.

Outline

Introduction to Missing Data

Introduction to (Conformal) Prediction

Prediction Sets for Missing Data

General Framework

Conclusions

Prediction for Missing Data

- ⋆ When the outcomes can be missing for some of the observations, we cannot implement conformal prediction as it is.
- ⋆ It can be reinterpreted in terms of estimating equations and applied.
- ⋆ But, first, note that our conformal guarantee is

$$\mathbb{P}(Y \in [\widehat{\mu}(X) \pm q_k]) = \frac{k}{n+1},$$

which means we need to take $k = \lceil (n+1)(1-\alpha) \rceil$ in order to ensure

$$\mathbb{P}(Y \in [\widehat{\mu}(X) \pm q_k]) \geq 1 - \alpha.$$

- ⋆ If we take $k' = \lceil n(1-\alpha) \rceil$, then $q_{k'}$ is the $(1-\alpha)$-th quantile on n points and we only get

$$\mathbb{P}(Y \in [\widehat{\mu} \pm q_{k'}]) \geq 1 - \alpha - \frac{1}{n+1}.$$

Reinterpretation of Conformal Prediction

⋆ If $Y_1, \ldots, Y_n$ is an IID sequence with population CDF $F(\cdot)$, then the oracle one-sided prediction interval would have been

$$(-\infty, q_\alpha],$$

where

$$\mathbb{P}(Y \leq q_\alpha) = 1 - \alpha \quad \equiv \quad \mathbb{E}[\mathbf{1}\{Y \leq q_\alpha\}] = 1 - \alpha.$$

⋆ Based on the data, we have

$$\widehat{F}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Y_i \leq \theta\}.$$

⋆ A natural estimator of q_α would be

$$\widehat{q}_\alpha = \inf \left\{ \theta : \widehat{F}_n(\theta) \geq 1 - \alpha \right\} = Y_{(\lceil n(1-\alpha) \rceil)}.$$

⋆ We already proved

$$\mathbb{E}[\mathbf{1}\{Y \leq \widehat{q}_\alpha\}] \; = \; \mathbb{E}[F(\widehat{q}_\alpha)] \; \geq \; 1 - \alpha - 1/(n+1).$$

Reinterpretation of Conformal Prediction

⋆ If $Y_1, \ldots, Y_n$ is an IID sequence with population CDF $F(\cdot)$, then the oracle one-sided prediction interval would have been

$$(-\infty, q_\alpha],$$

where

$$\mathbb{P}(Y \leq q_\alpha) = 1 - \alpha \quad \equiv \quad \mathbb{E}[\mathbf{1}\{Y \leq q_\alpha\}] = 1 - \alpha.$$

⋆ Based on the data, we have

$$\widehat{F}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Y_i \leq \theta\}.$$

⋆ A natural estimator of q_α would be

$$\widehat{q}_\alpha = \inf \left\{ \theta : \widehat{F}_n(\theta) \geq 1 - \alpha \right\} = Y_{(\lceil n(1-\alpha) \rceil)}.$$

⋆ We already proved

$$\mathbb{E}[\mathbf{1}\{Y \leq \widehat{q}_\alpha\}] \; = \; \mathbb{E}[F(\widehat{q}_\alpha)] \; \geq \; 1 - \alpha - 1/(n+1).$$

Some comments

- ⋆ Note that noting about the consistency or the rate of convergence of the quantile estimator $\widehat{q}_\alpha$ comes up.
- ⋆ This is surprising when thought from estimating equations point of view.
- ⋆ It can be reconciled using

$$\widehat{F}_n(\widehat{q}_\alpha) - F(\widehat{q}_\alpha) = \widehat{F}_n(q_\alpha) - F(q_\alpha) + [(\widehat{F}_n - F)(\widehat{q}_\alpha) - (\widehat{F}_n - F)(q_\alpha)].$$

- ⋆ The first term is mean zero (but is $O_p(n^{-1/2})$), the second term has a non-zero mean but is $O_p(n^{-1/2})\sqrt{|F(\widehat{q}_\alpha) - F(q_\alpha)|}$.

Back to Missing Data: Illustration with MAR

⋆ Suppose we observe

$$\mathcal{D}_{-1} := \{(X_j, T_j, T_j Y_j) : -n \leq j \leq 0\},$$
$$\mathcal{D}_n = \{(X_i, T_i, T_i Y_i) : 1 \leq i \leq n\}.$$

⋆ The oracle one-sided prediction set for Y would be $(-\infty, q_\alpha]$, where

$$\mathbb{P}(Y \leq q_\alpha) = 1 - \alpha \quad \equiv \quad \mathbb{E}[\mathbf{1}\{Y \leq q_\alpha\}] = 1 - \alpha.$$

⋆ Unlike the full data setting, we cannot estimate this expectation using

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Y_i \leq \theta\},$$

or

$$\frac{\sum_{i=1}^{n} T_i \mathbf{1}\{Y_i \leq \theta\}}{\sum_{i=1}^{n} T_i},$$

because they are inconsistent estimators, under MAR.

Illustration with MAR

⋆ Under MAR, we know that

$$\mathbb{E}[\mathbf{1}\{Y \le \theta\}] = \mathbb{E}\left[\frac{T\mathbf{1}\{Y \le \theta\}}{p^*(X)}\right], \quad p^*(x) = \mathbb{P}(T = 1|X = x).$$

⋆ One estimate of q_α (and the prediction interval) can be obtained as

$$\widehat{q}_\alpha^{\mathrm{IPW}} := \inf\left\{\theta : \frac{1}{n}\sum_{i=1}^{n}\frac{T_i\mathbf{1}\{Y_i \le \theta\}}{\widehat{p}(X_i)} \ge 1-\alpha\right\},$$

where $\widehat{p}(\cdot)$ is an estimator of $\mathbb{P}(T = 1|X = x)$ from $\mathcal{D}_{-1}$.

⋆ This essentially yields the weighted conformal prediction by Tibshirani et al. (2019).

⋆ One can also consider the Hajek estimator

$$\widehat{q}_\alpha^{\mathrm{IPW}} := \inf\left\{\theta : \frac{\sum_{i=1}^{n} T_i\mathbf{1}\{Y_i \le \theta\}/\widehat{p}(X_i)}{\sum_{i=1}^{n} T_i/\widehat{p}(X_i)} \ge 1-\alpha\right\},$$

Illustration with MAR

- It can be proved, using Angelopoulos et al. (2022), that

$$\mathbb{P}(Y \leq \widehat{q}_\alpha^{\mathrm{IPW}}) \geq 1 - \alpha - \frac{\mathfrak{C}\|1/\widehat{p}\|_\infty}{n} - \mathfrak{C}\left\|\frac{1}{\widehat{p}(X)} - \frac{1}{p^*(X)}\right\|_2,$$

  for some absolute constant $\mathfrak{C}$.

- The sample size dependent term on the right can be avoided by a simple correction, but the estimation error term for $p^*(\cdot)$ cannot be.

- Similarly, one can define the outcome regression version:

$$\widehat{q}_\alpha^{\mathrm{OR}} := \inf\left\{\theta : \frac{1}{n}\sum_{i=1}^{n} \widehat{m}(\theta, X_i) \geq 1 - \alpha\right\},$$

  where $\widehat{m}(\theta, x)$ is an estimator of $\mathbb{E}[\mathbf{1}\{Y \leq \theta\}|X = x]$ from $\mathcal{D}_{-1}$.

- Neither of these is the most efficient estimator of q_α under MAR.

Double Robust Prediction Sets with MAR

⋆ We can estimate $F(\theta) = \mathbb{P}(Y \leq \theta)$ using a double robust estimator as

$$\widehat{F}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{T_i}{\widehat{p}(X_i)}(\mathbf{1}\{Y_i \leq \theta\} - \widehat{m}(\theta, X_i)) + \widehat{m}(\theta, X_i)\right].$$

⋆ This is a double robust estimator converging to $F(\theta)$ as long as one of $\|\widehat{p} - p^*\|$ or $\|\widehat{m} - m^*\|$ is converging to zero.

⋆ Note that $\theta \mapsto \widehat{F}_n(\theta)$ is not monotonic. Define

$$\widehat{q}_\alpha^{\mathrm{DR}} := \inf\left\{\theta : \widehat{F}_n(\theta) \geq 1 - \alpha\right\}.$$

⋆ This was proposed in Yang et al. (2024, JRSSB), and using

$$\begin{aligned} F(\widehat{q}_\alpha^{\mathrm{DR}}) &= \widehat{F}_n(\widehat{q}_\alpha^{\mathrm{DR}}) - (\widehat{F}_n - F)(\widehat{q}_\alpha^{\mathrm{DR}}) \\ &\geq \alpha - \sup_\theta |(\widehat{F}_n - F)(\theta)|, \end{aligned}$$

we proved a coverage of at least $1 - \alpha - O(1/\sqrt{n}) - \|\widehat{p} - p^*\|\|\widehat{m} - m_n^*\|$.

Double Robust Prediction Sets with MAR

⋆ The IPW prediction set satisfies

$$\mathbb{P}(Y \leq \widehat{q}_\alpha^{\mathrm{IPW}}) \geq 1 - \alpha - O(1/n) - \|\widehat{p} - p^*\|,$$

while the DR prediction set satisfies

$$\mathbb{P}(Y \leq \widehat{q}_\alpha^{\mathrm{DR}}) \geq 1 - \alpha - O(1/\sqrt{n}) - \|\widehat{p} - p^*\|\|\widehat{m} - m^*\|.$$

⋆ The proof techniques are rather different: IPW coverage is proved using monotonicity of its estimate of CDF, while DR coverage is proved using empirical process techniques.

⋆ In the current work, we improved the guarantee of DR prediction set to

$$\mathbb{P}(Y \leq \widehat{q}_\alpha^{\mathrm{DR}}) \geq 1 - \alpha - O(a_p a_m + \Delta_n),$$

where $a_p = \|\widehat{p} - p^*\|, a_m = \|\widehat{m} - m^*\|$ and

$$\Delta_n = \log(n)\left[\frac{1}{n^{3/4}} + \frac{a_m}{\sqrt{n}}\right].$$

Double Robust Prediction Sets

⋆ This improvement in rate stems from replacing

$$\begin{aligned} F(\widehat{q}_\alpha^{\mathrm{DR}}) &= \widehat{F}_n(\widehat{q}_\alpha^{\mathrm{DR}}) - (\widehat{F}_n - F)(\widehat{q}_\alpha^{\mathrm{DR}}) \\ &\geq \alpha - \sup_\theta |(\widehat{F}_n - F)(\theta)|, \end{aligned}$$

with

$$\begin{aligned} F(\widehat{q}_\alpha^{\mathrm{DR}}) = \widehat{F}_n(\widehat{q}_\alpha^{\mathrm{DR}}) \\ - [\widehat{F}_n(q_\alpha) - F(q_\alpha)] \\ - [(\widehat{F}_n - F)(\widehat{q}_\alpha^{\mathrm{DR}}) - (\widehat{F}_n - F)(q_\alpha)], \end{aligned}$$

⋆ These rate improvements are only for double marginal coverage $\mathbb{P}(Y \leq \widehat{q}_\alpha)$, not for $F(\widehat{q}_\alpha)$.

⋆ $F(\widehat{q}_\alpha) - (1-\alpha)$ is always *at least* of order $1/\sqrt{n}$.

Classical missing data

Worst-case undercoverage	$a_p = 0,\ a_m > 0$	$a_p > 0,\ a_m = 0$	$a_p = 0,\ a_m = 0$
IPW/WCP (Tibshirani et al.)	0	$O(a_p)$	0
OR/CRC (Angelopoulos et al.)	$O(a_m)$	0	0
DR (bound in Yang et al.)	$O(n^{-1/2})$	$O(n^{-1/2})$	$O(n^{-1/2})$
DR (sharpened bound)	$O(n^{-3/4} + a_m n^{-1/2})$	$O(n^{-3/4})$	$O(n^{-3/4})$
RSCP (Not presented)	0	$O(a_p n^{-1/2})$	0

$a_p = \|\widehat{p} - p^*\|_2$ and $a_m = \sup_\theta \|\widehat{m}(\theta, \cdot) - m^*(\theta, \cdot)\|_2$.

Classical missing data (Simulation Study)

$$X_1, X_2, \varepsilon \overset{iid}{\sim} N(0,1), \qquad Y = 20 + 70X_1 - 50X_2 + \varepsilon.$$

$$p^*(X) = 0.2 + 0.6 \ \text{logit}^{-1}(0.4 + 0.6X_1 - 0.6X_2),$$
$$\widehat{p}(X) = 0.2 + 0.6 \ \text{logit}^{-1}(0.4 - 0.6X_1 + 0.6X_2),$$
$$\widehat{m}(\theta, X) = \Phi(\theta).$$

We use $\alpha = 0.1$, $n = 100$, and 100 repetitions.

Classical missing data (Simulation Study)

Method	$\widehat{\pi} = \pi^*$ $\widehat{m} \neq m^*$	$\widehat{\pi} \neq \pi^*$ $\widehat{m} = m^*$	$\widehat{\pi} = \pi^*$ $\widehat{m} = m^*$
IPW/WCP	0.879 ± 0.070	0.658 ± 0.075	0.879 ± 0.070
OR/CRC	0.337 ± 0.023	0.896 ± 0.045	0.896 ± 0.045
DR	0.876 ± 0.079	0.895 ± 0.046	0.886 ± 0.049
RSCP	0.932 ± 0.066	0.955 ± 0.066	0.948 ± 0.046

The mean coverage $\pm$ standard deviation is provided for all the methods across different scenarios; the target coverage level is 0.90. IPW/WCP fails when the propensity score is incorrectly estimated, OR/CRC fails when the outcome regression is incorrectly estimated, and both the EIF-based procedures are substantially more stable.

Outline

Introduction to Missing Data

Introduction to (Conformal) Prediction

Prediction Sets for Missing Data

General Framework

Conclusions

General Framework

- ⋆ Suppose we observe $O_{-n}, \ldots, O_0, O_1, \ldots, O_n$ and we want to construct a prediction set for the response.
- ⋆ Make an identifying assumption on the DGP so that

$$\mathbb{E}[\mathbf{1}\{R(Y;O) \leq \theta\}] = \mathbb{E}[f(\theta; O, \eta^*)],$$

for some known function f and some nuisance η^* identified from observed data distribution; well-studied in semiparametrics.

- ⋆ Then find

$$\widehat{q}_\alpha := \inf\left\{\theta : \frac{1}{n}\sum_{i=1}^{n} f(\theta; O_i, \widehat{\eta}) \geq 1-\alpha\right\}.$$

- ⋆ If $f(\cdot)$ is an influence function, then a second order error in terms of $\widehat{\eta}$ appears in the coverage error. If η^* is a vector of nuisances, then usually double/multiple robustness appears in the coverage error.

General Framework

- ⋆ If $\theta \mapsto f(\theta; O, \widehat{\eta})$ is monotonic, then the conformal risk control Angelopoulos et al. (2022) framework yields sharp miscoverage errors.
- ⋆ Without monotonicity, one needs to fall back on empirical process techniques.
- ⋆ Some recent works have relaxed monotonicity conditions of Angelopoulos et al. (2022), but they are still weaker than our result for the MAR setting.
- ⋆ Semiparametric theory offers a systematic way to obtained efficient prediction sets under coarsened data. In contrast, some part of ML/Stats literature relies on ad hoc constructions.

Outline

Introduction to Missing Data

Introduction to (Conformal) Prediction

Prediction Sets for Missing Data

General Framework

Conclusions

Conclusions

- ⋆ Semiparametric theory has a lot to offer in the way of constructing efficient prediction with better finite-sample properties.
- ⋆ Theoretical guarantees for methods based on efficient influence function are tricky when relying only on exchangeability based arguments.
- ⋆ While IPW and OR methods get exact finite-sample coverage when their respective nuisance functions are well-specified, AIPW does not get exact finite-sample validity when at least one of them is correct.
- ⋆ It is an interesting open question to check if this is possible.
- ⋆ Finally, the question of optimal shape of the prediction sets in the context of coarsened data is unclear.