Debiased Machine Learning for Conformal Prediction of Counterfactual Outcomes Under Runtime Confounding

CLeaR 2026

Keith Barnatchez, Kevin P. Josey, Rachel C. Nethery, and Giovanni Parmigiani

April 7th, 2025

Collaborators

Rachel Nethery

Giovanni Parmigiani

Kevin Josey

Roadmap

Roadmap

Motivation: Counterfactual predictions and prediction intervals

Growing set of methods for predicting **counterfactual outcomes** $Y(a)$ under different treatments $a \in \mathcal{A}$, given a covariate profile $\boldsymbol{X}$

$$\underbrace{\hat{\mathbb{E}}[\, Y(a) \mid \boldsymbol{X} \,]}_{\boldsymbol{X}-\text{conditional outcome under tmt. A=a}}$$

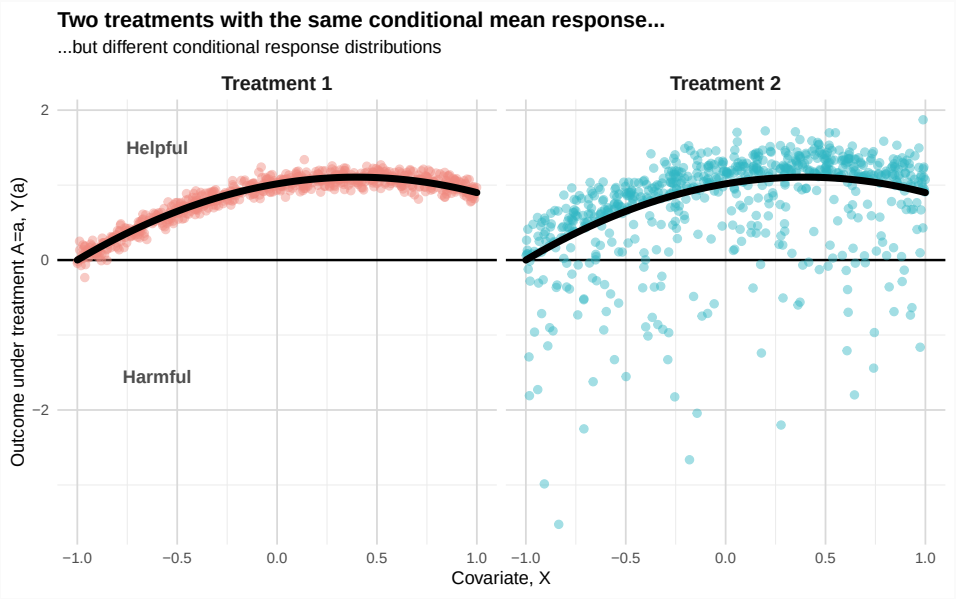
Motivation: Counterfactual predictions and prediction intervals

Growing set of methods for predicting **counterfactual outcomes** $Y(a)$ under different treatments $a \in \mathcal{A}$, given a covariate profile $\boldsymbol{X}$

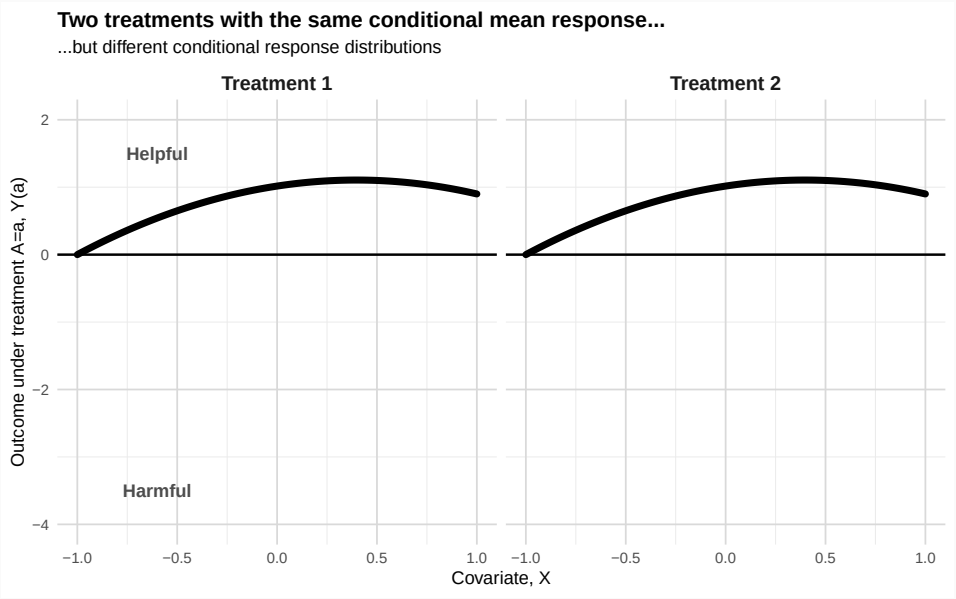
$$\underbrace{\hat{\mathbb{E}}[\, Y(a) \mid \boldsymbol{X} \,]}_{\boldsymbol{X}-\text{conditional outcome under tmt. A=a}}$$

Along with point predictions, prediction intervals can serve as key inputs for decision making

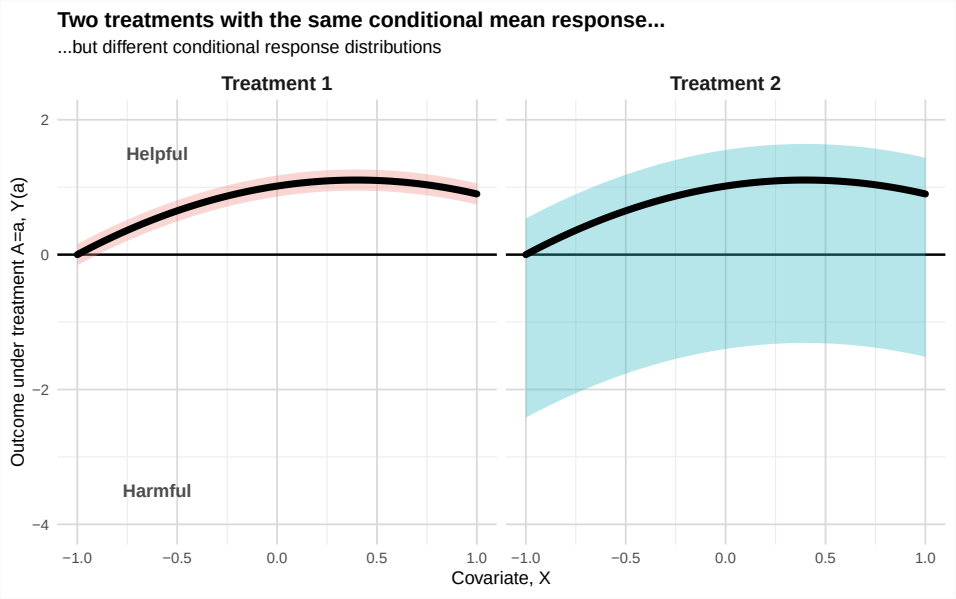
Counterfactual predictions vs. prediction intervals



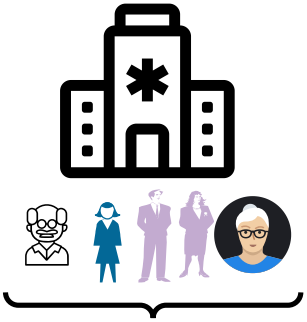
Counterfactual predictions vs. prediction intervals



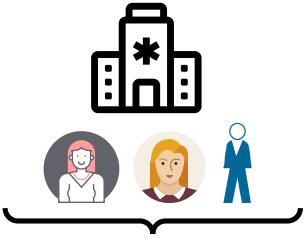
Counterfactual predictions vs. prediction intervals



Roadmap

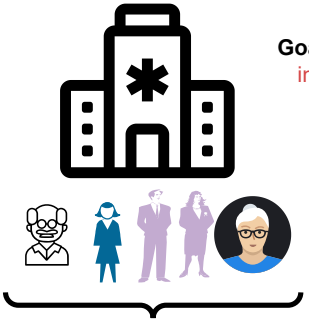


$(Y_i, A_i, \mathbf{X}_i)_{i=1}^{n_1}$

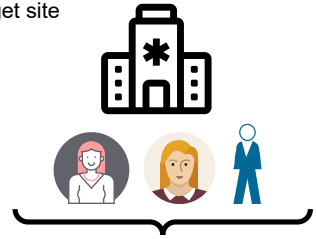


$(\mathbf{V}_i)_{i=1}^{n_0}$

where $\mathbf{V} \subset \mathbf{X}$

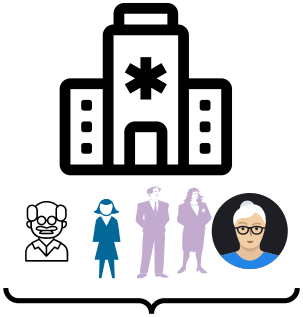


Goal: use info from source site to inform decisions in target site



$(Y_i, A_i, \mathbf{X}_i)_{i=1}^{n_1}$

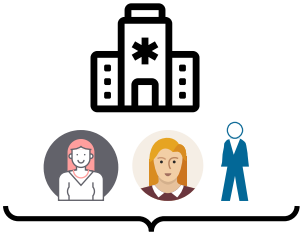
$(\mathbf{V}_i)_{i=1}^{n_0}$
where $\mathbf{V} \subset \mathbf{X}$



$(Y_i, A_i, \mathbf{X}_i)_{i=1}^{n_1}$

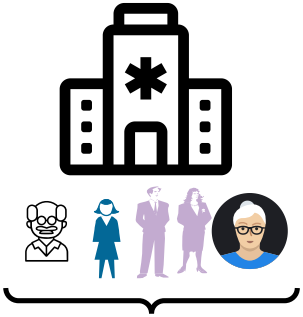
$\downarrow$

$\hat{m}_a(\mathbf{X}) = \hat{\mathbb{E}}[Y|A = a, \mathbf{X}]$

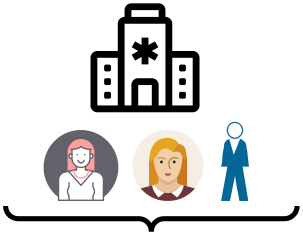


$(\mathbf{V}_i)_{i=1}^{n_0}$

where $\mathbf{V} \subset \mathbf{X}$



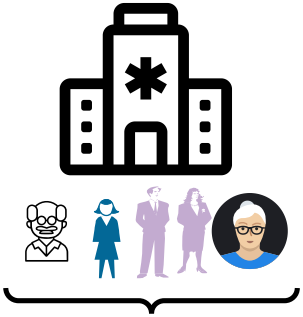
$(Y_i, A_i, \mathbf{X}_i)_{i=1}^{n_1}$



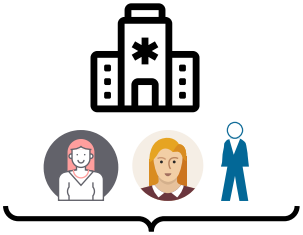
$(\mathbf{V}_i)_{i=1}^{n_0}$
where $\mathbf{V} \subset \mathbf{X}$

$\hat{m}_a(\mathbf{X}) = \hat{\mathbb{E}}[Y|A = a, \mathbf{X}]$

$\hat{m}_a(\mathbf{V})$



$(Y_i, A_i, \mathbf{X}_i)_{i=1}^{n_1}$

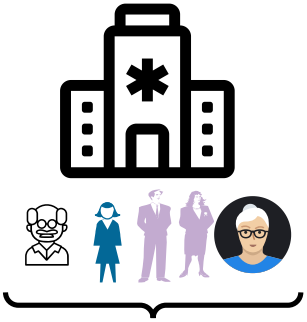


$(\mathbf{V}_i)_{i=1}^{n_0}$

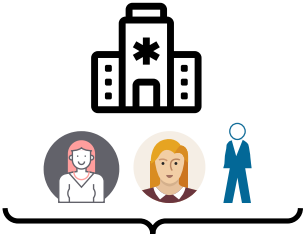
where $\mathbf{V} \subset \mathbf{X}$

$\hat{m}_a(\mathbf{X}) = \hat{\mathbb{E}}[Y|A = a, \mathbf{X}]$

$\hat{m}_a(\mathbf{V})$

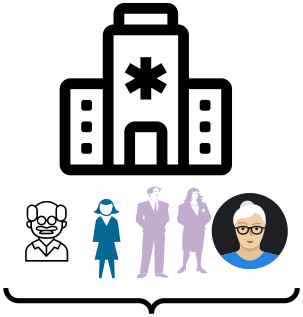


$(Y_i, A_i, \mathbf{X}_i)_{i=1}^{n_1}$



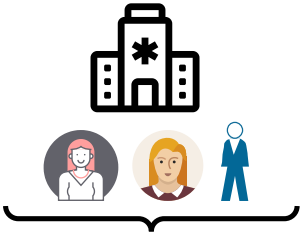
$(\mathbf{V}_i)_{i=1}^{n_0}$

where $\mathbf{V} \subset \mathbf{X}$



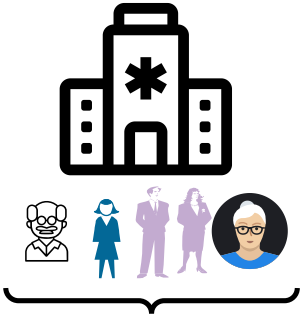
$(Y_i, A_i, \mathbf{X}_i)_{i=1}^{n_1}$

$\hat{m}_a(\mathbf{V}) = \hat{\mathbb{E}}[Y|A = a, \mathbf{V}]$



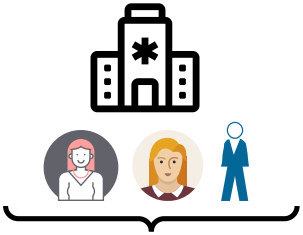
$(\mathbf{V}_i)_{i=1}^{n_0}$

where $\mathbf{V} \subset \mathbf{X}$



$(Y_i, A_i, \mathbf{X}_i)_{i=1}^{n_1}$

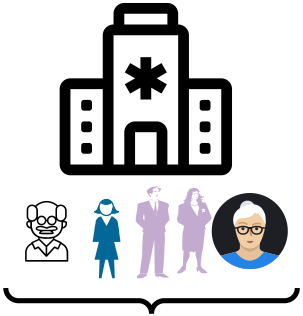
$\hat{m}_a(\mathbf{V}) = \hat{\mathbb{E}}[Y|A = a, \mathbf{V}]$



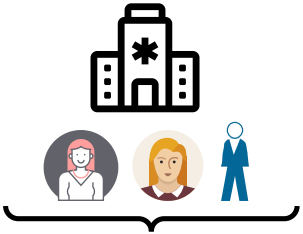
$(\mathbf{V}_i)_{i=1}^{n_0}$

where $\mathbf{V} \subset \mathbf{X}$

$\hat{m}_a(\mathbf{V})$



$(Y_i, A_i, \mathbf{X}_i)_{i=1}^{n_1}$

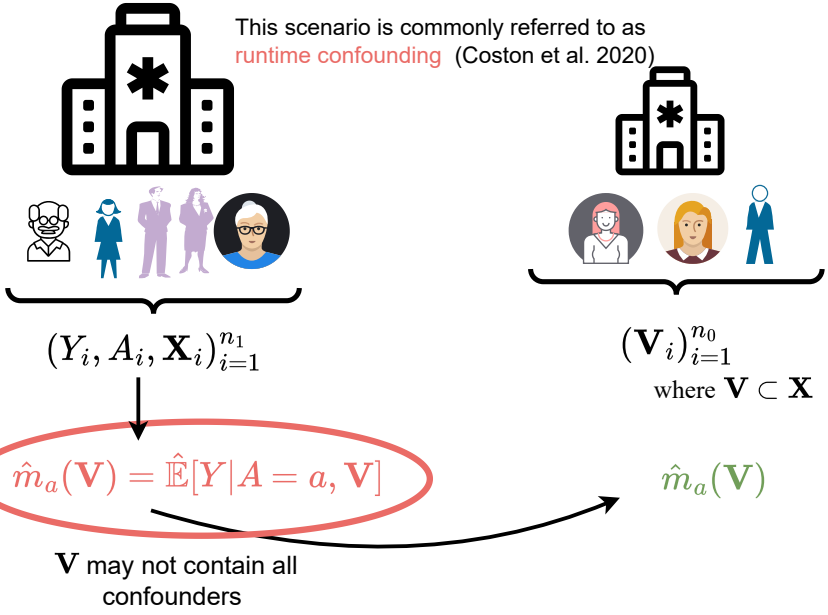


$(\mathbf{V}_i)_{i=1}^{n_0}$
where $\mathbf{V} \subset \mathbf{X}$

$\hat{m}_a(\mathbf{V}) = \hat{\mathbb{E}}[Y|A = a, \mathbf{V}]$

$\hat{m}_a(\mathbf{V})$

$\mathbf{V}$ may not contain all confounders


This scenario is commonly referred to as
runtime confounding (Coston et al. 2020)
$(Y_i, A_i, \mathbf{X}_i)_{i=1}^{n_1}$
$(\mathbf{V}_i)_{i=1}^{n_0}$
where $\mathbf{V} \subset \mathbf{X}$
$\hat{m}_a(\mathbf{V}) = \hat{\mathbb{E}}[Y|A = a, \mathbf{V}]$
$\hat{m}_a(\mathbf{V})$
$\mathbf{V}$ may not contain all confounders

Contributions

This paper...

- Proposes methods for constructing prediction intervals for counterfactual outcomes $Y(a)$ under runtime confounding

- We construct these intervals in a model-agnostic way that
  - Ensures their validity asymptotically, and
  - Relies on less data to obtain $\approx$ valid coverage than commonly-taken approaches

Roadmap

Roadmap

Problem setting

Suppose we observe...

- $(Y_i, A_i, \boldsymbol{X}_i),\ i = 1, \ldots, n_1$, from a source population
- $\boldsymbol{V}_i,\ \ i = 1, \ldots, n_0$ from a target population, where $\boldsymbol{V}$ is a subset of all the covariates collected in $\boldsymbol{X}$, where $\boldsymbol{X} = (\boldsymbol{V}, \boldsymbol{U})$

Problem setting

Suppose we observe...

- $(Y_i, A_i, \boldsymbol{X}_i),\ i = 1, \ldots, n_1$, from a source population
- $\boldsymbol{V}_i,\ \ i = 1, \ldots, n_0$ from a target population, where $\boldsymbol{V}$ is a subset of all the covariates collected in $\boldsymbol{X}$, where $\boldsymbol{X} = (\boldsymbol{V}, \boldsymbol{U})$

Interest lies in constructing prediction intervals for counterfactual outcomes in the target population which attain a desired marginal coverage probability of $1 - \alpha$:

$$\mathbb{P}\{\, Y_i(a) \in \hat{C}_a(\boldsymbol{V}_i) \mid \text{In target pop.}\, \} = 1 - \alpha, \qquad a \in \mathcal{A}$$

Problem setting

Suppose we observe...

- $(Y_i, A_i, \boldsymbol{X}_i),\ i = 1, \ldots, n_1$, from a source population
- $\boldsymbol{V}_i,\ \ i = 1, \ldots, n_0$ from a target population, where $\boldsymbol{V}$ is a subset of all the covariates collected in $\boldsymbol{X}$, where $\boldsymbol{X} = (\boldsymbol{V}, \boldsymbol{U})$

Interest lies in constructing prediction intervals for counterfactual outcomes in the target population which attain a desired marginal coverage probability of $1 - \alpha$:

$$\mathbb{P}\{\, Y_i(a) \in \hat{C}_a(\boldsymbol{V}_i) \mid \text{In target pop.}\, \} = 1 - \alpha, \qquad a \in \mathcal{A}$$

Y	A	$\boldsymbol{V}$	$\boldsymbol{U}$
Y_1	A_1	$\boldsymbol{V}_1$	$\boldsymbol{U}_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_{n_1}	A_{n_1}	$\boldsymbol{V}_{n_1}$	$\boldsymbol{U}_{n_1}$

Source sample

Y	A	$\boldsymbol{V}$	$\boldsymbol{U}$
NA	NA	$\boldsymbol{V}_{n_1+1}$	NA
$\vdots$	$\vdots$	$\vdots$	$\vdots$
NA	NA	$\boldsymbol{V}_{n_1+n_0}$	NA

Target sample

Problem setting

Suppose we observe...

- $(Y_i, A_i, \boldsymbol{X}_i),\ i = 1, \ldots, n_1$, from a source population
- $\boldsymbol{V}_i,\ \ i = 1, \ldots, n_0$ from a target population, where $\boldsymbol{V}$ is a subset of all the covariates collected in $\boldsymbol{X}$, where $\boldsymbol{X} = (\boldsymbol{V}, \boldsymbol{U})$

Interest lies in constructing prediction intervals for counterfactual outcomes in the target population which attain a desired marginal coverage probability of $1 - \alpha$:

$$\mathbb{P}\{\, Y_i(a) \in \hat{C}_a(\boldsymbol{V}_i) \mid S = 0 \,\} = 1 - \alpha, \qquad a \in \mathcal{A}$$

Y	A	$\boldsymbol{V}$	$\boldsymbol{U}$	S
Y_1	A_1	$\boldsymbol{V}_1$	$\boldsymbol{U}_1$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_{n_1}	A_{n_1}	$\boldsymbol{V}_{n_1}$	$\boldsymbol{U}_{n_1}$	1
NA	NA	$\boldsymbol{V}_{n_1+1}$	NA	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
NA	NA	$\boldsymbol{V}_{n_1+n_0}$	NA	0

Problem setting

Suppose we observe...

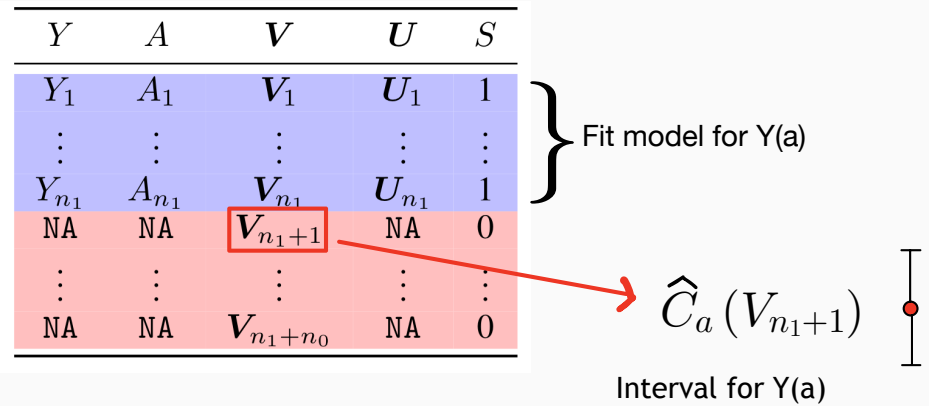
- $(Y_i, A_i, \boldsymbol{X}_i),\ i = 1, \ldots, n_1$, from a source population
- $\boldsymbol{V}_i,\ \ i = 1, \ldots, n_0$ from a target population, where $\boldsymbol{V}$ is a subset of all the covariates collected in $\boldsymbol{X}$, where $\boldsymbol{X} = (\boldsymbol{V}, \boldsymbol{U})$

Interest lies in constructing prediction intervals for counterfactual outcomes in the target population which attain a desired marginal coverage probability of $1 - \alpha$:

$$\mathbb{P}\{\, Y_i(a) \in \hat{C}_a(\boldsymbol{V}_i) \mid S = 0 \,\} = 1 - \alpha, \qquad a \in \mathcal{A}$$

This is an example of runtime confounding: the covariates we have access to at test time differ from those available to us at training time (Coston et al. 2020)

Counterfactual prediction intervals



Assumptions

1. **(Positivity in the source study)**: $0 < \mathbb{P}(A = a|\boldsymbol{X}, S = 1) < 1$
2. **(Consistency)**: $Y = \sum_{a \in \mathcal{A}} I(A = a) \cdot Y(a)$
3. **(Positivity of source membership)**: $0 < \mathbb{P}(S = 1|\boldsymbol{V}) < 1$
4. **Runtime confounding in source**: $Y(a) \perp\!\!\!\perp A|\boldsymbol{X}, S = 1$, but $Y(a) \not\perp\!\!\!\perp A|\boldsymbol{V}, S = 1$
5. **Source exchangeability**: $Y(a) \perp\!\!\!\perp S|\boldsymbol{V}$

Assumptions

1. **(Positivity in the source study)**: $0 < \mathbb{P}(A = a|\boldsymbol{X}, S = 1) < 1$
2. **(Consistency)**: $Y = \sum_{a \in \mathcal{A}} I(A = a) \cdot Y(a)$
3. **(Positivity of source membership)**: $0 < \mathbb{P}(S = 1|\boldsymbol{V}) < 1$
4. **Runtime confounding in source**: $Y(a) \perp\!\!\!\perp A|\boldsymbol{X}, S = 1$, but $Y(a) \not\perp\!\!\!\perp A|\boldsymbol{V}, S = 1$
5. **Source exchangeability**: $Y(a) \perp\!\!\!\perp S|\boldsymbol{V}$

Roadmap

Conformal prediction: a tool for constructing valid prediction intervals

Conformal prediction (Vovk et al., 2005) is a general framework that

- Takes an arbitrary prediction function $\hat{f}(\boldsymbol{X})$ for an outcome Y, and
- Outputs a prediction interval function $\hat{C}(\boldsymbol{X})$ for new observations such that

$$\mathbb{P}(Y \in \hat{C}(\boldsymbol{X})) \geq 1 - \alpha$$

Conformal prediction: a tool for constructing valid prediction intervals

Conformal prediction (Vovk et al., 2005) is a general framework that

- Takes an arbitrary prediction function $\hat{f}(\boldsymbol{X})$ for an outcome Y, and
- Outputs a prediction interval function $\hat{C}(\boldsymbol{X})$ for new observations such that

$$\mathbb{P}(Y \in \hat{C}(\boldsymbol{X})) \geq 1 - \alpha$$

When data are i.i.d. or exchangeable, above guarantee holds in finite samples

Conformal prediction: a tool for constructing valid prediction intervals

Conformal prediction (Vovk et al., 2005) is a general framework that

- Takes an arbitrary prediction function $\hat{f}(\boldsymbol{X})$ for an outcome Y, and
- Outputs a prediction interval function $\hat{C}(\boldsymbol{X})$ for new observations such that

$$\mathbb{P}(Y \in \hat{C}(\boldsymbol{X})) \geq 1 - \alpha$$

When data are i.i.d. or exchangeable, above guarantee holds in finite samples

In broader settings (e.g. covariate shift between train + test data), finite sample coverage is generally **not guaranteed** (Bhattacharyya and Barber, 2026)

- The source of distribution shift needs to be estimated

The key ingredient in conformal prediction

The key idea behind conformal prediction (no covariate shift, everything i.i.d.) is to

1. Obtain absolute residuals $R(\boldsymbol{X}_i, Y_i) = |Y_i - \hat{f}(\boldsymbol{X}_i)|$ (example of a conformity score)
2. Let $\hat{C}(\boldsymbol{X}_i) = \hat{f}(\boldsymbol{X}_i) \pm r_\alpha$, where r_α is the $1-\alpha$ quantile of the abs. residuals
3. Notice this will imply for a new observation that $\mathbb{P}(Y \in \hat{C}(\boldsymbol{X})) = 1 - \alpha$

The key ingredient in conformal prediction

The key idea behind conformal prediction (no covariate shift, everything i.i.d.) is to

1. Obtain absolute residuals $R(\boldsymbol{X}_i, Y_i) = |Y_i - \hat{f}(\boldsymbol{X}_i)|$ (example of a conformity score)
2. Let $\hat{C}(\boldsymbol{X}_i) = \hat{f}(\boldsymbol{X}_i) \pm r_\alpha$, where r_α is the $1-\alpha$ quantile of the abs. residuals
3. Notice this will imply for a new observation that $\mathbb{P}(Y \in \hat{C}(\boldsymbol{X})) = 1 - \alpha$

This is because

$$\{Y \in \hat{C}(\boldsymbol{X})\} = \{\text{The residual } R_i \leq r_\alpha\} = \{\text{The residual } R_i \text{ falls below its } 1-\alpha \text{ quantile}\}$$

Notice the probability of the last event is $1 - \alpha$

The key ingredient in conformal prediction

The key idea behind conformal prediction (no covariate shift, everything i.i.d.) is to

1. Obtain absolute residuals $R(\boldsymbol{X}_i, Y_i) = |Y_i - \hat{f}(\boldsymbol{X}_i)|$ (example of a conformity score)
2. Let $\hat{C}(\boldsymbol{X}_i) = \hat{f}(\boldsymbol{X}_i) \pm r_\alpha$, where r_α is the $1-\alpha$ quantile of the abs. residuals
3. Notice this will imply for a new observation that $\mathbb{P}(Y \in \hat{C}(\boldsymbol{X})) = 1 - \alpha$

The key ingredient in conformal prediction

The key idea behind conformal prediction (no covariate shift, everything i.i.d.) is to

1. Obtain absolute residuals $R(\boldsymbol{X}_i, Y_i) = |Y_i - \hat{f}(\boldsymbol{X}_i)|$ (example of a conformity score)
2. Let $\hat{C}(\boldsymbol{X}_i) = \hat{f}(\boldsymbol{X}_i) \pm r_\alpha$, where r_α is the $1-\alpha$ quantile of the abs. residuals
3. Notice this will imply for a new observation that $\mathbb{P}(Y \in \hat{C}(\boldsymbol{X})) = 1 - \alpha$

For our problem of interest, this means we seek an interval $\hat{C}_a(\boldsymbol{V}) = \hat{\mu}_a(\boldsymbol{V}) \pm r_{a,\alpha}$ such that

$$\mathbb{P}(R(\boldsymbol{V}_i, Y_i(a)) \leq r_{a,\alpha} | S = 0) = 1 - \alpha$$

The key ingredient in conformal prediction

The key idea behind conformal prediction (no covariate shift, everything i.i.d.) is to

1. Obtain absolute residuals $R(\boldsymbol{X}_i, Y_i) = |Y_i - \hat{f}(\boldsymbol{X}_i)|$ (example of a conformity score)
2. Let $\hat{C}(\boldsymbol{X}_i) = \hat{f}(\boldsymbol{X}_i) \pm r_\alpha$, where r_α is the $1-\alpha$ quantile of the abs. residuals
3. Notice this will imply for a new observation that $\mathbb{P}(Y \in \hat{C}(\boldsymbol{X})) = 1 - \alpha$

For our problem of interest, this means we seek an interval $\hat{C}_a(\boldsymbol{V}) = \hat{\mu}_a(\boldsymbol{V}) \pm r_{a,\alpha}$ such that

$$\mathbb{P}(R(\boldsymbol{V}_i, Y_i(a)) \leq r_{a,\alpha} | S = 0) = 1 - \alpha$$

Key problem: We don't observe outcomes, treatments, or all of $\boldsymbol{X}$ in the target pop, so we can't directly estimate $r_{a,\alpha}$ for our target population

- No way to compute residuals R_i, so there's no way to directly estimate $r_{a,\alpha}$

Roadmap

Quick aside: forming counterfactual predictions

While we remain agnostic about how predictions for $Y(a)$ are formed, it's useful to think about how to construct valid point predictions under runtime confounding

Quick aside: forming counterfactual predictions

While we remain agnostic about how predictions for $Y(a)$ are formed, it's useful to think about how to construct valid point predictions under runtime confounding

Coston et al. (2020) provide useful framework

1. First, fit $\hat{m}_a(\boldsymbol{X}) = \hat{\mathbb{E}}(Y|A = a, \boldsymbol{X})$ with the training data
2. Next, regress the estimated $\hat{m}_a(\boldsymbol{X})$ on $\boldsymbol{V}$, yielding $\hat{\mu}_a(\boldsymbol{V}) = \hat{\mathbb{E}}(\hat{m}_a(\boldsymbol{X})|\boldsymbol{V})$

Quick aside: forming counterfactual predictions

While we remain agnostic about how predictions for $Y(a)$ are formed, it's useful to think about how to construct valid point predictions under runtime confounding

Coston et al. (2020) provide useful framework

1. First, fit $\hat{m}_a(\boldsymbol{X}) = \hat{\mathbb{E}}(Y|A=a,\boldsymbol{X})$ with the training data
2. Next, regress the estimated $\hat{m}_a(\boldsymbol{X})$ on $\boldsymbol{V}$, yielding $\hat{\mu}_a(\boldsymbol{V}) = \hat{\mathbb{E}}(\hat{m}_a(\boldsymbol{X})|\boldsymbol{V})$

Approach is motivated by the observation that

$$\begin{aligned}\mathbb{E}(Y(a)|\boldsymbol{V}) \;&=\; \mathbb{E}[\,\mathbb{E}(Y(a)|\underbrace{\boldsymbol{V},\boldsymbol{U}}_{\boldsymbol{X}})\,|\boldsymbol{V}] \;=\; \mathbb{E}(\;\mathbb{E}(Y|A=a,\boldsymbol{X})\;\;|\boldsymbol{V}) \\ &= \mathbb{E}(m_a(\boldsymbol{X})\mid\boldsymbol{V})\end{aligned}$$

Interval construction relies on estimation of $r_{a,\alpha}$

Suppose we've fit a prediction model $\hat{\mu}_a(\boldsymbol{V})$ for $Y(a)$, let $R_a = |Y(a) - \hat{\mu}_a(\boldsymbol{V})|$, and recall our interest in the equation

$$\mathbb{P}(R_a(\boldsymbol{V}_i, Y_i) \leq r_{a,\alpha} | S = 0) = 1 - \alpha$$

Interval construction relies on estimation of $r_{a,\alpha}$

Suppose we've fit a prediction model $\hat{\mu}_a(\boldsymbol{V})$ for $Y(a)$, let $R_a = |Y(a) - \hat{\mu}_a(\boldsymbol{V})|$, and recall our interest in the equation

$$\mathbb{P}(R_a(\boldsymbol{V}_i, Y_i) \leq r_{a,\alpha} | S = 0) = 1 - \alpha$$

Notice $r_{a,\alpha}$ is the key quantity above – if we knew it, we could construct valid intervals for $Y(a)$ in the target population of the form $\hat{\mu}_a(\boldsymbol{V}) \pm r_{a,\alpha}$,

Interval construction relies on estimation of $r_{a,\alpha}$

Suppose we've fit a prediction model $\hat{\mu}_a(\boldsymbol{V})$ for $Y(a)$, let $R_a = |Y(a) - \hat{\mu}_a(\boldsymbol{V})|$, and recall our interest in the equation

$$\mathbb{P}(R_a(\boldsymbol{V}_i, Y_i) \leq r_{a,\alpha} | S = 0) = 1 - \alpha$$

Notice $r_{a,\alpha}$ is the key quantity above – if we knew it, we could construct valid intervals for $Y(a)$ in the target population of the form $\hat{\mu}_a(\boldsymbol{V}) \pm r_{a,\alpha}$,

- Our paper focuses on estimation of $r_{a,\alpha}$, with particular attention given to semiparametric efficient estimation, following Yang et al. (2024)

IPW identification of $r_{a,\alpha}$

Continuing to let $r_{a,\alpha}$ be the quantity which satisfies

$$\mathbb{P}(R(\boldsymbol{V}_i, Y_i(a)) \leq r_{a,\alpha} | S = 0) = 1 - \alpha$$

It can be shown that this same quantity *also* satisfies

$$\mathbb{E}\left[\underbrace{\frac{\mathbb{P}(S = 0 \mid \boldsymbol{V})}{\mathbb{P}(A = a \mid \boldsymbol{X}, S = 1)\mathbb{P}(S = 1 \mid \boldsymbol{V})}}_{\text{weights}} \underbrace{S\mathbb{I}(A = a)\mathbb{I}\{R_a(\boldsymbol{V}, Y \leq r_{a,\alpha}\}}_{\text{source pop. conformity scores}} \right] = 1 - \alpha$$

IPW identification of $r_{a,\alpha}$

Continuing to let $r_{a,\alpha}$ be the quantity which satisfies

$$\mathbb{P}(R(\boldsymbol{V}_i, Y_i(a)) \leq r_{a,\alpha} | S = 0) = 1 - \alpha$$

It can be shown that this same quantity *also* satisfies

$$\mathbb{E}\left[\underbrace{\frac{\mathbb{P}(S = 0 \mid \boldsymbol{V})}{\mathbb{P}(A = a \mid \boldsymbol{X}, S = 1)\mathbb{P}(S = 1 \mid \boldsymbol{V})}}_{\text{weights}} \underbrace{S\mathbb{I}(A = a)\mathbb{I}\{R_a(\boldsymbol{V}, Y \leq r_{a,\alpha}\}}_{\text{source pop. conformity scores}} \right] = 1 - \alpha$$

Intuition: (1) Take the conformity scores that you compute among those with $A = a$ in the source pop, and (2) find the $1 - \alpha$ quantile of its reweighted distribution

Regression-based identification of $r_{a,\alpha}$

Additionally, it can be shown $r_{a,\alpha}$ satisfies

$$\mathbb{E}[m_a(r_{a,\alpha}, \boldsymbol{V})|S = 0] = 1 - \alpha,$$

where for generic r,

$$q_a(r, \boldsymbol{X}) := \mathbb{P}(R_a(\boldsymbol{V}, Y) \leq r|\boldsymbol{X}, A = a, S = 1),$$
$$m_a(r, \boldsymbol{V}) := \mathbb{E}[\ q_a(r, \boldsymbol{X})|\boldsymbol{V}, S = 1)\ ].$$

Regression-based identification of $r_{a,\alpha}$

Additionally, it can be shown $r_{a,\alpha}$ satisfies

$$\mathbb{E}[m_a(r_{a,\alpha}, \boldsymbol{V})|S = 0] = 1 - \alpha,$$

where for generic r,

$$q_a(r, \boldsymbol{X}) := \mathbb{P}(R_a(\boldsymbol{V}, Y) \leq r|\boldsymbol{X}, A = a, S = 1),$$
$$m_a(r, \boldsymbol{V}) := \mathbb{E}[\ q_a(r, \boldsymbol{X})|\boldsymbol{V}, S = 1)\ ].$$

Intuition:

1. For an arbitrary r, regress the $\boldsymbol{X}$-conditional conformity score source CDF $q_a(r, \boldsymbol{X})$...
2. ...on $\boldsymbol{V}$ in the source population, yielding $m_a(r, \boldsymbol{V})$

$r_{a,\alpha}$ satisfies $\mathbb{E}[m_a(r_{a,\alpha}, \boldsymbol{V})|S = 0] = 1 - \alpha$

Stepping back

- Remember that our goal is to construct intervals for $Y(a)$ in the target population so that

$$\mathbb{P}(Y(a) \text{ is in the interval } | \underbrace{S = 0}_{\text{in the target pop.}}) = 1 - \alpha$$

Stepping back

- Remember that our goal is to construct intervals for $Y(a)$ in the target population so that

$$\mathbb{P}(Y(a) \text{ is in the interval} \mid \underbrace{S = 0}_{\text{in the target pop.}}) = 1 - \alpha$$

- Conformal prediction gives us a blueprint for constructing these intervals, while being agnostic to how we predict $Y(a)$

Stepping back

- Remember that our goal is to construct intervals for $Y(a)$ in the target population so that

$$\mathbb{P}(Y(a) \text{ is in the interval} \mid \underbrace{S = 0}_{\text{in the target pop.}}) = 1 - \alpha$$

- Conformal prediction gives us a blueprint for constructing these intervals, while being agnostic to how we predict $Y(a)$

- Once we have a prediction model for $Y(a)$ (takes in covariates $\boldsymbol{V} \subset \boldsymbol{X}$, outputs $\hat{Y}(a)$), constructing valid intervals amounts to finding $r_{a,\alpha}$ satisfying

$$\mathbb{P}(\underbrace{R(\boldsymbol{V}, Y(a))}_{\text{scores}} \leq r_{a,\alpha} | S = 0) = 1 - \alpha$$

Stepping back

- Remember that our goal is to construct intervals for $Y(a)$ in the target population so that

$$\mathbb{P}(Y(a) \text{ is in the interval} \mid \underbrace{S = 0}_{\text{in the target pop.}}) = 1 - \alpha$$

- Conformal prediction gives us a blueprint for constructing these intervals, while being agnostic to how we predict $Y(a)$

- Once we have a prediction model for $Y(a)$ (takes in covariates $\boldsymbol{V} \subset \boldsymbol{X}$, outputs $\hat{Y}(a)$), constructing valid intervals amounts to finding $r_{a,\alpha}$ satisfying

$$\mathbb{P}(\underbrace{R(\boldsymbol{V}, Y(a))}_{\text{scores}} \leq r_{a,\alpha} | S = 0) = 1 - \alpha$$

- So our goal boils down to estimating $r_{a,\alpha}$ well

Slow convergence of the IPW plug-in estimator

The earlier weighting expression suggests we can estimate $r_{a,\alpha}$ via the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\mathbb{P}}(S_i=0|\boldsymbol{V}_i)}{\hat{\mathbb{P}}(A_i=a|\boldsymbol{X}_i,S_i=1)\hat{\mathbb{P}}(S_i=1|\boldsymbol{V}_i)}\underbrace{S_i\mathbb{I}(A_i=a)\mathbb{I}\{R_a(Y_i,\boldsymbol{V}_i)\leq r_{a,\alpha}\}}_{\text{source pop scores}}\right)-(1-\alpha)=0$$

Slow convergence of the IPW plug-in estimator

The earlier weighting expression suggests we can estimate $r_{a,\alpha}$ via the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\mathbb{P}}(S_i=0|\boldsymbol{V}_i)}{\hat{\mathbb{P}}(A_i=a|\boldsymbol{X}_i,S_i=1)\hat{\mathbb{P}}(S_i=1|\boldsymbol{V}_i)}\underbrace{S_i\mathbb{I}(A_i=a)\mathbb{I}\{R_a(Y_i,\boldsymbol{V}_i)\leq r_{a,\alpha}\}}_{\text{source pop scores}}\right)-(1-\alpha)=0$$

But this suggests the rate at which we estimate $r_{a,\alpha}$ will be dictated by

$$\hat{r}_{a,\alpha}-r_{a,\alpha}=O_{\mathbb{P}}(1/\sqrt{n}+||\hat{g}_a-g_a||+||\hat{\kappa}-\kappa||)$$

where $\hat{g}_a(\boldsymbol{X})=\hat{\mathbb{P}}(A=a|S=1,\boldsymbol{X})$ and $\hat{\kappa}(\boldsymbol{V})=\hat{\mathbb{P}}(S=1|\boldsymbol{V})$

Slow convergence of the IPW plug-in estimator

The earlier weighting expression suggests we can estimate $r_{a,\alpha}$ via the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\mathbb{P}}(S_i=0|\boldsymbol{V}_i)}{\hat{\mathbb{P}}(A_i=a|\boldsymbol{X}_i,S_i=1)\hat{\mathbb{P}}(S_i=1|\boldsymbol{V}_i)}\underbrace{S_i\mathbb{I}(A_i=a)\mathbb{I}\{R_a(Y_i,\boldsymbol{V}_i)\leq r_{a,\alpha}\}}_{\text{source pop scores}}\right)-(1-\alpha)=0$$

But this suggests the rate at which we estimate $r_{a,\alpha}$ will be dictated by

$$\hat{r}_{a,\alpha}-r_{a,\alpha}=O_{\mathbb{P}}(1/\sqrt{n}+||\hat{g}_a-g_a||+||\hat{\kappa}-\kappa||)$$

where $\hat{g}_a(\boldsymbol{X})=\hat{\mathbb{P}}(A=a|S=1,\boldsymbol{X})$ and $\hat{\kappa}(\boldsymbol{V})=\hat{\mathbb{P}}(S=1|\boldsymbol{V})$

If we use flexible models, this can be notably slower than ideal $\sqrt{n}$ rates

Debiased estimator

Much like in ATE estimation, we can improve estimation of $r_{a,\alpha}$ by basing estimation on its EIF. Letting η collect all nuisance functions, we can choose $\hat{r}_{a,\alpha}$ so

$$\frac{1}{n}\sum_{i=1}^{n} \mathsf{EIF}(\hat{r}_{a,\alpha},\ \mathsf{data}_i,\ \hat{\eta}(\hat{r}_{a,\alpha})) = 0$$

Debiased estimator

Much like in ATE estimation, we can improve estimation of $r_{a,\alpha}$ by basing estimation on its EIF. Letting η collect all nuisance functions, we can choose $\hat{r}_{a,\alpha}$ so

$$\frac{1}{n}\sum_{i=1}^{n} \mathsf{EIF}(\hat{r}_{a,\alpha},\ \mathsf{data}_i,\ \hat{\eta}(\hat{r}_{a,\alpha})) = 0$$

Why? The population-level EIF for the true $r_{a,\alpha}$ is mean zero

Debiased estimator

Much like in ATE estimation, we can improve estimation of $r_{a,\alpha}$ by basing estimation on its EIF. Letting η collect all nuisance functions, we can choose $\hat{r}_{a,\alpha}$ so

$$\frac{1}{n}\sum_{i=1}^{n} \mathsf{EIF}(\hat{r}_{a,\alpha},\ \mathsf{data}_i,\ \hat{\eta}(\hat{r}_{a,\alpha})) = 0$$

Why? The population-level EIF for the true $r_{a,\alpha}$ is mean zero

$$\begin{aligned}\mathsf{EIF}(r_{a,\alpha},\ \mathsf{data}_i,\ \eta(r_{a,\alpha})) &= (1-S)\cdot\{m(r_{a,\alpha},\boldsymbol{X}) - (1-\alpha)\} \\ &+ \frac{(1-\kappa(\boldsymbol{V}))}{\kappa(\boldsymbol{V})} S\cdot\{q_a(r_{a,\alpha},\boldsymbol{X}) - m_a(r_{a,\alpha},\boldsymbol{V})\} \\ &+ \frac{(1-\kappa(\boldsymbol{V}))}{g_a(\boldsymbol{X})\kappa(\boldsymbol{V})}\mathbb{I}(A=a)S\cdot\{\mathbb{I}(R_a(Y,\boldsymbol{V})\le r_{a,\alpha}) - q_a(r_{a,\alpha},\boldsymbol{X})\}\end{aligned}$$

Debiased estimator

Much like in ATE estimation, we can improve estimation of $r_{a,\alpha}$ by basing estimation on its EIF. Letting η collect all nuisance functions, we can choose $\hat{r}_{a,\alpha}$ so

$$\frac{1}{n}\sum_{i=1}^{n} \mathsf{EIF}(\hat{r}_{a,\alpha},\ \mathsf{data}_i,\ \hat{\eta}(\hat{r}_{a,\alpha})) = 0$$

Why? The population-level EIF for the true $r_{a,\alpha}$ is mean zero

$$\begin{aligned}\mathsf{EIF}(r_{a,\alpha},\ \mathsf{data}_i,\ \eta(r_{a,\alpha})) &= (1-S)\cdot\{m(r_{a,\alpha},\boldsymbol{X}) - (1-\alpha)\} \\ &+ \frac{(1-\kappa(\boldsymbol{V}))}{\kappa(\boldsymbol{V})} S\cdot\{q_a(r_{a,\alpha},\boldsymbol{X}) - m_a(r_{a,\alpha},\boldsymbol{V})\} \\ &+ \frac{(1-\kappa(\boldsymbol{V}))}{g_a(\boldsymbol{X})\kappa(\boldsymbol{V})}\mathbb{I}(A=a)S\cdot\{\mathbb{I}(R_a(Y,\boldsymbol{V})\le r_{a,\alpha}) - q_a(r_{a,\alpha},\boldsymbol{X})\}\end{aligned}$$

Challenge: Solving above estimating equation requires repeatedly estimating $q_a(r,\boldsymbol{X})$ and $m_a(r,\boldsymbol{V})$ at different r

Debiased estimator

Much like in ATE estimation, we can improve estimation of $r_{a,\alpha}$ by basing estimation on its EIF. Letting η collect all nuisance functions, we can choose $\hat{r}_{a,\alpha}$ so

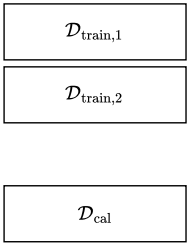
$$\frac{1}{n}\sum_{i=1}^{n}\mathsf{EIF}(\hat{r}_{a,\alpha},\ \mathsf{data}_i,\ \hat{\eta}(\hat{r}_{a,\alpha}))=0$$

Why? The population-level EIF for the true $r_{a,\alpha}$ is mean zero

$$\begin{aligned}\mathsf{EIF}(r_{a,\alpha},\ \mathsf{data}_i,\ \eta(r_{a,\alpha})) &= (1-S)\cdot\{m(r_{a,\alpha},\boldsymbol{X})-(1-\alpha)\}\\ &+\frac{(1-\kappa(\boldsymbol{V}))}{\kappa(\boldsymbol{V})}S\cdot\{q_a(r_{a,\alpha},\boldsymbol{X})-m_a(r_{a,\alpha},\boldsymbol{V})\}\\ &+\frac{(1-\kappa(\boldsymbol{V}))}{g_a(\boldsymbol{X})\kappa(\boldsymbol{V})}\mathbb{I}(A=a)S\cdot\{\mathbb{I}(R_a(Y,\boldsymbol{V})\le r_{a,\alpha})-q_a(r_{a,\alpha},\boldsymbol{X})\}\end{aligned}$$

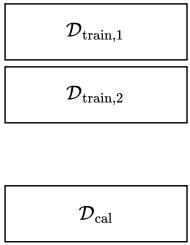
Can avoid estimating $q_a(r,\boldsymbol{X})$ and $m_a(r,\boldsymbol{V})$ at different r by adopting the localized DML framework from Kallus et al. (2024)

Localized DML Implementation



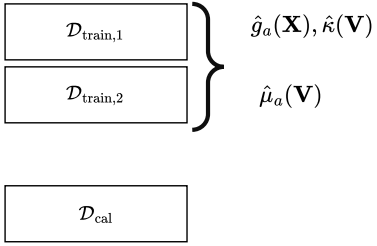
To avoid estimating $q_a(r, \boldsymbol{X})$ and $m_a(r, \boldsymbol{V})$ at different r, we adopt the localized DML framework from Kallus et al. (2024)

Localized DML Implementation



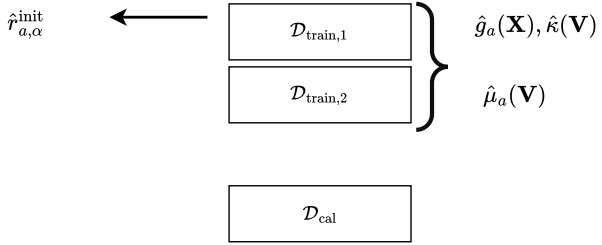
Idea: split data intro training and calibration folds

Localized DML Implementation



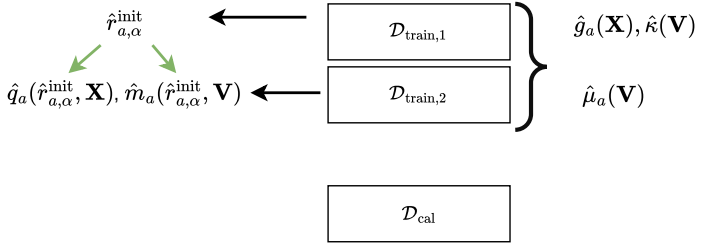
Using all of the training data, fit nuisances that don't depend on r

Localized DML Implementation



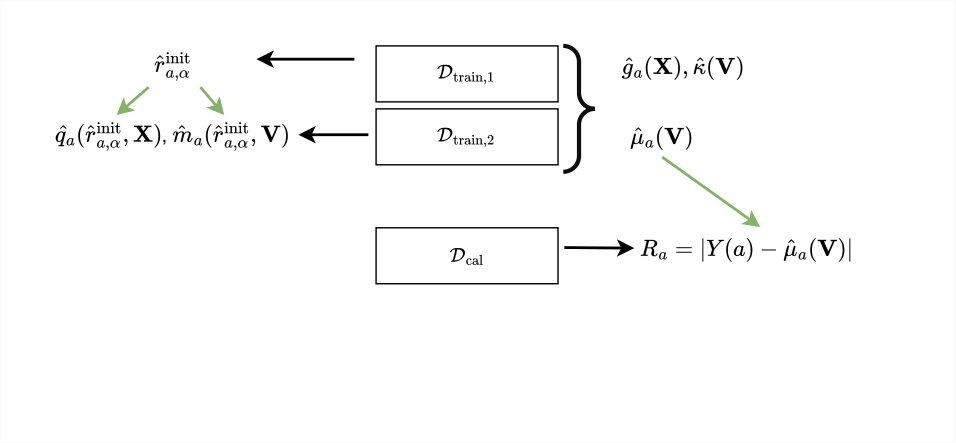
Obtain an initial estimate $\hat{r}^{\text{init}}_{a,\alpha}$ on first fold of the training data, $\mathcal{D}_{\text{train},1}$

Localized DML Implementation



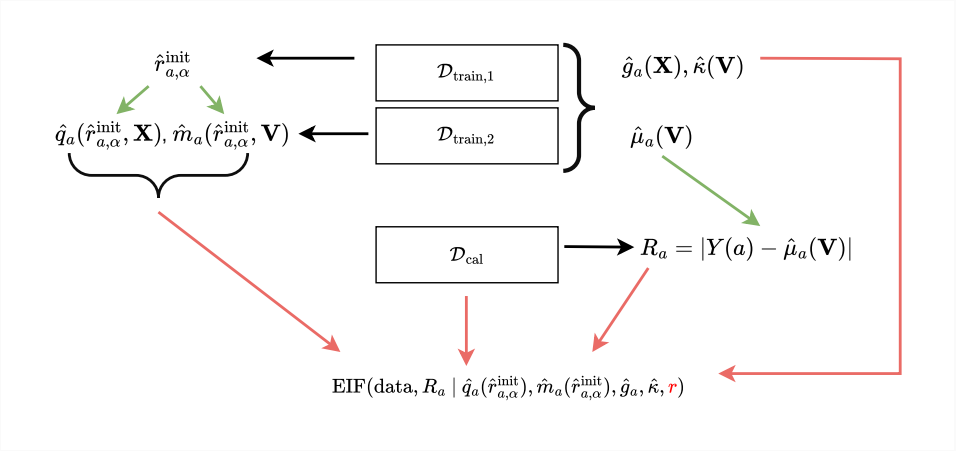
Train $\hat{q}_a(r, \boldsymbol{X})$ and $\hat{m}_a(r, \boldsymbol{V})$ in other training fold, $\mathcal{D}_{\text{train},2}$ at $\hat{r}^{\text{init}}_{a,\alpha}$

Localized DML Implementation



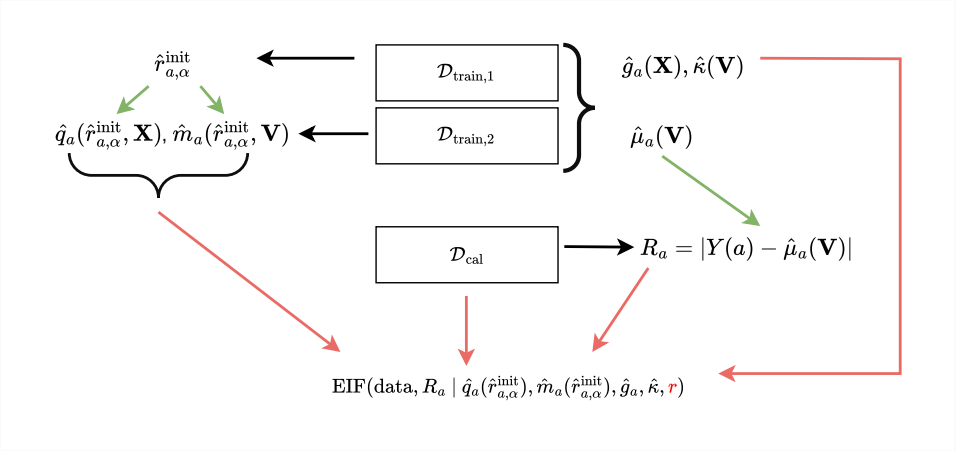
Using $\hat{\mu}_a(\boldsymbol{V})$ from training folds, construct conformity scores in the calibration set

Localized DML Implementation



Construct $\text{EIF}(r,\ \text{data}_i \mid \eta(r_{a,\alpha}))$ for all calibration set units

Localized DML Implementation



Solve $\sum_{i \in \mathcal{D}_{\text{cal}}} \text{EIF}(r,\ \text{data}_i \mid \hat{\eta}(\hat{r}^{\text{init}}_{a,\alpha})) = 0$

Coverage result

Theorem When $\hat{C}_a(\boldsymbol{V})$ is constructed according to our proposed DML algorithm, then

$$\mathbb{P}(Y(a) \in \hat{C}_a(\boldsymbol{V})|S=0) = 1-\alpha+O_{\mathbb{P}}(1/\sqrt{n} \ + \ R_n), \text{ where}$$

$$R_n = \sup_r ||\hat{q}_a(r,\cdot) - q_a(r,\cdot)|| \cdot ||\hat{g}_a - g_a|| + \sup_r ||\hat{m}_a(r,\cdot) - m_a(r,\cdot)|| \cdot ||\hat{\kappa} - \kappa||.$$

Coverage result

Theorem When $\hat{C}_a(\boldsymbol{V})$ is constructed according to our proposed DML algorithm, then

$$\mathbb{P}(Y(a) \in \hat{C}_a(\boldsymbol{V})|S=0) = 1-\alpha+O_{\mathbb{P}}(1/\sqrt{n} \ + \ R_n), \text{ where}$$

$$R_n = \sup_r ||\hat{q}_a(r,\cdot) - q_a(r,\cdot)|| \cdot ||\hat{g}_a - g_a|| + \sup_r ||\hat{m}_a(r,\cdot) - m_a(r,\cdot)|| \cdot ||\hat{\kappa} - \kappa||.$$

Bias decomposition implies coverage bias will shrink at ideal $n^{-1/2}$ rates so long as

1. $\sup_r ||\hat{q}_a(r,\cdot) - q_a(r,\cdot)|| \cdot ||\hat{g}_a - g_a|| = o_{\mathbb{P}}(1/\sqrt{n})$, and
2. $\sup_r ||\hat{m}_a(r,\cdot) - m_a(r,\cdot)|| \cdot ||\hat{\kappa} - \kappa|| = o_{\mathbb{P}}(1/\sqrt{n})$

Attainable if all nuisances are estimated at, e.g. $n^{-1/4}$ rates

Roadmap

Snapshot of synthetic data experiments

Compared our proposed method to

- Weighted conformal prediction (Tibshirani et al., 2019) based on earlier IPW expression
- A DML method based on Yang et al. (2024) that *ignores* runtime confounding, pretending $\boldsymbol{V}$ collects all confounders

Snapshot of synthetic data experiments

Compared our proposed method to

- Weighted conformal prediction (Tibshirani et al., 2019) based on earlier IPW expression
- A DML method based on Yang et al. (2024) that *ignores* runtime confounding, pretending $\boldsymbol{V}$ collects all confounders

Simulated data corresponding to a runtime confounding scenario, varying the sample size

Snapshot of synthetic data experiments

Compared our proposed method to

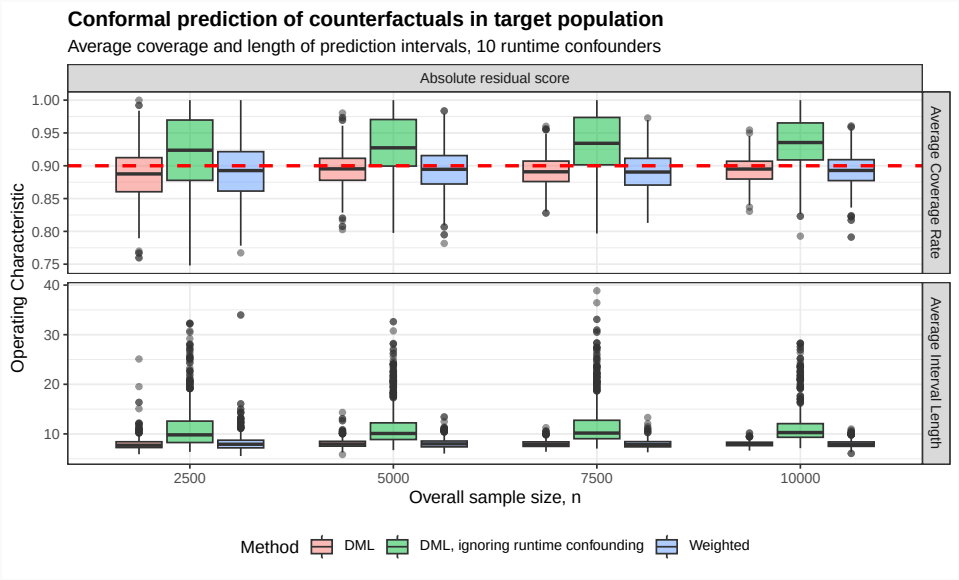
- Weighted conformal prediction (Tibshirani et al., 2019) based on earlier IPW expression
- A DML method based on Yang et al. (2024) that *ignores* runtime confounding, pretending $\boldsymbol{V}$ collects all confounders

Simulated data corresponding to a runtime confounding scenario, varying the sample size

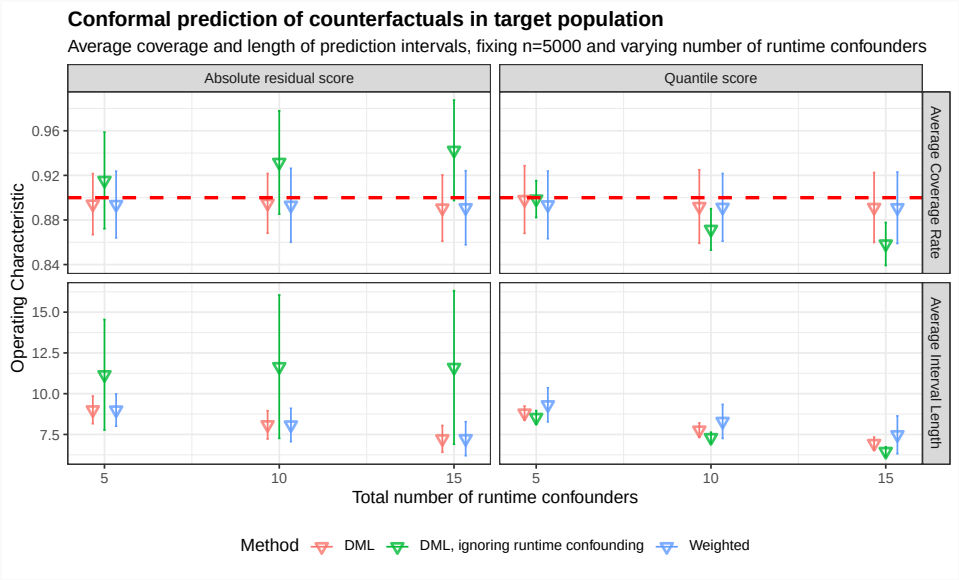
For simplicity, focused on producing 90% prediction intervals for $Y(1)$ and $Y(0)$ in target data. Across 1,000 iterations for each sample size considered, we compared

- The pooled average coverage rate of $Y(1)$ and $Y(0)$ in target data
- The pooled average interval width

Snapshot of synthetic experiments: varying n



Snapshot of synthetic experiments: varying number of runtime confounders



Roadmap

Discussion

Runtime confounding is a common challenge that can complicate the construction of valid prediction intervals for counterfactual outcomes

- Our proposed method enables construction of asymptotically valid prediction intervals for counterfactual outcomes under runtime confounding
- Theory + experiments demonstrate ignoring runtime confounding can result in intervals which considerably miscover

Many avenues for future work, including

- Methods which address instances of poor overlap between the source and target populations
- Sensitivity analyses to partially relax source independence assumption

Thank you!

https://kbarnatchez.com

Link to paper

References

Bhattacharyya, A. and Barber, R. F. (2026). Group-weighted conformal prediction. *Electronic Journal of Statistics*, 20(1):1171–1199.

Coston, A., Kennedy, E., and Chouldechova, A. (2020). Counterfactual predictions under runtime confounding. *Advances in neural information processing systems*, 33:4150–4162.

Kallus, N., Mao, X., and Uehara, M. (2024). Localized debiased machine learning: Efficient inference on quantile treatment effects and beyond. *Journal of Machine Learning Research*, 25(16):1–59.

Tibshirani, R. J., Foygel Barber, R., Candes, E., and Ramdas, A. (2019). Conformal prediction under covariate shift. *Advances in neural information processing systems*, 32.

Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic learning in a random world*, volume 29. Springer.

Yang, Y., Kuchibhotla, A. K., and Tchetgen Tchetgen, E. (2024). Doubly robust calibration of prediction sets under covariate shift. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkae009.

Appendix

Back to start

Roadmap

- Theorems
  - Identification of $r_{a,\alpha}$
  - EIF for $r_{a,\alpha}$
  - Robustness Theorem
- Assumptions
- Counterfactual Quantile Regression

Identification of $r_{a,\alpha}$

For compactness, let $R_a := R_a(Y, \boldsymbol{V})$ and notice

$$\begin{aligned}
\mathbb{P}(R_a \leq r_{a,\alpha} | S = 0) &= \mathbb{E}(\mathbb{I}(R_a \leq r_{a,\alpha}) | S = 0) \\
&= \mathbb{E}[\mathbb{E}(\mathbb{I}(R_a \leq r_{a,\alpha}) | \boldsymbol{X}, S = 0) | S = 0] \\
&= \mathbb{E}[\mathbb{E}(\mathbb{I}(R_a \leq r_{a,\alpha}) | \boldsymbol{X}, S = 1) | S = 0] \\
&= \mathbb{E}\{\mathbb{E}[\mathbb{E}(\mathbb{I}(R_a \leq r_{a,\alpha}) | \boldsymbol{X}, A = a, S = 1) | \boldsymbol{V}, S = 1] | S = 0\} \\
&= \mathbb{E}\{\mathbb{E}[q_a(r_{a,\alpha}, \boldsymbol{X}) | \boldsymbol{V}, S = 1] | S = 0\} \\
&= \mathbb{E}\{m_a(r, \boldsymbol{V}) | S = 0\}
\end{aligned}$$

EIF for $r_{a,\alpha}$

Theorem

Let $\eta_a(r) := (q_a(r), m_a(r), g_a, \kappa)$ and suppose $\boldsymbol{O} \sim \mathbb{P}$. The efficient influence curve for $r_{a,\alpha}$ in a nonparametric model for $\mathbb{P}$ is proportional to

$$\begin{aligned}
&\chi_a(r_{a,\alpha}, \boldsymbol{O}\,; \eta_a(r_{a,\alpha})) := \\
&(1-S)(m_a(r_{a,\alpha}, \boldsymbol{V}) - (1-\alpha)) + \frac{S(1-\kappa(\boldsymbol{V}))}{\kappa(\boldsymbol{V})}\{q_a(r_{a,\alpha}, \boldsymbol{X}) - m_a(r_{a,\alpha}, \boldsymbol{V})\} \\
&+ w_a(\boldsymbol{O})\{(R_a(Y, \boldsymbol{V}) \leq r_{a,\alpha}) - q_a(r_{a,\alpha}, \boldsymbol{X})\},
\end{aligned} \tag{1}$$

where $q_a(r, \boldsymbol{X})) = \mathbb{P}(R_a \leq r | \boldsymbol{X}, A = a, S = 1)$, $m_a(r, \boldsymbol{X}) = \mathbb{E}[q_a(r, \boldsymbol{X}) | \boldsymbol{V}, S = 1]$, $\kappa(\boldsymbol{V}) = \mathbb{P}(S = 1 | \boldsymbol{V})$, and $g_a(\boldsymbol{X}) = \mathbb{P}(A = a | \boldsymbol{X}, S = 1)$

Robustness Theorem

When $\hat{C}_a(\boldsymbol{V})$ is constructed according to our proposed debiased machine learning algorithm, then

$$\mathbb{P}(Y(a) \in \hat{C}_a(\boldsymbol{V})|S=0) = 1-\alpha+O_{\mathbb{P}}(1/\sqrt{n} \ + \ R_n), \text{ where}$$

$$R_n = \sup_r ||\hat{q}_a(r,\cdot) - q_a(r,\cdot)|| \cdot ||\hat{g}_a - g_a|| + \sup_r ||\hat{m}_a(r,\cdot) - m_a(r,\cdot)|| \cdot ||\hat{\kappa} - \kappa||.$$

Roadmap

- Theorems
  - Identification of $r_{a,\alpha}$
  - EIF for $r_{a,\alpha}$
  - Robustness Theorem
- Assumptions
- Counterfactual Quantile Regression

Assumptions

Assumption (Positivity)

$0 < \mathbb{P}(A = a|\boldsymbol{X} = \boldsymbol{x}) < 1$ *for all* $\boldsymbol{x}$ *with positive support*

Assumption (Consistency)

$Y = I(A = a) \cdot Y(a)$

Assumption (Unconfounedness)

$Y(a) \perp\!\!\!\perp A|\boldsymbol{X}$

Assumption (Source exchangeability)

$Y(a) \perp\!\!\!\perp S|\boldsymbol{V}$

Assumption (Source positivity)

$0 < \mathbb{P}(S = 1|\boldsymbol{V} = \boldsymbol{v}) < 1$ *for all* $\boldsymbol{v}$ *with positive support*

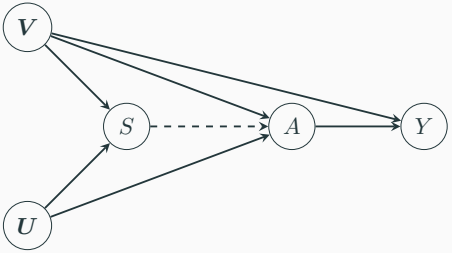
Revisiting Assumption 4

Can be difficult to assess when Assumption 4, $Y(a) \perp\!\!\!\perp S|\boldsymbol{V}$, will hold

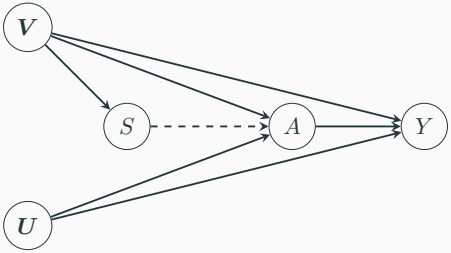
Often easier to instead think through the following two Assumptions, which imply Assumption 4

- $Y(a) \perp\!\!\!\perp S|\boldsymbol{X}$
- $\boldsymbol{U} \perp\!\!\!\perp S|\boldsymbol{V}$

Some DAGs consistent with independence assumptions

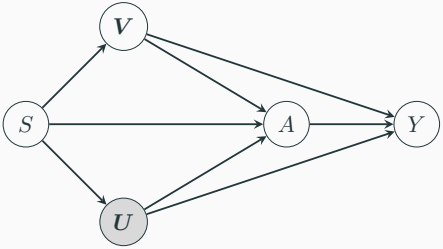


(a)

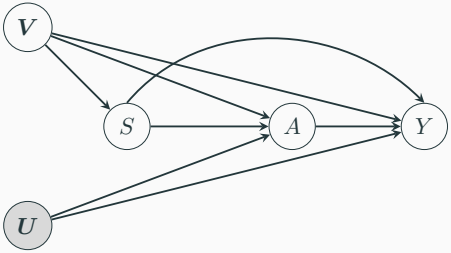


(b)

DAGs implying Assumption 4 is violated



(a)



(b)

Roadmap

Theorems

- Identification of $r_{a,\alpha}$
- EIF for $r_{a,\alpha}$
- Robustness Theorem

Assumptions

Counterfactual Quantile Regression

Quantile Regression

Earlier, we discussed simple way for constructing intervals:

1. First, fit $\hat{m}_a(\boldsymbol{X}) = \hat{\mathbb{E}}(Y|A = a, \boldsymbol{X})$ with the training data
2. Next, regress the estimated $\hat{m}_a(\boldsymbol{X})$ on $\boldsymbol{V}$, yielding $\hat{\mu}_a(\boldsymbol{V}) = \hat{\mathbb{E}}(\hat{m}_a(\boldsymbol{X})|\boldsymbol{V})$

Drawback of this approach is that interval will be **constant width** – won't adapt to local variability in $Y(a)$ around values of $\boldsymbol{V}$

Common approach in conformal prediction is to instead construct intervals based on a quantile conformity score

Quick aside: forming counterfactual predictions

Another common approach to form prediction intervals is to

- Predict $Q_{a,\alpha/2}(\boldsymbol{V}) := \text{Quantile}_{\alpha/2}(Y|\boldsymbol{V})$ and $Q_{a,1-\alpha/2}(\boldsymbol{V}) := \text{Quantile}_{1-\alpha/2}(Y|\boldsymbol{V})$
- Use tools from conformal to adjust naive intervals $(\hat{Q}_{a,\alpha/2}(\boldsymbol{V}), \hat{Q}_{a,1-\alpha/2}(\boldsymbol{V}))$

Challenge: $\mathbb{E}(Q_{a,\alpha}(\boldsymbol{X})|\boldsymbol{V}) \neq \text{Quantile}_{\alpha/2}(Y(a)|\boldsymbol{V})$ due to nonlinearity of the quantile function

- But we need to use all of $\boldsymbol{X}$ since $\boldsymbol{V}$ doesn't contain all confounders

In our paper, we derive a valid loss function for estimating $Q_{a,\alpha}(\boldsymbol{X})$ under runtime confounding settings

Quantile weighted loss function

If $Q_{a,\alpha}(\boldsymbol{v})$ satisfies $\mathbb{P}(Y(a) \leq Q_{a,\alpha}(\boldsymbol{v})|\boldsymbol{V} = \boldsymbol{v}, S = 0) = 1 - \alpha$ for all $\boldsymbol{v}$ with positive support, then it additionally satisfies

$$Q_{a,\alpha}(\boldsymbol{V}) = \underset{\tilde{Q}_{a,\alpha}}{\arg\min}\, \mathbb{E}\left[w_a(\boldsymbol{O})\rho_\alpha(Y - \tilde{Q}_{a,\alpha}(\boldsymbol{V}))\right], \tag{2}$$

where $\rho_\alpha(x) = \alpha|x|\mathbb{I}(x \geq 0) + (1-\alpha)|x|\mathbb{I}(x < 0)$ is the pinball loss function and

$$w_a(\boldsymbol{O}) := \frac{\mathbb{I}(A = a)S(1 - \kappa(\boldsymbol{V}))}{g_a(\boldsymbol{X})\kappa(\boldsymbol{V})}.$$