Conformal Inference of Counterfactuals and Individual Treatment Effects

Lihua Lei

Department of Statistics, Stanford University

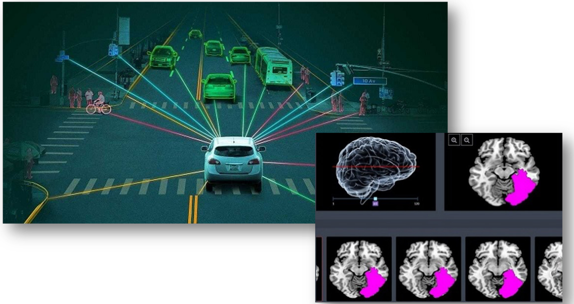
Center for Statistical Methodology, LSHTM, 2021

Collaborator

Emmanuel Candès

ML in critical applications

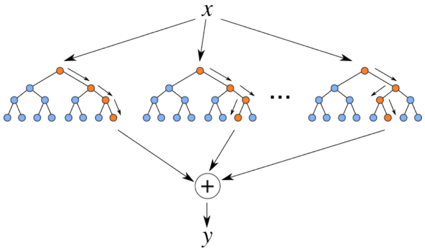
ML tools make potentially high-stakes decisions: self-driving cars, disease diagnosis, ...

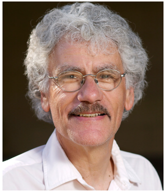


Can we have reliable uncertainty quantification (confidence) in these predictions?

Today's predictive algorithms

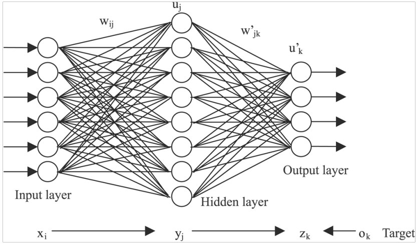
random forests, gradient boosting





Breiman and Friedman

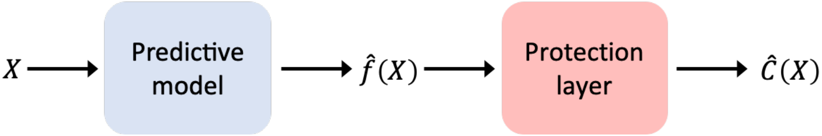
neural networks



LeCun, Hinton and Bengio

A snapshot of conformal inference

- Developed a predictive layer that returns valid prediction intervals

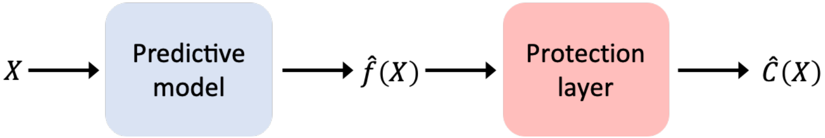


- Training samples $(X_i, Y_i), \; i = 1, \ldots, n$

- Test point $(X, Y = ?)$

A snapshot of conformal inference

- Developed a predictive layer that returns valid prediction intervals



- Training samples (X_i, Y_i), $i = 1, \ldots, n$
- Test point $(X, Y = ?)$
- Conformal inference Vovk et al. '99, Papadopoulos et al. '12, Lei et al. '18, Barber et al. '19, Romano et al. '19

  Constructs predictive interval $\hat{C}(x)$ with $\mathbb{P}\left(Y \in \hat{C}(X)\right) \geq 90\%$

- Holds in finite samples for any distribution of (X, Y) and any predictive algorithm $\hat{f}$

From factuals to counterfactuals

From factuals to counterfactuals

Counterfactual reasoning is ubiquitous in modern science

- Causal inference: what would have been one's response had one taken the treatment
- Offline policy evaluation: what would have been the outcome had the policy changed
- Algorithmic fairness: what would have been the prediction had one belonged to another group
- Explainable machine learning: what would have been the output had the input changed

Agenda

Part I: counterfactual predictive inference

Inference of counterfactuals?

- Potential outcome (PO) framework (Neyman, '23; Rubin, '74)
  - $T \in \{0, 1\}$ binary treatment
  - $Y(1), Y(0)$ potential outcomes
  - X covariates
- Assumptions: super-population (i.i.d.) + SUTVA + unconfoundedness $(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$

Inference of counterfactuals?

- Potential outcome (PO) framework (Neyman, '23; Rubin, '74)
  - $T \in \{0, 1\}$ binary treatment
  - $Y(1), Y(0)$ potential outcomes
  - X covariates
- Assumptions: super-population (i.i.d.) + SUTVA + unconfoundedness $(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$

Find interval estimate $\hat{C}_1(X)$ s.t. $\mathbb{P}(Y(1) \in \hat{C}_1(X) \mid T = 0) \geq 90\%$

Inference of counterfactuals?

- Causal diagram (DAG) framework (Pearl, '95)
  - $T \in \{0, 1\}$ binary treatment
  - Y_1, Y_0 counterfactuals
  - X covariates
- Assumptions: super-population (i.i.d.) $+$ X satisfying the backdoor criterion

Find interval estimate $\hat{C}_1(X)$ s.t. $\mathbb{P}(Y_1 \in \hat{C}_1(X) \mid T = 0) \geq 90\%$

Counterfactual inference

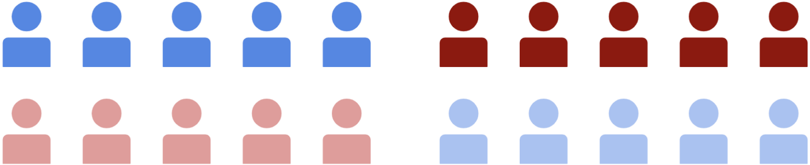
Assign treatment by a coin toss for each subject based on the **propensity score** $e(x)$

$$\mathbb{P}(\text{treated} \mid X = x) = e(x)$$

$$\mathbb{P}(\text{control} \mid X = x) = 1 - e(x)$$

Counterfactual inference

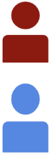
Each subject has potential outcomes $(Y(1), Y(0))$ and the observed outcome Y^{obs}



Observed

Unobserved

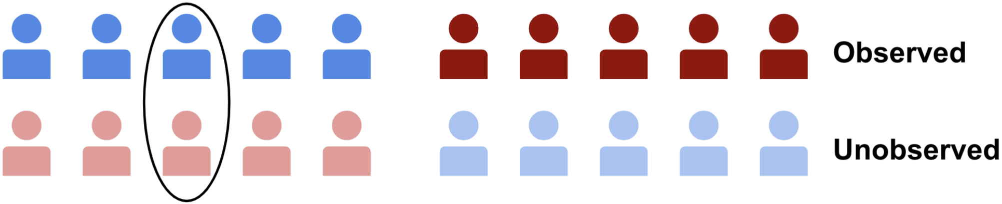
SUTVA



$$Y^{\text{obs}} = Y(1)$$

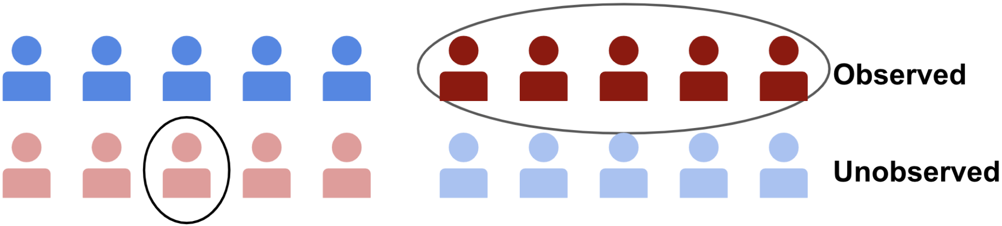
$$Y^{\text{obs}} = Y(0)$$

Counterfactual inference

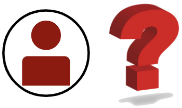


How to infer $Y(1)$ of 

Counterfactual inference

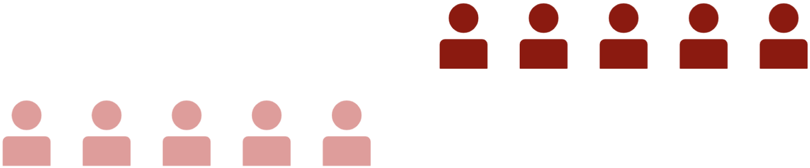


Use observed treated units



The counterfactual inference problem and covariate shift

$$P_{X|T=1} \times P_{Y(1)|X}$$



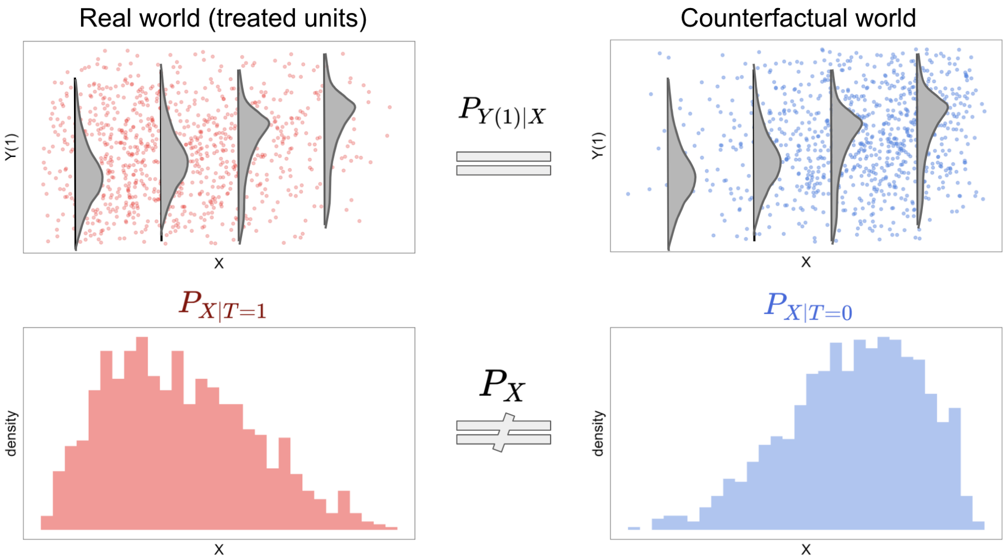
Observed

Unobserved

$$P_{X|T=0} \times P_{Y(1)|X}$$

Distribution mismatch! Covariate shift

The counterfactual inference problem and covariate shift



The counterfactual inference problem and covariate shift

Use i.i.d. samples (observed treated units) from $P_{X|T=1} \times P_{Y(1)|X}$ to construct $\hat{C}_1(X)$ with

$$\mathbb{P}(Y(1) \in \hat{C}_1(X)) \geq 90\% \quad \text{under } P_{X|T=0} \times P_{Y(1)|X}$$

The counterfactual inference problem and covariate shift

Use i.i.d. samples (observed treated units) from $P_{X|T=1} \times P_{Y(1)|X}$ to construct $\hat{C}_1(X)$ with

$$\mathbb{P}(Y(1) \in \hat{C}_1(X)) \geq 90\% \quad \text{under } P_{X|T=0} \times P_{Y(1)|X}$$

Covariate shift $\quad w(x) \triangleq \frac{dP_{X|T=0}}{dP_{X|T=1}}(x) \propto \frac{1 - e(x)}{e(x)}$

Conformal inference under covariate shift

Conformal Inference

Vovk et al. ('99), Papadopoulos et al. ('12), Lei et al. ('18)

$$(X_i, Y_i) \overset{i.i.d.}{\sim} P_X \times P_{Y|X} \Longrightarrow \mathbb{P}_{(X,Y)\sim P_X \times P_{Y|X}}(Y \in \hat{C}(X)) \geq 90\%$$

Conformal inference under covariate shift

Weighted Split Conformalized Quantile Regression (CQR)

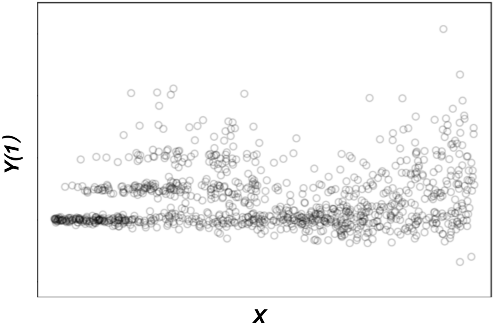
Tibshirani, Barber, Candès, Ramdas ('19); Romano, Patterson, Candès ('19)

$$(X_i, Y_i) \overset{i.i.d.}{\sim} P_X \times P_{Y|X} \Longrightarrow \mathbb{P}_{(X,Y)\sim Q_X \times P_{Y|X}}(Y \in \hat{C}(X)) \geq 90\%$$

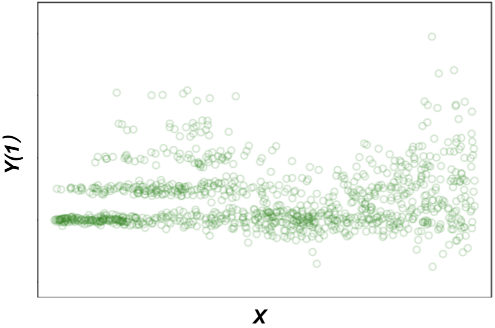
Conformal inference under covariate shift

Weighted Split Conformalized Quantile Regression (CQR)

Randomly split $(X_i, Y_i^{\mathrm{obs}})_{T_i=1}$ into two folds



Proper training set

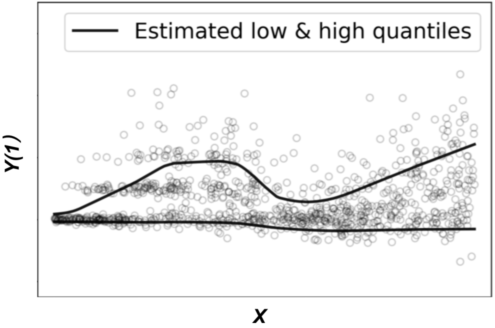


Calibration set

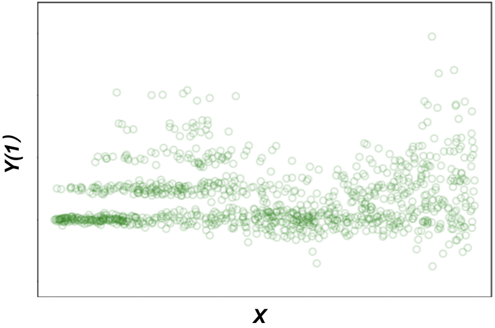
Conformal inference under covariate shift

Weighted Split Conformalized Quantile Regression (CQR)

Fit 5 & 95%-th quantiles of $Y(1) \mid X$ on training fold



Apply quantile regression

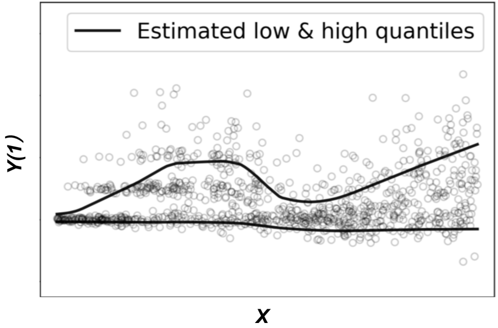


Calibration set

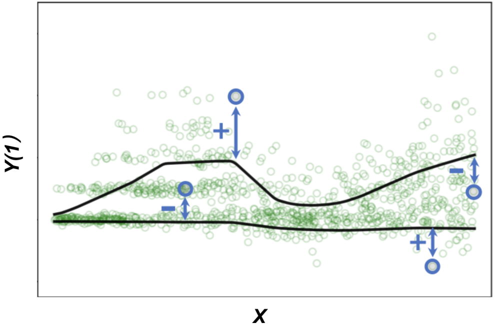
Conformal inference under covariate shift

Weighted Split Conformalized Quantile Regression (CQR)

Estimate 5 & 95%-th quantiles of $Y(1) \mid X$ on calibration fold



Apply quantile regression

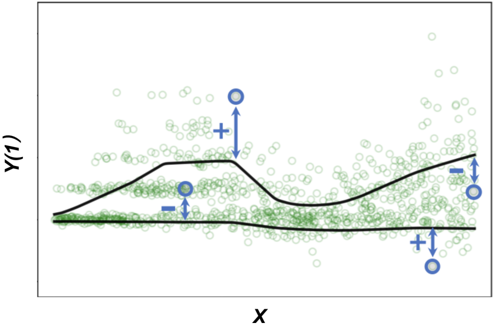


Calibrate

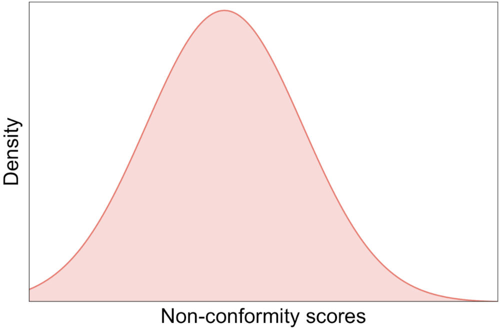
Conformal inference under covariate shift

Weighted Split Conformalized Quantile Regression (CQR)

Signed distance: $V_i \triangleq \max\{\hat{q}_{0.05}(X_i) - Y_i(1), Y_i(1) - \hat{q}_{0.95}(X_i)\}$



Calibrate

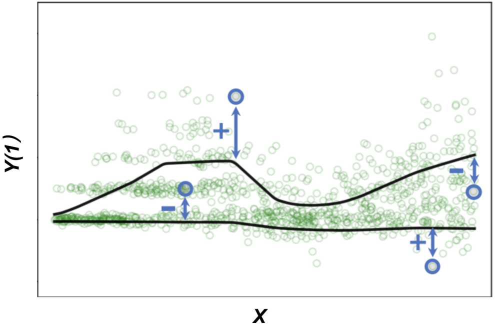


Histogram of signed distances

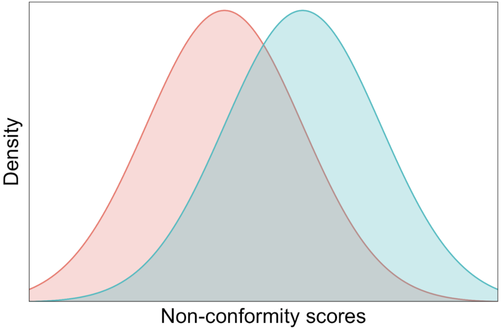
Conformal inference under covariate shift

Weighted Split Conformalized Quantile Regression (CQR)

Weighted dist.: $\sum_{i=1}^{n} p_i(x)\delta_{V_i} + p_\infty(x)\delta_\infty$ where $p_i(x) = w(X_i)/\left(\sum_{i=1}^{n} w(X_i) + w(x)\right)$



Calibrate

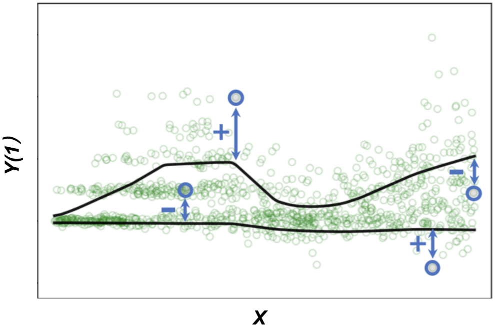


Histogram weighted by $w(x)$

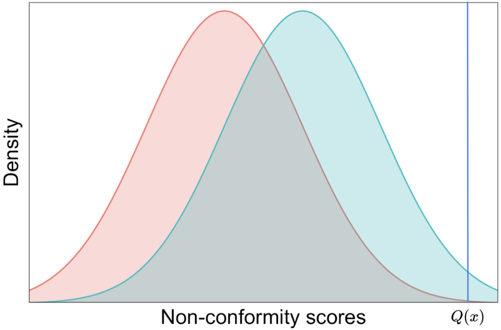
Conformal inference under covariate shift

Weighted Split Conformalized Quantile Regression (CQR)

Cutoff: $Q(x) \triangleq \texttt{Quantile}\left(90\%, \sum_{i=1}^{n} p_i(x)\delta_{V_i} + p_\infty(x)\delta_\infty\right)$



Calibrate

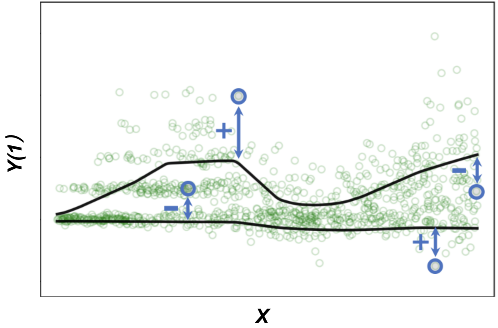


Find the 90%-th quantile $Q(x)$

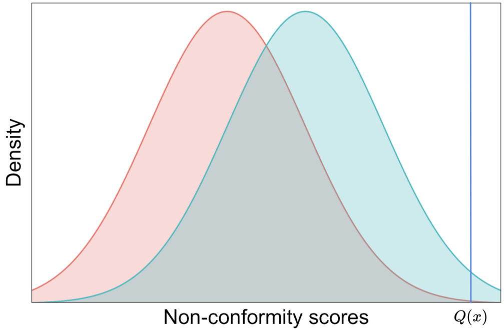
Conformal inference under covariate shift

Weighted Split Conformalized Quantile Regression (CQR)

Interval: $\hat{C}_1(x) = [\hat{q}_{0.05}(x) - Q(x), \hat{q}_{0.95}(x) + Q(x)]$



Calibrate



Find the 90%-th quantile $Q(x)$

Near-exact counterfactual inference in finite samples

Theorem (L. and Candès, 2020, for randomized experiments)

Set $w(x) = (1 - e(x))/e(x)$ *(*$e(x)$ *known) in weighted split-CQR. Then*

$$90\% \le \mathbb{P}(Y(1) \in \hat{C}_1(X) \mid T = 0) \le 90\% + c/n$$

- *Lower bound holds without extra assumption*
- *Upper bound holds if* V_i*'s are a.s. distinct & overlap holds, and* c *only depends on the overlap*

Near-exact counterfactual inference in finite samples

Theorem (L. and Candès, 2020, for randomized experiments)

Set $w(x) = (1 - e(x))/e(x)$ ($e(x)$ known) in weighted split-CQR. Then

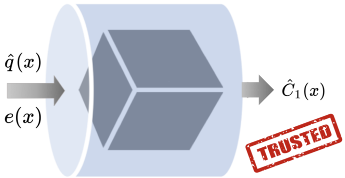
$$90\% \leq \mathbb{P}(Y(1) \in \hat{C}_1(X) \mid T = 0) \leq 90\% + c/n$$

- *Lower bound holds without extra assumption*
- *Upper bound holds if V_i's are a.s. distinct & overlap holds, and c only depends on the overlap*

✓ Any conditional distribution $P_{Y(1)|X}$

✓ Any sample size

✓ Any procedure to fit conditional quantiles



Approximate counterfactual inference

Theorem (informal, L. and Candès, 2020, for observational studies)

Let $\hat{e}(x)$ *be an estimate of* $e(x)$*. Set* $w(x) = (1 - \hat{e}(x))/\hat{e}(x)$ *in weighted split-CQR. Then*

$$\mathbb{P}(Y(1) \in \hat{C}_1(X) \mid T = 0) \approx 90\%$$

if (1) $\hat{e}(x) \approx e(x)$ **OR** *(2)* $\hat{q}_{0.05/0.95}(x) \approx q_{0.05/0.95}(x)$*. Under (2),*

$$\mathbb{P}(Y(1) \in \hat{C}_1(X) \mid T = 0, X) \approx 90\% \textit{ with high probability (conditional coverage!)}$$

Similar to the **double robustness** for ATE

Agenda

Part II: Empirical results on counterfactual inference

Simulation

- Variant of example from Wager and Athey ('18)
- $X \in \mathbb{R}^d$ Gaussian, independent or correlated, with $d \in \{10, 100\}$
- $Y(0) \equiv 0 \rightsquigarrow$ ITE inference is counterfactual inference
- $Y(1) \mid X \sim N(\mu(X), \sigma(X)^2)$:
  - $\mu(X)$ depends on X_1, X_2 smoothly
  - $\sigma(X) \equiv 1$ (homoscedastic) or $\sigma(X) = -\log(1 - \Phi(X_1))$ (heteroscedastic)
- $e(X) \in [0.25, 0.5]$ depends on X_1 smoothly

Our R package cfcausal (github.com/lihualei71/cfcausal)

cfcausal 0.2.0 | Reference | Articles

cfcausal

An R package for conformal inference of counterfactuals and individual treatment effects

Overview

This R package implements weighted conformal inference-based procedures for counterfactuals and individual treatment effects proposed in our paper: Conformal Inference of Counterfactuals and Individual Treatment Effects. It includes both the split conformal inference and cross-validation+. For each type of conformal inference, both conformalized quantile regression (CQR) and standard conformal inference are supported. It provides a pool of convenient learners and allows flexible user-defined learners for conditional mean and quantiles.

- `conformalCf()` produces intervals for counterfactuals or outcomes with missing values in general.
- `conformalIte()` produces intervals for individual treatment effects with a binary treatment under the potential outcome framework.
- `conformal()` provides a generic framework of weighted conformal inference for continuous outcomes.
- `conformalInt()` provides a generic framework of weighted conformal inference for interval outcomes.

Installation

```
if (!require("devtools")){
    install.packages("devtools")
}
devtools::install_github("lihualei71/cfcausal")
```

License

Full license

MIT + file LICENSE

Citation

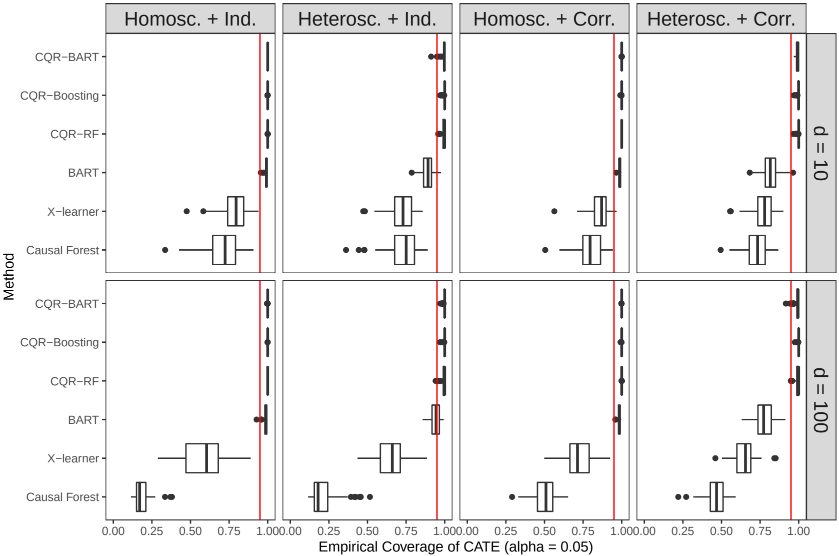
Citing cfcausal

Developers

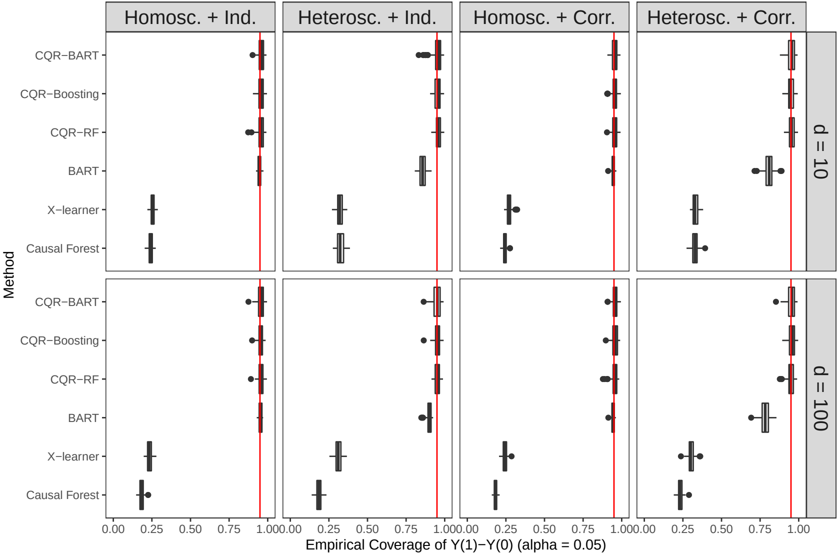
Lihua Lei

Maintainer

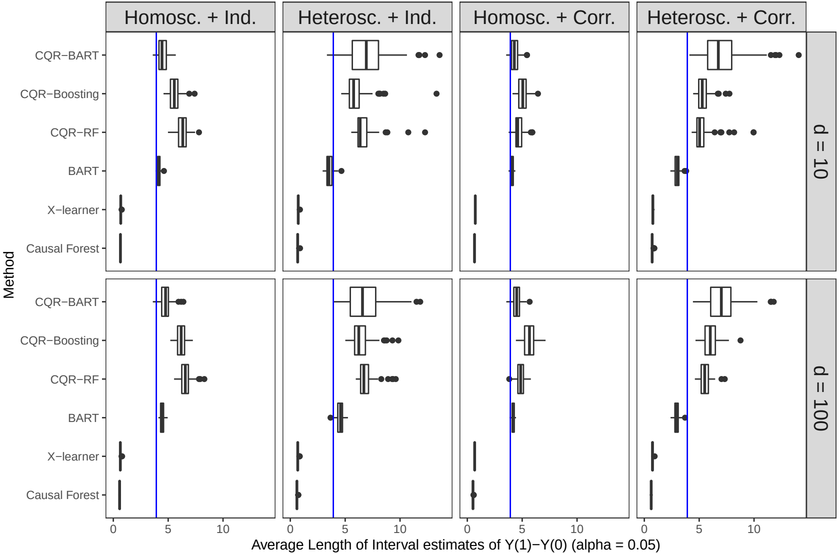
Marginal coverage of $\mathrm{CATE} = \mathbb{E}[Y(1) \mid X]$ (sanity check)



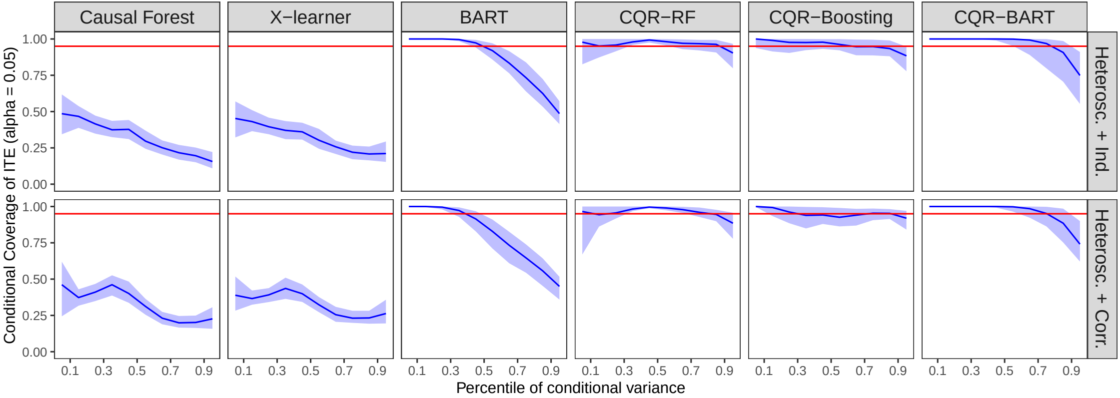
Marginal coverage of $Y(1)$



Average length of $\hat{C}_1(X)$



Conditional coverage of $Y(1)$



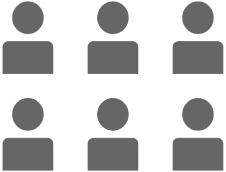
Agenda

Part III: from counterfactuals to individual treatment effects

The ITE inference problem

$$(X_i, Y_i^{\text{obs}}) \overset{i.i.d.}{\sim} P_{X|T=0} \times P_{Y(0)|X}$$

$$(X_i, Y_i^{\text{obs}}) \overset{i.i.d.}{\sim} P_{X|T=1} \times P_{Y(1)|X}$$



$$X \sim Q_X$$

L. and Candès, '20

Prediction interval for **individual treatment effect** $\text{ITE} = Y(1) - Y(0)$

$$\mathbb{P}_{X \sim Q_X}\left(\text{ITE} \in \hat{C}_{\text{ITE}}(X)\right) \geq 1 - \alpha$$

Contrast with conditional average treatment effects

Conditional average treatment effects (CATE)

$$\tau(x) \triangleq \mathbb{E}[\mathrm{ITE} \mid X = x] \neq \mathrm{ITE}$$

Contrast with conditional average treatment effects

Conditional average treatment effects (CATE)

$$\tau(x) \triangleq \mathbb{E}[\mathrm{ITE} \mid X = x] \neq \mathrm{ITE}$$

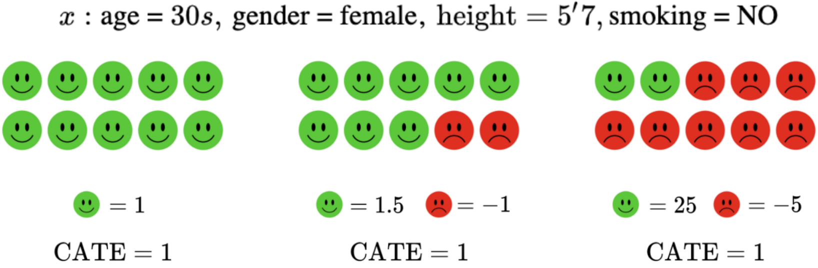
- Uncertainty of the response around the CATE function (ignored by CATE)

Contrast with conditional average treatment effects

Conditional average treatment effects (CATE)

$$\tau(x) \triangleq \mathbb{E}[\mathrm{ITE} \mid X = x] \neq \mathrm{ITE}$$

- Uncertainty of the response around the CATE function (ignored by CATE)



Contrast with conditional average treatment effects

Conditional average treatment effects (CATE)

$$\tau(x) \triangleq \mathbb{E}[\mathrm{ITE} \mid X = x] \neq \mathrm{ITE}$$

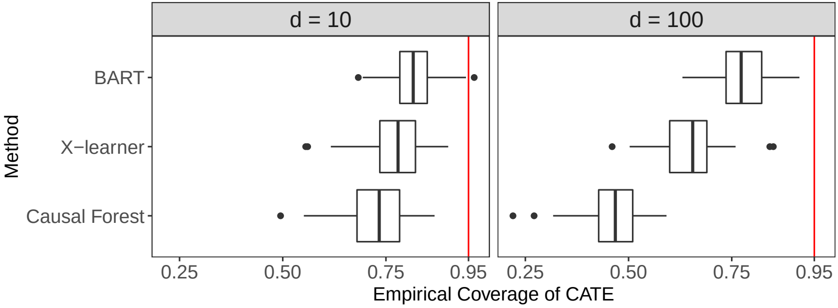
- Uncertainty of the response around the CATE function (ignored by CATE)
- Uncertainty of CATE estimators due to finite samples (impossibility result by Barber '20)

Contrast with conditional average treatment effects

Conditional average treatment effects (CATE)

$$\tau(x) \triangleq \mathbb{E}[\mathrm{ITE} \mid X = x] \neq \mathrm{ITE}$$

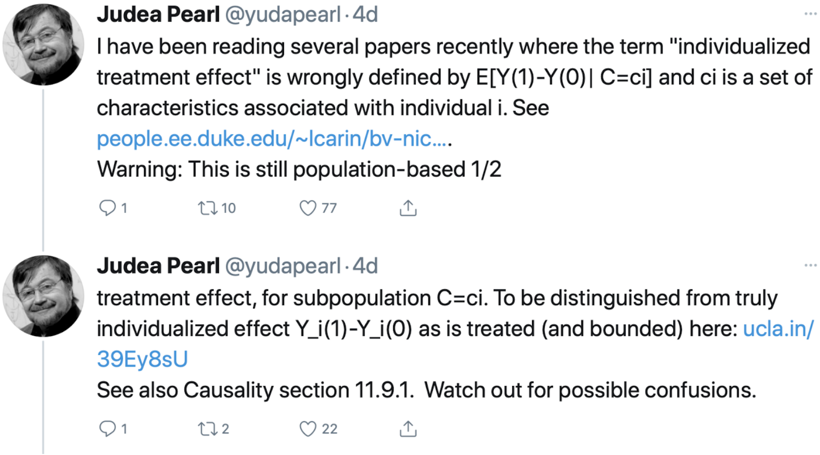
- Uncertainty of the response around the CATE function (ignored by CATE)
- Uncertainty of CATE estimators due to finite samples (impossibility result by Barber '20)



Contrast with conditional average treatment effects

Conditional average treatment effects (CATE)

$$\tau(x) \triangleq \mathbb{E}[\mathrm{ITE} \mid X = x] \neq \mathrm{ITE}$$



Judea Pearl @yudapearl · 4d

I have been reading several papers recently where the term "individualized treatment effect" is wrongly defined by E[Y(1)-Y(0)| C=ci] and ci is a set of characteristics associated with individual i. See people.ee.duke.edu/~lcarin/bv-nic....

Warning: This is still population-based 1/2

1 10 77

Judea Pearl @yudapearl · 4d

treatment effect, for subpopulation C=ci. To be distinguished from truly individualized effect Y_i(1)-Y_i(0) as is treated (and bounded) here: ucla.in/39Ey8sU

See also Causality section 11.9.1. Watch out for possible confusions.

1 2 22

Summary

Conformal inference of counterfactuals and individual treatment effects is reliable

- Randomized experiments: **near-exact** coverage in finite samples with any black-box
- Observational studies: **doubly robust** guarantees of coverage

Other Uses?

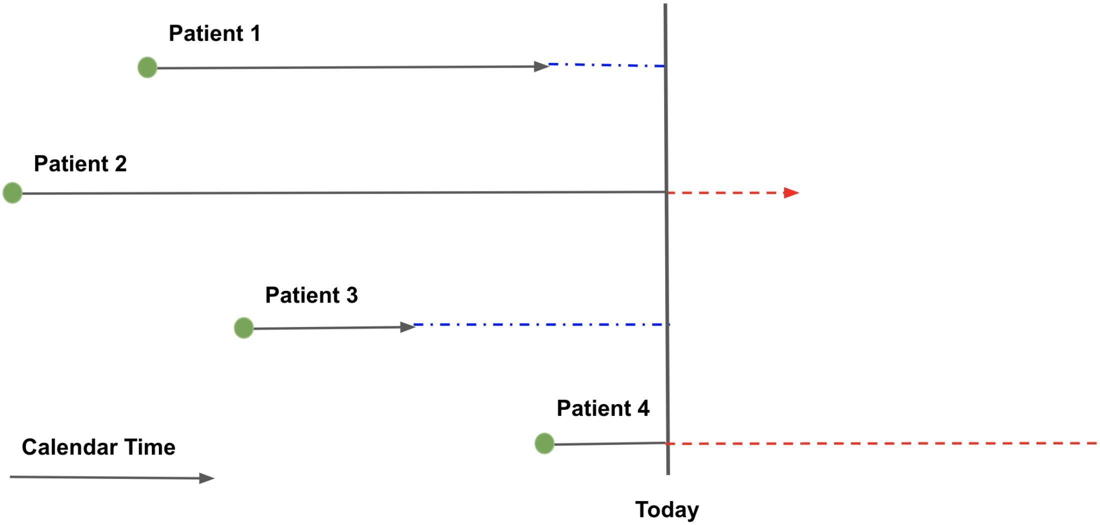
- Conformalized survival predictive analysis (w/ Emmanuel Candès and Zhimei Ren)
- Medical image analysis (w/ Stephen Bates, Anastasios Angelopoulos, Jitendra Malik, and Micheal Jordan)

Conformalized survival analysis

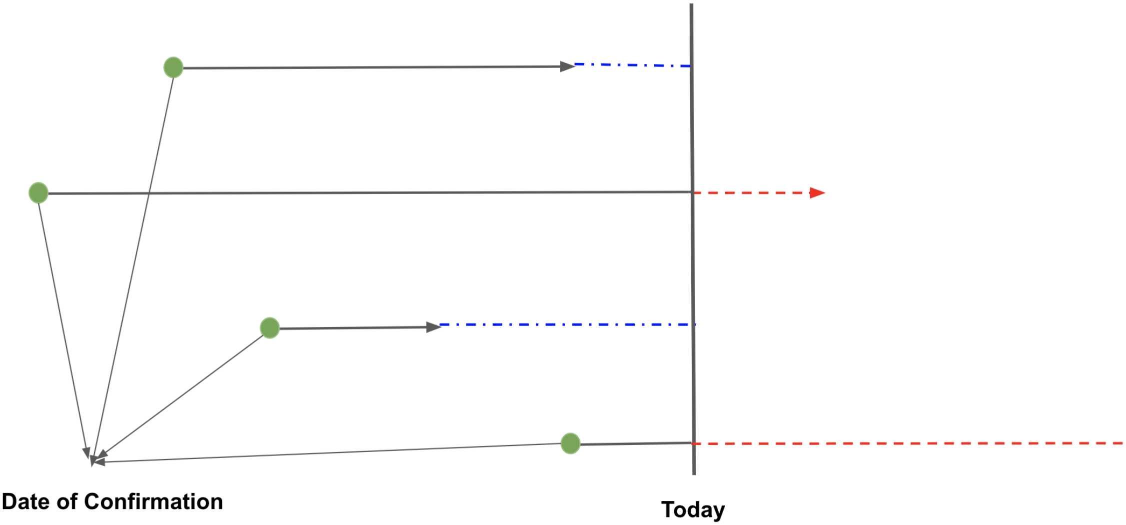
Zhimei Ren

Emmanuel Candès

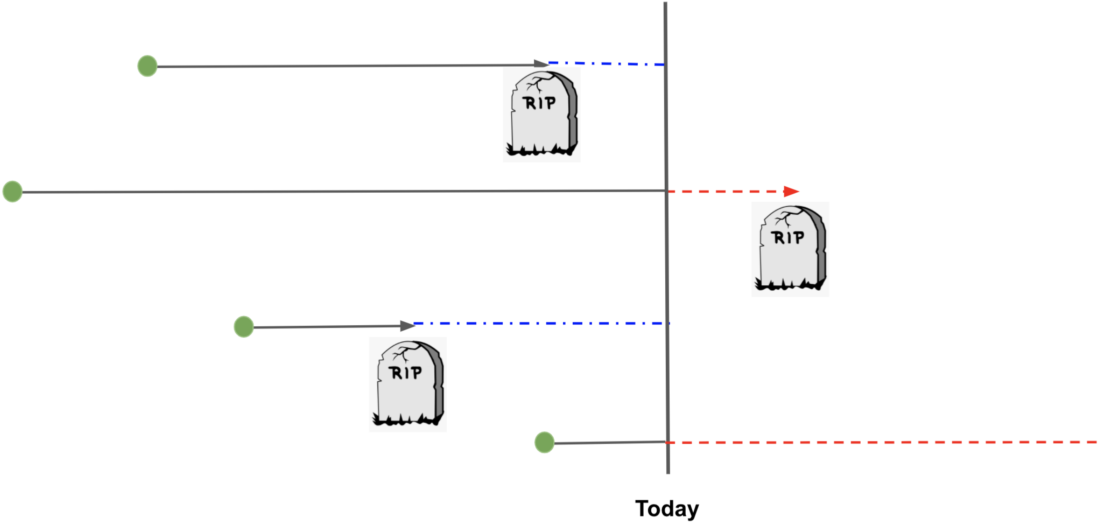
Right Censored Data: Type-I Censoring



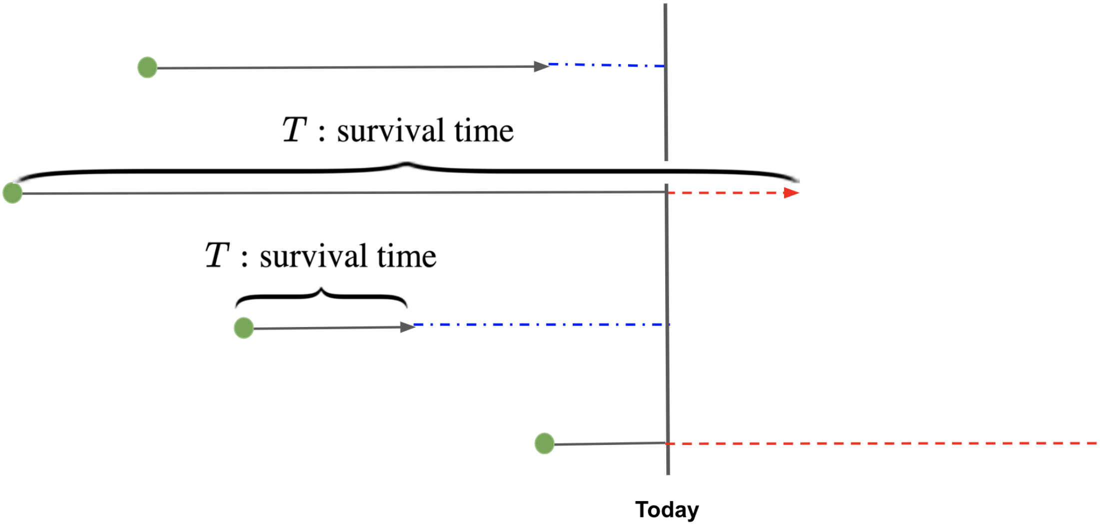
Right Censored Data: Type-I Censoring



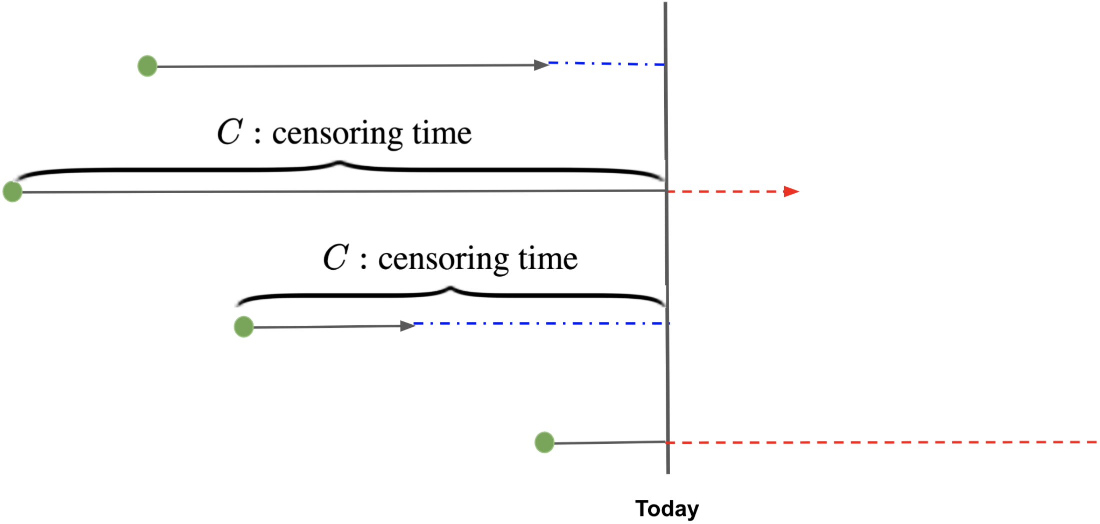
Right Censored Data: Type-I Censoring



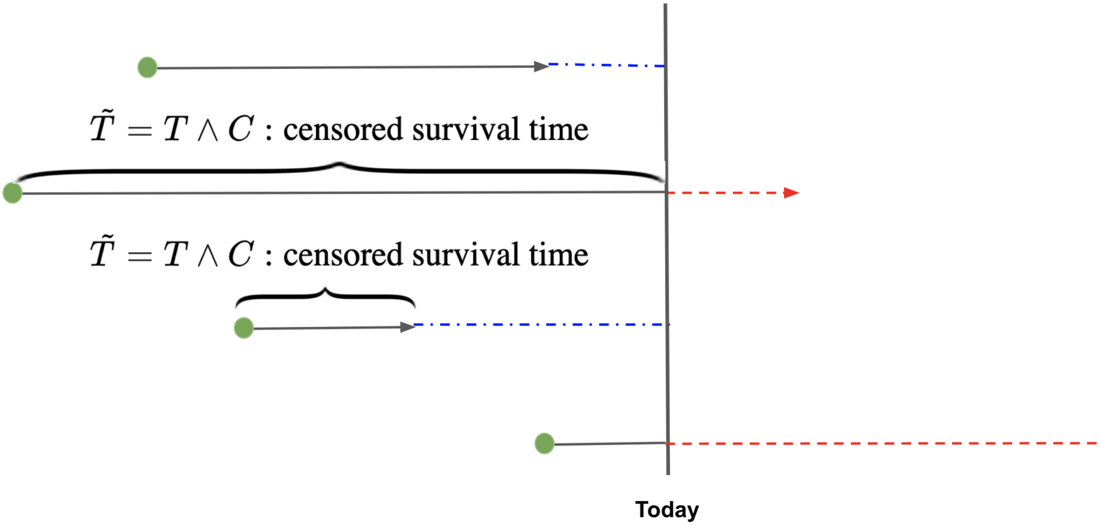
Right Censored Data: Type-I Censoring



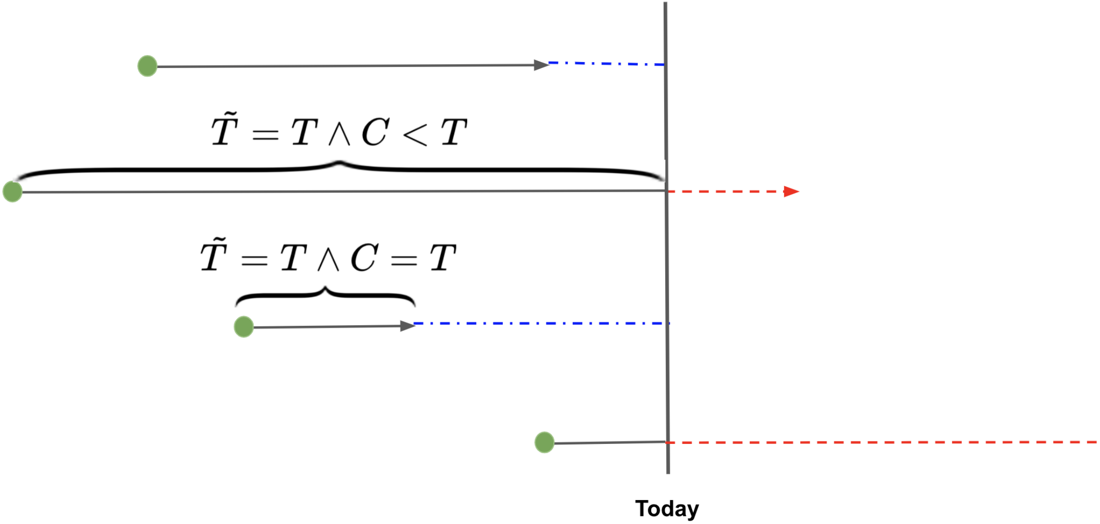
Right Censored Data: Type-I Censoring



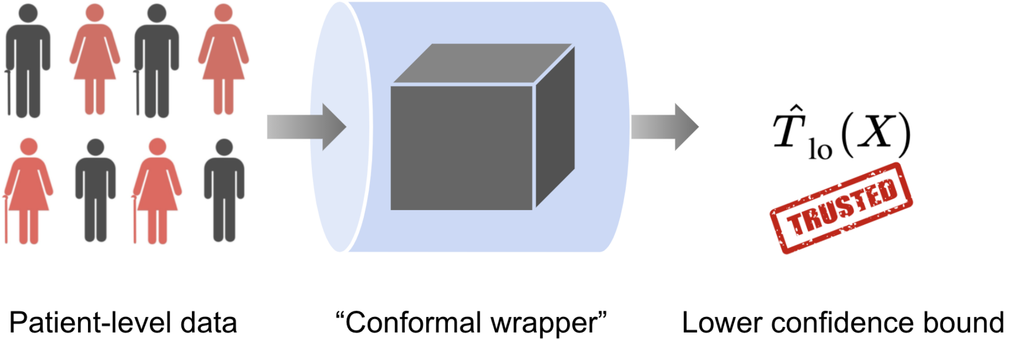
Right Censored Data: Type-I Censoring



Right Censored Data: Type-I Censoring



A reliable predictive system for survival times



Find lower predictive bound $\hat{T}_{\mathrm{lo}}(X)$, s.t. $\mathbb{P}(T \geq \hat{T}_{\mathrm{lo}}(X)) \geq 90\%$

Survival times as counterfactuals?

- Event indicator $\Delta = I(T < C)$:

$$\tilde{T} = \begin{cases} T & \text{if } \Delta = 1 \\ C & \text{if } \Delta = 0 \end{cases} .$$

- Treat T as a "potential outcome" under the "treatment" $\Delta = 1$?

- INVALID because "unconfoundedness" does not hold:

$$(T, C) \not\perp I(T < C) \mid X$$

- $(X_i, T_i)_{\Delta_i = 1}$ has shifts in both the covariate distribution and conditional survival function

Conformalized survival analysis

- Ignoring the censoring leads to a prediction problem

$$\mathbb{P}(\tilde{T} \geq \hat{T}_{\mathrm{lo}}(X)) \geq 90\% \Longrightarrow \mathbb{P}(T \geq \hat{T}_{\mathrm{lo}}(X)) \geq 90\%$$

- Potentially huge efficiency loss

Conformalized survival analysis

- Ignoring the censoring leads to a prediction problem

$$\mathbb{P}(\tilde{T} \geq \hat{T}_{\text{lo}}(X)) \geq 90\% \Longrightarrow \mathbb{P}(T \geq \hat{T}_{\text{lo}}(X)) \geq 90\%$$

- Potentially huge efficiency loss
- We apply **weighted conformal inference** on a carefully chosen subpopulation
- Near-exactness: $\hat{T}_{\text{lo}}(X)$ is valid if $P(C \mid X)$ is known (up to a multiplicative constant)
- Double robustness: $\hat{T}_{\text{lo}}(X)$ is approximately valid if $P(C \mid X)$ or $P(T \mid X)$ is estimated well
- Also useful beyond the type-I censoring

- Tutorial on conformal inference by Emmanuel Candès at Bernoulli-IMS One World Symposium
- *Conformal Inference of Counterfactuals and Individual Treatment Effects* (**L.** and Candès, '20)
- *Conformalized Survival Analysis* (Candès*, **L.***, and & Ren*, '21)
- *Distribution-Free, Risk-Controlling Prediction Sets* (Bates*, Angelopoulos*, **L.***, Malik, and Jordan, '21)

Thank you!

* alphabetical order or equal contribution

Double robustness of weighted split-CQR

Theorem (L. and Candès, '20)

Assume one of the following holds:

(1) $\mathbb{E}\left|1/\hat{e}(X) - 1/e(X)\right| = o(1)$;

(2) $\mathbb{P}(Y(1) = y \mid X = x)$ *uniformly bounded away from* 0 *and* ∞ *and there exists* $\delta > 0$

$$\mathbb{E}\left[1/\hat{e}(X)^{1+\delta}\right] = O(1), \quad \mathbb{E}\left[H(X)/\hat{e}(X)\right], \mathbb{E}\left[H(X)/e(X)\right] = o(1),$$

$$\textit{where } H(x) = \max\{|\hat{q}_{0.05}(x) - q_{0.05}(x)|, |\hat{q}_{0.95}(x) - q_{0.95}(x)|\}.$$

Then

$$\mathbb{P}(Y(1) \in \hat{C}_1(X) \mid T = 0) \geq 90\% - o(1).$$

Furthermore, if (2) holds, then

$$\mathbb{P}(Y(1) \in \hat{C}_1(X) \mid T = 0, X) \geq 90\% - o_{\mathbb{P}}(1).$$

The ITE inference problem

- **Naive approach**: get $\hat{C}_1(x)$ and $\hat{C}_0(x)$ by weighted split-CQR and set

$$\hat{C}_{\mathrm{ITE}}(x) = \hat{C}_1(x) - \hat{C}_0(x)$$

  - Apply Bonferroni correction (5% for each potential outcome)
  - $\mathbb{P}(Y(1) - Y(0) \in \hat{C}_{\mathrm{ITE}}(X)) \geq 90\%$ regardless of the correlation structure between $Y(1)$ and $Y(0)$

The ITE inference problem

- **Naive approach**: get $\hat{C}_1(x)$ and $\hat{C}_0(x)$ by weighted split-CQR and set

$$\hat{C}_{\mathrm{ITE}}(x) = \hat{C}_1(x) - \hat{C}_0(x)$$

  - Apply Bonferroni correction (5% for each potential outcome)
  - $\mathbb{P}(Y(1) - Y(0) \in \hat{C}_{\mathrm{ITE}}(X)) \geq 90\%$ regardless of the correlation structure between $Y(1)$ and $Y(0)$

- **Nested approach**: our focus
  - Use counterfactual inference to generate ITE intervals for subjects in the study
  - Generalize these intervals to subjects not in the study

$\rightsquigarrow$ Reduces conservatism of the naive approach