What Can Conformal Inference Offer to Statistics?

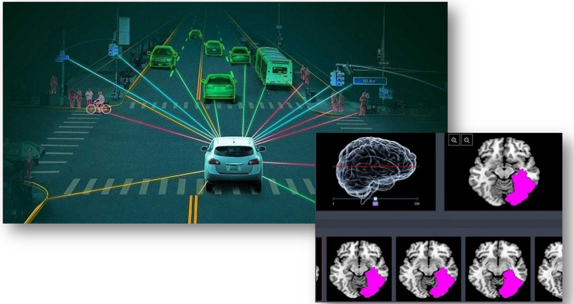
Lihua Lei

Department of Statistics, Stanford University

Job Talk, 2021-2022

ML in critical applications

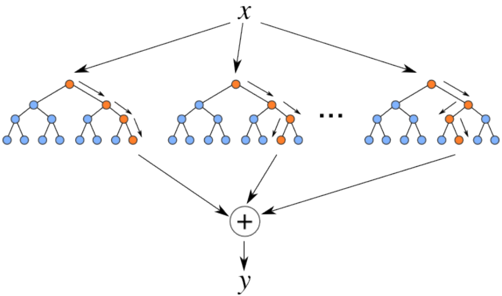
ML tools make potentially high-stakes decisions: self-driving cars, disease diagnosis, ...



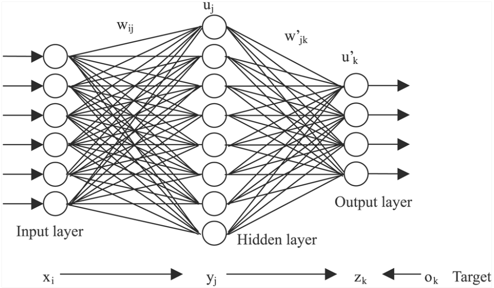
Can we have reliable uncertainty quantification (confidence) in these predictions?

Today's predictive algorithms

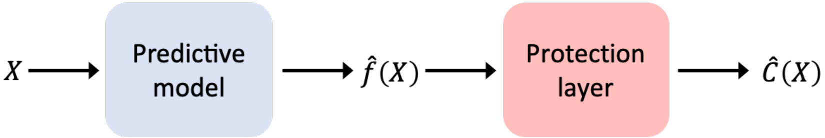
random forests, gradient boosting



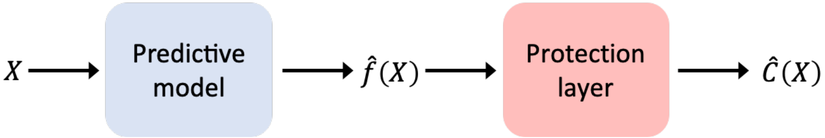
neural networks



Previous work on conformal inference



Previous work on conformal inference



- i.i.d. training samples (X_i, Y_i), $i = 1, \ldots, n$
- Test point $(X, Y = ?)$ from the same distribution
- 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 $\hat{f}$

Statistical inference is more complicated than predicting "the seen"

Statistical inference is more complicated than predicting "the seen"

- Counterfactual
  - Causal inference, offline policy evaluation, algorithmic fairness, ...
  - What would have been one's response had one taken the treatment
  - Observable for those in a particular "treatment arm"

Statistical inference is more complicated than predicting “the seen”

- Counterfactual
  - Causal inference, offline policy evaluation, algorithmic fairness, ...
  - What would have been one’s response had one taken the treatment
  - Observable for those in a particular “treatment arm”
- Time-to-event (survival) outcome
  - Survival analysis, industrial life testing, economics, ...
  - Censored by study termination, loss to follow-up, ...
  - Observable for those whose event (e.g., death) has occurred

Statistical inference is more complicated than predicting "the seen"

- Counterfactual
  - Causal inference, offline policy evaluation, algorithmic fairness, ...
  - What would have been one's response had one taken the treatment
  - Observable for those in a particular "treatment arm"
- Time-to-event (survival) outcome
  - Survival analysis, industrial life testing, economics, ...
  - Censored by study termination, loss to follow-up, ...
  - Observable for those whose event (e.g., death) has occurred

Goal: construct calibrated prediction intervals for partially observed outcomes

Part I: conformalized counterfactual prediction

Emmanuel Candès

Inference of counterfactuals? Potential outcomes Neyman '23, Rubin '74

Assumptions

- stable unit treatment values (SUTVA)

science table

Unit	x_i	T_i	$Y_i(1)$	$Y_i(0)$	Y_i^{obs}
Treatment Group					
1	✓	1	✓	×	$Y_1(1)$
2	✓	1	✓	×	$Y_2(1)$
3	✓	1	✓	×	$Y_3(1)$
4	✓	1	✓	×	$Y_4(1)$
5	✓	1	✓	×	$Y_5(1)$
Control Group					
6	✓	0	×	✓	$Y_6(0)$
7	✓	0	×	✓	$Y_7(0)$
8	✓	0	×	✓	$Y_8(0)$
9	✓	0	×	✓	$Y_9(0)$
10	✓	0	×	✓	$Y_{10}(0)$

Inference of counterfactuals? Potential outcomes Neyman '23, Rubin '74

Assumptions

- stable unit treatment values (SUTVA)
- super population (i.i.d.)
- unconfoundedness $(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$

science table

Unit	x_i	T_i	$Y_i(1)$	$Y_i(0)$	Y_i^{obs}
Treatment Group					
1	✓	1	✓	×	$Y_1(1)$
2	✓	1	✓	×	$Y_2(1)$
3	✓	1	✓	×	$Y_3(1)$
4	✓	1	✓	×	$Y_4(1)$
5	✓	1	✓	×	$Y_5(1)$
Control Group					
6	✓	0	×	✓	$Y_6(0)$
7	✓	0	×	✓	$Y_7(0)$
8	✓	0	×	✓	$Y_8(0)$
9	✓	0	×	✓	$Y_9(0)$
10	✓	0	×	✓	$Y_{10}(0)$

Inference of counterfactuals? Potential outcomes
Neyman '23, Rubin '74

Assumptions

- stable unit treatment values (SUTVA)
- super population (i.i.d.)
- unconfoundedness $(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$

Goal: find interval estimate $\hat{C}_1(X)$, s.t.,

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

science table

Unit	x_i	T_i	$Y_i(1)$	$Y_i(0)$	Y_i^{obs}
Treatment Group					
1	✓	1	✓	×	$Y_1(1)$
2	✓	1	✓	×	$Y_2(1)$
3	✓	1	✓	×	$Y_3(1)$
4	✓	1	✓	×	$Y_4(1)$
5	✓	1	✓	×	$Y_5(1)$
Control Group					
6	✓	0	×	✓	$Y_6(0)$
7	✓	0	×	✓	$Y_7(0)$
8	✓	0	×	✓	$Y_8(0)$
9	✓	0	×	✓	$Y_9(0)$
10	✓	0	×	✓	$Y_{10}(0)$

Counterfactual inference

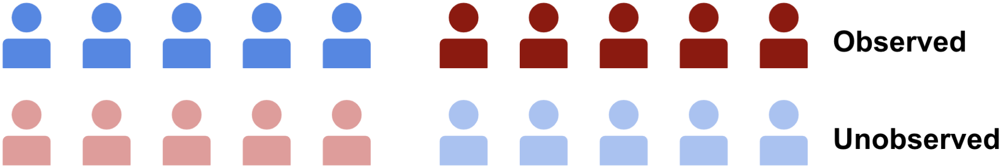
Assign treatment by a coin toss for each subject based on the **propensity score** $e(x)$

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

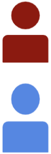
$$\mathbb{P}(\text{control} \mid X = x) = 1 - e(x)$$

Counterfactual inference

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



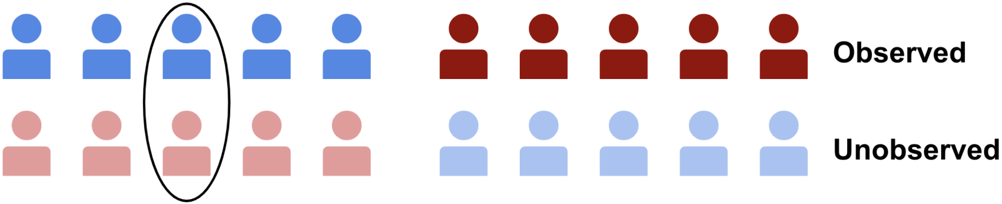
SUTVA



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

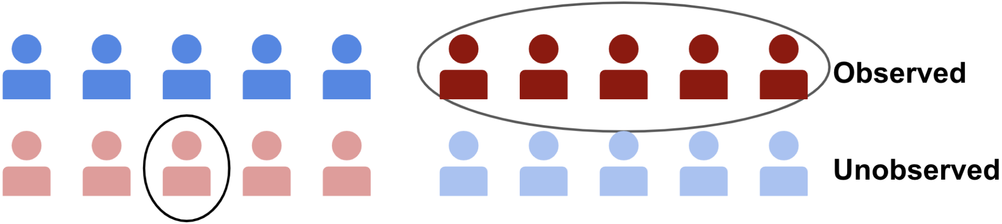
$$Y^{\mathrm{obs}} = Y(0)$$

Counterfactual inference



How to infer $Y(1)$ of  

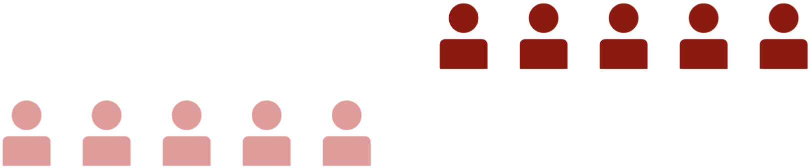
Counterfactual inference



Use observed treated units

Covariate shift under unconfoundedness $Y(1) \perp\!\!\!\perp T \mid X$

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



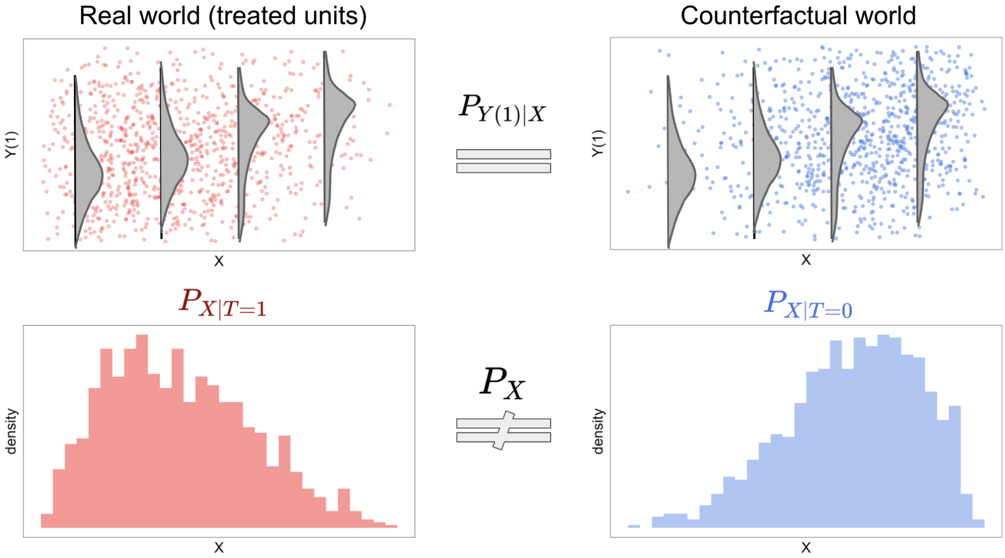
Observed

Unobserved

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

Distribution mismatch! Covariate shift

Covariate shift under unconfoundedness $Y(1) \perp\!\!\!\perp T \mid X$



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

Conformal inference under covariate shift

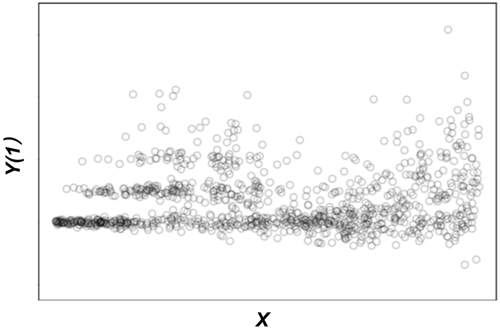
Weighted Split Conformalized Quantile Regression (CQR)

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

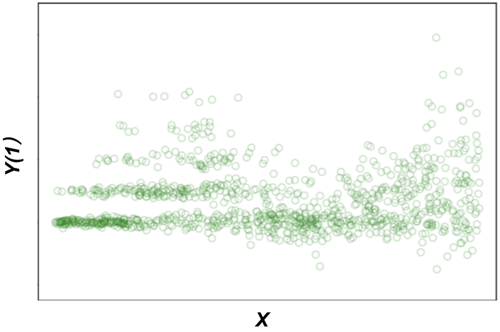
$$(X_i, Y_i) \overset{i.i.d.}{\sim} P_X \times P_{Y|X} \Longrightarrow \mathbb{P}_{(X,Y)\sim Q_X \times P_{Y|X}}(Y \in \hat{C}(X)) \geq 90\%$$

Weighted CQR

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



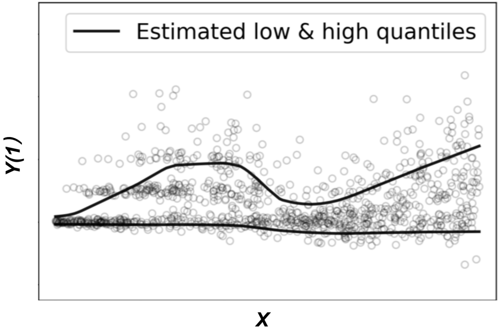
Proper training set



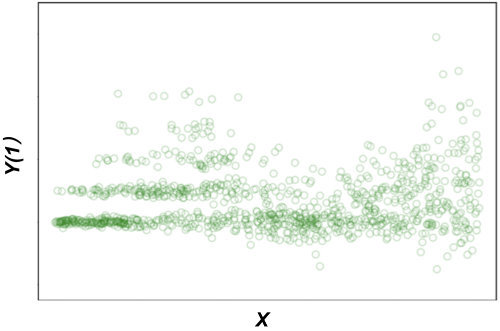
Calibration set

Weighted CQR

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

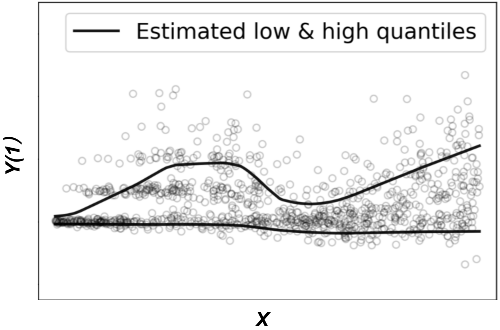


Apply quantile regression

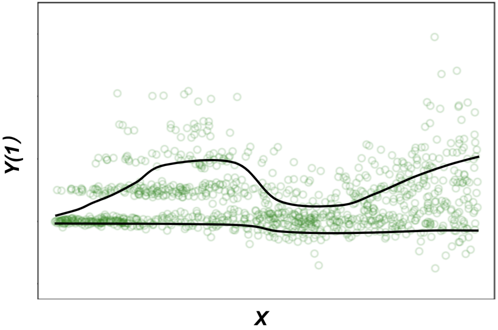


Calibration set

Weighted CQR



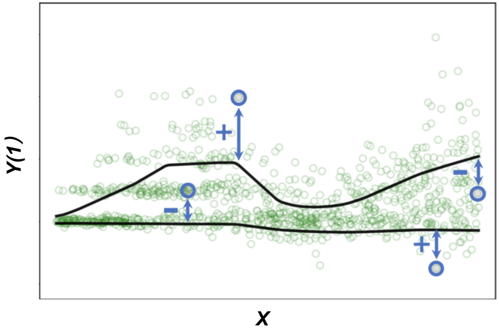
Apply quantile regression



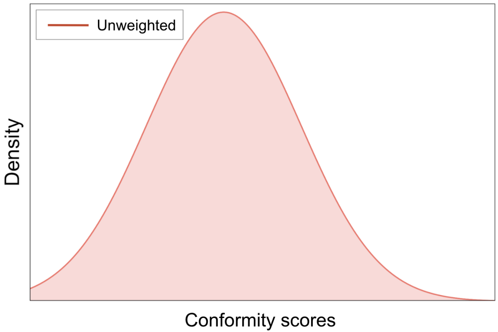
Calibration set

Weighted CQR

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



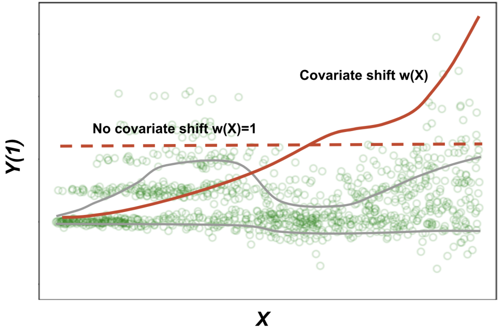
Calibrate



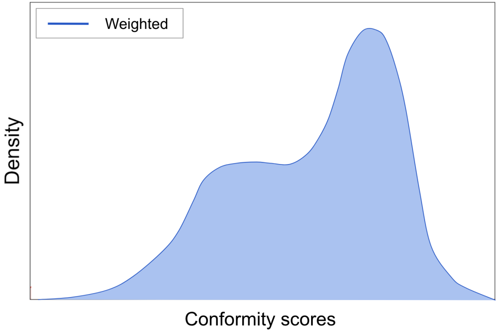
Histogram of signed distances

Weighted CQR

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



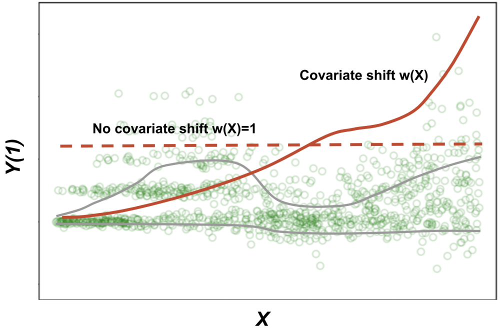
Calibrate



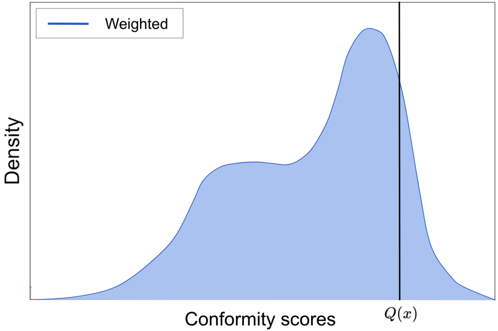
Histogram weighted by $w(x)$

Weighted CQR

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



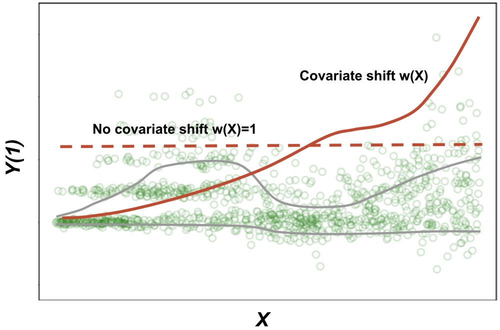
Calibrate



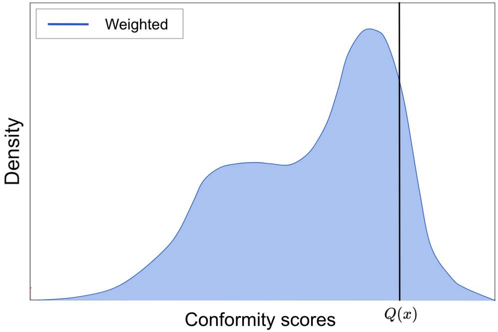
Find the 90%-th quantile $Q(x)$

Weighted 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, '20, for randomized experiments)

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

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

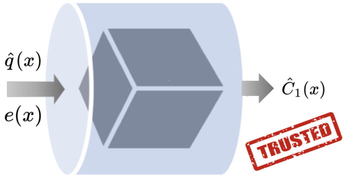
- *Lower bound holds without extra assumption*
- *Upper bound holds if V_i's are a.s. distinct & overlap holds, and c only depends on the overlap*

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

✓ Any sample size

✓ Any procedure to fit conditional quantiles

- Reconcile Bayesians and frequentists (if $Q(x) \ge 0$)



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)* $\hat{q}_{0.05/0.95}(x) \approx q_{0.05/0.95}(x)$*.*

Similar to the **double robustness** for ATE

Adaptivity to good outcome modelling

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

If $\hat{q}_{0.05/0.95}(x) \approx q_{0.05/0.95}(x)$,

$$\hat{C}_1(x) \approx [q_{0.05}(x), q_{0.95}(x)] \;\; (\textit{oracle counterfactual interval}),$$

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

- $[q_{0.05}(x), q_{0.95}(x)]$ is the optimal interval for symmetric unimodal conditional distribution
- Continue to hold if $(0.05, 0.95) \to (\beta, \beta + 1 - \alpha)$, e.g., $(0.01, 0.91)$ Romano and Sesia, '21
- Good outcome modelling $\Longrightarrow$ good intervals and conditional coverage!
- Robustness (marginal coverage) + adaptivity (efficiency and conditional coverage)

Technical conditions

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* $\delta > 0$

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

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

Then

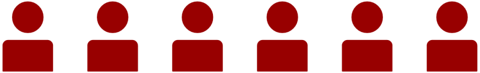
$$\mathbb{P}(Y(1) \in \hat{C}_1(X) \mid T = 0) \geq 90\% - o(1).$$

Furthermore, if (2) holds, then

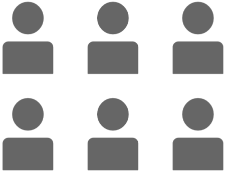
$$\mathbb{P}(Y(1) \in \hat{C}_1(X) \mid T = 0, X) \geq 90\% - o_{\mathbb{P}}(1).$$

From counterfactuals to individual treatment effects (ITE)

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



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



$$X \sim Q_X$$

L. and Candès, '20

Prediction interval for $\mathrm{ITE} = Y(1) - Y(0)$ (not $\mathrm{CATE} = \mathbb{E}[\mathrm{ITE} \mid X]$)

$$\mathbb{P}_{X \sim Q_X}(\mathrm{ITE} \in \hat{C}_{\mathrm{ITE}}(X)) \geq 90\%$$

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

Summary for conformalized counterfactual inference

Conformal inference of counterfactuals is reliable

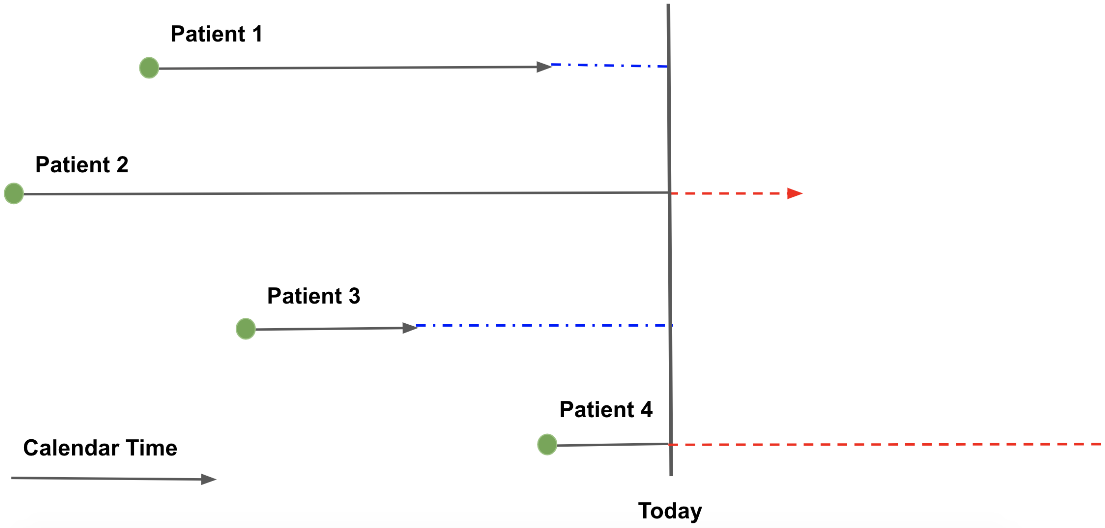
- Randomized experiments: **near-exact** coverage in finite samples with any black-box
- Observational studies: **doubly robust** guarantees of coverage

Part II: conformalized survival analysis

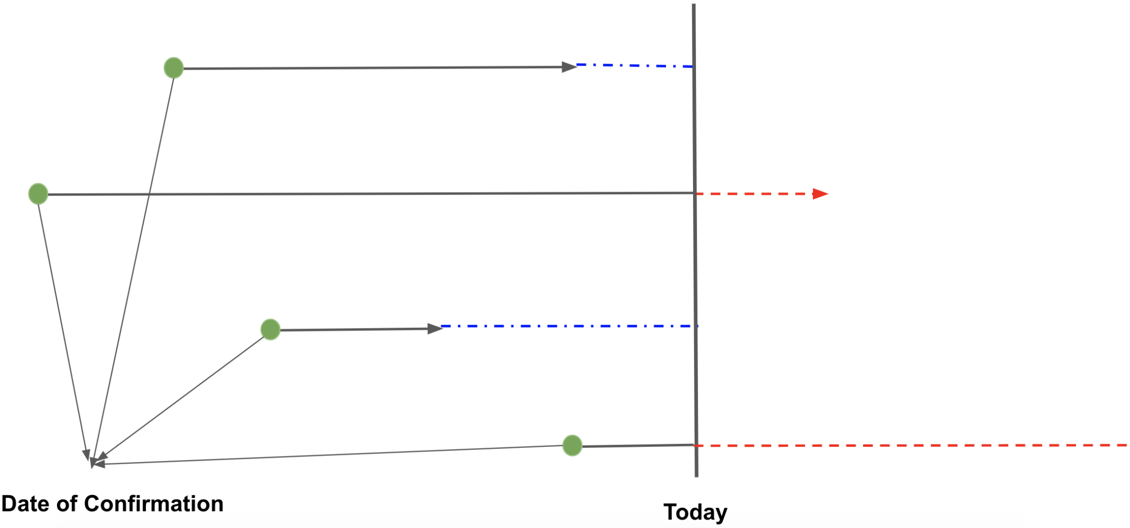
Zhimei Ren

Emmanuel Candès

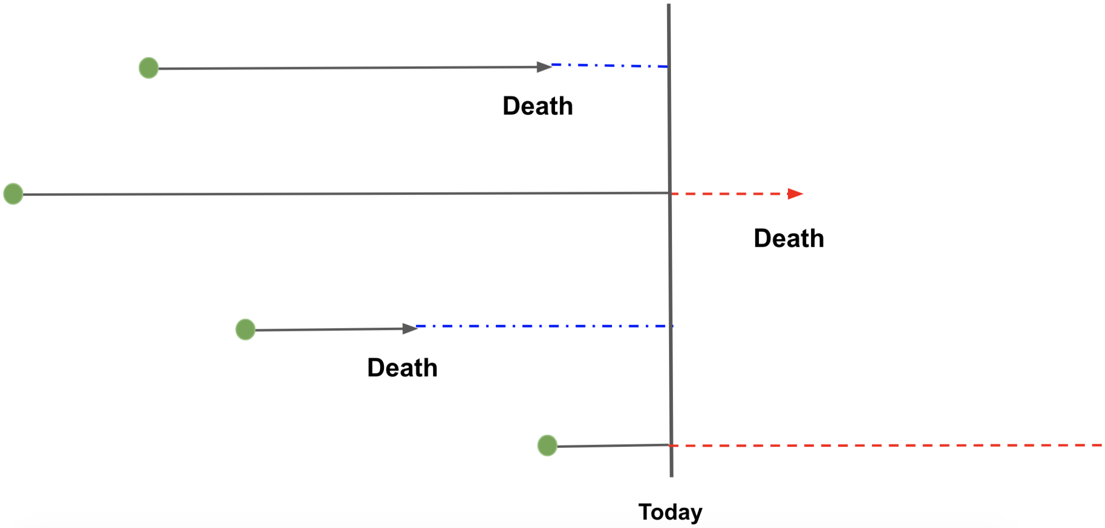
Right Censored Data: Type-I Censoring



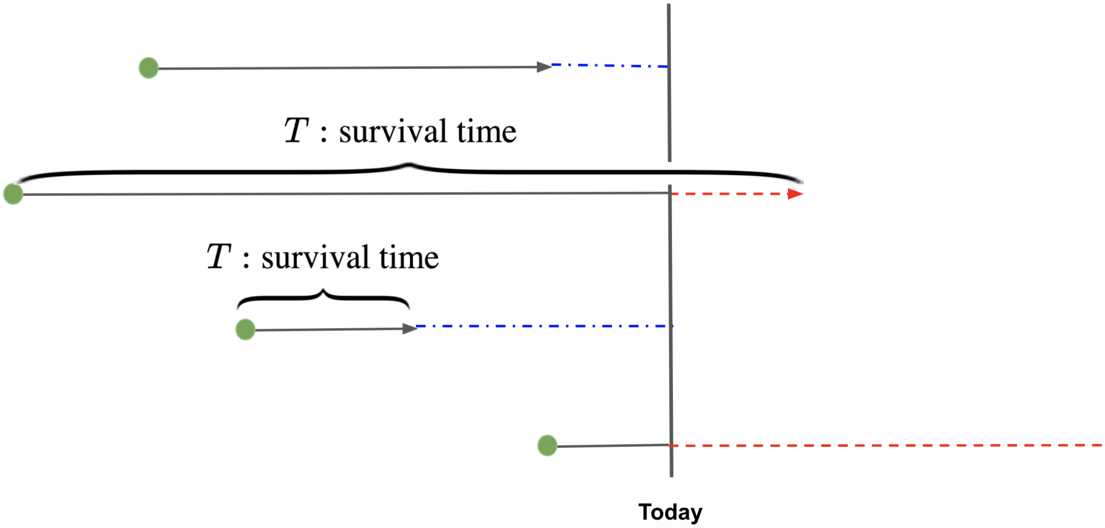
Right Censored Data: Type-I Censoring



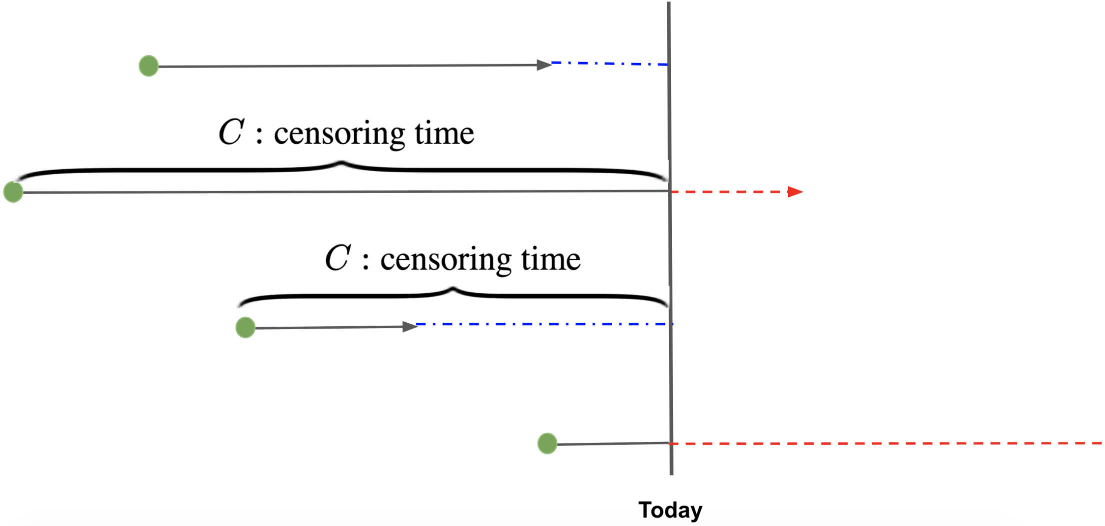
Right Censored Data: Type-I Censoring



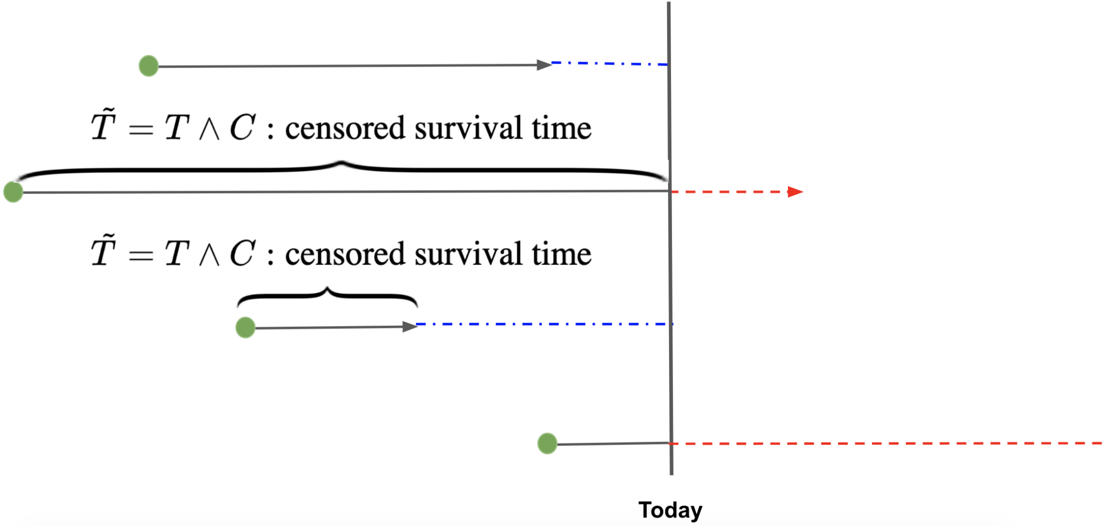
Right Censored Data: Type-I Censoring



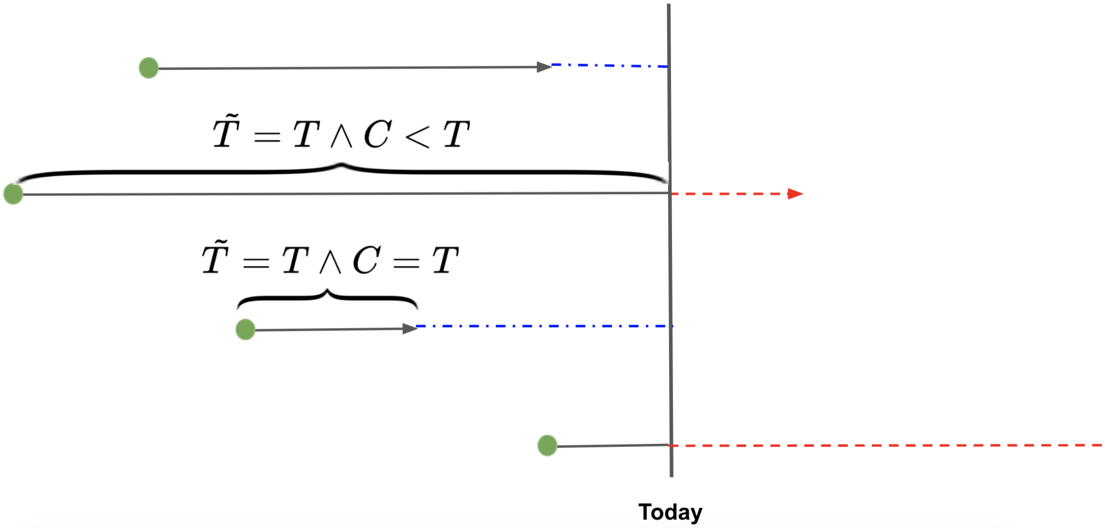
Right Censored Data: Type-I Censoring



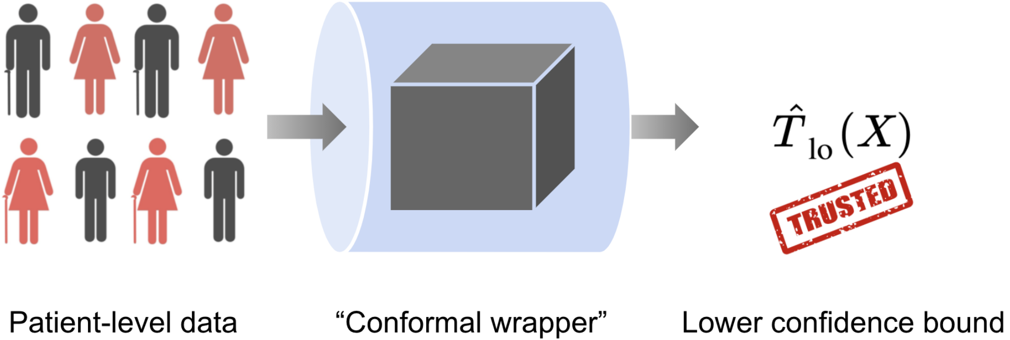
Right Censored Data: Type-I Censoring



Right Censored Data: Type-I Censoring



A reliable predictive system for survival times



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

First thought: 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” $\Delta = 1$?

First thought: survival times as counterfactuals?

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

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

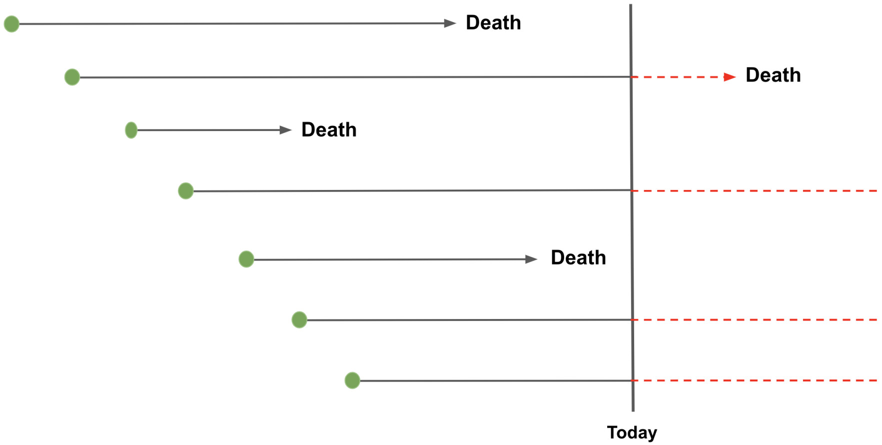
- Treat T as a "potential outcome" under the "treatment" $\Delta = 1$?
- **Invalid** because "unconfoundedness" does not hold:

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

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

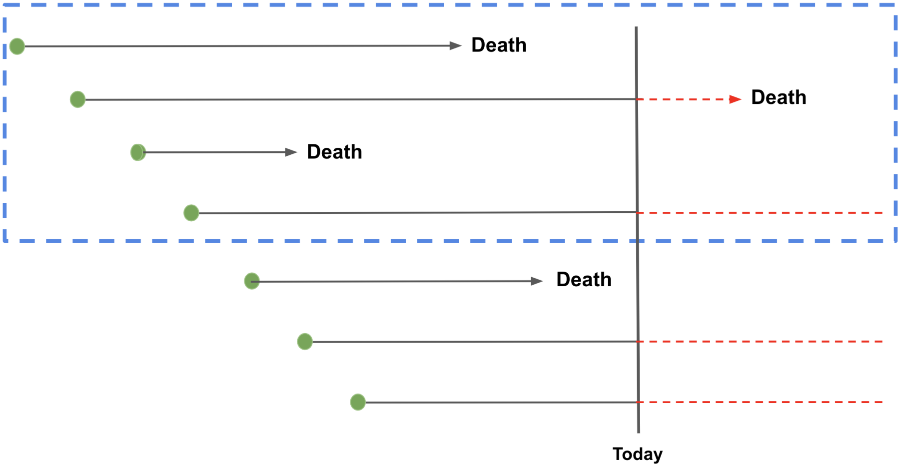
Conformalized survival analysis

Order the units by censoring times C_i



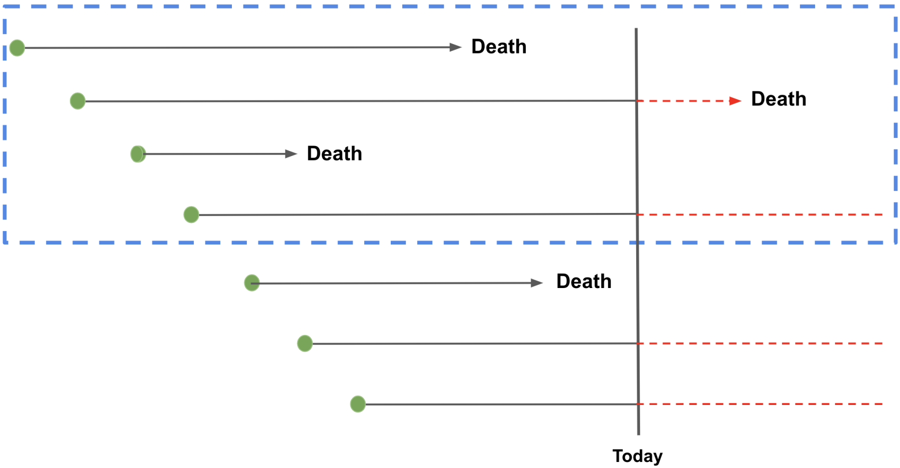
Conformalized survival analysis

Study population: $C_i \geq c_0$ (c_0 chosen via data splitting)



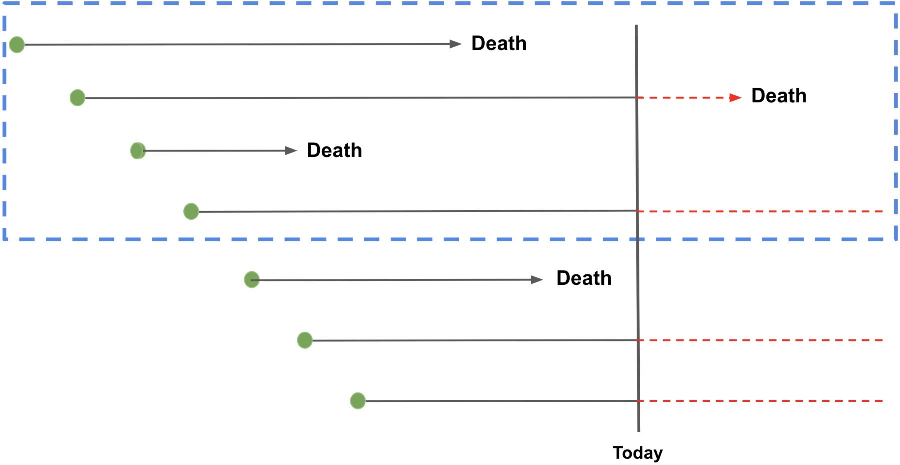
Conformalized survival analysis

On this population ($C \geq c_0$), the **surrogate outcome** $T \wedge c_0 = \tilde{T} \wedge c_0$ is always observable

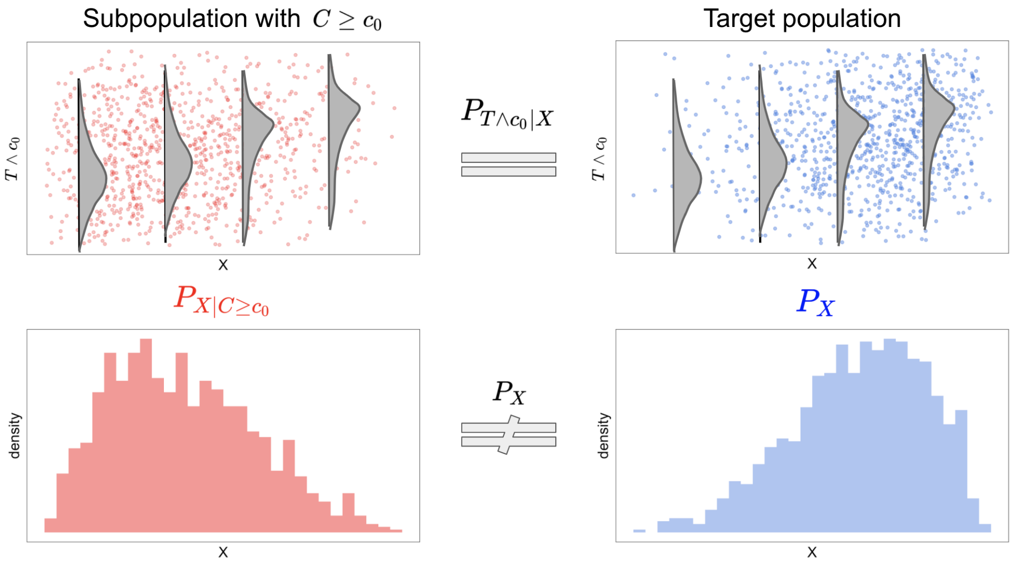


Conformalized survival analysis

$$\mathbb{P}(T \wedge c_0 \geq \hat{T}_{\mathrm{lo}}(X)) \geq 90\% \Longrightarrow \mathbb{P}(T \geq \hat{T}_{\mathrm{lo}}(X)) \geq 90\%$$



Covariate shift under conditionally independent censoring $T \perp\!\!\!\perp C \mid X$



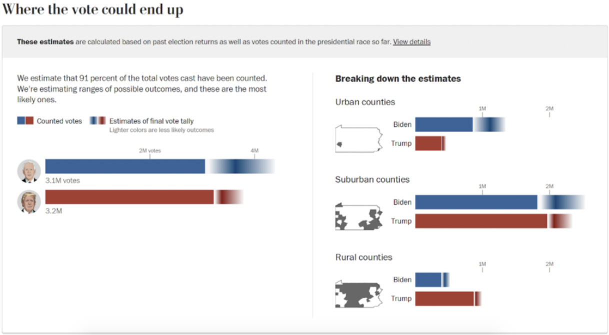
Conformalized survival analysis

- Assumptions:
  - (T_i, C_i, X_i) are i.i.d.
  - Conditionally independent censoring ($T \perp\!\!\!\perp C \mid X$)
  - Type-I censoring (also useful beyond this setting)
- Finite-sample validity: $\hat{T}_{\mathrm{lo}}(X)$ is valid if $P(C \mid X)$ is known
- Double robustness: $\hat{T}_{\mathrm{lo}}(X)$ is approximately valid if $P(C \mid X)$ or $P(T \mid X)$ is estimated well

Part III: what else?

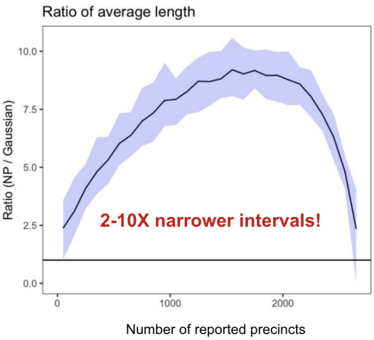
Election night model: prediction intervals for aggregated outcomes

The Washington Post Election Night Model



Georgia Senate Runoff Election

Parametric conformal inference

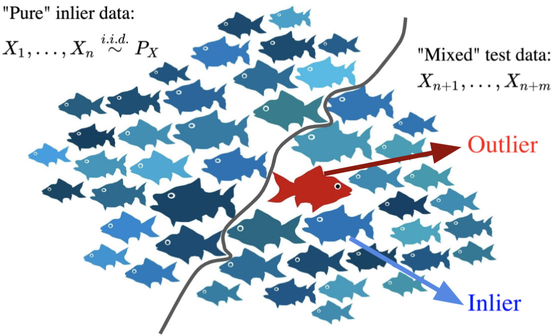


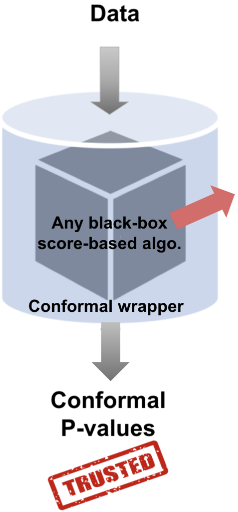
John Cherian

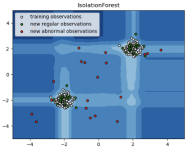
Lenny Bronner

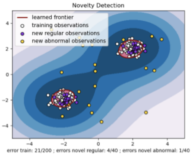
$$\mathbb{P}\left(\underbrace{\textstyle\sum_{i:\text{precincts}} \text{Votes}_i}_{\text{aggregated outcome}} \in \hat{C}(\{X_i\}_{\text{precincts}})\right) \geq 90\%$$

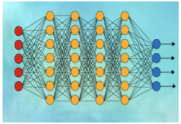
Outlier detection with conformal p-values











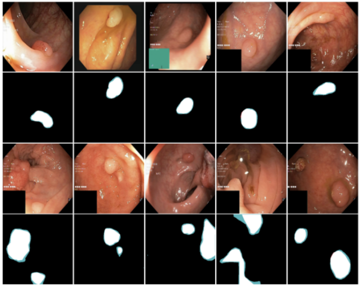
Stephen Bates

Emmanuel Candès

Yaniv Romano

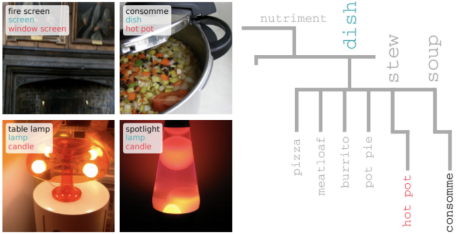
Matteo Sesia

Risk calibrated prediction



teal: false positive
black: true negative
white: true positive
red: false negative

Tumor/Polyp detection



black: correct label
red: raw output
teal: our output

Hierarchical classification


Anastasios Angelopoulos


Stephen Bates

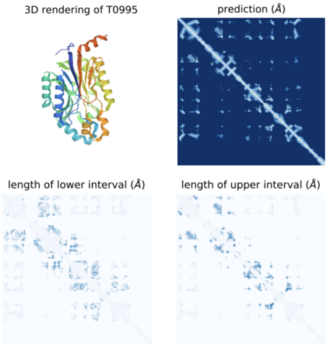

Emmanuel Candès


Michael Jordan


Jitendra Malik

$$\mathbb{P}\left(\underbrace{\mathbb{E}[L(Y,\hat{Y})]}_{\text{risk}} \leq \gamma\right) \geq 90\%$$

Risk calibrated prediction



Protein structure prediction

Object detection

Anastasios Angelopoulos

Stephen Bates

Emmanuel Candès

Michael Jordan

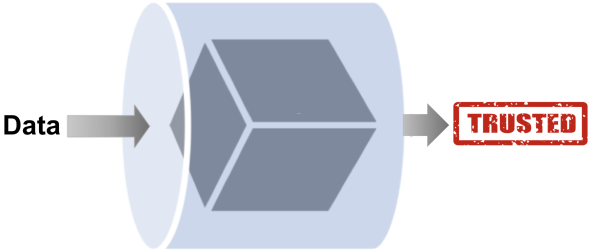
Jitendra Malik

$$\mathbb{P}\left(\underbrace{\mathbb{E}[L(Y,\hat{Y})]}_{\text{risk}} \leq \gamma\right) \geq 90\%$$

What can conformal inference offer to statistics?

What can conformal inference offer to statistics? A LOT!

- Causal inference
- Time-to-event analysis
- Election night model
- Outlier detection
- Risk calibration
- ...



Valid inference under partial/complete misspecification!

Long-term career goal: principles for inference under misspecification

- **Network**: (hierarchical) clustering under misspecified SBMs
  w/ Tianxi Li, Sharmo Bhattacharyya, Purna Sarkar, Peter Bickel, Liza Levina, Xiaodong Li, Xingmei Lou
- **Multiple testing**: FDR control with side information
  w/ Will Fithian, Aaditya Ramdas, Chiao-Yu Yang, Nhat Ho, Yixiang Luo
- **Causal inference**:
  - **Randomized experiments**: model-free regression adjustment
    w/ Peng Ding
  - **Observational studies**: distribution-free assessment of overlap
    w/ Alex D'Amour, Peng Ding, Avi Feller, Jas Sekhon
- **High-d inference**: finite-sample valid test for linear models with exchangeable errors
  w/ Peter Bickel
- **Econometrics**: panel data analysis under heterogeneous treatment effects
  w/ Dmitry Arkhangelsky, Guido Imbens, Xiaoman Luo

All models are wrong, but some are (hopefully) useful

All models are wrong, but we can make them safe and useful!

All models are wrong, but we can make them safe and useful!

Thank you!

Check out my CV and other works on my website!

lihualei71.github.io