Conformal Causal Inference

Ruocheng Guo

TikTok (London, UK)

Introduction

- Point estimate of outcome is often not enough for decision making in high-stake applications.

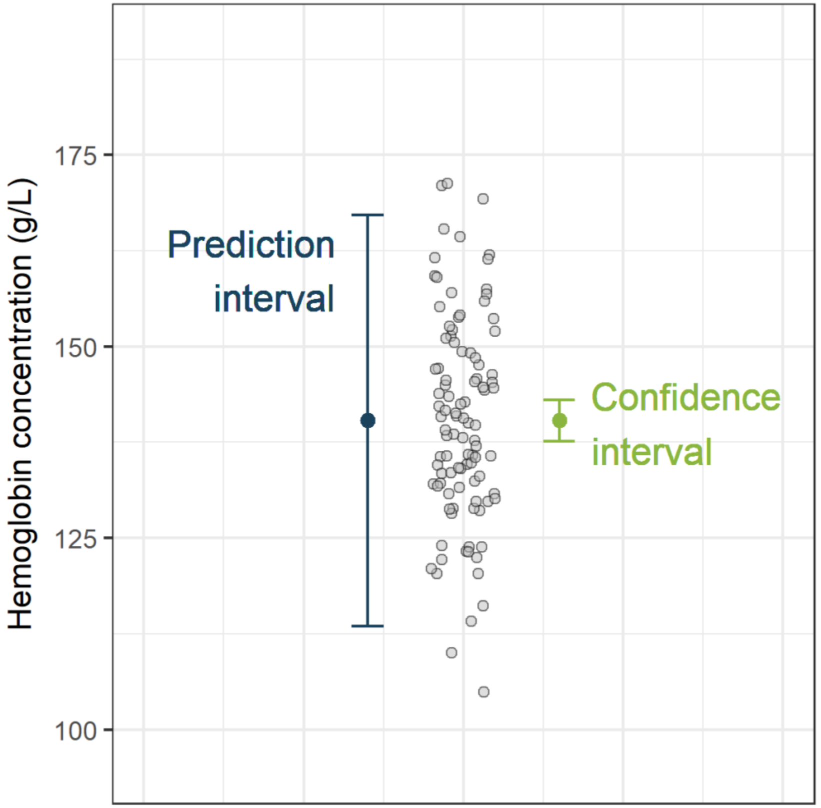
Example

A confidence interval, or at least a p-value, is required by the U.S. Food and Drug Administration to approve a drug, in order to guarantee sufficient evidence and confidence in favor of the drug [1].

[1] Lei, Lihua, and Emmanuel J. Candès. "Conformal inference of counterfactuals and individual treatment effects." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 83.5 (2021): 911-938.

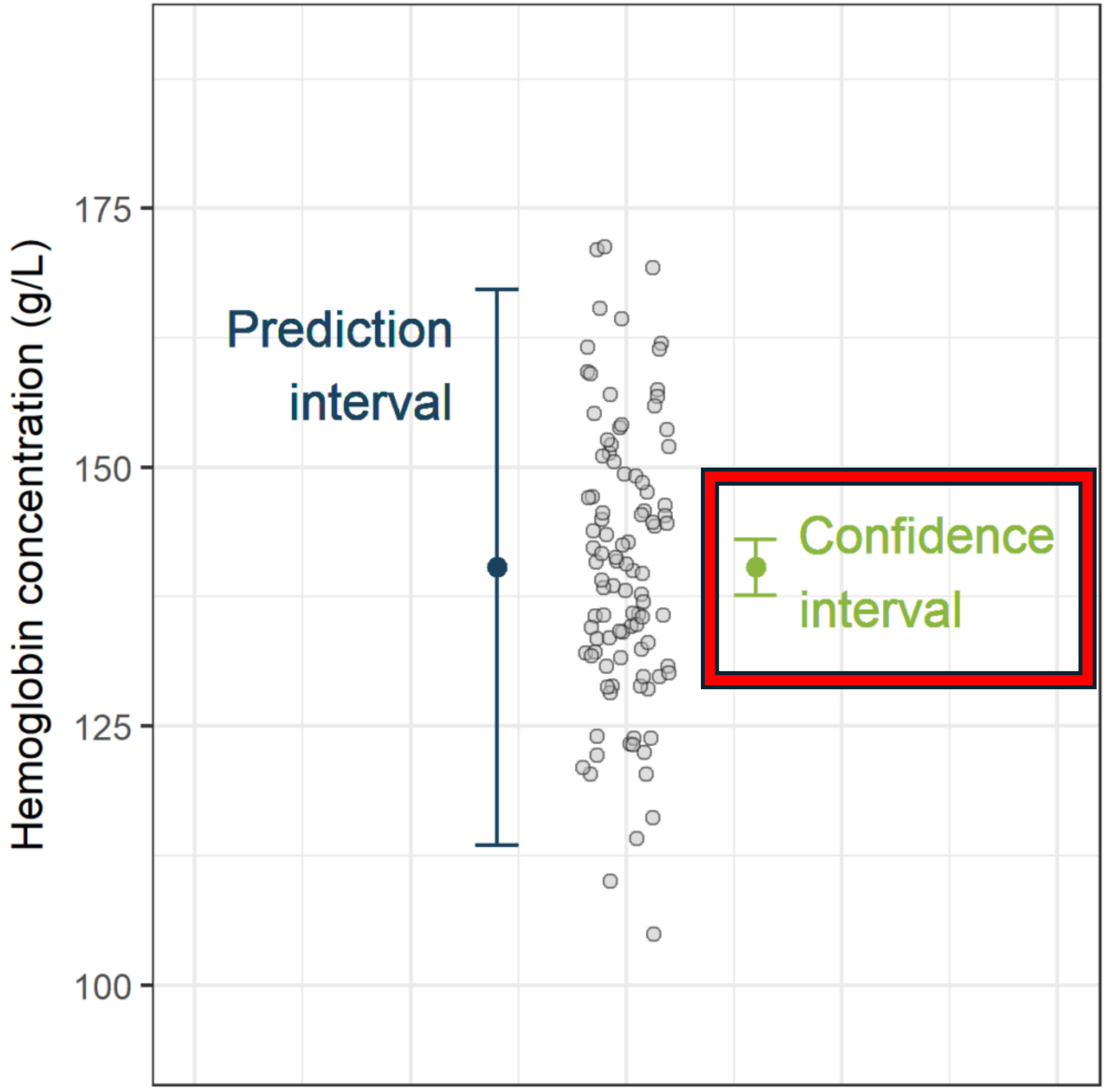
Uncertainty Quantification Basics: confidence interval and prediction interval

Confidence interval vs prediction interval



Confidence interval

- Interval for the true population parameter (e.g., mean)
- 95% confidence interval for the mean:
  - if we repeat our sampling process many times, 95% of the constructed confidence intervals would contain the true population mean.



Confidence interval

- How to compute?
  - Compute mean $\bar{y}$
  - Estimate standard error of the mean (SEM)
    - $SEM = \frac{s}{\sqrt{n}}$
    - Where s is the standard deviation and n is the sample size
  - Compute t value
    - t value is a function of confidence level (95%) and degree of freedom $(n-1)$, obtained by looking up the t value table.
    - The t-value tells us how many standard errors (SEM) we need to go above and below the sample mean to form our confidence interval.
    - With smaller degrees of freedom (smaller sample sizes), the t-value is larger, reflecting greater uncertainty in the estimate.
  - Get confidence interval
    - $CI = \bar{y} \pm t \cdot SEM$

Compute t value

- Suppose you have a sample size $n = 10$:
- Degrees of freedom $df = n - 1 = 10 - 1 = 9$.
- For a 95% confidence interval (two-tailed), look up the t-value for 9 degrees of freedom at the **0.025** (2.5%) significance level in each tail (total 5% for both tails).
- The table or software would give a t-value around 2.262 for 95% confidence with 9 degrees of freedom.

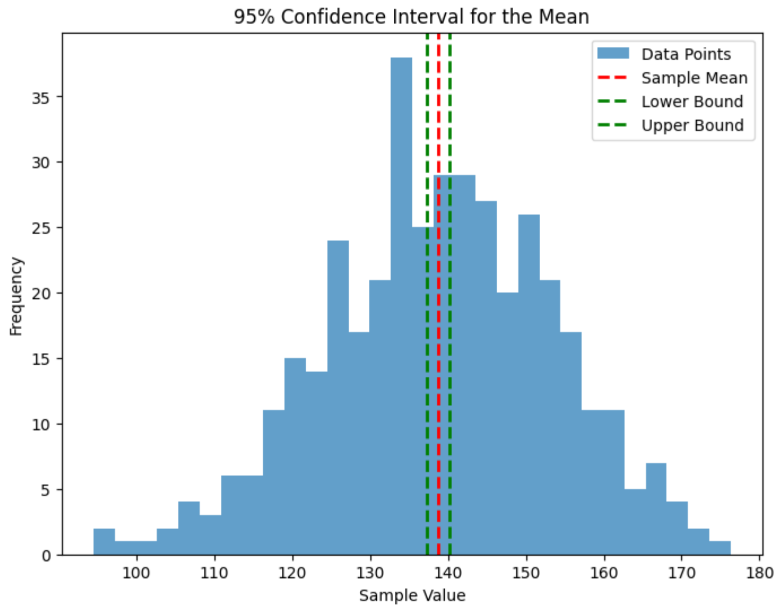
Demo

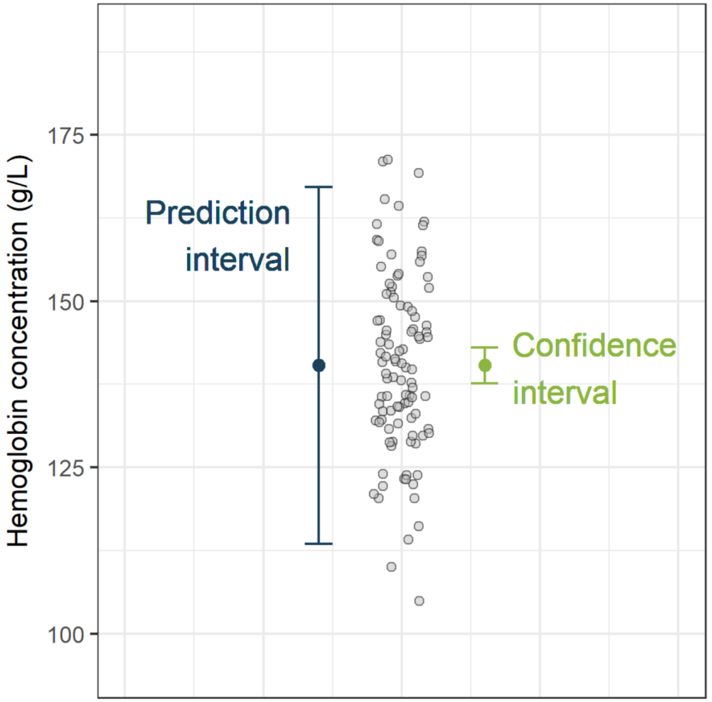
```python
# Generate 400 random samples from the Gaussian distribution
samples = np.random.normal(loc=mean, scale=sd, size=400)

# Calculate the sample mean and standard deviation
sample_mean = np.mean(samples)
sample_sd = np.std(samples, ddof=1)  # Use ddof=1 for sample standard deviation

# Calculate the standard error of the mean
sem = sample_sd / np.sqrt(len(samples))

# Degrees of freedom
df = len(samples) - 1

# Calculate the t-value for a 95% confidence interval (two-tailed)
# We use the percent point function (ppf) from the t-distribution
# to find the critical t-value.
# For a 95% confidence interval, we need the value that cuts off 2.5%
# in each tail (100% - 95% = 5%, so 5%/2 = 2.5% in each tail).
# Using a large number of samples to get an accurate t-value
t_value = np.abs(np.round(np.percentile(np.random.standard_t(df, size=100000), 97.5),3))

# Calculate the margin of error
margin_of_error = t_value * sem

# Calculate the confidence interval
confidence_interval = (sample_mean - margin_of_error, sample_mean + margin_of_error)
```

https://colab.research.google.com/drive/122Sf62ncPgNw1UTbEIRxoqRi_vdJ-jV7?usp=sharing

Demo

- Sample Mean: 138.7896358153267
- Sample Standard Deviation: 14.743598764414 23
- Standard Error of the Mean: 0.7371799382207115
- t-value (95% confidence interval): 1.959
- Margin of Error: 1.4441354989743738
- Confidence Interval: (137.34550031635234, 140.23377131430107)



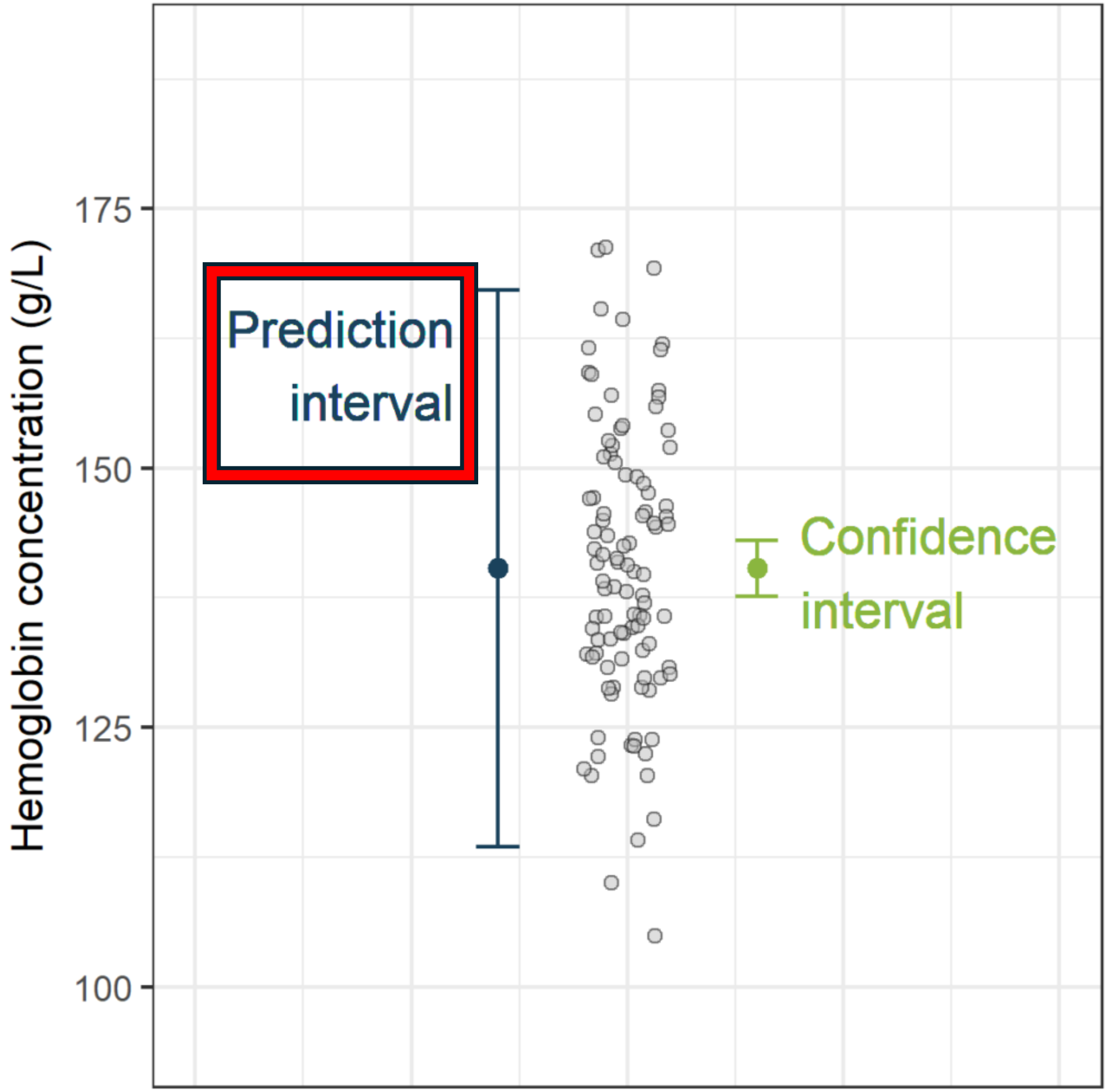


Issues of confidence interval

- Only for population parameters
  - What if I want uncertainty on an individual?
- Assumption
  - Normality: The population is assumed to be **normally distributed** (especially important for small samples).

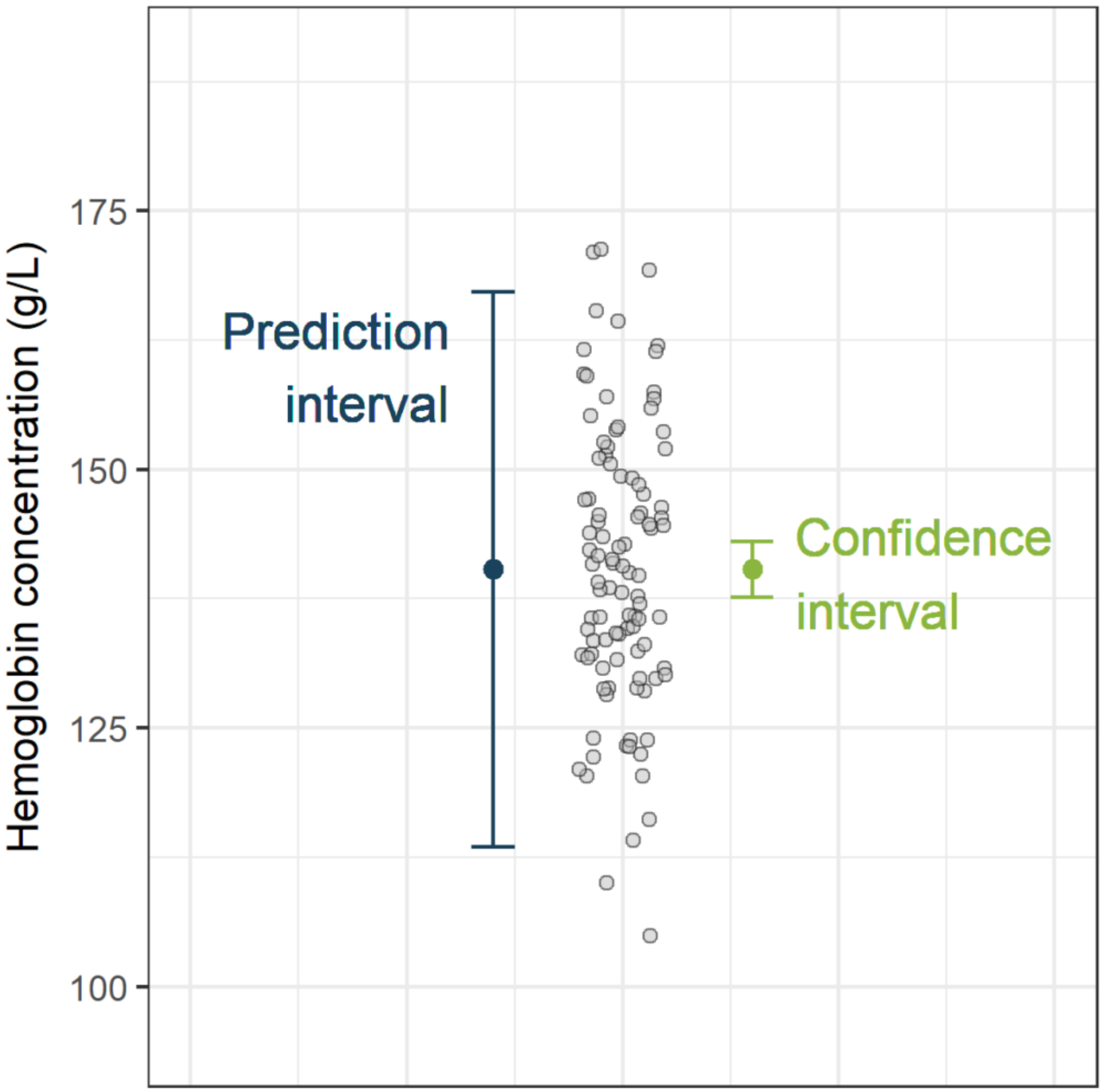
Prediction interval

- Estimates the range within which an individual observation is expected to fall.
- Why is it wider?
  - accounts uncertainty in the estimate of the mean **and the natural variability around individual points**.
- 95% prediction interval for a new observation would mean that there is a 95% chance that the new observation will fall within this interval.



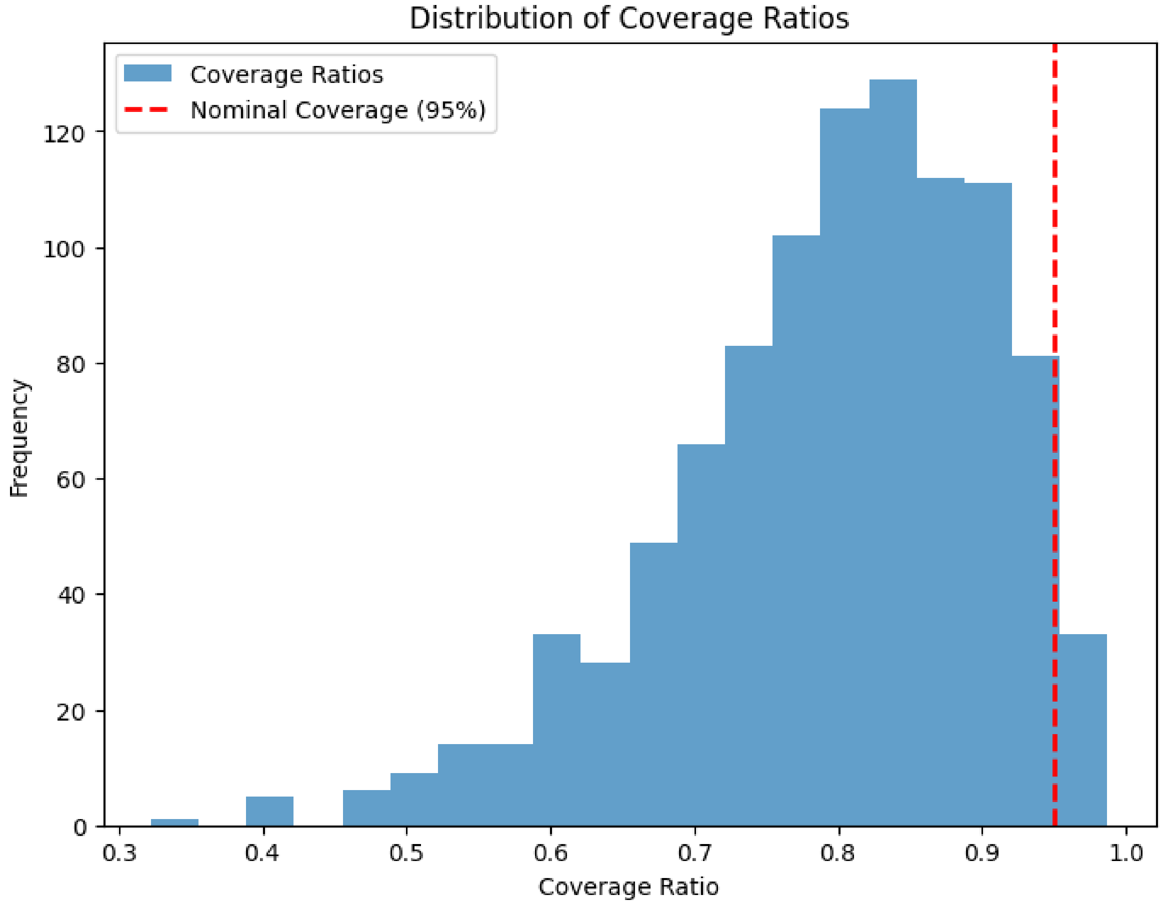
A naïve prediction interval

- Let the 2.5th and 97.5th percentiles of the observations be the lower and upper bounds.
- This assumes a new observation is randomly sampled from the existing observations.



Issue of naïve prediction interval

- In ML, in many cases, we create prediction interval for p(y|x) -> small sample size for each x.
- Small sample size -> unreliable coverage
  - If we only observe 10 data points and use its 95% naïve prediction interval, how likely our interval can cover 95% of the observations?
  - Nominal: 95%
  - What we get: 79.65%

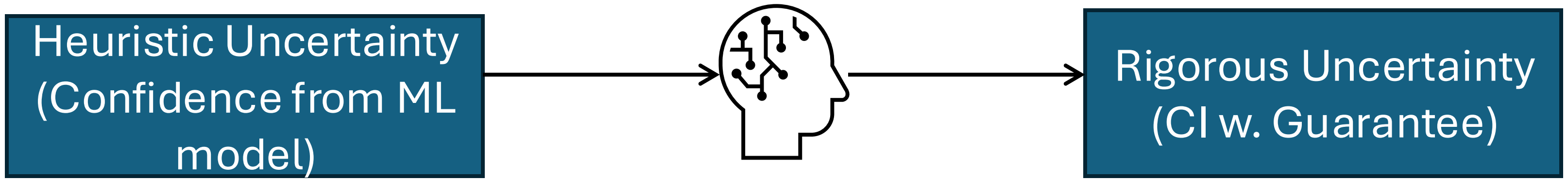


Average Coverage Ratio: 0.7965275

https://colab.research.google.com/drive/122Sf62ncPgNw1UTbEIRxoqRi_vdJ-jV7?usp=sharing

Conformal prediction

Conformal Prediction (CP)



- CP Predicts predictive interval that has **guaranteed probability to cover the ground truth**
  - Weak assumption: exchangeability between calibration and test data, no assumption on data/error distribution
  - Low cost: only need inference on the calibration set
  - Model agnostic: works with any ML models

How to construct a conformal interval?

- Assume we have
  - a pre-trained regression model
  - a calibration dataset with ground truth labels (exchangeable with the test set, for simplicity, you can understand it as the i.i.d. assumption)
  - a predefined coverage rate $1 - \alpha$

How to construct a conformal interval?

- Make predictions on a calibration dataset $\{\widehat{y}_i\}_{i=1}^{N}$

How to construct a conformal interval?

- Make predictions on a calibration dataset $\{\hat{y}_i\}_{i=1}^{N}$
- Compute distribution $\hat{F}$ of nonconformity scores (errors)

$$s_{\mathrm{i}} = |\hat{y}_i - y_i|$$

How to construct a conformal interval?

- Make predictions on a calibration dataset $\{\widehat{y}_i\}_{i=1}^{N}$
- Compute distribution $\hat{F}$ of nonconformity scores (errors)

$$s_{\mathrm{i}} = |\widehat{y}_i - y_i|$$

- Compute the $1 - \alpha$-th quantile $q_{\hat{F}}$ of nonconformity scores $\hat{F}$

How to construct a conformal interval?

- Make predictions on a calibration dataset $\{\widehat{y}_i\}_{i=1}^{N}$
- Compute distribution $\widehat{F}$ of nonconformity scores (errors)

$$\mathrm{s}_{\mathrm{i}} = |\widehat{y}_i - y_i|$$

- Compute the $1 - \alpha$-th quantile $q_{\widehat{F}}$ of nonconformity scores $\widehat{F}$
- Create conformal interval $C_{SCP}(x_i) = [\hat{y}_i - q_{\hat{F}}, \hat{y}_i + q_{\hat{F}}]$

How to construct a conformal interval?

- Make predictions on a calibration dataset $\{\widehat{y}_i\}_{i=1}^{N}$
- Compute distribution $\hat{F}$ of nonconformity scores (errors)

$$s_{\mathrm{i}} = |\widehat{y}_i - y_i|$$

- Compute the $1 - \alpha$-th quantile $q_{\hat{F}}$ of nonconformity scores $\hat{F}$
- Create conformal interval $C_{SCP}(x_i) = [\hat{y}_i - q_{\hat{F}}, \hat{y}_i + q_{\hat{F}}]$
- With exchangeability, test data shares same distribution $\hat{F}$ as calibration data, therefore, $C_{SCP}(x_i)$ has guaranteed marginal coverage rate $1 - \alpha$

(Interview) questions

- What is the difference between confidence interval, prediction interval, and conformal interval?
  - Which of them are for population/individual parameters?
    - A. confidence interval
    - B. naïve prediction interval
    - C. conformal interval

(Interview) questions

- What is the difference between confidence interval, prediction interval, and conformal interval?
  - Which of them are for population/individual parameters?
    - A. confidence interval (population)
    - B. naïve prediction interval (individual)
    - C. conformal interval (individual)

(Interview) questions

- What is the difference between confidence interval, prediction interval, and conformal interval?
  - Which of them **have the different intervals** for different test samples?
    - A. confidence interval
    - B. naïve prediction interval
    - C. conformal interval

(Interview) questions

- What is the difference between confidence interval, prediction interval, and conformal interval?
  - Which of them have **guaranteed coverage**?
    - A. confidence interval
    - B. naïve prediction interval
    - C. conformal interval

(Interview) questions

- What is the difference between confidence interval, prediction interval, and conformal interval?
  - Which of them have **guaranteed coverage**?
    - A. confidence interval
    - B. naïve prediction interval
    - C. conformal interval

Issues of the standard CP

- Only guarantees marginal coverage
- Needs exchangeability between calibration and test data

Marginal coverage

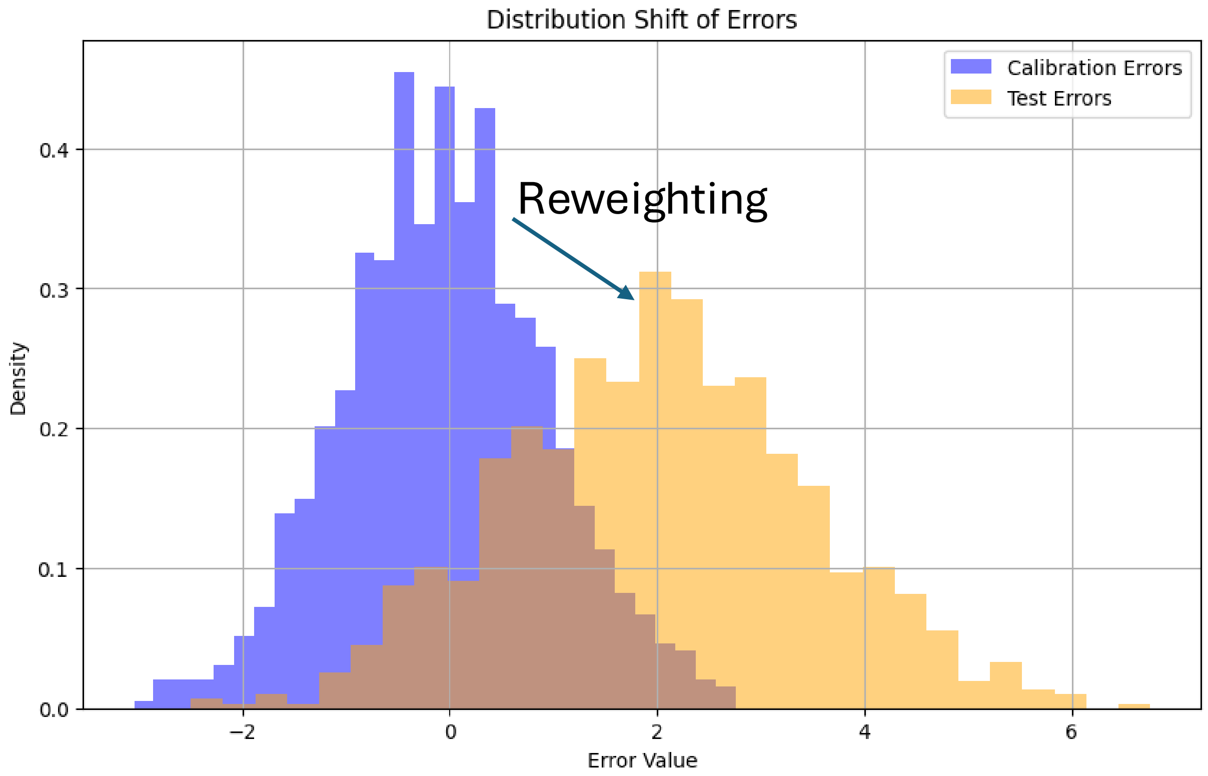
- Marginal coverage (coverage on average)
  - CP only guarantees coverage on average, that is
$$marginal\_coverage = \frac{\sum_i 1(y_i \in C_{SCP}(x_i))}{N}$$
where $1\left(y_i \in C_{SCP}(x_i)\right)$ is the indicator function
  - This implies the coverage might be lower/higher than $1 - \alpha$ if we only consider a subgroup of samples
  - There are recent papers that can achieve conditional coverage [3]

Exchangeability

- It does not hold in causal inference with observational data
  - Distribution shift as calibration set is observational but test set is counterfactual

Exchangeability

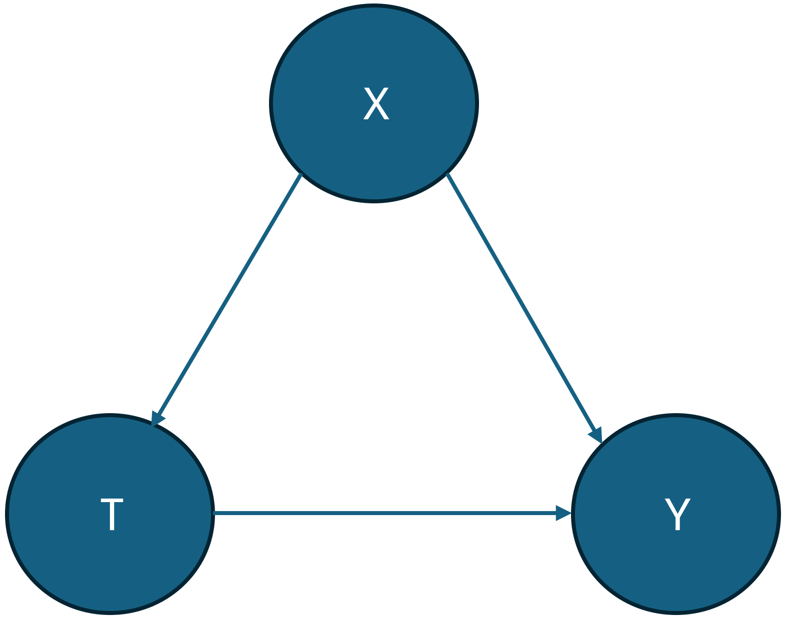
- Fortunately, Lei et al. [1] proposed weighted CP based methods to construct conformal intervals for potential outcomes



Conformal Causal Inference by Reweighting

Causal inference

- What is the effect of the treatment T (pills or surgery) on the outcome Y (recovery rate), given both observed confounding X (severity of disease)?
- For an individual, what is the estimated treatment effect? What is the prediction interval with guaranteed coverage of the estimate?



Weighted CP for causal inference

- Assume strong ignorability and consistency, causal inference is a covariate shift problem:
  - For the treated group $T = 1$, we observe samples from $P(X, Y|T = 1)$, our goal is to estimate CP interval for samples from interventional distribution $Q(X)P(Y(1)|X)$, e.g., $Q(X) = P(X)$ is the marginal distribution of X

Weighted CP for causal inference

- Assume strong ignorability and consistency, causal inference is a covariate shift problem:
  - For the treated group $T = 1$, we observe samples from $P(X, Y|T = 1)$, our goal is to estimate CP interval for samples from interventional distribution $Q(X)P(Y(1)|X)$, e.g., $Q(X) = P(X)$ is the marginal distribution of X
  - Decompose observational distribution as

$$P(X, Y|T = 1) = P(X|T = 1)P(Y|X, T = 1) = P(X|T = 1)P(Y(1)|X)$$

Weighted CP for causal inference

- Assume strong ignorability and consistency, causal inference is a covariate shift problem:
  - For the treated group $T = 1$, we observe samples from $P(X, Y|T = 1)$, our goal is to estimate CP interval for samples from interventional distribution $Q(X)P(Y(1)|X)$, e.g., $Q(X) = P(X)$ is the marginal distribution of X
  - Decompose observational distribution as

$$P(X, Y|T = 1) = P(X|T = 1)P(Y|X, T = 1) = P(X|T = 1)P(Y(1)|X)$$

Weighted CP for causal inference

- Assume strong ignorability and consistency, causal inference is a covariate shift problem:
  - For the treated group $T = 1$, we observe samples from $P(X, Y|T = 1)$, our goal is to estimate CP interval for samples from interventional distribution $Q(X)P(Y(1)|X)$, e.g., $Q(X) = P(X)$ is the marginal distribution of X
  - Decompose observational distribution as

$$P(X, Y|T = 1) = P(X|T = 1)P(Y|X, T = 1) = P(X|T = 1)P(Y(1)|X)$$

  - So, the difference is only covariate shift between $\mathrm{P}(X)$ and $P(X|T = 1)$

Weighted CP for causal inference

- Assume strong ignorability and consistency, causal inference is a covariate shift problem:
  - For the treated group $T = 1$, we observe samples from $P(X, Y|T = 1)$, our goal is to estimate CP interval for samples from interventional distribution $Q(X)P(Y(1)|X)$, e.g., $Q(X) = P(X)$ is the marginal distribution of X
  - Decompose observational distribution as

$$P(X, Y|T = 1) = P(X|T = 1)P(Y|X, T = 1) = P(X|T = 1)P(Y(1)|X)$$

  - So, the difference is only covariate shift between $\mathrm{P}(X)$ and $P(X|T = 1)$
  - This guides us to reweight the nonconformity scores using

$$w_{T=1}(x) = \frac{dP(X)}{dP(X|T = 1)} = \frac{P(T = 1)}{P(T = 1|X)}$$

    - Where $P(T = 1)$ is constant and $P(T = 1|X)$ is propensity score

Weighted CP for causal inference

- Assume strong ignorability and consistency, causal inference is a covariate shift problem:
  - For the treated group $T = 1$, we observe samples from $P(X, Y|T = 1)$, our goal is to estimate CP interval for samples from interventional distribution $Q(X)P(Y(1)|X)$, e.g., $Q(X) = P(X)$ is the marginal distribution of X
  - Decompose observational distribution as
    $$P(X, Y|T = 1) = P(X|T = 1)P(Y|X, T = 1) = P(X|T = 1)P(Y(1)|X)$$
  - So, the difference is only covariate shift between $\mathrm{P}(X)$ and $P(X|T = 1)$
  - This guides us to reweight the nonconformity scores using
    $$w_{T=1}(x) = \frac{dP(X)}{dP(X|T = 1)} = \frac{P(T = 1)}{P(T = 1|X)}$$
    - Where $P(T = 1)$ is constant and $P(T = 1|X)$ is propensity score
  - If $Q(X)$ is different from $P(X)$

Weighted CP for potential outcomes (stage 1)

- Stage 1 gets conformal intervals on calibration data
- Steps different from standard CP are highlighted

Weighted CP for potential outcomes (stage 1)

- Stage 1 gets conformal intervals on calibration data
- Steps different from standard CP are highlighted
- Obtain distribution $\widehat{F}'$ of nonconformity scores (errors) $s_{\mathrm{i}} = |\hat{y}_i - y_i|$

Weighted CP for potential outcomes (stage 1)

- Stage 1 gets conformal intervals on calibration data
- Steps different from standard CP are highlighted
- Obtain distribution $\widehat{F}'$ of nonconformity scores (errors) $s_i = |\hat{y}_i - y_i|$
- Reweight $\widehat{F}'$ to estimate distribution of nonconformity scores on interventional data $\hat{F} = \sum_{i \in \mathcal{D}_{cal}} p_i \delta_{s_i}$
  - Where δ_{s_i} is the density of s_i and $p_i = \frac{w(x_i)}{\sum_i w(x_i)}$ is the normalized weight

Weighted CP for potential outcomes (stage 1)

- Stage 1 gets conformal intervals on calibration data
- Steps different from standard CP are highlighted
- Obtain distribution $\widehat{F'}$ of nonconformity scores (errors) $s_{\mathrm{i}} = |\hat{y}_i - y_i|$
- Reweight $\widehat{F'}$ to estimate distribution of nonconformity scores on interventional data $\hat{F} = \sum_{i \in \mathcal{D}_{cal}} p_i \delta_{s_i}$
    - Where δ_{s_i} is the density of s_i and $p_i = \frac{w(x_i)}{\sum_i w(x_i)}$ is the normalized weight
- Compute the $1 - \alpha$-th quantile $q_{\hat{F}}$ of weighted nonconformity scores

Weighted CP for potential outcomes (stage 1)

- Stage 1 gets conformal intervals on calibration data
- Steps different from standard CP are highlighted
- Obtain distribution $\widehat{F'}$ of nonconformity scores (errors) $s_\mathrm{i} = |\widehat{y_i} - y_i|$
- Reweight $\widehat{F'}$ to estimate distribution of nonconformity scores on interventional data $\hat{F} = \sum_{i \in \mathcal{D}_{cal}} p_i \delta_{s_i}$
  - Where δ_{s_i} is the density of s_i and $p_i = \frac{w(x_i)}{\sum_i w(x_i)}$ is the normalized weight
- Compute the $1 - \alpha$-th quantile $q_{\hat{F}}$ of weighted nonconformity scores
- Confidence interval for calibration data $C_{SCP}(x_i) = [\hat{y}_i - q_{\hat{F}}, \hat{y}_i + q_{\hat{F}}]$

Weighted CP for potential outcomes (stage 2)

- Stage 2 generalizes CP interval from calibration to test
- Fit ML models to predict lower/upper bounds using datasets $\{(x_i, \hat{y}_i - q_{\hat{F}})\}_{i \in \mathcal{D}_{cal}}$ and $\{(x_i, \hat{y}_i + q_{\hat{F}})\}_{i \in \mathcal{D}_{cal}}$
- Use these models to predict lower/upper bounds for any test sample x_j

Results

- Left: Their methods (Inexact, Exact and Naïve) provide coverage while X-learner and Causal Forest failed to do that.
- Right: Different variants have different length of CI.

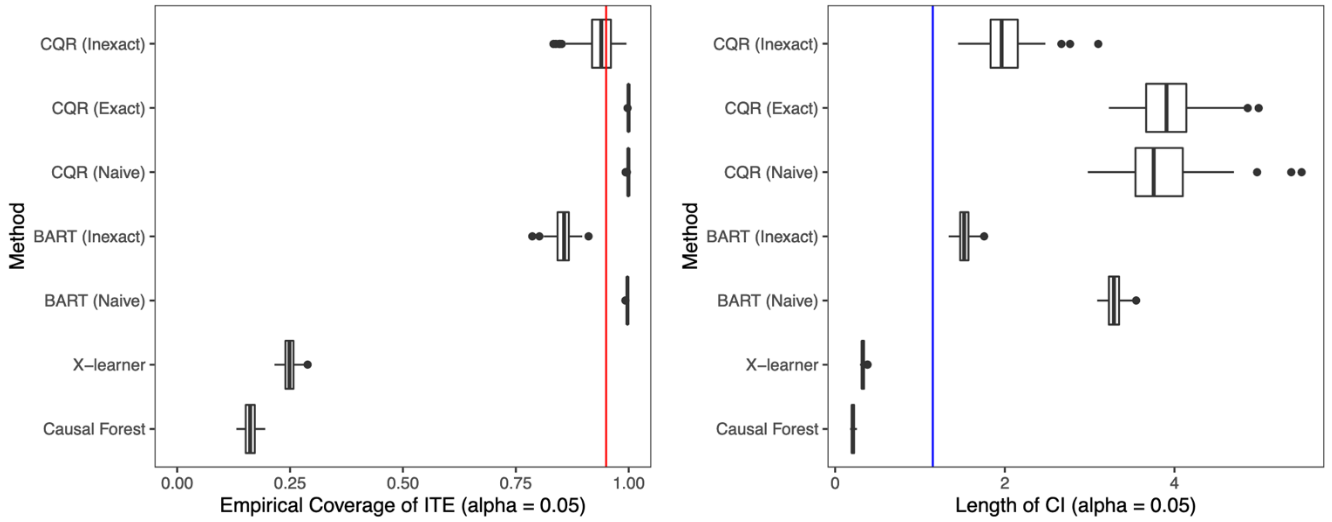


Fig. 5. Coverage (left) and average length (right) of intervals for ITE on synthetic data generated from the NLSM data. The red vertical line corresponds to the target coverage 95% and the blue vertical line corresponds to the average length of oracle intervals.

Take away

- Reweighting nonconformity scores can be used for conformal causal inference under strong ignorability.

Extra content

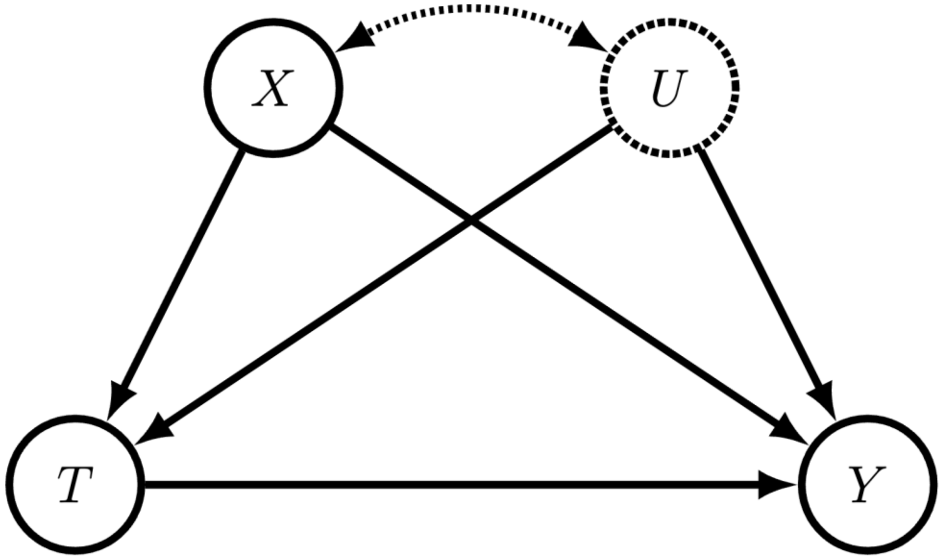
Conformal Counterfactual Inference under Hidden Confounding (KDD’24)

Zonghao Chen[1]*, **Ruocheng Guo[2]***, Jean-Francois Ton[2], Yang Liu[2]
1 UCL
2 ByteDance Research
* Equal contribution

Motivating example

- What is the effect of the treatment T (pills or surgery) on the outcome Y (recovery rate), given both observed confounding X (severity of disease) and unobserved confounding U (patient adherence to treatment)?

- For an individual, what is the estimated treatment effect? What is the confidence interval with guaranteed coverage of the estimate?



Problem

- A merged dataset with n observational and m interventional data ($n >> m$)
  - $\mathcal{D} = \{(x_i, y_i)_{i=1}^{n} \sim P_{\{X,Y\}}\} \cup \{(x_i, y_i)_{i=n+1}^{n+m} \sim P'_{\{X,Y\}}\}$.
- Goal: construct confidence interval $C(x_i)$ with guaranteed coverage rate $1 - \alpha$ for potential outcomes given an unseen test sample $x_i, i > n + m$.

Existing methods

Reweighting by Propensity

Weighted Conformal Prediction (WCP)

Nonconformity score distribution on calibration $\widehat{F_{cal}}$

Nonconformity score distribution on test $\widehat{F_{ts}}$

Observational data

Weighted conformal prediction with biased propensity scores

Coverage

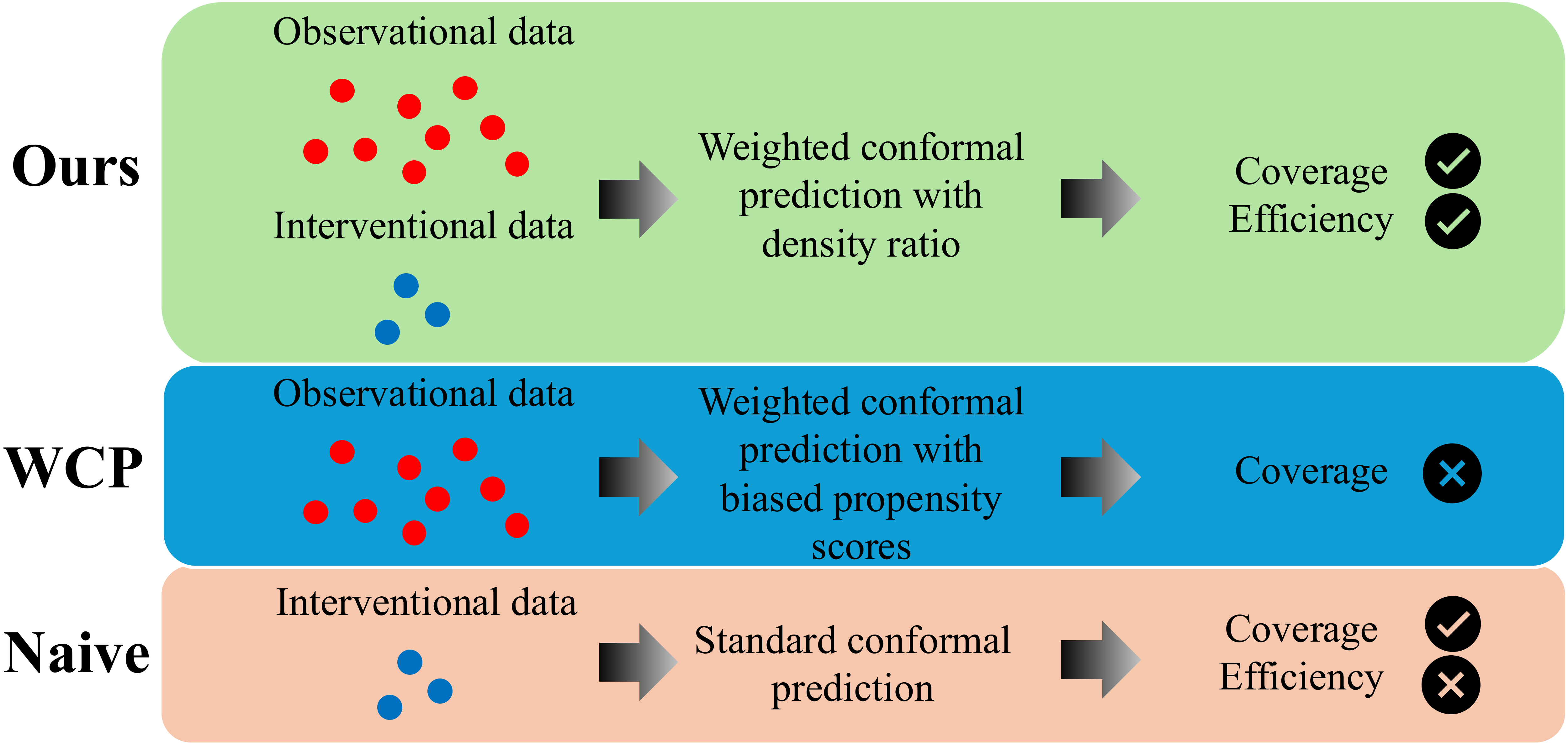
Naive

Interventional data

Standard conformal prediction

Coverage

Efficiency


Ours
Observational data
Interventional data
Weighted conformal prediction with density ratio
Coverage
Efficiency
WCP
Observational data
Weighted conformal prediction with biased propensity scores
Coverage
Naive
Interventional data
Standard conformal prediction
Coverage
Efficiency

Theoretical Results

- Our method has guaranteed coverage.
- Under additive Gaussian noise model, our method is highly likely to have narrower confidence intervals than the naïve method.

Our method (wSCP-DR-Inexact)

- A merged dataset with n observational and m interventional data
  - $\mathcal{D} \ = \ \{(x_i, y_i)_{i=1}^{n} \sim \ P_{\{X,Y\}}\} \cup \{\ (x_i, y_i)_{i=n+1}^{n+m} \sim \ P'_{\{X,Y\}}\}$.
- First, we compute weights, which will be used to handle distribution shift between observational and interventional data.
- Weight functions
  - $w(x, y) = \ 1$ if $(x, y) \sim \ P_{\{X,Y\}}$ and $w(x, y) \ = \ \frac{d\ P_{\{X,Y\}}}{d\ P'_{\{X,Y\}}}(x, y)$ if $(x, y) \ \sim \ P'_{\{X,Y\}}$
- Let ${p_i}_{i=1}^{|\mathcal{D}|}$ denote the "normalized" weight functions.

Our method (wSCP-DR-Inexact)

- Obtain distribution $\widehat{F'}$ of nonconformity scores $s_i = |\hat{y}_i - y_i|$
- Reweight $\widehat{F'}$ to estimate distribution of nonconformity scores on interventional data $\hat{F} = \sum_{i \in \mathcal{D}_{cal}} p_i \delta_{s_i}$
- Compute the $1 - \alpha$-th quantile of nonconformity scores $q_{\hat{F}}$
- Confidence interval for calibration $C_{SCP}(x_i) = [\hat{y}_i - q_{\hat{F}}, \hat{y}_i + q_{\hat{F}}]$
- Fit ML models to predict lower/upper bounds using datasets $\{(x_i, \hat{y}_i - q_{\hat{F}})\}_{i \in \mathcal{D}_{cal}}$ and $\{(x_i, \hat{y}_i + q_{\hat{F}})\}_{i \in \mathcal{D}_{cal}}$
- Use these models to predict lower/upper bounds for any test sample

Experiments

- Datasets
  - Synthetic data with controllable hidden confounding
  - Real-world recommendation datasets (rating prediction)
    - Yahoo!R3
    - Coat
- Evaluation metrics
  - Coverage rate
    - probability of true potential outcome / ITE in the predicted interval
  - Interval width

Results

- Synthetic data with hidden confounding

Table 2: Results for counterfactual outcomes and ITEs on the synthetic data. We compare our methods wSCP-DR (Inexact), wSCP-DR (Inexact), and wTCP-DR with baselines. Results are shown for coverage and confidence interval width on the synthetic data with $n = 10,000$ and $m = 250$. Boldface and underlining are used to highlight the top and second-best interval width among the methods with coverage close to 0.9.

Method	Coverage $Y(0)$ ↑	Interval Width $Y(0)$ ↓	Coverage $Y(1)$ ↑	Interval Width $Y(1)$ ↓	Coverage ITE ↑	Interval Width ITE ↓
wSCP-DR(Inexact)	0.891 ± 0.026	<u>0.414</u> ± 0.008	0.889 ± 0.019	**0.421** ± 0.013	0.942 ± 0.017	**0.835** ± 0.016
wSCP-DR(Exact)	0.934 ± 0.026	0.496 ± 0.010	0.935 ± 0.023	<u>0.503</u> ± 0.010	0.957 ± 0.018	0.998 ± 0.015
wTCP-DR	0.899 ± 0.028	**0.386** ± 0.013	0.923 ± 0.015	0.576 ± 0.066	0.953 ± 0.015	<u>0.962</u> ± 0.074
WCP	0.572 ± 0.039	0.222 ± 0.007	0.608 ± 0.042	0.227 ± 0.009	0.710 ± 0.027	0.449 ± 0.012
Naive	0.932 ± 0.018	0.508 ± 0.042	0.930 ± 0.023	0.560 ± 0.049	0.952 ± 0.018	1.068 ± 0.098

Empirical Results

- Recommendation system data
  - Rating prediction with distribution shift

Table 3: Coverage and interval width results on Yahoo and Coat. Boldface and underlining are used to highlight the top and second-best interval width among the methods with coverage close to 0.9.

	Yahoo		Coat	
Method	Coverage ↑	Interval Width ↓	Coverage ↑	Interval Width ↓
wSCP-DR(Inexact)	0.892 ± 0.019	**4.353** ± 0.019	0.919 ± 0.008	**3.787** ± 0.045
wSCP-DR(Exact)	0.952 ± 0.001	5.140 ± 0.001	0.959 ± 0.001	4.565 ± 0.228
wSCP-DR*(Inexact)	0.892 ± 0.020	**4.353** ± 0.020	0.919 ± 0.008	<u>3.789</u> ± 0.046
wSCP-DR*(Exact)	0.952 ± 0.001	5.140 ± 0.001	0.960 ± 0.001	4.571 ± 0.233
WCP-NB	0.825 ± 0.002	4.036 ± 0.002	0.912 ± 0.005	3.635 ± 0.040
Naive	0.899 ± 0.001	6.047 ± 0.001	0.896 ± 0.003	7.725 ± 0.018

Take away

- We propose a simple yet effective method to handle hidden confounding for conformal counterfactual inference.
- Our method reweights nonconformity scores with density ratio of joint distributions instead of propensity scores (WCP).
- Theoretically, we prove the proposed method guarantees coverage as well as is more efficient than the naïve method.
- Empirically, experimental results support our claims.

Our paper and code can be found at

https://arxiv.org/abs/2405.12387

https://github.com/rguo12/KDD24-Conformal

References

[1] Lihua Lei, Emmanuel J. Candès, Conformal Inference of Counterfactuals and Individual Treatment Effects, Journal of the Royal Statistical Society Series B: Statistical Methodology, Volume 83, Issue 5, November 2021, Pages 911–938, https://doi.org/10.1111/rssb.12445

[2] Angelopoulos, Anastasios N., and Stephen Bates. "A gentle introduction to conformal prediction and distribution-free uncertainty quantification." arXiv preprint arXiv:2107.07511 (2021).

[3] Feldman, Shai, Stephen Bates, and Yaniv Romano. "Improving conditional coverage via orthogonal quantile regression." *Advances in neural information processing systems* 34 (2021): 2060-2071.

[4] Wang, Fangxin, et al. "Equal opportunity of coverage in fair regression." *Advances in Neural Information Processing Systems* 36 (2024).

[5] Chen, Zonghao, et al. "Conformal counterfactual inference under hidden confounding." *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024.