Distribution-free Uncertainty Quantification: Impossibility and Possibility

Part I

Xinmeng Huang

Based on:
Distribution-free binary classification: prediction sets, confidence intervals and calibration
Is distribution-free inference possible for binary regression?

Overview

Motivation: why *distribution-free*?

- Many algorithms assume certain conditions on the underlying distribution.
- Unless we can verify these assumptions, the validity is not guaranteed in practice.
- A *distribution-free* algorithm makes no assumption about the distribution, and thus is valid universally for any underlying distribution.

Overview

Conformal Prediction

- Setting: observations $(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}$ with $(X_i, Y_i) \overset{i.i.d}{\sim} P$ for $i \in [n]$.

- Goal: Construct a data-dependent set $\hat{S}_n : \mathbb{R}^d \to \{\text{subsets of } \mathbb{R}\}$ using $(X_i, Y_i)_{i=1}^n$ such that for a new $(X_{n+1}, Y_{n+1}) \sim P$

$$\mathbb{P}\left(Y_{n+1} \in \hat{S}_n(X_{n+1})\right) \geq 1 - \alpha \quad \text{for all distributions } P \text{ on } \mathbb{R}^d \times \mathbb{R}.$$

$$\begin{array}{ccccc} \text{Training data} & & & & \text{Prediction set} \\ (X_i, Y_i)_{i=1}^n & \rightsquigarrow & \text{Algorithm } \hat{S}_n & \rightsquigarrow & \{\hat{S}_n(x)\}_{x \in \mathbb{R}^d} \end{array}$$

- We call such $\hat{S}_n$ a $(1-\alpha)$-DF (distribution-free) prediction set.

Conformal Prediction

- Setting: observations $(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}$ with $(X_i, Y_i) \overset{i.i.d}{\sim} P$ for $i \in [n]$.
- Goal: Construct a data-dependent set $\hat{S}_n : \mathbb{R}^d \to \{\text{subsets of } \mathbb{R}\}$ using $(X_i, Y_i)_{i=1}^n$ such that for a new $(X_{n+1}, Y_{n+1}) \sim P$

$$\mathbb{P}\left(Y_{n+1} \in \hat{S}_n(X_{n+1})\right) \geq 1 - \alpha \quad \text{for all distributions } P \text{ on } \mathbb{R}^d \times \mathbb{R}.$$

$$\begin{array}{c} \text{Training data} \\ (X_i, Y_i)_{i=1}^n \end{array} \rightsquigarrow \text{Algorithm } \hat{S}_n \rightsquigarrow \begin{array}{c} \text{Prediction set} \\ \{\hat{S}_n(x)\}_{x \in \mathbb{R}^d} \end{array}$$

- We call such $\hat{S}_n$ a $(1 - \alpha)$-DF (distribution-free) prediction set.

Discrete Labels

- $(1-\alpha)$-DF prediction set is well addressed by conformal prediction methods built on the rank statistics and exchangeability.

- In modern ML tasks, labels are usually discretized, where $(1-\alpha)$-DF prediction set may be pointless.

Example

Suppose $Y \in \{0,1\}$ and $Y \mid X = x \sim \text{Ber}(\pi(x))$ with $\pi(x) \in (\alpha, 1-\alpha)$, then any $(1-\alpha)$-DF prediction set has no choice but $\hat{S}_n = \{0,1\}$.

Discrete Labels

- $(1 - \alpha)$-DF prediction set is well addressed by conformal prediction methods built on the rank statistics and exchangeability.
- In modern ML tasks, labels are usually discretized, where $(1 - \alpha)$-DF prediction set may be pointless.

Example

Suppose $Y \in \{0, 1\}$ and $Y \mid X = x \sim$ Ber$(\pi(x))$ with $\pi(x) \in (\alpha, 1 - \alpha)$, then any $(1 - \alpha)$-DF prediction set has no choice but $\hat{S}_n = \{0, 1\}$.

Overview

1. Motivation
2. Background
3. DF Confidence Interval
4. Calibration
5. Summary

Confidence Interval

- Other candidates for the scenarios with discrete labels?

- A "reasonable" candidate is to estimate the conditional mean.

- Setting: Feature $X \in \mathbb{R}^d$ and label $Y \in \{0,1\}$, an unknown distribution $P = P_X \times \text{Ber}(\pi_P(X))$

  Marginal distrib. of X $\qquad$ $\pi_P(x) = \mathbb{P}(Y=1 \mid X=x) = \mathbb{E}_P[Y \mid X=x]$

- Goal: Construct a data-dependent interval $\hat{C}_n : \mathbb{R}^d \to \{\text{intervals of } \mathbb{R}\}$ using $(X_i, Y_i)_{i=1}^n$ such that for a new $(X_{n+1}, Y_{n+1}) \sim P$

$$\mathbb{P}\left(\pi_P(X_{n+1}) \in \hat{C}_n(X_{n+1})\right) \geq 1-\alpha \quad \text{for all distributions } P \text{ on } \mathbb{R}^d \times \{0,1\}.$$

- We call such $\hat{C}_n$ as a $(1-\alpha)$-DF (distribution-free) confidence interval.

Confidence Interval

- Other candidates for the scenarios with discrete labels?
- A "reasonable" candidate is to estimate the conditional mean.
- Setting: Feature $X \in \mathbb{R}^d$ and label $Y \in \{0,1\}$, an unknown distribution $P = \underset{\substack{\uparrow \\ \text{Marginal distrib. of } X}}{P_X} \times \text{Ber}(\pi_P(X))$

  $\pi_P(x) = \mathbb{P}(Y = 1 \mid X = x) = \mathbb{E}_P[Y \mid X = x]$
- Goal: Construct a data-dependent interval $\hat{C}_n : \mathbb{R}^d \to \{\text{intervals of } \mathbb{R}\}$ using $(X_i, Y_i)_{i=1}^n$ such that for a new $(X_{n+1}, Y_{n+1}) \sim P$

  $$\mathbb{P}\left(\pi_P(X_{n+1}) \in \hat{C}_n(X_{n+1})\right) \geq 1-\alpha \quad \text{for all distributions } P \text{ on } \mathbb{R}^d \times \{0,1\}.$$
- We call such $\hat{C}_n$ as a $(1-\alpha)$-DF (distribution-free) confidence interval.

Confidence Interval

- Other candidates for the scenarios with discrete labels?
- A "reasonable" candidate is to estimate the conditional mean.
- Setting: Feature $X \in \mathbb{R}^d$ and label $Y \in \{0,1\}$, an unknown distribution $P = P_X \times \text{Ber}(\pi_P(X))$

  Marginal distrib. of X $\qquad$ $\pi_P(x) = \mathbb{P}(Y = 1 \mid X = x) = \mathbb{E}_P[Y \mid X = x]$

- Goal: Construct a data-dependent interval $\hat{C}_n : \mathbb{R}^d \to$ {intervals of $\mathbb{R}$} using $(X_i, Y_i)_{i=1}^n$ such that for a new $(X_{n+1}, Y_{n+1}) \sim P$

$$\mathbb{P}\left(\pi_P(X_{n+1}) \in \hat{C}_n(X_{n+1})\right) \geq 1 - \alpha \quad \text{for all distributions } P \text{ on } \mathbb{R}^d \times \{0,1\}.$$

- We call such $\hat{C}_n$ as a $(1-\alpha)$-DF (distribution-free) confidence interval.

Confidence Interval

- Other candidates for the scenarios with discrete labels?
- A "reasonable" candidate is to estimate the conditional mean.
- Setting: Feature $X \in \mathbb{R}^d$ and label $Y \in \{0,1\}$, an unknown distribution $P = P_X \times \text{Ber}(\pi_P(X))$

  Marginal distrib. of X $\qquad$ $\pi_P(x) = \mathbb{P}(Y = 1 \mid X = x) = \mathbb{E}_P[Y \mid X = x]$

- Goal: Construct a data-dependent interval $\hat{C}_n : \mathbb{R}^d \to$ {intervals of $\mathbb{R}$} using $(X_i, Y_i)_{i=1}^n$ such that for a new $(X_{n+1}, Y_{n+1}) \sim P$

$$\mathbb{P}\left(\pi_P(X_{n+1}) \in \hat{C}_n(X_{n+1})\right) \geq 1-\alpha \quad \text{for all distributions } P \text{ on } \mathbb{R}^d \times \{0,1\}.$$

- We call such $\hat{C}_n$ as a $(1-\alpha)$-DF (distribution-free) confidence interval.

Confidence Interval

- Other candidates for the scenarios with discrete labels?
- A "reasonable" candidate is to estimate the conditional mean.
- Setting: Feature $X \in \mathbb{R}^d$ and label $Y \in \{0,1\}$, an unknown distribution $P = P_X \times \text{Ber}(\pi_P(X))$

  Marginal distrib. of X $\qquad$ $\pi_P(x) = \mathbb{P}(Y = 1 \mid X = x) = \mathbb{E}_P[Y \mid X = x]$

- Goal: Construct a data-dependent interval $\hat{C}_n : \mathbb{R}^d \to \{\text{intervals of } \mathbb{R}\}$ using $(X_i, Y_i)_{i=1}^n$ such that for a new $(X_{n+1}, Y_{n+1}) \sim P$

  $$\mathbb{P}\left(\pi_P(X_{n+1}) \in \hat{C}_n(X_{n+1})\right) \geq 1-\alpha \quad \text{for all distributions } P \text{ on } \mathbb{R}^d \times \{0,1\}.$$

- We call such $\hat{C}_n$ as a $(1-\alpha)$-DF (distribution-free) confidence interval.

Confidence Interval

- Other candidates for the scenarios with discrete labels?
- A "reasonable" candidate is to estimate the conditional mean.
- Setting: Feature $X \in \mathbb{R}^d$ and label $Y \in \{0,1\}$, an unknown distribution $P = \underbrace{P_X}_{\text{Marginal distrib. of } X} \times \text{Ber}(\underbrace{\pi_P(X)}_{\pi_P(x) = \mathbb{P}(Y=1 \mid X=x) = \mathbb{E}_P[Y \mid X=x]})$

  Marginal distrib. of X $\quad$ $\pi_P(x) = \mathbb{P}(Y = 1 \mid X = x) = \mathbb{E}_P[Y \mid X = x]$
- Goal: Construct a data-dependent interval $\hat{C}_n : \mathbb{R}^d \to \{\text{intervals of } \mathbb{R}\}$ using $(X_i, Y_i)_{i=1}^n$ such that for a new $(X_{n+1}, Y_{n+1}) \sim P$

$$\mathbb{P}\left(\pi_P(X_{n+1}) \in \hat{C}_n(X_{n+1})\right) \geq 1-\alpha \quad \text{for all distributions } P \text{ on } \mathbb{R}^d \times \{0,1\}.$$

- We call such $\hat{C}_n$ as a $(1-\alpha)$-DF (distribution-free) confidence interval.

Prediction vs Estimation

- Given $X_{n+1} = x$, the goal of
  $$\begin{cases} \text{prediction set: to predict } Y_{n+1}; \\ \text{confidence interval: to estimate } \pi_P(x) \triangleq \mathbb{E}[Y_{n+1} \mid X_{n+1} = x]. \end{cases}$$

- The task of prediction sets requires to capture most of variability of Y.

- Is the task of DF confidence interval easier than that of DF prediction set?

- Does there exist a $(1-\alpha)$-DF confidence interval with vanishing length, i.e., $\text{len}(\hat{C}_n(x)) \to 0$, at least for "nice" distributions P?

- How to construct a $(1-\alpha)$-DF confidence interval?

Prediction vs Estimation

- Given $X_{n+1} = x$, the goal of
  $$\begin{cases} \text{prediction set: to predict } Y_{n+1}; \\ \text{confidence interval: to estimate } \pi_P(x) \triangleq \mathbb{E}[Y_{n+1} \mid X_{n+1} = x]. \end{cases}$$

- The task of prediction sets requires to capture most of variability of Y.

- Is the task of DF confidence interval easier than that of DF prediction set?

- Does there exist a $(1-\alpha)$-DF confidence interval with vanishing length, i.e., $\text{len}(\hat{C}_n(x)) \to 0$, at least for "nice" distributions P?

- How to construct a $(1-\alpha)$-DF confidence interval?

Prediction vs Estimation

- Given $X_{n+1} = x$, the goal of
$$\begin{cases} \text{prediction set: to predict } Y_{n+1}; \\ \text{confidence interval: to estimate } \pi_P(x) \triangleq \mathbb{E}[Y_{n+1} \mid X_{n+1} = x]. \end{cases}$$

- The task of prediction sets requires to capture most of variability of Y.

- Is the task of DF confidence interval easier than that of DF prediction set?

- Does there exist a $(1-\alpha)$-DF confidence interval with vanishing length, i.e., len$(\hat{C}_n(x)) \to 0$, at least for "nice" distributions P?

- How to construct a $(1-\alpha)$-DF confidence interval?

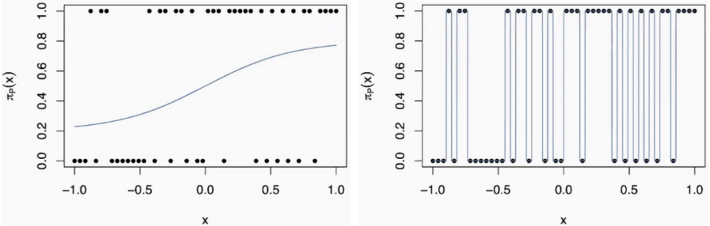
Prediction vs Estimation

- Given $X_{n+1} = x$, the goal of
  $$\begin{cases} \text{prediction set: to predict } Y_{n+1}; \\ \text{confidence interval: to estimate } \pi_P(x) \triangleq \mathbb{E}[Y_{n+1} \mid X_{n+1} = x]. \end{cases}$$
- The task of prediction sets requires to capture most of variability of Y.
- Is the task of DF confidence interval easier than that of DF prediction set?
- Does there exist a $(1-\alpha)$-DF confidence interval with vanishing length, i.e., len$(\hat{C}_n(x)) \to 0$, at least for "nice" distributions P?
- How to construct a $(1-\alpha)$-DF confidence interval?

Challenge of Estimating $\pi_P(X)$

- The challenge is that the observations can be consistent with many distributions.

- When the underlying P is nonatomic, all datapoints (features, $x \in \mathbb{R}^d$) cannot be observed twice.

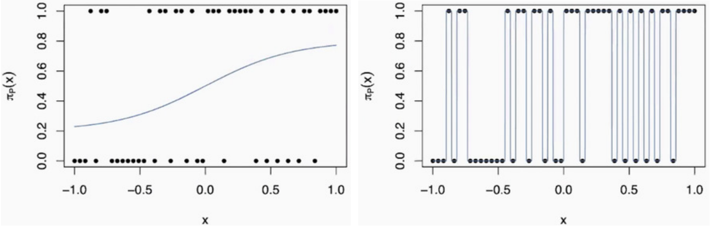


- Cannot tell if the data is generated from the left or the right curve?

Challenge of Estimating $\pi_P(X)$

- The challenge is that the observations can be consistent with many distributions.

- When the underlying P is nonatomic, all datapoints (features, $x \in \mathbb{R}^d$) cannot be observed twice.



- Cannot tell if the data is generated from the left or the right curve?

C.I. implies P.S.

Theorem

Suppose $\hat{C}_n$ be a $(1-\alpha)$-DF confidence interval. Then $\hat{S}_n := \hat{C}_n \cap \{0,1\}$ is a $(1-\alpha)$ prediction set uniformly over all nonatomic distributions P, i.e.,

$$\mathbb{P}_{(X_i,Y_i)\overset{\text{iid}}{\sim}P}\left(Y_{n+1} \in \hat{S}_n\left(X_{n+1}\right)\right) \geq 1-\alpha \quad \textit{for all nonatomic } P \textit{ on } \mathbb{R}^d \times \{0,1\}.$$

- Nonatomic: no $x \in \mathbb{R}^d$ such that $P(x) > 0$.
- It suffices to show $\hat{C}_n$ itself a is $(1-\alpha)$ prediction set.

C.I. implies P.S.

Theorem

Suppose $\hat{C}_n$ be a $(1-\alpha)$-DF confidence interval. Then $\hat{S}_n := \hat{C}_n \cap \{0,1\}$ is a $(1-\alpha)$ prediction set uniformly over all nonatomic distributions P, i.e.,

$$\mathbb{P}_{(X_i,Y_i)\overset{\text{iid}}{\sim} P}\left(Y_{n+1} \in \hat{S}_n\left(X_{n+1}\right)\right) \geq 1-\alpha \quad \textit{for all nonatomic } P \textit{ on } \mathbb{R}^d \times \{0,1\}.$$

- Nonatomic: no $x \in \mathbb{R}^d$ such that $P(x) > 0$.
- It suffices to show $\hat{C}_n$ itself a is $(1-\alpha)$ prediction set.

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof outline.** The main idea: applying distribution-free validty to specific distributions.
  - Point I: given any distinct observations $\mathcal{L} \triangleq (x_i, y_i)_{i=1}^{n+1}$ from P^{n+1}, $\mathcal{L}$ is consistent with $\text{Unif}_{n+1}(\mathcal{L})$ $\leftarrow$ uniformly draw $n+1$ i.i.d pairs from $\mathcal{L}$
  - Point II: $y_i = \mathbb{E}[Y \mid X = x_i]$ on $\text{Unif}_{n+1}(\mathcal{L})$, i.e., $Y = \pi_{\text{Unif}_{n+1}(\mathcal{L})}(X)$.

  Applying DF-validty of $\hat{C}_n$ over $\text{Unif}_{n+1}(\mathcal{L})$ yields

$$\mathbb{P}_{\text{Unif}_{n+1}(\mathcal{L})}\left(Y_{n+1} \in \hat{C}_n\left(X_{n+1}\right)\right) \geq 1-\alpha. \tag{1}$$

  Finally, marginalize $\mathcal{L} \sim P^{n+1}$ for (1) to get the result for P^{n+1}.
  Marginalizing $\mathcal{L} \sim P^{n+1}$ for $\text{Unif}_{n+1}(\mathcal{L})$ is not P^{n+1}, i.e.,
  $\mathbb{E}_{\mathcal{L}\sim P^{n+1}}[\text{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$!

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof outline.** The main idea: applying distribution-free validty to specific distributions.
  - Point I: given any distinct observations $\mathcal{L} \triangleq (x_i, y_i)_{i=1}^{n+1}$ from P^{n+1}, $\mathcal{L}$ is consistent with $\mathsf{Unif}_{n+1}(\mathcal{L})$ $\leftarrow$ uniformly draw $n+1$ i.i.d pairs from $\mathcal{L}$
  - Point II: $y_i = \mathbb{E}[Y \mid X = x_i]$ on $\mathsf{Unif}_{n+1}(\mathcal{L})$, i.e., $Y = \pi_{\mathsf{Unif}_{n+1}(\mathcal{L})}(X)$.

  Applying DF-validty of $\hat{C}_n$ over $\mathsf{Unif}_{n+1}(\mathcal{L})$ yields

$$\mathbb{P}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left(Y_{n+1} \in \hat{C}_n\left(X_{n+1}\right)\right) \geq 1-\alpha.$$

  Finally, marginalize $\mathcal{L} \sim P^{n+1}$ for (1) to get the result for P^{n+1}.

  Marginalizing $\mathcal{L} \sim P^{n+1}$ for $\mathsf{Unif}_{n+1}(\mathcal{L})$ is not P^{n+1}, i.e.,

  $\mathbb{E}_{\mathcal{L} \sim P^{n+1}}[\mathsf{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$!

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof outline.** The main idea: applying distribution-free validty to specific distributions.
  - Point I: given any distinct observations $\mathcal{L} \triangleq (x_i, y_i)_{i=1}^{n+1}$ from P^{n+1}, $\mathcal{L}$ is consistent with $\text{Unif}_{n+1}(\mathcal{L})$ $\leftarrow$ uniformly draw $n+1$ i.i.d pairs from $\mathcal{L}$
  - Point II: $y_i = \mathbb{E}[Y \mid X = x_i]$ on $\text{Unif}_{n+1}(\mathcal{L})$, i.e., $Y = \pi_{\text{Unif}_{n+1}(\mathcal{L})}(X)$.

  Applying DF-validty of $\hat{C}_n$ over $\text{Unif}_{n+1}(\mathcal{L})$ yields

$$\mathbb{P}_{\text{Unif}_{n+1}(\mathcal{L})}\left(Y_{n+1} \in \hat{C}_n\left(X_{n+1}\right)\right) \geq 1-\alpha.$$

  Finally, marginalize $\mathcal{L} \sim P^{n+1}$ for (1) to get the result for P^{n+1}.

  Marginalizing $\mathcal{L} \sim P^{n+1}$ for $\text{Unif}_{n+1}(\mathcal{L})$ is not P^{n+1}, i.e.,

  $\mathbb{E}_{\mathcal{L}\sim P^{n+1}}[\text{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$!

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof outline.** The main idea: applying distribution-free validty to specific distributions.
    - Point I: given any distinct observations $\mathcal{L} \triangleq (x_i, y_i)_{i=1}^{n+1}$ from P^{n+1}, $\mathcal{L}$ is consistent with $\mathsf{Unif}_{n+1}(\mathcal{L})$ $\leftarrow$ uniformly draw $n+1$ i.i.d pairs from $\mathcal{L}$
    - Point II: $y_i = \mathbb{E}[Y \mid X = x_i]$ on $\mathsf{Unif}_{n+1}(\mathcal{L})$, i.e., $Y = \pi_{\mathsf{Unif}_{n+1}(\mathcal{L})}(X)$.

Applying DF-validty of $\hat{C}_n$ over $\mathsf{Unif}_{n+1}(\mathcal{L})$ yields

$$\mathbb{P}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left(Y_{n+1} \in \hat{C}_n\left(X_{n+1}\right)\right) \geq 1-\alpha.$$

Finally, marginalize $\mathcal{L} \sim P^{n+1}$ for (1) to get the result for P^{n+1}.

Marginalizing $\mathcal{L} \sim P^{n+1}$ for $\mathsf{Unif}_{n+1}(\mathcal{L})$ is not P^{n+1}, i.e.,

$\mathbb{E}_{\mathcal{L} \sim P^{n+1}}[\mathsf{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$!

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof outline.** The main idea: applying distribution-free validty to specific distributions.
  - Point I: given any distinct observations $\mathcal{L} \triangleq (x_i, y_i)_{i=1}^{n+1}$ from P^{n+1}, $\mathcal{L}$ is consistent with $\mathsf{Unif}_{n+1}(\mathcal{L})$ $\leftarrow$ uniformly draw $n+1$ i.i.d pairs from $\mathcal{L}$
  - Point II: $y_i = \mathbb{E}[Y \mid X = x_i]$ on $\mathsf{Unif}_{n+1}(\mathcal{L})$, i.e., $Y = \pi_{\mathsf{Unif}_{n+1}(\mathcal{L})}(X)$.

  Applying DF-validty of $\hat{C}_n$ over $\mathsf{Unif}_{n+1}(\mathcal{L})$ yields

$$\mathbb{P}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left(Y_{n+1} \in \hat{C}_n\left(X_{n+1}\right)\right) \geq 1-\alpha.$$

  Finally, marginalize $\mathcal{L} \sim P^{n+1}$ for (1) to get the result for P^{n+1}.

  Marginalizing $\mathcal{L} \sim P^{n+1}$ for $\mathsf{Unif}_{n+1}(\mathcal{L})$ is not P^{n+1}, i.e.,

  $\mathbb{E}_{\mathcal{L} \sim P^{n+1}}[\mathsf{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$!

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof outline.** The main idea: applying distribution-free validty to specific distributions.
  - Point I: given any distinct observations $\mathcal{L} \triangleq (x_i, y_i)_{i=1}^{n+1}$ from P^{n+1}, $\mathcal{L}$ is consistent with $\mathsf{Unif}_{n+1}(\mathcal{L})$ $\leftarrow$ uniformly draw $n+1$ i.i.d pairs from $\mathcal{L}$
  - Point II: $y_i = \mathbb{E}[Y \mid X = x_i]$ on $\mathsf{Unif}_{n+1}(\mathcal{L})$, i.e., $Y = \pi_{\mathsf{Unif}_{n+1}(\mathcal{L})}(X)$.

  Applying DF-validty of $\hat{C}_n$ over $\mathsf{Unif}_{n+1}(\mathcal{L})$ yields

$$\mathbb{P}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left(Y_{n+1} \in \hat{C}_n\left(X_{n+1}\right)\right) \geq 1-\alpha.$$

  Finally, marginalize $\mathcal{L} \sim P^{n+1}$ for (1) to get the result for P^{n+1}.

  Marginalizing $\mathcal{L} \sim P^{n+1}$ for $\mathsf{Unif}_{n+1}(\mathcal{L})$ is not P^{n+1}, i.e., $\mathbb{E}_{\mathcal{L} \sim P^{n+1}}[\mathsf{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$!

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof outline.** The main idea: applying distribution-free validty to specific distributions.
  - Point I: given any distinct observations $\mathcal{L} \triangleq (x_i, y_i)_{i=1}^{n+1}$ from P^{n+1}, $\mathcal{L}$ is consistent with $\text{Unif}_{n+1}(\mathcal{L})$ $\leftarrow$ uniformly draw $n+1$ i.i.d pairs from $\mathcal{L}$
  - Point II: $y_i = \mathbb{E}[Y \mid X = x_i]$ on $\text{Unif}_{n+1}(\mathcal{L})$, i.e., $Y = \pi_{\text{Unif}_{n+1}(\mathcal{L})}(X)$.

  Applying DF-validty of $\hat{C}_n$ over $\text{Unif}_{n+1}(\mathcal{L})$ yields

$$\mathbb{P}_{\text{Unif}_{n+1}(\mathcal{L})}\left(Y_{n+1} \in \hat{C}_n\left(X_{n+1}\right)\right) \geq 1-\alpha.$$

  Finally, marginalize $\mathcal{L} \sim P^{n+1}$ for (1) to get the result for P^{n+1}.
  Marginalizing $\mathcal{L} \sim P^{n+1}$ for $\text{Unif}_{n+1}(\mathcal{L})$ is not P^{n+1}, i.e.,
  $\mathbb{E}_{\mathcal{L}\sim P^{n+1}}[\text{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$!

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof outline.** $\mathbb{E}_{\mathcal{L}\sim P^{n+1}}[\mathsf{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$

  $\mathsf{Unif}_{n+1}(\mathcal{L})(\text{repetition}) > 0$ $\quad$ $P^{n+1}(\text{repetition}) = 0$

  - Fact I: $P^{n+1} \overset{(d)}{\neq} \mathsf{Unif}_{n+1}(\mathcal{L} \sim P^{n+1})$.
  - Fact II: $P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(\mathcal{L} \sim P^{n+1})$.

    subsample $n+1$ pairs without replacement from $\mathcal{L}$

  - Fact III: for any $m \geq n+1$,

    $P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(\mathcal{L} \sim P^m) \overset{(d)}{\approx} \mathsf{Unif}_{n+1}(\mathcal{L} \sim P^m)$.

    when $m \to \infty$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.

- **Proof outline.** $\mathbb{E}_{\mathcal{L}\sim P^{n+1}}[\mathsf{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$

  $\mathsf{Unif}_{n+1}(\mathcal{L})(\text{repetition}) > 0$ $\qquad$ $P^{n+1}(\text{repetition}) = 0$

  - Fact I: $P^{n+1} \overset{(d)}{\neq} \mathsf{Unif}_{n+1}(\mathcal{L} \sim P^{n+1})$.
  - Fact II: $P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(\mathcal{L} \sim P^{n+1})$.

    subsample $n+1$ pairs without replacement from $\mathcal{L}$

  - Fact III: for any $m \geq n+1$,

    $P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(\mathcal{L} \sim P^m) \overset{(d)}{\approx} \mathsf{Unif}_{n+1}(\mathcal{L} \sim P^m)$.

    when $m \to \infty$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof outline.** $\mathbb{E}_{\mathcal{L}\sim P^{n+1}}[\mathsf{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$

  $\mathsf{Unif}_{n+1}(\mathcal{L})(\text{repetition}) > 0$ $P^{n+1}(\text{repetition}) = 0$

  - Fact I: $P^{n+1} \overset{(d)}{\neq} \mathsf{Unif}_{n+1}(\mathcal{L} \sim P^{n+1})$.
  - Fact II: $P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(\mathcal{L} \sim P^{n+1})$.

    subsample $n+1$ pairs without replacement from $\mathcal{L}$

  - Fact III: for any $m \geq n+1$,

    $P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(\mathcal{L} \sim P^m) \overset{(d)}{\approx} \mathsf{Unif}_{n+1}(\mathcal{L} \sim P^m)$.

    when $m \to \infty$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof outline.** $\mathbb{E}_{\mathcal{L}\sim P^{n+1}}[\mathsf{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$

  $\mathsf{Unif}_{n+1}(\mathcal{L})(\text{repetition}) > 0$ $\quad$ $P^{n+1}(\text{repetition}) = 0$

  - Fact I: $P^{n+1} \overset{(d)}{\neq} \mathsf{Unif}_{n+1}(\mathcal{L}\sim P^{n+1})$.
  - Fact II: $P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(\mathcal{L}\sim P^{n+1})$.

    subsample $n+1$ pairs without replacement from $\mathcal{L}$

  - Fact III: for any $m \geq n+1$,
    $P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(\mathcal{L}\sim P^{m}) \overset{(d)}{\approx} \mathsf{Unif}_{n+1}(\mathcal{L}\sim P^{m})$.

    when $m \to \infty$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.

- **Proof outline.** $\mathbb{E}_{\mathcal{L}\sim P^{n+1}}[\mathsf{Unif}_{n+1}(\mathcal{L})] \overset{(d)}{\neq} P^{n+1}$

  $\mathsf{Unif}_{n+1}(\mathcal{L})(\text{repetition}) > 0$ $\qquad$ $P^{n+1}(\text{repetition}) = 0$

  - Fact I: $P^{n+1} \overset{(d)}{\neq} \mathsf{Unif}_{n+1}(\mathcal{L} \sim P^{n+1})$.
  - Fact II: $P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(\mathcal{L} \sim P^{n+1})$.

    subsample $n+1$ pairs without replacement from $\mathcal{L}$

  - Fact III: for any $m \geq n+1$,

    $P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(\mathcal{L} \sim P^m) \overset{(d)}{\approx} \mathsf{Unif}_{n+1}(\mathcal{L} \sim P^m)$.

    when $m \to \infty$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof summary.** $E_1 \triangleq \{\pi(Y_{n+1}) \in \hat{C}_n(X_{n+1})\}$, $E_2 \triangleq \{Y_{n+1} \in \hat{C}_n(X_{n+1})\}$.

$$
\begin{aligned}
&\mathbb{P}_{P^{n+1}}(E_1) \overset{m \geq n+1}{=} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) \quad (\because P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(P^m)) \\
=&\lim_{m\to\infty} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) = \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \quad (\because \mathsf{Sub}_{n+1}(\mathcal{L}) \underset{m\to\infty}{\overset{(d)}{=}} \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \quad (\because E_1 = E_2 \text{ on } \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\cdots = \mathbb{P}_{P^{n+1}}(E_2)
\end{aligned}
$$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof summary.** $E_1 \triangleq \{\pi(Y_{n+1}) \in \hat{C}_n(X_{n+1})\}$, $E_2 \triangleq \{Y_{n+1} \in \hat{C}_n(X_{n+1})\}$.

$$
\begin{aligned}
&\mathbb{P}_{P^{n+1}}(E_1) \overset{m \geq n+1}{=} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) \quad (\because P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(P^m)) \\
=&\lim_{m\to\infty} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) = \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \quad (\because \mathsf{Sub}_{n+1}(\mathcal{L}) \underset{m\to\infty}{\overset{(d)}{=}} \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \quad (\because E_1 = E_2 \text{ on } \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\cdots = \mathbb{P}_{P^{n+1}}(E_2)
\end{aligned}
$$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof summary.** $E_1 \triangleq \{\pi(Y_{n+1}) \in \hat{C}_n(X_{n+1})\}$, $E_2 \triangleq \{Y_{n+1} \in \hat{C}_n(X_{n+1})\}$.

$$
\begin{aligned}
&\mathbb{P}_{P^{n+1}}(E_1) \overset{m \geq n+1}{=} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) \quad (\because P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(P^m)) \\
=&\lim_{m\to\infty} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) = \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \quad (\because \mathsf{Sub}_{n+1}(\mathcal{L}) \underset{m\to\infty}{\overset{(d)}{=}} \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \quad (\because E_1 = E_2 \text{ on } \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\cdots = \mathbb{P}_{P^{n+1}}(E_2)
\end{aligned}
$$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof summary.** $E_1 \triangleq \{\pi(Y_{n+1}) \in \hat{C}_n(X_{n+1})\}$, $E_2 \triangleq \{Y_{n+1} \in \hat{C}_n(X_{n+1})\}$.

$$
\begin{aligned}
&\mathbb{P}_{P^{n+1}}(E_1) \overset{m \geq n+1}{=} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) \quad (\because P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(P^m)) \\
=&\lim_{m\to\infty} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) = \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \quad (\because \mathsf{Sub}_{n+1}(\mathcal{L}) \underset{m\to\infty}{\overset{(d)}{=}} \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \quad (\because E_1 = E_2 \text{ on } \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\cdots = \mathbb{P}_{P^{n+1}}(E_2)
\end{aligned}
$$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof summary.** $E_1 \triangleq \{\pi(Y_{n+1}) \in \hat{C}_n(X_{n+1})\}$, $E_2 \triangleq \{Y_{n+1} \in \hat{C}_n(X_{n+1})\}$.

$$\mathbb{P}_{P^{n+1}}(E_1) \overset{m \geq n+1}{=} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) \quad (\because P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(P^m))$$

$$= \lim_{m\to\infty} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) = \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right]$$

$$= \mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right]$$

$$= \mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \quad (\because \mathsf{Sub}_{n+1}(\mathcal{L}) \underset{m\to\infty}{\overset{(d)}{=}} \mathsf{Unif}_{n+1}(\mathcal{L}))$$

$$= \mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \quad (\because E_1 = E_2 \text{ on } \mathsf{Unif}_{n+1}(\mathcal{L}))$$

$$= \mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right]$$

$$= \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right]$$

$$= \cdots = \mathbb{P}_{P^{n+1}}(E_2)$$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof summary.** $E_1 \triangleq \{\pi(Y_{n+1}) \in \hat{C}_n(X_{n+1})\}$, $E_2 \triangleq \{Y_{n+1} \in \hat{C}_n(X_{n+1})\}$.

$$
\begin{aligned}
&\mathbb{P}_{P^{n+1}}(E_1) \overset{m \geq n+1}{=} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) \quad (\because P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(P^m)) \\
=&\lim_{m\to\infty} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) = \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \quad (\because \mathsf{Sub}_{n+1}(\mathcal{L}) \underset{m\to\infty}{\overset{(d)}{=}} \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \quad (\because E_1 = E_2 \text{ on } \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\cdots = \mathbb{P}_{P^{n+1}}(E_2)
\end{aligned}
$$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof summary.** $E_1 \triangleq \{\pi(Y_{n+1}) \in \hat{C}_n(X_{n+1})\}$, $E_2 \triangleq \{Y_{n+1} \in \hat{C}_n(X_{n+1})\}$.

$$
\begin{aligned}
&\mathbb{P}_{P^{n+1}}(E_1) \overset{m \geq n+1}{=} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) \quad (\because P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(P^m)) \\
=&\lim_{m\to\infty} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) = \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \quad (\because \mathsf{Sub}_{n+1}(\mathcal{L}) \underset{m\to\infty}{\overset{(d)}{=}} \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \quad (\because E_1 = E_2 \text{ on } \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\cdots = \mathbb{P}_{P^{n+1}}(E_2)
\end{aligned}
$$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof summary.** $E_1 \triangleq \{\pi(Y_{n+1}) \in \hat{C}_n(X_{n+1})\}$, $E_2 \triangleq \{Y_{n+1} \in \hat{C}_n(X_{n+1})\}$.

$$
\begin{aligned}
\mathbb{P}_{P^{n+1}}(E_1) &\overset{m \geq n+1}{=} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) \quad (\because P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(P^m)) \\
&= \lim_{m\to\infty} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) = \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
&= \mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
&= \mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \quad (\because \mathsf{Sub}_{n+1}(\mathcal{L}) \underset{m\to\infty}{\overset{(d)}{=}} \mathsf{Unif}_{n+1}(\mathcal{L})) \\
&= \mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \quad (\because E_1 = E_2 \text{ on } \mathsf{Unif}_{n+1}(\mathcal{L})) \\
&= \mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
&= \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
&= \cdots = \mathbb{P}_{P^{n+1}}(E_2)
\end{aligned}
$$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof summary.** $E_1 \triangleq \{\pi(Y_{n+1}) \in \hat{C}_n(X_{n+1})\}$, $E_2 \triangleq \{Y_{n+1} \in \hat{C}_n(X_{n+1})\}$.

$$
\begin{aligned}
&\mathbb{P}_{P^{n+1}}(E_1) \overset{m \geq n+1}{=} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) \quad (\because P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(P^m)) \\
=&\lim_{m\to\infty} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) = \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \quad (\because \mathsf{Sub}_{n+1}(\mathcal{L}) \underset{m\to\infty}{\overset{(d)}{=}} \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \quad (\because E_1 = E_2 \text{ on } \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\cdots = \mathbb{P}_{P^{n+1}}(E_2)
\end{aligned}
$$

Proof

- **Theorem.** $(1-\alpha)$-DF confidence interval implies $(1-\alpha)$ prediction set, valid for all nonatomic distributions.
- **Proof summary.** $E_1 \triangleq \{\pi(Y_{n+1}) \in \hat{C}_n(X_{n+1})\}$, $E_2 \triangleq \{Y_{n+1} \in \hat{C}_n(X_{n+1})\}$.

$$
\begin{aligned}
&\mathbb{P}_{P^{n+1}}(E_1) \overset{m \geq n+1}{=} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) \quad (\because P^{n+1} \overset{(d)}{=} \mathsf{Sub}_{n+1}(P^m)) \\
=&\lim_{m\to\infty} \mathbb{P}_{\mathsf{Sub}_{n+1}(P^m)}(E_1) = \lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_1} \mid \mathcal{L}\right]\right] \quad (\because \mathsf{Sub}_{n+1}(\mathcal{L}) \underset{m\to\infty}{\overset{(d)}{=}} \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Unif}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \quad (\because E_1 = E_2 \text{ on } \mathsf{Unif}_{n+1}(\mathcal{L})) \\
=&\mathbb{E}_{\mathcal{L}\sim P^m}\left[\lim_{m\to\infty} \mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\lim_{m\to\infty} \mathbb{E}_{\mathcal{L}\sim P^m}\left[\mathbb{E}_{\mathsf{Sub}_{n+1}(\mathcal{L})}\left[\mathbb{1}_{E_2} \mid \mathcal{L}\right]\right] \\
=&\cdots = \mathbb{P}_{P^{n+1}}(E_2)
\end{aligned}
$$

A Simple Corollary

Corollary

For any $(1-\alpha)$*-DF confidence interval* $\hat{C}_n$ *with* $\alpha \in (0, 0.5)$. *Then there exists* *one distribution* P *such that*

$$\textit{len}_{n,\alpha}(\hat{C}_n, P) := \mathbb{E}_{P^{n+1}}\left[|\hat{C}_n(X_{n+1})|\right] \geq 0.5 - \alpha.$$

- This lower bound does not vanish as $n \to \infty$!
- **Proof outline.** Let $P = P_X \times \text{Ber}(0.5)$ with P_X nonatomic.

  Theorem 2 implies $\mathbb{P}(\{\mathbb{E}[Y_{n+1} \mid X_{n+1}], Y_{n+1}\} \subseteq \hat{C}_n(X_{n+1})) \geq 1 - 2\alpha$.

  W.p at least $1 - 2\alpha$, as an interval,
  $\text{len}(\hat{C}_n(X_{n+1})) \geq |\mathbb{E}[Y_{n+1} \mid X_{n+1}] - Y_{n+1}| = |0.5 - 1 \text{ or } 0| = 0.5$.

  $\mathbb{E}_{P^{n+1}}[\text{len}(\hat{C}_n(X_{n+1}))] \geq 0.5(1 - 2\alpha) = \frac{1}{2} - \alpha$.

A Simple Corollary

Corollary

For any $(1-\alpha)$*-DF confidence interval* $\hat{C}_n$ *with* $\alpha \in (0, 0.5)$*. Then there exists* *one distribution* P *such that*

$$\textit{len}_{n,\alpha}(\hat{C}_n, P) := \mathbb{E}_{P^{n+1}}[|\hat{C}_n(X_{n+1})|] \geq 0.5 - \alpha.$$

- This lower bound does not vanish as $n \to \infty$!
- **Proof outline.** Let $P = P_X \times \text{Ber}(0.5)$ with P_X nonatomic.

  Theorem 2 implies $\mathbb{P}(\{\mathbb{E}[Y_{n+1} \mid X_{n+1}], Y_{n+1}\} \subseteq \hat{C}_n(X_{n+1})) \geq 1 - 2\alpha$.

  W.p at least $1 - 2\alpha$, as an interval,
  $\text{len}(\hat{C}_n(X_{n+1})) \geq |\mathbb{E}[Y_{n+1} \mid X_{n+1}] - Y_{n+1}| = |0.5 - 1 \text{ or } 0| = 0.5$.

  $\mathbb{E}_{P^{n+1}}[\text{len}(\hat{C}_n(X_{n+1}))] \geq 0.5(1 - 2\alpha) = \frac{1}{2} - \alpha$.

A Simple Corollary

Corollary

For any $(1-\alpha)$*-DF confidence interval* $\hat{C}_n$ *with* $\alpha \in (0, 0.5)$*. Then there exists* *one distribution* P *such that*

$$\textit{len}_{n,\alpha}(\hat{C}_n, P) := \mathbb{E}_{P^{n+1}}[|\hat{C}_n(X_{n+1})|] \geq 0.5 - \alpha.$$

- This lower bound does not vanish as $n \to \infty$!
- **Proof outline.** Let $P = P_X \times \text{Ber}(0.5)$ with P_X nonatomic.

  Theorem 2 implies $\mathbb{P}(\{\mathbb{E}[Y_{n+1} \mid X_{n+1}], Y_{n+1}\} \subseteq \hat{C}_n(X_{n+1})) \geq 1 - 2\alpha$.

  W.p at least $1 - 2\alpha$, as an interval,
  $\text{len}(\hat{C}_n(X_{n+1})) \geq |\mathbb{E}[Y_{n+1} \mid X_{n+1}] - Y_{n+1}| = |0.5 - 1 \text{ or } 0| = 0.5.$

  $\mathbb{E}_{P^{n+1}}[\text{len}(\hat{C}_n(X_{n+1}))] \geq 0.5(1 - 2\alpha) = \frac{1}{2} - \alpha.$

A Simple Corollary

Corollary

For any $(1-\alpha)$*-DF confidence interval* $\hat{C}_n$ *with* $\alpha \in (0, 0.5)$*. Then there exists* *one distribution* P *such that*

$$\textit{len}_{n,\alpha}(\hat{C}_n, P) := \mathbb{E}_{P^{n+1}}[|\hat{C}_n(X_{n+1})|] \geq 0.5 - \alpha.$$

- This lower bound does not vanish as $n \to \infty$!
- **Proof outline.** Let $P = P_X \times \mathrm{Ber}(0.5)$ with P_X nonatomic.

  Theorem 2 implies $\mathbb{P}(\{\mathbb{E}[Y_{n+1} \mid X_{n+1}], Y_{n+1}\} \subseteq \hat{C}_n(X_{n+1})) \geq 1 - 2\alpha$.

  W.p at least $1 - 2\alpha$, as an interval,
  $\mathrm{len}(\hat{C}_n(X_{n+1})) \geq |\mathbb{E}[Y_{n+1} \mid X_{n+1}] - Y_{n+1}| = |0.5 - 1 \text{ or } 0| = 0.5$.

  $\mathbb{E}_{P^{n+1}}[\mathrm{len}(\hat{C}_n(X_{n+1}))] \geq 0.5(1 - 2\alpha) = \frac{1}{2} - \alpha$.

A Simple Corollary

Corollary

For any $(1-\alpha)$*-DF confidence interval* $\hat{C}_n$ *with* $\alpha \in (0, 0.5)$*. Then there exists* *one distribution* P *such that*

$$\textit{len}_{n,\alpha}(\hat{C}_n, P) := \mathbb{E}_{P^{n+1}}[|\hat{C}_n(X_{n+1})|] \geq 0.5 - \alpha.$$

- This lower bound does not vanish as $n \to \infty$!
- **Proof outline.** Let $P = P_X \times \text{Ber}(0.5)$ with P_X nonatomic.

  Theorem 2 implies $\mathbb{P}(\{\mathbb{E}[Y_{n+1} \mid X_{n+1}], Y_{n+1}\} \subseteq \hat{C}_n(X_{n+1})) \geq 1 - 2\alpha$.

  W.p at least $1 - 2\alpha$, as an interval,
  $\text{len}(\hat{C}_n(X_{n+1})) \geq |\mathbb{E}[Y_{n+1} \mid X_{n+1}] - Y_{n+1}| = |0.5 - 1 \text{ or } 0| = 0.5.$

  $\mathbb{E}_{P^{n+1}}[\text{len}(\hat{C}_n(X_{n+1}))] \geq 0.5(1 - 2\alpha) = \frac{1}{2} - \alpha.$

A Stronger Lemma

Lemma

Let $(X_i, Y_i)_{i=1}^{n+1} \overset{i.i.d}{\sim} P$ *(nonatomic) and* $Z \perp (X_i, Y_i)_{i=1}^{n}$ *with*

$$\mathbb{E}[Z \mid X_{n+1}] = \pi_P(X_{n+1}).$$

Then

$$\mathbb{P}\left(Z \in \hat{C}_n(X_{n+1})\right) \geq 1 - \alpha$$

- E.g., $Z = Y_{n+1}$ yields the previous theorem.
- **Proof outline.** Let $\tilde{P}_{X,Z}$ be the joint distri. of (X_{n+1}, Z_{n+1}) and define

$$\tilde{P} : (X, Z) \sim \tilde{P}_{X,Z} \text{ and } Y \mid (X, Z) = \text{Ber}(Z).$$

Marginalizing $\tilde{P}$ over Z, $(X, Y) \sim P$. $\leftarrow \mathbb{E}_Z[\text{Ber}(Z)] \overset{(d)}{=} \text{Ber}(\mathbb{E}[Z] = \pi_P(X))$
Applying DF validty to distribuitions conditioned on $(X_i, Z_i)_{i=1}^{n+1}$.

A Stronger Lemma

Lemma

Let $(X_i, Y_i)_{i=1}^{n+1} \overset{i.i.d}{\sim} P$ *(nonatomic) and* $Z \perp (X_i, Y_i)_{i=1}^{n}$ *with*

$$\mathbb{E}[Z \mid X_{n+1}] = \pi_P(X_{n+1}).$$

Then

$$\mathbb{P}\left(Z \in \hat{C}_n(X_{n+1})\right) \geq 1 - \alpha$$

- E.g., $Z = Y_{n+1}$ yields the previous theorem.
- **Proof outline.** Let $\tilde{P}_{X,Z}$ be the joint distri. of (X_{n+1}, Z_{n+1}) and define

$$\tilde{P} : (X, Z) \sim \tilde{P}_{X,Z} \text{ and } Y \mid (X, Z) = \text{Ber}(Z).$$

Marginalizing $\tilde{P}$ over Z, $(X, Y) \sim P$. $\leftarrow \mathbb{E}_Z[\text{Ber}(Z)] \overset{(d)}{=} \text{Ber}(\mathbb{E}[Z] = \pi_P(X))$
Applying DF validty to distribuitions conditioned on $(X_i, Z_i)_{i=1}^{n+1}$.

A Stronger Lemma

Lemma

Let $(X_i, Y_i)_{i=1}^{n+1} \overset{i.i.d}{\sim} P$ *(nonatomic) and* $Z \perp (X_i, Y_i)_{i=1}^{n}$ *with*

$$\mathbb{E}[Z \mid X_{n+1}] = \pi_P(X_{n+1}).$$

Then

$$\mathbb{P}\left(Z \in \hat{C}_n(X_{n+1})\right) \geq 1 - \alpha$$

- E.g., $Z = Y_{n+1}$ yields the previous theorem.
- **Proof outline.** Let $\tilde{P}_{X,Z}$ be the joint distri. of (X_{n+1}, Z_{n+1}) and define

$$\tilde{P} : (X, Z) \sim \tilde{P}_{X,Z} \text{ and } Y \mid (X, Z) = \mathsf{Ber}(Z).$$

Marginalizing $\tilde{P}$ over Z, $(X, Y) \sim P$. $\leftarrow \mathbb{E}_Z[\mathsf{Ber}(Z)] \overset{(d)}{=} \mathsf{Ber}(\mathbb{E}[Z] = \pi_P(X))$
Applying DF validty to distribuitions conditioned on $(X_i, Z_i)_{i=1}^{n+1}$.

A Stronger Lemma

Lemma

Let $(X_i, Y_i)_{i=1}^{n+1} \overset{i.i.d}{\sim} P$ *(nonatomic) and* $Z \perp (X_i, Y_i)_{i=1}^{n}$ *with*

$$\mathbb{E}[Z \mid X_{n+1}] = \pi_P(X_{n+1}).$$

Then

$$\mathbb{P}\left(Z \in \hat{C}_n(X_{n+1})\right) \geq 1 - \alpha$$

- E.g., $Z = Y_{n+1}$ yields the previous theorem.
- **Proof outline.** Let $\tilde{P}_{X,Z}$ be the joint distri. of (X_{n+1}, Z_{n+1}) and define

$$\tilde{P} : (X, Z) \sim \tilde{P}_{X,Z} \text{ and } Y \mid (X, Z) = \text{Ber}(Z).$$

Marginalizing $\tilde{P}$ over Z, $(X, Y) \sim P$. $\leftarrow \mathbb{E}_Z[\text{Ber}(Z)] \overset{(d)}{=} \text{Ber}(\mathbb{E}[Z] = \pi_P(X))$
Applying DF validty to distribuitions conditioned on $(X_i, Z_i)_{i=1}^{n+1}$.

A Stronger Lemma

Lemma

Let $(X_i, Y_i)_{i=1}^{n+1} \overset{i.i.d}{\sim} P$ *(nonatomic) and* $Z \perp (X_i, Y_i)_{i=1}^{n}$ *with*

$$\mathbb{E}[Z \mid X_{n+1}] = \pi_P(X_{n+1}).$$

Then

$$\mathbb{P}\left(Z \in \hat{C}_n(X_{n+1})\right) \geq 1 - \alpha$$

- E.g., $Z = Y_{n+1}$ yields the previous theorem.
- **Proof outline.** Let $\tilde{P}_{X,Z}$ be the joint distri. of (X_{n+1}, Z_{n+1}) and define

$$\tilde{P} : (X, Z) \sim \tilde{P}_{X,Z} \text{ and } Y \mid (X, Z) = \text{Ber}(Z).$$

Marginalizing $\tilde{P}$ over Z, $(X, Y) \sim P$. $\leftarrow \mathbb{E}_Z[\text{Ber}(Z)] \overset{(d)}{=} \text{Ber}(\mathbb{E}[Z] = \pi_P(X))$

Applying DF validty to distribuitions conditioned on $(X_i, Z_i)_{i=1}^{n+1}$.

A Stronger Lemma

Lemma

Let $(X_i, Y_i)_{i=1}^{n+1} \overset{i.i.d}{\sim} P$ *(nonatomic) and* $Z \perp (X_i, Y_i)_{i=1}^{n}$ *with*

$$\mathbb{E}[Z \mid X_{n+1}] = \pi_P(X_{n+1}).$$

Then

$$\mathbb{P}\left(Z \in \hat{C}_n(X_{n+1})\right) \geq 1 - \alpha$$

- E.g., $Z = Y_{n+1}$ yields the previous theorem.
- **Proof outline.** Let $\tilde{P}_{X,Z}$ be the joint distri. of (X_{n+1}, Z_{n+1}) and define

$$\tilde{P} : (X, Z) \sim \tilde{P}_{X,Z} \text{ and } Y \mid (X, Z) = \text{Ber}(Z).$$

Marginalizing $\tilde{P}$ over Z, $(X, Y) \sim P$. $\leftarrow \mathbb{E}_Z[\text{Ber}(Z)] \overset{(d)}{=} \text{Ber}(\mathbb{E}[Z] = \pi_P(X))$
Applying DF validty to distribuitions conditioned on $(X_i, Z_i)_{i=1}^{n+1}$.

General and Sharp Lower Bounds

Theorem

For any nonatomic P, len$_{n,\alpha}(\hat{C}_n, P) \geq L_\alpha(P) > 0$ where

$$L_\alpha(P) = \inf_{a:\mathbb{R}^d \to [0,1]} \{\mathbb{E}_P[\ell(\pi_P(X), a(X))] : \mathbb{E}_P[a(X)] \leq \alpha\}$$

with $\ell : [0,1] \to [0,1]$ fixed.

- Distribution-specific lower bound, not vanishing as $n \to \infty$.
- **Proof outline.**

  $\mathbb{P}(Z \in \hat{C}_n(X)) \geq 1 - \alpha$ for all Z such that $\mathbb{E}[Z \mid X] = \pi_P(X)$.

  Using "worst" Z's distribution (uniform + discrete) with $\mathbb{E}[Z \mid X] = \pi_P(X)$.

General and Sharp Lower Bounds

Theorem

For any nonatomic P, len$_{n,\alpha}(\hat{C}_n, P) \geq L_\alpha(P) > 0$ where

$$L_\alpha(P) = \inf_{a:\mathbb{R}^d \to [0,1]} \{\mathbb{E}_P[\ell(\pi_P(X), a(X))] : \mathbb{E}_P[a(X)] \leq \alpha\}$$

with $\ell : [0,1] \to [0,1]$ fixed.

- Distribution-specific lower bound, not vanishing as $n \to \infty$.
- **Proof outline.**

  $\mathbb{P}(Z \in \hat{C}_n(X)) \geq 1 - \alpha$ for all Z such that $\mathbb{E}[Z \mid X] = \pi_P(X)$.

  Using "worst" Z's distribution (uniform + discrete) with $\mathbb{E}[Z \mid X] = \pi_P(X)$.

General and Sharp Lower Bounds

Theorem

For any nonatomic P, len$_{n,\alpha}(\hat{C}_n, P) \geq L_\alpha(P) > 0$ where

$$L_\alpha(P) = \inf_{a:\mathbb{R}^d \to [0,1]} \{\mathbb{E}_P[\ell(\pi_P(X), a(X))] : \mathbb{E}_P[a(X)] \leq \alpha\}$$

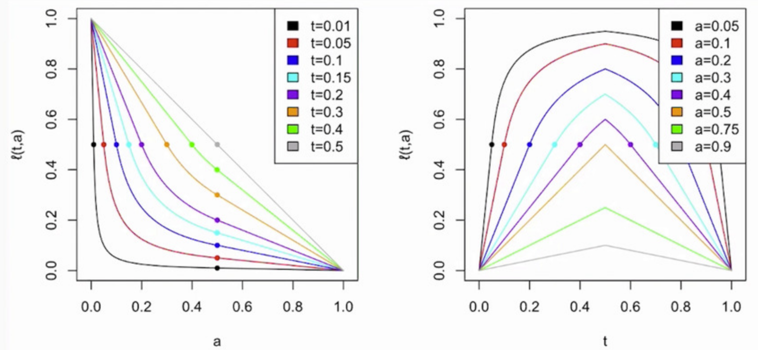
with $\ell : [0,1] \to [0,1]$ fixed.

- Distribution-specific lower bound, not vanishing as $n \to \infty$.
- **Proof outline.**

  $\mathbb{P}(Z \in \hat{C}_n(X)) \geq 1 - \alpha$ for all Z such that $\mathbb{E}[Z \mid X] = \pi_P(X)$.

  Using "worst" Z's distribution (uniform + discrete) with $\mathbb{E}[Z \mid X] = \pi_P(X)$.

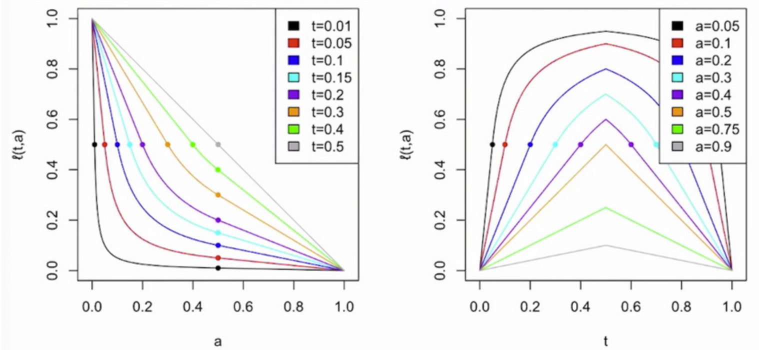
$\ell(t,a)$ Function



$$\ell(t,a) = \begin{cases} 2(1-a)t, & t \leq \frac{1}{2} \leq a \\ t/2a, & 0 < t \leq a < \frac{1}{2} \\ 1 - a/2t, & a < t \leq \frac{1}{2} \\ 0, & a = t = 0 \\ \text{(symmetric)}, & t \geq \frac{1}{2} \end{cases}$$

- $L_\alpha(P_X \times \mathrm{Ber}(0.5)) = 1 - \alpha$ because of $\ell(0.5, \alpha) = 1 - \alpha$.

$\ell(t, a)$ Function



$$\ell(t, a) = \begin{cases} 2(1-a)t, & t \leq \frac{1}{2} \leq a \\ t/2a, & 0 < t \leq a < \frac{1}{2} \\ 1 - a/2t, & a < t \leq \frac{1}{2} \\ 0, & a = t = 0 \\ \text{(symmetric)}, & t \geq \frac{1}{2} \end{cases}$$

- $L_\alpha(P_X \times \text{Ber}(0.5)) = 1 - \alpha$ because of $\ell(0.5, \alpha) = 1 - \alpha$.

A Matching Upper Bound?

Theorem

The proposed binning-based algorithm $\hat{C}_n$ satisfies

- *Distribution-free validity, i.e., coverage $\geq 1-\alpha$ w.r.t. P.*
- *Near-optimal length if the partition is "good":*

$$\mathbb{E}_P\left[\mathit{len}_{n,\alpha}(\hat{C}_n, P)\right] \leq L_\alpha(P) + \sqrt{2\alpha^{-1}\cdot\mathbb{E}_P\left[\left|\pi_P(X) - \pi_{m(X)|}\right|\right]} + \mathcal{O}\left(\sqrt{\frac{M\log n}{\alpha n}}\right)$$

partition (discretization) error $\nearrow$ — $\uparrow$ *concentration type error*

- Shuo will give a more detailed presentation about "upper bounds".

A Matching Upper Bound?

Theorem

The proposed binning-based algorithm $\hat{C}_n$ satisfies

- *Distribution-free validity, i.e., coverage $\geq 1 - \alpha$ w.r.t. P.*
- *Near-optimal length if the partition is "good":*

$$\mathbb{E}_P\left[\mathit{len}_{n,\alpha}(\hat{C}_n, P)\right] \leq L_\alpha(P) +$$

$$\underbrace{\sqrt{2\alpha^{-1} \cdot \mathbb{E}_P\left[\left|\pi_P(X) - \pi_{m(X)|}\right|\right]}}_{\text{partition (discretization) error}} + \underbrace{\mathcal{O}\left(\sqrt{\frac{M \log n}{\alpha n}}\right)}_{\text{concentration type error}}$$

- Shuo will give a more detailed presentation about "upper bounds".

Overview

Calibration

- A predictor $f : \mathbb{R}^d \to [0,1]$ is (perfectly) calibrated if $\mathbb{E}[Y \mid f(X)] = f(X)$.
- Approximate calibration: (ε, α)-calibrated if

$$\mathbb{P}\left(|\mathbb{E}[Y \mid f(X)] - f(X)| \leq \varepsilon\right) \geq 1 - \alpha.$$

- $\underbrace{|\mathbb{E}[Y \mid f(x)] - f(x)| \leq \varepsilon}_{\text{calibration}} \Rightarrow \underbrace{\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x))}_{\text{CI w.r.t. } P_{(f(X),Y)}} := [f(x) - \varepsilon, f(x) + \varepsilon]$

  (ε, α)-calibration implies $(1-\alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$.

- Conversely, if $\hat{C}$ is $(1-\alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$,
  $\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x)) \Rightarrow |\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] - m_{\hat{C}}(f(x))| \leq 0.5|\hat{C}(f(x))|$
  where $m_{\hat{C}}(f(X))$ is the middle point of $\hat{C}(f(x))$.

  $\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] = \mathbb{E}[Y \mid f(x)]$ if $m_{\hat{C}}$ is injective

  $(1-\alpha)$-DF conf. interval implies $(0.5 \sup_{f(x)} |\hat{C}(f(x))|, \alpha)$-calibration

Calibration

- A predictor $f : \mathbb{R}^d \to [0,1]$ is (perfectly) calibrated if $\mathbb{E}[Y \mid f(X)] = f(X)$.
- Approximate calibration: (ε, α)-calibrated if

$$\mathbb{P}\left(|\mathbb{E}[Y \mid f(X)] - f(X)| \le \varepsilon\right) \ge 1 - \alpha.$$

- $\underbrace{|\mathbb{E}[Y \mid f(x)] - f(x)| \le \varepsilon}_{\text{calibration}} \;\Rightarrow\; \underbrace{\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x))}_{\text{CI w.r.t. } P_{(f(X),Y)}} := [f(x) - \varepsilon, f(x) + \varepsilon]$

  (ε, α)-calibration implies $(1-\alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$.

- Conversely, if $\hat{C}$ is $(1-\alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$,
  $\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x)) \;\Rightarrow\; |\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] - m_{\hat{C}}(f(x))| \le 0.5|\hat{C}(f(x))|$
  where $m_{\hat{C}}(f(X))$ is the middle point of $\hat{C}(f(x))$.

  $\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] = \mathbb{E}[Y \mid f(x)]$ if $m_{\hat{C}}$ is injective

  $(1-\alpha)$-DF conf. interval implies $(0.5 \sup_{f(x)} |\hat{C}(f(x))|, \alpha)$-calibration

Calibration

- A predictor $f : \mathbb{R}^d \to [0,1]$ is (perfectly) calibrated if $\mathbb{E}[Y \mid f(X)] = f(X)$.
- Approximate calibration: (ε, α)-calibrated if

$$\mathbb{P}\left(|\mathbb{E}[Y \mid f(X)] - f(X)| \le \varepsilon\right) \ge 1 - \alpha.$$

- $\underbrace{|\mathbb{E}[Y \mid f(x)] - f(x)| \le \varepsilon}_{\text{calibration}} \;\Rightarrow\; \underbrace{\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x))}_{\text{CI w.r.t. } P_{(f(X),Y)}} := [f(x) - \varepsilon, f(x) + \varepsilon]$

  (ε, α)-calibration implies $(1 - \alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$.

- Conversely, if $\hat{C}$ is $(1-\alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$,

  $\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x)) \;\Rightarrow\; |\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] - m_{\hat{C}}(f(x))| \le 0.5|\hat{C}(f(x))|$

  where $m_{\hat{C}}(f(X))$ is the middle point of $\hat{C}(f(x))$.

  $\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] = \mathbb{E}[Y \mid f(x)]$ if $m_{\hat{C}}$ is injective

  $(1-\alpha)$-DF conf. interval implies $(0.5 \sup_{f(x)} |\hat{C}(f(x))|, \alpha)$-calibration

Calibration

- A predictor $f : \mathbb{R}^d \to [0,1]$ is (perfectly) calibrated if $\mathbb{E}[Y \mid f(X)] = f(X)$.
- Approximate calibration: (ε, α)-calibrated if

$$\mathbb{P}\left(|\mathbb{E}[Y \mid f(X)] - f(X)| \leq \varepsilon\right) \geq 1 - \alpha.$$

- $\underbrace{|\mathbb{E}[Y \mid f(x)] - f(x)| \leq \varepsilon}_{\text{calibration}} \;\; \Rightarrow \;\; \underbrace{\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x))}_{\text{CI w.r.t. } P_{(f(X),Y)}} := [f(x) - \varepsilon, f(x) + \varepsilon]$

  (ε, α)-calibration implies $(1-\alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$.

- Conversely, if $\hat{C}$ is $(1-\alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$,
  $\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x)) \;\; \Rightarrow \;\; |\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] - m_{\hat{C}}(f(x))| \leq 0.5|\hat{C}(f(x))|$
  where $m_{\hat{C}}(f(X))$ is the middle point of $\hat{C}(f(x))$.

  $\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] = \mathbb{E}[Y \mid f(x)]$ if $m_{\hat{C}}$ is injective

  $(1-\alpha)$-DF conf. interval implies $(0.5 \sup_{f(x)} |\hat{C}(f(x))|, \alpha)$-calibration

Calibration

- A predictor $f : \mathbb{R}^d \to [0,1]$ is (perfectly) calibrated if $\mathbb{E}[Y \mid f(X)] = f(X)$.
- Approximate calibration: (ε, α)-calibrated if

$$\mathbb{P}\left(|\mathbb{E}[Y \mid f(X)] - f(X)| \leq \varepsilon\right) \geq 1 - \alpha.$$

- $\underbrace{|\mathbb{E}[Y \mid f(x)] - f(x)| \leq \varepsilon}_{\text{calibration}} \;\Rightarrow\; \underbrace{\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x))}_{\text{CI w.r.t. } P_{(f(X),Y)}} := [f(x) - \varepsilon, f(x) + \varepsilon]$
  (ε, α)-calibration implies $(1-\alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$.

- Conversely, if $\hat{C}$ is $(1-\alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$,
  $\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x)) \;\Rightarrow\; |\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] - m_{\hat{C}}(f(x))| \leq 0.5|\hat{C}(f(x))|$
  where $m_{\hat{C}}(f(X))$ is the middle point of $\hat{C}(f(x))$.
  $\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] = \mathbb{E}[Y \mid f(x)]$ if $m_{\hat{C}}$ is injective
  $(1-\alpha)$-DF conf. interval implies $(0.5 \sup_{f(x)} |\hat{C}(f(x))|, \alpha)$-calibration

Calibration

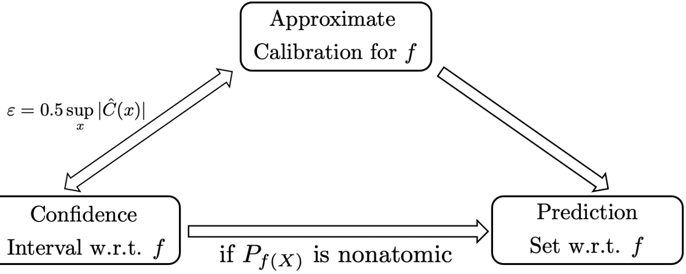
- A predictor $f : \mathbb{R}^d \to [0, 1]$ is (perfectly) calibrated if $\mathbb{E}[Y \mid f(X)] = f(X)$.
- Approximate calibration: (ε, α)-calibrated if

$$\mathbb{P}\left(|\mathbb{E}[Y \mid f(X)] - f(X)| \le \varepsilon\right) \ge 1 - \alpha.$$

- $\underbrace{|\mathbb{E}[Y \mid f(x)] - f(x)| \le \varepsilon}_{\text{calibration}} \;\Rightarrow\; \underbrace{\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x))}_{\text{CI w.r.t. } P_{(f(X),Y)}} := [f(x) - \varepsilon, f(x) + \varepsilon]$
  (ε, α)-calibration implies $(1 - \alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$.

- Conversely, if $\hat{C}$ is $(1 - \alpha)$-DF confidence interval w.r.t $P_{(f(X),Y)}$,
  $\mathbb{E}[Y \mid f(x)] \in \hat{C}(f(x)) \;\Rightarrow\; |\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] - m_{\hat{C}}(f(x))| \le 0.5|\hat{C}(f(x))|$
  where $m_{\hat{C}}(f(X))$ is the middle point of $\hat{C}(f(x))$.
  $\mathbb{E}[Y \mid m_{\hat{C}}(f(x))] = \mathbb{E}[Y \mid f(x)]$ if $m_{\hat{C}}$ is injective
  $(1 - \alpha)$-DF conf. interval implies $(0.5 \sup_{f(x)} |\hat{C}(f(x))|, \alpha)$-calibration

A “Tripod” Result



Impossibility of Asymptotic Calibration

Theorem (Informal)

It is impossible for an injective post-hoc calibration algorithm to be distribution-free asymptotically calibrated. (Given α fixed, $\varepsilon \to 0$ as $n \to \infty$.)

- Injectivity is a must: discretized recalibration map works.

Proof outline.

- Let $g_n : [0,1] \to [0,1]$ (with n observations) be the recalibration map.
- injectivity $\Rightarrow P_{g_n \circ f(X)}$ nonatomic for all n and nonatomic $P_{f(X)}$
- If $g_n \circ f$, with calib. error ε_n, is asymptotically calibrated, then

$$\hat{C}_n(f(X)) := [g_n \circ f(X) - \varepsilon_n, g_n \circ f(X) + \varepsilon_n]$$

  is a $(1-\alpha)$-DF confidence interval.
- $\text{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) = 2\varepsilon_n \to 0$ contradicts $\text{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) \geq L_\alpha(P_{f(X)})$.

Impossibility of Asymptotic Calibration

Theorem (Informal)

It is impossible for an injective post-hoc calibration algorithm to be distribution-free asymptotically calibrated. (Given α fixed, $\varepsilon \to 0$ as $n \to \infty$.)

- Injectivity is a must: discretized recalibration map works.

Proof outline.

- Let $g_n : [0,1] \to [0,1]$ (with n observations) be the recalibration map.
- injectivity $\Rightarrow P_{g_n \circ f(X)}$ nonatomic for all n and nonatomic $P_{f(X)}$
- If $g_n \circ f$, with calib. error ε_n, is asymptotically calibrated, then

$$\hat{C}_n(f(X)) := [g_n \circ f(X) - \varepsilon_n, g_n \circ f(X) + \varepsilon_n]$$

  is a $(1-\alpha)$-DF confidence interval.
- $\text{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) = 2\varepsilon_n \to 0$ contradicts $\text{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) \geq L_\alpha(P_{f(X)})$.

Impossibility of Asymptotic Calibration

Theorem (Informal)

It is impossible for an injective post-hoc calibration algorithm to be distribution-free asymptotically calibrated. (Given α fixed, $\varepsilon \to 0$ as $n \to \infty$.)

- Injectivity is a must: discretized recalibration map works.

Proof outline.

- Let $g_n : [0,1] \to [0,1]$ (with n observations) be the recalibration map.
- injectivity $\Rightarrow P_{g_n \circ f(X)}$ nonatomic for all n and nonatomic $P_{f(X)}$
- If $g_n \circ f$, with calib. error ε_n, is asymptotically calibrated, then

$$\hat{C}_n(f(X)) := [g_n \circ f(X) - \varepsilon_n, g_n \circ f(X) + \varepsilon_n]$$

  is a $(1-\alpha)$-DF confidence interval.
- $\text{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) = 2\varepsilon_n \to 0$ contradicts $\text{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) \geq L_\alpha(P_{f(X)})$.

Impossibility of Asymptotic Calibration

Theorem (Informal)

It is impossible for an injective post-hoc calibration algorithm to be distribution-free asymptotically calibrated. (Given α fixed, $\varepsilon \to 0$ as $n \to \infty$.)

- Injectivity is a must: discretized recalibration map works.

Proof outline.

- Let $g_n : [0,1] \to [0,1]$ (with n observations) be the recalibration map.
- injectivity $\Rightarrow P_{g_n \circ f(X)}$ nonatomic for all n and nonatomic $P_{f(X)}$
- If $g_n \circ f$, with calib. error ε_n, is asymptotically calibrated, then

$$\hat{C}_n(f(X)) := [g_n \circ f(X) - \varepsilon_n, g_n \circ f(X) + \varepsilon_n]$$

  is a $(1-\alpha)$-DF confidence interval.
- $\text{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) = 2\varepsilon_n \to 0$ contradicts $\text{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) \geq L_\alpha(P_{f(X)})$.

Impossibility of Asymptotic Calibration

Theorem (Informal)

It is impossible for an injective post-hoc calibration algorithm to be distribution-free asymptotically calibrated. (Given α fixed, $\varepsilon \to 0$ as $n \to \infty$.)

- Injectivity is a must: discretized recalibration map works.

Proof outline.

- Let $g_n : [0,1] \to [0,1]$ (with n observations) be the recalibration map.
- injectivity $\Rightarrow P_{g_n \circ f(X)}$ nonatomic for all n and nonatomic $P_{f(X)}$
- If $g_n \circ f$, with calib. error ε_n, is asymptotically calibrated, then

$$\hat{C}_n(f(X)) := [g_n \circ f(X) - \varepsilon_n, g_n \circ f(X) + \varepsilon_n]$$

is a $(1-\alpha)$-DF confidence interval.

- $\text{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) = 2\varepsilon_n \to 0$ contradicts $\text{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) \geq L_\alpha(P_{f(X)})$.

Impossibility of Asymptotic Calibration

Theorem (Informal)

It is impossible for an injective post-hoc calibration algorithm to be distribution-free asymptotically calibrated. (Given α fixed, $\varepsilon \to 0$ as $n \to \infty$.)

- Injectivity is a must: discretized recalibration map works.

Proof outline.

- Let $g_n : [0,1] \to [0,1]$ (with n observations) be the recalibration map.
- injectivity $\Rightarrow P_{g_n \circ f(X)}$ nonatomic for all n and nonatomic $P_{f(X)}$
- If $g_n \circ f$, with calib. error ε_n, is asymptotically calibrated, then

$$\hat{C}_n(f(X)) := [g_n \circ f(X) - \varepsilon_n, g_n \circ f(X) + \varepsilon_n]$$

  is a $(1-\alpha)$-DF confidence interval.
- $\mathrm{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) = 2\varepsilon_n \to 0$ contradicts $\mathrm{len}_{n,\alpha}(\hat{C}_n, P_{f(X)}) \geq L_\alpha(P_{f(X)})$.

A Small Advertisement

T-Cal: An optimal test for the calibration of predictive models

https://arxiv.org/abs/2203.01850

What we show (parts):

- Given a predictor f and a finite sample (any large),

  verifying f is (perfectly) calibrated or not is impossible in general!

- Assuming $\mathbb{E}[Y \mid f(X)]$ is Hölder smooth w.r.t $f(x)$,

  we propose T-Cal to test its calibration (large or small?),

  and show its minimax optimality!

Overview

Summary

Main takeaways:

- Notions (C.I., P.S., and Calib.) in uncertainty quantification are closely related. Tripod
- Impossibility (in the DF sense) of

$$\begin{cases} \text{vanishing length of confidence interval} \\ \text{calibrating a predictor asymptotically} \\ \text{verifying the calibration of a predictor} \end{cases}$$

  Main challenge: you only observe each datapoint (x or $f(x)$) once for nonatomic distributions
- Possibility exists for certain classes of distributions, e.g., smooth, discrete.

Summary

Main takeaways:

- Notions (C.I., P.S., and Calib.) in uncertainty quantification are closely related. Tripod
- Impossibility (in the DF sense) of

$$\begin{cases} \text{vanishing length of confidence interval} \\ \text{calibrating a predictor asymptotically} \\ \text{verifying the calibration of a predictor} \end{cases}$$

  Main challenge: you only observe each datapoint (x or $f(x)$) once for nonatomic distributions

- Possibility exists for certain classes of distributions, e.g., smooth, discrete.

Summary

Main takeaways:

- Notions (C.I., P.S., and Calib.) in uncertainty quantification are closely related. Tripod
- Impossibility (in the DF sense) of

$$\begin{cases} \text{vanishing length of confidence interval} \\ \text{calibrating a predictor asymptotically} \\ \text{verifying the calibration of a predictor} \end{cases}$$

  Main challenge: you only observe each datapoint (x or $f(x)$) once for nonatomic distributions

- Possibility exists for certain classes of distributions, e.g., smooth, discrete.

Summary

Main takeaways:

- Notions (C.I., P.S., and Calib.) in uncertainty quantification are closely related. Tripod
- Impossibility (in the DF sense) of

$$\begin{cases} \text{vanishing length of confidence interval} \\ \text{calibrating a predictor asymptotically} \\ \text{verifying the calibration of a predictor} \end{cases}$$

  Main challenge: you only observe each datapoint (x or $f(x)$) once for nonatomic distributions

- Possibility exists for certain classes of distributions, e.g., smooth, discrete.

Summary

Main takeaways:

- Notions (C.I., P.S., and Calib.) in uncertainty quantification are closely related. Tripod
- Impossibility (in the DF sense) of

$$\begin{cases} \text{vanishing length of confidence interval} \\ \text{calibrating a predictor asymptotically} \\ \text{verifying the calibration of a predictor} \end{cases}$$

  Main challenge: you only observe each datapoint (x or $f(x)$) once for nonatomic distributions

- Possibility exists for certain classes of distributions, e.g., smooth, discrete.

Thank you!