Conformal Prediction with missing values (2023 ICML)

Hwichang Jeong

February 26, 2025

Table of Contents

Introduction

Background

- Conventional quantile regression-based conformal prediction methods tend to construct prediction intervals that undercover the response conditionally to some missing patterns.

Contribution

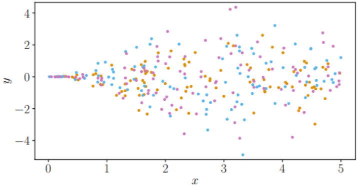
- Suggest novel conformalized quantile regression framework, missing data augmentation, which yields prediction intervals that are valid conditionally to the patterns of missing values.
- Prove that the proposed algorithms satisfy desirable theoretical properties.

Table of Contents

Notation

- $X = (X_1, \cdots, X_d) \in \mathbb{R}^d$: d-dimensional feature vector.
- $M = (M_1, \cdots, M_d) \in \{0, 1\}^d$: A mask such that M_j=0 when X_j is observed and $M_j = 1$ otherwise.
- $Y \in \mathbb{R}$: Outcome.
- $\mathcal{M} = \{0, 1\}^d$: Set of masks.
- For a mask $m \in \mathcal{M}$, $X_{obs(m)}$ is the random vector of observed components, and $X_{mis(m)}$ is the random vector of unobserved one.
- For $(\mathring{m}, \breve{m}) \in \mathcal{M}^2$, $\mathring{m} \subset \breve{m}$ denotes $\mathring{m}_j = 1$ then $\breve{m}_j = 1$ for any $j \in \{1, \cdots, d\}$, i.e. $\breve{m}$ includes at least the same missing values than $\mathring{m}$.
- α : Miscoverage rate

(Split) Conformalized Quantile Regression (CQR)

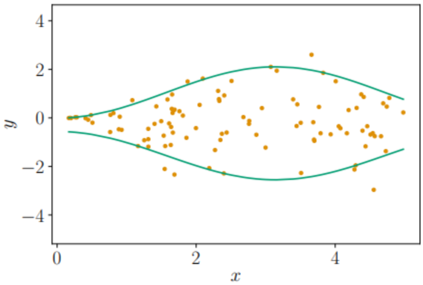


- Create a proper training set, a calibration set, and keep your test set, by randomly splitting your data set.

Proper training set : Tr / Calibration set : Cal / Test set : Te

(Split) Conformalized Quantile Regression (CQR)

Step 1



On the proper training set:

- Learn $\hat{q}_{\text{low}}$ and $\hat{q}_{\text{upp}}$

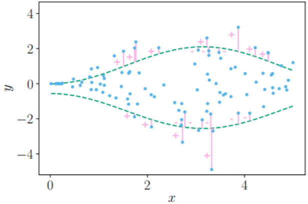
$$\text{low} = \alpha/2, \text{high} = 1 - \alpha/2$$

$$\hat{q}_{\text{low}}(X) = \text{QuantileRegression}(X, \alpha/2)$$

$$\hat{q}_{\text{high}}(X) = \text{QuantileRegression}(X, 1 - \alpha/2)$$

(Split) Conformalized Quantile Regression (CQR)

Step 2

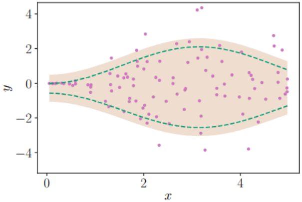


On the calibration set:

- Predict with $\hat{q}_{\text{low}}$ and $\hat{q}_{\text{upp}}$
- Get the scores
  $s^{(k)} = \max\left\{\hat{q}_{\text{low}}\left(x^{(k)}\right) - y^{(k)}, y^{(k)} - \hat{q}_{\text{upp}}\left(x^{(k)}\right)\right\}$
- Compute the $(1-\alpha) \times (1 + \frac{1}{\#\text{Cal}})$ empirical quantile of the $s^{(k)}$, noted $\widehat{Q}_{1-\hat{\alpha}}(S)$

(Split) Conformalized Quantile Regression (CQR)

Step 3



On the test set:

- Predict with $\hat{q}_{\text{low}}$ and $\hat{q}_{\text{upp}}$
- Build $\hat{C}_{\hat{\alpha}}(x)$: $[\hat{q}_{\text{low}}(x) - \widehat{Q}_{1-\hat{\alpha}}(S), \hat{q}_{\text{upp}}(x) + \widehat{Q}_{1-\hat{\alpha}}(S)]$

(Split) Conformalized Quantile Regression (CQR)

Theorem (1)

If $\left(X^{(k)}, Y^{(k)}\right), k = 1, \ldots, n+1$ *are exchangeable, then the prediction interval* $\hat{C}\left(X^{(n+1)}\right)$ *constructed by the split CQR algorithm satisfies*

$$\mathbb{P}\left\{Y^{(n+1)} \in \hat{C}\left(X^{(n+1)}\right)\right\} \geq 1 - \alpha$$

Moreover, if the conformity scores $s^{(k)}$ *are almost surely distinct, then the prediction interval is nearly perfectly calibrated:*

$$\mathbb{P}\left\{Y_{n+1} \in \hat{C}\left(X_{n+1}\right)\right\} \leq 1 - \alpha + 1/\left(\#Cal + 1\right)$$

Validity with Missing values

$$1 - \alpha \leq \mathbb{P}\left\{Y_{n+1} \in \hat{C}\left(X^{(n+1)}, M^{(n+1)}\right)\right\} \leq 1 - \alpha + \frac{1}{(\#\text{Cal} + 1)}$$

Impute then predict + conformalization

- Let $\phi^m : \mathbb{R}^{|obs(m)|} \to \mathbb{R}^{|mis(m)|}$ be a imputation function.
- Let $\Phi = (\Phi_1, \cdots, \Phi_d) : \mathbb{R}^d \times \{0,1\}^d \to \mathbb{R}^d$ be a function such that

$$\Phi_j(X, M) = X_j \mathbb{I}_{M_j=0} + \phi_j^M(X_{obs(M)}) \mathbb{I}_{M_j=1}$$

Theorem (2)

Assume exchangeability holds and the imputation function Φ is the output of an algorithm $\mathcal{I}$ treating its input data points symmetrically: $\mathcal{I}\left(\left(X^{(\sigma(k))}, M^{(\sigma(k))}\right)_{k=1}^{n+1}\right) \overset{(d)}{=} \mathcal{I}\left(\left(X^{(k)}, M^{(k)}\right)_{k=1}^{n+1}\right)$ conditionally on $\left(X^{(k)}, M^{(k)}\right)_{k=1}^{n+1}$ and for any permutation σ. Then, the prediction interval $\hat{C}\left(X^{(n+1)}, M^{(n+1)}\right)$ satisfies

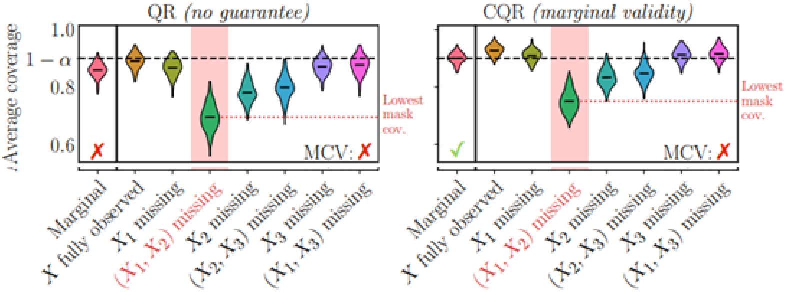
$$1 - \alpha \leq \mathbb{P}\left\{Y_{n+1} \in \hat{C}\left(X^{(n+1)}, M^{(n+1)}\right)\right\}$$

. Also if the conformity scores $s^{(k)}$ are almost surely distinct,

$$1 - \alpha \leq \mathbb{P}\left\{Y_{n+1} \in \hat{C}\left(X^{(n+1)}, M^{(n+1)}\right)\right\} \leq 1 - \alpha + \frac{1}{(\#Cal + 1)}$$

is holds.

Mask-Conditional-Validity(MCV)



- Gaussian linear regression with $d = 3$.
- Missingness mechanism : MCAR / Missing rate : 20%.
- Imputation : Iterative regression.

Mask-Conditional-Validity(MCV)

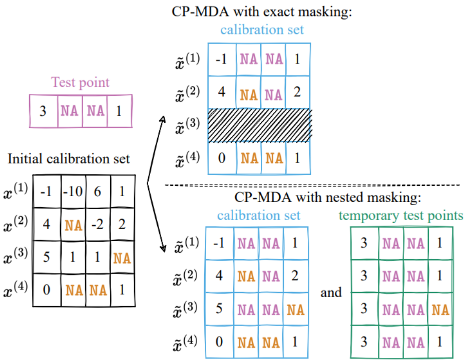
For any $m \in \mathcal{M}$,

$$1-\alpha \leq \mathbb{P}\left(Y^{(n+1)} \in \widehat{C}_{\alpha}\left(X^{(n+1)}, m\right) \mid M^{(n+1)} = m\right) \leq 1-\alpha+\frac{1}{\#\mathrm{Cal}^{\mathrm{m}}+1},$$

where $\mathrm{Cal}^{\mathrm{m}} = \left\{\mathrm{k} \in \mathrm{Cal} \text{ such that } \mathrm{m}^{(\mathrm{k})} \subset \mathrm{m}\right\}$.

MCV

- CP-MDA-Exact : Select the data in the calibration set (Cal) that contains missing values included in the missing columns of the test data + Additional Masking
- CP-MDA-Nested : Masking the calibration set with max-mask + temporary test point (quantile of confidence interval induced by TTP).



Theorem - MCV

Assume missing mechanism is MCAR and $(Y \perp M)|X$

- For all $m \in \mathcal{M}$, CP-MDA-Exact satisfies

$$1-\alpha \leq \mathbb{P}\left(Y^{(n+1)} \in \widehat{C}_\alpha\left(X^{(n+1)}, m\right) \mid M^{(n+1)} = m\right),$$

- If the conformity scores $s^{(k)}$ are almost surely distinct, for all $m \in \mathcal{M}$, CP-MDA-Exact satisfies,

$$1-\alpha \leq \mathbb{P}\left(Y^{(n+1)} \in \widehat{C}_\alpha\left(X^{(n+1)}, m\right) \mid M^{(n+1)} = m\right) \leq 1-\alpha+\frac{1}{\#\mathrm{Cal}^{\mathrm{m}}+1},$$

where $\mathrm{Cal}^{\mathrm{m}} = \{\mathrm{k} \in \mathrm{Cal} \text{ such that } \mathrm{m}^{(\mathrm{k})} \subset \mathrm{m}\}$.

Theorem - MCV

Assume missing mechanism is MCAR and $(Y \perp M)|X$

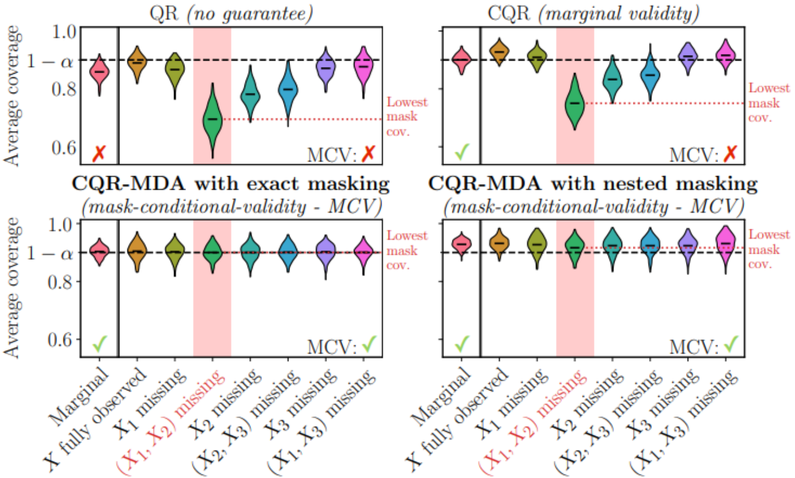
- Let $(\mathring{m}, \breve{m}) \in \mathcal{M}^2$. If $\mathring{m} \subset \breve{m}$ then for any $\delta \in [0, 0.5]$, $q_{1-\delta/2}^{Y|\left(X_{\text{obs}(\mathring{m})}, M=\breve{m}\right)} \leq q_{1-\delta/2}^{Y|\left(X_{\text{obs}(\breve{m})}, M=\breve{m}\right)}, q_{\delta/2}^{Y|\left(X_{\text{obs}(\mathring{m})}, M=\breve{m}\right)} \geq q_{\delta/2}^{Y|\left(X_{\text{obs}(\breve{m})}, M=\breve{m}\right)}$ are hold.
  Then, CP-MDA-Nested satisfies

$$1 - \alpha \leq \mathbb{P}\left(Y^{(n+1)} \in \widehat{C}_\alpha\left(X^{(n+1)}, m\right) \mid M^{(n+1)} = m\right),$$

up to a technical minor modification of the output.

Table of Contents

Experiment



QR (no guarantee)
CQR (marginal validity)
CQR-MDA with exact masking
(mask-conditional-validity - MCV)
CQR-MDA with nested masking
(mask-conditional-validity - MCV)
Average coverage
1.0
1 − α
0.8
0.6
Lowest mask cov.
MCV: ✗
MCV: ✓
Marginal
X fully observed
X1 missing
(X1, X2) missing
X2 missing
(X2, X3) missing
X3 missing
(X1, X3) missing