# Predictive uncertainty quantification with missing covariates

Margaux Zaffran

June 13, 2024

DATA Seminar

# Who am I?

- 3rd (last) year statistics PhD Student, @ INRIA & École Polytechnique
- Funded by Électricité de France
- My advisors:

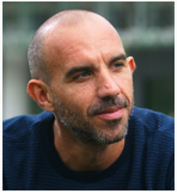
**Aymeric Dieuleveut**

*École Polytechnique*

**Olivier Féron**

*EDF R&D*

*FiME*



**Yannig Goude**

*EDF R&D*

*LMO*

**Julie Josse**

*PreMeDICaL*

*INRIA*

- Research interests:
  - Distribution-free uncertainty quantification
  - Time series data
  - Missing values
  - Real life applications (energy, environmental, medical and societal domains)

# Predictive Uncertainty Quantification with Missing Covariates

**Yaniv Romano**

*Technion - Israel Institute of Technology*


**Julie Josse**

*PreMeDICaL*

*INRIA*

**Aymeric Dieuleveut**

*École Polytechnique*

*Conformal Prediction with Missing Values*, ICML 2023

*Predictive Uncertainty Quantification with Missing Covariates*, submitted in 2024

# Quantifying predictive uncertainty

- $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ random variables
- $n$ training samples $\left(X^{(k)}, Y^{(k)}\right)_{k=1}^{n}$
- Goal: predict an unseen point $Y^{(n+1)}$ at $X^{(n+1)}$ with **confidence**
- How? Given a miscoverage level $\alpha \in [0, 1]$, build a predictive set $\mathcal{C}_\alpha$ such that:

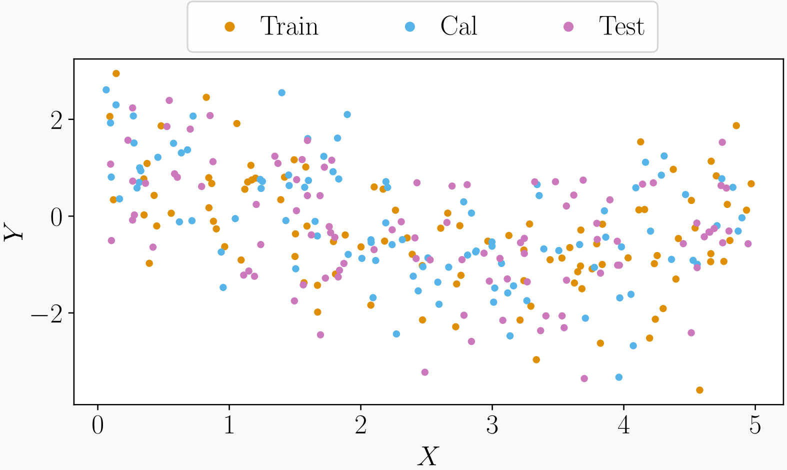
$$\mathbb{P}\left\{Y^{(n+1)} \in \mathcal{C}_\alpha\left(X^{(n+1)}\right)\right\} \geq 1 - \alpha, \tag{1}$$

  and $\mathcal{C}_\alpha$ should be as small as possible, in order to be informative.

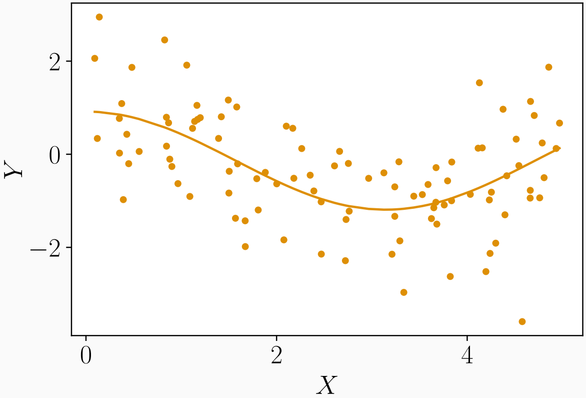
  *For example:* $\alpha = 0.1$ and obtain a 90% coverage interval

- Construction of the predictive intervals should be
  - agnostic to the model
  - agnostic to the data distribution
  - valid in finite samples

# Split Conformal Prediction (SCP)[1,2,3]: toy example



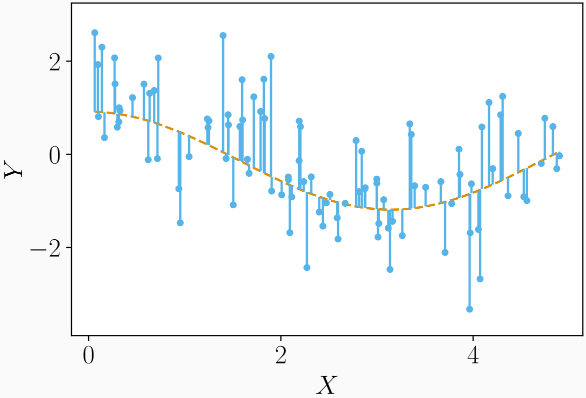

[1] Vovk et al. (2005), *Algorithmic Learning in a Random World*
[2] Papadopoulos et al. (2002), *Inductive Confidence Machines for Regression*, ECML
[3] Lei et al. (2018), *Distribution-Free Predictive Inference for Regression*, JRSS B




- Learn (or get) $\hat{\mu}$

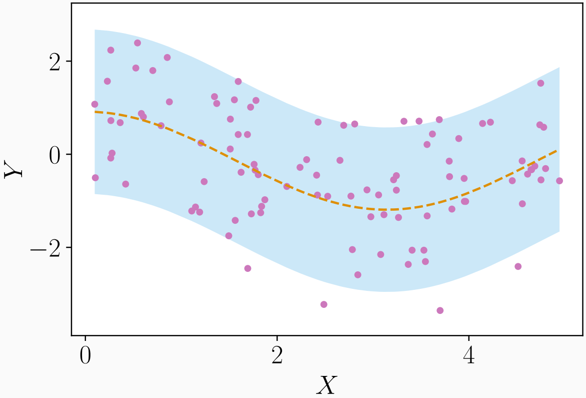
[1] Vovk et al. (2005), *Algorithmic Learning in a Random World*
[2] Papadopoulos et al. (2002), *Inductive Confidence Machines for Regression*, ECML
[3] Lei et al. (2018), *Distribution-Free Predictive Inference for Regression*, JRSS B

# Split Conformal Prediction (SCP)[1,2,3]: toy example




- Predict with $\hat{\mu}$
- Get the $|$residuals$|$, a.k.a. conformity scores
- Compute the $(1-\alpha)$ empirical quantile of $\mathcal{S} = \{|\text{residuals}|\}_{\text{Cal}} \cup \{+\infty\}$, noted $q_{1-\alpha}(\mathcal{S})$

[1] Vovk et al. (2005), *Algorithmic Learning in a Random World*
[2] Papadopoulos et al. (2002), *Inductive Confidence Machines for Regression*, ECML
[3] Lei et al. (2018), *Distribution-Free Predictive Inference for Regression*, JRSS B




- Predict with $\hat{\mu}$
- Build $\widehat{C}_\alpha(x)$: $[\hat{\mu}(x) \pm q_{1-\alpha}(\mathcal{S})]$

[1] Vovk et al. (2005), *Algorithmic Learning in a Random World*
[2] Papadopoulos et al. (2002), *Inductive Confidence Machines for Regression*, ECML
[3] Lei et al. (2018), *Distribution-Free Predictive Inference for Regression*, JRSS B

## Standard mean-regression SCP: implementation details

1. Randomly split the training data into a proper training set (size $\#\mathrm{Tr}$) and a calibration set (size $\#\mathrm{Cal}$)
2. Get $\hat{\mu}$ *by training the algorithm* $\mathcal{A}$ *on the proper training set*
3. On the calibration set, get prediction values with $\hat{\mu}$
4. Obtain a set of $\#\mathrm{Cal} + 1$ conformity scores:

$$\mathcal{S} = \{S_i = |\hat{\mu}(X_i) - Y_i|, i \in \mathrm{Cal}\} \cup \{+\infty\}$$

(+ worst-case scenario)

5. Compute the $1 - \alpha$ empirical quantile of these scores, noted $q_{1-\alpha}(\mathcal{S})$
6. For a new point $X_{n+1}$, return

$$\widehat{C}_\alpha(X_{n+1}) = [\hat{\mu}(X_{n+1}) - q_{1-\alpha}(\mathcal{S}); \hat{\mu}(X_{n+1}) + q_{1-\alpha}(\mathcal{S})]$$

# Standard mean-regression SCP: theoretical foundation

**Exchangeability**

$\left(X^{(k)}, Y^{(k)}\right)_{k=1}^{n}$ are exchangeable if for any permutation $\sigma$ of $[\![1, n]\!]$ we have:

$$\left(X^{(1)}, Y^{(1)}\right), \ldots, \left(X^{(n)}, Y^{(n)}\right) \stackrel{d}{=} \left(X^{(\sigma(1))}, Y^{(\sigma(1))}\right), \ldots, \left(X^{(\sigma(n))}, Y^{(\sigma(n))}\right).$$

**Examples of exchangeable sequences**

- i.i.d. samples
- The components of $\mathcal{N}\left(\begin{pmatrix} m \\ \vdots \\ \vdots \\ m \end{pmatrix}, \begin{pmatrix} \sigma^2 & & \gamma^2 & \\ & \ddots & & \\ \gamma^2 & & \ddots & \\ & & & \sigma^2 \end{pmatrix}\right)$

# Standard mean-regression SCP: theoretical guarantees

**Standard mean-regression SCP marginal validity (Vovk et al., 2005; Lei et al., 2018)**

Suppose $\left(X^{(k)}, Y^{(k)}\right)_{k=1}^{n+1}$ are exchangeable (or i.i.d.)[a]. Standard mean-regression SCP applied on $\left(X^{(k)}, Y^{(k)}\right)_{k=1}^{n}$ outputs $\widehat{C}_\alpha(\cdot)$ such that:

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

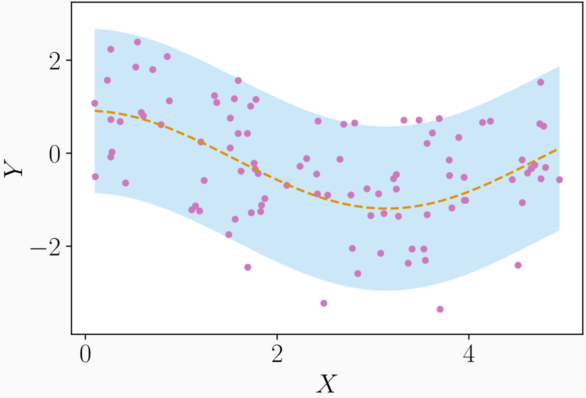
Additionally, if the scores $\left\{S^{(k)}\right\}_{k\in\mathrm{Cal}} \cup \left\{S^{(n+1)}\right\}$ are a.s. distinct:

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

[a]Only the calibration and test data need to be exchangeable.

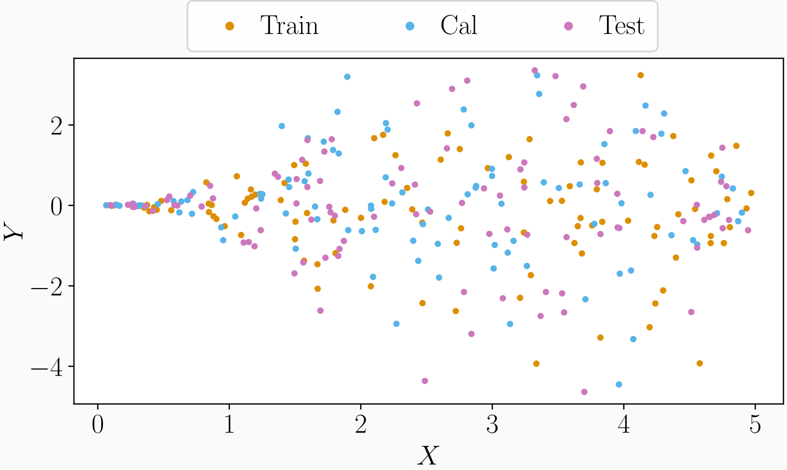
✗ Marginal coverage: $\mathbb{P}\left\{Y^{(n+1)} \in \widehat{C}_\alpha\left(X^{(n+1)}\right) \cancel{\mid X^{(n+1)} = x}\right\} \geq 1-\alpha$

# Standard mean-regression SCP is not adaptive

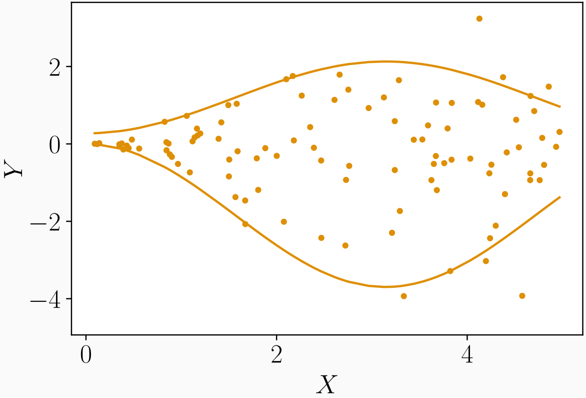



- Predict with $\hat{\mu}$
- Build $\widehat{C}_\alpha(x)$: $[\hat{\mu}(x) \pm q_{1-\alpha}(\mathcal{S})]$

# Conformalized Quantile Regression (CQR)[4]

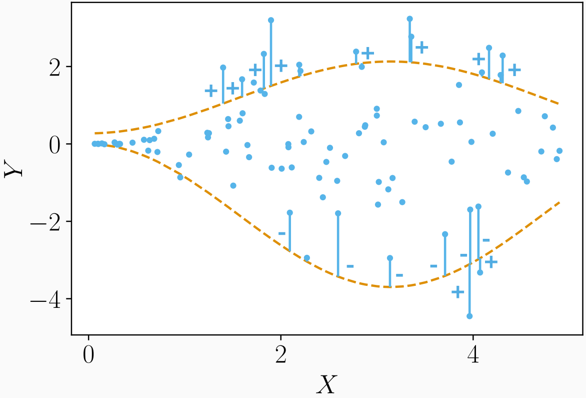



[4] Romano et al. (2019), *Conformalized Quantile Regression*, NeurIPS




- Learn (or get) $\widehat{\mathrm{QR}}_{\mathrm{lower}}$ and $\widehat{\mathrm{QR}}_{\mathrm{upper}}$

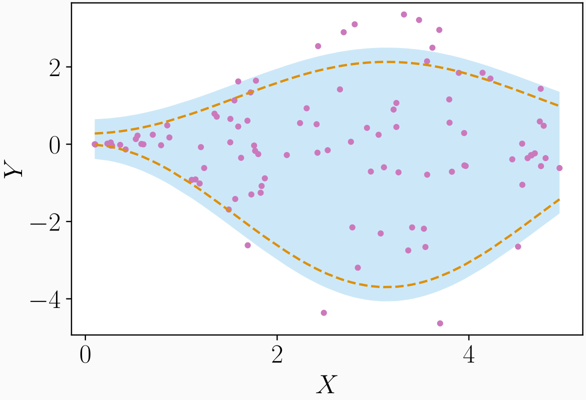
[4] Romano et al. (2019), *Conformalized Quantile Regression*, NeurIPS




- Predict with $\widehat{\mathrm{QR}}_{\mathrm{lower}}$ and $\widehat{\mathrm{QR}}_{\mathrm{upper}}$
- Get the scores $\mathcal{S} = \{S_i\}_{\mathrm{Cal}} \cup \{+\infty\}$
- Compute the $(1-\alpha)$ empirical quantile of $\mathcal{S}$, noted $q_{1-\alpha}(\mathcal{S})$

$$\hookrightarrow \quad S_i := \max\left\{\widehat{\mathrm{QR}}_{\mathrm{lower}}(X_i) - Y_i,\, Y_i - \widehat{\mathrm{QR}}_{\mathrm{upper}}(X_i)\right\}$$

[4] Romano et al. (2019), *Conformalized Quantile Regression*, NeurIPS




- Predict with $\widehat{\text{QR}}_{\text{lower}}$ and $\widehat{\text{QR}}_{\text{upper}}$

- Build

$$\widehat{C}_\alpha(x) = [\widehat{\text{QR}}_{\text{lower}}(x) - q_{1-\alpha}(\mathcal{S}); \widehat{\text{QR}}_{\text{upper}}(x) + q_{1-\alpha}(\mathcal{S})]$$

[4]Romano et al. (2019), *Conformalized Quantile Regression*, NeurIPS

# CQR: theoretical guarantees

**CQR marginal validity (Romano et al., 2019)**

Suppose $\left(X^{(k)}, Y^{(k)}\right)_{k=1}^{n+1}$ are exchangeable (or i.i.d.)[a]. CQR applied on $\left(X^{(k)}, Y^{(k)}\right)_{k=1}^{n}$ outputs $\widehat{C}_\alpha(\cdot)$ such that:

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

Additionally, if the scores $\left\{S^{(k)}\right\}_{k \in \mathrm{Cal}} \cup \left\{S^{(n+1)}\right\}$ are a.s. distinct:

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

[a] Only the calibration and test data need to be exchangeable.

✗ Marginal coverage: $\mathbb{P}\left\{Y_{n+1} \in \widehat{C}_\alpha\left(X_{n+1}\right) \cancel{| X_{n+1}=x}\right\} \geq 1-\alpha$

# SCP is defined by the conformity score function

1. Randomly split the training data into a proper training set (size $\#\mathrm{Tr}$) and a calibration set (size $\#\mathrm{Cal}$)
2. Get $\hat{A}$ *by training the algorithm* $\mathcal{A}$ *on the proper training set*
3. On the calibration set, obtain $\#\mathrm{Cal} + 1$ conformity scores

$$\mathcal{S} = \{S_i = \mathrm{s}(\hat{A}(X_i), Y_i), i \in \mathrm{Cal}\} \cup \{+\infty\}$$

Ex 1: $\mathrm{s}(\hat{A}(X_i), Y_i) := |\hat{\mu}(X_i) - Y_i|$ in regression with standard scores

Ex 2: $\mathrm{s}(\hat{A}(X_i), Y_i) := \max\left(\widehat{\mathrm{QR}}_{\mathrm{lower}}(X_i) - Y_i, Y_i - \widehat{\mathrm{QR}}_{\mathrm{upper}}(X_i)\right)$ in CQR

4. Compute the $1 - \alpha$ empirical quantile of these scores, noted $q_{1-\alpha}(\mathcal{S})$
5. For a new point $X_{n+1}$, return

$$\widehat{C}_\alpha(X_{n+1}) = \{y \text{ such that } \mathrm{s}(\hat{A}(X_{n+1}), y) \leq q_{1-\alpha}(\mathcal{S})\}$$

↪ The definition of the conformity scores is crucial, as they incorporate almost all the information: data + underlying model

**SCP marginal validity (Vovk et al., 2005)**

Suppose $\left(X^{(k)}, Y^{(k)}\right)_{k=1}^{n+1}$ are exchangeable (or i.i.d.)[a]. SCP applied on $\left(X^{(k)}, Y^{(k)}\right)_{k=1}^{n}$ outputs $\widehat{C}_\alpha(\cdot)$ such that:

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

Additionally, if the scores $\left\{S^{(k)}\right\}_{k \in \mathrm{Cal}} \cup \left\{S^{(n+1)}\right\}$ are a.s. distinct:

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

[a] Only the calibration and test data need to be exchangeable.

✗ Marginal coverage: $\mathbb{P}\left\{Y_{n+1} \in \widehat{C}_\alpha\left(X_{n+1}\right) \cancel{| X_{n+1} = x}\right\} \geq 1 - \alpha$

## SCP: summary

Split Conformal Prediction is simple to compute and works:

- ✓ any regression (and classification) algorithm (neural nets, random forest...);
- ✓ distribution-free as long as the data is exchangeable;
- ✓ finite sample.

✗ Note that the theoretical guarantee is **marginal** over the joint distribution of $(X, Y)$, and **not conditional**. In particular, features conditional validity is not ensured: there is no guarantee that for any $x \in \mathcal{X}$

$$\mathbb{P}\left\{Y^{(n+1)} \in \widehat{C}_\alpha\left(X^{(n+1)}\right) | X^{(n+1)} = x\right\} \geq 1 - \alpha.$$

# Informative conditional coverage as such is impossible

- Impossibility results

  $\hookrightarrow$ Vovk (2012); Lei and Wasserman (2014); Barber et al. (2021)

> Without distribution assumption, in finite sample, a perfectly features conditionally valid $\widehat{C}_\alpha$ is such that $\mathbb{P}\left\{\text{mes}\left(\widehat{C}_\alpha(x)\right) = \infty\right\} \geq 1-\alpha$ for any non-atomic $x$.

- Approximate conditional coverage

  $\hookrightarrow$ Romano et al. (2020); Guan (2022); Jung et al. (2023); Gibbs et al. (2023)

  Target $\mathbb{P}(Y_{n+1} \in \widehat{C}_\alpha | X_{n+1} \in \mathcal{R}(x)) \geq 1-\alpha$

- Asymptotic (with the sample size) conditional coverage

  $\hookrightarrow$ **Romano et al. (2019)**; Kivaranovic et al. (2020); Chernozhukov et al. (2021); Sesia and Romano (2021); Izbicki et al. (2022)

Non exhaustive references.

# Missing values are ubiquitous and challenging

**Data:** $\left(X^{(k)}, M^{(k)}, Y^{(k)}\right)_{k=1}^{n}$

| $Y$ | $X_1$ | $X_2$ | $X_3$ |
|---|---|---|---|
| 22 | 5 | 6 | 3 |
| 19 | 6 | 8 | NA |
| 19 | 5 | 3 | 6 |
| 7 | NA | 9 | NA |
| 13 | 4 | 9 | 0 |
| 20 | NA | NA | 1 |
| 9 | 8 | NA | 4 |

Mask $M = (M_1\ M_2\ M_3)$

| $M_1$ | $M_2$ | $M_3$ |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 0 | 1 |
| 0 | 0 | 0 |
| 1 | 0 | 1 |
| 0 | 0 | 0 |
| 1 | 1 | 0 |
| 0 | 1 | 0 |

$\hookrightarrow$ $2^d$ potential masks.

$\hookrightarrow$ $M$ can depend on $X$ or $Y$ (depending on the missing mechanism[5]).

$\Rightarrow$ Statistical and computational challenges.

[5] Three mechanisms connecting $X$ and $M$ from Rubin (1976), *Inference and missing data*, Biometrika

# Supervised learning with missing values: impute-then-regress

Impute-then-regress procedures are widely used.

1. Replace NA using an imputation function (e.g. the mean), noted $\phi$.

| | | | | |
|---|---|---|---|---|
| $x^{(1)}$ | -1 | -10 | 6 | 0 |
| $x^{(2)}$ | 4 | NA | -2 | 2 |
| $x^{(3)}$ | 5 | 1 | 2 | NA |
| $x^{(4)}$ | 0 | NA | NA | 1 |

$\xrightarrow{\ \phi\ }$

| | | | | |
|---|---|---|---|---|
| $u^{(1)}$ | -1 | -10 | 6 | 0 |
| $u^{(2)}$ | 4 | -4.5 | -2 | 2 |
| $u^{(3)}$ | 5 | 1 | 2 | 1 |
| $u^{(4)}$ | 0 | -4.5 | 3 | 1 |

2. Train your algorithm (Random Forest, Neural Nets, etc.) on the imputed data:

$$\left\{ \underbrace{\phi\left(X^{(k)}_{\text{obs}(M^{(k)})}, M^{(k)}\right)}_{U^{(k)} = \text{imputed } X^{(k)}}, Y^{(k)} \right\}_{k=1}^{n}.$$

$\hookrightarrow$ we consider an impute-then-regress pipeline in this work.

# Goals of predictive uncertainty quantification with missing values

**Goal:** predict $Y^{(n+1)}$ with **confidence** $1-\alpha$, i.e. build the smallest $\mathcal{C}_\alpha$ such that:

**1. Marginal Validity (MV)**

$$\mathbb{P}\left\{Y^{(n+1)} \in \mathcal{C}_\alpha\left(X^{(n+1)}, M^{(n+1)}\right)\right\} \geq 1-\alpha. \qquad \text{(MV)}$$

*For example:* $\alpha = 0.1$ and obtain a 90% coverage interval.

**2. Mask-Conditional-Validity (MCV)**

$$\mathbb{P}\left\{Y^{(n+1)} \in \mathcal{C}_\alpha\left(X^{(n+1)}, m\right) | M^{(n+1)}\right\} \overset{a.s.}{\geq} 1-\alpha. \qquad \text{(MCV)}$$

# CP is marginally valid (MV) after imputation

**Exchangeability after imputation (Z., Dieuleveut, Josse and Romano, 2023)**

Assume $\left(X^{(k)}, M^{(k)}, Y^{(k)}\right)_{k=1}^{n}$ are i.i.d. (or exchangeable).
Then, for any missing mechanism, for almost all imputation function[a] $\phi$:
$\left(\phi\left(X^{(k)}_{\text{obs}(M^{(k)})}, M^{(k)}\right), Y^{(k)}\right)_{k=1}^{n}$ are **exchangeable**.

[a] Even if the imputation is not accurate, the guarantee will hold.

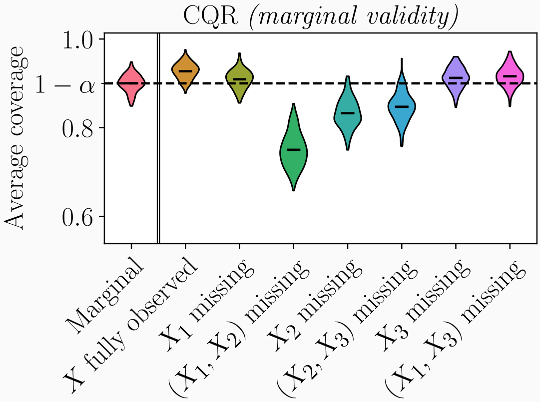
$\Rightarrow$ CQR, and Conformal Prediction, applied on an imputed data set still enjoys marginal guarantees[6]:

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

[6] The upper bound also holds under continuously distributed scores.

# CQR is marginally valid on imputed data sets

$Y = \beta^T X + \varepsilon$, $\beta = (1, 2, -1)^T$, $X$ and $\varepsilon$ Gaussian.




✓ Marginal (i.e. average) coverage (MV) is indeed recovered!

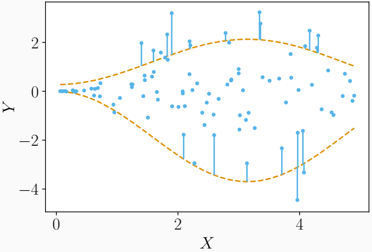
✗ Mask-conditional-validity (MCV) is not attained

↪ Missing values induce heteroskedasticity

*(supported by theory under (non-)parametric assumptions)*

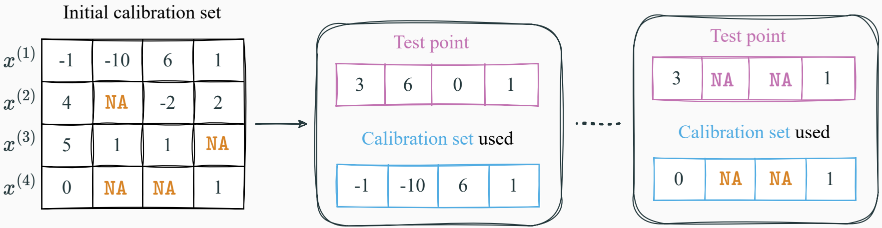
# Conformalization step is independent of the important variable: the mask!

**Observation:** the $\alpha$-correction term is computed among all the data points, regardless of their mask!




**Warning:** $2^d$ possible masks

$\Rightarrow$ Splitting the calibration set by mask is infeasible (lack of data)!



# Fully distribution-free MCV is necessarily uninformative

**General MCV hardness result** **(Z., Josse, Romano and Dieuleveut, 2024)**[7]

If any $\widehat{C}_\alpha$ is distribution-free MCV then **for any distribution** $P$, for any mask $m$ such that $P_M(m) > 0$, it holds:

$$\mathbb{P}_{P^{\otimes(n+1)}}\left(\text{mes}\left(\widehat{C}_\alpha\left(X_{n+1}, m\right)\right) = \infty\right) \geq 1 - \alpha - \Delta_{m,n} \geq 1 - \alpha - P_M(m)\sqrt{n+1}.$$

Irreducible term: consider $\widehat{C}_\alpha$ outputting $\mathcal{Y}$ with probability $1 - \alpha$ and $\emptyset$ otherwise.

$\Delta_{m,n}$ term: smaller than $P_M(m)\sqrt{n+1}$

↪ gets negligible (making the lower bound nearly $1 - \alpha$) for low probability masks compared to $n$;

↪ gets large (making the lower bound trivial because negative) for high probability masks compared to $n$.

[7] An analogous statement is also available for the classification framework.

**$M \perp\!\!\!\perp X$ hardness result (Z., Josse, Romano and Dieuleveut, 2024)**

If any $\widehat{C}_\alpha$ is MCV under $M \perp\!\!\!\perp X$, then for any distribution $P$ such that $M \perp\!\!\!\perp X$, for any mask $m$ such that $P_M(m) > 0$, it holds:

$$\mathbb{P}_{P^{\otimes(n+1)}}\left(\text{mes}\left(\widehat{C}_\alpha\left(X_{n+1}, m\right)\right) = \infty\right) \geq 1-\alpha-\Delta_{m,n} \geq 1-\alpha-P_M(m)\sqrt{n+1}.$$

**$Y \perp\!\!\!\perp M \mid X$ hardness result (Z., Josse, Romano and Dieuleveut, 2024)**

If any $\widehat{C}_\alpha$ is MCV under $Y \perp\!\!\!\perp M \mid X$, then for any distribution $P$ such that $Y \perp\!\!\!\perp M \mid X$, for any mask $m$ such that $\frac{1}{\sqrt{2}} \geq P_M(m) > 0$, it holds:

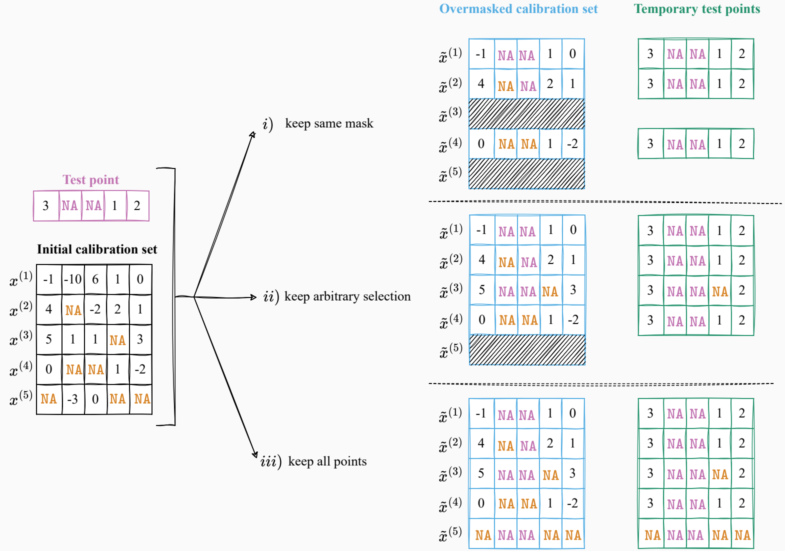
$$\mathbb{P}_{P^{\otimes(n+1)}}\left(\text{mes}\left(\widehat{C}_\alpha\left(X_{n+1}, m\right)\right) = \infty\right) \geq 1-\alpha-\Delta_{m,n} \geq 1-\alpha-2P_M(m)\sqrt{n+1}.$$

$\Rightarrow$ need to restrict both the link between $M$ and $X$, as well as between $M$ and $Y$.

Analogous statements are also available for the classification framework.

# CP-MDA-Nested* (Missing Data Augmentation)

**Idea:** for each test point, modify the calibration points to mimic the test mask



**Mask-conditional-validity of CP-MDA-Nested***
**(Z., Josse, Romano and Dieuleveut, 2024)**

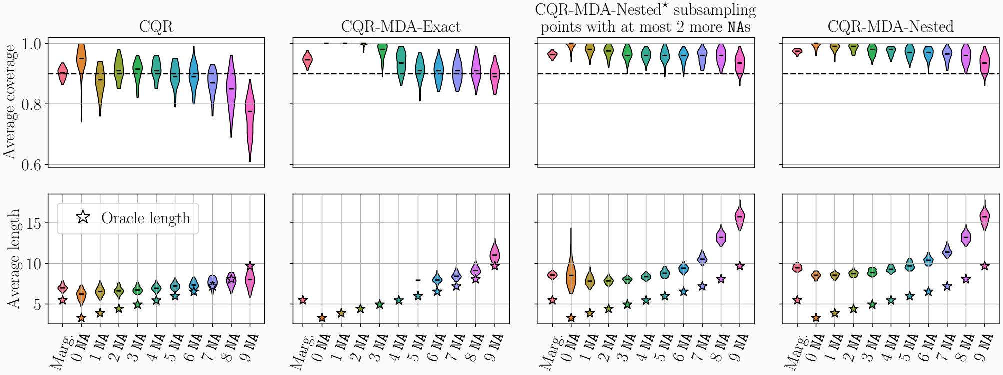
Under the assumptions that:

- $M \perp\!\!\!\perp X$,
- $Y \perp\!\!\!\perp M \,|X$,
- $\left(X^{(k)}, M^{(k)}, Y^{(k)}\right)_{k=1}^{n+1}$ are i.i.d.,
- the subsampling scheme is independent of $\left(X^{(k)}, Y^{(k)}\right)_{k=1}^{n+1}$,

then, for almost all imputation function, CP-MDA-Nested* reaches (MCV) at the level $1-2\alpha$, that is:

$$\mathbb{P}\left\{Y^{(n+1)} \in \widehat{C}_\alpha\left(X^{(n+1)}, m\right) |M^{(n+1)}\right\} \overset{a.s.}{\geq} 1-2\alpha.$$

# Experiments on $M \perp\!\!\!\perp X$ and $Y \perp\!\!\!\perp M \mid X$ Gaussian linear data in dimension 10




40% of missing values

# Experiments beyond independence

- Under various MAR and MNAR mechanisms, `CP-MDA-Nested`$^\star$ maintains empirical MCV;
- When $Y \not\perp\!\!\!\perp M \mid X$ and the imputation is not accurate enough, `CP-MDA-Nested`$^\star$ fails to empirically ensure MCV, with a loss of coverage that is more critical when subsampling.

## Take-home-messages

- CP marginal guarantees hold on the imputed data set.
- CQR (and more generally CP) fails to attain coverage conditional on the missing pattern, i.e. MCV.
- Missingness introduces additional heteroskedasticity.
- MCV is impossible to ensure in an informative way without restricting both the dependence between $M$ and $X$, and between $M$ and $Y$.
- `CP-MDA-Nested`$^\star$ (Missing Data Augmentation) is the first method to output predictive intervals with missing values.
- `CP-MDA-Nested`$^\star$ attains conditional coverage with respect to the missing pattern (in MCAR and $Y \perp\!\!\!\perp M \mid X$ setting).
- `CP-MDA-Nested`$^\star$ is empirically robust to non-MCAR scenarii.

# Extensions

- Consistency of universal quantile learner when chained with almost any imputation function (Z., Dieuleveut, Josse and Romano, 2023)
- (Non-)Parametric modelizations of the missing covariates' influence on predictive uncertainty (Z., Josse, Romano and Dieuleveut, 2024)
- Other theoretical guarantees on `CP-MDA-Nested`$^\star$ (Z., Dieuleveut, Josse and Romano, 2023; Z., Josse, Romano and Dieuleveut, 2024)
- Critical care medical data experiments (Z., Dieuleveut, Josse and Romano, 2023)

**A natural open direction:** is it possible to achieve MCV under MAR and $Y \perp\!\!\!\perp M \mid X$ assumptions?

# References i


Barber, R. F., Candès, E. J., Ramdas, A., and Tibshirani, R. J. (2021). The limits of distribution-free conditional predictive inference. *Information and Inference: A Journal of the IMA*, 10(2).

Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021). Distributional conformal prediction. *Proceedings of the National Academy of Sciences*, 118(48).

Gibbs, I., Cherian, J. J., and Candès, E. J. (2023). Conformal prediction with conditional guarantees. arXiv: 2305.12616.

Guan, L. (2022). Localized conformal prediction: a generalized inference framework for conformal prediction. *Biometrika*, 110(1).

Izbicki, R., Shimizu, G., and Stern, R. B. (2022). CD-split and HPD-split: Efficient conformal regions in high dimensions. *Journal of Machine Learning Research*, 23(87).

Jung, C., Noarov, G., Ramalingam, R., and Roth, A. (2023). Batch multivalid conformal prediction. In *International Conference on Learning Representations*.

Kivaranovic, D., Johnson, K. D., and Leeb, H. (2020). Adaptive, Distribution-Free Prediction Intervals for Deep Networks. In *International Conference on Artificial Intelligence and Statistics*. PMLR.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*.

Lei, J. and Wasserman, L. (2014). Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1).

Papadopoulos, H., Proedrou, K., Vovk, V., and Gammerman, A. (2002). Inductive Confidence Machines for Regression. In *Machine Learning: ECML*. Springer.

Romano, Y., Barber, R. F., Sabatti, C., and Candès, E. (2020). With Malice Toward None: Assessing Uncertainty via Equalized Coverage. *Harvard Data Science Review*, 2(2).

Romano, Y., Patterson, E., and Candès, E. (2019). Conformalized Quantile Regression. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3).

Sesia, M. and Romano, Y. (2021). Conformal prediction using conditional histograms. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Vovk, V. (2012). Conditional Validity of Inductive Conformal Predictors. In *Asian Conference on Machine Learning*. PMLR.

Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer US.