

Distribution-Free Uncertainty Quantification

A short introduction to Conformal Prediction

Margaux Zaffran

Useful resources on Conformal Prediction (non exhaustive)

- Book reference: Vovk et al. (2005) (*new edition in 2022*)
- A gentle tutorial:
 - Angelopoulos and Bates (2023)
 - [Videos playlist](#)
- Another tutorial: Fontana et al. (2023)
- Ryan Tibshirani [introductory lecture's notes](#)
- [GitHub repository](#) with plenty of links: Manokhin (2022)

All material is freely accessible on [this webpage](#), including sources.

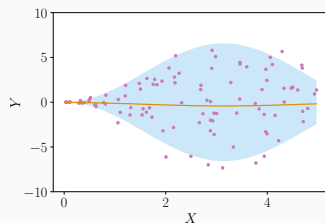
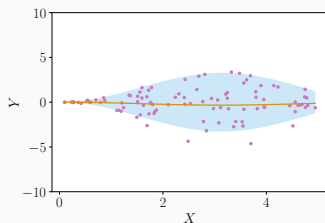
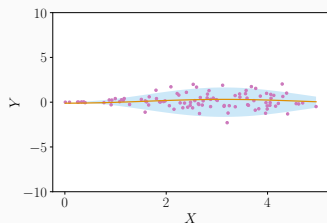
Feel free to reuse it for presentations or teaching! If you do, please credit this version, and include a link to this webpage.



This tutorial by Margaux Zaffran (2023) is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

<https://conformalpredictionintro.github.io>

On the importance of quantifying uncertainty



↪ Same **predictions**, yet 3 distinct underlying phenomena!

⇒ **Quantifying uncertainty** conveys this information.

Quantile Regression

Split Conformal Prediction (SCP)

Avoiding data splitting: full conformal and out-of-bags approaches

Beyond exchangeability

Reminder about quantiles

- Quantile level $\beta \in [0, 1]$
- $Q_X(\beta) := \inf\{x \in \mathbb{R}, \mathbb{P}(X \leq x) \geq \beta\}$
 $:= \inf\{x \in \mathbb{R}, F_X(x) \geq \beta\}$
- Empirical quantile $q_\beta(X_1, \dots, X_n)$
 $:= \lceil \beta \times n \rceil$ smallest value of (X_1, \dots, X_n)

Example of quantile: the median

$$\beta = 0.5$$

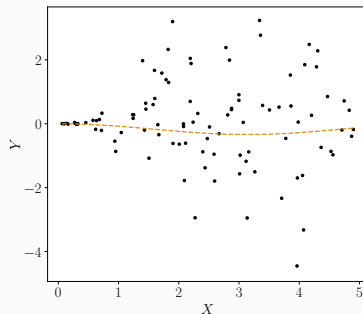
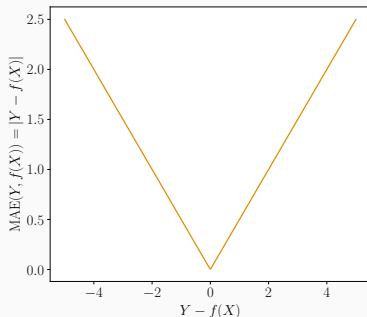
$\hookrightarrow q_{0.5}(X_1, \dots, X_n)$ is the empirical median of (X_1, \dots, X_n) ;

$\hookrightarrow Q_X(0.5)$ represents the median of the distribution of X .

Similarly, let $q_{\beta, \inf}(X_1, \dots, X_n) := \lfloor \beta \times n \rfloor$ smallest value of (X_1, \dots, X_n)

Median regression

- The Bayes predictor depends on the chosen **loss function**.
 \hookrightarrow **Bayes predictor** $f^* \in \operatorname{argmin}_f \operatorname{Risk}_\ell(f)$
$$:= \operatorname{argmin}_f \mathbb{E} [\ell(Y, f(X))]$$
- Mean Absolute Error (MAE):** $\ell(Y, Y') = |Y - Y'|$ Associated risk:
 $\operatorname{Risk}_\ell(f) = \mathbb{E} [|Y - f(X)|]$
$$\Rightarrow f^*(X) = \operatorname{median} [Y|X] = Q_{Y|X}(0.5)$$



Generalization: Quantile regression

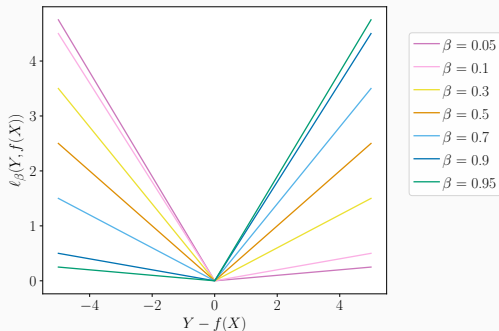
- Quantile level $\beta \in [0, 1]$
- Pinball loss

$$\ell_{\beta}(Y, Y') = \beta |Y - Y'| \mathbf{1}_{\{|Y - Y'| \geq 0\}} + (1 - \beta) |Y - Y'| \mathbf{1}_{\{|Y - Y'| \leq 0\}}$$

Associated risk: $\text{Risk}_{\ell_{\beta}}(f) = \mathbb{E}[\ell_{\beta}(Y, f(X))]$

Bayes predictor: $f^* \in \underset{f}{\operatorname{argmin}} \text{Risk}_{\ell_{\beta}}(f)$

$$\Rightarrow f^*(X) = Q_{Y|X}(\beta)$$



- Link between the **pinball loss** and the **quantiles**?

Set $q^* \in \arg \min_q \mathbb{E}[\ell_\beta(Y - q)]$. Then,

$$\begin{aligned} 0 &= \int_{-\infty}^{+\infty} \ell'_\beta(y - q^*) df_Y(y) \\ &= (\beta - 1) \int_{-\infty}^{q^*} df_Y(y) + \beta \int_{q^*}^{+\infty} df_Y(y) \end{aligned}$$

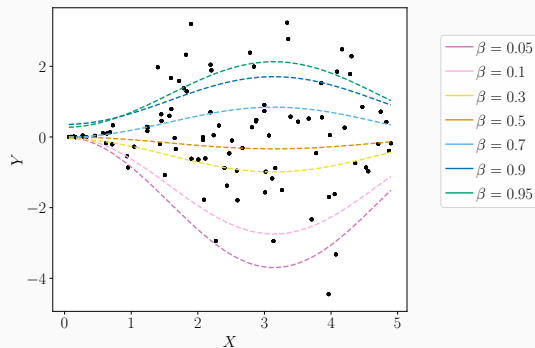
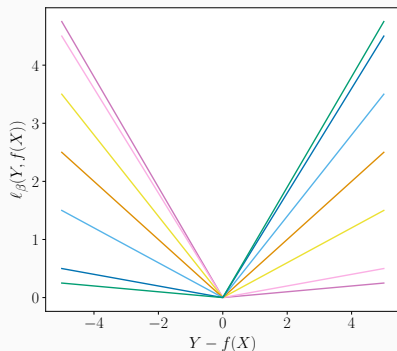
$$0 = (\beta - 1)F_Y(q^*) + \beta(1 - F_Y(q^*))$$

$$(1 - \beta)F_Y(q^*) = \beta(1 - F_Y(q^*))$$

$$\beta = F_Y(q^*)$$

$$\Leftrightarrow q^* = F_Y^{-1}(\beta)$$

Quantile regression: visualisation



Warning

No theoretical guarantee with a finite sample!

$$\mathbb{P} \left(Y \in \left[\hat{Q}_{Y|X}(\beta/2); \hat{Q}_{Y|X}(1 - \beta/2) \right] \right) \neq 1 - \beta$$

Quantile Regression

Split Conformal Prediction (SCP)

- Standard regression case

- Conformalized Quantile Regression (CQR)

- Generalization of SCP: going beyond regression

Avoiding data splitting: full conformal and out-of-bags approaches

Beyond exchangeability

Quantifying predictive uncertainty

- $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ random variables
- n training samples $(X_i, Y_i)_{i=1}^n$
- **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}_α such that:

$$\mathbb{P}\{Y_{n+1} \in \mathcal{C}_\alpha(X_{n+1})\} \geq 1 - \alpha, \quad (1)$$

and \mathcal{C}_α 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

Quantile Regression

Split Conformal Prediction (SCP)

Standard regression case

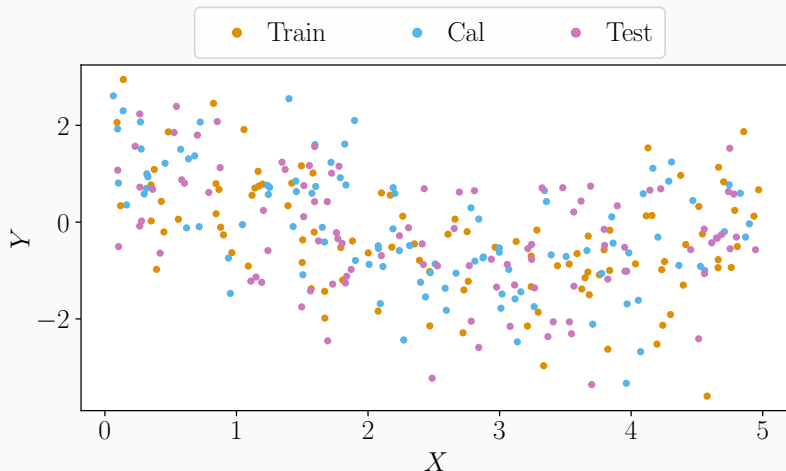
Conformalized Quantile Regression (CQR)

Generalization of SCP: going beyond regression

Avoiding data splitting: full conformal and out-of-bags approaches

Beyond exchangeability

Split Conformal Prediction (SCP)^{1,2,3}: toy example

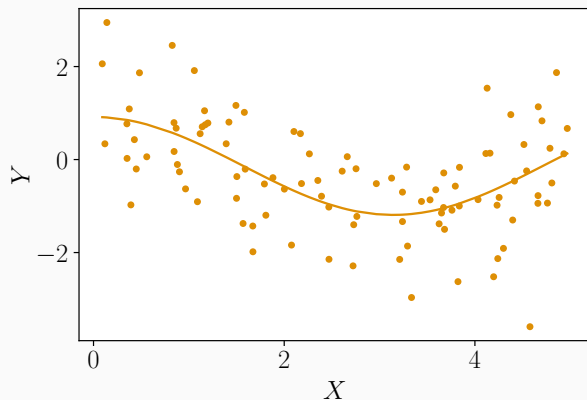


¹Vovk et al. (2005), *Algorithmic Learning in a Random World*

²Papadopoulos et al. (2002), *Inductive Confidence Machines for Regression*, ECML

³Lei et al. (2018), *Distribution-Free Predictive Inference for Regression*, JRSS B

Split Conformal Prediction (SCP)^{1,2,3}: training step



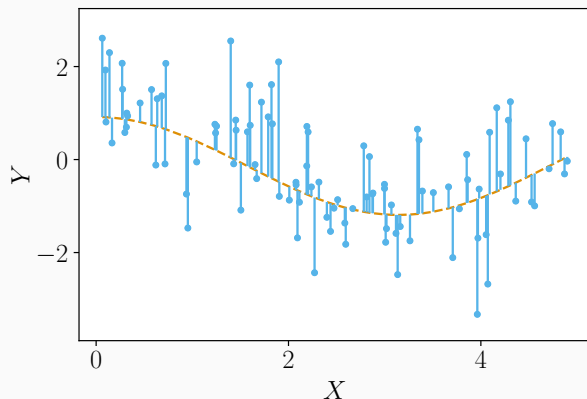
► Learn (or get) $\hat{\mu}$

¹Vovk et al. (2005), *Algorithmic Learning in a Random World*

²Papadopoulos et al. (2002), *Inductive Confidence Machines for Regression*, ECML

³Lei et al. (2018), *Distribution-Free Predictive Inference for Regression*, JRSS B

Split Conformal Prediction (SCP)^{1,2,3}: calibration step



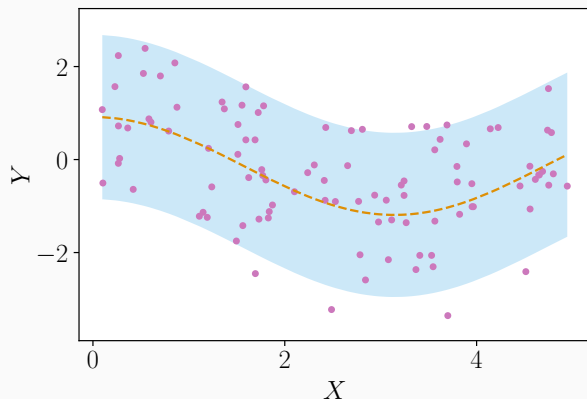
- Predict with $\hat{\mu}$
- Get the `|residuals|`, a.k.a. conformity scores
- Compute the $(1 - \alpha)$ empirical quantile of $\mathcal{S} = \{|residuals|\}_{\text{Cal}} \cup \{+\infty\}$, noted $q_{1-\alpha}(\mathcal{S})$

¹Vovk et al. (2005), *Algorithmic Learning in a Random World*

²Papadopoulos et al. (2002), *Inductive Confidence Machines for Regression*, ECML

³Lei et al. (2018), *Distribution-Free Predictive Inference for Regression*, JRSS B

Split Conformal Prediction (SCP)^{1,2,3}: prediction step



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

¹Vovk et al. (2005), *Algorithmic Learning in a Random World*

²Papadopoulos et al. (2002), *Inductive Confidence Machines for Regression*, ECML

³Lei et al. (2018), *Distribution-Free Predictive Inference for Regression*, JRSS B

SCP: implementation details



1. Randomly split the training data into a **proper training set** (size $\#Tr$) and a **calibration set** (size $\#Cal$)
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 $\#Cal + 1$ **conformity scores**:

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

(+ worst-case scenario)

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

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

SCP: implementation details



1. Randomly split the training data into a **proper training set** (size $\#Tr$) and a **calibration set** (size $\#Cal$)
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 $\#Cal$ **conformity scores**:

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

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

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

Definition (Exchangeability)

$(X_i, Y_i)_{i=1}^n$ are **exchangeable** if, for any permutation σ of $\llbracket 1, n \rrbracket$:

$$\mathcal{L}((X_1, Y_1), \dots, (X_n, Y_n)) = \mathcal{L}((X_{\sigma(1)}, Y_{\sigma(1)}), \dots, (X_{\sigma(n)}, Y_{\sigma(n)})),$$

where \mathcal{L} designates the joint distribution.

Examples of exchangeable sequences

- i.i.d. samples

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

SCP enjoys finite sample guarantees proved in Vovk et al. (2005); Lei et al. (2018).

Theorem

Suppose $(X_i, Y_i)_{i=1}^{n+1}$ are *exchangeable*⁴. SCP applied on $(X_i, Y_i)_{i=1}^n$ outputs $\hat{C}_\alpha(\cdot)$ such that:

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

Additionally, if the scores $\{S_i\}_{i \in \text{Cal}} \cup \{S_{n+1}\}$ are a.s. distinct:

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

⁴Only the calibration and test data need to be exchangeable.

Lemma (Quantile lemma)

If $(U_1, \dots, U_n, U_{n+1})$ are *exchangeable*, then for any $\beta \in]0, 1[$:

$$\mathbb{P}(U_{n+1} \leq q_\beta(U_1, \dots, U_n, +\infty)) \geq \beta.$$

Additionally, if U_1, \dots, U_n, U_{n+1} are almost surely distinct, then:

$$\mathbb{P}(U_{n+1} \leq q_\beta(U_1, \dots, U_n, +\infty)) \leq \beta + \frac{1}{n+1}.$$

When $(X_i, Y_i)_{i=1}^{n+1}$ are exchangeable, the scores $\{S_i\}_{i \in \text{Cal}} \cup \{S_{n+1}\}$ are exchangeable.

\hookrightarrow applying the quantile lemma to the scores concludes the proof.

Proof of the quantile lemma

First note that $U_{n+1} \leq q_\beta(U_1, \dots, U_n, +\infty) \iff U_{n+1} \leq q_\beta(U_1, \dots, U_n, U_{n+1})$.

Then, by definition of q_β :

$$U_{n+1} \leq q_\beta(U_1, \dots, U_n, U_{n+1}) \iff \text{rank}(U_{n+1}) \leq \lceil \beta(n+1) \rceil$$

By **exchangeability**, $\text{rank}(U_{n+1}) \sim \mathcal{U}\{1, \dots, n+1\}$. Thus:

$$\mathbb{P}(\text{rank}(U_{n+1}) \leq \lceil \beta(n+1) \rceil) \geq \frac{\lceil \beta(n+1) \rceil}{n+1} \geq \beta.$$

If U_1, \dots, U_n, U_{n+1} are **almost surely distinct (without ties)**:

$$\begin{aligned} \mathbb{P}(\text{rank}(U_{n+1}) \leq \lceil \beta(n+1) \rceil) &= \frac{\lceil \beta(n+1) \rceil}{n+1} \\ &\leq \frac{1 + \beta(n+1)}{n+1} = \beta + \frac{1}{n+1}. \end{aligned}$$

□

SCP enjoys finite sample guarantees proved in Vovk et al. (2005); Lei et al. (2018).

Theorem

Suppose $(X_i, Y_i)_{i=1}^{n+1}$ are *exchangeable*⁴. SCP applied on $(X_i, Y_i)_{i=1}^n$ outputs $\hat{C}_\alpha(\cdot)$ such that:

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

Additionally, if the scores $\{S_i\}_{i \in \text{Cal}} \cup \{S_{n+1}\}$ are a.s. distinct:

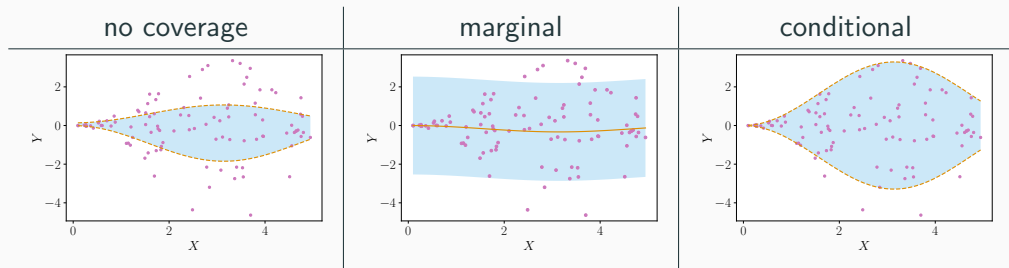
$$\mathbb{P} \left\{ Y_{n+1} \in \hat{C}_\alpha(X_{n+1}) \right\} \leq 1 - \alpha + \frac{1}{\#\text{Cal} + 1}.$$

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

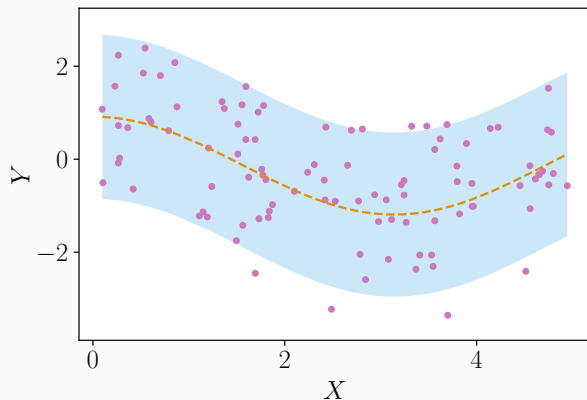
⁴Only the calibration and test data need to be exchangeable.

Conditional coverage implies adaptiveness

- **Marginal** coverage: $\mathbb{P} \left\{ Y_{n+1} \in \hat{C}_\alpha(X_{n+1}) \right\}$ the errors may differ across regions of the input space (i.e. non-adaptive)
- **Conditional** coverage: $\mathbb{P} \left\{ Y_{n+1} \in \hat{C}_\alpha(X_{n+1}) | X_{n+1} \right\}$ errors are evenly distributed (i.e. fully adaptive)
- Conditional coverage is **stronger** than marginal coverage



Standard mean-regression SCP is not adaptive



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

Informative conditional coverage as such is impossible

- Impossibility results

↪ Lei and Wasserman (2014); Vovk (2012); Barber et al. (2021a)

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

- Approximate conditional coverage

↪ Romano et al. (2020a); Guan (2022); Jung et al. (2023); Gibbs et al. (2023)

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

- Asymptotic (with the sample size) conditional coverage

↪ Romano et al. (2019); Kivaranovic et al. (2020); Chernozhukov et al. (2021); Sesia and Romano (2021); Izbicki et al. (2022)

Quantile Regression

Split Conformal Prediction (SCP)

Standard regression case

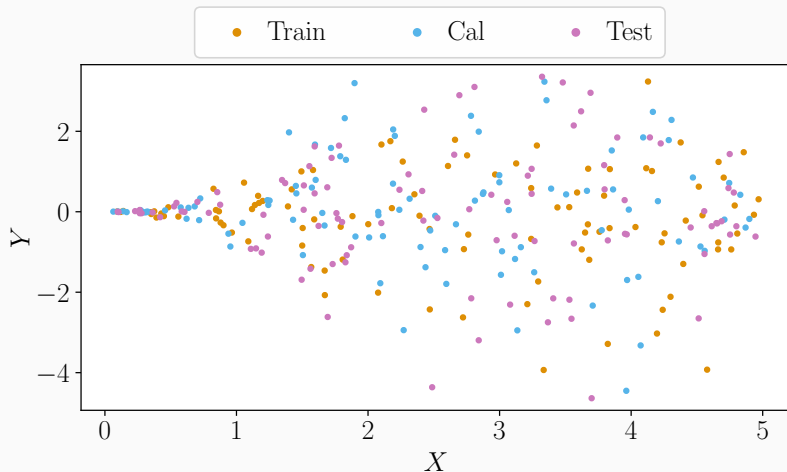
Conformalized Quantile Regression (CQR)

Generalization of SCP: going beyond regression

Avoiding data splitting: full conformal and out-of-bags approaches

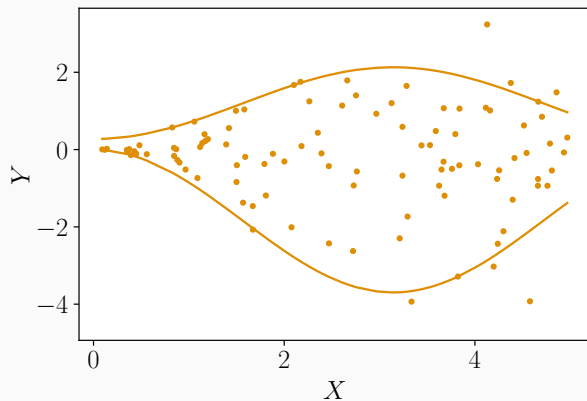
Beyond exchangeability

Conformalized Quantile Regression (CQR)⁵



⁵Romano et al. (2019), *Conformalized Quantile Regression*, NeurIPS

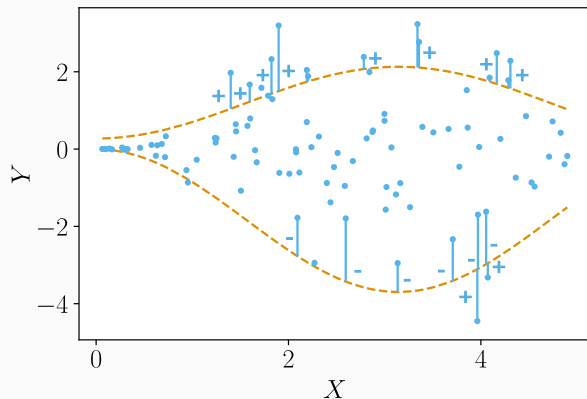
Conformalized Quantile Regression (CQR)⁵: training step



► Learn (or get) $\widehat{QR}_{\text{lower}}$ and $\widehat{QR}_{\text{upper}}$

⁵Romano et al. (2019), *Conformalized Quantile Regression*, NeurIPS

Conformalized Quantile Regression (CQR)⁵: calibration step

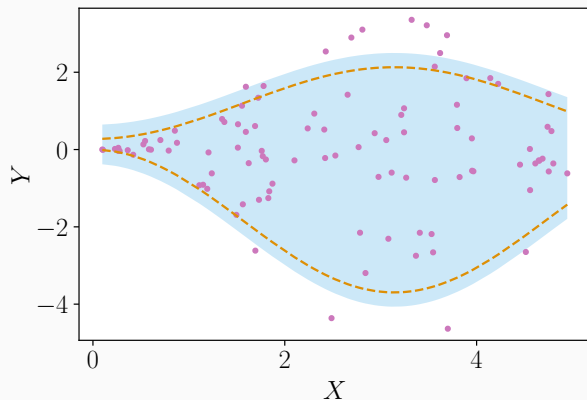


- ▶ Predict with \widehat{QR}_{lower} and \widehat{QR}_{upper}
- ▶ Get the scores
 $\mathcal{S} = \{S_i\}_{\text{Cal}} \cup \{+\infty\}$
- ▶ Compute the $(1 - \alpha)$ empirical quantile of \mathcal{S} , noted $q_{1-\alpha}(\mathcal{S})$

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

⁵Romano et al. (2019), *Conformalized Quantile Regression*, NeurIPS

Conformalized Quantile Regression (CQR)⁵: prediction step



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

► Build

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

⁵Romano et al. (2019), *Conformalized Quantile Regression*, NeurIPS



1. Randomly split the training data into a **proper training set** (size $\#Tr$) and a **calibration set** (size $\#Cal$)
 2. Get \widehat{QR}_{lower} and \widehat{QR}_{upper} by training the algorithm \mathcal{A} on the **proper training set**
 3. Obtain a set of $\#Cal + 1$ conformity scores \mathcal{S} :
- $\mathcal{S} = \{S_i = \max(\widehat{QR}_{lower}(X_i) - Y_i, Y_i - \widehat{QR}_{upper}(X_i)), i \in \text{Cal}\} \cup \{+\infty\}$
4. Compute the $1 - \alpha$ quantile of these scores, noted $q_{1-\alpha}(\mathcal{S})$
 5. For a new point X_{n+1} , return

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

1. Randomly split the training data into a **proper training set** (size $\#Tr$) and a **calibration set** (size $\#Cal$)
2. Get \widehat{QR}_{lower} and \widehat{QR}_{upper} by training the algorithm \mathcal{A} on the **proper training set**

3. Obtain a set of $\#Cal$ **conformity scores** \mathcal{S} :

$$\mathcal{S} = \{S_i = \max\left(\widehat{QR}_{lower}(X_i) - Y_i, Y_i - \widehat{QR}_{upper}(X_i)\right), i \in Cal\}$$

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

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

This procedure enjoys the finite sample guarantee proposed and proved in Romano et al. (2019).

Theorem

Suppose $(X_i, Y_i)_{i=1}^{n+1}$ are *exchangeable*⁶. CQR on $(X_i, Y_i)_{i=1}^n$ outputs $\hat{C}_\alpha(\cdot)$ such that:

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

If, in addition, the scores $\{S_i\}_{i \in \text{Cal}} \cup \{S_{n+1}\}$ are almost surely distinct, then

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

Proof: application of the quantile lemma.

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

⁶Only the calibration and test data need to be exchangeable.

Quantile Regression

Split Conformal Prediction (SCP)

Standard regression case

Conformalized Quantile Regression (CQR)

Generalization of SCP: going beyond regression

Avoiding data splitting: full conformal and out-of-bags approaches

Beyond exchangeability

SCP is defined by the conformity score function



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

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

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

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

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

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

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

This procedure enjoys the finite sample guarantee proposed and proved in Vovk et al. (2005).

Theorem

Suppose $(X_i, Y_i)_{i=1}^{n+1}$ are *exchangeable*⁷. SCP on $(X_i, Y_i)_{i=1}^n$ outputs $\hat{C}_\alpha(\cdot)$ such that:

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

If, in addition, the scores $\{S_i\}_{i \in \text{Cal}} \cup \{S_{n+1}\}$ are almost surely distinct, then

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

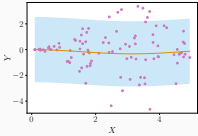
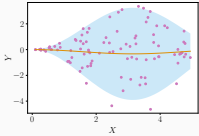
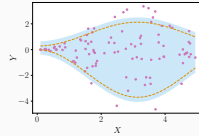
Proof: application of the quantile lemma.

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

⁷Only the calibration and test data need to be exchangeable.

SCP: what choices for the regression scores?

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

	Standard SCP Vovk et al. (2005)	Locally weighted SCP Lei et al. (2018)	CQR Romano et al. (2019)
$s(\hat{A}(X), Y)$	$ \hat{\mu}(X) - Y $	$\frac{ \hat{\mu}(X) - Y }{\hat{\rho}(X)}$	$\max(\hat{Q}R_{\text{lower}}(X) - Y, Y - \hat{Q}R_{\text{upper}}(X))$
$\hat{C}_\alpha(x)$	$[\hat{\mu}(x) \pm q_{1-\alpha}(S)]$	$[\hat{\mu}(x) \pm q_{1-\alpha}(S)\hat{\rho}(x)]$	$[\hat{Q}R_{\text{lower}}(x) - q_{1-\alpha}(S); \hat{Q}R_{\text{upper}}(x) + q_{1-\alpha}(S)]$
Visu.			
✓	black-box around a “usable” prediction	black-box around a “usable” prediction	adaptive
✗	not adaptive	limited adaptiveness	no black-box around a “usable” prediction

- $Y \in \{1, \dots, C\}$ (C classes)
- $\hat{A}(X) = (\hat{p}_1(X), \dots, \hat{p}_C(X))$ (estimated probabilities)
- $\text{s}(\hat{A}(X), Y) := 1 - (\hat{A}(X))_Y$
- For a new point X_{n+1} , return
$$\hat{C}_\alpha(X_{n+1}) = \{y \text{ such that } \text{s}(\hat{A}(X_{n+1}), y) \leq q_{1-\alpha}(\mathcal{S})\}$$

SCP: standard classification in practice

Ex: $Y_i \in \{\text{"dog"}, \text{"tiger"}, \text{"cat"}\}$, with $\alpha = 0.1$

- Scores on the calibration set

Cal_i	"dog"	"dog"	"dog"	"tiger"	"tiger"	"tiger"	"tiger"	"cat"	"cat"	"cat"
$\hat{p}_{\text{dog}}(X_i)$	0.95	0.90	0.85	0.15	0.15	0.20	0.15	0.15	0.25	0.20
$\hat{p}_{\text{tiger}}(X_i)$	0.02	0.05	0.10	0.60	0.55	0.50	0.45	0.40	0.35	0.45
$\hat{p}_{\text{cat}}(X_i)$	0.03	0.05	0.05	0.25	0.30	0.30	0.40	0.45	0.40	0.35
S_i	0.05	0.1	0.15	0.40	0.45	0.50	0.55	0.55	0.6	0.65

- $q_{1-\alpha}(\mathcal{S}) = 0.65$
- $\hat{A}(X_{n+1}) = (0.05, 0.60, 0.35)$
 - $\hookrightarrow s(\hat{A}(X_{n+1}), \text{"dog"}) = 0.95$
 - $\hookrightarrow s(\hat{A}(X_{n+1}), \text{"tiger"}) = 0.40 \leq q_{1-\alpha}(\mathcal{S})$
 - $\hookrightarrow s(\hat{A}(X_{n+1}), \text{"cat"}) = 0.65 \leq q_{1-\alpha}(\mathcal{S})$
- $\hat{C}_\alpha(X_{n+1}) = \{\text{"tiger"}, \text{"cat"}\}$

"dog" $\notin \hat{C}_\alpha(X_{n+1})$
"tiger" $\in \hat{C}_\alpha(X_{n+1})$
"cat" $\in \hat{C}_\alpha(X_{n+1})$

Ex: $Y \in \{\text{"dog"}, \text{"tiger"}, \text{"cat"}\}$, with $\alpha = 0.1$

- Scores on the calibration set

Cal _i	"dog"	"dog"	"dog"	"tiger"	"tiger"	"tiger"	"tiger"	"cat"	"cat"	"cat"
$\hat{p}_{\text{dog}}(X_i)$	0.95	0.90	0.85	0.05	0.05	0.05	0.05	0.10	0.10	0.15
$\hat{p}_{\text{tiger}}(X_i)$	0.02	0.05	0.10	0.85	0.80	0.75	0.70	0.25	0.30	0.30
$\hat{p}_{\text{cat}}(X_i)$	0.03	0.05	0.05	0.10	0.15	0.20	0.25	0.65	0.60	0.55
S_i	0.05	0.1	0.15	0.15	0.20	0.25	0.30	0.35	0.40	0.45

- $q_{1-\alpha}(\mathcal{S}) = 0.45$

- $\hat{A}(X_{n+1}) = (0.05, 0.60, 0.35)$

$$\hookrightarrow s(\hat{A}(X_{n+1}), \text{"dog"}) = 0.95$$

$$\hookrightarrow s(\hat{A}(X_{n+1}), \text{"tiger"}) = 0.40 \leq q_{1-\alpha}(\mathcal{S})$$

$$\hookrightarrow s(\hat{A}(X_{n+1}), \text{"cat"}) = 0.65$$

$$\text{"dog"} \notin \hat{C}_\alpha(X_{n+1})$$

$$\text{"tiger"} \in \hat{C}_\alpha(X_{n+1})$$

$$\text{"cat"} \notin \hat{C}_\alpha(X_{n+1})$$

- $\hat{C}_\alpha(X_{n+1}) = \{\text{"tiger"}\}$

The standard classification conformity score function leads to:

- ✓ smallest prediction sets on average

- ✗ undercovering (overcovering) hard (easy) subgroups

(similar to the standard mean regression case!)

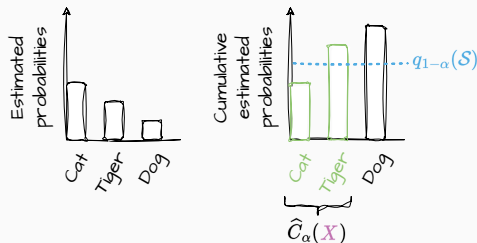
⇒ Other score functions can be built to improve adaptiveness

(as in regression with localized scores)

SCP: classification with Adaptive Prediction Sets⁸

1. Sort in decreasing order $\hat{p}_{\sigma(1)}(X) \geq \dots \geq \hat{p}_{\sigma(C)}(X)$
2. $\mathbf{s}(\hat{A}(X), Y) := \sum_{k=1}^{\sigma^{-1}(Y)} \hat{p}_{\sigma(k)}(X)$ (sum of the estimated probabilities associated to classes at least as large as that of the true class Y)
3. Return the set of classes $\{\sigma_{n+1}(1), \dots, \sigma_{n+1}(r^*)\}$, where

$$r^* = \arg \max_{1 \leq r \leq C} \left\{ \sum_{k=1}^r \hat{p}_{\sigma_{n+1}(k)}(X_{n+1}) < q_{1-\alpha}(\mathcal{S}) \right\} + 1$$



⁸Romano et al. (2020b), *Classification with Valid and Adaptive Coverage*, NeurIPS

Figure highly inspired by Angelopoulos and Bates (2023).

SCP: classification with Adaptive Prediction Sets in practice

Ex: $Y \in \{\text{"dog"}, \text{"tiger"}, \text{"cat"}\}$, with $\alpha = 0.1$

- Scores on the calibration set

Cal _i	"dog"	"dog"	"dog"	"tiger"	"tiger"	"tiger"	"tiger"	"cat"	"cat"	"cat"
$\hat{p}_{\text{dog}}(X_i)$	0.95	0.90	0.85	0.05	0.05	0.05	0.10	0.25	0.10	0.15
$\hat{p}_{\text{tiger}}(X_i)$	0.02	0.05	0.10	0.85	0.80	0.75	0.75	0.40	0.30	0.30
$\hat{p}_{\text{cat}}(X_i)$	0.03	0.05	0.05	0.10	0.15	0.20	0.15	0.35	0.60	0.55
S_i	0.95	0.90	0.85	0.85	0.80	0.75	0.75	0.75	0.60	0.55

- $q_{1-\alpha}(\mathcal{S}) = 0.95$

\hookrightarrow Ex 1: $\hat{A}(X_{n+1}) = (0.05, 0.45, 0.5), r^* = 2$

$$\hat{C}_\alpha(X_{n+1}) = \{\text{"tiger"}, \text{"cat"}\}$$

\hookrightarrow Ex 2: $\hat{A}(X_{n+1}) = (0.03, 0.95, 0.02), r^* = 1$

$$\hat{C}_\alpha(X_{n+1}) = \{\text{"tiger"}\}$$

- **Simple** procedure which quantifies the uncertainty of **any** predictive model \hat{A} by returning predictive regions
- **Finite-sample** guarantees
- **Distribution-free** as long as the data are **exchangeable** (and so are the scores)
- **Marginal** theoretical guarantee over the joint (X, Y) distribution, and **not conditional**, i.e., no guarantee that for any x :

$$\mathbb{P} \left\{ Y_{n+1} \in \hat{C}_\alpha(X_{n+1}) \mid X_{n+1} = x \right\} \geq 1 - \alpha.$$

↪ marginal also over the whole calibration set and the test point!

Challenges: open questions (non exhaustive!)

- Conditional coverage
- Computational cost vs statistical power
- Exchangeability

(~ Previous Section)

(Next Section)

(Last Section)

Quantile Regression

Split Conformal Prediction (SCP)

Avoiding data splitting: full conformal and out-of-bags approaches

Full Conformal Prediction

Jackknife+

Beyond exchangeability

Quantile Regression

Split Conformal Prediction (SCP)

Avoiding data splitting: full conformal and out-of-bags approaches

Full Conformal Prediction

Jackknife+

Beyond exchangeability

SCP suffers from data splitting:

- lower statistical efficiency (lower model accuracy and higher predictive set size)
- higher statistical variability

Can we avoid splitting the data set?

The naive idea does not enjoy valid coverage (even empirically)

- A naive idea:
 - Get \hat{A} by training the algorithm \mathcal{A} on $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$.
 - compute the empirical quantile $q_{1-\alpha}(\mathcal{S})$ of the set of scores

$$\mathcal{S} = \left\{ \text{s} \left(\hat{A}(X_i), Y_i \right) \right\}_{i=1}^n \cup \{\infty\}.$$

- output the set $\left\{ y \text{ such that } \text{s} \left(\hat{A}(X_{n+1}), y \right) \leq q_{1-\alpha}(\mathcal{S}) \right\}.$

✗ \hat{A} has been obtained using the training set $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ but did not use X_{n+1} .

$\Rightarrow \text{s} \left(\hat{A}(X_{n+1}), y \right)$ stochastically dominates any element of $\left\{ \text{s} \left(\hat{A}(X_i), Y_i \right) \right\}_{i=1}^n$.

Full Conformal Prediction⁹ does not discard training points!

- Full (or transductive) Conformal Prediction
 - avoids data splitting
 - at the cost of many more model fits
- **Idea:** the most probable labels Y_{n+1} live in \mathcal{Y} , and have a low enough conformity score. By looping over all possible $y \in \mathcal{Y}$, the ones leading to the smallest conformity scores will be found.

⁹Vovk et al. (2005), *Algorithmic Learning in a Random World*

Full Conformal Prediction (CP): recovering exchangeability

For any candidate (X_{n+1}, y) :

1. Get \hat{A}_y by training \mathcal{A} on $\{(X_1, Y_1), \dots, (X_n, Y_n)\} \cup \{(X_{n+1}, y)\}$
2. Obtain a set of training scores

$$\mathcal{S}_y^{(\text{train})} = \left\{ s(\hat{A}_y(X_i), Y_i) \right\}_{i=1}^n \cup \{s(\hat{A}_y(X_{n+1}), y)\}$$

and compute their $1 - \alpha$ empirical quantile $q_{1-\alpha}(\mathcal{S}_y^{(\text{train})})$

Output the set $\left\{ y \text{ such that } s(\hat{A}_y(X_{n+1}), y) \leq q_{1-\alpha}(\mathcal{S}_y^{(\text{train})}) \right\}$.

- ✓ Test point treated in the same way than train points
- ✗ Computationally costly

Definition (Symmetrical algorithm)

A deterministic algorithm $\mathcal{A} : (U_1, \dots, U_n) \mapsto \hat{A}$ is **symmetric** if for any permutation σ of $\llbracket 1, n \rrbracket$:

$$\mathcal{A}(U_1, \dots, U_n) \stackrel{\text{a.s.}}{=} \mathcal{A}(U_{\sigma(1)}, \dots, U_{\sigma(n)}) .$$

Full CP enjoys finite sample guarantees proved in Vovk et al. (2005).

Theorem

Suppose that

- (i) $(X_i, Y_i)_{i=1}^{n+1}$ are *exchangeable*,
- (ii) the algorithm \mathcal{A} is *symmetric*.

Full CP applied on $(X_i, Y_i)_{i=1}^n \cup \{X_{n+1}\}$ outputs $\hat{C}_\alpha(\cdot)$ such that:

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

Additionally, if the scores are a.s. distinct:

$$\mathbb{P} \left\{ Y_{n+1} \in \hat{C}_\alpha(X_{n+1}) \right\} \leq 1 - \alpha + \frac{1}{n+1}.$$

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

FCP sets with an interpolating algorithm

Assume \mathcal{A} interpolates:

- $\hat{A} = \mathcal{A}((x_1, y_1), \dots, (x_{n+1}, y_{n+1}))$
- $\hat{A}(x_k) - y_k = 0$ for any $k \in \llbracket 1, n+1 \rrbracket$

\Rightarrow Full Conformal Prediction (*with standard score functions*) outputs \mathcal{Y} (the whole label space) for any new test point!

Quantile Regression

Split Conformal Prediction (SCP)

Avoiding data splitting: full conformal and out-of-bags approaches

Full Conformal Prediction

Jackknife+

Beyond exchangeability

Jackknife: the naive idea does not enjoy valid coverage

- Based on leave-one-out (LOO) residuals



- $\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ training data
- Get \hat{A}_{-i} by training \mathcal{A} on $\mathcal{D}_n \setminus (X_i, Y_i)$
- LOO scores $\mathcal{S} = \left\{ |\hat{A}_{-i}(X_i) - Y_i| \right\}_i \cup \{+\infty\}$ (in standard mean regression)
- Get \hat{A} by training \mathcal{A} on \mathcal{D}_n
- Build the predictive interval: $\left[\hat{A}(X_{n+1}) \pm q_{1-\alpha}(\mathcal{S}) \right]$

Warning

No guarantee on the prediction of \hat{A} with scores based on $(\hat{A}_{-i})_i$, without assuming a form of **stability** on \mathcal{A} .

- Based on **leave-one-out (LOO) residuals**



- $\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ training data

- Get \hat{A}_{-i} by training \mathcal{A} on $\mathcal{D}_n \setminus (X_i, Y_i)$

- LOO predictions / predictive intervals**

$$\mathcal{S}_{\text{up/down}} = \left\{ \hat{A}_{-i}(X_{n+1}) \pm |\hat{A}_{-i}(X_i) - Y_i| \right\}_i \cup \{\pm\infty\}$$

(in standard mean regression)

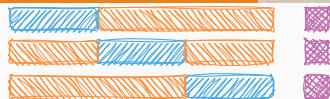
- Build the predictive interval: $[q_{\alpha, \text{inf}}(\mathcal{S}_{\text{down}}); q_{1-\alpha}(\mathcal{S}_{\text{up}})]$

Theorem

If $\mathcal{D}_n \cup (X_{n+1}, Y_{n+1})$ are exchangeable and \mathcal{A} is symmetric: $\mathbb{P}(Y_{n+1} \in \hat{\mathcal{C}}_{\alpha}(X_{n+1})) \geq 1 - 2\alpha$.

¹⁰Barber et al. (2021b), *Predictive Inference with the jackknife+*, The Annals of Statistics

Recall $q_{\beta, \text{inf}}(X_1, \dots, X_n) := \lfloor \beta \times n \rfloor$ smallest value of (X_1, \dots, X_n)



- Based on **cross-validation residuals**
- $\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ training data
- Split \mathcal{D}_n into K folds F_1, \dots, F_K
- Get \hat{A}_{-F_k} by training \mathcal{A} on $\mathcal{D}_n \setminus F_k$
- **Cross-val predictions / predictive intervals**

$$\mathcal{S}_{\text{up/down}} = \left\{ \left\{ \hat{A}_{-F_k}(X_{n+1}) \pm |\hat{A}_{-F_k}(X_i) - Y_i| \right\}_{i \in F_k} \right\}_k \cup \{\pm\infty\}$$

(in standard mean regression)

- Build the predictive interval: $[q_{\alpha, \text{inf}}(\mathcal{S}_{\text{down}}); q_{1-\alpha}(\mathcal{S}_{\text{up}})]$

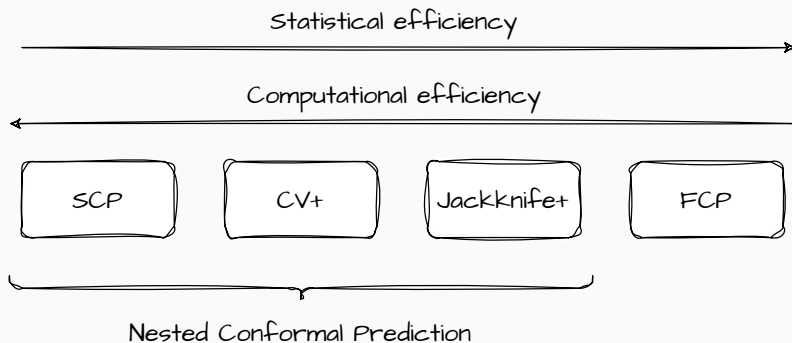
Theorem

If $\mathcal{D}_n \cup (X_{n+1}, Y_{n+1})$ are exchangeable and \mathcal{A} is symmetric:

$$\mathbb{P}(Y_{n+1} \in \hat{C}_\alpha(X_{n+1})) \geq 1 - 2\alpha - \min \left(\frac{2(1 - 1/K)}{n/K + 1}, \frac{1 - K/n}{K + 1} \right) \geq 1 - 2\alpha - \sqrt{2/n}.$$

¹¹Barber et al. (2021b), *Predictive Inference with the jackknife+*, The Annals of Statistics

Recall $q_{\beta, \text{inf}}(X_1, \dots, X_n) := \lfloor \beta \times n \rfloor$ smallest value of (X_1, \dots, X_n)



- Generalized framework encapsulating out-of-sample methods: Nested CP (Gupta et al., 2022)
- Accelerating FCP: Nouretdinov et al. (2001); Lei (2019); Ndiaye and Takeuchi (2019); Cherubin et al. (2021); Ndiaye and Takeuchi (2022); Ndiaye (2022)

Quantile Regression

Split Conformal Prediction (SCP)

Avoiding data splitting: full conformal and out-of-bags approaches

Beyond exchangeability

Exchangeability does not hold in many practical applications

- CP requires **exchangeable** data points to ensure validity
- ✗ Covariate shift, i.e. \mathcal{L}_X changes but $\mathcal{L}_{Y|X}$ stays constant
- ✗ Label shift, i.e. \mathcal{L}_Y changes but $\mathcal{L}_{X|Y}$ stays constant
- ✗ Arbitrary distribution shift
- ✗ Possibly many shifts, not only one

- **Setting:**
 - $(X_1, Y_1), \dots, (X_n, Y_n) \stackrel{i.i.d.}{\sim} P_X \times P_{Y|X}$
 - $(X_{n+1}, Y_{n+1}) \sim \tilde{P}_X \times P_{Y|X}$
- **Idea:** give more importance to calibration points that are closer in distribution to the test point
- **In practice:**

1. estimate the **likelihood ratio** $w(X_i) = \frac{d\tilde{P}_X(X_i)}{dP_X(X_i)}$
2. normalize the weights, i.e. $\omega_i = \omega(X_i) = \frac{w(X_i)}{\sum_{j=1}^{n+1} w(X_j)}$
3. outputs $\hat{C}_\alpha(X_{n+1}) = \left\{ y : \mathbf{s}(\hat{A}(X_{n+1}), y) \leq q_{1-\alpha}(\{\omega_i S_i\}_{i \in \text{Cal}} \cup \{+\infty\}) \right\}$

¹²Tibshirani et al. (2019), *Conformal Prediction Under Covariate Shift*, NeurIPS

- **Setting:**
 - $(X_1, Y_1), \dots, (X_n, Y_n) \stackrel{i.i.d.}{\sim} P_{X|Y} \times P_Y$
 - $(X_{n+1}, Y_{n+1}) \sim P_{X|Y} \times \tilde{P}_Y$
 - **Classification**
- **Idea:** give more importance to calibration points that are closer in distribution to the test point
- **Trouble:** the actual test labels are **unknown**
- **In practice:**

1. estimate the **likelihood ratio** $w(Y_i) = \frac{d\tilde{P}_Y(Y_i)}{dP_Y(Y_i)}$ using algorithms from the existing label shift literature

2. normalize the weights, i.e. $\omega_i^y = \omega^y(X_i) = \frac{w(Y_i)}{\sum_{j=1}^n w(Y_j) + w(y)}$

3. outputs $\hat{C}_\alpha(X_{n+1}) =$

$$\left\{ y : \mathbb{s}(\hat{A}(X_{n+1}), y) \leq q_{1-\alpha}(\{\omega_i^y S_i\}_{i \in \text{Cal}} \cup \{+\infty\}) \right\}$$

¹³Podkopaev and Ramdas (2021), *Distribution-free uncertainty quantification for classification under label shift*, UAI

- Arbitrary distribution shift: Cauchois et al. (2020) leverages ideas from the distributionally robust optimization literature
- Two major **general theoretical results** beyond exchangeability:
 - Chernozhukov et al. (2018)
 - ↪ If the learnt model is accurate and the data noise is strongly mixing, then CP is valid asymptotically ✓
 - Barber et al. (2022)
 - ↪ Quantifies the coverage loss depending on the strength of exchangeability violation
 - $$\mathbb{P}(Y_{n+1} \in \hat{C}_\alpha(X_{n+1})) \geq 1 - \alpha - \text{average violation of exchangeability by each calibration point}$$
 - ↪ proposed algorithm: **reweighting** again!
 - e.g., in a temporal setting, give higher weights to more recent points.

- **Data:** T_0 random variables $(X_1, Y_1), \dots, (X_{T_0}, Y_{T_0})$ in $\mathbb{R}^d \times \mathbb{R}$
- **Aim:** predict the response values as well as predictive intervals for T_1 subsequent observations $X_{T_0+1}, \dots, X_{T_0+T_1}$ sequentially: at any prediction step $t \in \llbracket T_0 + 1, T_0 + T_1 \rrbracket$, $Y_{t-T_0}, \dots, Y_{t-1}$ have been revealed
- Build the smallest interval \hat{C}_α^t such that:

$$\mathbb{P} \left\{ Y_t \in \hat{C}_\alpha^t(X_t) \right\} \geq 1 - \alpha, \text{ for } t \in \llbracket T_0 + 1, T_0 + T_1 \rrbracket,$$

often simplified in:

$$\frac{1}{T_1} \sum_{t=T_0+1}^{T_0+T_1} \mathbb{1} \left\{ Y_t \in \hat{C}_\alpha^t(X_t) \right\} \approx 1 - \alpha.$$

Recent developments

- Consider splitting strategies that respect the temporal structure
- Gibbs and Candès (2021) propose a method which reacts faster to temporal evolution
 - **Idea**: track the previous coverages of the predictive intervals ($\mathbb{1}\{Y_t \in \hat{C}_\alpha(X_t)\}$)
 - **Tool**: update the empirical quantile level with a learning rate γ
 - Asymptotic guarantee (on average) for **any distribution** (even adversarial)
- Zaffran et al. (2022) studies the influence of this learning rate γ and proposes, along with Gibbs and Candès (2022), a method not requiring to choose γ
- Bhatnagar et al. (2023) enjoys **anytime** regret bound, by leveraging tools from the strongly adaptive regret minimization literature
- Bastani et al. (2022) proposes an algorithm achieving stronger coverage guarantees (conditional on specified overlapping subsets, and threshold calibrated) without hold-out set
- Angelopoulos et al. (2023) combines CP ideas with control theory ones, to adaptively improve the predictive intervals depending on the errors structure

- Angelopoulos, A. N. and Bates, S. (2023). Conformal prediction: A gentle introduction. *Foundations and Trends® in Machine Learning*, 16(4).
- Angelopoulos, A. N., Candès, E. J., and Tibshirani, R. J. (2023). Conformal pid control for time series prediction. arXiv: 2307.16895.
- Barber, R. F., Candès, E. J., Ramdas, A., and Tibshirani, R. J. (2021a). The limits of distribution-free conditional predictive inference. *Information and Inference: A Journal of the IMA*, 10(2).
- Barber, R. F., Candès, E. J., Ramdas, A., and Tibshirani, R. J. (2021b). Predictive inference with the jackknife+. *The Annals of Statistics*, 49(1).
- Barber, R. F., Candès, E. J., Ramdas, A., and Tibshirani, R. J. (2022). Conformal prediction beyond exchangeability. To appear in *Annals of Statistics (2023)*.

- Bastani, O., Gupta, V., Jung, C., Noarov, G., Ramalingam, R., and Roth, A. (2022). Practical adversarial multivalid conformal prediction. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Bhatnagar, A., Wang, H., Xiong, C., and Bai, Y. (2023). Improved online conformal prediction via strongly adaptive online learning. In *Proceedings of the 40th International Conference on Machine Learning*. PMLR.
- Cauchois, M., Gupta, S., Ali, A., and Duchi, J. C. (2020). Robust Validation: Confident Predictions Even When Distributions Shift. arXiv: 2008.04267.
- Chernozhukov, V., Wüthrich, K., and Yinchu, Z. (2018). Exact and Robust Conformal Inference Methods for Predictive Machine Learning with Dependent Data. In *Conference On Learning Theory*. PMLR.
- Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021). Distributional conformal prediction. *Proceedings of the National Academy of Sciences*, 118(48).

- Cherubin, G., Chatzikokolakis, K., and Jaggi, M. (2021). Exact optimization of conformal predictors via incremental and decremental learning. In *Proceedings of the 38th International Conference on Machine Learning*. PMLR.
- Fontana, M., Zeni, G., and Vantini, S. (2023). Conformal prediction: A unified review of theory and new challenges. *Bernoulli*, 29(1).
- Gibbs, I. and Candès, E. (2021). Adaptive conformal inference under distribution shift. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Gibbs, I. and Candès, E. (2022). Conformal inference for online prediction with arbitrary distribution shifts. arXiv: 2208.08401.
- 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).
- Gupta, C., Kuchibhotla, A. K., and Ramdas, A. (2022). Nested conformal prediction and quantile out-of-bag ensemble methods. *Pattern Recognition*, 127.
- 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. (2019). Fast exact conformalization of the lasso using piecewise linear homotopy. *Biometrika*, 106(4).
- 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).
- Manokhin, V. (2022). Awesome conformal prediction.
<https://github.com/valeman/awesome-conformal-prediction>.
- Ndiaye, E. (2022). Stable conformal prediction sets. In *Proceedings of the 39th International Conference on Machine Learning*. PMLR.

- Ndiaye, E. and Takeuchi, I. (2019). Computing full conformal prediction set with approximate homotopy. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Ndiaye, E. and Takeuchi, I. (2022). Root-finding approaches for computing conformal prediction set. *Machine Learning*, 112(1).
- Nouretdinov, I., Melliush, T., and Vovk, V. (2001). Ridge regression confidence machine. In *Proceedings of the 18th International Conference on Machine Learning*.
- Papadopoulos, H., Proedrou, K., Vovk, V., and Gammerman, A. (2002). Inductive Confidence Machines for Regression. In *Machine Learning: ECML*. Springer.
- Podkopaev, A. and Ramdas, A. (2021). Distribution-free uncertainty quantification for classification under label shift. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*. PMLR.

- Romano, Y., Barber, R. F., Sabatti, C., and Candès, E. (2020a). 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.
- Romano, Y., Sesia, M., and Candès, E. (2020b). Classification with valid and adaptive coverage. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Sesia, M. and Romano, Y. (2021). Conformal prediction using conditional histograms. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

- Tibshirani, R. J., Barber, R. F., Candes, E., and Ramdas, A. (2019). Conformal Prediction Under Covariate Shift. 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. (2015). Cross-conformal predictors. *Annals of Mathematics and Artificial Intelligence*, 74(1-2).
- Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer US.
- Zaffran, M., Féron, O., Goude, Y., Josse, J., and Dieuleveut, A. (2022). Adaptive conformal predictions for time series. In *Proceedings of the 39th International Conference on Machine Learning*. PMLR.



This tutorial by Margaux Zaffran (2023) is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

<https://conformalpredictionintro.github.io>