BETA-OPTIM: UNCERTAINTY QUANTIFICATION METHOD FOR MULTI-OUTPUT REGRESSION PROBLEMS

Jaad BELHOUARI-DURAND

Supervised by Paul Malfrait and Athmane BAKHTA
2DS/CODA/SGR Paris

Collaboration with Matthieu Gallezot and Sylvain Berger
Acoustics/CODA/SGR Paris

SAINT-GOBAIN

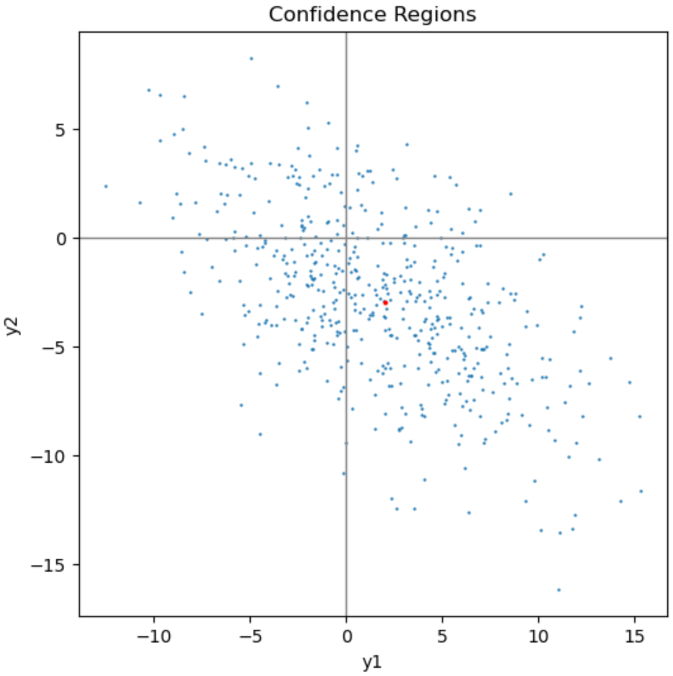
PART 1

INTRODUCTION TO THE PROBLEM

MULTI-DIMENSIONAL CONFIDENCE INTERVALS

2 dimensional case

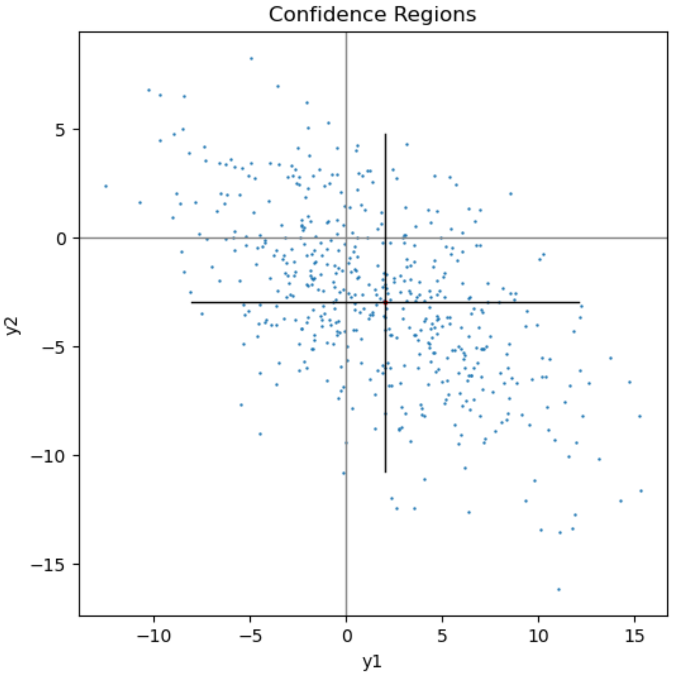
How to quantify the uncertainties in multiple dimension ?



MULTI-DIMENSIONAL CONFIDENCE INTERVALS

2 dimensional case

How to quantify the uncertainties in multiple dimension ?

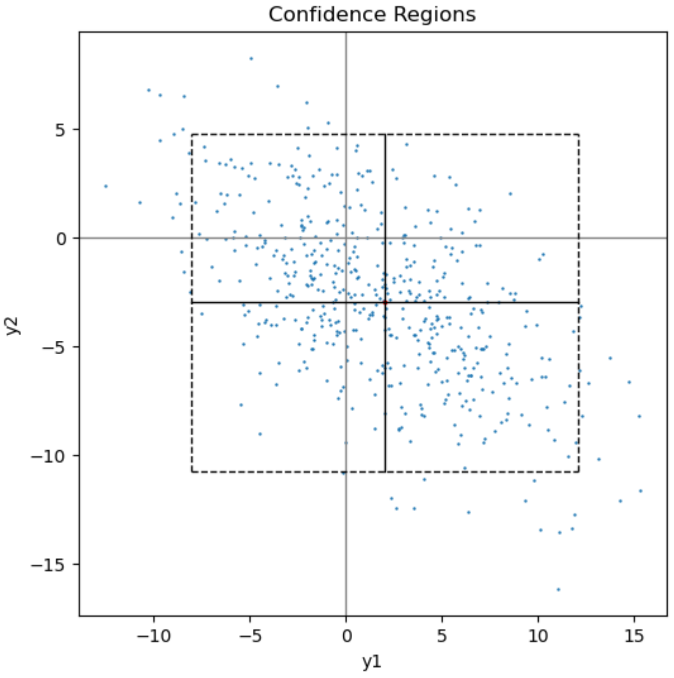


95% Confidence region for **each** dimension

MULTI-DIMENSIONAL CONFIDENCE INTERVALS

2 dimensional case

How to quantify the uncertainties in multiple dimension ?



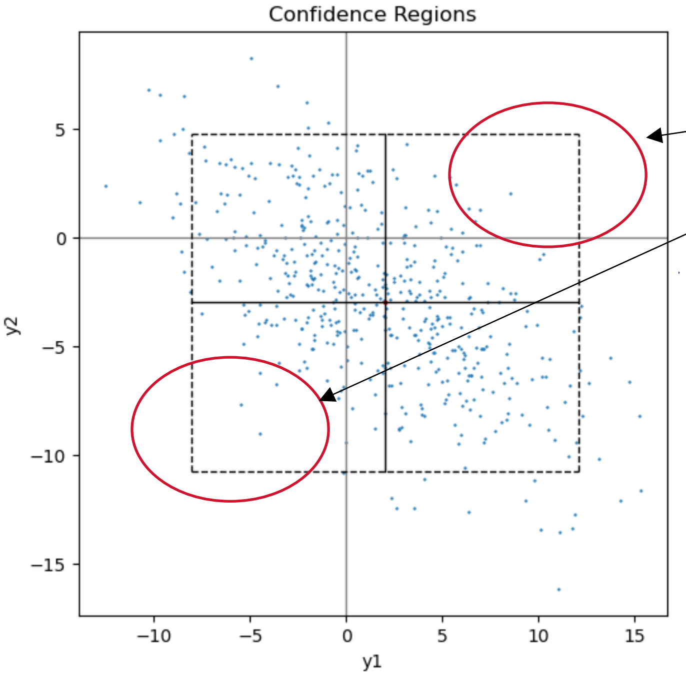
95% Confidence region for **each** dimension

Create a box around

MULTI-DIMENSIONAL CONFIDENCE INTERVALS

2 dimensional case

➡ *How to quantify the uncertainties in multiple dimension ?*



Almost no data

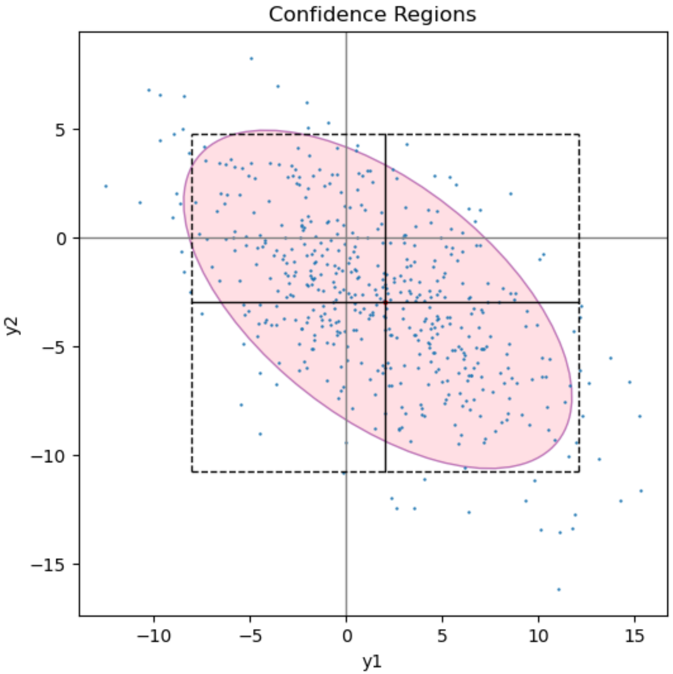
95% Confidence region for **each** dimension

Create a box around

MULTI-DIMENSIONAL CONFIDENCE INTERVALS

2 dimensional case

Can we do better ?



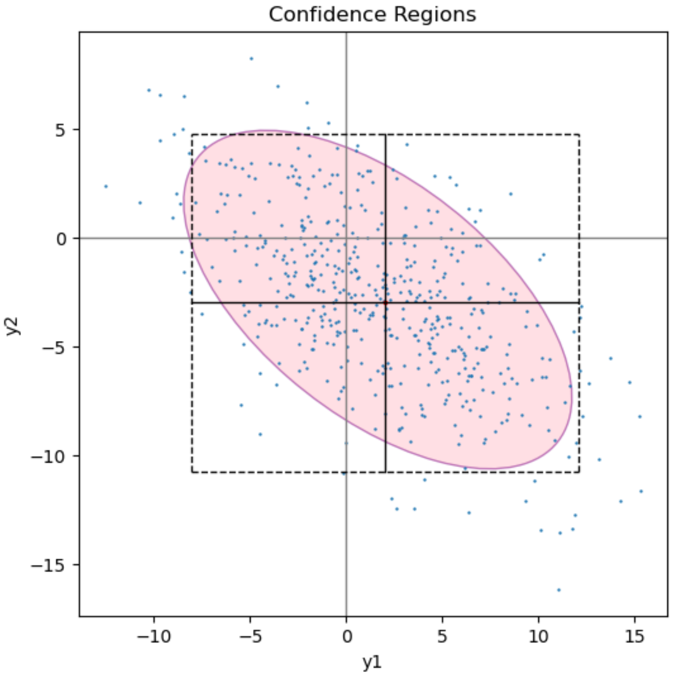
95% Confidence region for **each** dimension

Ellipse region

MULTI-DIMENSIONAL CONFIDENCE INTERVALS

2 dimensional case

*Am I sure that **both y1 and y2** are **simultaneously** inside the confidence region at level 5% ?*



95% Confidence region for **each** dimension

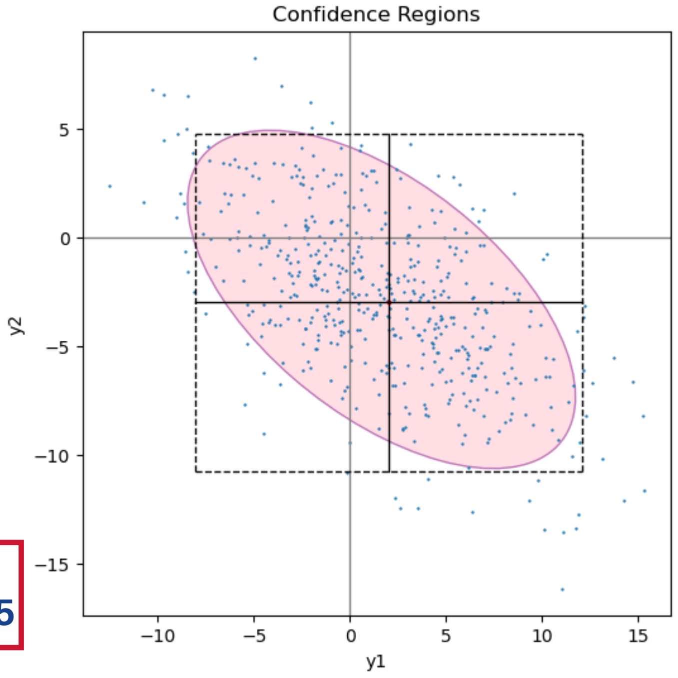
Ellipse region

MULTI-DIMENSIONAL CONFIDENCE INTERVALS

2 dimensional case

*Am I sure that **both y1 and y2** are **simultaneously** inside the confidence region at level 5% ?*

Independent intervals:
(0.95x0.95)x100 = 90.25



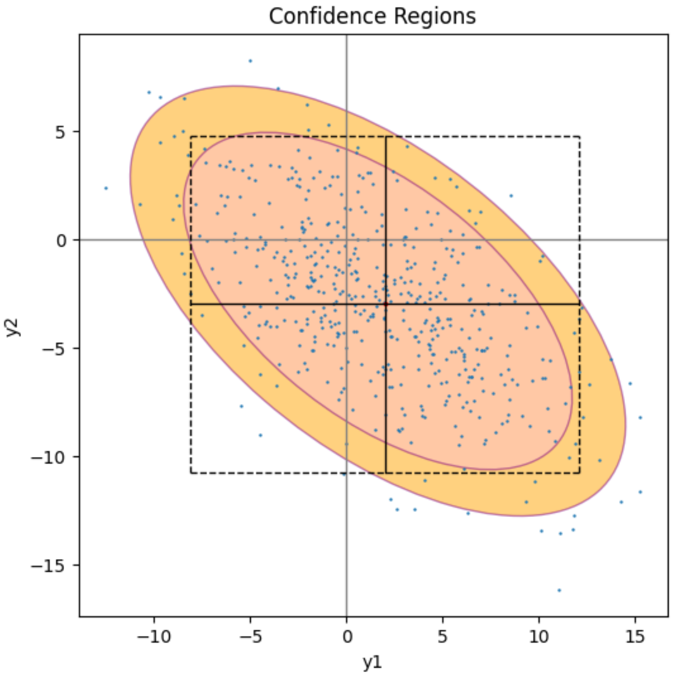
95% Confidence region for **each** dimension

Ellipse region

MULTI-DIMENSIONAL CONFIDENCE INTERVALS

2 dimensional case

*Am I sure that **both y1 and y2** are **simultaneously** inside the confidence region at level 5% ?*



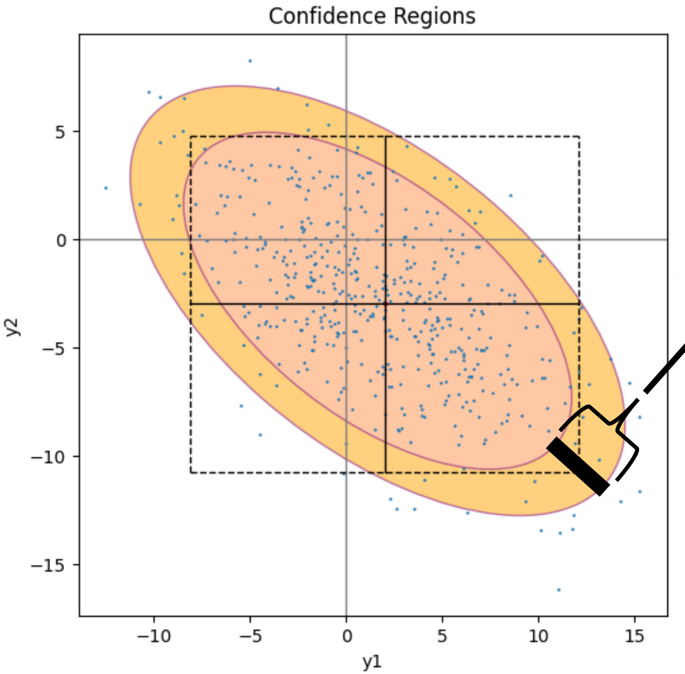
95% Simultaneous Confidence region

Ellipse region

MULTI-DIMENSIONAL CONFIDENCE INTERVALS

2 dimensional case

*Am I sure that **both y1 and y2** are **simultaneously** inside the confidence region at level 5% ?*



__How much__ should I __increase__ the __size__ of the ellipse region ?

95% Simultaneous Confidence region

Ellipse region

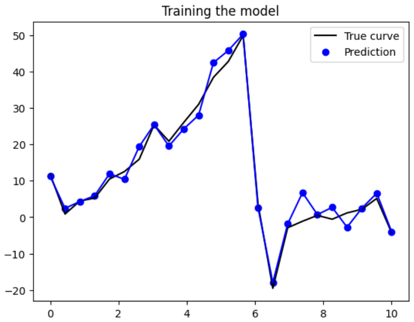
PART 2

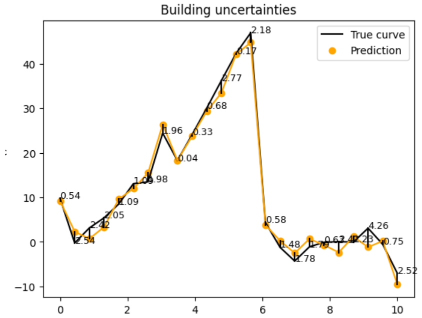
BRIEF INTRODUCTION TO « CONFORMAL PREDICTION »

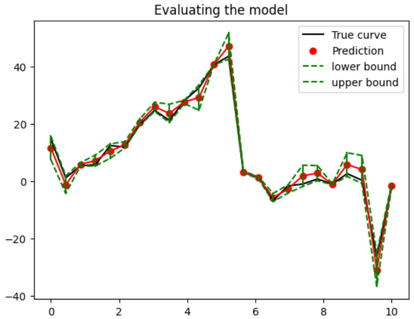
SAINT-GOBAIN

Data Split

Train 70%	Calibration 20%	Test 10%





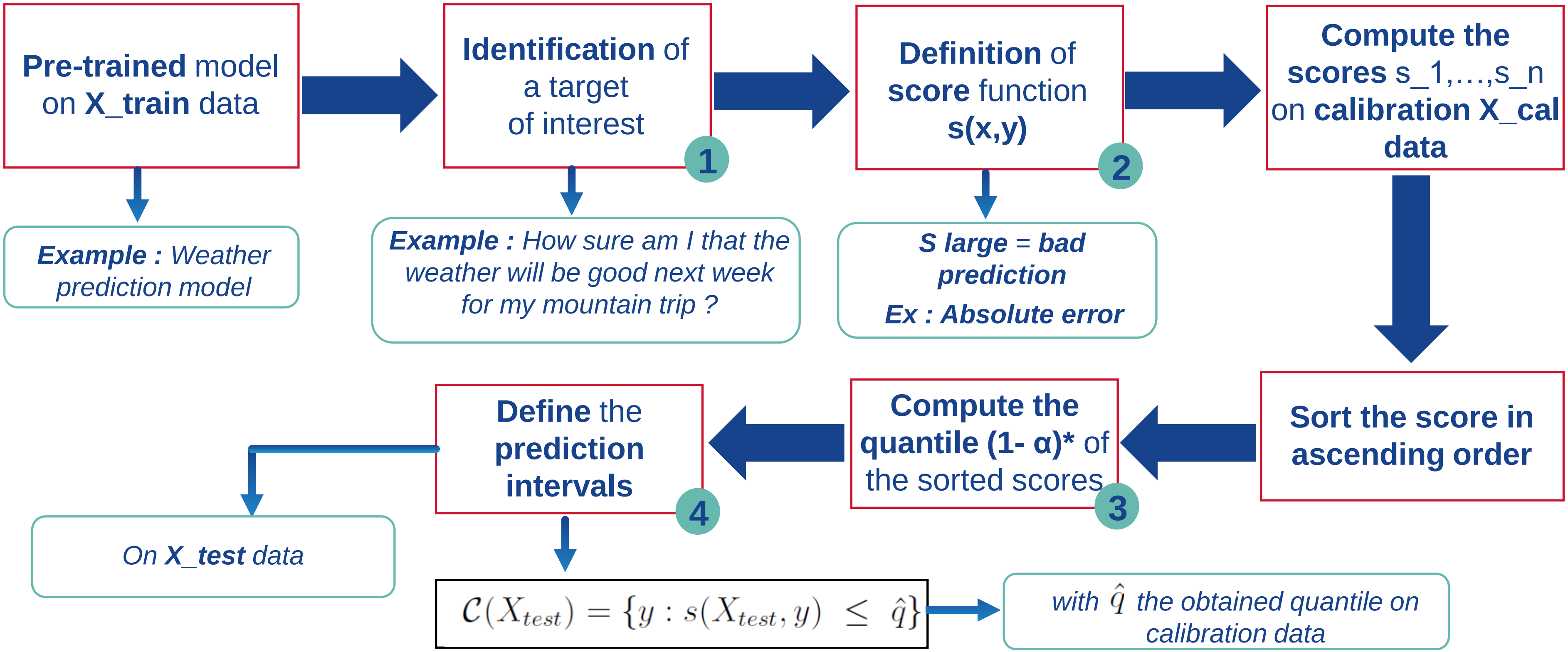


To **train** the **model**

To **compute** the **scores** and get the **uncertainties**

To **evaluate** model **performances** and quality of **prediction intervals**

HOW DOES IT WORK?



2 Step 2 : Define a score function

$$s(x_i, y_i) = |\hat{f}(x_i) - y_i| \quad \forall i \in \mathcal{C}_{calibration}$$

CONFORMAL PREDICTION METHOD : BIG PICTURE

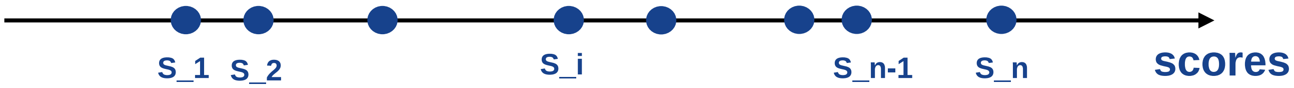
2 Step 2 : Define a score function

$$s(x_i, y_i) = |\hat{f}(x_i) - y_i| \quad \forall i \in \mathcal{C}_{calibration}$$

→ **Local Error Rate α**

CONFORMAL PREDICTION METHOD : BIG PICTURE

2 Step 2 : Compute and sort the scores

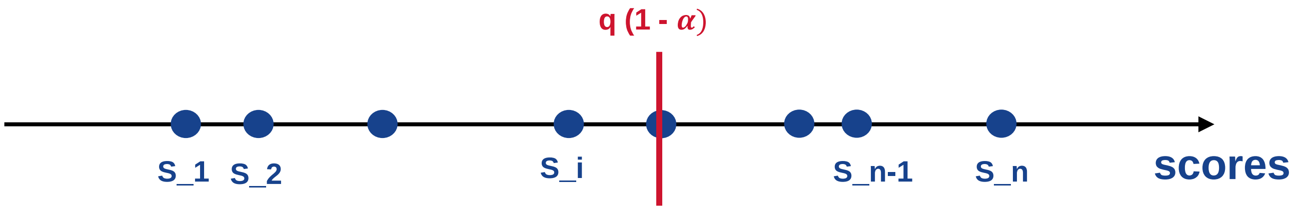


$$s(x_i, y_i) = |\hat{f}(x_i) - y_i| \quad \forall i \in \mathcal{C}_{calibration}$$

→ **Local Error Rate α**

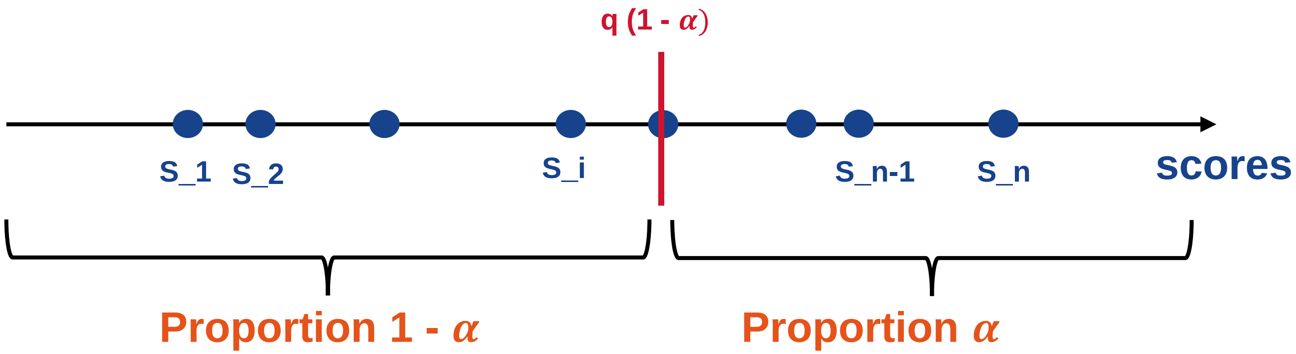
CONFORMAL PREDICTION METHOD : BIG PICTURE

3 Step 3 : Get the quantile



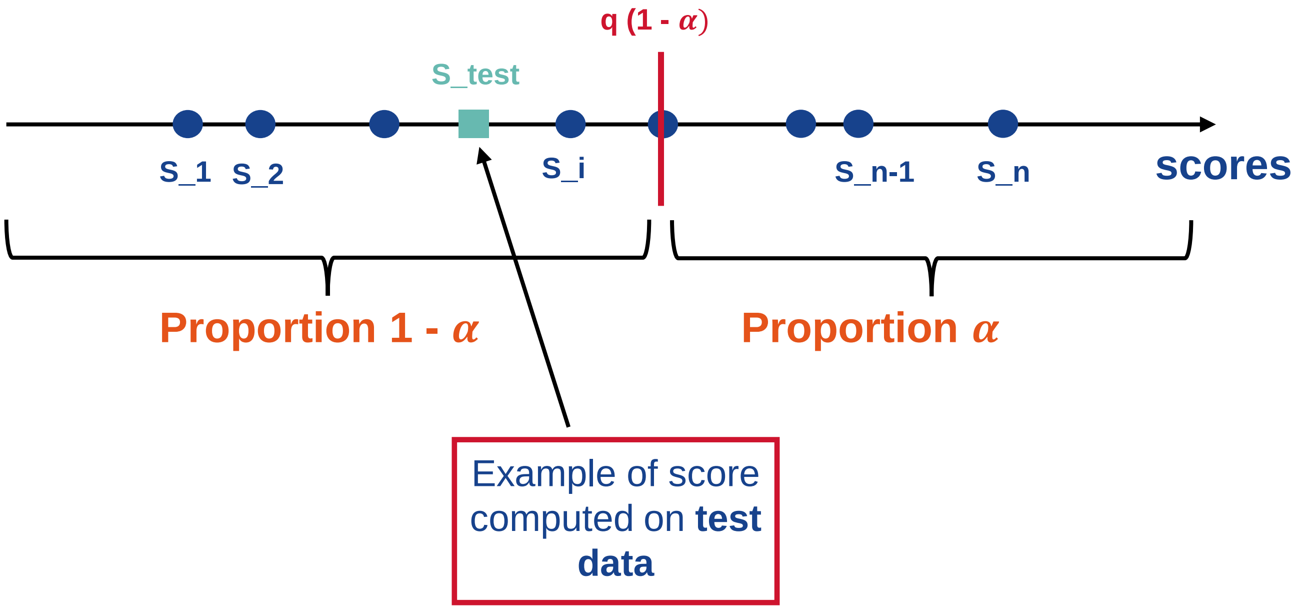
CONFORMAL PREDICTION METHOD : BIG PICTURE

3 Step 3 : Get the quantile



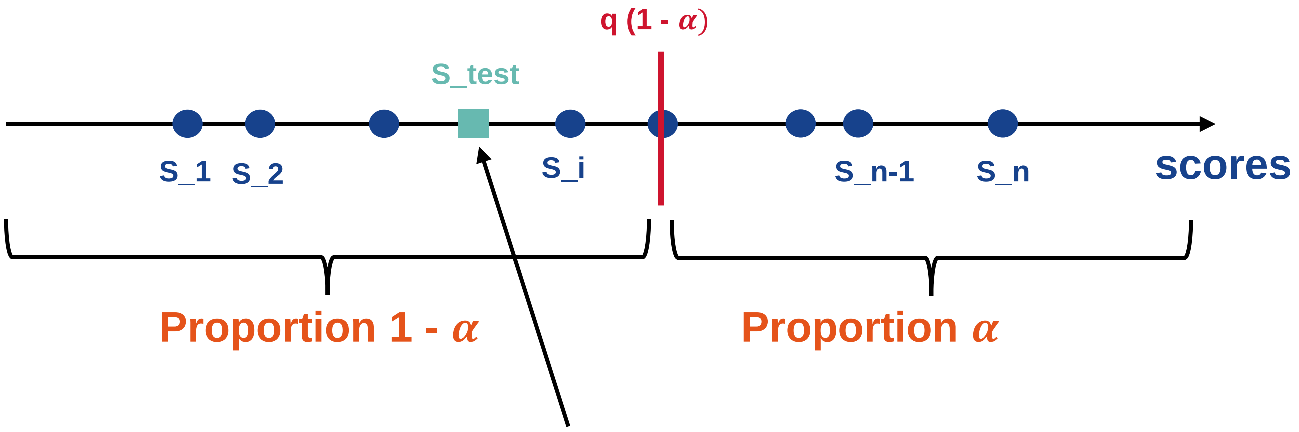
CONFORMAL PREDICTION METHOD : BIG PICTURE

4 Step 4 : Define the Prediction Intervals



CONFORMAL PREDICTION METHOD : BIG PICTURE

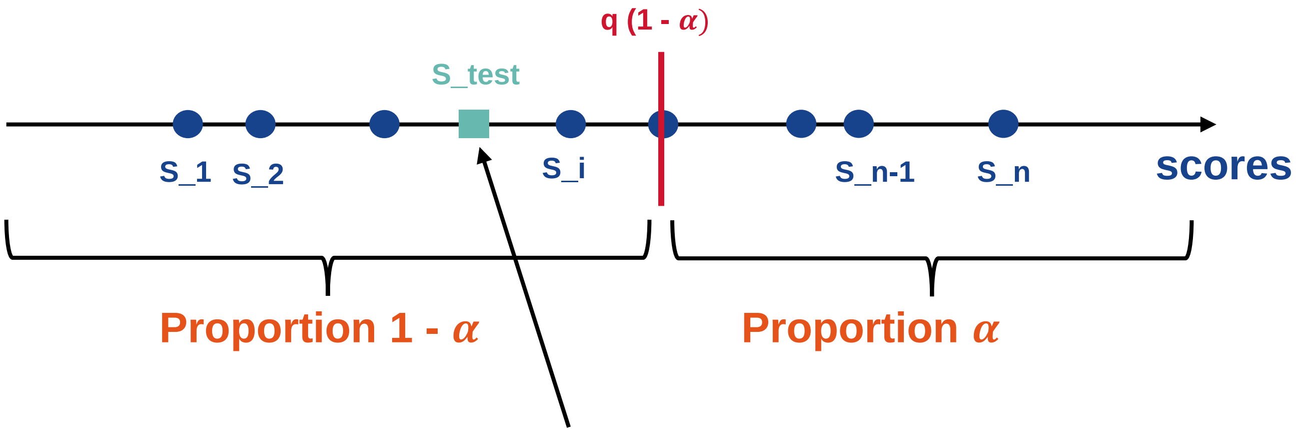
4 Step 4 : Define the Prediction Intervals



Example of score computed on **test data**

$P(S\_test < q\,(1-\alpha)) = 1 - \alpha$

4 Step 4 : Define the Prediction Intervals



$$C(x) = [f(x) - q(1-\alpha)\,;\, f(x) + q(1-\alpha)]$$

Prediction interval

Example of score computed on **test data**

$$P(S\_test < q\,(1 - \alpha)) = 1 - \alpha$$

THEOREM : CONFORMAL COVERAGE GUARANTEE

Theorem : Conformal coverage guarantee

Suppose (X_i, Y_i) for all $i = 1, ..., n$ *and* (X_{test}, Y_{test}) *are i.i.d.,we define* $\hat{q}$ *as in step 3 above and* $\mathcal{C}(X_{test})$ *as in step 4 above. Then the following holds :*

$$1 - \alpha \leq \mathbb{P}(Y_{test} \in \mathcal{C}(X_{test})) \leq 1 - \alpha + \frac{1}{n+1}$$

SOURCE : V. Vovk, A. Gammerman, and C. Saunders, “Machine-learning applications of algorithmic randomness,” in *International Conference on Machine Learning*, 1999, pp. 444–453

THEOREM : CONFORMAL COVERAGE GUARANTEE

Theorem : Conformal coverage guarantee

Suppose (X_i, Y_i) for all $i = 1, ..., n$ *and* (X_{test}, Y_{test}) *are i.i.d.,we define* $\hat{q}$ *as in step 3 above and* $\mathcal{C}(X_{test})$ *as in step 4 above. Then the following holds :*

$$1 - \alpha \leq \mathbb{P}(Y_{test} \in \mathcal{C}(X_{test})) \leq 1 - \alpha + \frac{1}{n+1}$$

Calibration data

SOURCE : V. Vovk, A. Gammerman, and C. Saunders, “Machine-learning applications of algorithmic randomness,” in *International Conference on Machine Learning*, 1999, pp. 444–453

THEOREM : CONFORMAL COVERAGE GUARANTEE

Theorem : Conformal coverage guarantee

Suppose (X_i, Y_i) for all $i = 1, ..., n$ *and* (X_{test}, Y_{test}) *are i.i.d.,we define* $\hat{q}$ *as in step 3 above and* $\mathcal{C}(X_{test})$ *as in step 4 above. Then the following holds :*

$$1 - \alpha \leq \mathbb{P}(Y_{test} \in \mathcal{C}(X_{test})) \leq 1 - \alpha + \frac{1}{n+1}$$

Calibration data

We can use a weaker hypothesis too !

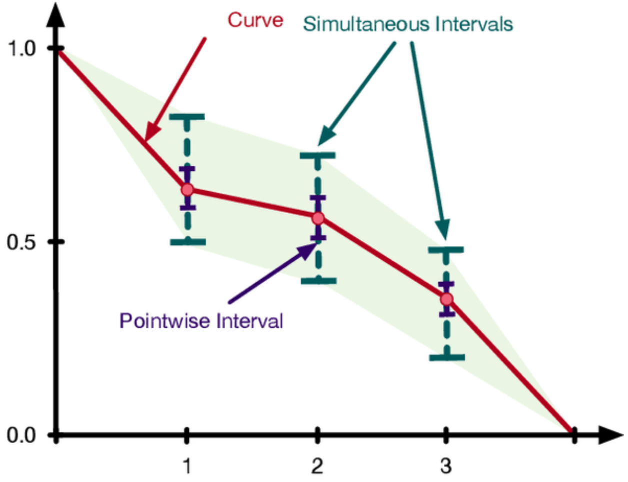
SOURCE : V. Vovk, A. Gammerman, and C. Saunders, “Machine-learning applications of algorithmic randomness,” in *International Conference on Machine Learning*, 1999, pp. 444–453

PART 3

OUR BETA-OPTIM METHOD

PROBLEM : SIMULTANEAOUS PREDICTION INTERVAL

Local vs Global



One-at-a-time (local)

$$\mathbb{P}(Y_{test,i} \in \mathcal{C}(X_{test,i})) \geq 1 - \alpha, \quad \forall\, i \in \{1, \ldots, d\}$$

Simultaneous (global)

$$\mathbb{P}(\forall\, i \in \{1, \ldots, d\}, Y_{test,i} \in \mathcal{C}(X_{test,i})) \geq 1 - \alpha$$

METHOD COMBINAISON: TFP X CP

Presentation

- <u>**Split Method**</u> **:**
  - *Train, Test & Calibration*
- <u>**Score Function**</u>**:**
  - Residual normalized score : $\frac{|Y - \hat{\mu}(X)|}{\hat{\sigma}(X)}$
- <u>**Prediction intervals**</u> **:**

$$[\hat{\mu}(X) - q(s) * \hat{\sigma}(X), \hat{\mu}(X) + q(s) * \hat{\sigma}(X)]$$

with **q(s)** the **(1 – α) quantile** of the **sorted scores.**

BETA-OPTIM STRATEGY

Build simultaneous prediction intervals for Y

$$Y = (y_1, y_2, \ldots, y_p) \in \mathbb{R}^p$$

BETA-OPTIM STRATEGY

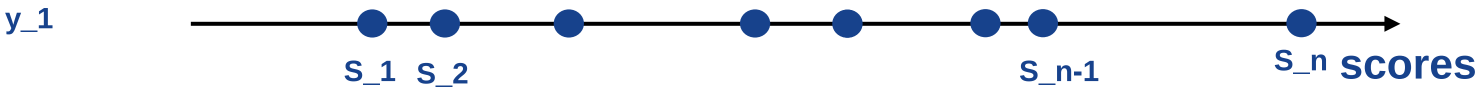
Build simultaneous prediction intervals for Y

$$Y = (y_1, y_2, \ldots, y_p) \in \mathbb{R}^p$$

→ **Local Error Rate β**

Build simultaneous prediction intervals for Y

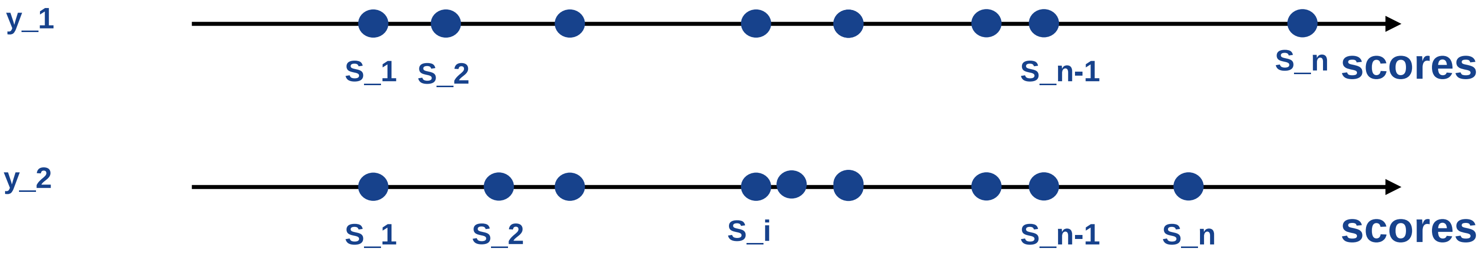
$$Y = (y_1, y_2, \ldots, y_p) \in \mathbb{R}^p$$



BETA-OPTIM STRATEGY

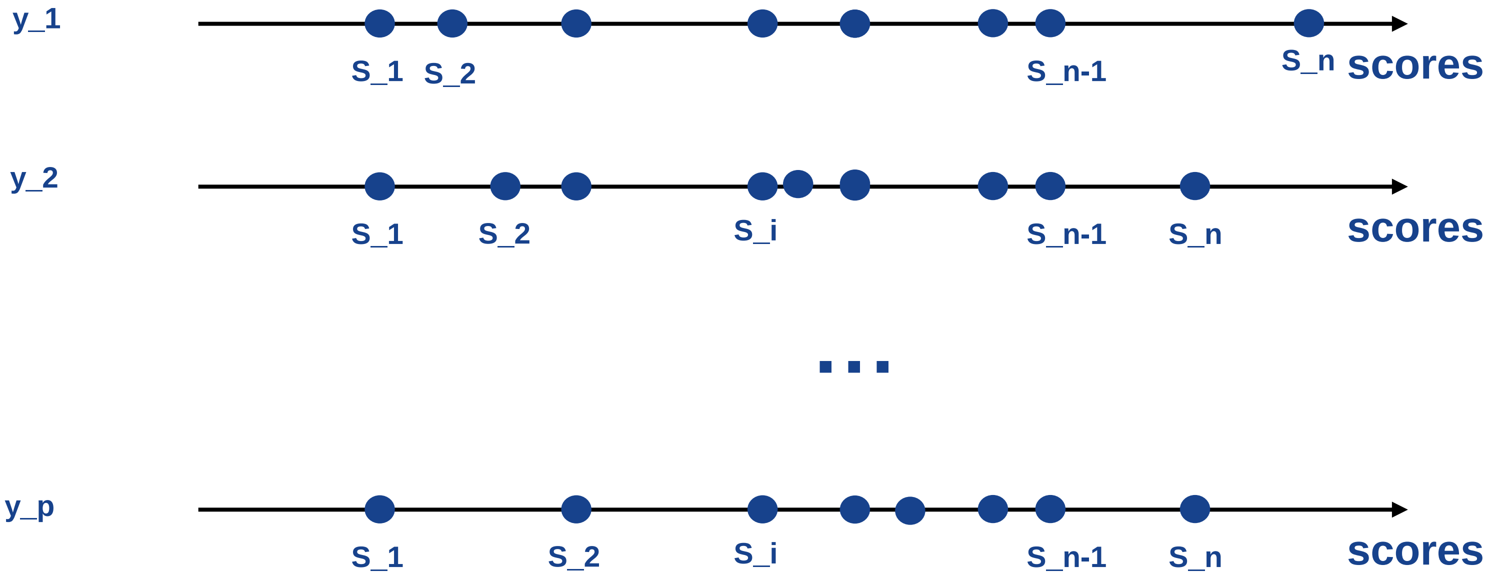
Build simultaneous prediction intervals for Y

$$Y = (y_1, y_2, \ldots, y_p) \in \mathbb{R}^p$$



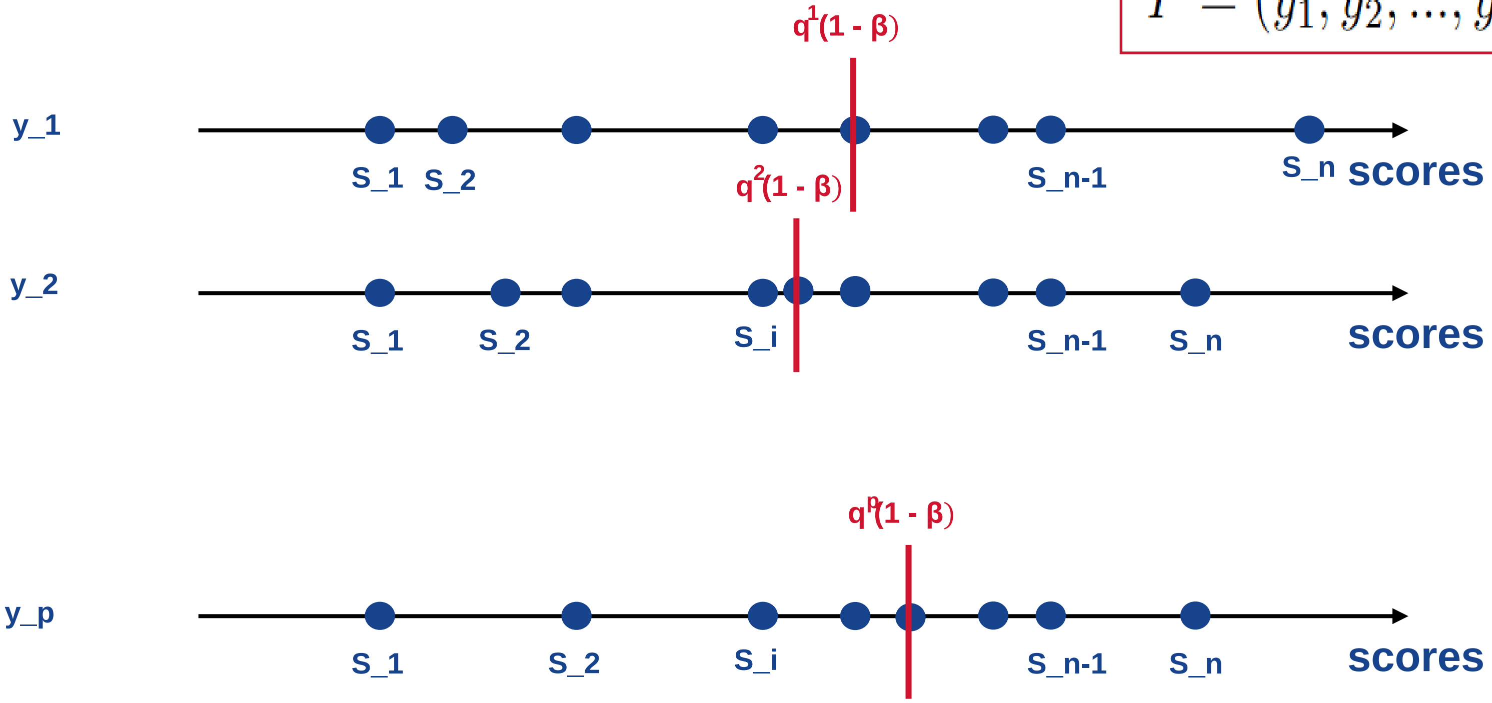
Build simultaneous prediction intervals for Y

$$Y = (y_1, y_2, \ldots, y_p) \in \mathbb{R}^p$$



Build simultaneous prediction intervals for Y

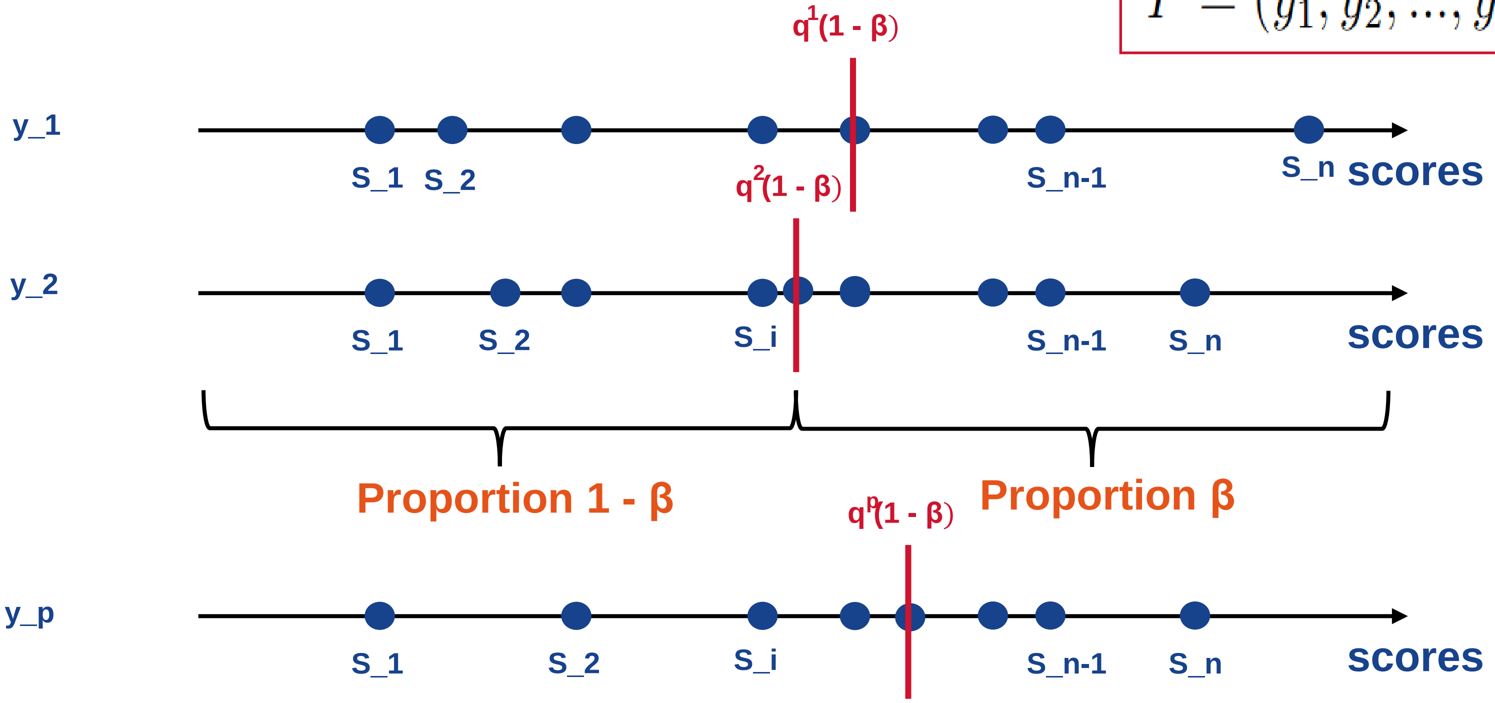
$$Y = (y_1, y_2, \ldots, y_p) \in \mathbb{R}^p$$



BETA-OPTIM STRATEGY

Build simultaneous prediction intervals for Y

$$Y = (y_1, y_2, \dots, y_p) \in \mathbb{R}^p$$



BETA-OPTIM STRATEGY

Build simultaneous prediction intervals for Y

dim

y_1

y_2

⋮

y_p

BETA-OPTIM STRATEGY

Build simultaneous prediction intervals for Y

dim	quantiles
y_1	$q^1(1 - \beta)$
y_2	$q^2(1 - \beta)$
⋮	⋮
y_p	$q^p(1 - \beta)$

BETA-OPTIM STRATEGY

Build simultaneous prediction intervals for Y

dim	quantiles	Prediction intervals
y_1	$q^1(1 - \beta)$	$C^1_\beta(x)$
y_2	$q^2(1 - \beta)$	$C^2_\beta(x)$
⋮	⋮	⋮
y_p	$q^p(1 - \beta)$	$C^p_\beta(x)$

Build simultaneous prediction intervals for Y

dim	quantiles	Prediction intervals
y_1	$q^1(1 - \beta)$	$C^1_\beta(x)$
y_2	$q^2(1 - \beta)$	$C^2_\beta(x)$
⋮	⋮	⋮
y_p	$q^p(1 - \beta)$	$C^p_\beta(x)$

Simultaneous Coverage Evaluation

Build simultaneous prediction intervals for Y

dim	quantiles	Prediction intervals
y_1	$q^1(1 - \beta)$	$C^1_\beta(x)$
y_2	$q^2(1 - \beta)$	$C^2_\beta(x)$
⋮	⋮	⋮
y_p	$q^p(1 - \beta)$	$C^p_\beta(x)$

Simultaneous Coverage Evaluation

$$(x, y) \in D_{Cal}\ , for\ all\ i\ \in \{1, \dots, p\}\ , y_i \in C^i_\beta(x)$$

Build simultaneous prediction intervals for Y

dim	quantiles	Prediction intervals
y_1	$q^1(1 - \beta)$	$C_\beta^1(x)$
y_2	$q^2(1 - \beta)$	$C_\beta^2(x)$
⋮	⋮	⋮
y_p	$q^p(1 - \beta)$	$C_\beta^p(x)$

Simultaneous Coverage Evaluation

↓

$$(x, y) \in D_{Cal}\ , for\ all\ i\ \in \{1, \dots, p\}\ , y_i \in C_\beta^i(x)$$

↓

In which proportion ?

BETA-OPTIM STRATEGY

Build simultaneous prediction intervals for Y

dim	quantiles	Prediction intervals
y_1	$q^1(1 - \beta)$	$C^1_\beta(x)$
y_2	$q^2(1 - \beta)$	$C^2_\beta(x)$
⋮	⋮	⋮
y_p	$q^p(1 - \beta)$	$C^p_\beta(x)$

Simultaneous Coverage Evaluation

↓

$$(x, y) \in D_{Cal}\ , for\ all\ i\ \in \{1, \dots, p\}\ , y_i \in C^i_\beta(x)$$

↓

In which proportion ?

↓

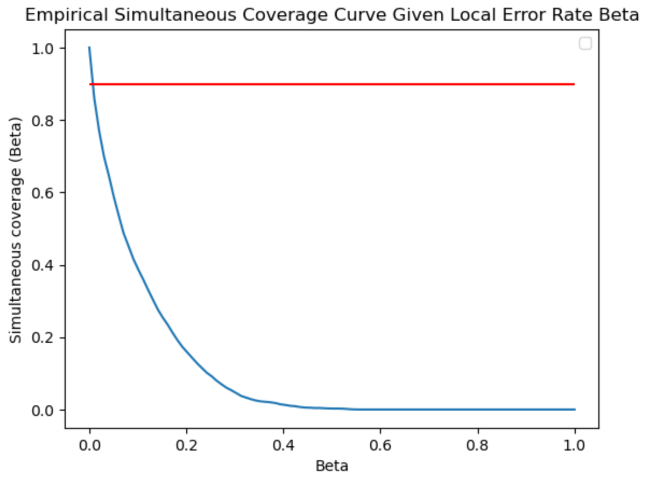
Find **local error** rate **β** such that we achieve **global error α on calibration data**

Objective function

$$\mathcal{J}(\beta) = |simcov(\beta) - (1-\alpha)|$$

with

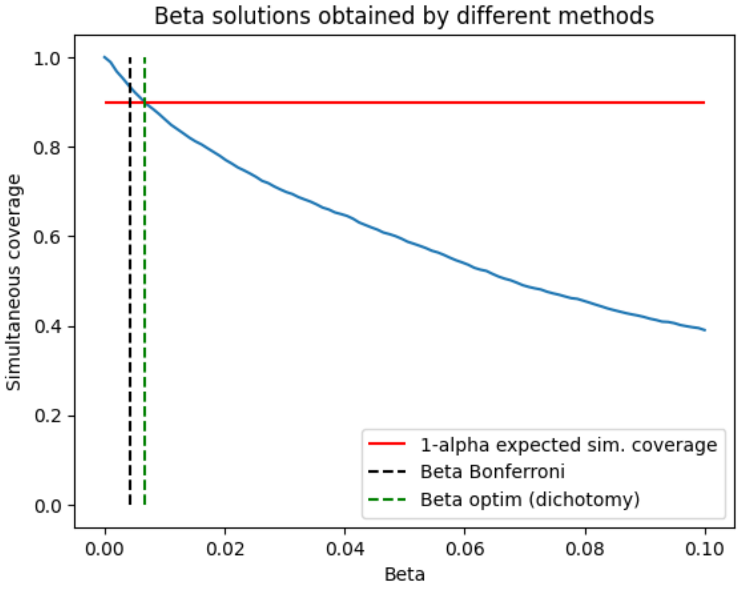
$$simcov(\beta) = \frac{\sum_{(x,y)\in\mathcal{D}_{cal}} \prod_{i\in\{1,\dots,p\}} \mathbb{1}\{y_i \in \mathcal{C}_\beta(x_i)\}}{|\mathcal{D}_{cal}|}$$



Find **local error** rate **β** such that we achieve **global error α on calibration data**

RESULTS

Results $(\alpha = 0.1)$



Method	Emp. Sim. Coverage
Bonferroni	**0,928**
Beta optim ★	**0,909**

Exact sim. coverage achieved !

PART 4

CONCLUSION

CONCLUSION

What ?

Simultaneous Prediction Intervals

Multi-Output Regression Problem

Conformal Split Method
+
Optimization Problem

Beta-Optim

CONCLUSION

What ?

Simultaneous Prediction Intervals

Multi-Output Regression Problem

Conformal Split Method
+
Optimization Problem

Beta-Optim

Characteristics ?

Agnostic to the Model

Distribution-Free

Valid in Finite Sample

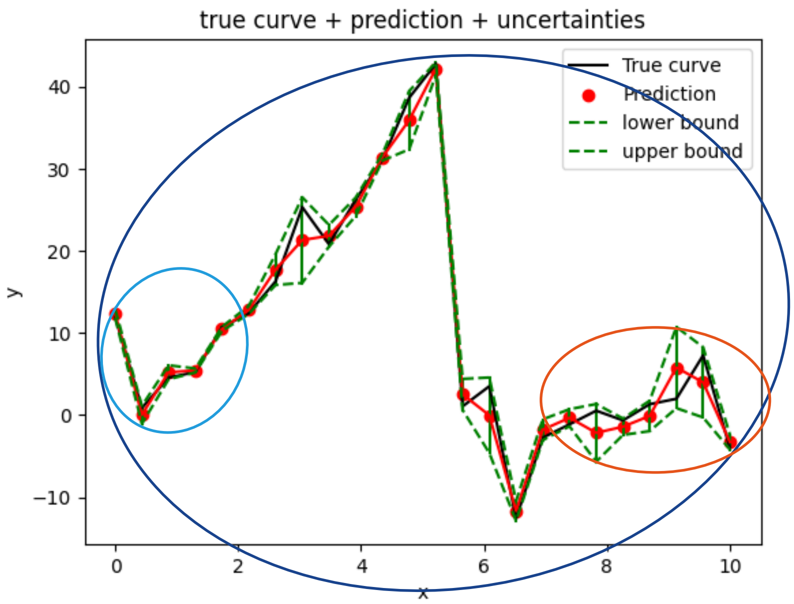
Conformal Pediction

CONCLUSION

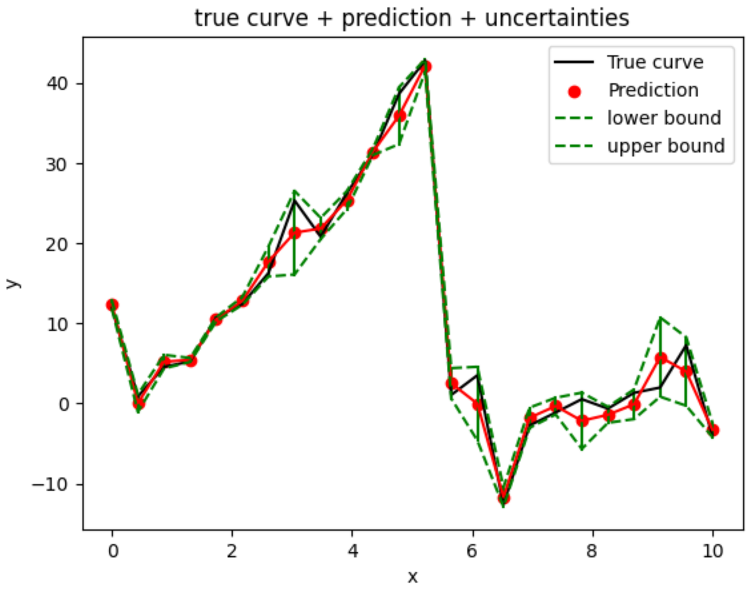
What ?	Characteristics ?	Properties ?
Simultaneous Prediction Intervals	Agnostic to the Model	Adaptative
Multi-Output Regression Problem	Distribution-Free	Thinner when I’m sure, Larger Otherwise
Conformal Split Method + Optimization Problem	Valid in Finite Sample	Robust to Heteroscedasticity
Beta-Optim	Conformal Pediction	Standard Deviation

CONCLUSION

What ?	Characteristics ?	Properties ?
Simultaneous Prediction Intervals	Agnostic to the Model	Adaptative
Multi-Output Regression Problem	Distribution-Free	Thinner when I'm sure, Larger Otherwise
Conformal Split Method + Optimization Problem	Valid in Finite Sample	Robust to Heteroscedasticity
Beta-Optim	Conformal Pediction	Standard Deviation



PERSPECTIVES



How to transition from a **pointwise** to a **continious uncertainty** with our curves ?

Deepen the concept and theoretical results about **adaptativity**

BETA-OPTIM: UNCERTAINTY QUANTIFICATION METHOD FOR MULTI-OUTPUT REGRESSION PROBLEMS

Jaad BELHOUARI-DURAND

Supervised by Paul Malfrait and Athmane BAKHTA,
2DS/CODA/SGR Paris

Collaboration with Matthieu Gallezot and Sylvain Berger
Acoustics/CODA/SGR Paris

SAINT-GOBAIN

APPENDIX
ADAPTATIVITY

NOTION OF ADAPTATIVITY

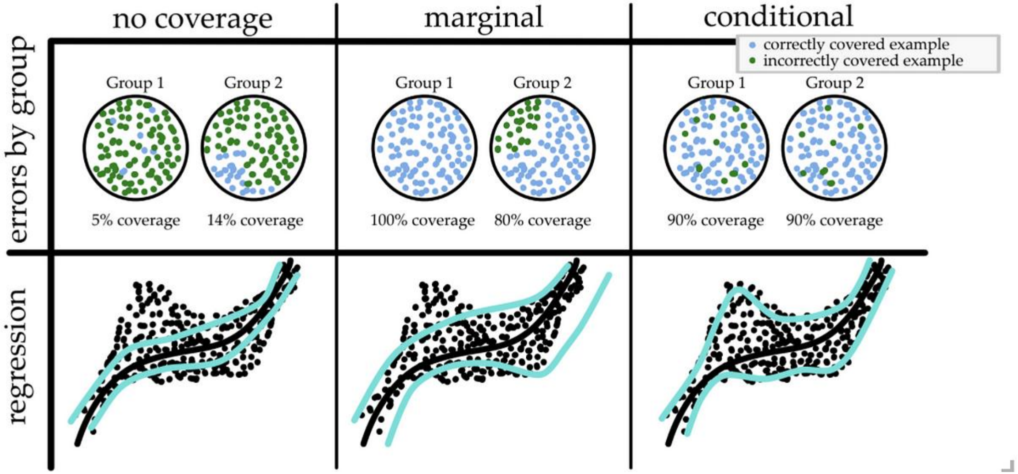
Conditional coverage

$$\mathbb{P}[Y_{test} \in \mathcal{C}(X_{test}) | X_{test}] \geq 1 - \alpha$$

I want my **prediciton interval** to **adapted to the input** given

Marginal coverage

$$1 - \alpha \leq \mathbb{P}(Y_{test} \in \mathcal{C}(X_{test}))$$

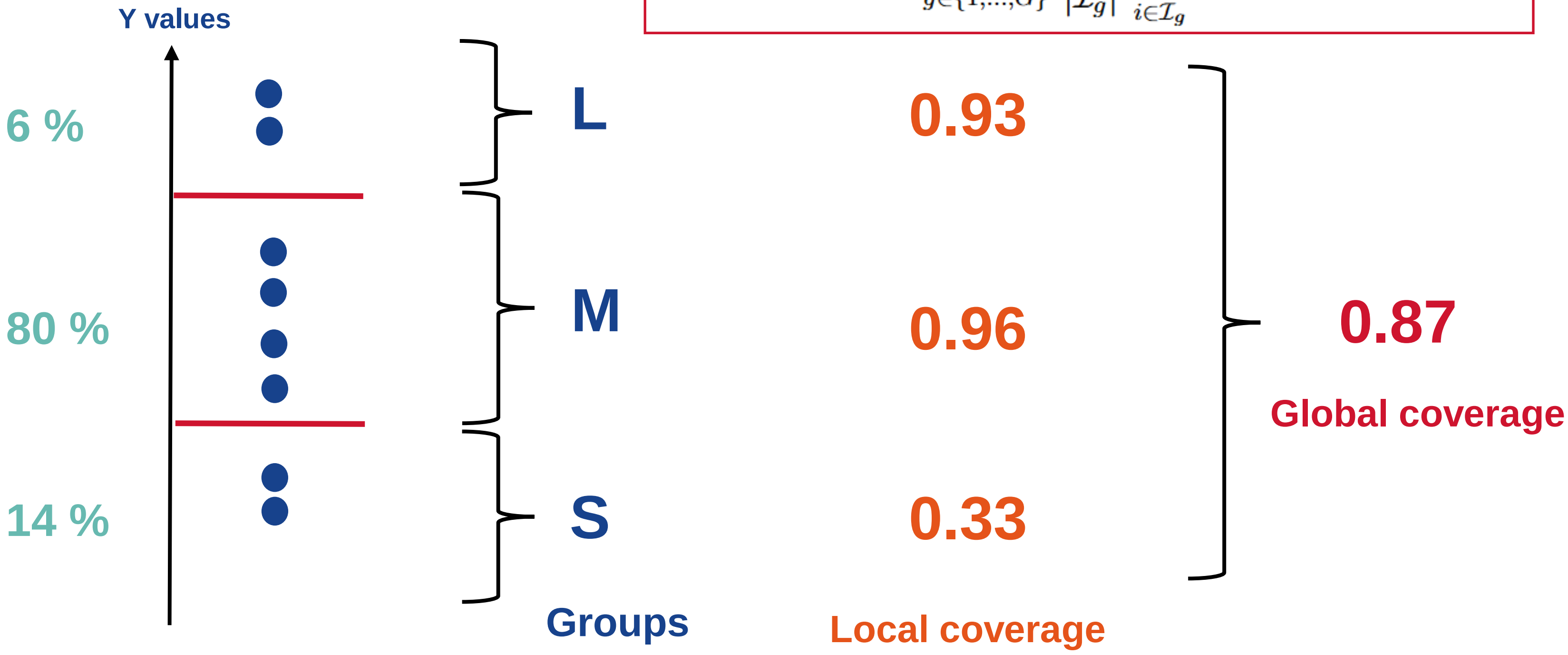


SOURCE : ANASTASIOS N. ANGELOPOULOS

CONFORMAL PREDICTION METHOD : BIG PICTURE

- **TSC metric:**

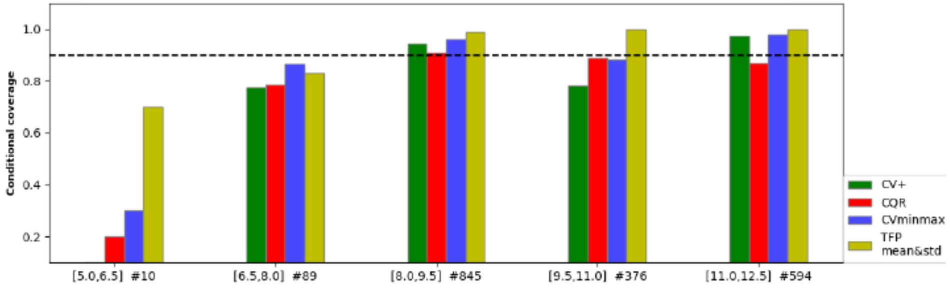
$$TSC\ metric: \quad \min_{g \in \{1,\ldots,G\}} \frac{1}{|\mathcal{I}_g|} \sum_{i \in \mathcal{I}_g} \mathbb{1}\{Y_i^{(val)} \in \mathcal{C}(X_i^{(val)})\}$$



ADAPTATIVITY ?

Target-Stratified Coverage (TSC)

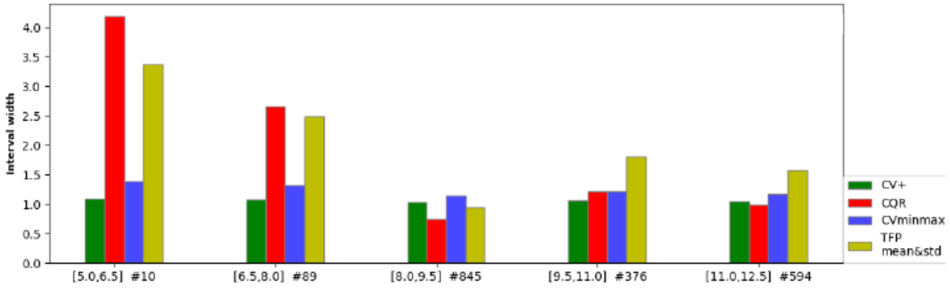
- **Idea** :
  - We build **5 groups {XS, S, M, L, XL}** regardig the **magnitude** of the prediction



(a) Conditional coverage

Observations:

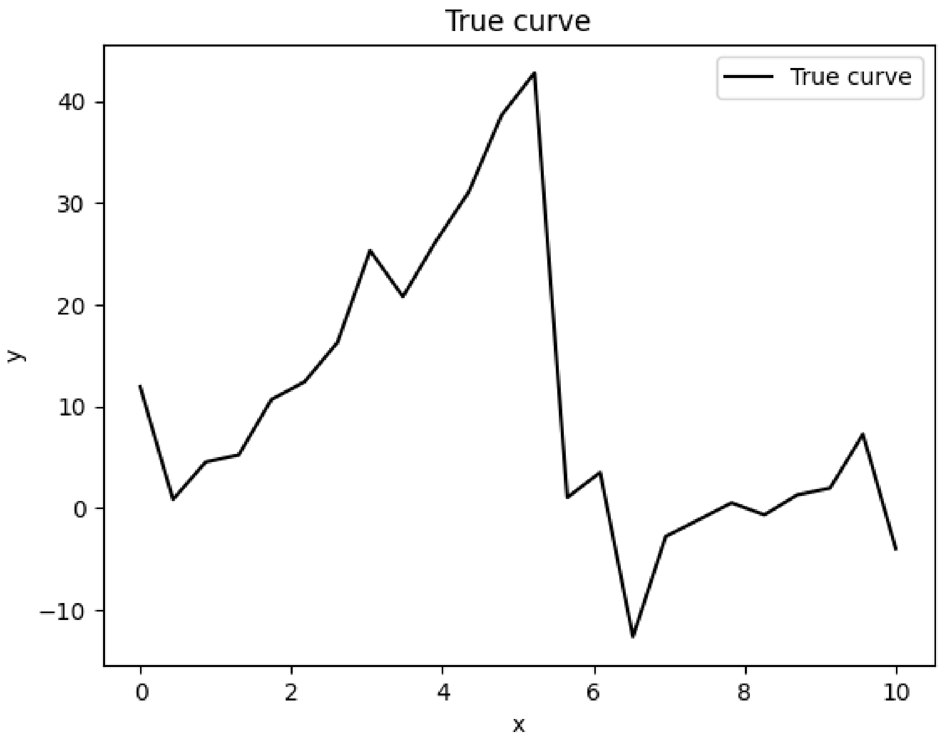
- **TFP and CQR** are the only method providing **adaptative prediction intervals**
- TFP method is the only one close to the **desired coverage**



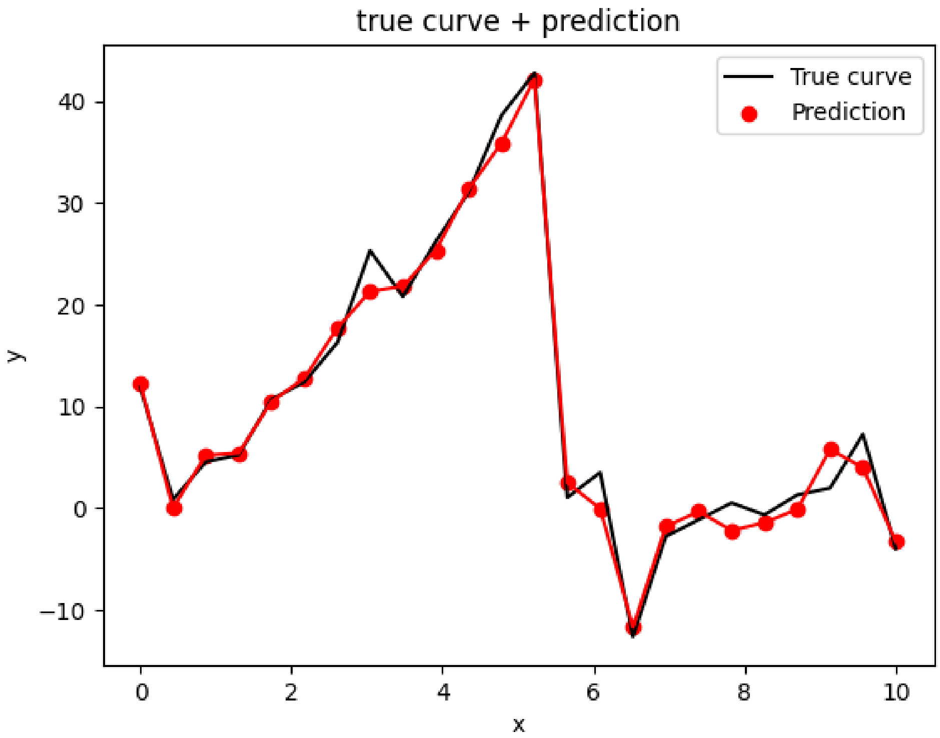
(b) Conditional interval width

APPENDIX
CURVES

RECAP

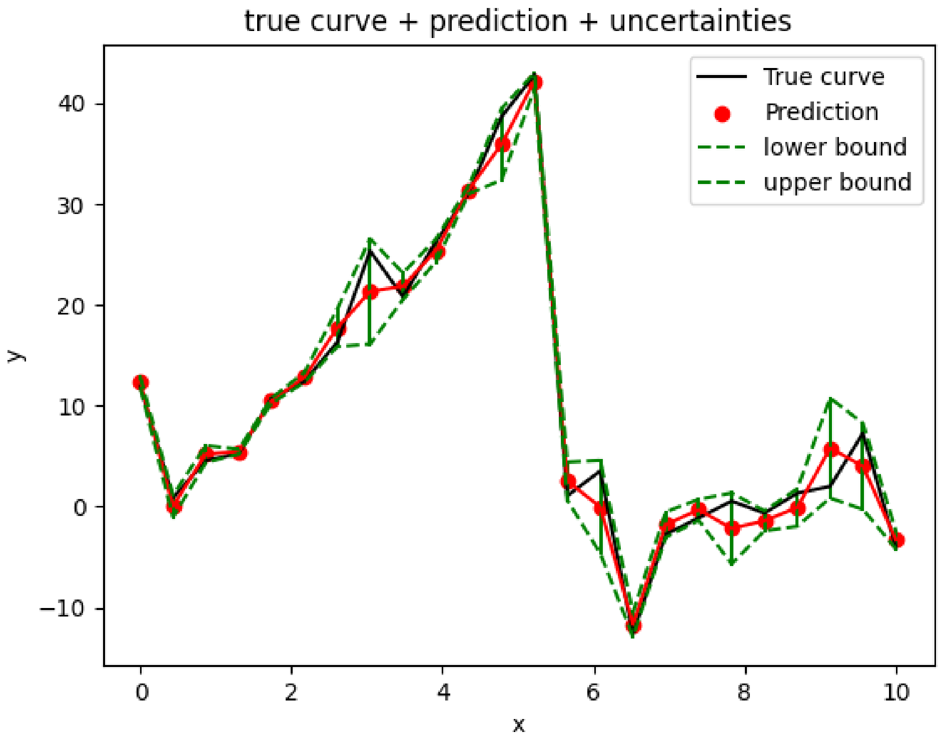


RECAP



true curve + prediction
True curve
Prediction
40
30
20
10
0
−10
y
0
2
4
6
8
10
x

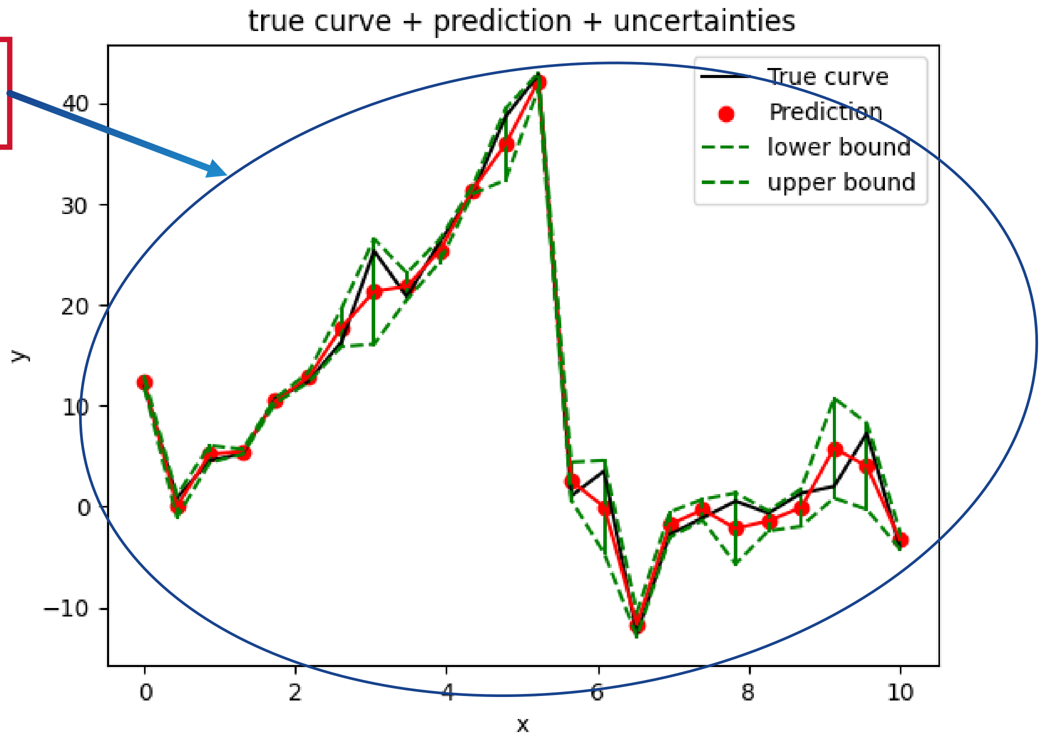
RECAP



true curve + prediction + uncertainties
True curve
Prediction
lower bound
upper bound
40
30
20
10
0
−10
y
0
2
4
6
8
10
x

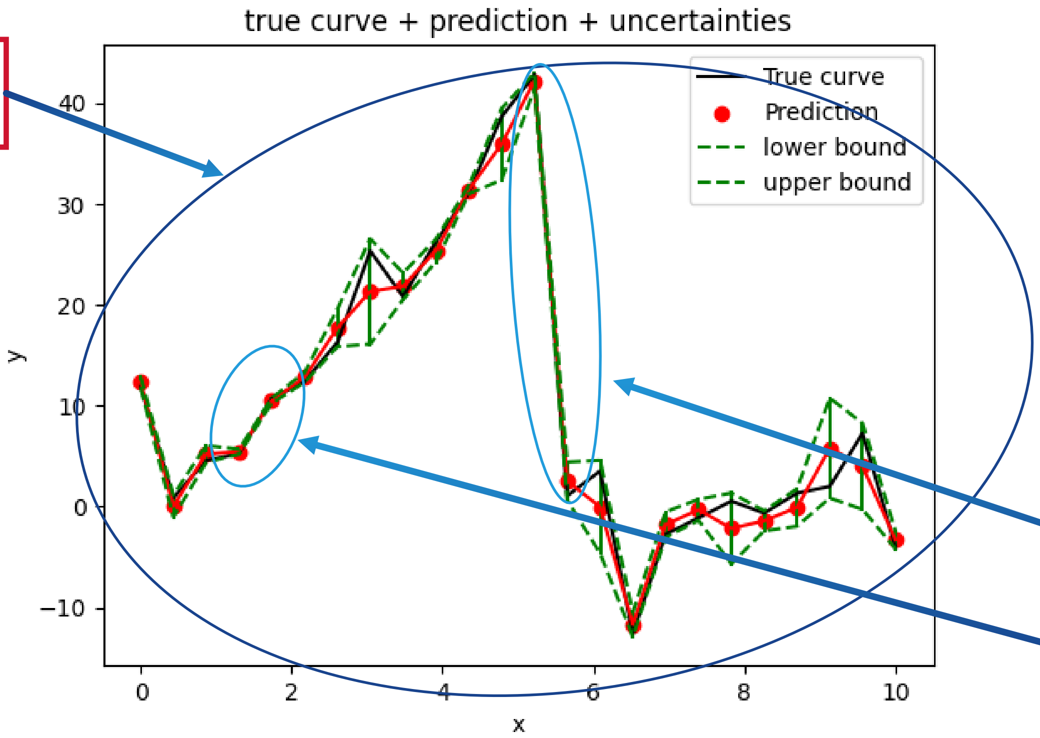
RECAP

« **Contains** the true curve »



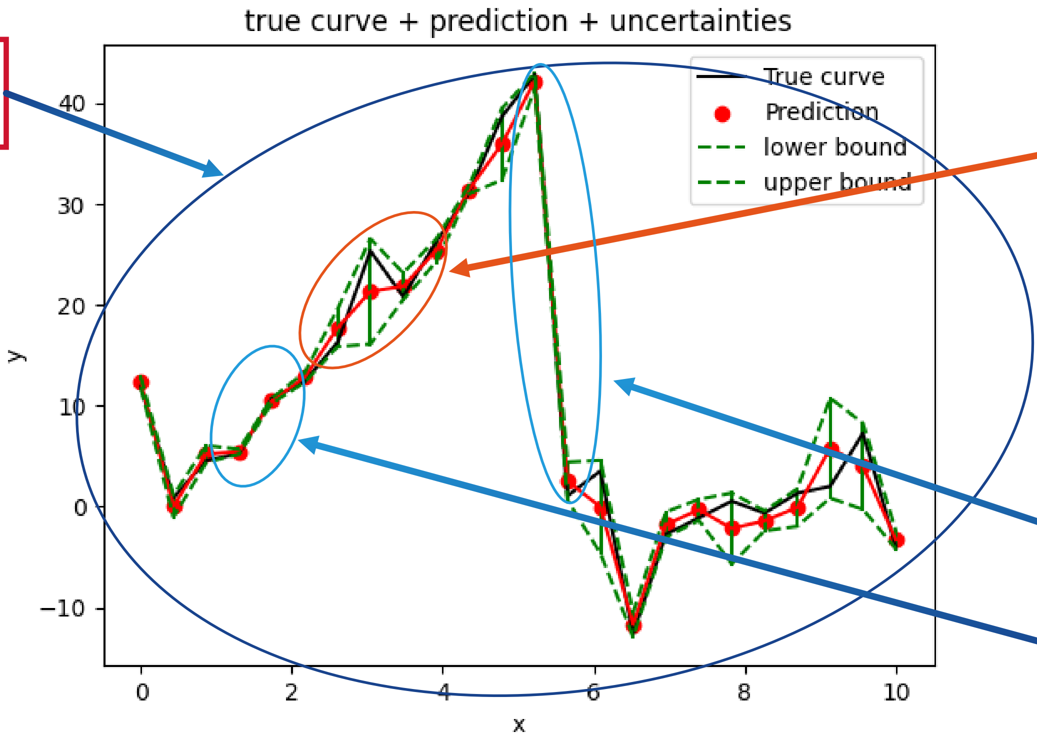
RECAP

« **Contains** the true curve »



« **thin** when I am **sure** »

« **Contains** the true curve »



« **Wider** when I am **more uncertain** »

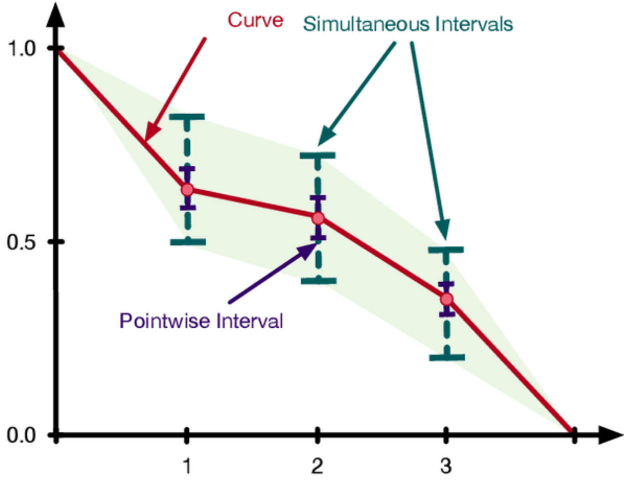
« **thin** when I am **sure** »

APPENDIX
BONFERRONI CORRECTION

SAINT-GOBAIN

PROBLEM : SIMULTANEAOUS PREDICTION INTERVAL

Local vs Global



One-at-a-time (local)

$$\mathbb{P}(Y_{test,i} \in \mathcal{C}(X_{test,i})) \geq 1 - \alpha, \ \ \forall\, i \in \{1, \ldots, d\}$$

Simultaneous (global)

$$\mathbb{P}(\forall\, i \in \{1, \ldots, d\}, Y_{test,i} \in \mathcal{C}(X_{test,i})) \geq 1 - \alpha$$

FIRST SOLUTION: BONFERRONI

- **Goal : Make the entire curve fall within the prediction interval with probability 1 - α**

  - **Bonferroni Strategy :**
    - Find a local error **β_Bonf** such that the **FWER** is **α**
    - **β_Bonf = α/d** with d the number of prediction intervals

Bonferroni inequality:

$$\mathbb{P}(\forall\ i \in \{1, ..., d\}, Y_{test,i} \in \mathcal{C}(X_{test,i}))\ \geq\ 1 - d \times \alpha$$

Results:

- **Empirical simultaneous coverage: 0.92**

Empirical probability that my whole curve fall within the prediction interval

FIRST SOLUTION: BONFERRONI

- **Goal : Make the entire curve fall within the prediction interval with probability 1 - α**

  - **Bonferroni Strategy :**
    - Find a local error **β_Bonf** such that the **FWER** is **α**
    - **β_Bonf = α/d** with d the number of prediction intervals

Bonferroni inequality:

$$\mathbb{P}(\forall\ i \in \{1, ..., d\}, Y_{test,i} \in \mathcal{C}(X_{test,i}))\ \geq\ 1 - d \times \alpha$$

Results:

- **Empirical simultaneous coverage: 0.92**

Too conservative !

Empirical probability that my whole curve fall within the prediction interval