Tutoriel : Prédiction Conforme

Joseba Dalmau et Mouhcine Mendil

DEEL
Dependable, Explainable & Embeddable Learning

IRT SAINT EXUPÉRY

PFIA 2024
Plate-forme Intelligence Artificielle

IRT SAINT EXUPÉRY

DEEL

Dependable, Explainable & Embeddable Learning

MODÈLES D'APPRENTISSAGE ET QUANTIFICATION D'INCERTITUDES

- Peut-on faire confiance aux prédiction de notre modèle ?
- Dans quelles circonstances ?
- À quel point ?

TECNIQUES CLASSIQUES DE QUANTIFICATION D'INCERTITUDES

- Méthodes Bayesiennes
- Méthodes Ensemblistes
- Techniques de Dropout
- ...

NON POST-HOC

SANS GARANTIES

Prédiction conforme : But et garantie

Données : un prédicteur $\hat{f} : \mathcal{X} \longrightarrow \mathcal{Y}$ et un taux d'erreur nominale α

Construire : $\hat{C}_\alpha : \mathcal{X} \longrightarrow \mathcal{P}(\mathcal{Y})$

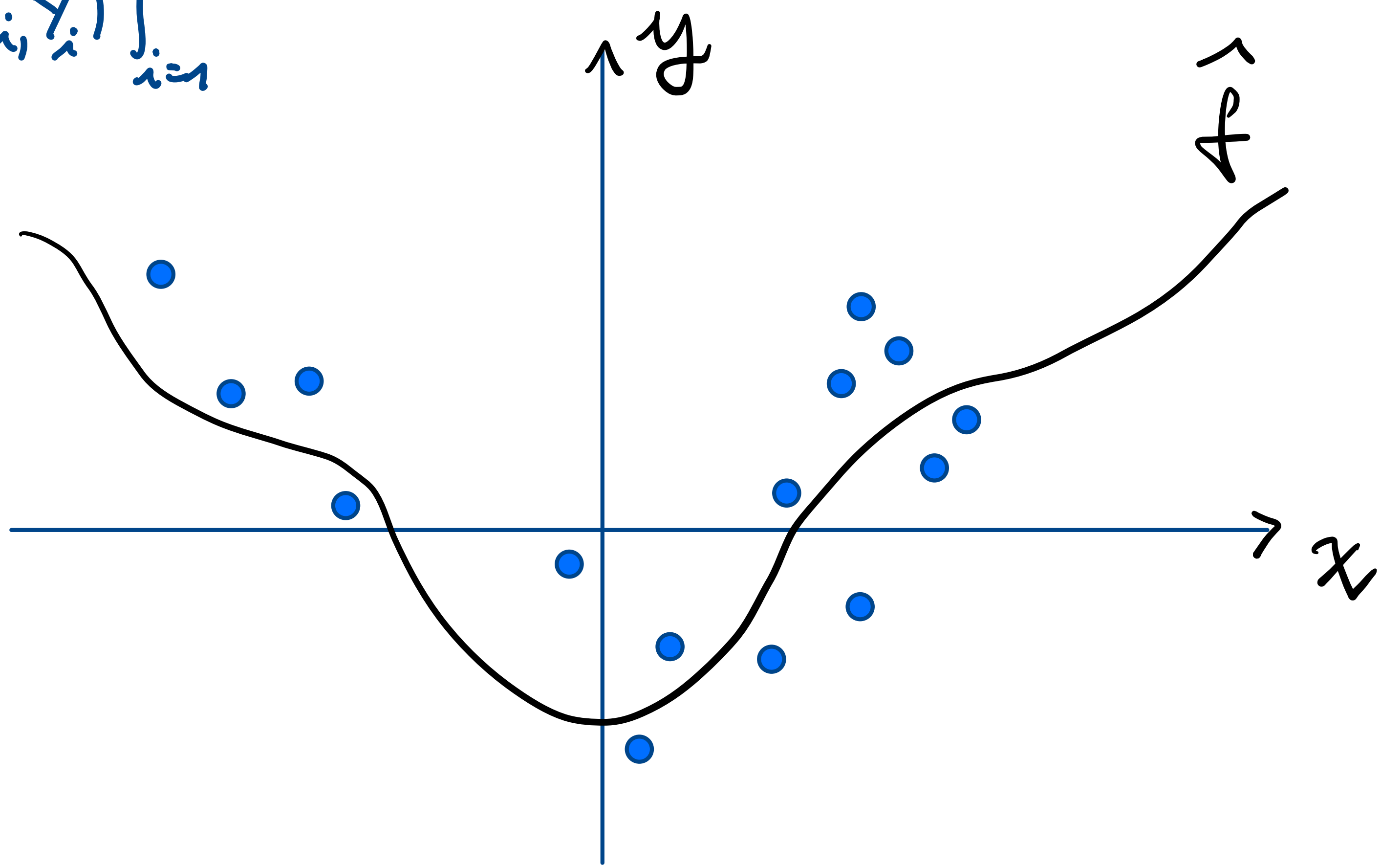
Avec la garantie :

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

CALIBRATION

On utilise un ensemble de calibration

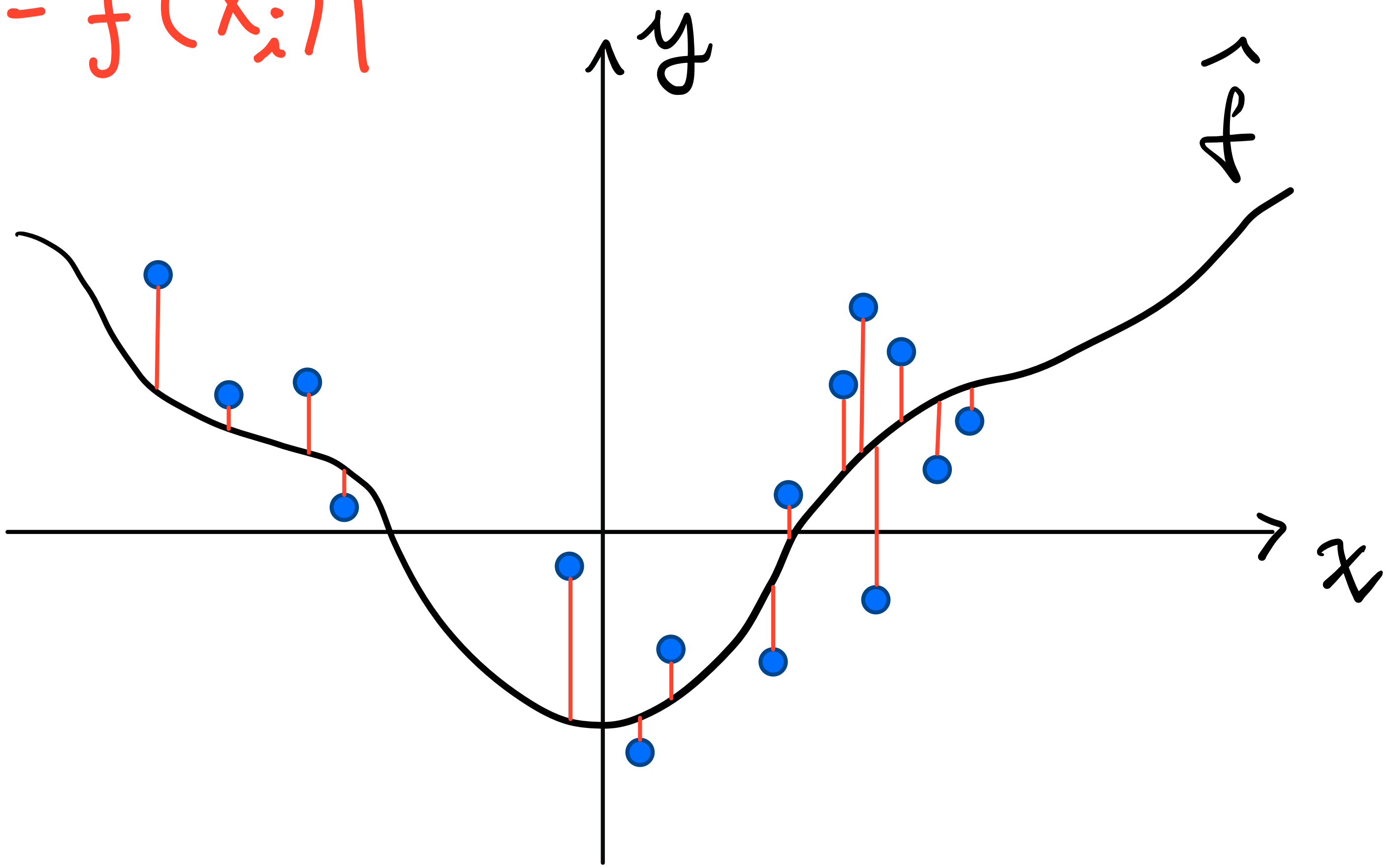
$D_{calib} = \{(X_i, Y_i)\}_{i=1}^{n}$



CALIBRATION

On mésure les scores (erreurs)

$$S_i = |y_i - \hat{f}(x_i)|$$



CALIBRATION

On calcule :

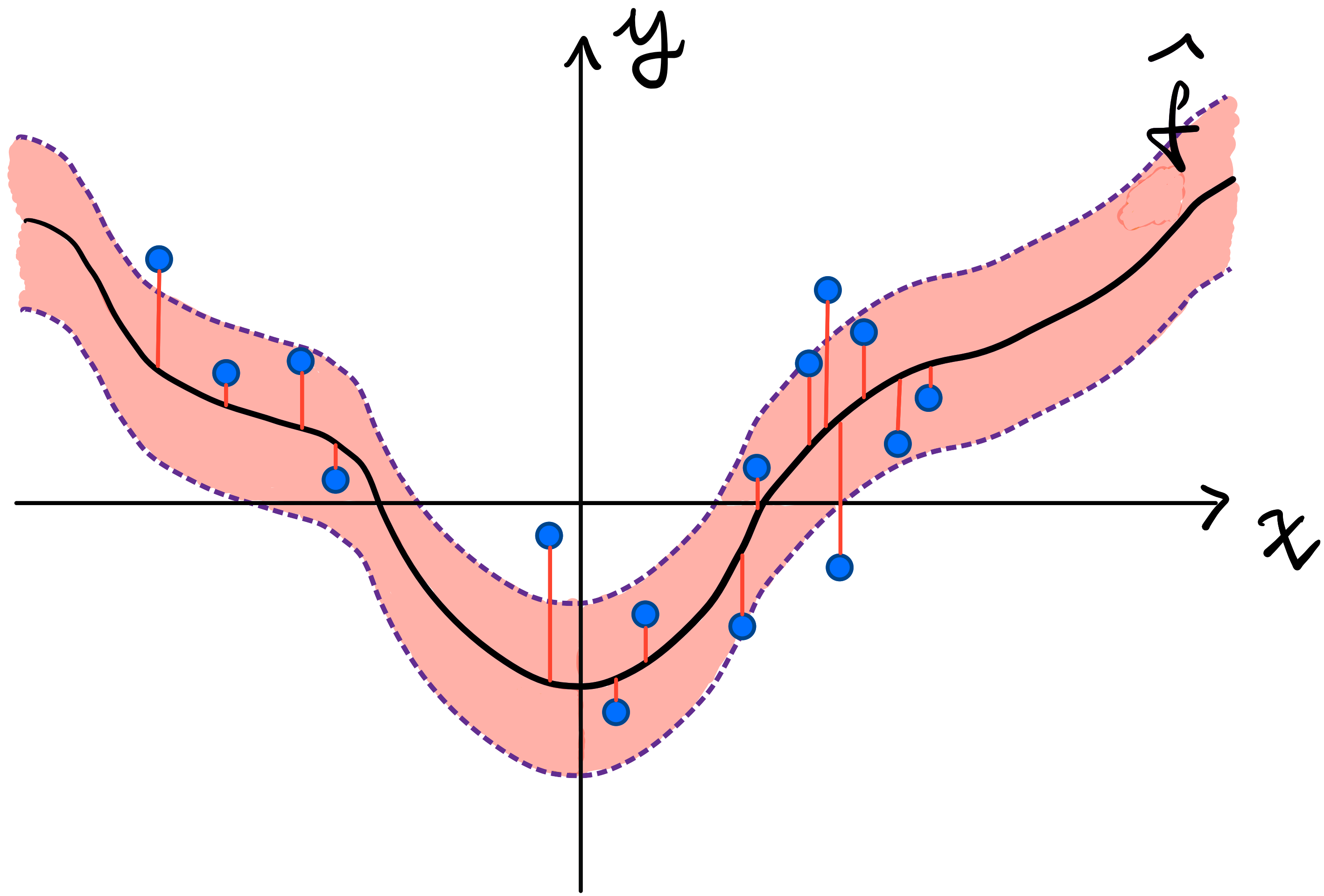
$$\delta_\alpha := \text{le } \frac{\lceil (n+1)(1-\alpha) \rceil}{n}\text{-ème}$$

quantile des scores $S_1, \dots, S_n$

i.e. le $\lceil (n+1)(1-\alpha) \rceil$-ème plus petit score.

CALIBRATION

On prédit : $\hat{C}_\alpha(x) = \left[\hat{f}(x) - \delta_\alpha, \hat{f}(x) + \delta_\alpha\right]$

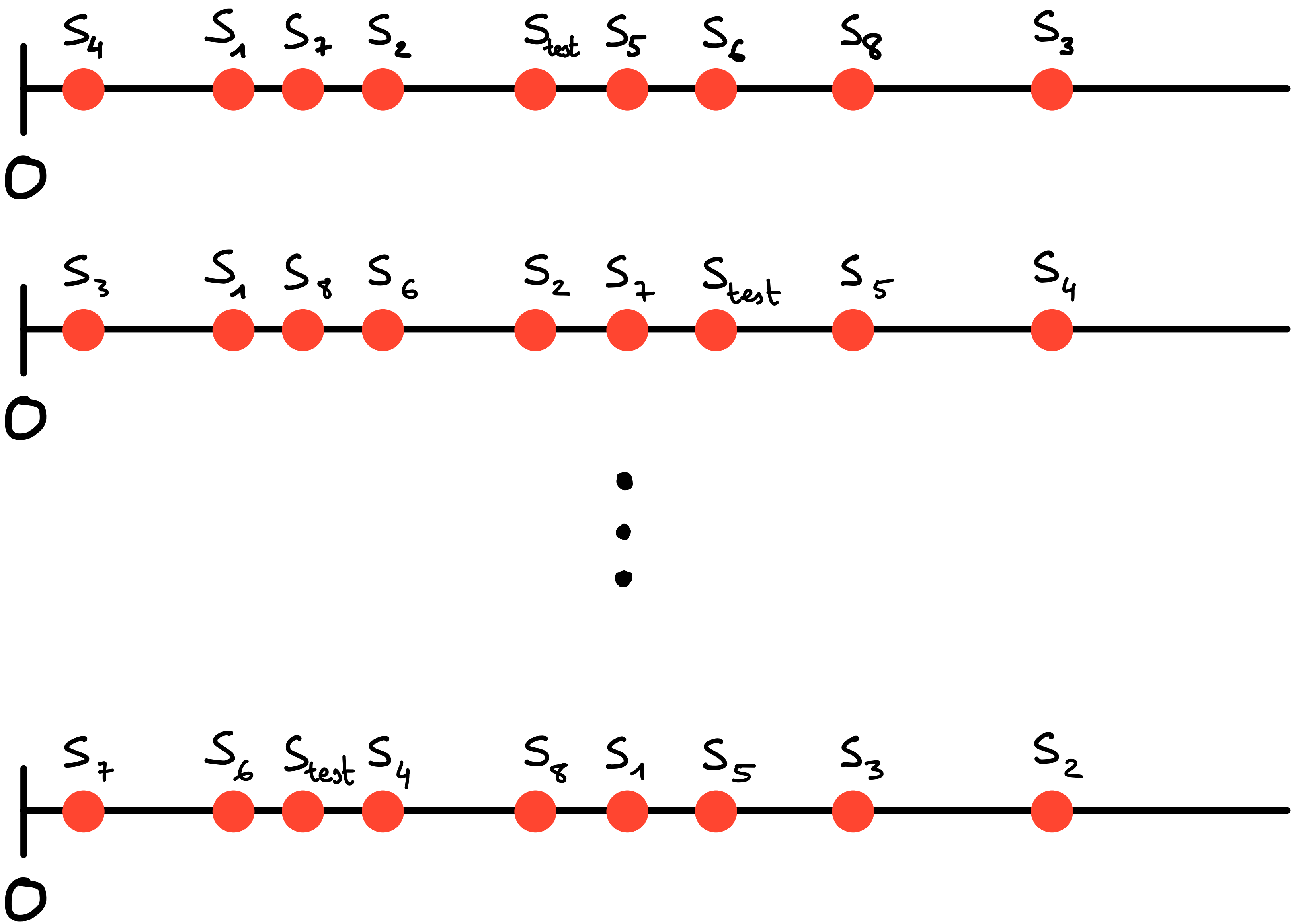


GARANTIE

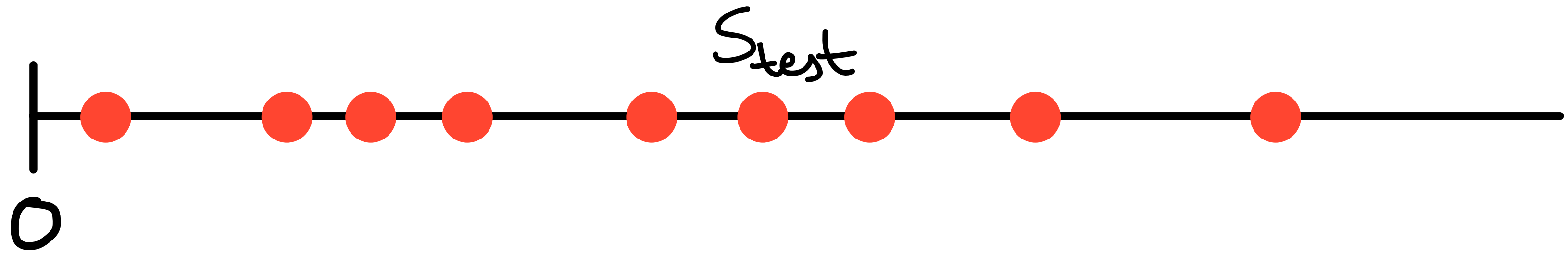
<u>Théorème</u>.– Si $\{(X_i, Y_i)\}_{i=1}^{n+1}$ sont échangeables, alors

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

DÉMONSTRATION



DÉMONSTRATION



$$\mathbb{P}\left(\text{Rang de } S_{test} = k\right)$$

$$= \frac{n!}{(n+1)!} = \frac{1}{n+1}$$

$$\Rightarrow \text{Rang de } S_{test} \sim \mathrm{Unif}\left(\{1, \ldots, n+1\}\right)$$

DÉMONSTRATION

Rang de $S_{test} \sim \text{Unif}\{1, \dots, n+1\}$

$$\Rightarrow \mathbb{P}\left(\text{Rang de } S_{test} \leq K \right) = \frac{K}{n+1}$$

On choisit le plus petit K t.q.

$\frac{K}{n+1} \geq 1-\alpha$ i.e. $K = \lceil (1-\alpha)(n+1) \rceil$

TOO GOOD TO BE TRUE ?

Si $\hat{f}$ mauvais

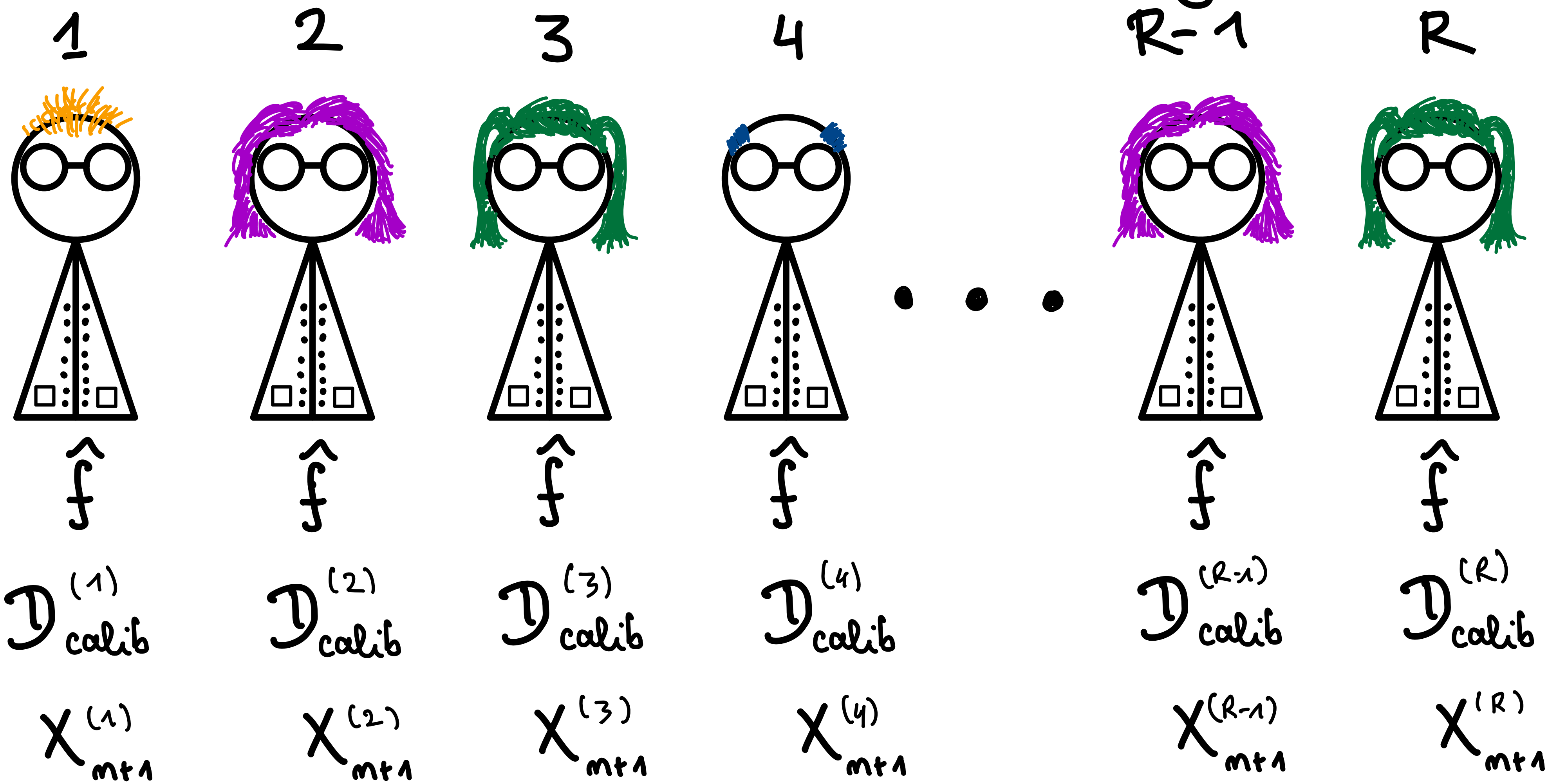
$\Longrightarrow \hat{C}_\alpha$ très grand

AVANTAGES

- Post-hoc
- Distribution-free
- Hypothèses minimales
- Garanties à échantillon fini

LIMITATIONS

Garantie calibration-marginale

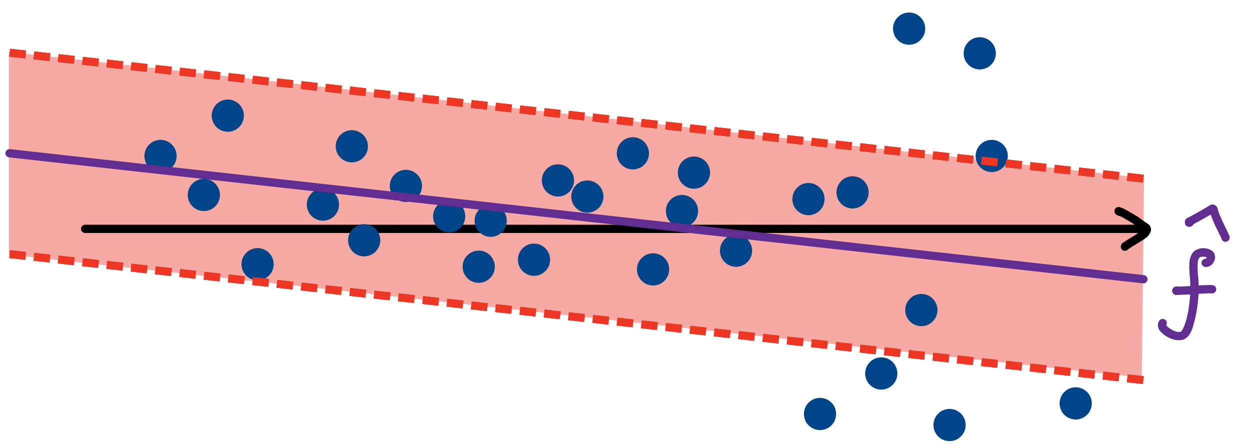


LIMITATIONS

Garantie non-conditionelle

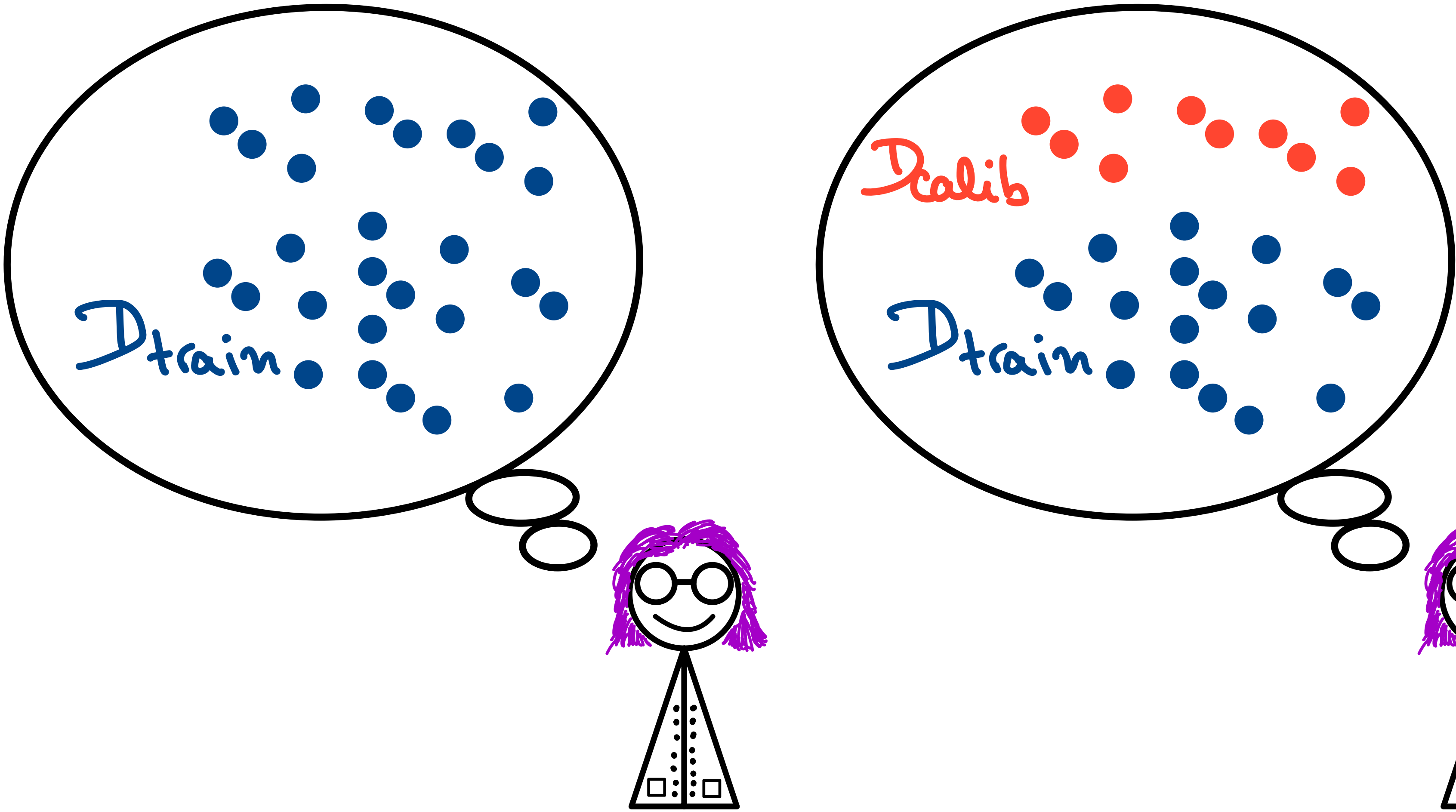
$$\mathbb{P}\left(Y_{n+1} \in \hat{C}_\alpha(X_{n+1})\right) \geq 1-\alpha$$

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



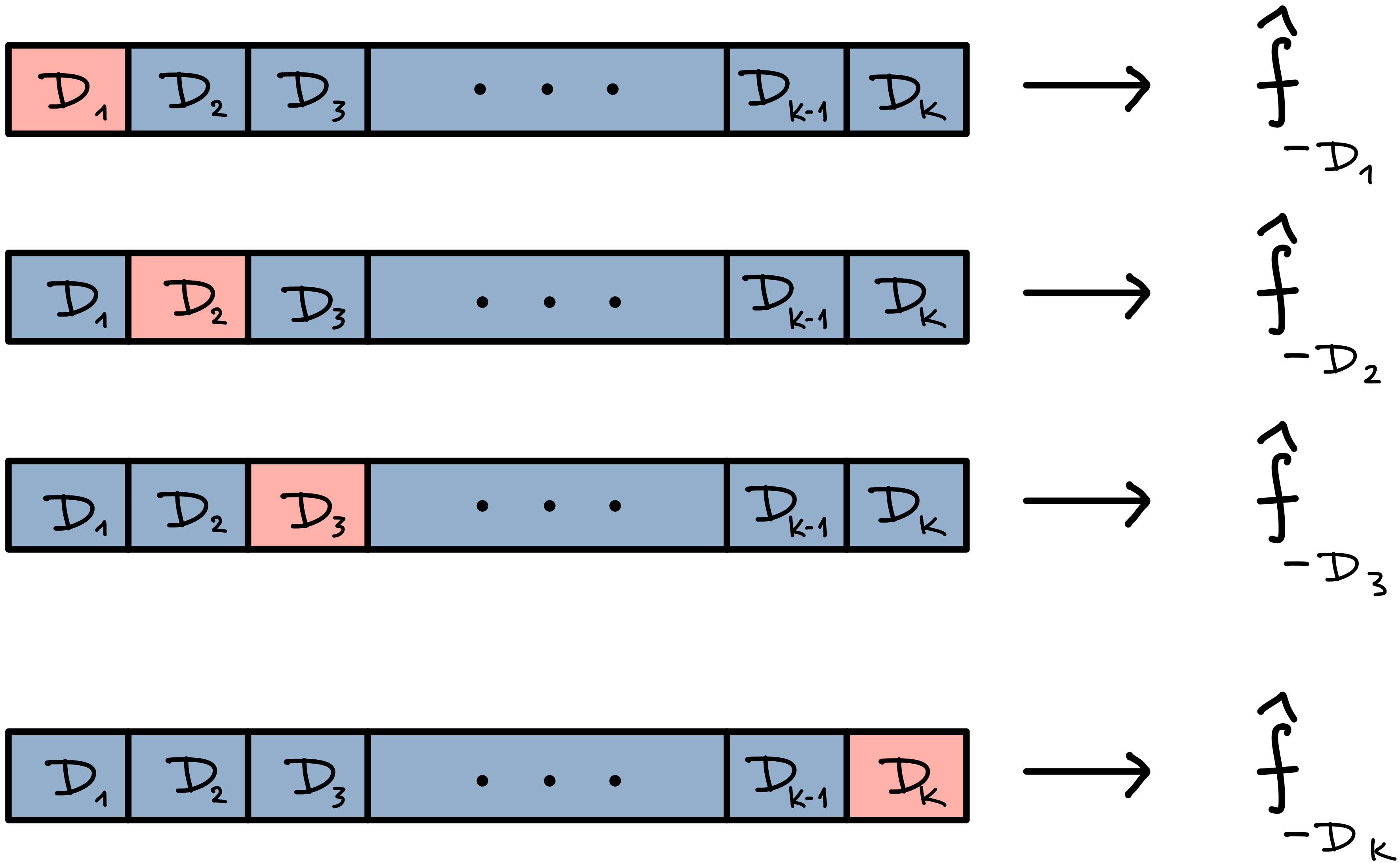
LIMITATIONS

Besoin de données de calibration



JACKNIFE+, CROSS VALIDATION+

Partition des données :



JACKNIFE+, CROSS VALIDATION+

Calibration:

$$S_i^{CV} = \left| Y_i - \hat{f}_{-D_{ind(i)}}(X_i) \right|, \qquad i = 1, \dots, n$$

JACKNIFE+, CROSS VALIDATION+

Inférence :

$\hat{\ell}_\alpha(x) := \lfloor \alpha(n+1) \rfloor$-ème plus petite valeur de

$$\hat{f}_{-S_{ind(1)}}(x) - S_1^{cv}, \dots, \hat{f}_{-S_{ind(n)}}(x) - S_n^{cv}$$

$\hat{u}_\alpha(x) := \lceil (1-\alpha)(n+1) \rceil$-ème plus petite valeur de

$$\hat{f}_{-S_{ind(1)}}(x) + S_1^{cv}, \dots, \hat{f}_{-S_{ind(n)}}(x) + S_n^{cv}$$

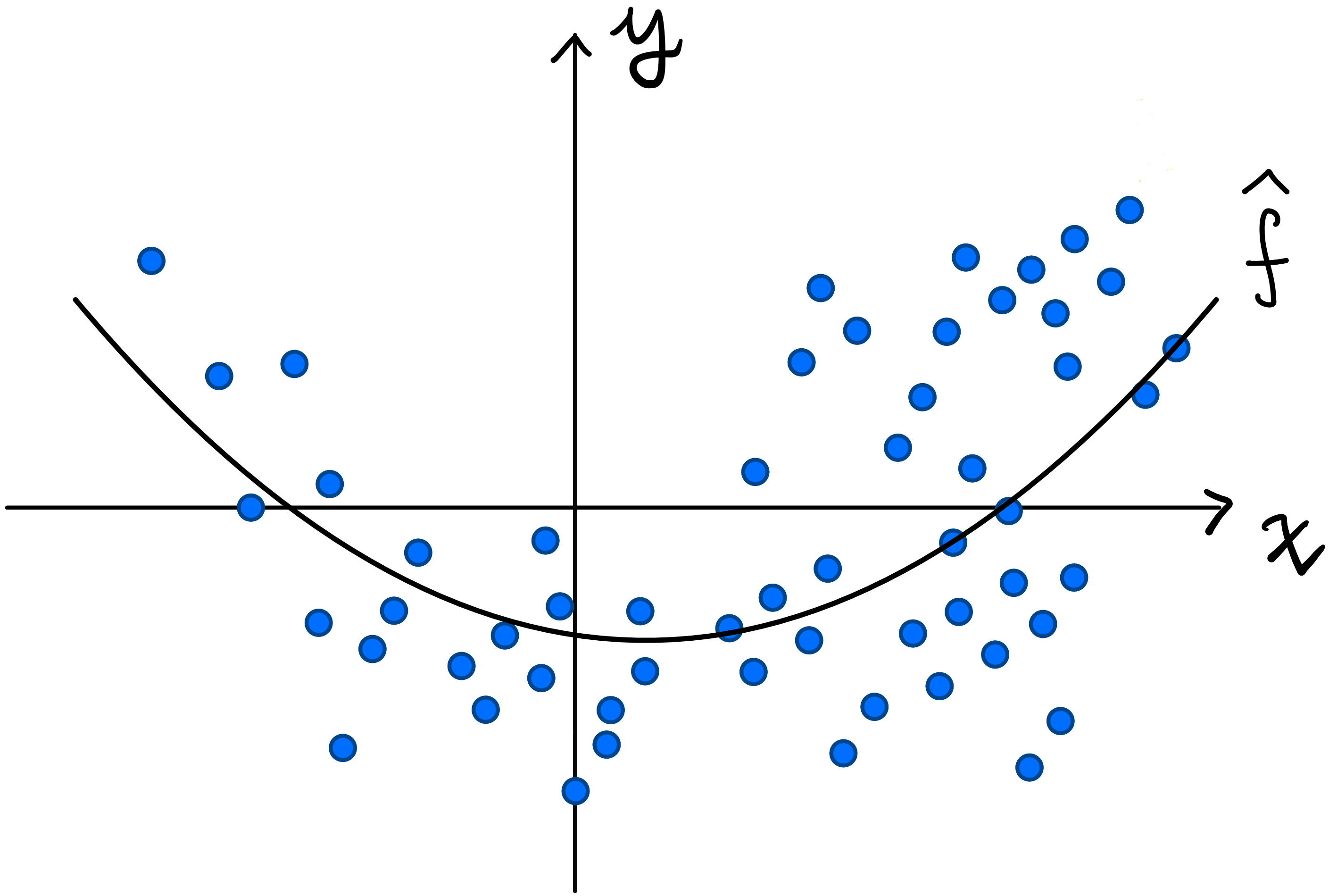
$$\hat{C}_\alpha(x) := \left[\hat{\ell}_\alpha(x), \hat{u}_\alpha(x) \right]$$

JACKKNIFE+, CROSS VALIDATION+

Garantie :

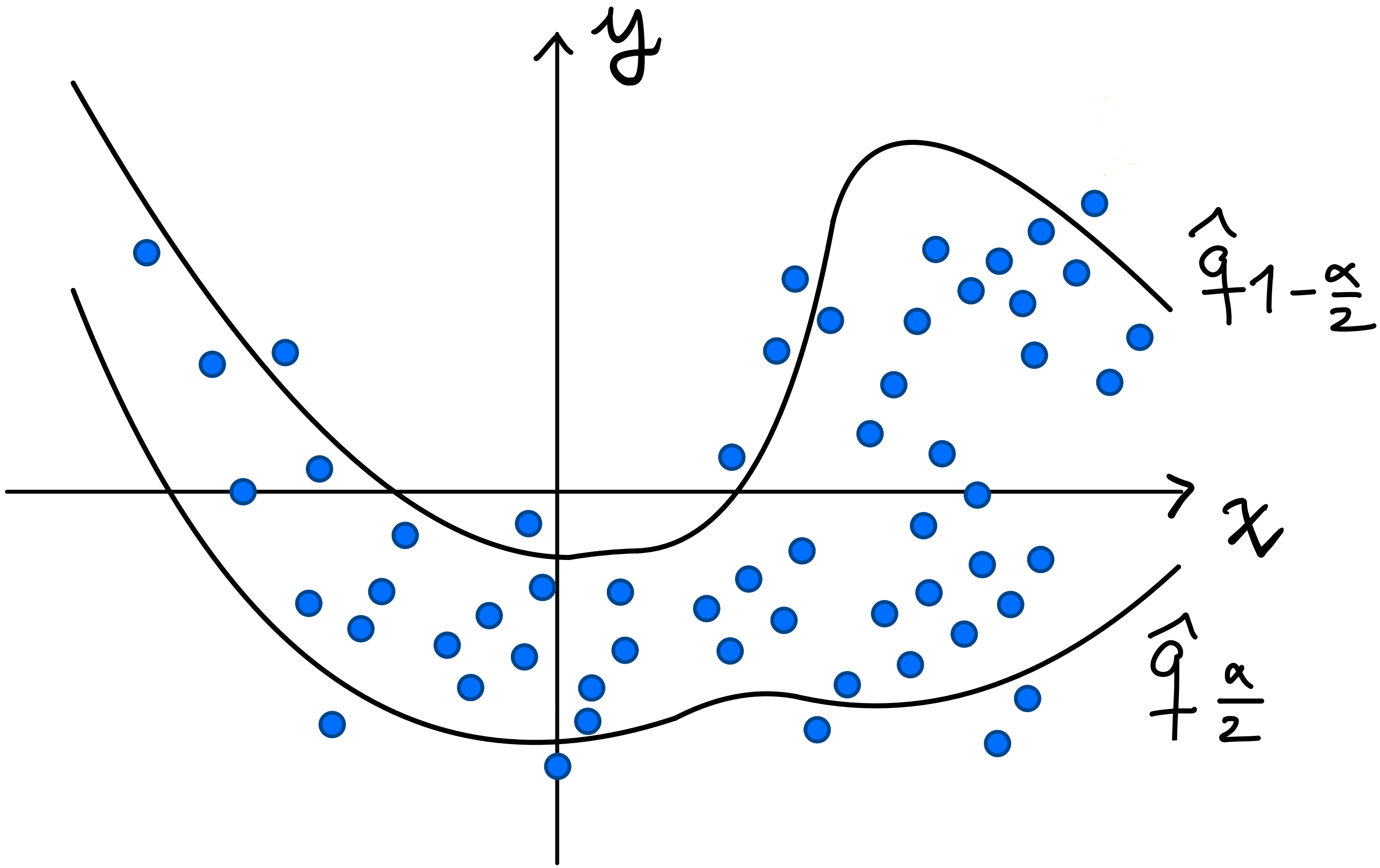
$$\mathbb{P}\left(Y_{n+1} \in \hat{C}_\alpha(X_{n+1})\right) \geq 1 - 2\alpha$$

CQR: Conformal Quantile Regression



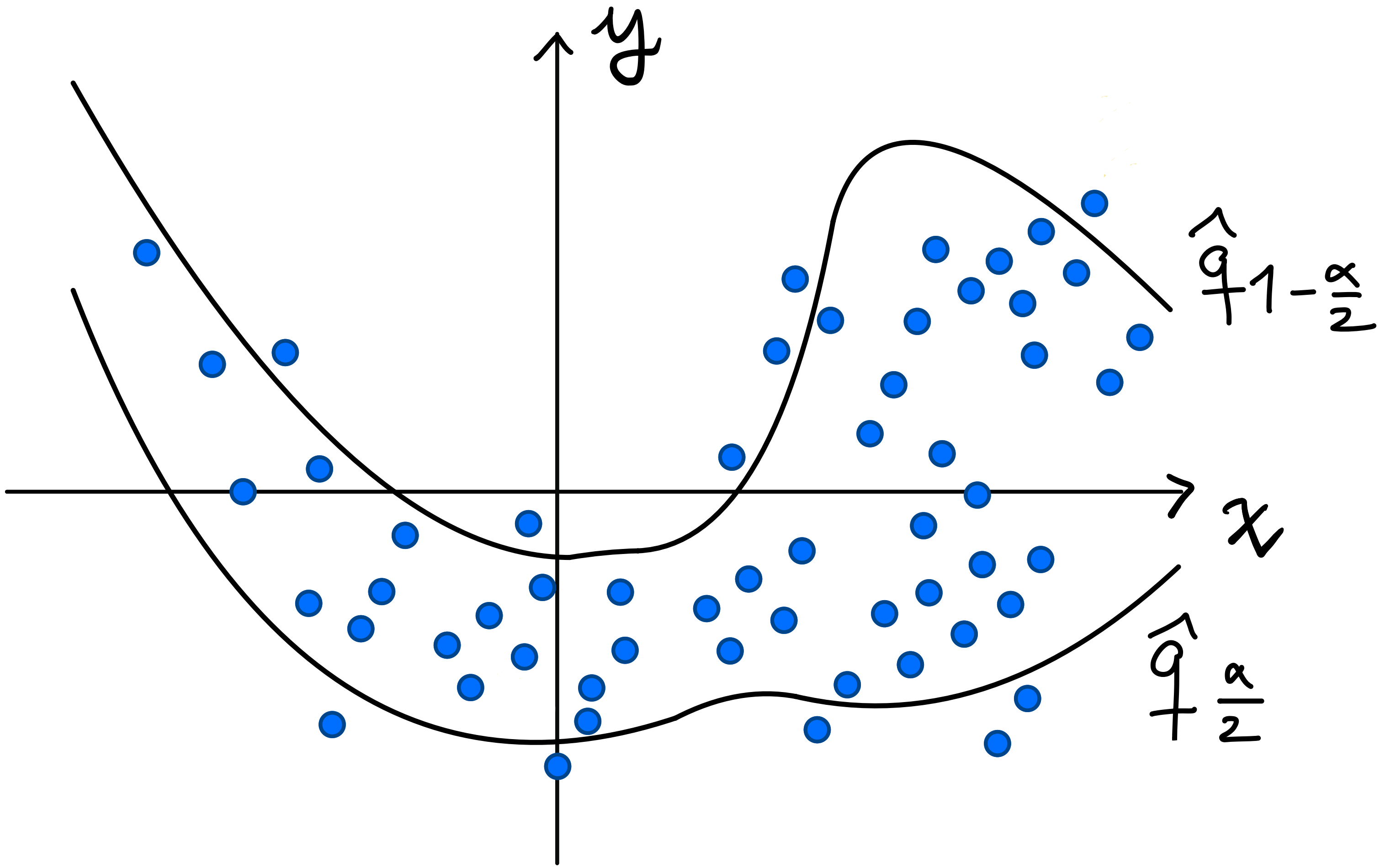
CQR

$$\hat{q}_t(x) \simeq \mathbb{P}(Y \leq t \mid X = x)$$



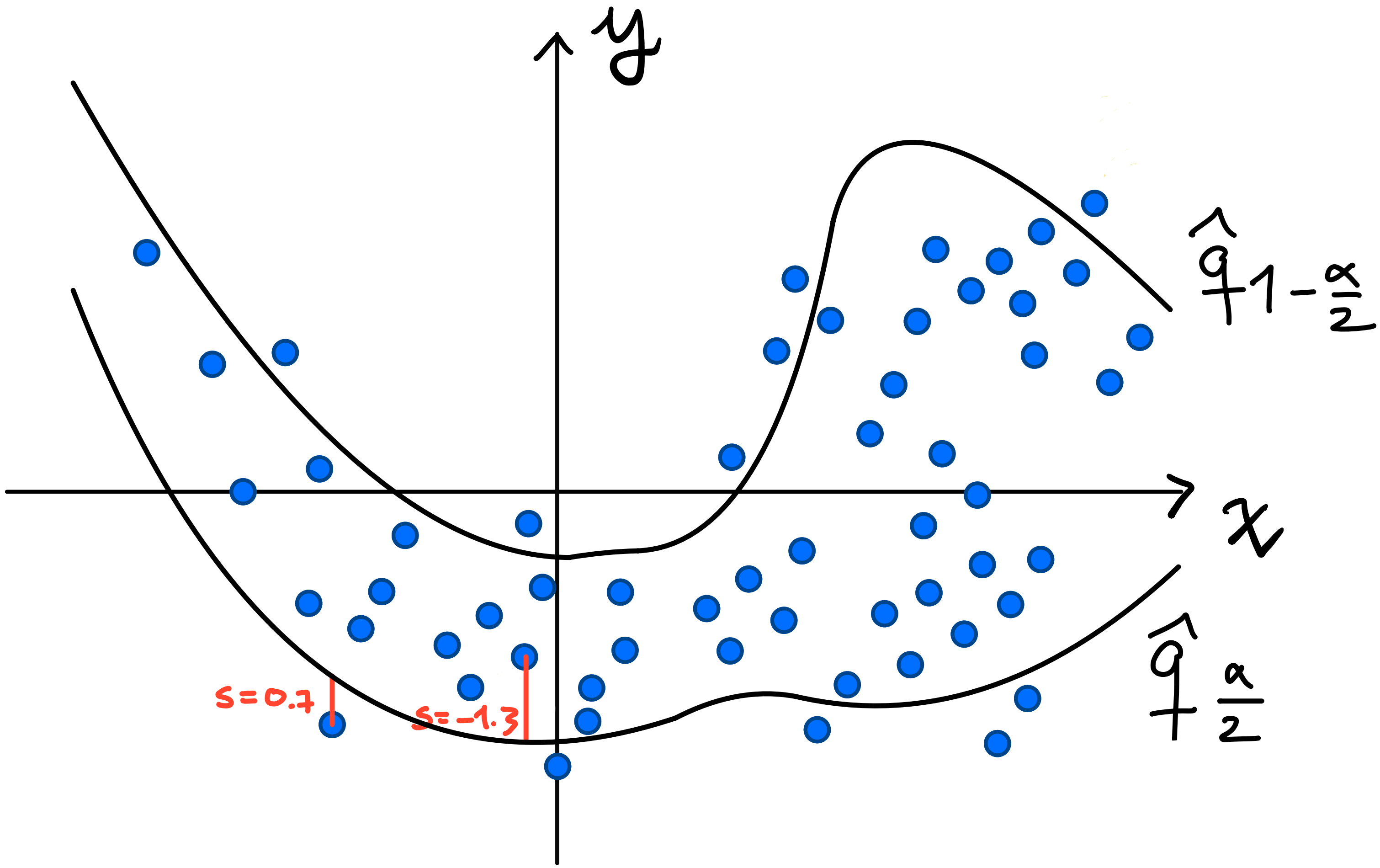
CQR

$$\mathbb{P}\left(Y \in \left[q_{1-\frac{\alpha}{2}}(x), q_{1+\frac{\alpha}{2}}(x)\right]\right) = ?$$



CQR : CALIBRATION

$$S_i = \max\left\{ q_{\frac{\alpha}{2}}(x_i) - y_i \,,\; y_i - q_{1-\frac{\alpha}{2}}(x_i) \right\}$$



CQR : CALIBRATION

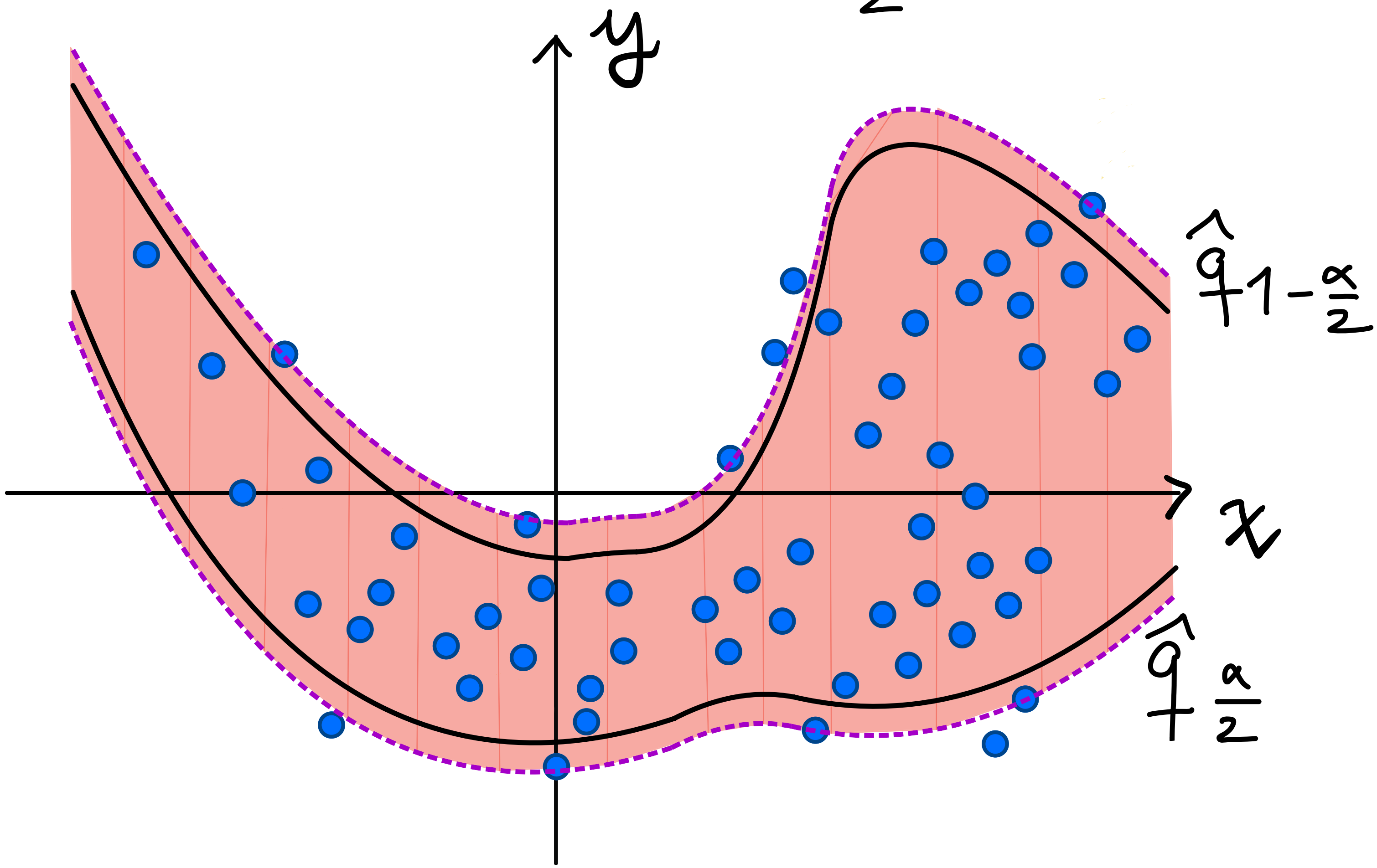
Calibration :

$$\delta_\alpha := \text{le } \left\lceil \frac{(n+1)(1-\alpha)}{n} \right\rceil \text{-ème}$$

quantile des scores $S_1, \ldots, S_n$

CQR : CALIBRATION

$$C_\alpha(x) = \left[\hat{q}_{\frac{\alpha}{2}}(x) - \delta_\alpha \, , \, \hat{q}_{1-\frac{\alpha}{2}}(x) + \delta_\alpha\right]$$



CLASSIFICATION CONFORME

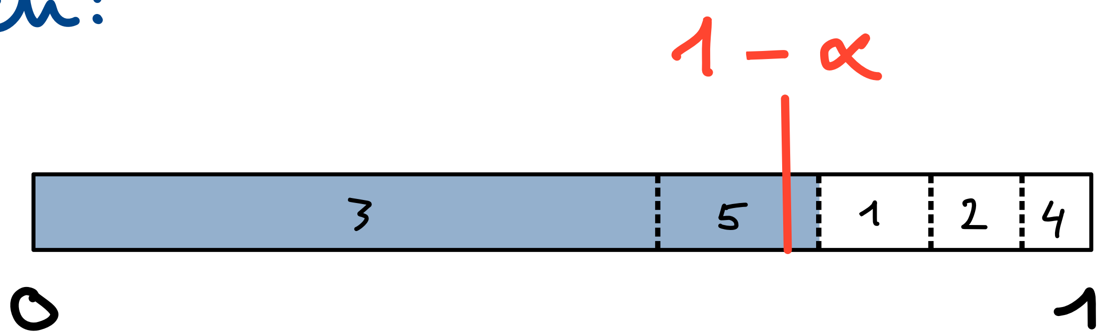
$$\hat{\pi} : \mathcal{X} \longrightarrow \mathcal{P}(\{1, \dots, K\})$$

$$x \longmapsto (\hat{\pi}_1(x), \dots, \hat{\pi}_K(x))$$

Rank : $\hat{\pi}_{(1)}(x) \geq \hat{\pi}_{(2)}(x) \geq \dots \geq \hat{\pi}_{(K)}(x)$

Naïve approach:

$\hat{\pi}(x)$



CLASSIFICATION CONFORME

$$\hat{\pi} : \mathcal{X} \longrightarrow \mathcal{P}(\{1, \dots, K\})$$

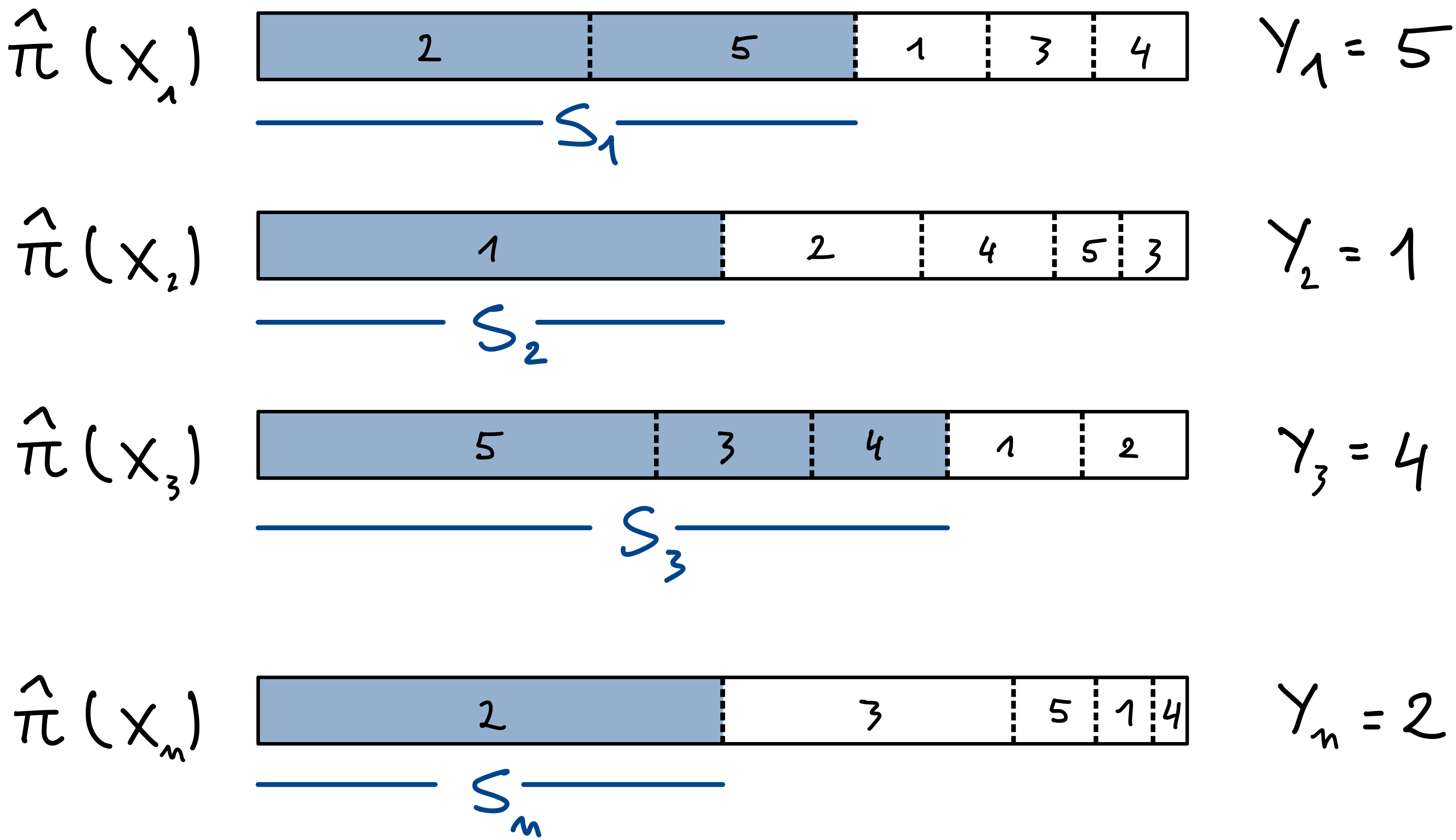
$$x \longmapsto (\hat{\pi}_1(x), \dots, \hat{\pi}_K(x))$$

Rank : $\hat{\pi}_{(1)}(x) \geq \hat{\pi}_{(2)}(x) \geq \dots \geq \hat{\pi}_{(K)}(x)$

$$L(x, \pi, \tau) := \min_{c \in \{1, \dots, K\}} \left\{ \pi_{(1)}(x) + \dots + \pi_{(c)}(x) \geq \tau \right\}$$

Classification Conforme

Calibration



CLASSIFICATION CONFORME

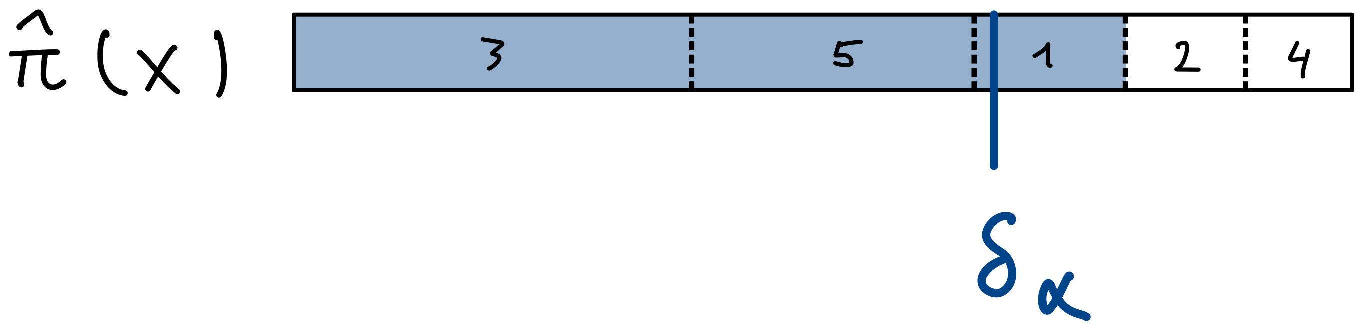
Calibration :

$$\delta_\alpha := \text{le } \left\lceil \frac{(n+1)(1-\alpha)}{n} \right\rceil \text{-ème}$$

quantile des scores $S_1, \ldots, S_n$

CLASSIFICATION CONFORME

Inférence :



$$C_\alpha(x) = \{1, 3, 5\}$$

Garantie :

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