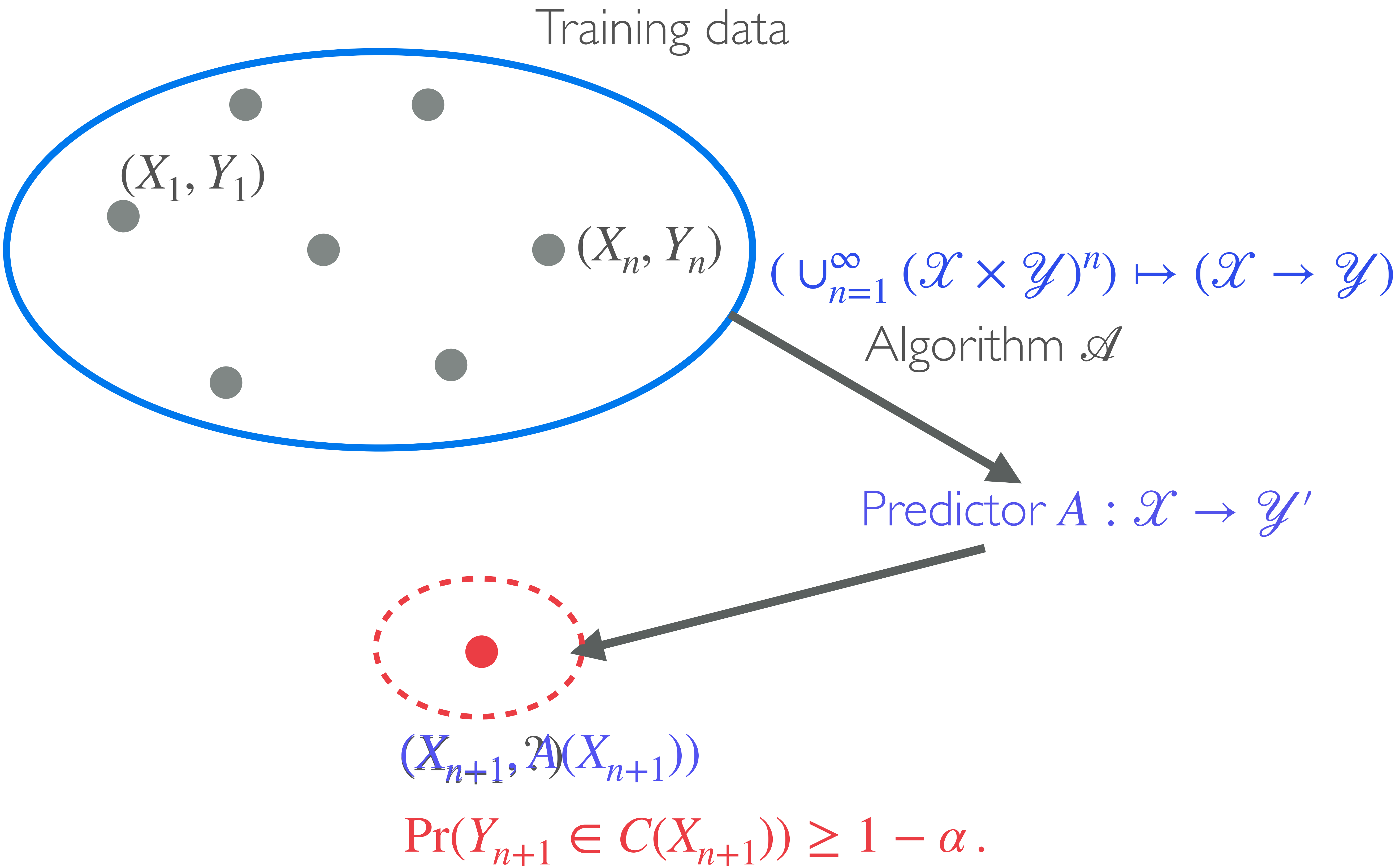
Assumption-free uncertainty quantification for black box models

Aaditya Ramdas

Outline

1. ***What does "predictive inference" mean?***
2. Methods for assumption-free predictive inference
3. Empirical observations
4. Generalizations
5. Open problems + summary

Prediction vs "Predictive Inference"



Why "Predictive Inference"?

Why do we want $\Pr(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha$?

eg: time taken to airport (Y) at some time of day (X)

A "mean" prediction of 80mins is not useful because 80 +/- 5 mins is very different from 80 +/- 40 mins.

To make real-world decisions based on predictions, we need to quantify uncertainty of those predictions.

Prediction interval vs Confidence interval

Predictor (prediction algorithm)

$$(\cup_{n=1}^{\infty} (\mathscr{X} \times \mathscr{Y})^n) \mapsto (\mathscr{X} \to \mathscr{Y}')$$

Prediction interval (or set)

$\Pr(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha\,.$ vol$(C_n) \geq q_{1-\alpha}(P_{Y|X=X_{n+1}})\,.$

Estimator (estimation algorithm)

$$(\cup_{n=1}^{\infty} (\mathscr{X} \times \mathscr{Y})^n) \mapsto \Theta$$

Confidence interval (or set)

$\Pr(\theta \in C_n) \geq 1 - \alpha\,.$ Often, vol$(C_n) \to 0$ as $n \to \infty\,.$

Need to assume a model relating Y to X.
θ is a subset of model parameters.
eg: assume $Y = X\theta +$ noise.

"Assumption-free" Predictive Inference

Given data $D_n \equiv (X_1, Y_1), \ldots, (X_n, Y_n) \sim P_X \times P_{Y|X} \equiv P_{XY}$,

any algorithm $\mathscr{A} : (\cup_{n=1}^{\infty} (\mathscr{X} \times \mathscr{Y})^n) \mapsto (\mathscr{X} \to \mathscr{Y}')$,

and $X_{n+1} \sim P_X$, produce a set $C(X_{n+1}) \equiv C_{\mathscr{A}, D_n}(X_{n+1})$ s.t.

for all P_{XY}, algorithms $\mathscr{A}$, $\Pr(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha$.

Remarks

(a) *don't need iid data, just exchangeable*

(b) *algorithm optional, assumed permutation-invariant*

(c) *trivial without size/efficiency requirement*

(d) *probability is over all training, test data, algorithm*

(e) *not conditional on* X_{n+1}

Vovk, Shafer, and others, 2000-19

Conditional coverage is impossible without assumptions

We will achieve $\Pr(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha$.

Not $\forall x$, $\Pr(Y_{n+1} \in C(X_{n+1}) \mid X_{n+1} = x) \geq 1 - \alpha$.

In fact, one can prove that assumption-free conditional inference is impossible.
(the expected volume of the set must be infinite)

Balasubramanian, Ho and Vovk '14

Lei, Wasserman, Rinaldo, Tibshirani '18

BaCaRaTi'19

Outline

1. What does "predictive inference" mean?

2. Methods for assumption-free predictive inference

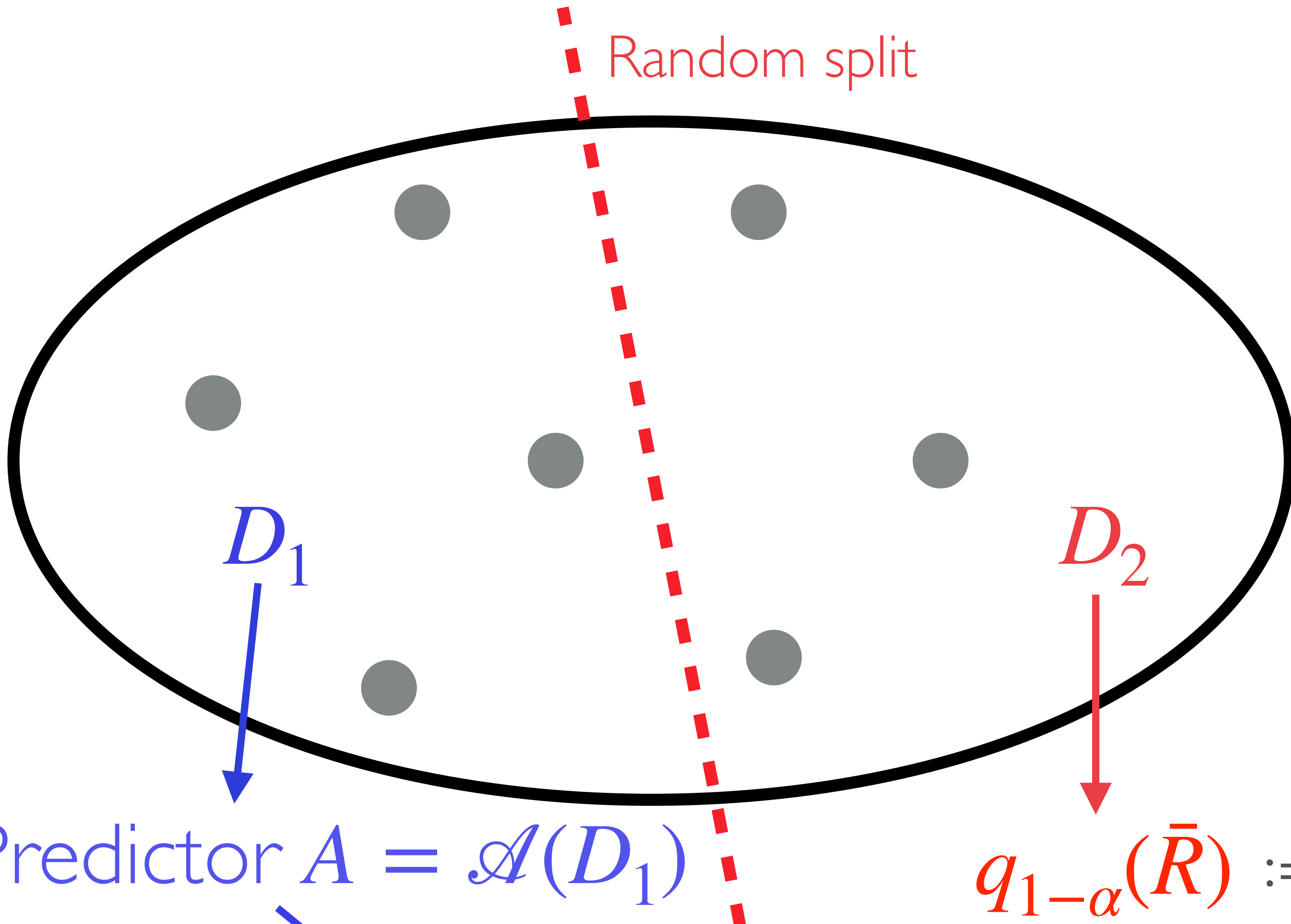
3. Empirical observations

4. Generalizations

5. Open problems + summary

Split/"Inductive" Conformal Prediction

Training data



Hold-out scores $R_i \equiv R_A(X_i, Y_i)$

eg: residual $R_i := |Y_i - A(X_i)|$

Let $\bar{R} = \{R_i\}_{i \in D_2}$

$R_{n+1} = |Y_{n+1} - A(X_{n+1})|$

$R_1, \ldots, R_{n_2}, R_{n+1}$ exchangeable

$\implies$ rank(R_{n+1}) is uniform ★

Predictor $A = \mathscr{A}(D_1)$

$q_{1-\alpha}(\bar{R})$ $:= \lceil (1-\alpha)(n_2+1) \rceil$ ranked element of $\bar{R}$

w.h.p. $R_{n+1} \equiv |Y_{n+1} - A(X_{n+1})| \leq q_{1-\alpha}(\bar{R})$

$\implies$ w.h.p. $Y_{n+1} \in \{y : |y - A(X_{n+1})| \leq q_{1-\alpha}(\bar{R})\}$

$C(X_{n+1}) := [A(X_{n+1}) \pm q_{1-\alpha}(\bar{R})]$ $\equiv \{y : |y - A(X_{n+1})| \leq q_{1-\alpha}(\bar{R})\}$

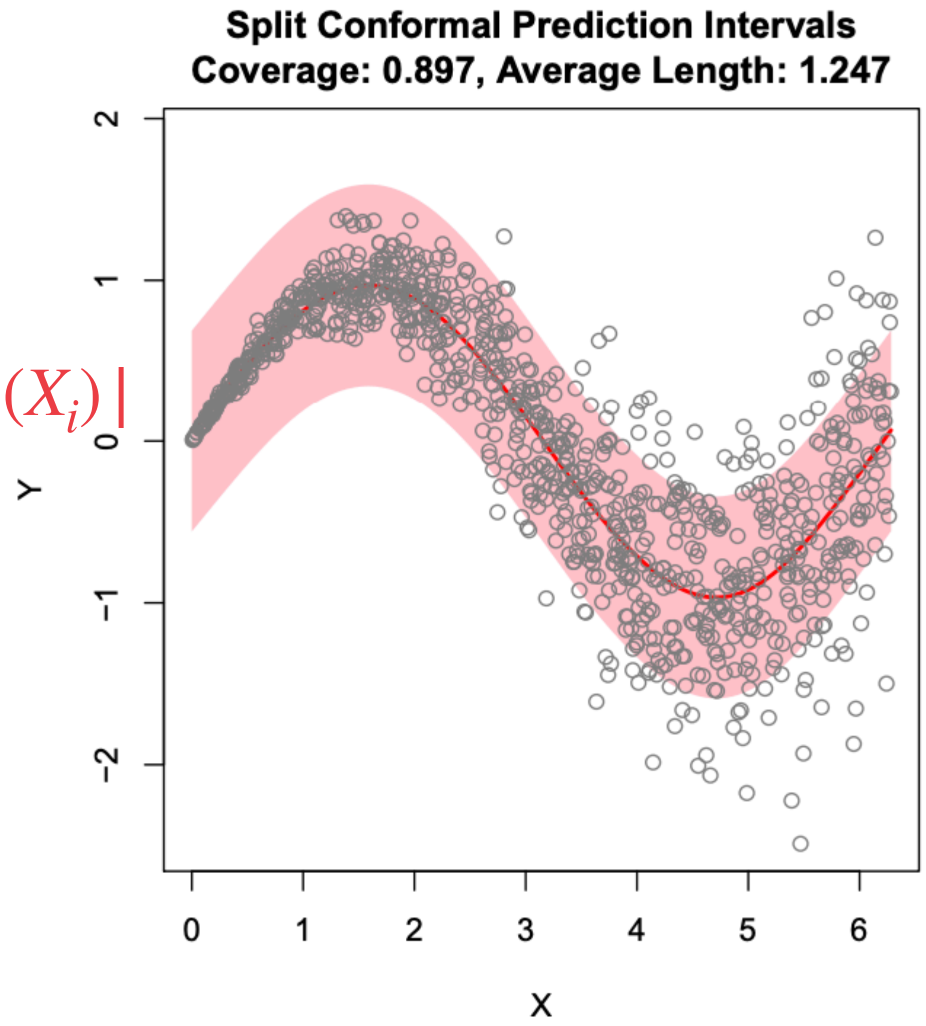
$\Pr(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha$.

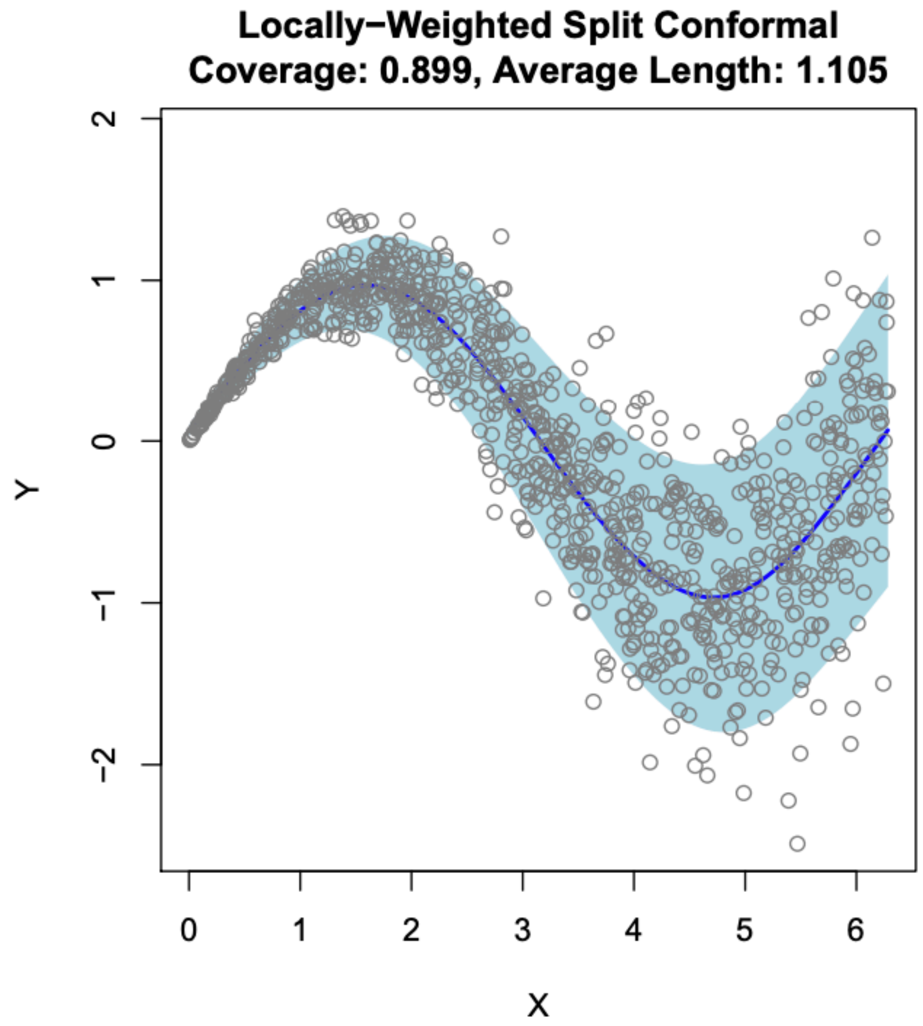
$\leq^* 1 - \alpha + \frac{1}{n_2 + 1}$.

Papadopoulos, Proedrou, Vovk, Gammerman '02
Lei, Wasserman, Rinaldo, Tibshirani '18

LWRT'18

$R_i := |Y_i - A(X_i)|$





$$R_i := \frac{|Y_i - A(X_i)|}{\sigma(X_i)}$$

Classifier $A : \mathscr{X} \to \Delta(\mathscr{Y})$ or $\mathbb{R}^{\mathscr{Y}}$

$$R_i := \sum_{k=1}^{K} [A(X_i)]_k \, \mathbf{1}_{[A(X_i)]_k > [A(X_i)]_{Y_i}}$$

or $R_i :=$ rank(Y_i) in $A(X_i)$