Backward Conformal Prediction

Etienne Gauthier[1] **Francis Bach**[1] **Michael I. Jordan**[1,2]

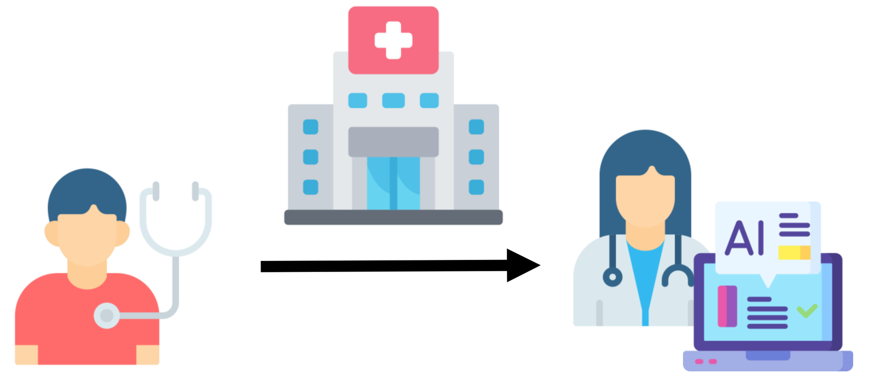
[1] Inria, ENS, PSL University [2] UC Berkeley

Motivation

Conformal Prediction

- **Setup:** score function $S\colon X \times Y \to \mathbb{R}$, calibration set $\{(X_i, Y_i)\}_{i=1}^n$, test point (X_{test}, Y_{test}), user pre-specified miscoverage α.
- **Assumption:** data drawn exchangeably $\sim \mathbb{P}$.
- **Goal:** build a prediction set $\hat{C}_n^\alpha(X_{test})$ achieving marginal coverage:

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



$\alpha = 1\%$ → $|\hat{C}_n^\alpha(X_{test})| = 50$ ✗

$\alpha = 2\%$ → $|\hat{C}_n^\alpha(X_{test})| = 5$ ✓

- **Small values of α** can lead to **uninformative** prediction sets.
- One may instead aim for prediction sets of **maximal allowed size**.

Approach

- **Size constraint rule:** $\mathcal{T}(\{(X_i, Y_i)\}_{i=1}^n, X_{test}) \in \{1, \dots, |Y|\}$.
- **Define:** $\tilde{\alpha} \coloneqq \inf\left\{\alpha \in (0,1) : |\hat{C}_n^\alpha(X_{test})| \leq \mathcal{T}(\{(X_i, Y_i)\}_{i=1}^n, X_{test})\right\}$,

the smallest α such that the prediction set satisfies the size constraint.

Guarantees

- **Size:** $|\hat{C}_n^{\tilde{\alpha}}(X_{test})| \leq \mathcal{T}(\{(X_i, Y_i)\}_{i=1}^n, X_{test})$.
- **Marginal coverage:** $\mathbb{P}\left(Y_{test} \in \hat{C}_n^{\tilde{\alpha}}(X_{test})\right) \gtrsim 1 - \mathbb{E}[\tilde{\alpha}]$.

$\mathbb{E}[\tilde{\alpha}]$ is unknown but it can be estimated using $\{(X_i, Y_i)\}_{i=1}^n$.

Works for all e-value-based prediction sets of the form:

$\hat{C}_n^\alpha(X_{test}) \coloneqq \{y : E_n < 1/\alpha\}$ with $\mathbb{E}[E_n] \leq 1$.

Example: $E_n = \dfrac{(n+1) S(X_{test}, Y_{test})}{\sum_{i=1}^n S(X_i, Y_i) + S(X_{test}, Y_{test})}$.

Conformal Prediction	Backward Conformal Prediction
Input: α	Input: size constraint $\mathcal{T}$
Output: prediction set + size	Output: implied $\tilde{\alpha}$ + prediction set
Control coverage → size follows	**Control size → coverage follows**
Guarantee: coverage exact, size unknown	**Guarantee:** size exact, coverage estimated

Coverage Estimation

Strategy: Leave-One-Out

(X_1, Y_1) (X_2, Y_2) (X_3, Y_3) ... (X_n, Y_n) ⟶ $\tilde{\alpha}_1$

(X_1, Y_1) (X_2, Y_2) (X_3, Y_3) ... (X_n, Y_n) ⟶ $\tilde{\alpha}_2$

(X_1, Y_1) (X_2, Y_2) (X_3, Y_3) ... (X_n, Y_n) ⟶ $\tilde{\alpha}_3$

(X_1, Y_1) (X_2, Y_2) (X_3, Y_3) ... (X_n, Y_n) ⟶ $\tilde{\alpha}_n$

$$\hat{\alpha}^{LOO} \coloneqq \frac{1}{n}\sum_{i=1}^n \tilde{\alpha}_i$$

Intuition: mimics test-time behavior by treating each calibration point as unseen.

Theoretical Properties

Under mild regularity assumptions:

$$|\hat{\alpha}^{LOO} - \mathbb{E}[\tilde{\alpha}]| = O_P(n^{-1/2}),$$

$$\mathrm{Var}(\hat{\alpha}^{LOO}) = O_P(n^{-1}).$$

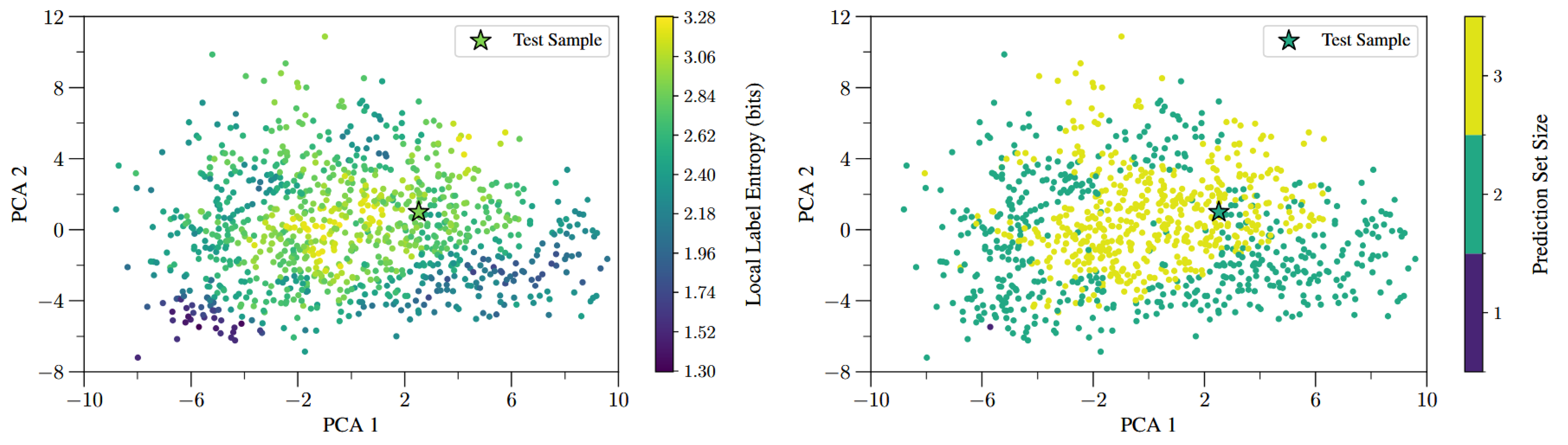
Experiments

Setup

- **Dataset:** CIFAR-10 (50K training / 10K test images, 10 classes).
- **Score:** cross-entropy $S(x,y) = -\log p_f(y|x)$, for pretrained model f.

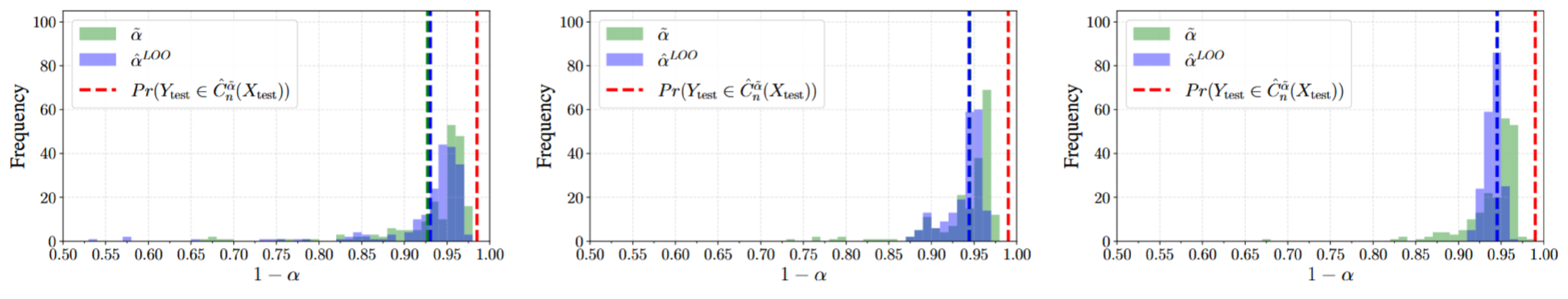
Size constraint rule

- Feature extraction + PCA + k-NN based local entropy + binning.
- **Intuition:** features associated with high label variability, as estimated on the calibration set, correspond to greater uncertainty and should be assigned larger prediction sets.



Local entropy of test point: 2.648. Prediction set size for the test sample: 2.

Estimator performance



$n = 100$ $n = 1000$ $n = 5000$

Takeaway: a principled alternative framework for decision-making under resource constraints.

PAPER CODE