Non-Exchangeable Conformal Risk Control

António Farinhas[1,2], Chrysoula Zerva[1,2], Dennis Ulmer[3,4], André F. T. Martins[1,2,5]

[1]Instituto de Telecomunicações, [2]Instituto Superior Técnico, Universidade de Lisboa (Lisbon ELLIS Unit),
[3]IT University of Copenhagen, [4]Pioneer Centre for Artificial Intelligence, [5]Unbabel

Conformal prediction has sparked a lot of interest but:
- **What if coverage is not your main concern?**
- **What if the data is not i.i.d.?**

Motivation

Conformal prediction provides prediction sets/intervals that are guaranteed to cover the ground truth in expectation:

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

Limitations:

- The data is assumed to be exchangeable;
- The adequate notion of error control may not be coverage.

Barber et al. (2024) and Angelopoulos et al. (2024) offer solutions to these problems separately.

We extend these lines of work and propose **non-exchangeable conformal risk control**.

Multilabel classification in a time series (synthetic)

We sample N=2000 datapoints $(X_i, Y_i) \in \mathbb{R}^M \times \mathbb{R}^M$, where $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_M)$ and $Y_i \sim \mathbf{sign}(\mathbf{W}X_i - \mathbf{0.5} + .1\mathcal{N}(\mathbf{0}, \mathbf{I}_M))$

1. exchageable data: $\boldsymbol{W} = \mathbf{I}_M$
2. changepoints: we start with $\mathbf{W}^{(0)} = \mathbf{I}_M$ and for every change point $k > 0$ we rotate the coefficients s.t. $\mathbf{W}^{(k)}_{i,j} = \mathbf{W}^{(k-1)}_{i-1,j}$ for $i > j$ and $\mathbf{W}^{(k)}_{1,j} = \mathbf{W}^{(k-1)}_{M,j}$
3. distribution drift: we start with $\mathbf{W}^{(0)} = \mathbf{I}_M$ and set $\mathbf{W}^{(N)}$ to the last matrix of (2); we compute each $\mathbf{W}^{(k)}$ by linearly interpolating between $\mathbf{W}^{(0)}$ and $\mathbf{W}^{(N)}$

- We fit M independent logistic regression models
- Prediction sets of the form $\mathcal{C}_\lambda(X_i) := \{m \in [M] \;:\; f_m(X_i) \geq 1-\lambda\}$
- We use weights $w_i = 0.99^{n+1-i}$
- We minimize the false negative rate $L(\lambda; (X_i, Y_i)) = 1 - \frac{|Y_i \cap \mathcal{C}_\lambda(X_i)|}{|Y_i|}$

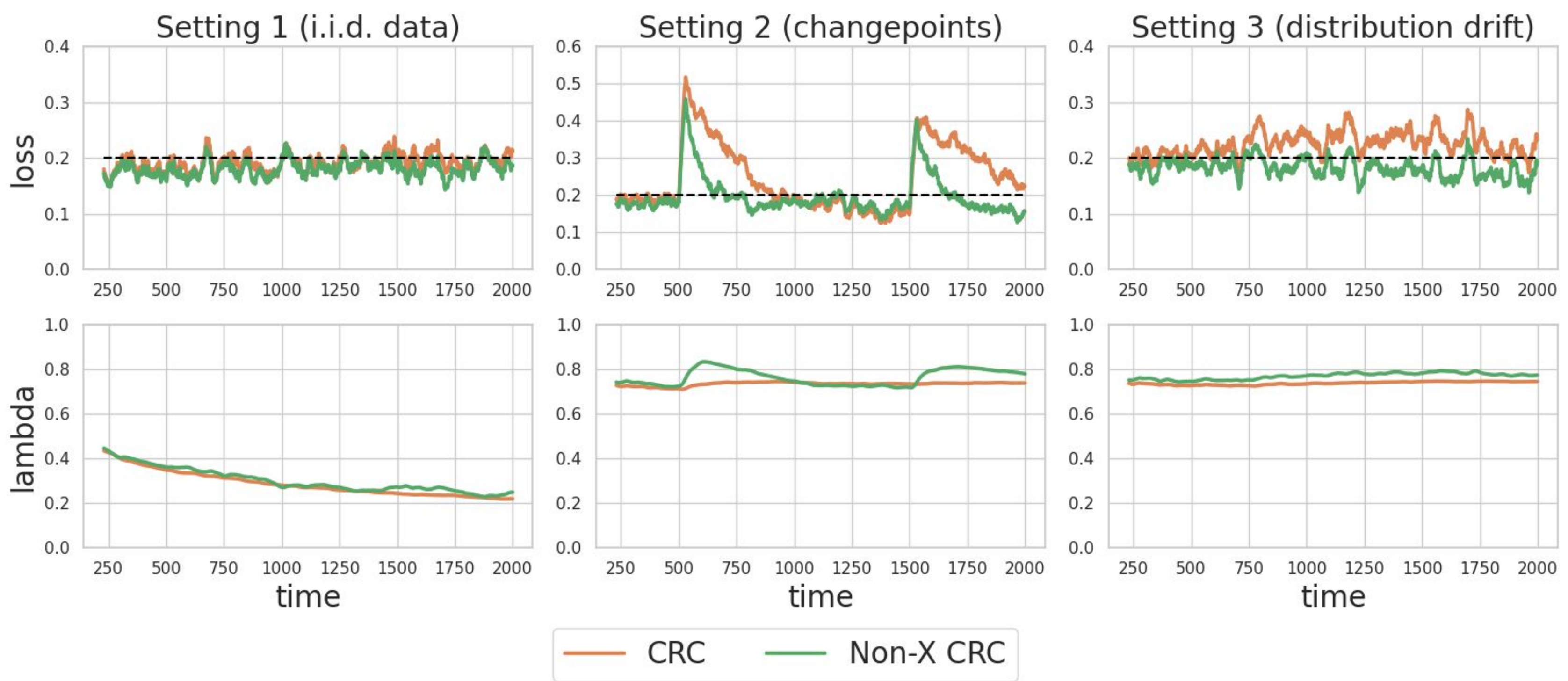


Table: Mean/median for settings (1), (2), and (3)

Method	Setting 1 (i.i.d. data)	Setting 2 (changepoints)	Setting 3 (distribution drift)
CRC	0.191 / 0.183	0.246 / 0.228	0.225 / 0.218
non-X CRC	**0.181 / 0.175**	**0.196 / 0.183**	**0.182 / 0.175**

Our proposal

We define prediction sets $\mathcal{C}_\lambda(\cdot)$, where λ is a parameter such that $\lambda \leq \lambda' \implies \mathcal{C}_\lambda(\cdot) \subseteq \mathcal{C}_{\lambda'}(\cdot)$, and provide guarantees of the form:

$$\mathbb{E}[L(\hat{\lambda}; (X_{n+1}, Y_{n+1}))] \leq \alpha + (B-A)\textstyle\sum_{i=1}^n \tilde{w}_i d_{\text{TV}}(Z, Z^i)$$

Assumptions:

- Monotonically nonincreasing loss wrt to λ (shrinks as the prediction set grows).

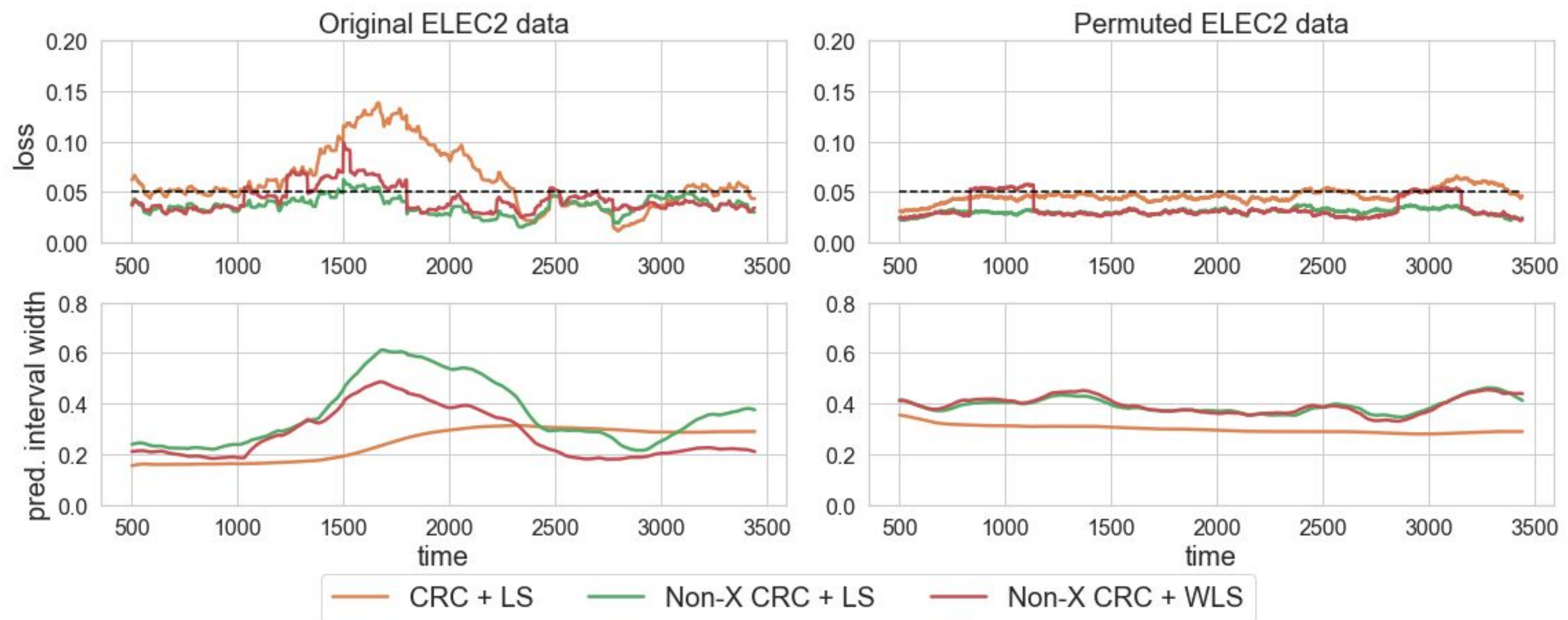
This is accomplished by using

$$\hat{\lambda} = \inf\left\{\lambda : \frac{N_w}{N_w+1}\hat{R}_n(\lambda) + \frac{B}{N_w+1} \leq \alpha\right\}, \quad \hat{R}_n(\lambda) = \frac{1}{N_w}\sum_{i=1}^n w_i L(\lambda; (x_i, y_i))$$

where $N_w := \sum_{i=1}^N w_i$.

Monitoring electricity usage

- We fit a least squares regression model at each time step
- We use weights $w_i = 0.99^{n+1-i}$
- Prediction sets of the form $\mathcal{C}_\lambda(x_i) = [f(x_i) - \lambda, f(x_i) + \lambda]$
- We minimize $L(\lambda; (x_i, y_i)) = \begin{cases} 0, & \text{if } |f(x_i) - y_i| \leq \lambda, \\ |f(x_i) - y_i| - \lambda, & \text{otherwise.} \end{cases}$

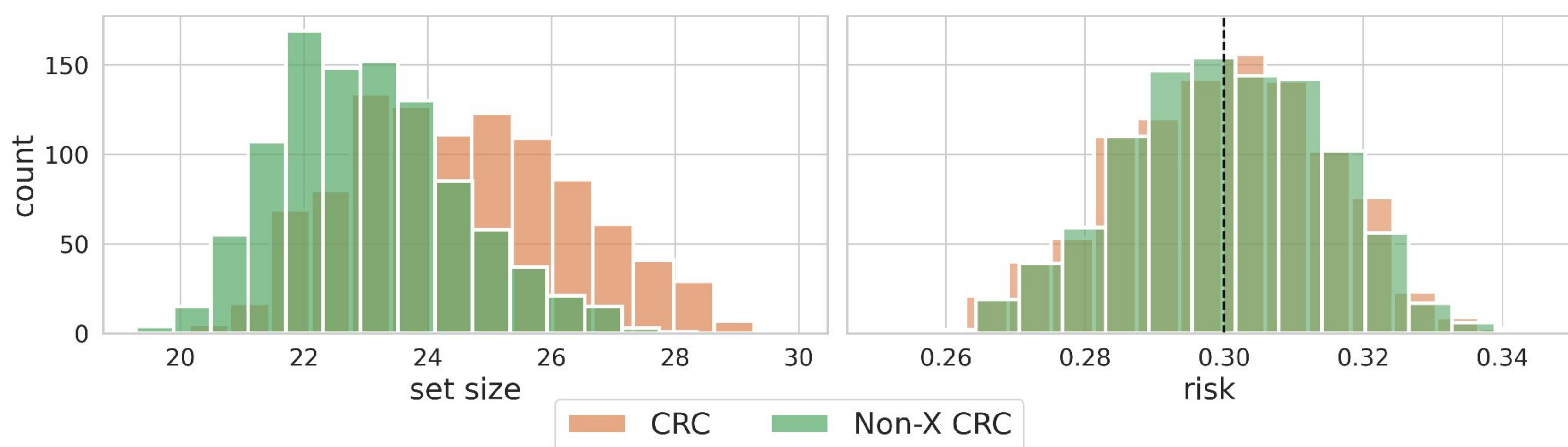


Open-domain question answering

- Two stages: retriever model + reader model
- We calibrate the best token-based F1-score

$$L(\lambda; (X_i, Y_i)) = 1 - \max\{F_1(a, c) : c \in \mathcal{C}_\lambda(X_i), a \in Y_i\},$$
$$\mathcal{C}_\lambda = \{y : f(X_i, y) \geq \lambda\}$$

- We use weights based on sentence similarity, relaxing the assumption of data-independent weights



Acknowledgments: This work was supported by EU's Horizon Europe Research and Innovation Actions (UTTER, contract 101070631), by DECOLLAGE (ERC-2022-CoG 101088763), by the Portuguese Recovery and Resilience Plan (project C645008882-00000055; Center for Responsible AI), and by Fundação para a Ciência e Tecnologia (contract UIDB/50008/2020).