Summary of *Adaptive Conformal Predictions for Time Series*

Margaux Zaffran[*] Olivier Féron Yannig Goude
Julie Josse Aymeric Dieuleveut

39th International Conference on Machine Learning

Théo Remlinger @theo.remlinger

Context

In many areas such as healthcare or energy, the lack of uncertainty quantification of predictive models is a major barrier to the adoption of powerful machine learning methods. The emergent field of conformal prediction (CP, Vovk et al., 2005) is a promising framework for distribution-free uncertainty quantification. It is a general procedure to build predictive intervals for any (black box) predictive model, which are *valid* (i.e. achieve nominal marginal coverage), in finite sample, and without any distributional assumptions except that the data are exchangeable. The goal is to build a predictive interval $\mathcal{C}_\alpha$ such that: $\mathbb{P}\left\{Y_{n+1} \in \mathcal{C}_\alpha\left(X_{n+1}\right)\right\} \geq 1-\alpha$.

To achieve this, Split CP (SCP, Papadopoulos et al., 2002) first splits the n points of the training set in two sets $\mathrm{Tr}, \mathrm{Cal} \subset [1, \mathrm{n}]$ to create a *proper training set*, Tr, and a *calibration set*, Cal. On the proper training set a regression model $\hat{\mu}$ (chosen by the user) is fitted, and then used to predict on the calibration set. A *conformity score* is applied to assess the conformity between the calibration's response values and the predicted values, giving $S_{\mathrm{Cal}} = \{(s_i)_{i \in \mathrm{Cal}}\}$. In regression, usually the absolute value of the residuals is used, i.e. $|\hat{\mu}(x_i) - y_i|$. Finally, a corrected $(1-\hat{\alpha})$-th quantile of these scores $\widehat{Q}_{1-\hat{\alpha}}(S_{\mathrm{Cal}})$ is computed to define the size of the interval. In its simplest form, it is centered on the predicted value: $\mathcal{C}_\alpha\left(x_{n+1}\right) = \widehat{C}_{\hat{\alpha}}(x_{n+1}) := [\hat{\mu}(x_{n+1}) \pm \widehat{Q}_{1-\hat{\alpha}}(S_{\mathrm{Cal}})]$.

Given the non-exchangeability of time series data, SCP can not be applied as such to forecasting tasks. To achieve this, we study and extend Adaptive Conformal Inference (ACI, Gibbs and Candès, 2021) in the context of time series with general dependency. ACI is a method designed to handle an online setting with distributional shift. ACI relies on using an adaptive miscoverage rate α_t, that is updated according to previous performances and to a learning rate $\gamma \geq 0$. Concretely, at each time step t where a prediction is given, $\hat{\alpha} := \alpha_t$ and $\alpha_{t+1} = \alpha_t + \gamma(\alpha - \mathbb{1}\{y_t \notin \hat{C}_{\alpha_t}(x_t)\})$: if ACI does not cover at time t, then $\alpha_{t+1} \leq \alpha_t$, thus $\widehat{Q}_{1-\alpha_{t+1}} \geq \widehat{Q}_{1-\alpha_t}$, and the size of the predictive interval increases; conversely when it covers. Unlike SCP, ACI is asymptotically valid, regardless of the data distribution.

*margaux.zaffran@inria.fr

❶ Theory

First, we study theoretically, using Markov Chain theory, the impact of γ on the length of the predictive intervals, in order to describe not only the *validity* but also the *efficiency* of ACI. Moreover, ACI is usually applied without knowing the type of data it will encounter. If the scores are actually exchangeable, ACI's *validity* would not improve upon SCP (known to be *valid*), thus assessing ACI's impact on *efficiency* is necessary. Thereby, we first provide an analysis in the exchangeable case, then in the auto-regressive one (time series).

Theorem 1 (informal)**.** *If the scores are exchangeable and the calibration is perfect, then the average length of ACI's intervals worsen linearly with γ with respect to classical SCP.*

Theorem 2 (informal)**.** *If the residuals are auto-regressive of coefficient φ (the higher the more important the temporal dependence) and the calibration is perfect, then there exists an optimal $\gamma^* > 0$ minimizing the average length for high φ.*

These results stress that choosing γ is crucial but difficult.

❷ Algorithm

Second, we design AgACI, a parameter-free method using online expert aggregation (Cesa-Bianchi and Lugosi, 2006). Based on the pinball loss of level $1 - \frac{\alpha}{2}$ (resp. $\frac{\alpha}{2}$), AgACI assigns weights to each expert (an expert is a version of ACI with some γ) depending on their previous performances in order to output a unique upper bound (resp. lower bound) which is the weighted mean of the experts upper (resp. lower) bounds.

❸ Numerical experiments

Third, we compare ACI with various γ, AgACI and benchmark methods, on extensive synthetic experiments of increasing temporal dependence and on the task on forecasting French electricity prices.

These experiments highlight that:

- Benchmark methods are not robust to the increase of the temporal dependence;
- ACI is robust to this increase, maintaining validity in all settings with a well-chosen γ;
- AgACI is robust to this increase without choosing γ, at the cost of not being the smallest.

Discover more in the paper!

Paper ⟶
5' talk ⟶
45' talk ⟶
Code ⟶

References

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games.* Cambridge University Press.

Gibbs, I. and Candès, E. (2021). Adaptive Conformal Inference Under Distribution Shift. *Advances in Neural Information Processing Systems.*

Papadopoulos, H., Proedrou, K., Vovk, V., and Gammerman, A. (2002). Inductive Confidence Machines for Regression. In *Machine Learning: ECML 2002.* Springer.

Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic Learning in a Random World.* Springer US.