Adaptive Conformal Predictions for Time Series

Margaux Zaffran

39th International Conference on Machine Learning

Olivier Féron

EDF R&D
FiME
Paris

Yannig Goude

EDF R&D
LMO
Paris

Julie Josse

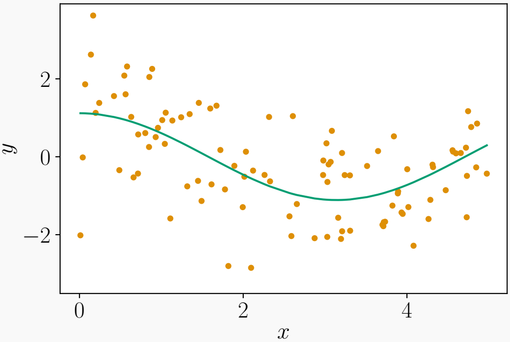
INRIA
IDESP
Montpellier

Aymeric Dieuleveut

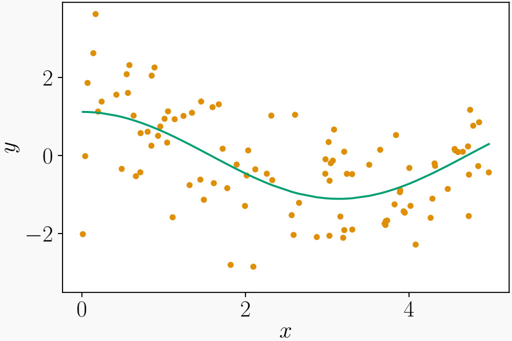
Ecole Polytechnique
Paris

France

Usual statistical learning

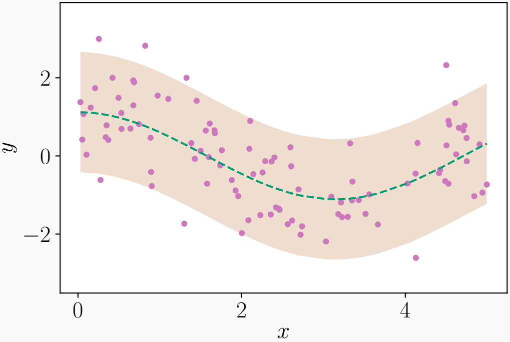


Usual statistical learning

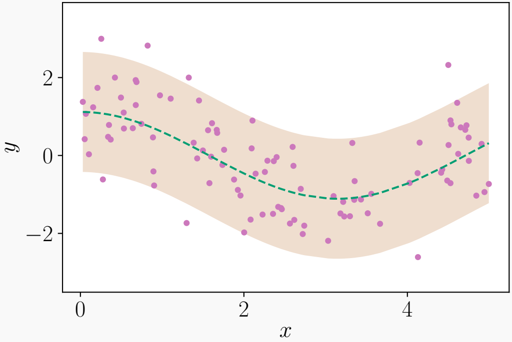


Unquantified uncertainty $\Rightarrow$ incapacity of knowing if you can trust these predictions

Split conformal prediction



Split conformal prediction



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

For example: $\alpha = 0.1$ and obtain a 90% coverage interval.

Conformal prediction: summary

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

Split conformal prediction is simple to compute and works:

- finite sample;
- any regression algorithm (neural nets, random forest...);
- distribution-free as long as the data is exchangeable.

Time series are not exchangeable

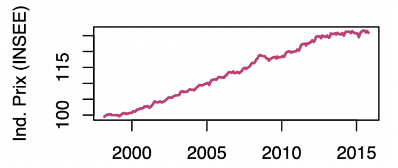


Figure 1: Trend[1]

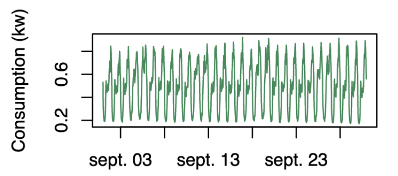


Figure 2: Seasonality[2]

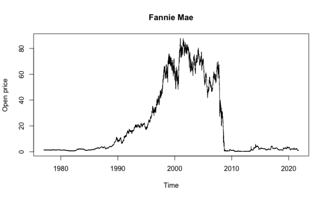


Figure 3: Shift

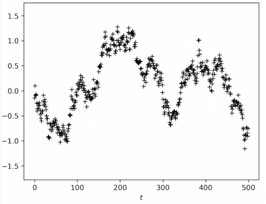


Figure 4: Time dependence

[1] Images from Yannig Goude class material.

Framework and notations

- Data: T_0 observations $(x_1, y_1), \ldots, (x_{T_0}, y_{T_0})$ in $\mathbb{R}^d \times \mathbb{R}$
- Aim: predict the response values as well as predictive intervals for T_1 subsequent observations $x_{T_0+1}, \ldots, x_{T_0+T_1}$

$\hookrightarrow$ Build the smallest interval Interval_α^t such that:

$$\mathbb{P}\left\{Y_t \in \text{Interval}_\alpha^t(X_t)\right\} \geq 1 - \alpha, \quad \text{for } t \in [\![T_0 + 1, T_0 + T_1]\!].$$

Extensions of conformal prediction to forecasting time series

- Chernozhukov et al. (2018)
- Wisniewski et al. (2020) and Kath and Ziel (2021)
- Xu and Xie (2021)
- Gibbs and Candès (2021)

Adaptive Conformal Inference (ACI), Gibbs and Candès (2021)

- Chernozhukov et al. (2018)
- Wisniewski et al. (2020) and Kath and Ziel (2021)
- Xu and Xie (2021)
- $\boxed{\text{Gibbs and Candès (2021)}}$

Adaptive Conformal Inference (ACI), Gibbs and Candès (2021)

- Chernozhukov et al. (2018)
- Wisniewski et al. (2020) and Kath and Ziel (2021)
- Xu and Xie (2021)
- Gibbs and Candès (2021)

Initially proposed to handle distribution shift.

Adaptive Conformal Inference (ACI), Gibbs and Candès (2021)

- Chernozhukov et al. (2018)
- Wisniewski et al. (2020) and Kath and Ziel (2021)
- Xu and Xie (2021)
- Gibbs and Candès (2021)

Initially proposed to handle distribution shift.

The proposed update scheme is the following:

$$\alpha_{t+1} := \alpha_t + \gamma \left(\alpha - \text{error}_t \right)$$

with some chosen $\gamma \geq 0$.

ACI asymptotic result

The proposed update scheme is the following:

$$\alpha_{t+1} := \alpha_t + \gamma \left(\alpha - \mathrm{error}_t \right)$$

with some chosen $\gamma \geq 0$.

Gibbs and Candès (2021) provide an asymptotic validity result for any distribution.

$$\frac{1}{T_1} \sum_{t=T_0+1}^{T_0+T_1} \mathbb{1} \left\{ y_t \in \mathrm{Interval}_t(x_t) \right\} \underset{T_1 \to +\infty}{\longrightarrow} 1 - \alpha \quad \textit{e.g. 90\%}$$

ACI asymptotic result

The proposed update scheme is the following:

$$\alpha_{t+1} := \alpha_t + \gamma \left(\alpha - \text{error}_t\right)$$

with some chosen $\gamma \geq 0$.

Gibbs and Candès (2021) provide an asymptotic validity result for any distribution.

$$\left| \frac{1}{T_1} \sum_{t=T_0+1}^{T_0+T_1} \mathbb{1}\left\{ y_t \in \text{Interval}_t(x_t) \right\} - (1-\alpha) \right| \leq \frac{2}{\gamma T_1}$$

ACI asymptotic result

The proposed update scheme is the following:

$$\alpha_{t+1} := \alpha_t + \gamma \left(\alpha - \mathrm{error}_t \right)$$

with some chosen $\gamma \geq 0$.

Gibbs and Candès (2021) provide an asymptotic validity result for any distribution.

$$\left| \frac{1}{T_1} \sum_{t=T_0+1}^{T_0+T_1} \mathbb{1} \left\{ y_t \in \mathrm{Interval}_t(x_t) \right\} - (1 - \alpha) \right| \leq \frac{2}{\gamma T_1}$$

$\Rightarrow$ favors large γ.

ACI asymptotic result

The proposed update scheme is the following:

$$\alpha_{t+1} := \alpha_t + \gamma \left(\alpha - \mathrm{error}_t \right)$$

with some chosen $\gamma \geq 0$.

Gibbs and Candès (2021) provide an asymptotic validity result for any distribution.

$$\left| \frac{1}{T_1} \sum_{t=T_0+1}^{T_0+T_1} \mathbb{1} \left\{ y_t \in \mathrm{Interval}_t(x_t) \right\} - (1 - \alpha) \right| \leq \frac{2}{\gamma T_1}$$

$\Rightarrow$ favors large γ. But, the higher γ, the more frequent are the infinite intervals.

Theoretical analysis of ACI's length

Approach

- Consider extreme cases (useful in an adversarial context) with simple theoretical distributions (additional assumptions);

Approach

- Consider extreme cases (useful in an adversarial context) with simple theoretical distributions (additional assumptions);
- Assume the calibration is perfect (and more), to rely on Markov Chain theory.

Theoretical analysis of ACI's length: exchangeable case

Theorem (Informal)

If the data is exchangeable and if the calibration is perfect, then as $\gamma \to 0$:

$$\textit{Average length of intervals from ACI using } \gamma$$

$$=$$

$$\textit{Average length of intervals from Split Conformal Prediction}$$

$$+ \gamma \times \mathcal{C}(\alpha, \textit{distribution of the data}),$$

where $\mathcal{C}(\alpha, \textit{distribution of the data}) > 0$ in non-atypical cases.

Theoretical and numerical analysis of ACI's length: AR(1) case

$$\varepsilon_{t+1} = \varphi\varepsilon_t + \xi_{t+1}$$

Theoretical and numerical analysis of ACI's length: AR(1) case

$$\varepsilon_{t+1} = \varphi \varepsilon_t + \xi_{t+1}$$

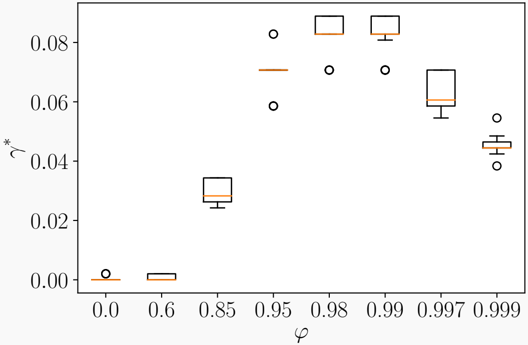


Figure 5: γ^* minimizing the average length for each φ.

AgACI

AgACI: adaptive wrapper around ACI

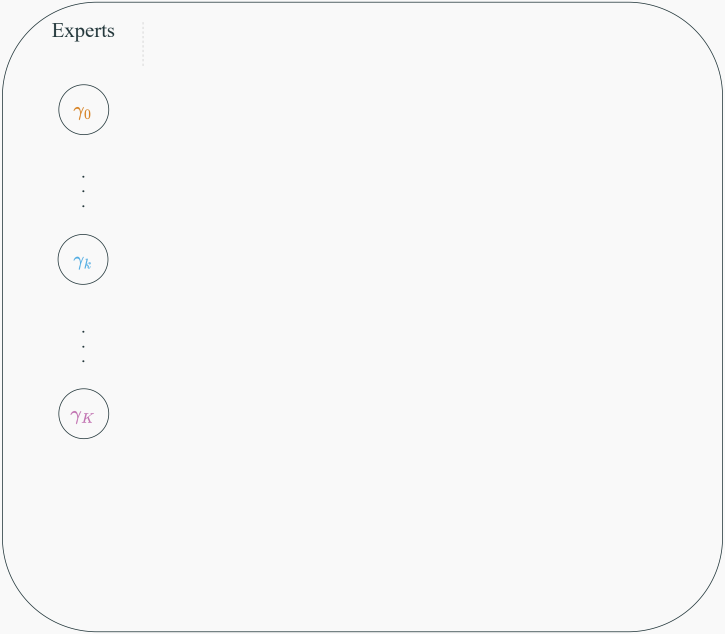
Online aggregation under expert advice (Cesa-Bianchi and Lugosi, 2006) computes an optimal weighted mean of experts.

AgACI: adaptive wrapper around ACI

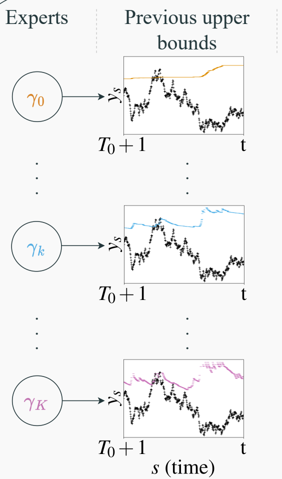
Online aggregation under expert advice (Cesa-Bianchi and Lugosi, 2006) computes an optimal weighted mean of experts.

`AgACI` performs 2 independent aggregations: one for each bound (the upper and lower ones).

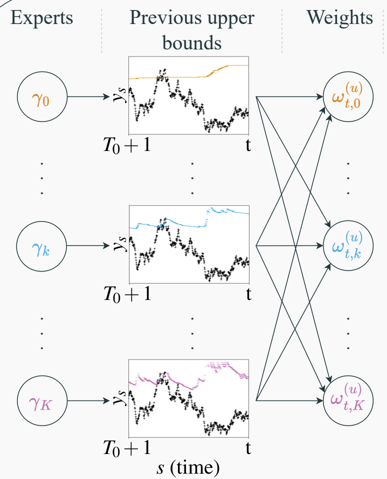
AgACI: adaptive wrapper around ACI, scheme (upper bound)



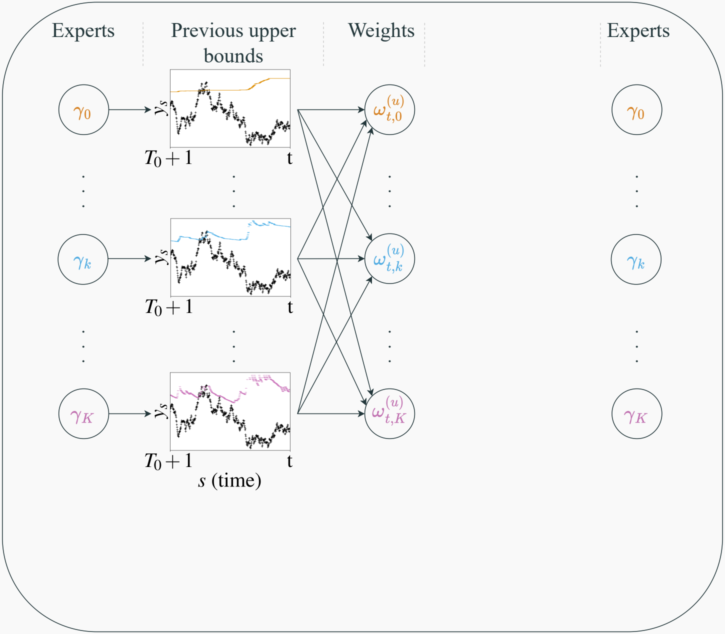
AgACI: adaptive wrapper around ACI, scheme (upper bound)



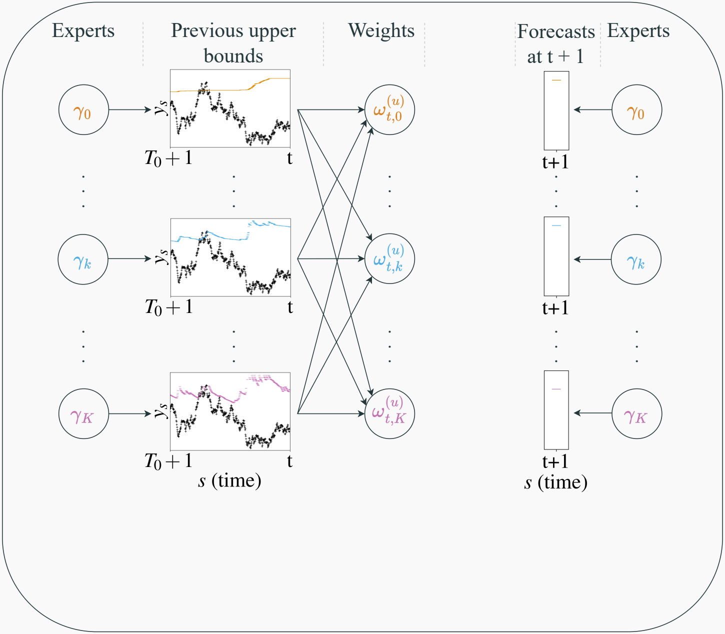
AgACI: adaptive wrapper around ACI, scheme (upper bound)



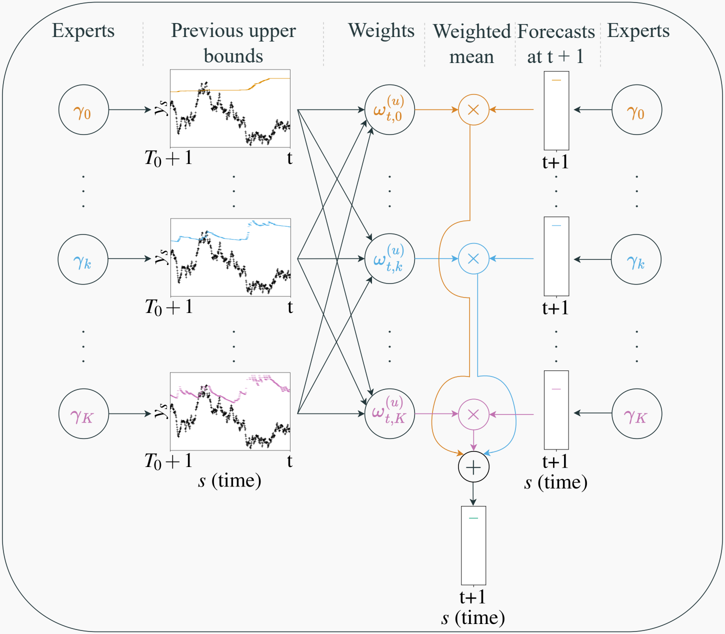
AgACI: adaptive wrapper around ACI, scheme (upper bound)



AgACI: adaptive wrapper around ACI, scheme (upper bound)



AgACI: adaptive wrapper around ACI, scheme (upper bound)



Numerical experiments

Simulated data and French electricity price forecasting

Experimental take-away messages

- Benchmarks are not robust to the increase in the temporal dependence;

Experimental take-away messages

- Benchmarks are not robust to the increase in the temporal dependence;
- ACI is robust, maintaining validity, with an appropriate γ;

Experimental take-away messages

- Benchmarks are not robust to the increase in the temporal dependence;
- ACI is robust, maintaining validity, with an appropriate γ;
- AgACI is robust, maintaining validity, not the smallest;
- more on the paper!

Thanks for listening and feel free to reach out to us, have a look at the paper for more details, and join us at the poster session!