Adaptive Conformal Inference through the Lens of Blackwell Approachability

Guillaume Principato Gilles Stoltz

March 21, 2026

My thesis in buzz words

PhD objective: Probabilistic forecasting for EV smart charging

Hierarchical data

EV charging

Conformal prediction

Time series

My thesis in buzz words

PhD objective: Probabilistic forecasting for EV smart charging

~~Hierarchical data~~

~~EV charging~~

Conformal prediction

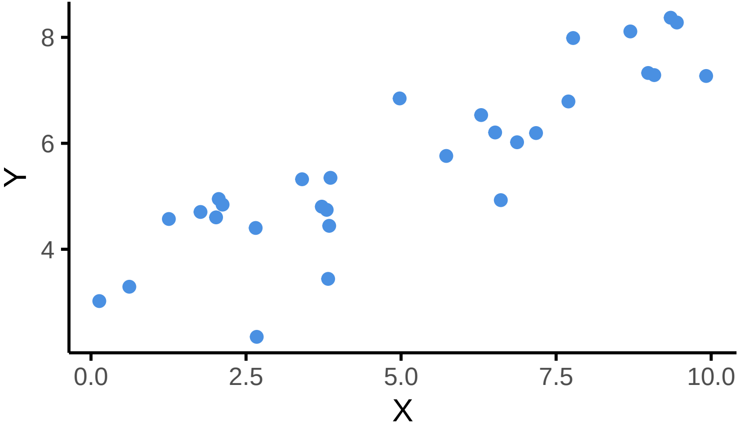
Beyond exchangeability

This talk:
Adaptive Conformal Inference through
the Lens of Blackwell Approachability

Outline

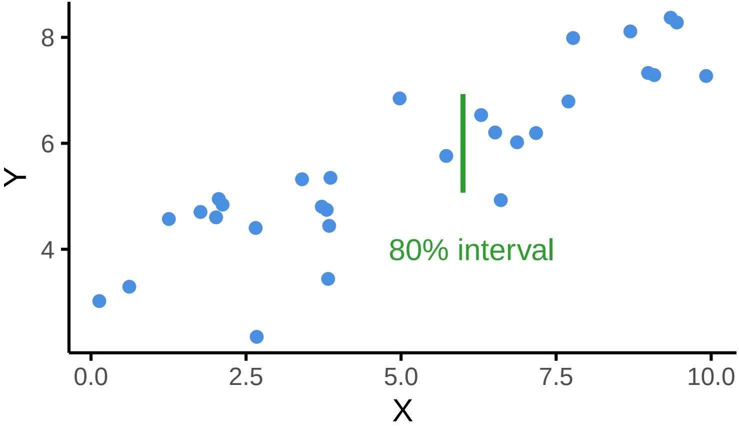
General setting for SCP

Regression setting with covariates $x_t \in \mathbb{R}$ and targets $y_t \in \mathcal{Y}$. Past observations are available. We receive x_t and want to predict y_t with confidence $1 - \alpha = 80\%$.



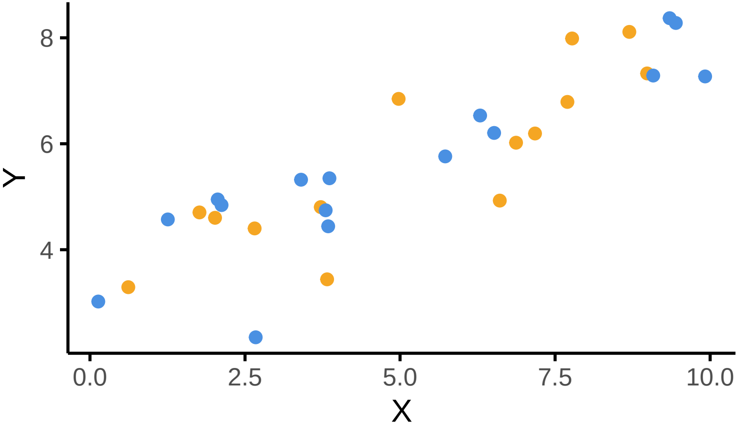
General setting for SCP

Regression setting with covariates $x_t \in \mathbb{R}$ and targets $y_t \in \mathcal{Y}$. Past observations are available. We receive x_t and want to predict y_t with confidence $1 - \alpha = 80\%$.



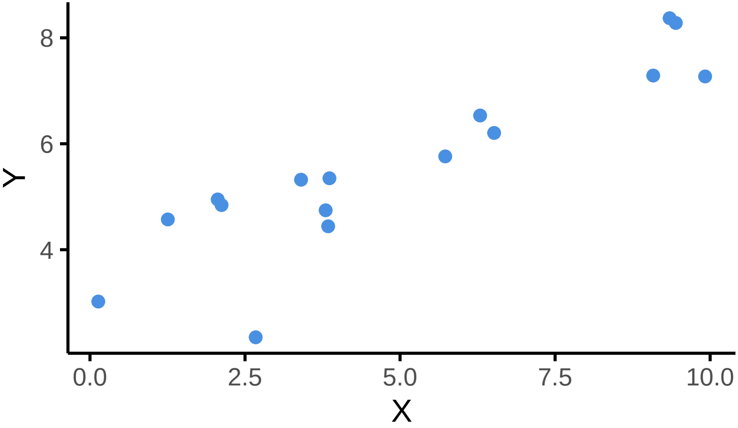
SCP with absolute residuals

Step 1. Split the data between $\mathcal{D}_{\text{train}}$ and $\mathcal{D}_{\text{calib}}$.



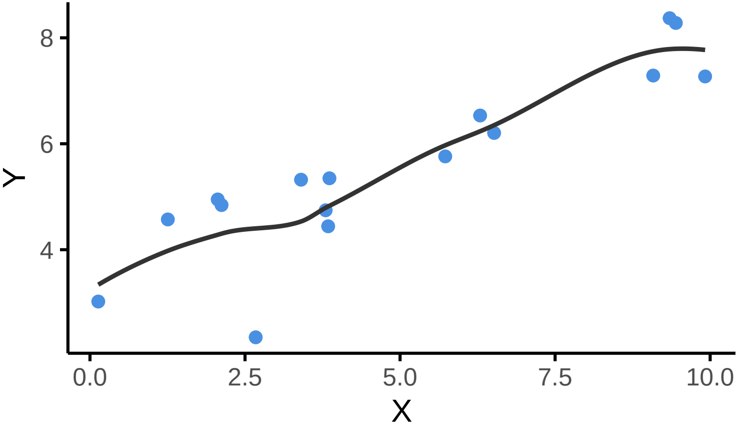
SCP with absolute residuals

Step 2. Learn a regression function $\widehat{\mu}$ on $\mathcal{D}_{\text{train}}$.



SCP with absolute residuals

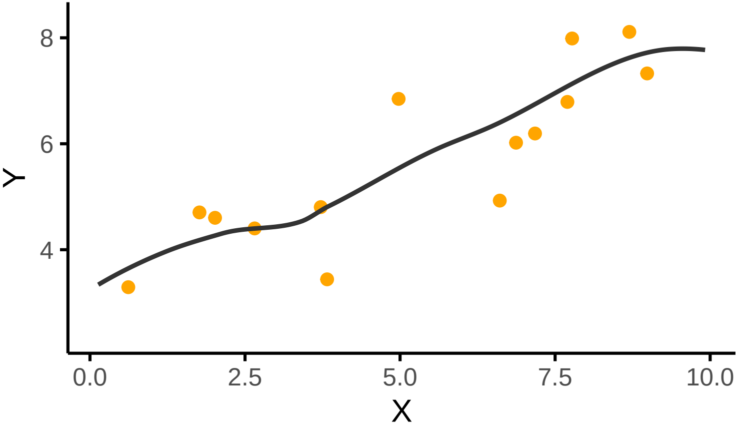
Step 2. Learn a regression function $\widehat{\mu}$ on $\mathcal{D}_{\text{train}}$.



SCP with absolute residuals

Step 3. Compute conformity scores on $\mathcal{D}_{\text{calib}}$, e.g., for $k \in \mathcal{D}_{\text{calib}}$,

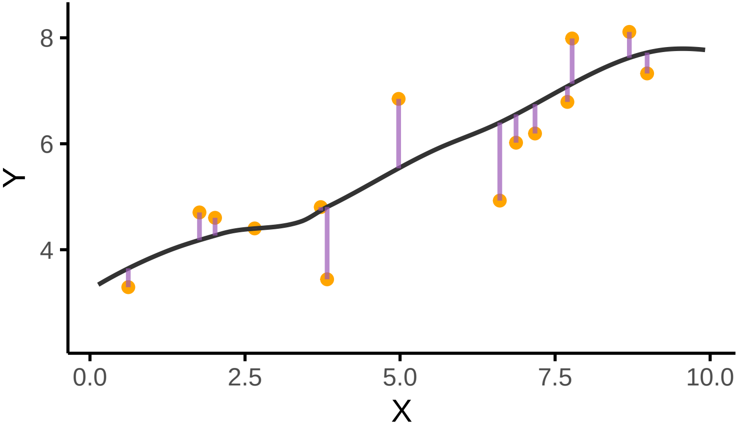
$$s_k = \left| y_k - \widehat{\mu}(\boldsymbol{x}_k) \right| .$$



SCP with absolute residuals

Step 3. Compute conformity scores on $\mathcal{D}_{\text{calib}}$, e.g., for $k \in \mathcal{D}_{\text{calib}}$,

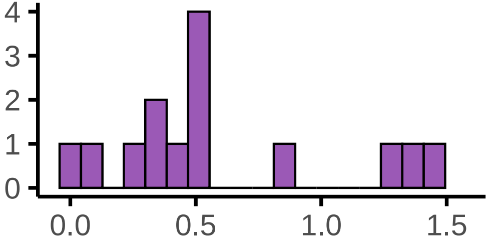
$$s_k = \left| y_k - \widehat{\mu}(\boldsymbol{x}_k) \right| .$$

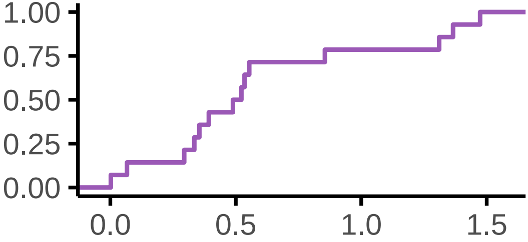


SCP with absolute residuals

Step 4. Compute the empirical quantile q of the conformity scores, e.g.,

$$q = s_{\left(\lceil (|\mathcal{D}_{\text{calib}}|+1)(1-\alpha)\rceil\right)} \,.$$

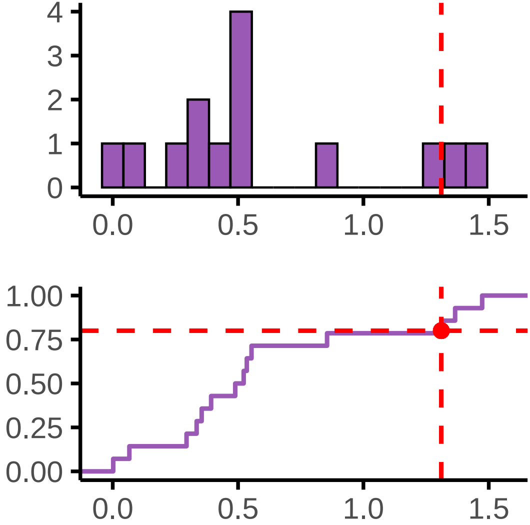




SCP with absolute residuals

Step 4. Compute the empirical quantile q of the conformity scores, e.g.,

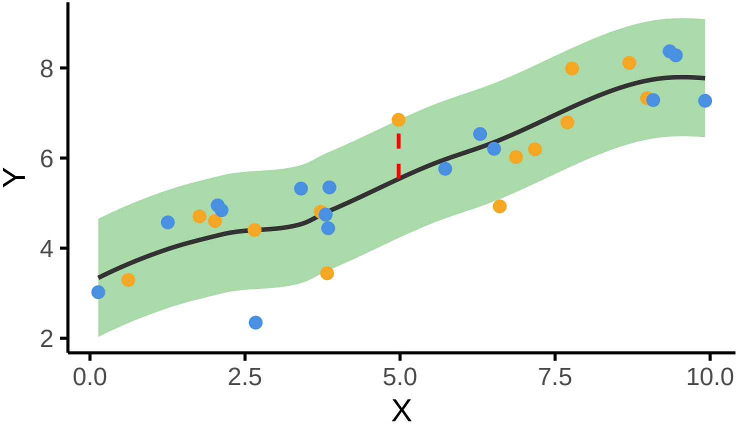
$$q = s_{\left(\lceil (|\mathcal{D}_{\mathrm{calib}}|+1)(1-\alpha)\rceil\right)} .$$



SCP with absolute residuals

Step 5. Output the prediction set

$$C_{1-\alpha}(\boldsymbol{x}_t) = \left[\widehat{\mu}(\boldsymbol{x}_t) \pm q\right].$$



Coverage guarantee of SCP

SCP[1] achieves the following objective

Objective: At time t, build a prediction set $C_{1-\alpha}(x_t)$ such that:

Theorem 1: Coverage of SCP [Vovk et al. (2005)]

If $(x_k, y_k)_{k \in \mathcal{D}_{\mathrm{calib}} \cup \{t\}}$ are i.i.d. (or exchangeable), then

$$\mathbb{P}\big(y_t \in C_{1-\alpha}(x_t)\big) \geqslant 1 - \alpha\,.$$

Moreover, under an additional continuity assumption

$$\mathbb{P}\big(y_t \in C_{1-\alpha}(x_t)\big) \leqslant 1 - \alpha + \mathcal{O}\left(\frac{1}{|\mathcal{D}_{\mathrm{calib}}|}\right)\,.$$

[1] the name SCP comes from Lei et al. (2018) and is a rebranding of inductive conformal prediction.

General summary of SCP

Split the data between $\mathcal{D}_{\text{train}}$ and $\mathcal{D}_{\text{calib}}$

↓

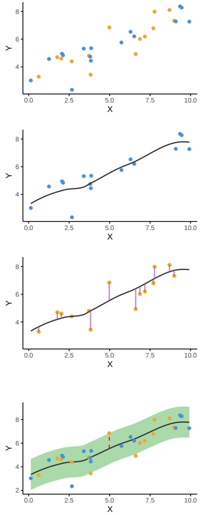
Learn the regression function $\widehat{\mu}$ on $\mathcal{D}_{\text{train}}$

↓

Compute conformity scores on $\mathcal{D}_{\text{calib}}$

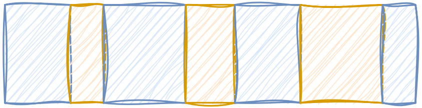
↓

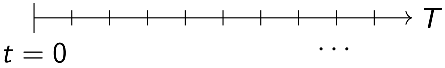
Derive the prediction set using the empirical quantiles of the scores



Outline

New paradigm: sequential setting (Gibbs and Candès, 2021)





Split past data into $\mathcal{D}_{\mathrm{train}}$ and $\mathcal{D}_{\mathrm{calib}}$.

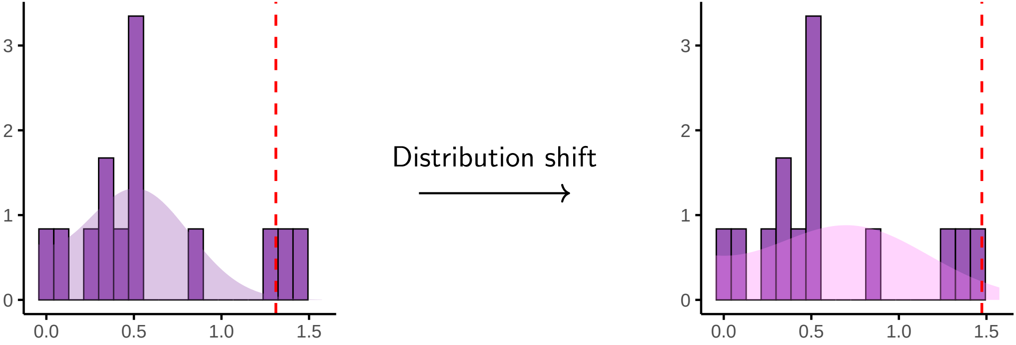
It induces candidate prediction sets $C(\cdot)$ based on $\widehat{\mu}(\cdot)$ and on the empirical distribution of the conformity scores on $\mathcal{D}_{\mathrm{calib}}$.

At each round $t \geqslant 1$, the learner

- observes x_t
- outputs a prediction set $C(x_t)$
- observes y_t and sees whether $y_t \in C(x_t)$ or not.

Beyond exchangeability: the need for adaptivity

Motivation: The i.i.d. assumption is not realistic in many applications, e.g., due to distribution shifts (such as during Covid).



The idea of Gibbs and Candès (2021) is to adapt the miscoverage level α at each round t by picking α_t based on past performance.

Clarification on the candidate prediction sets

Assumptions: We fix $\mathcal{D}_{\mathrm{train}}$ and $\mathcal{D}_{\mathrm{calib}}$ at the beginning of the game, and assume the conformity scores are almost surely distinct and bounded by $L < \infty$.

The candidate prediction sets at round $t \geqslant 1$ are of the form:

$$C_{1-\alpha_t}(x_t) = \left[\widehat{\mu}(x_t) \pm s_{\left(\lceil (|\mathcal{D}_{\mathrm{calib}}|+1)(1-\alpha_t) \rceil\right)}\right], \qquad \text{for some } \alpha_t \in [0,1] \, .$$

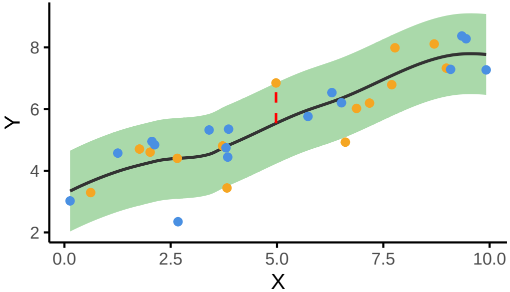
Therefore, in the setting we consider (Gibbs and Candès, 2021), our only degree of freedom is to pick α_t at each round t.

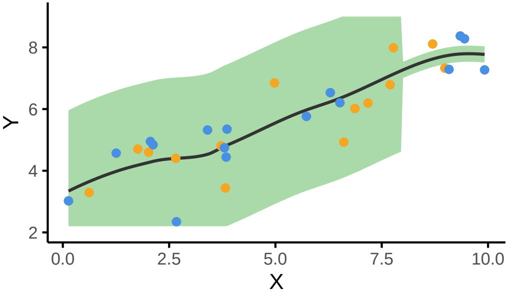
Remark on the edge cases

By convention, we set $s_{(0)} = 0$ and $s_{(|\mathcal{D}_{\mathrm{calib}}|+1)} = L$, so that

$$\begin{aligned} \alpha_t \geqslant 1 \quad &\Longrightarrow \quad C_{1-\alpha_t}(x_t) = C_0(x_t) = \emptyset\,, \\ \alpha_t \leqslant 0 \quad &\Longrightarrow \quad C_{1-\alpha_t}(x_t) = C_1(x_t) = \left[\widehat{\mu}(x_t) \pm L\right]. \end{aligned}$$

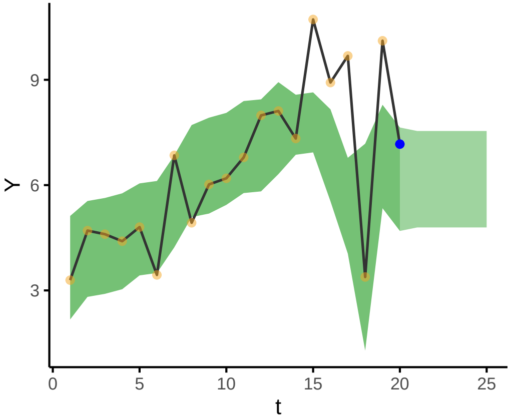
We can force the coverage frequency, but it may come at the cost of large prediction sets.

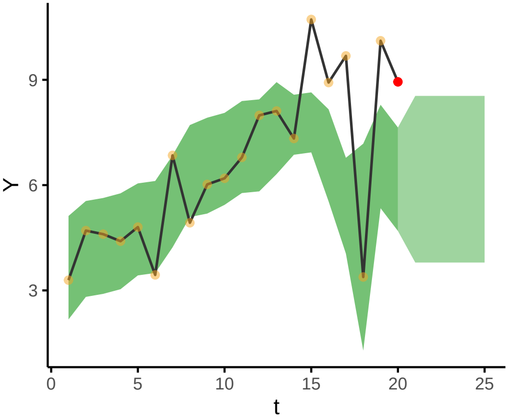




Presentation of the ACI algorithm (Gibbs and Candès, 2021)

$$\alpha_{t+1} = \alpha_t - \gamma\left(\mathbf{1}_{y_t \notin C_{1-\alpha_t}(x_t)} - \alpha\right), \qquad \text{for some learning rate } \gamma > 0\,.$$





What Guarantees for ACI?

The main theoretical result of Gibbs and Candès (2021) is that ACI achieves asymptotic validity even for abitrary sequences $(x_t, y_t)_{t\geqslant 1}$, i.e., it satisfies

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{y_t \notin C_{1-\alpha_t}(x_t)} \leqslant \alpha \qquad \text{a.s.}$$

Sketch of proof: Recall the update rule of ACI: $\alpha_{t+1} = \alpha_t - \gamma\left(\mathbf{1}_{y_t \notin C_{1-\alpha_t}(x_t)} - \alpha\right)$.
By noticing a telescoping sum, we get

$$\left|\frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{y_t \notin C_{1-\alpha_t}(x_t)} - \alpha\right| \leqslant \frac{|\alpha_1 - \alpha_{T+1}|}{\gamma T} \leqslant \frac{\max\{\alpha_1, 1-\alpha_1\} + \gamma}{\gamma T} \qquad \text{a.s.,}$$

where the second inequality is due to the fact that $\alpha_t \in [-\gamma, 1+\gamma]$ almost surely.

Extensions of ACI

There is a trade-off for the choice of the learning rate γ:

- If γ is too small, then the convergence to the target miscoverage α is slow.
- If γ is too large, then α_t can become negative or larger than 1, inducing variability.

Recent works have focused on a sequential choice of γ_t based on the interpretation of ACI as a online gradient descent algorithm. Notably, this adaptation can be done through:

- expert aggregation (Zaffran et al., 2022)
- sleeping expert (Gibbs and Candès, 2024)
- decaying step size (Angelopoulos et al., 2024)

Open questions and limitations of ACI

Open questions:

- Can we also control the size of the prediction sets?
- Interpretation of ACI as a no-regret algorithm with respect to the pinball loss (Gibbs and Candès, 2024).
- Is their a general class of algorithms that can achieve the same guarantees as ACI?

Limitations of ACI:

- Is ACI still conformal prediction?
- ACI is not adaptive to the difficulty of the problem.

Outline

Definition of the objectives

Our goal is twofold:

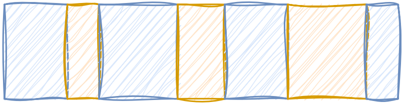
$$\text{ensure} \qquad \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{y_t \notin C_{1-\alpha_t}(x_t)} \leqslant \alpha\,, \tag{$\star$}$$

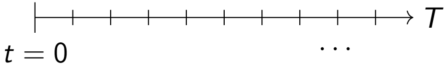
$$\text{control} \qquad \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathfrak{L}\big(C_{1-\alpha_t}\big)\,. \tag{$\star\star$}$$

We refer to ($\star$) as validity, and to ($\star\star$) as efficiency.

To that end, we formulate ACI as a repeated finite game, and express these objectives in terms of the approachability of some target set.

Formulation of ACI as a Repeated Finite Game





Split past data into $\mathcal{D}_{\mathrm{train}}$ and $\mathcal{D}_{\mathrm{calib}}$.

It induces a finite number of candidate prediction sets $C_{1-r(\alpha)}(\cdot)$ with $r(\alpha) \in \mathcal{A}$ based on $\widehat{\mu}(\cdot)$ and on the empirical distribution of the conformity scores on $\mathcal{D}_{\mathrm{calib}}$.

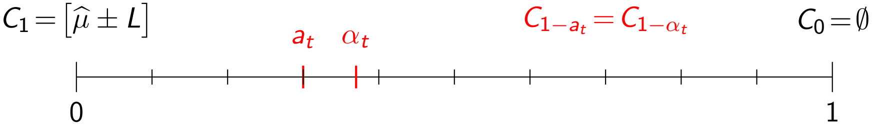
At each round $t \geqslant 1$, the learner

- observes x_t
- picks $a_t \in \mathcal{A}$ and outputs a prediction set $C_{1-a_t}(x_t)$
- observes y_t and a corresponding $b_t \in \mathcal{B}$ and sees whether $y_t \in C_{1-a_t}(x_t)$ or not based on a_t and b_t.

Formulation of ACI as a Repeated Finite Game

Recall that a candidate prediction set is of the form

$$C_{1-\alpha_t}(x_t) = \left[\widehat{\mu}(x_t) \pm s_{\left(\lceil(|\mathcal{D}_{\mathrm{calib}}|+1)(1-\alpha_t)\rceil\right)}\right], \qquad \text{for some } \alpha_t \in [0,1],$$

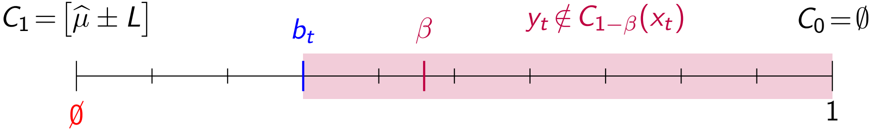


$$\mathcal{A} = \left\{ \frac{k}{|\mathcal{D}_{\mathrm{calib}}|+1} : k = 0, 1, \ldots, |\mathcal{D}_{\mathrm{calib}}| + 1 \right\}$$

Formulation of ACI as a Repeated Finite Game

Following Gibbs and Candès (2024), we can think about the target y_t as being generated by an opponent who picks some $b_t \in \mathcal{B}$ defined as

$$\begin{aligned} b_t &:= \sup\left\{\beta \in [0,1] : y_t \in C_{1-\beta}(x_t)\right\} \\ &= \min\left\{\beta \in \mathcal{B} : y_t \notin C_{1-\beta}(x_t)\right\}. \end{aligned}$$

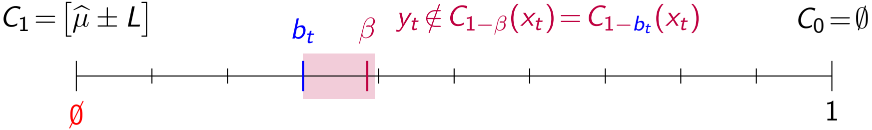


$$\mathcal{B} = \left\{\frac{k}{|\mathcal{D}_{\mathrm{calib}}|+1} : k = \cancel{0}, 1, \ldots, |\mathcal{D}_{\mathrm{calib}}| + 1\right\}$$

Formulation of ACI as a Repeated Finite Game

Following Gibbs and Candès (2024), we can think about the target y_t as being generated by an opponent who picks some $b_t \in \mathcal{B}$ defined as

$$\begin{aligned} b_t &:= \sup\left\{\beta \in [0,1] : y_t \in C_{1-\beta}(x_t)\right\} \\ &= \min\left\{\beta \in \mathcal{B} : y_t \notin C_{1-\beta}(x_t)\right\}. \end{aligned}$$

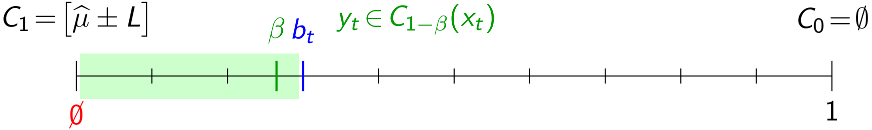


$$\mathcal{B} = \left\{\frac{k}{|\mathcal{D}_{\mathrm{calib}}|+1} : k = \emptyset, 1, \ldots, |\mathcal{D}_{\mathrm{calib}}| + 1\right\}$$

Formulation of ACI as a Repeated Finite Game

Following Gibbs and Candès (2024), we can think about the target y_t as being generated by an opponent who picks some $b_t \in \mathcal{B}$ defined as

$$
\begin{aligned}
b_t &:= \sup\Big\{\beta \in [0,1] : y_t \in C_{1-\beta}(x_t)\Big\} \\
&= \min\Big\{\beta \in \mathcal{B} : y_t \notin C_{1-\beta}(x_t)\Big\}\,.
\end{aligned}
$$



$$\mathcal{B} = \Big\{\frac{k}{|\mathcal{D}_{\mathrm{calib}}|+1} : k = \cancel{0}, 1, \ldots, |\mathcal{D}_{\mathrm{calib}}| + 1\Big\}$$

Formulation of ACI as a Repeated Finite Game

This rewriting is useful to encode different settings:

Exchangeable setting.

At each round t, the opponent picks b_t uniformly at random in $\mathcal{B}$, i.e., plays

$$b_t \sim \boldsymbol{q}_t = \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} \boldsymbol{\delta}_b \,.$$

Adversarial setting.

At each round t, the opponent picks b_t according to an arbitrary distribution $\boldsymbol{q}_t \in \Delta(\mathcal{B})$.

In between settings. ?

Formulation of ACI as a Repeated Finite Game

We recall our objectives $(\star)$ and $(\star\star)$

$$\text{ensure} \qquad \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{y_t \notin C_{1-\alpha_t}(x_t)} \leqslant \alpha\,, \tag{$\star$}$$

$$\text{control} \qquad \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathfrak{L}\big(C_{1-\alpha_t}\big)\,, \tag{$\star\star$}$$

and by definition of b_t, rewrite them as

$$\text{ensure} \qquad \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{b_t \leqslant a_t} \leqslant \alpha\,, \tag{$\star$}$$

$$\text{control} \qquad \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathfrak{L}\big(C_{1-a_t}\big)\,. \tag{$\star\star$}$$

Formulation of ACI as a Repeated Finite Game

We can now rewrite $(\star)$ and $(\star\star)$ with the following vector-valued payoff function

$$m : (a, b) \in \mathcal{A} \times \mathcal{B} \mapsto \begin{pmatrix} \mathbf{1}_{b \leqslant a} \\ \mathfrak{L}(C_{1-a}) \end{pmatrix}$$

by studying the convergence of the average payoff

$$\overline{m}_T := \frac{1}{T} \sum_{t=1}^{T} m(a_t, b_t) \qquad \text{to some target set } \mathcal{C} = \begin{pmatrix} [0, \alpha] \\ ? \end{pmatrix} .$$

Blackwell's theorem provides a necessary and sufficient condition for the approachability of $\mathcal{C}$ in a finite game, and an explicit strategy for the learner to ensure it.

Blackwell's Approachability theorem

A closed convex set $\mathcal{C}$ is said to be approachable if there exists a strategy for the learner such that, for any strategy of the opponent,

$$\lim_{T \to \infty} d\left(\overline{m}_T, \mathcal{C}\right) = 0 \qquad \text{a.s.}$$

Theorem 2: Approachability theorem [Blackwell (1956)]

A closed convex set $\mathcal{C}$ is approachable if and only if

$$\forall \boldsymbol{q} \in \Delta(\mathcal{B}), \exists \boldsymbol{p} \in \Delta(\mathcal{A}), \qquad m(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{C}\,. \tag{1}$$

Application of Theorem 2 includes e.g., external regret (Hannan, 1957) and calibration (Foster, 1999; Mannor and Stoltz, 2010).

Remark: The condition (1) is referred to as the dual condition.

Summary and overview of the rest of the talk

In the real world:

- observe x_t and output $C_t(x_t)$
- observe y_t

so that $\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{y_t \notin C_t(x_t)} \leqslant \alpha$

and $\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathfrak{L}\big(C_t\big)$ is "small".

In the repeated finite game:

- pick $a_t \in \mathcal{A}$
- observe $b_t \in \mathcal{B}$

so that $\lim_{T\to\infty} d\big(\overline{m}_T, \mathcal{C}\big) = 0$

where $\mathcal{C}$ needs to be defined to encode the objectives of validity and efficiency.

What's next?

- We present a new strategy.
- We show that it ensures the approachability of a satisfying target set.
- We give a taste of the general results.

Outline

BO-ACI: a new strategy for ACI

We leverage opportunistic approachability (Bernstein et al., 2014), i.e., the dual condition of Theorem 2. We define the function

$$\text{mscv} : (\alpha, \boldsymbol{q}) \longmapsto \text{“largest valid } a \in \mathcal{A}\text{, on average over } \boldsymbol{q}\text{”} .$$

Inputs: $\alpha \in [0, 1]$; auxiliary sequential forecaster $\mathcal{F}$ with outputs in $\Delta(\mathcal{B})$

Initialization: play any $a_1 \in \mathcal{A}$

At time steps $t = 2, 3, \dots$:

1. Get from $\mathcal{F}$ a probabilistic forecast $\boldsymbol{z}_t \in \Delta(\mathcal{B})$ of b_t.
2. Play $a_t = \text{mscv}(\alpha, \boldsymbol{z}_t)$, observe x_t and output $C_{1-a_t}(x_t)$
3. Observe b_t (i.e., y_t), and feed $\mathcal{F}$ with b_t

Theoretical guarantees of BO-ACI

For the sake of exposure, we present informal statements of the results.

Theorem 3: BO-ACI in the exchangeable setting (Principato and Stoltz, 2025)

The BO-ACI strategy, when run with a calibrated[2] forecaster, satisfies,

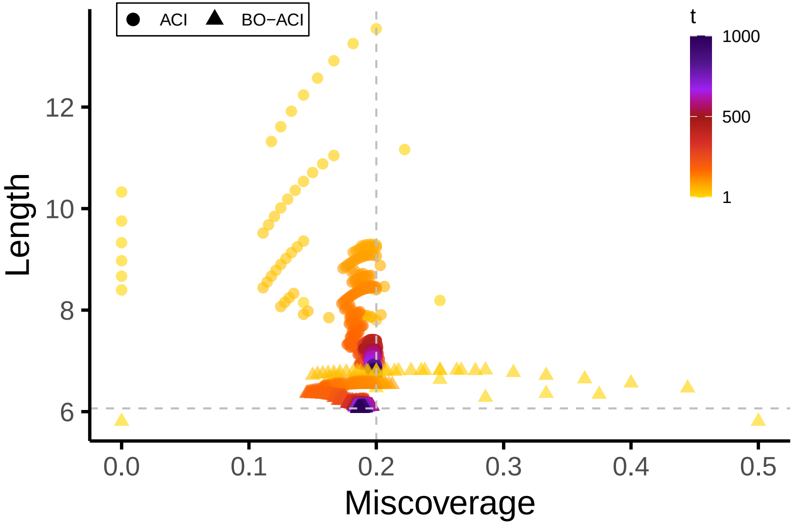
$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{y_t \notin C_{1-a_t}(x_t)} \leqslant \alpha \qquad \text{a.s.,}$$

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathfrak{L}\big(C_{1-a_t}\big) = \mathfrak{L}\big(C_{1-\alpha}\big) \qquad \text{a.s.}$$

This theoretical result is stronger than the one of ACI, see Zaffran et al. (2022)

[2] i.e., such that forecast probabilities match empirical frequencies (Dawid, 1982; Foster, 1999).

Comparison of BO-ACI with ACI in the exchangeable setting



Theoretical guarantees of BO-ACI

Theorem 4: BO-ACI in the adversarial setting (Principato and Stoltz, 2025)

The BO-ACI strategy, when run with a calibrated forecaster, satisfies,

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{y_t \notin C_{1-a_t}(x_t)} \leqslant \alpha \qquad \text{a.s.},$$

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathfrak{L}\big(C_{1-a_t}\big) \leqslant 2L \qquad \text{a.s.}$$

This efficiency control is useless, but we show that it is impossible to ensure a better one.

The remaining question is: what happens in between these two settings?

Q-restricted opponent's play

Let $Q \subseteq \Delta(\mathcal{B})$ be any convex subset. The opponent is said to be Q–restricted if,

$$\lim_{T\to\infty} \sum_{t=1}^{T} d(\boldsymbol{q}_t, Q) = 0 \qquad \text{a.s.}$$

Examples.

- Exchangeable setting: $Q = \{\boldsymbol{q}_0\}$ with $\boldsymbol{q}_0$ the uniform distribution on $\mathcal{B}$.
- Adversarial setting: $Q = \Delta(\mathcal{B})$.
- Near exchangeable setting: $Q = \{\boldsymbol{q} \in \Delta(\mathcal{B}) : d(\boldsymbol{q}, \boldsymbol{q}_0) \leqslant \varepsilon\}$ for some $\varepsilon > 0$.
- Regime switching setting: $Q = \operatorname{conv}\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_K\}$ for some $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_K \in \Delta(\mathcal{B})$.

General results of BO-ACI

We recall that $\mathrm{mscv}(\alpha, \boldsymbol{q})$ is the largest valid $a \in \mathcal{A}$, on average over $\boldsymbol{q}$.

Theorem 5: Opportunistic Approachability of BO-ACI (Principato and Stoltz, 2025)

The BO-ACI strategy, when run with a calibrated forecaster, satisfies, whenever the opponent is Q–restricted,

$$\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{y_t \notin C_{1-a_t}(x_t)} \leqslant \alpha \qquad \text{a.s.},$$

$$\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathfrak{L}\big(C_{1-a_t}\big) \leqslant \max_{\boldsymbol{q} \in Q} \mathfrak{L}\big(C_{1-\mathrm{mscv}(\alpha, \boldsymbol{q})}\big) \qquad \text{a.s.}$$

Remark: Theorem 3–4 are actually corollaries of Theorem 5.

Insight on the tools for Theorem 5

We follow the framework of Bernstein et al. (2014) and define a response-function $p^\star$ that maps any $\boldsymbol{q} \in \Delta(\mathcal{B})$ to a satisfactory distribution over $\mathcal{A}$, i.e., so that

$$m^\star(\boldsymbol{q}) = m\big(p^\star(\boldsymbol{q}), \boldsymbol{q}\big) \qquad \text{is a good payoff}\,.$$

We can now define[3] an opportunistically approachable target set as

$$\mathcal{C}_Q = \mathrm{conv}\Big(\big\{m^\star(\boldsymbol{q}) : d(\boldsymbol{q}, Q) = 0\big\}\Big)\,,$$

which corresponds to the result of Theorem 5 for $p^\star(\boldsymbol{q}) = \boldsymbol{\delta}_{\mathrm{mscv}(\alpha,\boldsymbol{q})}$.

[3] the formula is actually much more complex for technical reasons, but this is the main idea.

Sketch of proof of Theorem 5

Lemma 1. The Q–restriction property propagates to the calibrated forecaster.

Proof. Denote by $(\boldsymbol{z}_t)_{t \geqslant 1}$ the sequence of forecasts output by the calibrated forecaster.

$$d(\overline{m}_T, \mathcal{C}_Q) \leqslant \left\| \overline{m}_T - \frac{1}{T} \sum_{t=1}^{T} m^\star(\boldsymbol{z}_t) \right\| + d\left(\frac{1}{T} \sum_{t=1}^{T} m^\star(\boldsymbol{z}_t), \mathcal{C}_Q \right) .$$

We control (easily) the first term by calibration.

We control the second term with Lemma 1 together with geometric arguments.

Conclusion and future works

Summary:

- We showed that the ACI framework can be formulated as a repeated finite game, and that its objectives can be encoded with Blackwell approachability.
- We presented a new strategy, BO-ACI, and showed that it ensures the approachability of a satisfying target set.

On going work: Use case of Principato et al. (2024) on EV charging – application of conformal prediction to hierarchical time series with a hierarchical structure that varies over time.

References I

Angelopoulos, A. N., Barber, R., and Bates, S. (2024). Online conformal prediction with decaying step sizes. In *International Conference on Machine Learning*, pages 1616–1630. PMLR.

Bernstein, A., Mannor, S., and Shimkin, N. (2014). Opportunistic approachability and generalized no-regret problems. *Mathematics of Operations Research*, 39(4):1057–1083.

Blackwell, D. (1956). An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8.

Dawid, A. P. (1982). The well-calibrated bayesian. *Journal of the American statistical Association*, 77(379):605–610.

Foster, D. P. (1999). A proof of calibration via blackwell's approachability theorem. *Games and Economic Behavior*, 29(1-2):73–78.

Gibbs, I. and Candès, E. J. (2021). Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, 34:1660–1672.

Gibbs, I. and Candès, E. J. (2024). Conformal inference for online prediction with arbitrary distribution shifts. *Journal of Machine Learning Research*, 25(162):1–36.

References II

Hannan, J. (1957). Approximation to bayes risk in repeated play. *Contributions to the Theory of Games*, 3(2):97–139.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111.

Mannor, S. and Stoltz, G. (2010). A geometric proof of calibration. *Mathematics of Operations Research*, 35(4):721–727.

Principato, G. and Stoltz, G. (2025). Blackwell's approachability for sequential conformal inference. Preprint arXiv:2510.15824.

Principato, G., Stoltz, G., Amara-Ouali, Y., Goude, Y., Hamrouche, B., and Poggi, J.-M. (2024). Conformal prediction for hierarchical data. Preprint arXiv:2411.13479.

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

Zaffran, M., Féron, O., Goude, Y., Josse, J., and Dieuleveut, A. (2022). Adaptive conformal predictions for time series. In *International Conference on Machine Learning*, pages 25834–25866. PMLR.