Recent Advances in Conformal Prediction: Multivariate Predictions and Anytime Guarantees

Francis Bach

Inria - Ecole Normale Supérieure

ICSDS Conference, Seville, Spain - December 16, 2025

Recent Advances in Conformal Prediction: Multivariate Predictions and Anytime Guarantees

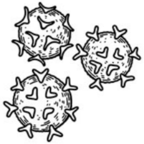
Francis Bach

Inria - Ecole Normale Supérieure

Joint work with **Sacha Braun**, **Etienne Gauthier**, **Eugène Berta**, **David Holzmüller**, **Liviu Aolaritei**, and **Michael Jordan**

ICSDS Conference, Seville, Spain - December 16, 2025

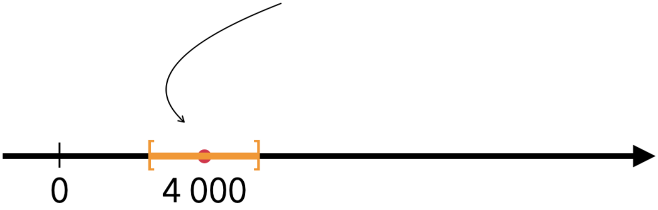
Need for uncertainty quantification

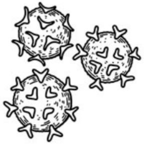


White blood cells per microliter for one individual

Need for uncertainty quantification

$$\mathbb{P}(\text{White blood cells} \in IC) = 1 - \alpha$$

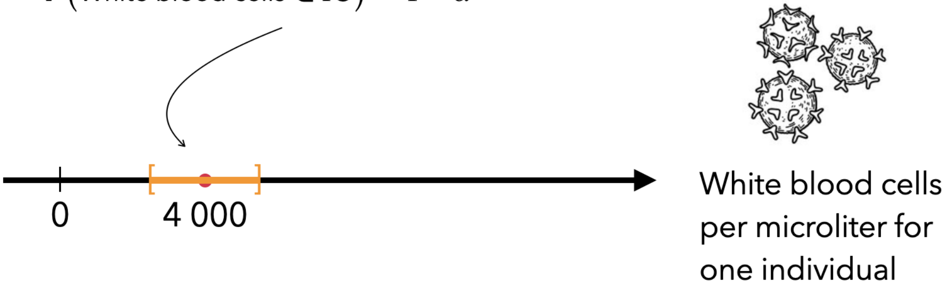




White blood cells per microliter for one individual

Need for uncertainty quantification

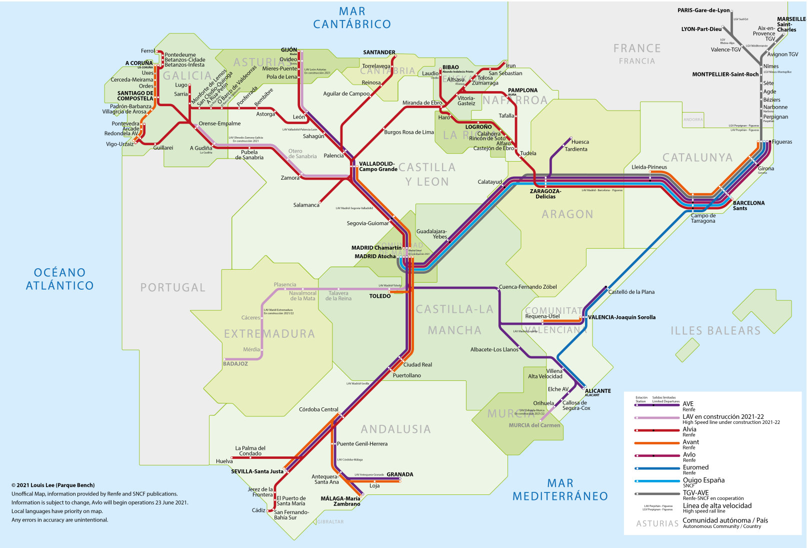
$$\mathbb{P}\big(\text{White blood cells} \in IC\big) = 1 - \alpha$$



⚠ Difficulty: High-dimensional inputs/outputs, no distributional assumptions

Need for uncertainty quantification

- Predicting delays on train networks



Need for uncertainty quantification

- Reinforcement learning and robotics

ChatGPT

Table of Contents

Foundations of conformal prediction

Conformal prediction yields finite sample guarantees

- **Observation**: dataset $\mathcal{D}$ of n independent and identically distributed (i.i.d.) samples $(X_i, Y_i) \sim \mathbb{P}_{X,Y}$ with $(X_i, Y_i) \in \mathcal{X} \times \mathcal{Y}$
- **Goal**: Given a target miscoverage level $\alpha \in (0, 1)$, construct a set-valued predictor $C_\alpha(X)$ such that

$$\mathbb{P}_{X,Y}(Y_{n+1} \in C_\alpha(X_{n+1})) \approx 1 - \alpha$$

Conformal prediction yields finite sample guarantees

- **Observation**: dataset $\mathcal{D}$ of n independent and identically distributed (i.i.d.) samples $(X_i, Y_i) \sim \mathbb{P}_{X,Y}$ with $(X_i, Y_i) \in \mathcal{X} \times \mathcal{Y}$
- **Goal**: Given a target miscoverage level $\alpha \in (0, 1)$, construct a set-valued predictor $C_\alpha(X)$ such that

$$\mathbb{P}_{X,Y}(Y_{n+1} \in C_\alpha(X_{n+1})) \approx 1 - \alpha$$

- **Conformal prediction allows finite sample guarantees** (Vovk et al., 2005; Shafer and Vovk, 2008; Angelopoulos and Bates, 2023; Angelopoulos et al., 2024)

Conformal prediction yields finite sample guarantees

- **Observation**: dataset $\mathcal{D}$ of n independent and identically distributed (i.i.d.) samples $(X_i, Y_i) \sim \mathbb{P}_{X,Y}$ with $(X_i, Y_i) \in \mathcal{X} \times \mathcal{Y}$
- **Goal**: Given a target miscoverage level $\alpha \in (0, 1)$, construct a set-valued predictor $C_\alpha(X)$ such that

$$\mathbb{P}_{X,Y}(Y_{n+1} \in C_\alpha(X_{n+1})) \approx 1 - \alpha$$

- **Conformal prediction allows finite sample guarantees** (Vovk et al., 2005; Shafer and Vovk, 2008; Angelopoulos and Bates, 2023; Angelopoulos et al., 2024)
- **Alternative frameworks**
  - Bayesian, resampling, PCS (predictability, computability, and stability), etc.

Proof of the finite sample guarantees

Probabilistic lemma (uniform rank statistics).
Let $n+1$ independent and identically distributed (i.i.d.) samples $S_i \sim \mathbb{P}_S$ with $S_i \in \mathbb{R}$.
Assume there are almost surely no ties between S_i.
Then $\mathrm{rank}(S_{n+1})$ is uniformly distributed in $[\![1, n+1]\!]$.

Proof of the finite sample guarantees

Probabilistic lemma (uniform rank statistics).
Let $n+1$ independent and identically distributed (i.i.d.) samples $S_i \sim \mathbb{P}_S$ with $S_i \in \mathbb{R}$.
Assume there are almost surely no ties between S_i.
Then $\mathrm{rank}(S_{n+1})$ is uniformly distributed in $[\![1, n+1]\!]$.

Proof: For $k \in [\![1, n+1]\!]$, because of exchangeability,

$$\mathbb{P}(\mathrm{rank}(S_{n+1}) = k) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbb{P}(\mathrm{rank}(S_i) = k) = \frac{1}{n+1}$$

Proof of the finite sample guarantees

Proposition. Let n i.i.d. samples $(X_i, Y_i) \sim \mathbb{P}_{X,Y}$, and a fixed $S : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.
Define $S_i = S(X_i, Y_i)$, and define order statistics as $S_{(1)} < S_{(2)} < \ldots < S_{(n)}$ (no ties a.s.).
For $r \leq n$, define $C_{1-r/(n+1)}(X) := \{y, S(X, y) \leq S_{(r)}\}$.
Then for a new test sample $(X_{n+1}, Y_{n+1}) \sim \mathbb{P}_{X,Y}$,

$$\mathbb{P}\Big(Y_{n+1} \in C_{1-\frac{r}{n+1}}(X_{n+1})\Big) = \frac{r}{n+1}.$$

Proof of the finite sample guarantees

Proposition. Let n i.i.d. samples $(X_i, Y_i) \sim \mathbb{P}_{X,Y}$, and a fixed $S : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.
Define $S_i = S(X_i, Y_i)$, and define order statistics as $S_{(1)} < S_{(2)} < \ldots < S_{(n)}$ (no ties a.s.).
For $r \leq n$, define $C_{1-r/(n+1)}(X) := \{y, S(X, y) \leq S_{(r)}\}$.
Then for a new test sample $(X_{n+1}, Y_{n+1}) \sim \mathbb{P}_{X,Y}$,

$$\mathbb{P}\Big(Y_{n+1} \in C_{1-\frac{r}{n+1}}(X_{n+1})\Big) = \frac{r}{n+1}.$$

Proof:

$$\begin{aligned}
\mathbb{P}\left(Y_{n+1} \in C_{1-\frac{r}{n+1}}(X_{n+1})\right) &= \mathbb{P}\left(S(X_{n+1}, Y_{n+1}) \leq S_{(r)} = r \text{ smallest of } S_1, \ldots, S_n\right) \\
&= \mathbb{P}\left(S_{n+1} = S(X_{n+1}, Y_{n+1}) \text{ among the } r \text{ smallest of } S_1, \ldots, S_n, S_{n+1}\right) \\
&= \mathbb{P}(\mathrm{rank}(S_i) \leq r) = \frac{r}{n+1}
\end{aligned}$$

Proof of the finite sample guarantees

Proposition. Let n i.i.d. samples $(X_i, Y_i) \sim \mathbb{P}_{X,Y}$, and a fixed $S : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.
Define $S_i = S(X_i, Y_i)$, and define order statistics as $S_{(1)} < S_{(2)} < \ldots < S_{(n)}$ (no ties a.s.).
For $r \leq n$, define $C_{1-r/(n+1)}(X) := \{y, S(X, y) \leq S_{(r)}\}$.
Then for a new test sample $(X_{n+1}, Y_{n+1}) \sim \mathbb{P}_{X,Y}$,

$$\mathbb{P}\Big(Y_{n+1} \in C_{1-\frac{r}{n+1}}(X_{n+1})\Big) = \frac{r}{n+1}.$$

Proof:

$$\begin{aligned}
\mathbb{P}\left(Y_{n+1} \in C_{1-\frac{r}{n+1}}(X_{n+1})\right) &= \mathbb{P}\left(S(X_{n+1}, Y_{n+1}) \leq S_{(r)} = r \text{ smallest of } S_1, \ldots, S_n\right) \\
&= \mathbb{P}\left(S_{n+1} = S(X_{n+1}, Y_{n+1}) \text{ among the } r \text{ smallest of } S_1, \ldots, S_n, S_{n+1}\right) \\
&= \mathbb{P}(\mathrm{rank}(S_i) \leq r) = \frac{r}{n+1}
\end{aligned}$$

By choosing $r = \lceil (1-\alpha)(n+1) \rceil$, we get

$$\mathbb{P}\left(Y_{n+1} \in C_\alpha(X_{n+1})\right) \in \left[1-\alpha,\ 1-\alpha+\frac{1}{n+1}\right)$$

Proof of the finite sample guarantees

Proposition. Let n i.i.d. samples $(X_i, Y_i) \sim \mathbb{P}_{X,Y}$, and a fixed $S : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.
Define $S_i = S(X_i, Y_i)$, and define order statistics as $S_{(1)} < S_{(2)} < \ldots < S_{(n)}$ (no ties a.s.).
For $r \leq n$, define $C_{1-r/(n+1)}(X) := \{y, S(X, y) \leq S_{(r)}\}$.
Then for a new test sample $(X_{n+1}, Y_{n+1}) \sim \mathbb{P}_{X,Y}$,

$$\mathbb{P}\Big(Y_{n+1} \in C_{1-\frac{r}{n+1}}(X_{n+1})\Big) = \frac{r}{n+1}.$$

Proof:

$$\begin{aligned}
\mathbb{P}\left(Y_{n+1} \in C_{1-\frac{r}{n+1}}(X_{n+1})\right) &= \mathbb{P}\left(S(X_{n+1}, Y_{n+1}) \leq S_{(r)} = r \text{ smallest of } S_1, \ldots, S_n\right) \\
&= \mathbb{P}\left(S_{n+1} = S(X_{n+1}, Y_{n+1}) \text{ among the } r \text{ smallest of } S_1, \ldots, S_n, S_{n+1}\right) \\
&= \mathbb{P}(\mathrm{rank}(S_i) \leq r) = \frac{r}{n+1} \quad \text{(NB: know more than probability)}
\end{aligned}$$

By choosing $r = \lceil (1-\alpha)(n+1) \rceil$, we get

$$\mathbb{P}\left(Y_{n+1} \in C_\alpha(X_{n+1})\right) \in \left[1-\alpha,\ 1-\alpha+\frac{1}{n+1}\right)$$

We get marginal guarantees and no conditional ones

- Given fixed "nonconformity score function" $S : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$,
  - $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1})) \approx 1 - \alpha$ ✓

We get marginal guarantees and no conditional ones

- Given fixed "nonconformity score function" $S : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$,
  - $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1})) \approx 1 - \alpha$ ✓
  - NB: Different from $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1}) | X_1, Y_1, \ldots, X_n, Y_n) \approx 1 - \alpha$ almost surely

We get marginal guarantees and no conditional ones

- Given fixed "nonconformity score function" $S : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$,
  - $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1})) \approx 1 - \alpha$ ✓
  - NB: Different from $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1}) | X_1, Y_1, \ldots, X_n, Y_n) \approx 1 - \alpha$ almost surely
  - $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1}) \mid X_{n+1}) \approx 1 - \alpha$ almost surely ✗

We get marginal guarantees and no conditional ones

- Given fixed "nonconformity score function" $S : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$,
  - $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1})) \approx 1 - \alpha$ ✓
  - NB: Different from $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1}) | X_1, Y_1, \ldots, X_n, Y_n) \approx 1 - \alpha$ almost surely
  - $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1}) \mid X_{n+1}) \approx 1 - \alpha$ almost surely ✗
- Impossible to get conditional validity without additional assumptions (see, e.g., Vovk 2012; Foygel Barber et al. 2021; Lei and Wasserman 2014)

We get marginal guarantees and no conditional ones

- Given fixed "nonconformity score function" $S : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$,
  - $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1})) \approx 1 - \alpha$ ✓
  - NB: Different from $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1}) | X_1, Y_1, \dots, X_n, Y_n) \approx 1 - \alpha$ almost surely
  - $\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1}) \mid X_{n+1}) \approx 1 - \alpha$ almost surely ✗
- Impossible to get conditional validity without additional assumptions (see, e.g., Vovk 2012; Foygel Barber et al. 2021; Lei and Wasserman 2014)
- Good conditional coverage $\Leftrightarrow$ score functions that "capture uncertainty" and have weak dependence on X
- Choice of score function is crucial
  - Any (even estimated) score can be "conformalized" using split conformal prediction
  - How to choose a good score?

Split conformal prediction

- Suppose we have n samples i.i.d. $(X_i, Y_i) \sim \mathbb{P}_{X,Y}$
- Split the dataset in two:
  - $\mathcal{D}_1$ = training set with $\text{Card}(\mathcal{D}_1) = n_1$
  - $\mathcal{D}_2$ = calibration set with $\text{Card}(\mathcal{D}_2) = n_2$
- Learn any model f on the training dataset

Split conformal prediction

- Suppose we have n samples i.i.d. $(X_i, Y_i) \sim \mathbb{P}_{X,Y}$
- Split the dataset in two:
  - $\mathcal{D}_1$ = training set with $\text{Card}(\mathcal{D}_1) = n_1$
  - $\mathcal{D}_2$ = calibration set with $\text{Card}(\mathcal{D}_2) = n_2$
- Learn any model f on the training dataset
- Define the nonconformity score:

$$S(X, Y) = |Y - f(X)|$$

- Let $\hat{q}_\alpha$ be the $\lceil (1-\alpha)(n_2+1) \rceil$-smallest value of all n_2 values of $S(X_i, Y_i)$ of the calibration set

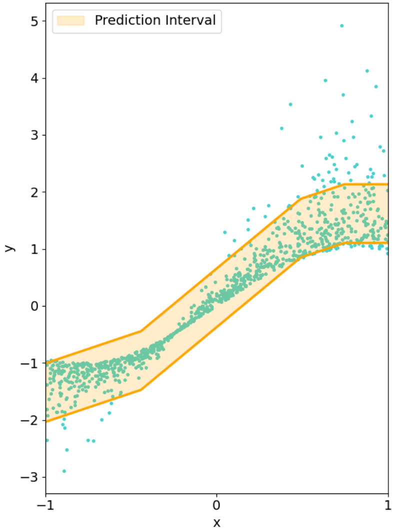
Split conformal prediction

- Suppose we have n samples i.i.d. $(X_i, Y_i) \sim \mathbb{P}_{X,Y}$
- Split the dataset in two:
  - $\mathcal{D}_1$ = training set with $\text{Card}(\mathcal{D}_1) = n_1$
  - $\mathcal{D}_2$ = calibration set with $\text{Card}(\mathcal{D}_2) = n_2$
- Learn any model f on the training dataset
- Define the nonconformity score:

$$S(X, Y) = |Y - f(X)|$$

- Let $\hat{q}_\alpha$ be the $\lceil (1-\alpha)(n_2+1) \rceil$-smallest value of all n_2 values of $S(X_i, Y_i)$ of the calibration set
- Then, with $C_\alpha(X) = \{y,\ f(X) - \hat{q}_\alpha \leq y \leq f(X) + \hat{q}_\alpha\}$,

$$\mathbb{P}(Y_{n+1} \in C_\alpha(X_{n+1}) \mid \mathcal{D}_1) \in \left[1-\alpha,\ 1-\alpha+\frac{1}{n_2+1}\right)$$



Improved conditional coverage with standardized residuals

- Also learn $\sigma(X)$, a "spread predictor" to predict the standard deviation of $|Y - f(X)|$
- Define the non-conformity score (Lei et al., 2018)

$$S(X, Y) = \frac{|Y - f(X)|}{\sigma(X)}$$

- Let $\hat{q}_\alpha$ be the $\lceil (1-\alpha)(n_2+1) \rceil$-smallest value of all n_2 values of $S(X_i, Y_i)$ of the calibration set

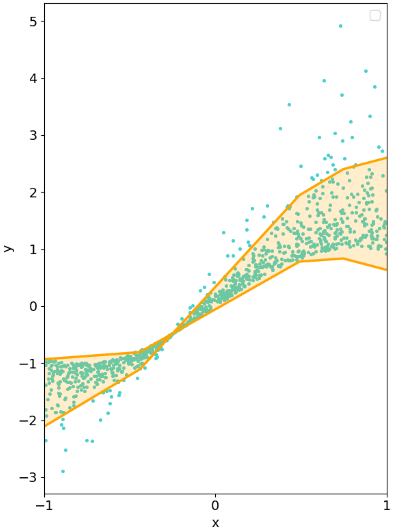
Improved conditional coverage with standardized residuals

- Also learn $\sigma(X)$, a "spread predictor" to predict the standard deviation of $|Y - f(X)|$
- Define the non-conformity score (Lei et al., 2018)

$$S(X, Y) = \frac{|Y - f(X)|}{\sigma(X)}$$

- Let $\hat{q}_\alpha$ be the $\lceil (1-\alpha)(n_2+1) \rceil$-smallest value of all n_2 values of $S(X_i, Y_i)$ of the calibration set
- Then, with $C_\alpha(X) = \{y,\ f(X) - \sigma(X)\hat{q}_\alpha \le y \le f(X) + \sigma(X)\hat{q}_\alpha\}$,

$$\mathbb{P}_{X,Y}(Y_{n+1} \in C_\alpha(X_{n+1}) \mid \mathcal{D}_1) \in \left[1-\alpha,\ 1-\alpha+\frac{1}{n_2+1}\right)$$



Quantile regression

- Learn $\hat{q}_{\alpha/2}(X)$ and $\hat{q}_{1-\alpha/2}(X)$ by minimizing the pinball loss on the training set, for $\tau = \alpha/2$ and $\tau = 1 - \alpha/2$:

$$L_\tau(y, \hat{y}) = \begin{cases} \tau(y - \hat{y}) & \text{if } y \geq \hat{y}, \\ (1-\tau)(\hat{y} - y) & \text{if } y < \hat{y}. \end{cases}$$

- Define the score (Romano et al., 2019)

$$S(X, Y) = \max\{\hat{q}_{\alpha/2}(X) - Y, Y - \hat{q}_{1-\alpha/2}(X)\}$$

Quantile regression

- Learn $\hat{q}_{\alpha/2}(X)$ and $\hat{q}_{1-\alpha/2}(X)$ by minimizing the pinball loss on the training set, for $\tau = \alpha/2$ and $\tau = 1 - \alpha/2$:

$$L_\tau(y, \hat{y}) = \begin{cases} \tau(y - \hat{y}) & \text{if } y \geq \hat{y}, \\ (1-\tau)(\hat{y} - y) & \text{if } y < \hat{y}. \end{cases}$$

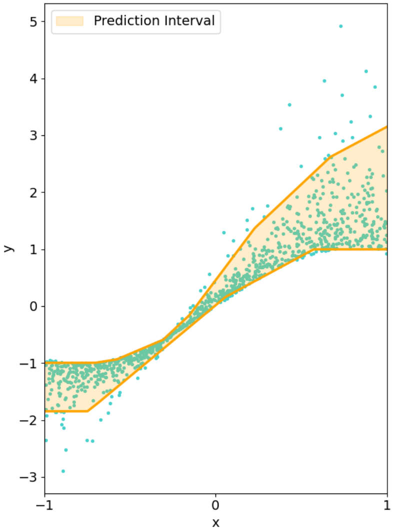
- Define the score (Romano et al., 2019)

$$S(X, Y) = \max\{\hat{q}_{\alpha/2}(X) - Y, Y - \hat{q}_{1-\alpha/2}(X)\}$$

- Let $\hat{q}_\alpha$ be the $\lceil (1-\alpha)(n_2+1) \rceil$-smallest value of all n_2 values of $S(X_i, Y_i)$ of the calibration set
- Then, with

$$C_\alpha(X) = \{y,\ \hat{q}_{\alpha/2}(X) - \hat{q}_\alpha \leq y \leq \hat{q}_{1-\alpha/2}(X) + \hat{q}_\alpha\},$$

$$\mathbb{P}_{X,Y}(Y_{n+1} \in C_\alpha(X_{n+1}) \mid \mathcal{D}_1) \in \left[1-\alpha,\ 1-\alpha+\frac{1}{n_2+1}\right)$$



Challenges in Conformal Prediction

- **Guarantees for conditional coverage** (Gibbs et al., 2025; Plassier et al., 2025)

Challenges in Conformal Prediction

- **Guarantees for conditional coverage** (Gibbs et al., 2025; Plassier et al., 2025)
- **Multivariate problems**
  - Choice of conformity score based on generalized ellipsoids
  - Trade-off size vs. miscoverage
  - Improving coverage and "flexibility"
  - Braun, Aolaritei, Berta, Jordan, Bach (2025)

Challenges in Conformal Prediction

- **Guarantees for conditional coverage** (Gibbs et al., 2025; Plassier et al., 2025)
- **Multivariate problems**
  - Choice of conformity score based on generalized ellipsoids
  - Trade-off size vs. miscoverage
  - Improving coverage and "flexibility"
  - Braun, Aolaritei, Berta, Jordan, Bach (2025)
- **Beyond i.i.d. (or exchangeability)**
  - Standard CP requires fixing the miscoverage level α and the batch size n in advance
  - How to handle adaptive goals (e.g., constraints on prediction set size) and sequential streams (anytime-validity)
  - Extensive use of *e*-values
  - Gauthier, Jordan, Bach (2025)

Challenges in Conformal Prediction

- **Guarantees for conditional coverage** (Gibbs et al., 2025; Plassier et al., 2025)
- **Multivariate problems**
  - Choice of conformity score based on generalized ellipsoids
  - Trade-off size vs. miscoverage
  - Improving coverage and "flexibility"
  - Braun, Aolaritei, Berta, Jordan, Bach (2025)
- **Beyond i.i.d. (or exchangeability)**
  - Standard CP requires fixing the miscoverage level α and the batch size n in advance
  - How to handle adaptive goals (e.g., constraints on prediction set size) and sequential streams (anytime-validity)
  - Extensive use of e-values
  - Gauthier, Jordan, Bach (2025)
- **Bonus:** How can we evaluate the conditional validity of a strategy?

Minimum volume covering sets

Minimum volume covering set (Braun et al., 2025a)

- **Goal**: Given a set of points $\{y_1, \dots, y_n\}$ in $\mathbb{R}^k$,
    - Find the smallest set that contains at least a fraction $1 - \alpha$ of points, $\alpha \in (0, 1)$

Minimum volume covering set (Braun et al., 2025a)

- **Goal**: Given a set of points $\{y_1, \ldots, y_n\}$ in $\mathbb{R}^k$,
  - Find the smallest set that contains at least a fraction $1 - \alpha$ of points, $\alpha \in (0, 1)$
- Define sets, with $M \in S_k^+$ (positive definite matrix), $\mu \in \mathbb{R}^k$, $\|\cdot\|$ a norm,

$$\mathbb{B}(\|\cdot\|, M, \mu) := \{y \in \mathbb{R}^k \mid \|M(y - \mu)\| \leq 1\}$$

Minimum volume covering set (Braun et al., 2025a)

- **Goal**: Given a set of points $\{y_1, \ldots, y_n\}$ in $\mathbb{R}^k$,
  - Find the smallest set that contains at least a fraction $1-\alpha$ of points, $\alpha \in (0,1)$
- Define sets, with $M \in S_k^+$ (positive definite matrix), $\mu \in \mathbb{R}^k$, $\|\cdot\|$ a norm,

$$\mathbb{B}(\|\cdot\|, M, \mu) := \{y \in \mathbb{R}^k \mid \|M(y-\mu)\| \leq 1\}$$

- **Formulation as optimization problem (MVCS)**:

$$\begin{aligned} \min \quad & \mathrm{Vol}(\mathbb{B}(\|\cdot\|, M, \mu)) = \lambda(B_{\|\cdot\|}(1)) \cdot \det(M)^{-1} \\ \text{s.t.} \quad & M \succcurlyeq 0,\ \mu \in \mathbb{R}^k, \\ & \mathrm{Card}\,\{i \in [n] \mid \|M(y_i - \mu)\| \leq 1\} \geq n - r + 1 \end{aligned} \tag{1}$$

Minimum volume covering set (Braun et al., 2025a)

Proposition (unconstrained formulation). Problem (MVCS) is equivalent to

$$\begin{aligned} \min \quad & -\log\det(\Lambda) + \sigma_r\left\{\|\Lambda y_i + \eta\|\right\} + \log\lambda(B_{\|\cdot\|}(1)) \\ \text{s.t.} \quad & \Lambda \succcurlyeq 0,\ \eta \in \mathbb{R}^k, \end{aligned}$$

where $\sigma_r\{a_i\}$ is the r-th largest element of a set $\{a_i\}_{i=1}^n$ with $a_i \in \mathbb{R}$.

Proof: Change of coordinates $\Lambda := \nu M$, for $\nu > 0$, and $\eta := -\Lambda\mu$

Minimum volume covering set (Braun et al., 2025a)

Proposition (unconstrained formulation). Problem (MVCS) is equivalent to

$$\begin{aligned} \min \quad & -\log\det(\Lambda) + \sigma_r\left\{\|\Lambda y_i + \eta\|\right\} + \log\lambda(B_{\|\cdot\|}(1)) \\ \text{s.t.} \quad & \Lambda \succcurlyeq 0,\ \eta \in \mathbb{R}^k, \end{aligned}$$

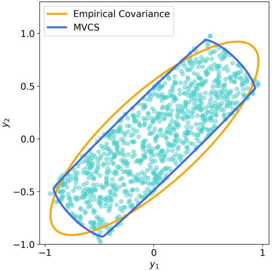
where $\sigma_r\{a_i\}$ is the r-th largest element of a set $\{a_i\}_{i=1}^n$ with $a_i \in \mathbb{R}$.

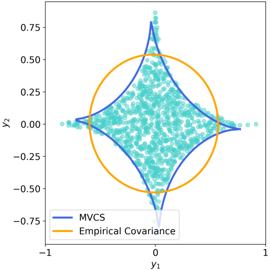
Proof: Change of coordinates $\Lambda := \nu M$, for $\nu > 0$, and $\eta := -\Lambda\mu$

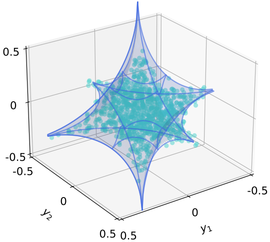
- Non-convex optimization problem, in fact NP-hard (Ahmadi et al., 2014)
- Convex relaxation akin to pinball loss: $\sigma_r\{a_i\} \leq$ sum of r largest elements
- Classical convex problem for $r = 1$ and $\|\cdot\| = \|\cdot\|_2$ (minimum-volume enclosing ellipsoid)

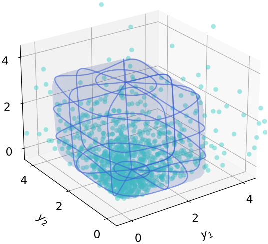
Shape obtained for different distributions of points

- $\|\cdot\| = \|\cdot\|_p$ does not need to be a norm (p can be less than one)









Minimum volume covering set (input / output)

Let $\mathbb{B}(M, \mu, p) := \{y \in \mathbb{R}^k \mid \|M(y-\mu)\|_p \leq 1\}$, where $M \succcurlyeq 0$, $\mu \in \mathbb{R}^k$, $p > 0$.

$$\begin{aligned} \min \quad & \mathbb{E}\left[\mathrm{Vol}(\mathbb{B}(M(X), f_\theta(X), p))\right] \\ \text{s.t.} \quad & \mathbb{P}\left(Y \in \mathbb{B}(p, M(X), f_\theta(X))\right) \geq 1-\alpha \end{aligned} \tag{MVCS-cond.}$$

Minimum volume covering set (input / output)

Let $\mathbb{B}(M, \mu, p) := \{y \in \mathbb{R}^k \mid \|M(y-\mu)\|_p \leq 1\}$, where $M \succcurlyeq 0$, $\mu \in \mathbb{R}^k$, $p > 0$.

$$\begin{aligned} \min \quad & \mathbb{E}\left[\mathrm{Vol}(\mathbb{B}(M(X), f_\theta(X), p))\right] \\ \text{s.t.} \quad & \mathbb{P}\left(Y \in \mathbb{B}(p, M(X), f_\theta(X))\right) \geq 1-\alpha \end{aligned} \qquad \text{(MVCS-cond.)}$$

Proposition Let $(X_i, Y_i)_{1 \leq i \leq n}$ be n i.i.d. samples, with $X_i \in \mathbb{R}^d$, $Y_i \in \mathbb{R}^k$.
Then (MVCS-cond.) is equivalent to:

$$\begin{aligned} \min \quad & \log\left(\sum_{i=1}^{n} \frac{1}{\det(\Lambda_\phi(x_i))}\right) + k \log \sigma_r \left\{\|\Lambda_\phi(x_i)(y_i - f_\theta(x_i))\|_p\right\} + \log \lambda(B_{\|\cdot\|_p}(1)) \\ \text{s.t.} \quad & \Lambda_\phi(\cdot) \succcurlyeq 0, \; p > 0, \; \theta, \phi \in \Theta. \end{aligned}$$

Minimum volume covering set (input / output)

Let $\mathbb{B}(M, \mu, p) := \{y \in \mathbb{R}^k \mid \|M(y-\mu)\|_p \leq 1\}$, where $M \succcurlyeq 0$, $\mu \in \mathbb{R}^k$, $p > 0$.

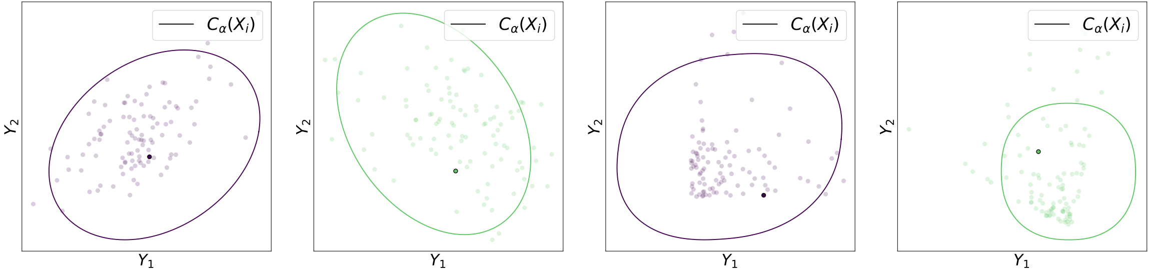
$$\begin{aligned} \min \quad & \mathbb{E}\left[\mathrm{Vol}(\mathbb{B}(M(X), f_\theta(X), p))\right] \\ \text{s.t.} \quad & \mathbb{P}\left(Y \in \mathbb{B}(p, M(X), f_\theta(X))\right) \geq 1 - \alpha \end{aligned} \qquad \text{(MVCS-cond.)}$$

Proposition Let $(X_i, Y_i)_{1 \leq i \leq n}$ be n i.i.d. samples, with $X_i \in \mathbb{R}^d$, $Y_i \in \mathbb{R}^k$.
Then (MVCS-cond.) is equivalent to:

$$\begin{aligned} \min \quad & \log\left(\sum_{i=1}^{n} \frac{1}{\det(\Lambda_\phi(x_i))}\right) + k \log \sigma_r \left\{\|\Lambda_\phi(x_i)(y_i - f_\theta(x_i))\|_p\right\} + \log \lambda(B_{\|\cdot\|_p}(1)) \\ \text{s.t.} \quad & \Lambda_\phi(\cdot) \succcurlyeq 0,\ p > 0,\ \theta, \phi \in \Theta. \end{aligned}$$

- The conformalisation step is made with the score $S(X, Y) = \|\Lambda_\phi(X)(Y - f_\theta(X))\|_p$
- Only marginal coverage

Empirical illustration for different distributions



Left: Gaussian noise | Right: exponential noise

Samples from the distribution of $Y|X \sim f(X) + B(X)$ where $B(X)$ is an heteroskedastic noise

Gaussian conformal prediction

Extending standardized residuals to multivariate regression

Assume

$$Y|X \sim \mathcal{N}(f(X), \Sigma(X)),$$

where $Y \in \mathbb{R}^k$. We can define the Mahalanobis score (Braun et al., 2025b)

$$S_{\mathrm{Mah}}(X, Y) = \left\|\Sigma(X)^{-1/2}(Y - f(X))\right\|_2$$

- Generalization of the standardized residuals (Lei et al., 2018) in multivariate regression
- Under the Gaussian assumption, $S_{\mathrm{Mah}}(X, Y)|X \sim \chi_2(k)$ which is independent of X

Extending standardized residuals to multivariate regression

Assume

$$Y|X \sim \mathcal{N}(f(X), \Sigma(X)),$$

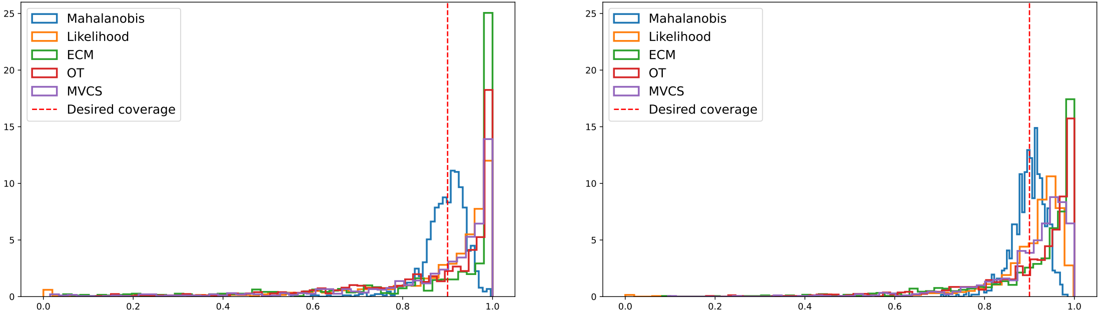
where $Y \in \mathbb{R}^k$. We can define the Mahalanobis score (Braun et al., 2025b)

$$S_{\mathrm{Mah}}(X, Y) = \left\|\Sigma(X)^{-1/2}(Y - f(X))\right\|_2$$

- Generalization of the standardized residuals (Lei et al., 2018) in multivariate regression
- Under the Gaussian assumption, $S_{\mathrm{Mah}}(X, Y)|X \sim \chi_2(k)$ which is independent of X
- Even if the distribution is not Gaussian, we expect our score to be less dependent on X
  - Let $f(X) = \mathbb{E}[Y|X]$ and $\Sigma(X) = \mathrm{cov}(Y|X)$, and $U = \Sigma(X)^{-1/2}(Y - f(X))$
  - Then $\mathbb{E}[U|X] = 0$ and $\mathrm{cov}(U|X) = I_k$ so the distribution of $S_{\mathrm{Mah}}(X, Y) = \|U\|_2$ should less depend on X
- NB: all developments can be extended to elliptical distributions

Improved Conditional Coverage

- Modeling via ellipsoids ⇒ estimation of variance along all directions



Histograms of conditional coverage $\mathbb{P}(Y \in C_\alpha(X_i) \mid X_i)$ on a synthetic dataset ($\alpha = 0.1$)
(left) heteroskedastic Gaussian noise (right) heteroskedastic exponential noise

- Baselines: empirical covariance matrix (Johnstone and Cox, 2021), optimal transport (Thurin et al., 2025)

Partially revealed outputs: motivation

- **Example**: predict fasting blood glucose (Y^1) and cholesterol (Y^2)
  - If conformal set too wide, a costly test (e.g., Y^1) may be performed
  - Once Y^1 is revealed, we want to **refine prediction for** Y^2
  - Standard approach: retrain model with $(X, Y^1) \Rightarrow$ expensive

Partially revealed outputs: motivation

- **Example**: predict fasting blood glucose (Y^1) and cholesterol (Y^2)
  - If conformal set too wide, a costly test (e.g., Y^1) may be performed
  - Once Y^1 is revealed, we want to **refine prediction for** Y^2
  - Standard approach: retrain model with $(X, Y^1) \Rightarrow$ expensive
- **Goal**: **update conformal sets directly** using the probabilistic model
  - Leverage classical conditioning properties of the Gaussian distribution

Method: Gaussian conditioning

- Assuming a Gaussian predictive model:

$$\hat{p}(Y|X) = \mathcal{N}\left(\begin{pmatrix} f^r(X) \\ f^h(X) \end{pmatrix}, \begin{pmatrix} \Sigma^{rr}(X) & \Sigma^{rh}(X) \\ \Sigma^{hr}(X) & \Sigma^{hh}(X) \end{pmatrix}\right)$$

- Partition output into revealed Y^r and hidden Y^h

Method: Gaussian conditioning

- Assuming a Gaussian predictive model:

$$\hat{p}(Y|X) = \mathcal{N}\left(\begin{pmatrix} f^r(X) \\ f^h(X) \end{pmatrix}, \begin{pmatrix} \Sigma^{rr}(X) & \Sigma^{rh}(X) \\ \Sigma^{hr}(X) & \Sigma^{hh}(X) \end{pmatrix}\right)$$

- Partition output into revealed Y^r and hidden Y^h
- Conditional distribution (closed form):

$$\hat{p}(Y^h|X, Y^r) = \mathcal{N}\left(\tilde{f}(X), \tilde{\Sigma}(X)\right)$$

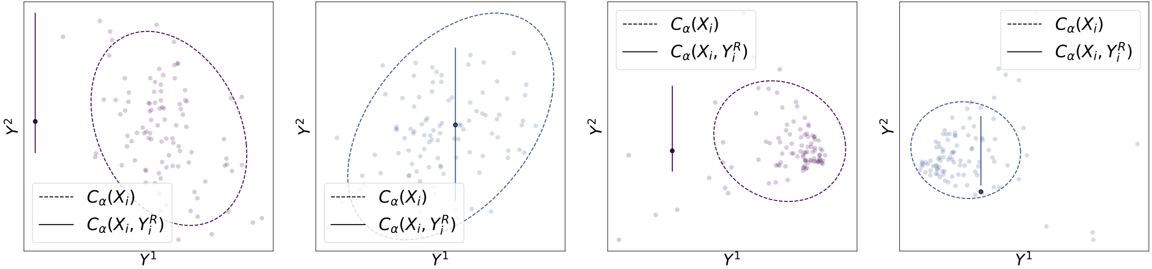
$$\tilde{f}(X) = f^h(X) + \Sigma^{hr}{\Sigma^{rr}}^{-1}(Y^r - f^r(X)), \quad \tilde{\Sigma}(X) = \Sigma^{hh} - \Sigma^{hr}{\Sigma^{rr}}^{-1}\Sigma^{rh}$$

- Non-conformity score:

$$S_{\text{Revealed}}(X, Y^r, Y^h) = \|\tilde{\Sigma}(X)^{-1/2}(Y^h - \tilde{f}(X))\|_2$$

Illustration of benefits

- Uses revealed Y^r without retraining
- Produces **narrower, adaptive sets**, and avoids empty sets (robustness)



Left: Gaussian noise

Right: exponential noise.

Different samples from the distribution of $Y|X \sim f(X) + B(X)$
($B(X)$=heteroskesdatic noise and Y^1 is revealed)

Projection of the output: motivation

- High-dimensional predictions: $Y \in \mathbb{R}^k$
- Often (at test time) only a subset or linear combination $\varphi(Y)$ is of interest
- Example:
  - Finance: portfolio returns $(R^1, R^2) = (rY^1, pY^2 + qY^3)$.

Projection of the output: motivation

- High-dimensional predictions: $Y \in \mathbb{R}^k$
- Often (at test time) only a subset or linear combination $\varphi(Y)$ is of interest
- Example:
  - Finance: portfolio returns $(R^1, R^2) = (rY^1, pY^2 + qY^3)$.
- **Naive approach**: project conformal set $\tilde{C}_\alpha(X)$ onto $\varphi(Y)$
  $\Rightarrow$ conservative (overly large) sets
- **Goal**: construct **direct conformal sets for** $\varphi(Y)$ without retraining

Method: linear transformations

- Trained Gaussian model:
$$\hat{p}(Y|X) = \mathcal{N}(f(X), \Sigma(X))$$
- For linear $\varphi(Y) = MY$ $(M \in \mathbb{R}^{p \times k})$:
$$\hat{p}(\varphi(Y)|X) = \mathcal{N}(Mf(X),\ M\Sigma(X)M^\top)$$

Method: linear transformations

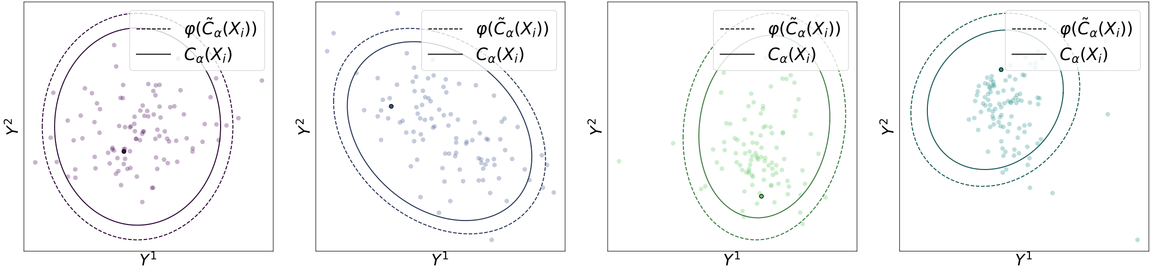
- Trained Gaussian model:
$$\hat{p}(Y|X) = \mathcal{N}(f(X), \Sigma(X))$$
- For linear $\varphi(Y) = MY$ ($M \in \mathbb{R}^{p \times k}$):
$$\hat{p}(\varphi(Y)|X) = \mathcal{N}(Mf(X),\ M\Sigma(X)M^\top)$$
- Non-conformity score in transformed space:
$$S_{\text{Lin.Trans.}}(X, Y) = \|(M\Sigma(X)M^\top)^{-1/2}(MY - Mf(X))\|_2$$
- Coverage guarantee:
$$\mathbb{P}(MY \in C_\alpha(X)) \in \left[1 - \alpha,\ 1 - \alpha + \tfrac{1}{n_2+1}\right)$$

Results: projected vs. direct sets

- Direct conformalization with $S_{\text{Lin.Trans.}}$ (full line) gives:
  - **Tighter sets**, correct coverage (if well-specified)
  - Better adaptation to heteroskedasticity
- Projecting $\tilde{C}_\alpha(X)$ (dashed line) gives **conservative sets**

Results: projected vs. direct sets

- Direct conformalization with $S_{\text{Lin. Trans.}}$ (full line) gives:
  - **Tighter sets**, correct coverage (if well-specified)
  - Better adaptation to heteroskedasticity
- Projecting $\tilde{C}_\alpha(X)$ (dashed line) gives **conservative sets**



Left: Gaussian noise Right: exponential noise

Projection of samples from the distribution of $Y|X \sim f(X) + B(X)$
($B(X)$=heteroskedastic noise)

Conformal prediction for multivariate predictions

- **Minimum volume "ellipsoid"**
  - Explicit optimization problem (for marginal coverage only)
  - Beyond ellipsoids

Conformal prediction for multivariate predictions

- **Minimum volume “ellipsoid”**
  - Explicit optimization problem (for marginal coverage only)
  - Beyond ellipsoids
- **Elliptical conditional model**
  - Allow conditioning and linear projection without retraining
  - Added flexibility

Conformal Prediction with E-values

Standard conformal prediction: a p-value perspective

Proposition (reformulation). Under the exchangeability assumption,

$$P = \frac{1 + \sum_{i=1}^{n} \mathbb{1}\{S(X_i, Y_i) \geq S(X_{n+1}, Y_{n+1})\}}{n+1}$$

is a p-value: $\forall \alpha$, $\mathbb{P}(P \leq \alpha) \leq \alpha$.

Standard conformal prediction: a p-value perspective

Proposition (reformulation). Under the exchangeability assumption,

$$P = \frac{1 + \sum_{i=1}^{n} \mathbb{1}\{S(X_i, Y_i) \geq S(X_{n+1}, Y_{n+1})\}}{n+1}$$

is a p-value: $\forall \alpha$, $\mathbb{P}(P \leq \alpha) \leq \alpha$.

Proof:

- The quantity $\sum_{i=1}^{n} \mathbb{1}\{S(X_i, Y_i) \geq S(X_{n+1}, Y_{n+1})\}$ is $n - R$ where R is the rank of $S(X_{n+1}, Y_{n+1})$ among $S_1, \ldots, S_n$
- Thus, $P = 1 - \frac{R}{n+1}$
- From previous guarantees, $\mathbb{P}(P \leq \alpha) = \mathbb{P}(1 - \frac{R}{n+1} \leq \alpha) \in (\alpha - \frac{1}{n+1}, \alpha]$

Limitations of the p-value framework

- Reformulation as *p*-value

$$\mathbb{P}(Y_{n+1} \notin C_\alpha(X_{n+1})) \leq \alpha \quad \Longleftrightarrow \quad \mathbb{P}(P \leq \alpha) \leq \alpha$$

Limitations of the p-value framework

- Reformulation as *p*-value

$$\mathbb{P}(Y_{n+1} \notin C_\alpha(X_{n+1})) \leq \alpha \quad \Longleftrightarrow \quad \mathbb{P}(P \leq \alpha) \leq \alpha$$

- Rigid assumptions:
  - **Fixed α:** The miscoverage level α must be fixed *before* observing the data
  - **Fixed batch:** It assumes a fixed batch of exchangeable data. In sequential settings, continuously monitoring p-values leads to validity issues

Limitations of the p-value framework

- Reformulation as *p*-value

$$\mathbb{P}(Y_{n+1} \notin C_\alpha(X_{n+1})) \leq \alpha \quad \Longleftrightarrow \quad \mathbb{P}(P \leq \alpha) \leq \alpha$$

- Rigid assumptions:
  - **Fixed α:** The miscoverage level α must be fixed *before* observing the data
  - **Fixed batch:** It assumes a fixed batch of exchangeable data. In sequential settings, continuously monitoring p-values leads to validity issues

- To enable **post-hoc** choices of α and **anytime-valid** monitoring
  $\Rightarrow$ Need a more flexible evidence measure: **e-values**
  (Grünwald et al., 2020; Wasserman et al., 2020; Shafer, 2021; Vovk and Wang, 2021)

Post-hoc guarantees and e-values

Standard CP (fixed α):

The standard validity $\mathbb{P}(Y_{n+1} \notin C_\alpha(X_{n+1})) = \mathbb{P}(P \leq \alpha) \leq \alpha$ for all α is equivalent to:

$$\sup_{\alpha} \mathbb{E}\left[\frac{\mathbb{1}\{Y_{n+1} \notin C_\alpha(X_{n+1})\}}{\alpha}\right] \leq 1$$

Post-hoc guarantees and e-values

Standard CP (fixed α):
The standard validity $\mathbb{P}(Y_{n+1} \notin C_\alpha(X_{n+1})) = \mathbb{P}(P \leq \alpha) \leq \alpha$ for all α is equivalent to:

$$\sup_{\alpha} \mathbb{E}\left[\frac{\mathbb{1}\{Y_{n+1} \notin C_\alpha(X_{n+1})\}}{\alpha}\right] \leq 1$$

Post-hoc CP (data-dependent $\tilde{\alpha}$):
To enable data-dependent decision making, we seek a stronger **post-hoc guarantee**:

$$\sup_{\tilde{\alpha}} \mathbb{E}\left[\frac{\mathbb{P}(\{Y_{n+1} \notin C_{\tilde{\alpha}}(X_{n+1})|\tilde{\alpha})}{\tilde{\alpha}}\right] = \sup_{\tilde{\alpha}} \mathbb{E}\left[\frac{\mathbb{1}\{Y_{n+1} \notin C_{\tilde{\alpha}}(X_{n+1})\}}{\tilde{\alpha}}\right] = \mathbb{E}\left[\sup_{\alpha} \frac{\mathbb{1}\{Y_{n+1} \notin C_\alpha(X_{n+1})\}}{\alpha}\right] \leq 1$$

Post-hoc guarantees and e-values

Standard CP (fixed α):

The standard validity $\mathbb{P}(Y_{n+1} \notin C_\alpha(X_{n+1})) = \mathbb{P}(P \le \alpha) \le \alpha$ for all α is equivalent to:

$$\sup_{\alpha} \mathbb{E}\left[\frac{\mathbb{1}\{Y_{n+1} \notin C_\alpha(X_{n+1})\}}{\alpha}\right] \le 1$$

Post-hoc CP (data-dependent $\tilde{\alpha}$):

To enable data-dependent decision making, we seek a stronger **post-hoc guarantee**:

$$\sup_{\tilde{\alpha}} \mathbb{E}\left[\frac{\mathbb{P}(\{Y_{n+1} \notin C_{\tilde{\alpha}}(X_{n+1})|\tilde{\alpha})}{\tilde{\alpha}}\right] = \sup_{\tilde{\alpha}} \mathbb{E}\left[\frac{\mathbb{1}\{Y_{n+1} \notin C_{\tilde{\alpha}}(X_{n+1})\}}{\tilde{\alpha}}\right] = \mathbb{E}\left[\sup_{\alpha} \frac{\mathbb{1}\{Y_{n+1} \notin C_\alpha(X_{n+1})\}}{\alpha}\right] \le 1$$

This requires the concept of **post-hoc p-values** and **e-values**.

Definitions (Ramdas and Wang, 2025)

- An **e-value** E is a non-negative random variable such that $\mathbb{E}[E] \le 1$
- A **post-hoc p-value** P satisfies the post-hoc guarantee above: $\mathbb{E}\left[\sup_\alpha \mathbb{1}\{P \le \alpha\}/\alpha\right] \le 1$

Theorem (Koning, 2024): Post-hoc p-values are exactly the inverses of e-values ($P = 1/E$).

Conformal prediction with adaptive coverage

Here, the e-value is the *soft-rank* e-value (Wang and Ramdas, 2022; Koning, 2025):

$$E = \frac{(n+1)S(X_{n+1}, Y_{n+1})}{\sum_{i=1}^{n} S(X_i, Y_i) + S(X_{n+1}, Y_{n+1})} = \frac{S(X_{n+1}, Y_{n+1})}{\frac{1}{n+1}\sum_{i=1}^{n} S(X_i, Y_i) + \frac{1}{n+1}S(X_{n+1}, Y_{n+1})}$$

Conformal prediction with adaptive coverage

Adaptive Coverage Guarantee (Gauthier et al., 2025b)

Let $\tilde{\alpha}$ be any (possibly data-dependent) miscoverage level that can depend on the calibration data $\{(X_i, Y_i)\}_{i=1}^n$ and the test feature X_{n+1}. Then we have:

$$\mathbb{E}\left[\frac{\mathbb{1}\{Y_{n+1} \notin \mathcal{C}_{\tilde{\alpha}}(X_{n+1})\}}{\tilde{\alpha}}\right] \leq 1,$$

where the prediction set $\mathcal{C}_{\tilde{\alpha}}$ is defined as:

$$\mathcal{C}_{\tilde{\alpha}}(X_{n+1}) = \left\{y \in \mathcal{Y} : \frac{(n+1)S(X_{n+1}, y)}{\sum_{i=1}^n S(X_i, Y_i) + S(X_{n+1}, y)} < \frac{1}{\tilde{\alpha}}\right\}.$$

Here, the e-value is the *soft-rank* e-value (Wang and Ramdas, 2022; Koning, 2025):

$$E = \frac{(n+1)S(X_{n+1}, Y_{n+1})}{\sum_{i=1}^n S(X_i, Y_i) + S(X_{n+1}, Y_{n+1})} = \frac{S(X_{n+1}, Y_{n+1})}{\frac{1}{n+1}\sum_{i=1}^n S(X_i, Y_i) + \frac{1}{n+1}S(X_{n+1}, Y_{n+1})}$$

Application: backward conformal prediction (Gauthier et al., 2025a)

- **Motivation:** Standard CP fixes coverage and lets size vary. Real-world systems often have physical constraints that dictate a size budget k, which may depend on the data

Application: backward conformal prediction (Gauthier et al., 2025a)

- **Motivation:** Standard CP fixes coverage and lets size vary. Real-world systems often have physical constraints that dictate a size budget k, which may depend on the data
- **Example:** In a recommender system, the number of items k we display depends on:
  - **Device type (X_{n+1}):** mobile (small screen) vs. desktop (large screen)
  - **Calibration uncertainty ($\mathcal{D}_{cal}$):** high variance $\implies$ increase the budget k

Application: backward conformal prediction (Gauthier et al., 2025a)

- **Motivation:** Standard CP fixes coverage and lets size vary. Real-world systems often have physical constraints that dictate a size budget k, which may depend on the data
- **Example:** In a recommender system, the number of items k we display depends on:
  - **Device type (X_{n+1}):** mobile (small screen) vs. desktop (large screen)
  - **Calibration uncertainty ($\mathcal{D}_{cal}$):** high variance $\implies$ increase the budget k
- **Backward CP**: fix size $k(X_{n+1}, \mathcal{D}_{cal})$, and adjust miscoverage:

$$\tilde{\alpha} = \inf \left\{ \alpha \in (0,1) : |\mathcal{C}_\alpha(X_{n+1})| \leq k(X_{n+1}, \mathcal{D}_{cal}) \right\}.$$

- **Guarantee**: $\mathbb{E}[\mathbb{1}\{Y_{n+1} \notin C_{\tilde{\alpha}}(X_{n+1})\}/\tilde{\alpha}] \leq 1 \quad \implies \quad \mathbb{P}(Y_{n+1} \in C_{\tilde{\alpha}}(X_{n+1})) \gtrapprox 1 - \mathbb{E}[\tilde{\alpha}]$

Estimating expected miscoverage $\mathbb{E}[\tilde{\alpha}]$

Problem: Coverage depends on unknown $\mathbb{E}[\tilde{\alpha}]$

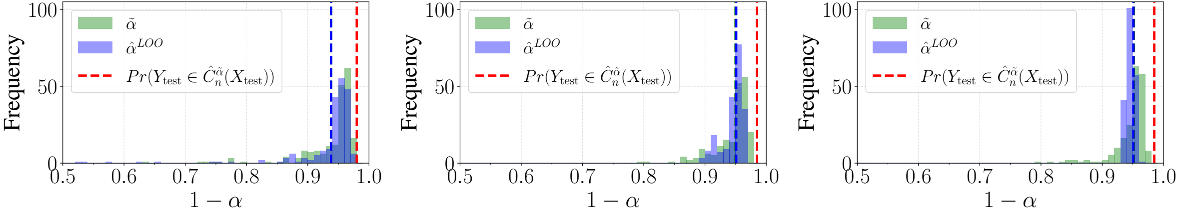
Solution: Estimate via **Leave-One-Out (LOO)** on calibration set

LOO Estimator: For each $i = 1, \ldots, n$:

1. Treat (X_i, Y_i) as pseudo-test point
2. Compute $\tilde{\alpha}_i$ using remaining data $\mathcal{D}_{cal} \setminus \{(X_i, Y_i)\}$

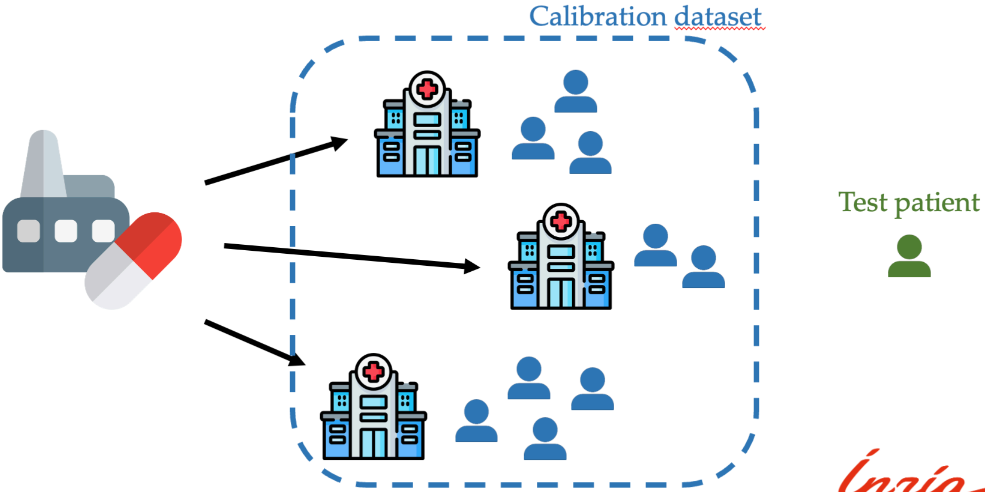
$$\widehat{\alpha}_{\text{LOO}} = \frac{1}{n} \sum_{i=1}^{n} \tilde{\alpha}_i$$

Guarantee: $|\widehat{\alpha}_{\text{LOO}} - \mathbb{E}[\tilde{\alpha}]| = O_P(1/\sqrt{n})$, $\quad \text{Var}(\widehat{\alpha}_{\text{LOO}}) = O_P(1/n)$.

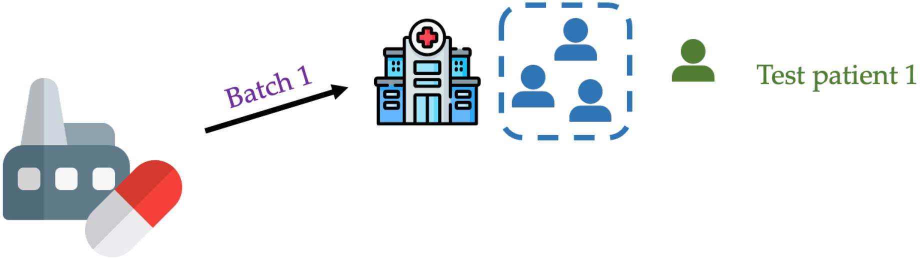


$n = 100$ $\quad$ $n = 1000$ $\quad$ $n = 5000$

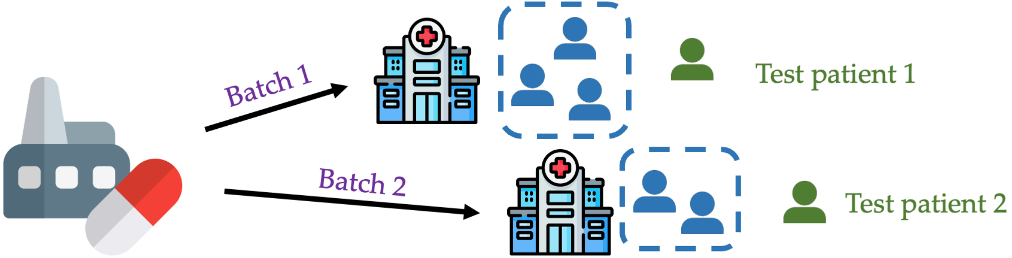
Anytime-Valid Guarantees with E-values (Gauthier et al., 2025b)



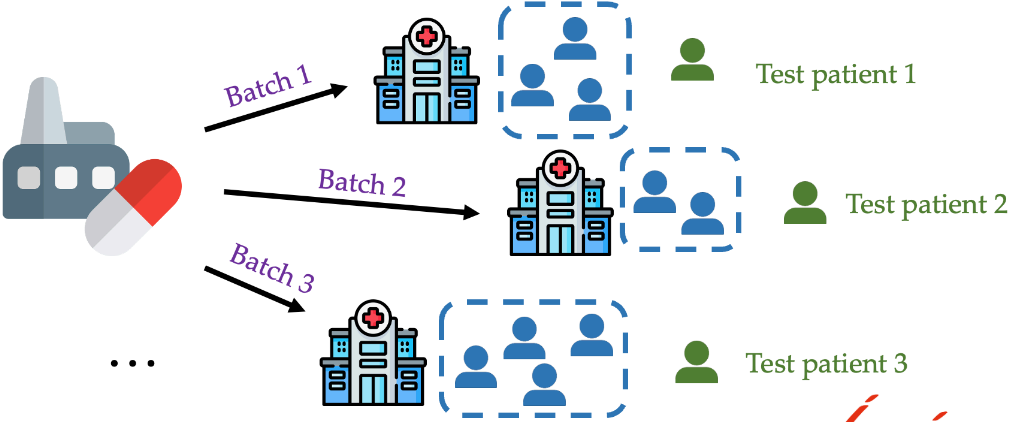
Anytime-Valid Guarantees with E-values (Gauthier et al., 2025b)



Anytime-Valid Guarantees with E-values (Gauthier et al., 2025b)



Anytime-Valid Guarantees with E-values (Gauthier et al., 2025b)



Inria

Anytime-Valid Guarantees with E-values (Gauthier et al., 2025b)

Insight: Beyond post-hoc guarantees, e-values naturally enable anytime-valid coverage

Anytime-Valid Guarantees with E-values (Gauthier et al., 2025b)

Insight: Beyond post-hoc guarantees, e-values naturally enable anytime-valid coverage

Scenario: We perform conformal prediction on multiple data batches sequentially. At each step t, we process a batch (where we aim to cover the target $Y^t_{n_t+1}$):

$$\{(X^t_1, Y^t_1), \ldots, (X^t_{n_t}, Y^t_{n_t}), (X^t_{n_t+1}, Y^t_{n_t+1})\}$$

Problem: Standard CP validity holds marginally for a fixed batch

Anytime-Valid Guarantees with E-values (Gauthier et al., 2025b)

Insight: Beyond post-hoc guarantees, e-values naturally enable anytime-valid coverage

Scenario: We perform conformal prediction on multiple data batches sequentially. At each step t, we process a batch (where we aim to cover the target $Y^t_{n_t+1}$):

$$\{(X^t_1, Y^t_1), \dots, (X^t_{n_t}, Y^t_{n_t}), (X^t_{n_t+1}, Y^t_{n_t+1})\}$$

Problem: Standard CP validity holds marginally for a fixed batch

Solution: Anytime-validity with e-values

- Construct a martingale $(M_t)_{t \geq 0}$ accumulating evidence over batches:

$$M_t = \prod_{s=1}^{t} (1 - \beta_s + \beta_s E_b), \quad \text{where } E_s = \frac{(n_s + 1) S(X^s_{n_s+1}, Y^s_{n_s+1})}{\sum_{i=1}^{n_s} S(X^s_i, Y^s_i) + S(X^s_{n_s+1}, Y^s_{n_s+1})}$$

Anytime-Valid Guarantees with E-values (Gauthier et al., 2025b)

Insight: Beyond post-hoc guarantees, e-values naturally enable anytime-valid coverage

Scenario: We perform conformal prediction on multiple data batches sequentially. At each step t, we process a batch (where we aim to cover the target $Y^t_{n_t+1}$):

$$\{(X^t_1, Y^t_1), \ldots, (X^t_{n_t}, Y^t_{n_t}), (X^t_{n_t+1}, Y^t_{n_t+1})\}$$

Problem: Standard CP validity holds marginally for a fixed batch

Solution: Anytime-validity with e-values

- Construct a martingale $(M_t)_{t\geq 0}$ accumulating evidence over batches:

$$M_t = \prod_{s=1}^{t}(1 - \beta_s + \beta_s E_b), \quad \text{where } E_s = \frac{(n_s + 1)S(X^s_{n_s+1}, Y^s_{n_s+1})}{\sum_{i=1}^{n_s} S(X^s_i, Y^s_i) + S(X^s_{n_s+1}, Y^s_{n_s+1})}$$

- **Ville's Inequality for supermartingales** $\mathbb{P}(\sup_t M_t \geqslant \alpha) \leqslant \mathbb{E}[M_0]/\alpha$

Anytime-Valid Guarantees with E-values (Gauthier et al., 2025b)

Insight: Beyond post-hoc guarantees, e-values naturally enable anytime-valid coverage

Scenario: We perform conformal prediction on multiple data batches sequentially. At each step t, we process a batch (where we aim to cover the target $Y^t_{n_t+1}$):

$$\{(X^t_1, Y^t_1), \ldots, (X^t_{n_t}, Y^t_{n_t}), (X^t_{n_t+1}, Y^t_{n_t+1})\}$$

Problem: Standard CP validity holds marginally for a fixed batch

Solution: Anytime-validity with e-values

- Construct a martingale $(M_t)_{t\geq 0}$ accumulating evidence over batches:

$$M_t = \prod_{s=1}^{t}(1-\beta_s+\beta_s E_b), \quad \text{where } E_s = \frac{(n_s+1)S(X^s_{n_s+1}, Y^s_{n_s+1})}{\sum_{i=1}^{n_s} S(X^s_i, Y^s_i) + S(X^s_{n_s+1}, Y^s_{n_s+1})}$$

- **Ville's Inequality for supermartingales** $\mathbb{P}(\sup_t M_t \geqslant \alpha) \leqslant \mathbb{E}[M_0]/\alpha$, ensures the coverage holds *simultaneously* for all time steps t (provided that the data in each batch is exchangeable conditional on the history:

$$\mathbb{P}(\forall t, Y_{n_t+1} \in C_t(X_{n_t+1})) \geq 1-\alpha$$

Bonus: Evaluating conditional miscoverage

Excess risk of the target coverage

- For any data-dependent set C_α, let $Z = \mathbb{1}\{Y \in C_\alpha(X)\} \in \{0,1\}$
- Perfect conditional coverage if $p(X) := \mathbb{P}(Y \in C_\alpha(X)|X) = \mathbb{P}(Z = 1|X) = 1 - \alpha$ a.s.

Excess risk of the target coverage

- For any data-dependent set C_α, let $Z = \mathbb{1}\{Y \in C_\alpha(X)\} \in \{0, 1\}$
- Perfect conditional coverage if $p(X) := \mathbb{P}(Y \in C_\alpha(X)|X) = \mathbb{P}(Z = 1|X) = 1 - \alpha$ a.s.
- **Key idea**: Estimating $\mathbb{P}(Z = 1|X = x)$ is a binary classification problem
  - Classification task equivalent to minimizing the risk for a proper score ℓ

$$\mathcal{R}_\ell(h) = \mathbb{E}[\ell(h(X), Z)]$$

  - Minimizer of $\mathcal{R}_\ell$ sastisfies $h^*(X) \in \arg\min_{q\in[0,1]} \mathbb{E}[\ell(q, Z) \mid X] = \mathbb{E}[Z \mid X] = p(X)$

Excess risk of the target coverage

- For any data-dependent set C_α, let $Z = \mathbb{1}\{Y \in C_\alpha(X)\} \in \{0, 1\}$
- Perfect conditional coverage if $p(X) := \mathbb{P}(Y \in C_\alpha(X)|X) = \mathbb{P}(Z = 1|X) = 1 - \alpha$ a.s.

- **Key idea**: Estimating $\mathbb{P}(Z = 1|X = x)$ is a binary classification problem
  - Classification task equivalent to minimizing the risk for a proper score ℓ

$$\mathcal{R}_\ell(h) = \mathbb{E}[\ell(h(X), Z)]$$

  - Minimizer of $\mathcal{R}_\ell$ sastisfies $h^*(X) \in \arg\min_{q \in [0,1]} \mathbb{E}[\ell(q, Z) \mid X] = \mathbb{E}[Z \mid X] = p(X)$

- **"Excess risk of target coverage"** (Braun et al., 2025c)

$$\ell\text{-ERT} = \mathcal{R}_\ell(1 - \alpha) - \mathcal{R}_\ell(p) = \mathcal{R}_\ell(1 - \alpha) - \inf_h \mathcal{R}_\ell(h)$$

  - Provides a measure of miscoverage and a detailed diagnostic of over/under coverage
  - Can be estimated with standard tools from non-parametric estimation

Benefits: a set of interpretable metrics

- Property of excess risk for a proper loss

$$\ell\text{-ERT} = \mathcal{R}_\ell(1-\alpha) - \mathcal{R}_\ell(p) = \mathbb{E}_X[d_\ell(1-\alpha, p(X))]$$

where d_ℓ is the associated Bregman divergence

Benefits: a set of interpretable metrics

- Property of excess risk for a proper loss

$$\ell\text{-ERT} = \mathcal{R}_\ell(1-\alpha) - \mathcal{R}_\ell(p) = \mathbb{E}_X[d_\ell(1-\alpha, p(X))]$$

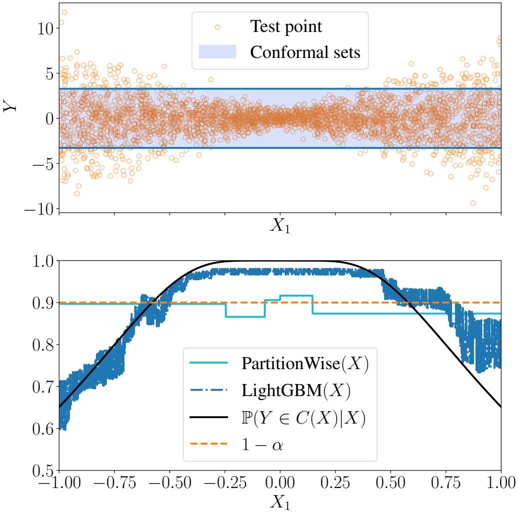
where d_ℓ is the associated Bregman divergence

- Examples of proper scoring rules and their associated ERT scores

Name	**Proper score** $\ell(p, y)$	ℓ-ERT
L_1-ERT	$\text{sgn}(p-(1-\alpha))(1-\alpha-y)$	$\mathbb{E}_X[\|1-\alpha-p(X)\|]$
L_2-ERT (Brier score)	$(y-p)^2$	$\mathbb{E}_X[(1-\alpha-p(X))^2]$
KL-ERT (Log loss)	$-\log p_y$	$\mathbb{E}_X[D_{\mathrm{KL}}(p(X)\|1-\alpha)]$

Metrics comparison without conditional coverage

- $X \sim \mathcal{U}\left([-1,1]^8\right)$
- $Y \sim \mathcal{N}(0, \sigma(X_1))$
- Sets are obtained by doing conformal prediction with the score $S(X, Y) = |Y|$
- L_1-ERT(LightGBM) $= 0.09$ (boosting with decision trees)
- L_1-ERT(PartitionWise) $= 0.01$ (a classical diagnostic measure)



Conclusion

Conclusion - Challenges in conformal prediction

- **Multivariate problems**
  - Choice of conformity score based on generalized ellipsoids
  - Trade-offs between volume, coverage, and “flexibility”
  - Braun, Aolaritei, Berta, Jordan, Bach (2025)

Conclusion - Challenges in conformal prediction

- **Multivariate problems**
  - Choice of conformity score based on generalized ellipsoids
  - Trade-offs between volume, coverage, and "flexibility"
  - Braun, Aolaritei, Berta, Jordan, Bach (2025)
- **Beyond i.i.d. (or exchangeability)**
  - Handling adaptive goals and sequential streams (anytime-validity)
  - Extensive use of e-values
  - Gauthier, Jordan, Bach (2025)

Conclusion - Challenges in conformal prediction

- **Multivariate problems**
  - Choice of conformity score based on generalized ellipsoids
  - Trade-offs between volume, coverage, and "flexibility"
  - Braun, Aolaritei, Berta, Jordan, Bach (2025)
- **Beyond i.i.d. (or exchangeability)**
  - Handling adaptive goals and sequential streams (anytime-validity)
  - Extensive use of e-values
  - Gauthier, Jordan, Bach (2025)
- **Bonus:** Evaluating conditional miscoverage using machine learning

Conclusion - Challenges in conformal prediction

- **Multivariate problems**
  - Choice of conformity score based on generalized ellipsoids
  - Trade-offs between volume, coverage, and "flexibility"
  - Braun, Aolaritei, Berta, Jordan, Bach (2025)
- **Beyond i.i.d. (or exchangeability)**
  - Handling adaptive goals and sequential streams (anytime-validity)
  - Extensive use of *e*-values
  - Gauthier, Jordan, Bach (2025)
- **Bonus:** Evaluating conditional miscoverage using machine learning
- **Many open problems**
  - Guarantees for conditional coverage (Gibbs et al., 2025; Plassier et al., 2025)
  - Conditional guarantees for trade-offs between size and miscoverage
  - Extension of conformal prediction to reinforcement learning

Conclusion - Challenges in conformal prediction

- **Multivariate problems**
  - Choice of conformity score based on generalized ellipsoids
  - Trade-offs between volume, coverage, and "flexibility"
  - Braun, Aolaritei, Berta, Jordan, Bach (2025)
- **Beyond i.i.d. (or exchangeability)**
  - Handling adaptive goals and sequential streams (anytime-validity)
  - Extensive use of e-values
  - Gauthier, Jordan, Bach (2025)
- **Bonus:** Evaluating conditional miscoverage using machine learning
- **Many open problems**
  - Guarantees for conditional coverage (Gibbs et al., 2025; Plassier et al., 2025)
  - Conditional guarantees for trade-offs between size and miscoverage
  - Extension of conformal prediction to reinforcement learning

References I

Ahmadi, A. A., Malioutov, D., and Luss, R. (2014). Robust minimum volume ellipsoids and higher-order polynomial level sets. In *NIPS workshop on optimization for machine learning, Montreal, Quebec, Canada*.

Angelopoulos, A. N., Barber, R. F., and Bates, S. (2024). Theoretical foundations of conformal prediction. *arXiv preprint arXiv:2411.11824*.

Angelopoulos, A. N. and Bates, S. (2023). Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591.

Braun, S., Aolaritei, L., Jordan, M. I., and Bach, F. (2025a). Minimum volume conformal sets for multivariate regression. *arXiv preprint arXiv:2503.19068*.

Braun, S., Berta, E., Jordan, M. I., and Bach, F. (2025b). Multivariate conformal prediction via conformalized gaussian scoring. *arXiv preprint arXiv:2507.20941*.

Braun, S., Holzmüller, D., Jordan, M. I., and Bach, F. (2025c). Conditional coverage diagnostics for conformal prediction. *arXiv preprint arXiv:2512.11779*.

References II

Foygel Barber, R., Candès, E. J., Ramdas, A., and Tibshirani, R. J. (2021). The limits of distribution-free conditional predictive inference. *Information and Inference: A Journal of the IMA*, 10(2):455–482.

Gauthier, E., Bach, F., and Jordan, M. I. (2025a). Backward conformal prediction. In *Advances in Neural Information Processing Systems*.

Gauthier, E., Bach, F., and Jordan, M. I. (2025b). E-values expand the scope of conformal prediction. *arXiv preprint arXiv:2503.13050*.

Gibbs, I., Cherian, J. J., and Candès, E. J. (2025). Conformal prediction with conditional guarantees. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(4):1100—-1126.

Grünwald, P., de Heide, R., and Koolen, W. M. (2020). Safe testing. In *2020 Information Theory and Applications Workshop (ITA)*, pages 1–54.

Johnstone, C. and Cox, B. (2021). Conformal uncertainty sets for robust optimization. In *Conformal and Probabilistic Prediction and Applications*, volume 152, pages 72–90.

Koning, N. W. (2024). Post-hoc α hypothesis testing and the post-hoc p-value. *arXiv preprint arXiv:2312.08040*.

References III

Koning, N. W. (2025). Measuring evidence against exchangeability and group invariance with e-values. *arXiv preprint arXiv:2310.01153*.

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.

Lei, J. and Wasserman, L. (2014). Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):71–96.

Plassier, V., Fishkov, A., Dheur, V., Guizani, M., Taieb, S. B., Panov, M., and Moulines, E. (2025). Rectifying conformity scores for better conditional coverage. In *International Conference on Machine Learning*.

Ramdas, A. and Wang, R. (2025). Hypothesis testing with e-values. *Foundations and Trends in Statistics*, 1(1-2):1–390.

Romano, Y., Patterson, E., and Candès, E. (2019). Conformalized quantile regression. *Advances in Neural Information Processing systems*, 32.

Shafer, G. (2021). Testing by betting: A strategy for statistical and scientific communication. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184:407–431.

References IV

Shafer, G. and Vovk, V. (2008). A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3):371–421.

Thurin, G., Nadjahi, K., and Boyer, C. (2025). Optimal transport-based conformal prediction. *International Conference on Machine Learning*.

Vovk, V. (2012). Conditional validity of inductive conformal predictors. In *Asian Conference on Machine Learning*, pages 475–490.

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

Vovk, V. and Wang, R. (2021). E-values: Calibration, combination and applications. *The Annals of Statistics*, 49(3):1736 – 1754.

Wang, R. and Ramdas, A. (2022). False discovery rate control with e-values. *Journal of the Royal Statistical Society Series B*, 84(3):822–852.

Wasserman, L., Ramdas, A., and Balakrishnan, S. (2020). Universal inference. *Proceedings of the National Academy of Sciences*, 117(29):16880–16890.