Recent advances in conformal prediction with E-values

Etienne Gauthier, Francis Bach, Michael I. Jordan
(INRIA, Ecole Normale Supérieure)

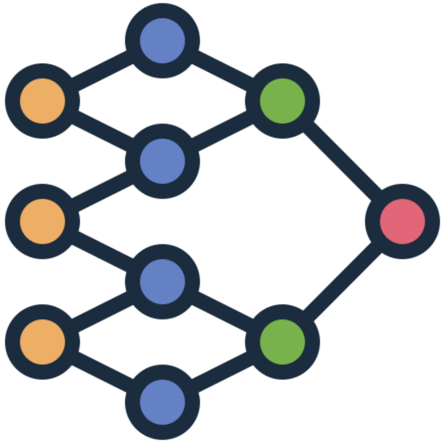
https://arxiv.org/abs/2503.13050
https://arxiv.org/abs/2505.13732

Overview – From P-values to E-values

- Basics of Conformal Prediction
- Batch Anytime-valid Conformal Prediction
- Conformal Prediction with Adaptive Coverage
- Backward Conformal Prediction
- Conformal Prediction under Ambiguous Ground Truth

Conformal Prediction

Motivation



$$\left\{\begin{array}{c}\text{airplane}\\ \text{automobile}\\ \text{bird}\\ \text{cat}\\ \text{deer}\\ \text{dog}\\ \text{frog}\\ \text{horse}\\ \text{ship}\\ \text{truck}\end{array}\right\} \leftarrow Y_{test}$$

- **Goal:** build a prediction set $C(X_{test})$ that contains Y_{test} with high probability: $\mathbb{P}(Y_{test} \in C(X_{test})) \geq 1 - \alpha$

Te

(Ground truth label Y_{test} = frog)

Predictor f

Prediction

$f(X_{test}) = \hat{Y}_{test}$

Main idea

Symmetric residuals: scores and calibration set

- **Score function $\boldsymbol{S}: \boldsymbol{X} \times \boldsymbol{Y} \to \mathbb{R}_+^*$**
  - Measures how well the predicted label aligns with the true label
  - Ex: $S(x, y) = (y - f(x))^2$ in regression, $-\log p_f(y|x)$ in classification

See Angelopoulos and Bates. Conformal prediction: A gentle introduction. FOT 2023

Main idea

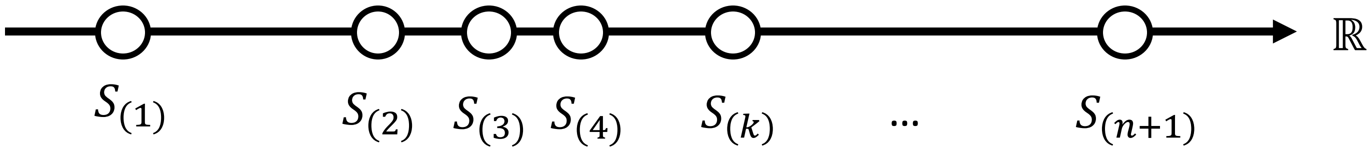
Symmetric residuals: scores and calibration set

- **Score function $\boldsymbol{S}: \boldsymbol{X} \times \boldsymbol{Y} \to \mathbb{R}_+^*$**
  - Measures how well the predicted label aligns with the true label
  - Ex: $S(x, y) = (y - f(x))^2$ in regression, $-\log p_f(y|x)$ in classification

> **Conformal Prediction (Vovk et al., 2005):**
>
> $$\mathbb{P}(Y_{test} \in C(X_{test})) \geq 1 - \alpha,$$
>
> where $C(X_{test}) = \{y : \text{rank}\big(S(X_{test}, y)\big) \leq \lceil (1-\alpha)(n+1) \rceil\}$.

- **Observation:** the $S(X_i, Y_i)$ and $S(X_{test}, Y_{test})$ are i.i.d:

$$\mathbb{P}\Big(\text{rank}\big(S(X_{test}, Y_{test})\big) \leq k\Big) = \frac{k}{n+1} \quad \to \quad k = \lceil (1-\alpha)(n+1) \rceil$$



See Angelopoulos and Bates. Conformal prediction: A gentle introduction. FOT 2023

P-values! And E-values…

Alternative formulation:

$$\mathbb{P}\left(\boxed{\frac{1+\sum_{i=1}^{n}\mathbb{1}\{S(X_i,Y_i)>S(X_{test},Y_{test})\}}{n+1}}\leq\alpha\right)\leq\alpha$$

p-value

$$\mathbb{1}\{S(X_i,Y_i)>S(X_{test},Y_{test})\}=\mathbb{1}\left\{\frac{S(X_i,Y_i)}{S(X_{test},Y_{test})}>1\right\}\leq\frac{S(X_i,Y_i)}{S(X_{test},Y_{test})}$$

$\leq 1/E$ where E is the soft-rank e-value [Wang & Ramdas 2020, Koning 2023, Balinsky & Balinsky 2024]:

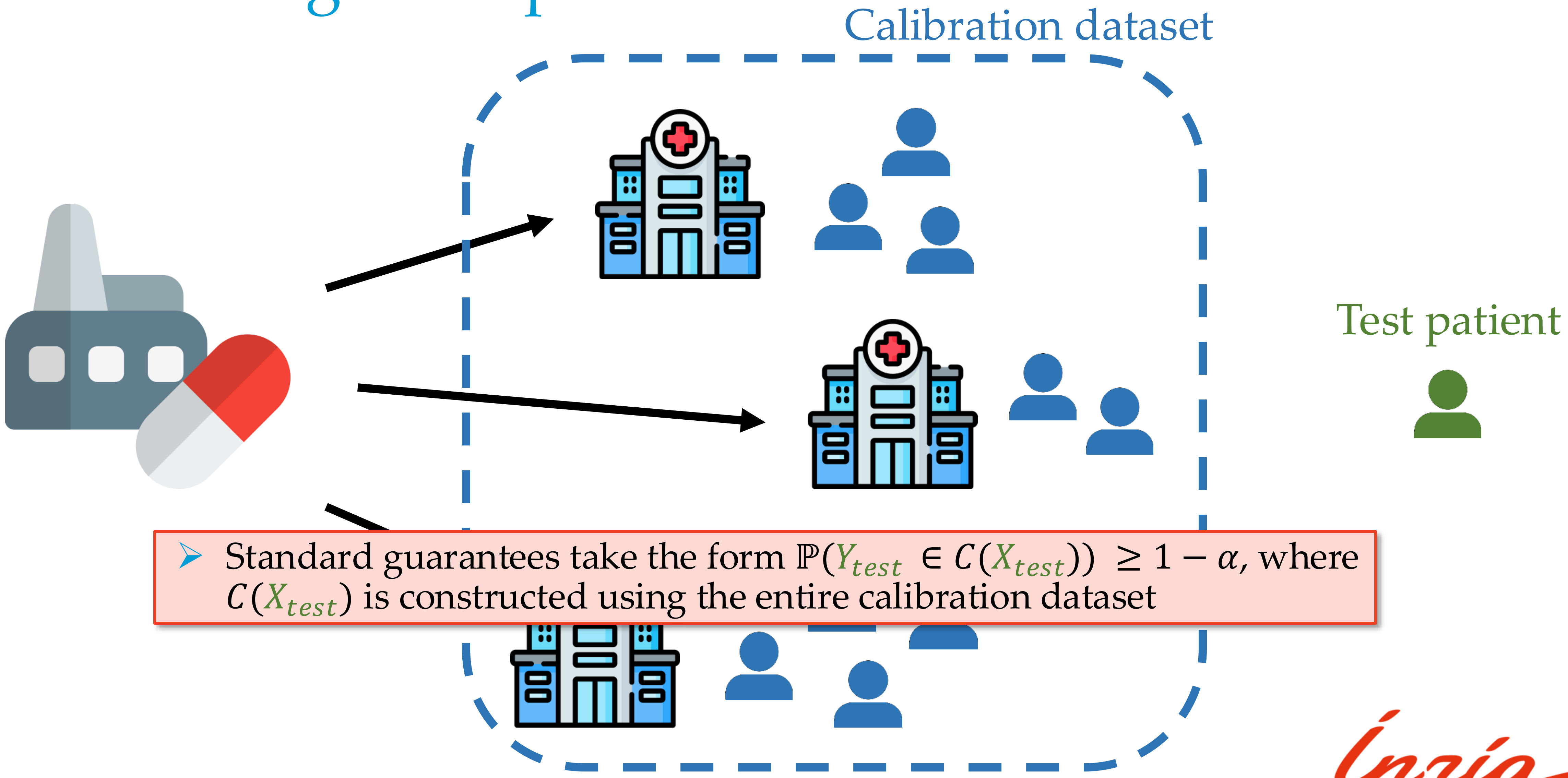
$$E=\frac{S(X_{test},Y_{test})}{\frac{1}{n+1}\left(\sum_{i=1}^{n}S(X_i,Y_i)+S(X_{test},Y_{test})\right)}\quad\text{with } \mathbb{P}\left(E\leq\frac{1}{\alpha}\right)\leq\alpha$$

➢ Conformal e-prediction and e-variables [Vovk 2024]

Overview

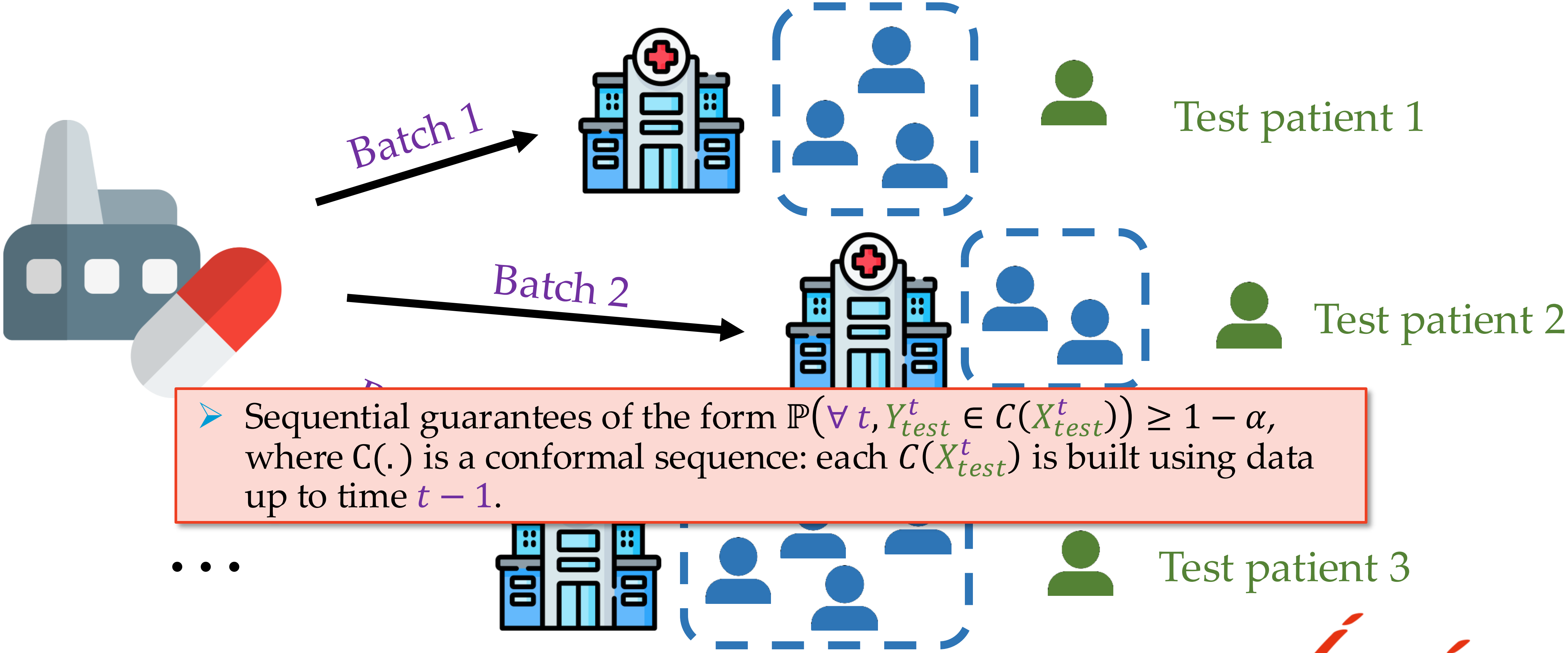
- Basics of Conformal Prediction
- Batch Anytime-valid Conformal Prediction
- Conformal Prediction with Adaptive Coverage
- Backward Conformal Prediction
- Conformal Prediction under Ambiguous Ground Truth

Motivating Example



➢ Standard guarantees take the form $\mathbb{P}(Y_{test} \in C(X_{test})) \geq 1 - \alpha$, where $C(X_{test})$ is constructed using the entire calibration dataset

Motivating Example



- Sequential guarantees of the form $\mathbb{P}\big(\forall\, t, Y^t_{test} \in C(X^t_{test})\big) \geq 1-\alpha$, where $C(.)$ is a conformal sequence: each $C(X^t_{test})$ is built using data up to time $t-1$.

Main Result

Supermartingale defined from e-variables + Ville's inequality

$$M_t = \prod_{b=1}^{t} (1 - \lambda_b + \lambda_b E_b),$$

$$E_b = \frac{s(X_{test}^b, Y_{test}^b)}{\frac{1}{n_b + 1}\left(\sum_{i=1}^{n_b} s(X_i^b, Y_i^b) + s(X_{test}^b, Y_{test}^b)\right)}$$

- **Batch Anytime-valid Conformal Prediction:**

$$\mathbb{P}\left(\forall\, t, Y_{test}^t \in C(X_{test}^t)\right) \geq 1 - \alpha,$$

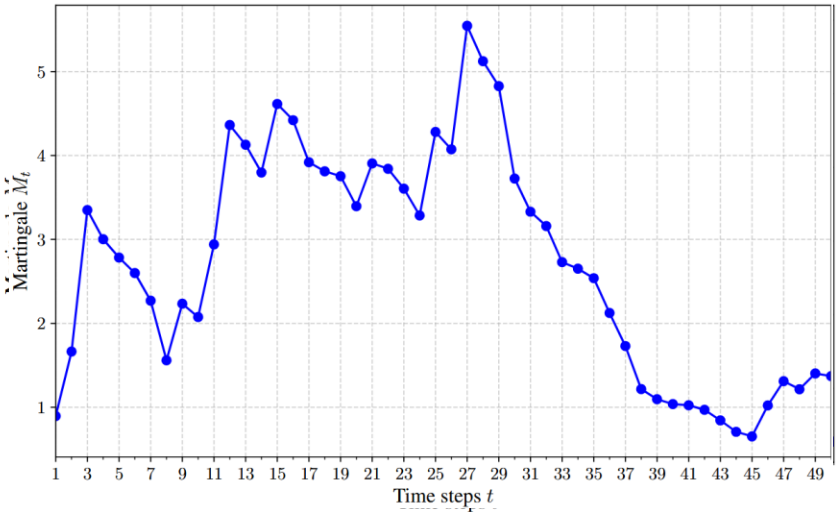
where $C(X_{test}^t) = \left\{ y : \prod_{b=1}^{t-1}(1 - \lambda_b + \lambda_b E_b) \times \frac{s(X_{test}^t, y)}{\frac{1}{n_t+1}\left(\sum_{i=1}^{n_t} s(X_i^t, Y_i^t) + s(X_{test}^t, y)\right)} < 1/\alpha \right\}.$

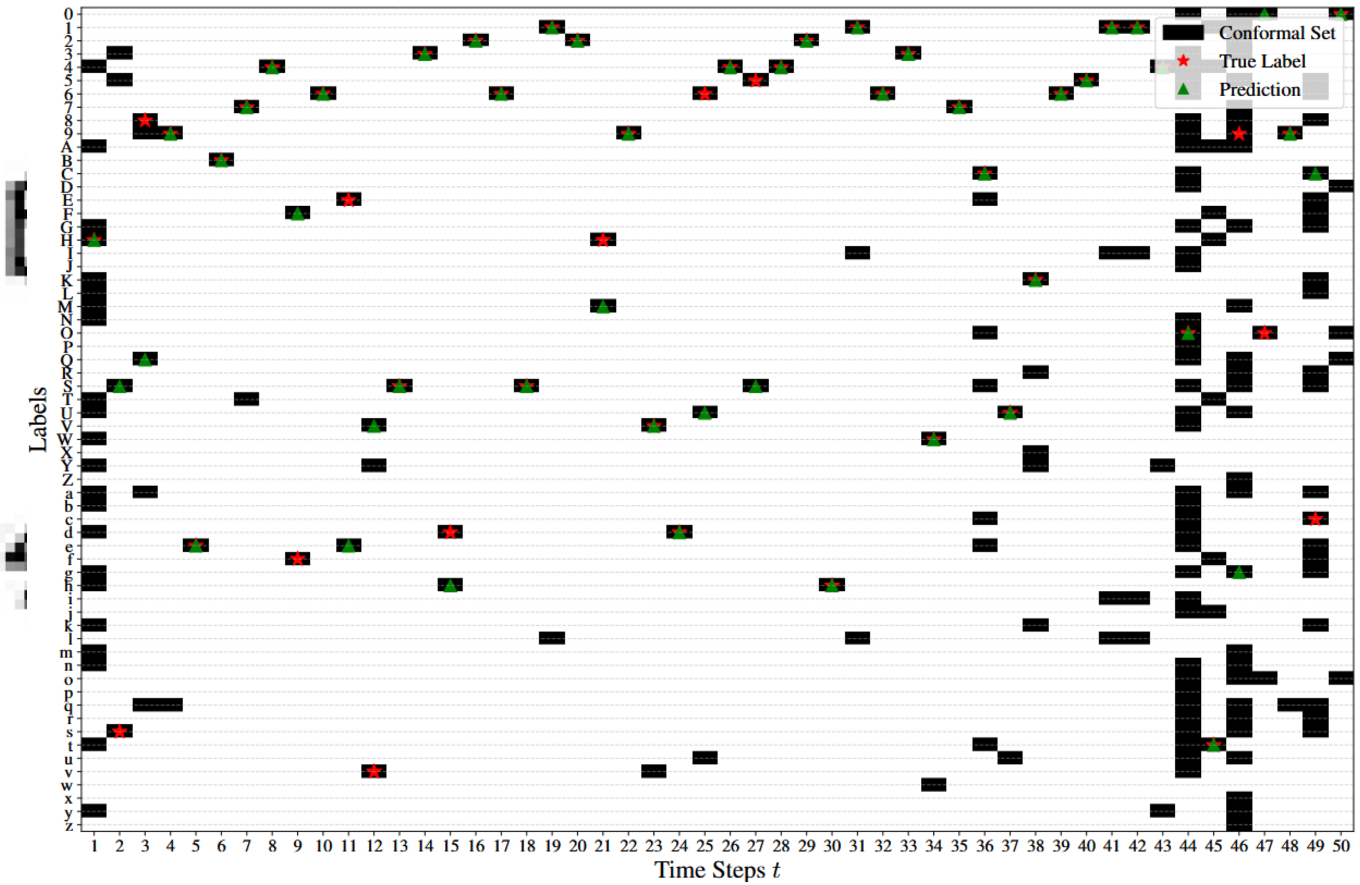
Ville's inequality for nonnegative supermartingale Xn (1939)

$$\mathbb{P}\left(\sup_t M_t \geq \alpha\right) \leq \mathbf{E}[M_0]/\alpha$$

Experiments

- FEMNIST dataset
- $S(x,y) = \frac{1}{p_f(y|x)^{1/4}}$
- $\lambda = 1$
- $\alpha = 0.15$

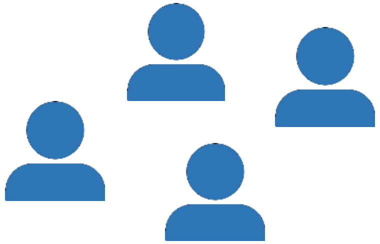




Today's Agenda

- ❑ Basics of Conformal Prediction
- ❑ Batch Anytime-valid Conformal Prediction
- ❑ Conformal Prediction with Adaptive Coverage
- ❑ Backward Conformal Prediction
- ❑ Conformal Prediction under Ambiguous Ground Truth

Motivating Example



Calibration dataset

Test patient

$\alpha = 0.01$

$$\left\{ \begin{array}{c} \text{Pneumonia} \\ \text{COVID-19} \\ \text{Asthma} \\ \text{COPD} \\ \text{Tuberculosis} \\ \text{Lung Cancer} \\ \text{Pulmonary Embolism} \\ \text{Bronchitis} \\ \text{Influenza} \\ \text{Sarcoidosis} \end{array} \right\}$$

$|C(X_{test})| = 8$

Main Result

Post-hoc guarantees

Post-hoc p-variables:

$$\sup \mathbb{E}\left[\frac{\mathbb{P}(P \leq \tilde{\alpha} \,|\tilde{\alpha})}{\ldots}\right] \leq 1$$

P

➢ **Conformal Prediction with Adaptive Coverage:**

$$\mathbb{E}\left[\frac{\mathbb{P}(Y_{test} \notin C(X_{test})|\tilde{\alpha})}{\tilde{\alpha}}\right] \leq 1,$$

for any adaptive (possibly data-dependent) miscoverage $\tilde{\alpha} > 0$, where:

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

First-order Taylor approximation: $\mathbb{E}\left[\frac{\mathbb{P}(Y_{test} \notin C(X_{test})|\tilde{\alpha})}{\tilde{\alpha}}\right] \approx \frac{\mathbb{E}[\mathbb{P}(Y_{test} \notin C(X_{test})|\tilde{\alpha})]}{\mathbb{E}[\tilde{\alpha}]} = \frac{\mathbb{P}(Y_{test} \notin C(X_{test}))}{\mathbb{E}[\tilde{\alpha}]}$

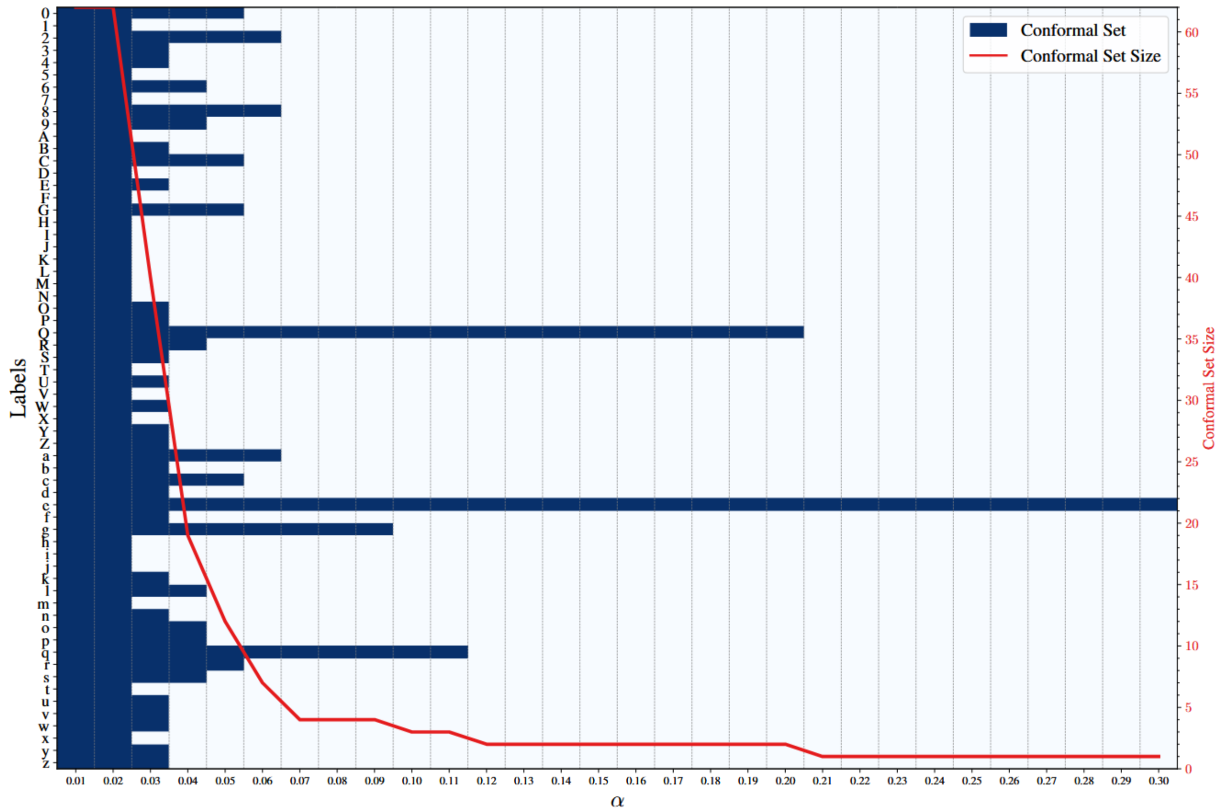
$$\mathbb{P}(Y_{test} \in C(X_{test})) \geq 1 - \boxed{\mathbb{E}[\tilde{\alpha}]}$$

Can be estimated using the calibration see [Gauthier, Bach & Jordan 2025]

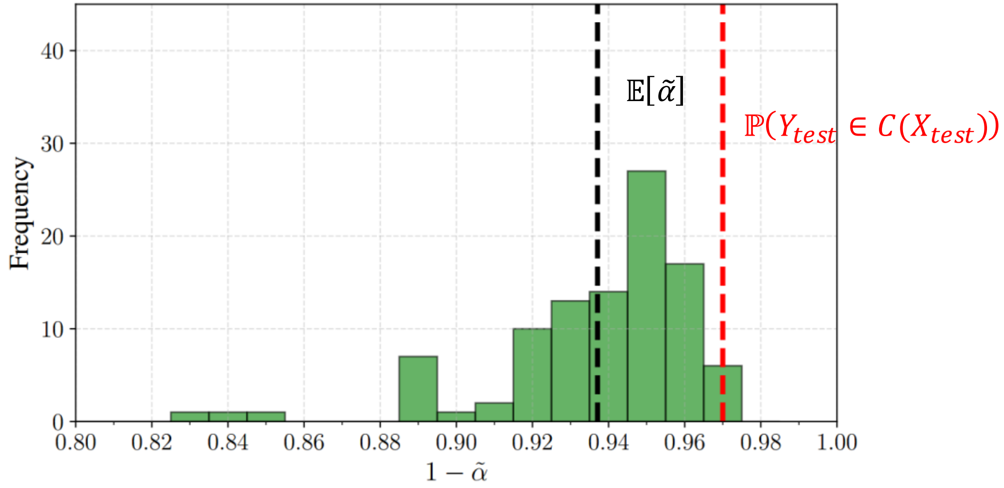
See Koning (2024), Gazin, Blanchard, Roquain (2023)

Experiments

$$\tilde{\alpha} = \inf\left\{\alpha \in (0,1): \#\left\{y : \frac{S(X_{test}, y)}{\frac{1}{n+1}\left(\sum_{i=1}^{n} S(X_i, Y_i) + S(X_{test}, y)\right)} < \frac{1}{\alpha}\right\} \leq C(\{(X_i, Y_i)\}, X_{test})\right\}$$



$$\mathbb{P}(Y_{test} \in C(X_{test})) \geq 1 - \mathbb{E}[\tilde{\alpha}]$$

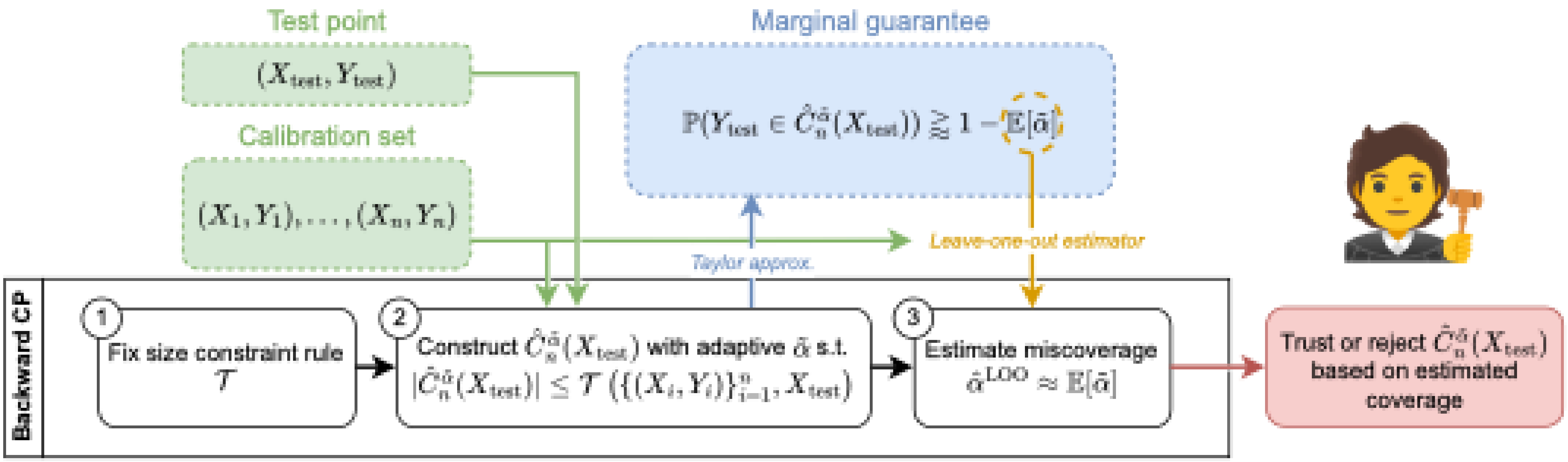


$C = 3$

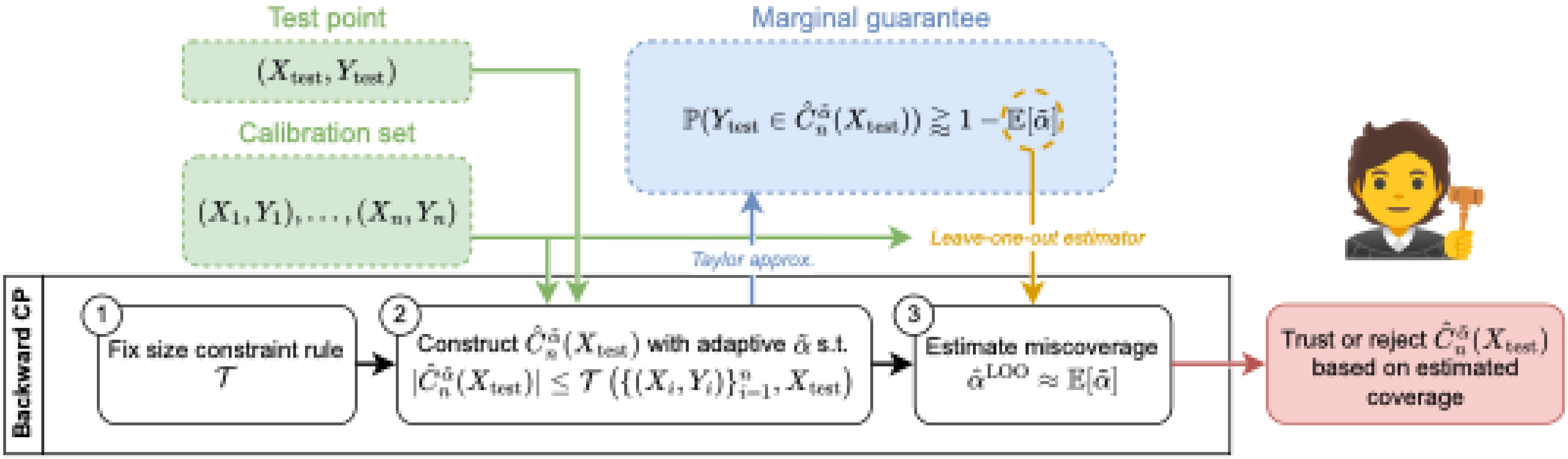
Overview

- Basics of Conformal Prediction
- Batch Anytime-valid Conformal Prediction
- Conformal Prediction with Adaptive Coverage
- Backward Conformal Prediction
- Conformal Prediction under Ambiguous Ground Truth

Backward Conformal Prediction



Backward Conformal Prediction



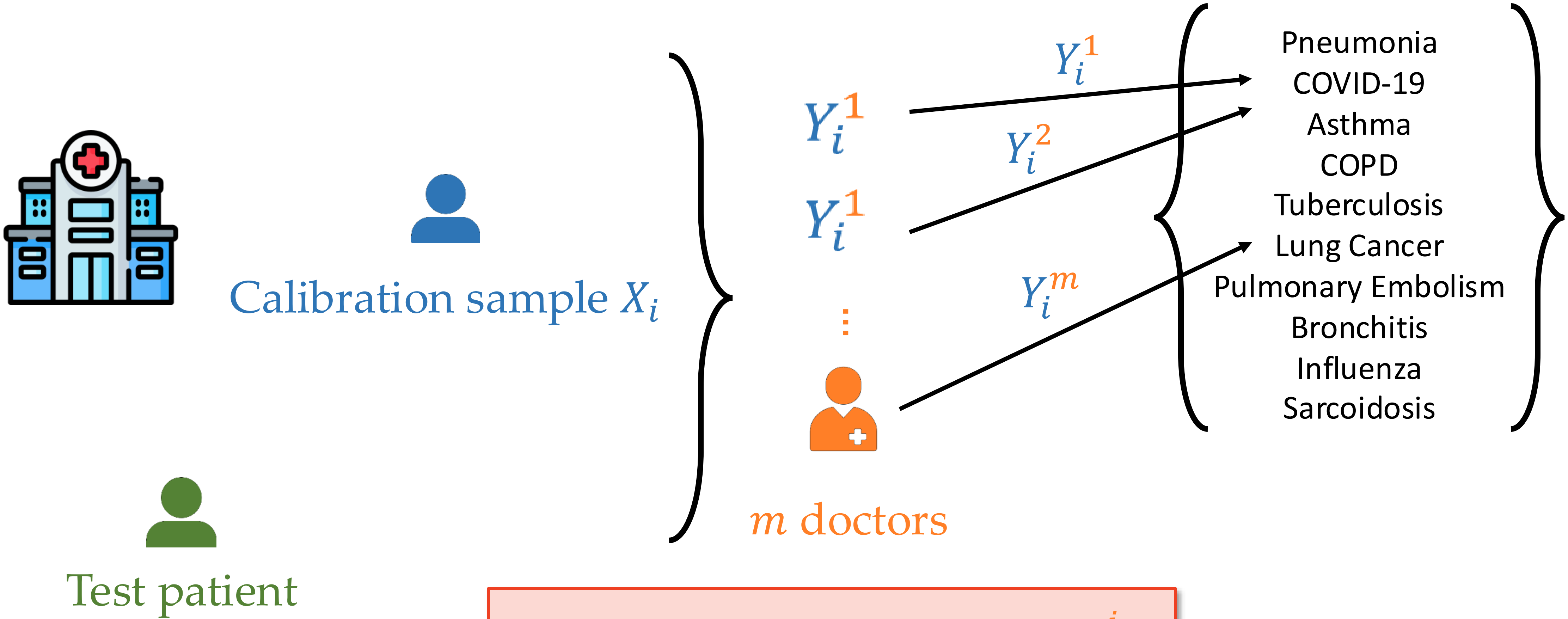
Leave-one-out estimator

$$\left|\hat{\alpha}^{\text{LOO}} - \mathbb{E}[\tilde{\alpha}]\right| = O_P\left(\frac{1}{\sqrt{n}}\right)$$

Overview

- Basics of Conformal Prediction
- Batch Anytime-valid Conformal Prediction
- Conformal Prediction with Adaptive Coverage
- Backward Conformal Prediction
- Conformal Prediction under Ambiguous Ground Truth

Motivating Example



➢ Calibration dataset: (X_i, Y_i^j)

See Stutz et al. (2023)

Main Result

Average of e-values = e-value

$$E = \frac{1}{m}\sum_{j=1}^{m} E^j$$

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

➢ **Conformal Prediction under Ambiguous Ground Truth:**

$$\mathbb{P}\big(Y_{test} \in C(X_{test})\big) \geq 1 - \alpha,$$

Where $C(X_{test}) = \left\{ y : \frac{1}{m}\sum_{j=1}^{m} \frac{S(X_{test}, Y_{test})}{\frac{1}{n+1}\sum_{i=1}^{n} S(X_i, Y_i^j) + S(X_{test}, Y_{test})} < 1/\alpha \right\}.$

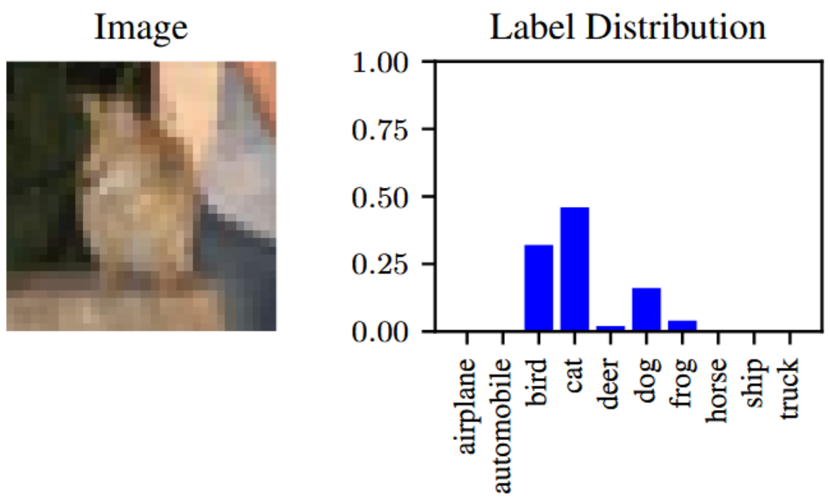
Markov's *inequality*

See Vovk and Wang (2020)

Inria

Experiments

- CIFAR-10H dataset (filtered)
- $S(x, y) = -\log p_f(y|x)$



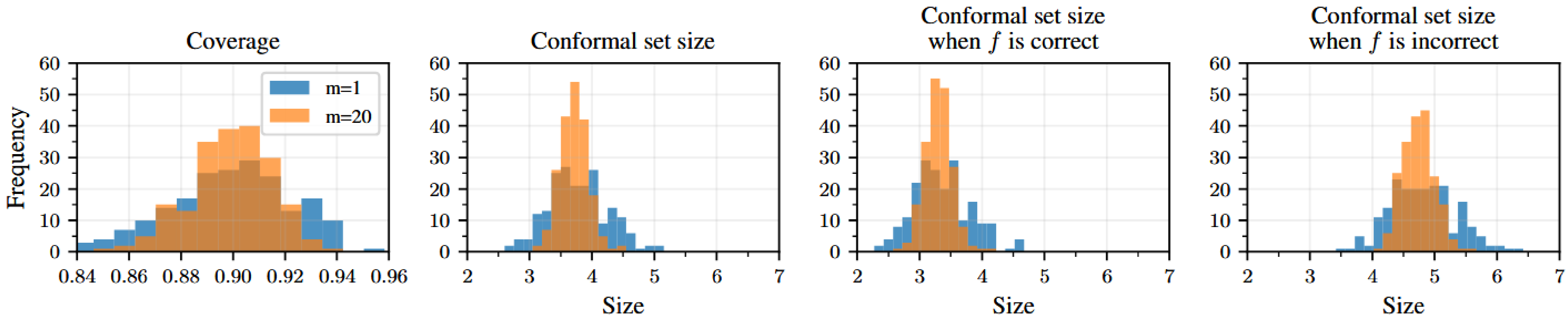


Figure 6: Comparison of coverage and conformal set sizes when using e-variables in Monte Carlo conformal prediction with $m = 1$ or $m = 20$ experts, with $\alpha = 0.3$, from Theorem 15.

Conclusion

- Explored **e-values for conformal prediction**, enabling more flexible inference
- Enables **online conformal methods** with anytime−valid guarantees
- Enables **data-dependent coverage guarantees**, allowing more adaptive and
  inf
- Fa n
  cas
- Op
  - (
- Op
  - Choice of the score function in the soft-rank e-value?
  - Choice of the e-value?
  - Conditional guarantees (Gibbs et al., 2024)

> **Poster to check out!**
> Sacha Braun, Minimum volume conformal sets for multivariate regression

> **Poster to check out!**
> Eugène Berta, Rethinking Early Stopping: Refine, Then Calibrate