

Statistical Learning: Anytime Valid Conformal Prediction

Barabás Eszter

ELTE Eötvös Loránd Tudományegyetem

Supervisor: Csáji Balázs Csanád

Budapest, 2025.

Introduction and Motivation

- Continuation of previous semester's work on conformal prediction
- Focus on advanced methods for uncertainty quantification
- Conformal prediction produces prediction sets with guaranteed coverage probability
- Classical approach: conformal predictors based on p-values
- This work explores an alternative: conformal e-prediction (using e-values)

Introduction and Motivation

- Limitations of p-values:
 - ▶ Susceptible to misuse (e.g., p-hacking)
 - ▶ Selective reporting and inflated Type I error
 - ▶ Dependence on fixed sampling schemes
- Advantages of e-values:
 - ▶ Valid under flexible or unknown sampling schemes
 - ▶ Allow data-dependent (post-hoc) choice of significance level
 - ▶ Can be combined across experiments (via multiplication)
 - ▶ Remain valid under sequential data collection (anytime validity)
- Key idea:
 - ▶ E-values can be constructed as supermartingales
 - ▶ Useful for adaptive procedures, online learning, and sequential experiments

Definitions

We begin with a sample space Ω equipped with a σ -algebra \mathcal{F} , and the set \mathcal{M} of all probability measures on (Ω, \mathcal{F}) , whose elements are the distributions.

Let $\mathcal{P} \subset \mathcal{M}$ set of probability measures representing the null hypothesis.

Definition

An *e-variable* E for \mathcal{P} is a $[0, \infty]$ -valued random variable satisfying

$$\mathbb{E}_P[E] \leq 1 \quad \text{for all } P \in \mathcal{P}.$$

An e-variable E is *exact* if

$$\mathbb{E}_P[E] = 1 \quad \text{for all } P \in \mathcal{P}.$$

Definition

A *p-variable* P for \mathcal{P} is a $[0, \infty)$ -valued random variable satisfying

$$P(P \leq \alpha) \leq \alpha \quad \text{for all } \alpha \in (0, 1) \text{ and all } P \in \mathcal{P}.$$

A p-variable P is *exact* if

$$P(P \leq \alpha) = \alpha \quad \text{for all } \alpha \in (0, 1) \text{ and all } P \in \mathcal{P}.$$

E-values and P-values

E-variables and p-variables are random variables, while e-values and p-values refer to their realized values after observing the data.

Proposition (Markov's inequality for e-values)

Let E be an e-variable for \mathcal{P} . Then

$$P\left(E \geq \frac{1}{\alpha}\right) \leq \alpha \quad \text{for all } P \in \mathcal{P} \text{ and } \alpha \in (0, 1]. \quad (1)$$

Hence, $1/E$ is a p-variable, $(\alpha E) \wedge 1$ is a level- α test, and $1_{\{E \geq 1/\alpha\}}$ is a level- α binary test.

This allows for the construction of conformal sets. Conformal e-prediction refers to conformal prediction methods based on e-variables.

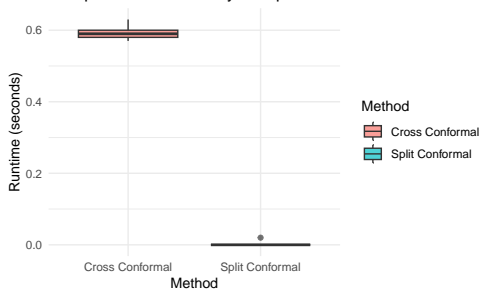
Conformal e-Prediction Setup

- Training data: $z_i = (x_i, y_i)$ for $i = 1, \dots, n$
- Goal: predict label y for a new object x
- Assign a score function:
 - ▶ $f(z_1, \dots, z_n, x, y)$ measures plausibility of label y
- Data space:
 - ▶ Objects: \mathcal{X} , labels: \mathcal{Y}
 - ▶ Observations: $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$
- Nonconformity e-measure:
 - ▶ Maps sequences (z_1, \dots, z_m) to scores $(\alpha_1, \dots, \alpha_m)$
 - ▶ Permutation invariant
 - ▶ Average constraint: $\frac{1}{m} \sum_{i=1}^m \alpha_i \leq 1$
- Conformal e-predictor:
 - ▶ $f(z_1, \dots, z_n, x, y) = \alpha_{n+1}$
 - ▶ Gives a score for each possible label

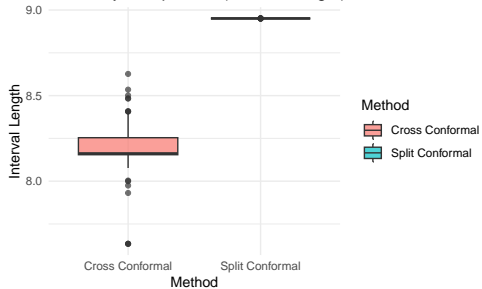
Split and Cross Conformal e-Prediction

- Key components:
 - ▶ Summary space Σ
 - ▶ Nonconformity map: $A : \mathcal{Z}^+ \rightarrow \Sigma$
 - ▶ Normalization: $N : \Sigma^+ \rightarrow [0, \infty)^+$ (ensures average ≤ 1)
- Split conformal:
 - ▶ Split data into training and calibration sets
 - ▶ Compute summaries for calibration and test candidate
 - ▶ Normalize to obtain conformal e-values α_y
 - ▶ Predictor: $f(z_1, \dots, z_n, x, y) = \alpha_y$
 - ▶ Computationally efficient under suitable A and N
- Cross-conformal:
 - ▶ Split data into K folds
 - ▶ Train/calibrate across all folds
 - ▶ Average fold-wise e-values
 - ▶ Improves predictive efficiency compared to split method

Computational Efficiency Comparison



Efficiency Comparison (Interval Length)



Batch Anytime-Valid Conformal Prediction

- Data arrive sequentially in batches with possible distribution shifts.
- Goal: construct prediction sets \hat{C}_t with global guarantee

$$\mathbb{P}(\forall t \geq 1, S_{t, n_{t+1}} \in \hat{C}_t) \geq 1 - \alpha.$$

- Standard conformal prediction ensures per-batch coverage, but **fails** to guarantee validity over time.
- Solution: use **test supermartingales** and **Ville's inequality**.

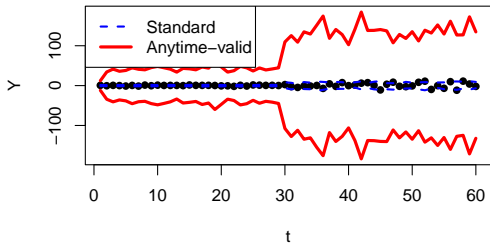
Theorem (Ville's Inequality)

Let $\{M_t\}_{t \geq 0}$ be a nonnegative supermartingale. Then, for any $\alpha \in (0, 1)$,

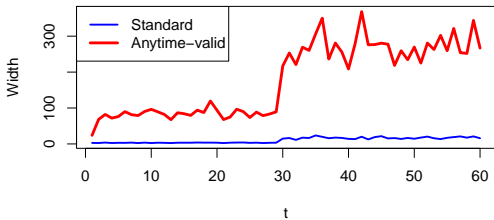
$$\mathbb{P}\left(\forall t \geq 0, M_t < \frac{1}{\alpha}\right) \geq 1 - \alpha.$$

- Construct e-values $E_s = \frac{S_s^{n_s+1}}{\frac{1}{n_s+1} \sum_{j=1}^{n_s+1} S_s^j}$ and martingale $M_t = \prod_{s=1}^t E_s$.
- Define \hat{C}_t via constraint $M_t < 1/\alpha \Rightarrow$ anytime-valid coverage.

Standard (blue) vs Anytime-valid (red)



Interval Widths



Data-Dependent Conformal Prediction via E-values

- Classical conformal prediction fixes $\alpha \Rightarrow$ guarantees $1 - \alpha$ coverage but **no control on set size**.
- Idea: allow α to be **data-dependent** ($\tilde{\alpha}$).

Definition

We say that a nonnegative random variable P is a post-hoc p-variable if

$$\sup_{\tilde{\alpha}} \mathbb{E} \left[\frac{\mathbb{P}(P \leq \tilde{\alpha} \mid \tilde{\alpha})}{\tilde{\alpha}} \right] \leq 1,$$

where the supremum is over every random variable $\tilde{\alpha} > 0$.

Theorem

P is a post-hoc p-variable if and only if

$$\mathbb{E} \left[\frac{1}{P} \right] \leq 1.$$

Data-Dependent Conformal Prediction via E-values

- Using e-values:

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

gives adaptive validity via $P = 1/E$.

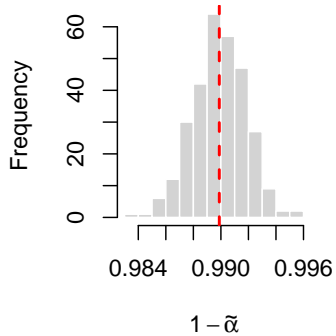
- Guarantee (data-dependent coverage):

$$\mathbb{E} \left(\frac{\mathbb{P}(Y_{n+1} \notin \hat{C}_n^{\tilde{\alpha}} \mid \tilde{\alpha})}{\tilde{\alpha}} \right) \leq 1.$$

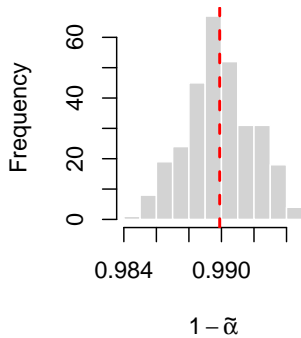
- Enables:

- ▶ Adaptive choice of $\tilde{\alpha}$
- ▶ Control of set size (e.g. enforce $|\hat{C}_n^\alpha| \leq C$)

C = 3



C = 5



Monte Carlo Conformal E-Prediction

- Setting: each input X_i has multiple labels $Y_i^{(j)}$, $j = 1, \dots, m$ (e.g. multiple experts).
- Data are **not exchangeable** \Rightarrow standard conformal prediction fails.
- Idea: use all labels via **e-values** instead of selecting one.
- For each expert:

$$E_j = \frac{S_{n+1}}{\frac{1}{n+1} \left(\sum_{i=1}^n S_i^{(j)} + S_{n+1} \right)}$$

- Average preserves validity:

$$\bar{E} = \frac{1}{m} \sum_{j=1}^m E_j \quad \text{is an e-variable.}$$

- Conformal set:

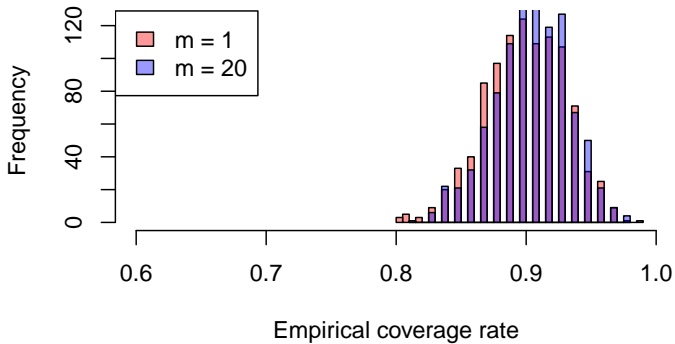
$$\hat{C}_n(x) = \{y : \bar{E} < 1/\alpha\}$$

- Guarantee:

$$\mathbb{P}(Y_{n+1} \in \hat{C}_n(X_{n+1})) \geq 1 - \alpha$$

- Key advantage: uses all experts \Rightarrow reduces variability while maintaining $1 - \alpha$ coverage.

Number of experts: $m = 1$ vs $m = 20$










AI use

I used Generative AI in my report

- to generate parts of the code for experiments and visualization
- for checking grammatical errors

References

-  Vovk, Vladimir, Alexander Gammerman, and Glenn Shafer. Algorithmic learning in a random world. Boston, MA: Springer US, 2005.
-  Ramdas, Aaditya, and Ruodu Wang. "Hypothesis testing with e-values." *Foundations and Trends® in Statistics* 1.1-2 (2025): 1-390.
-  Gauthier, Etienne, Francis Bach, and Michael I. Jordan. "E-values expand the scope of conformal prediction." *arXiv preprint arXiv:2503.13050* (2025).
-  Vovk, Vladimir. "Conformal e-prediction." *Pattern Recognition* 166 (2025): 111674.
-  Balinsky, A. A., and Balinsky, A. D. (2024). Enhancing conformal prediction using e-test statistics. In *Proceedings of the Thirteenth Symposium on Conformal and Probabilistic Prediction with Applications (Proceedings of Machine Learning Research, Vol. 230, pp. 65–72)*.
-  Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., and Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, 13(3), e1002106.
-  Csillag, Daniel, et al. "Extending Prediction-Powered Inference through Conformal Prediction." *arXiv preprint arXiv:2510.16166* (2025).