

Ch 8. Online Conformal Prediction

Sungwoo Park

Uncertainty Quantification Lab
Seoul National University

November 17th, 2025

Table of contents

1 Introduction

2 Online CP with Exchangeable Data

3 Conclusion

Contents

1 Introduction

2 Online CP with Exchangeable Data

3 Conclusion

Motivation: From Batch to Online

Batch Setting (Previous Chapters)

- Training dataset $\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ collected first
- Train conformal prediction on this dataset
- Provide predictive inference on future test points
- Data is fixed and static

Online Setting (This Chapter)

- Observe data points **sequentially**
- At time t : observed $(X_1, Y_1), \dots, (X_{t-1}, Y_{t-1})$ and X_t
- Need to construct prediction set $\mathcal{C}_t(X_t)$ for Y_t
- At time $t + 1$: add (X_t, Y_t) to dataset and repeat

Key Challenges: Multiple prediction sets, data reuse, potentially shifting distributions

Contents

1 Introduction

2 Online CP with Exchangeable Data

3 Conclusion

Setting and Goal

Setup

- Sequence of data points $(X_1, Y_1), (X_2, Y_2), \dots, (X_T, Y_T)$ that are **exchangeable**
- At each time step t : have seen $(X_1, Y_1), \dots, (X_{t-1}, Y_{t-1})$ and X_t
- Score function s (symmetric)

Goal

Construct prediction set $\mathcal{C}_t(X_t)$ for Y_t with correct coverage:

$$\mathbb{P}(Y_t \in \mathcal{C}_t(X_t)) \geq 1 - \alpha$$

Setting and Goal

Approach

Apply **full conformal prediction** (Algorithm 3.3) at each time step:

- Training data: $(X_1, Y_1), \dots, (X_{t-1}, Y_{t-1})$
- Test point: X_t
- At $t = 1$: zero training points $\Rightarrow C_1(X_1) = \mathcal{Y}$ (uninformative but well-defined)

Miscoverage Events

Definition (Miscoverage Indicator)

$$\text{err}_t = \begin{cases} 1 & \text{if } Y_t \notin \mathcal{C}_t(X_t) \\ 0 & \text{if } Y_t \in \mathcal{C}_t(X_t) \end{cases}$$

Basic Marginal Coverage

From Theorem 3.2 (validity of conformal prediction):

$$\mathbb{E}[\text{err}_t] \leq \alpha \quad \text{for all } t$$

Question: Does average coverage $\frac{1}{T} \sum_{t=1}^T \text{err}_t$ concentrate around $1 - \alpha$?

Challenge: Each data point t is reused in constructing $\mathcal{C}_{t'}$ for all $t' > t \Rightarrow$ coverage events appear to be dependent in complex ways.

Contents

1 Introduction

2 Online CP with Exchangeable Data

3 Conclusion

Conclusion

- 안녕

Thank you!

References I

- [1] Anastasios N Angelopoulos, Rina Foygel Barber, and Stephen Bates.
Theoretical foundations of conformal prediction.