Ch 8. Online Conformal Prediction

Sungwoo Park

Uncertainty Quantification Lab
Seoul National University

November 17th, 2025

Table of contents

Contents

From Batch to Online Setting

Batch Setting (Previous Chapters)

- Train on dataset $\mathcal{D}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$
- Provide inference on future test points
- One-time prediction

Online Setting (This Chapter)

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

Online Prediction Illustration

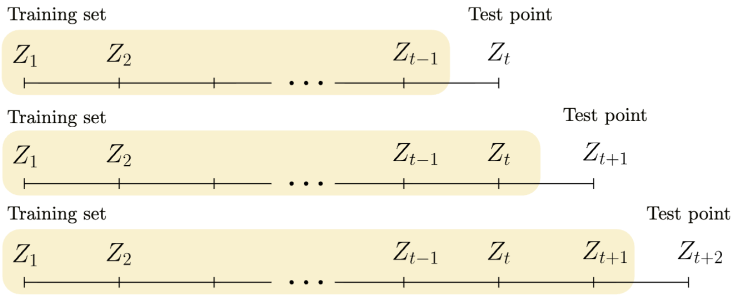

Figure 1: Online prediction loop — observe X_t, produce $\mathcal{C}_t(X_t)$, reveal Y_t, and update the dataset. [1]

New Challenges in Online Setting

1. **Data Reuse:** Each data point is used multiple times
   - (X_t, Y_t) is a test point at time t
   - (X_t, Y_t) becomes training data for times $t' > t$
2. **Multiple Prediction Sets:** Constructing many prediction sets over time
   - Need coverage guarantees for each $C_t(X_t)$
   - Need long-run average coverage guarantees
3. **Distribution Shift:** Distributions may change over time
   - Exchangeability may not hold
   - Need robust methods for non-stationary data

Q1: Why Study Sequences of Prediction Sets?

Question

$X_1, X_2, X_3, \cdots, X_t$ 로부터 score function을 이용하여 $C(X_t)$을 구성하는 Conformal Prediction Flow에서 Online Comformal Prediction을 통한 추정구간의 열을 구하는 것이 실질적으로 의미가 있는 것인지 알고 싶습니다.

Answer

- Online Data의 특성상 열로의 관찰이 필요함
- Corollary 8.3에서 제시된 수렴성을 확인할 수 있음
- Supermartingale의 특징을 활용해 이상치 관측 가능

Online Data Examples

Example 1: Medical ICU Monitoring

- Patient vital signs measured every hour
- Need prediction interval for next hour's blood pressure
- **Can't wait** to collect all 24 hours of data before making first prediction
- Need sequence: $C_1(X_1), C_2(X_2), \ldots, C_{24}(X_{24})$
- If coverage fails repeatedly, may indicate patient deterioration

Example 2: Algorithmic Trading

- Stock price predictions every minute
- Trading decisions based on prediction intervals
- Need to detect when market regime changes (Section 8.2 test)
- Online conformal adapts to volatility changes (Section 8.3)

Q2: Can We Use Split Conformal Online?

Question

Online conformal prediction은 Split conformal에는 적용할 수 없나요? 만약 적용할 수 있다면 어떤 방식으로 매 시점의 예측집합을 만드나요?

Answer: Maybe... NO

Split Conformal에서는 Training set과 Calibration set이 나누어지는데, 이를 고정시켜 prediction을 하기에 연속성이 중요한 Online setting에서는 적합하지 않다고 생각합니다.

Contents

Setup: Exchangeable Sequences

Assumptions

- Sequence $(X_1, Y_1), (X_2, Y_2), \ldots, (X_T, Y_T)$ is **exchangeable**
- At time t: observed $(X_1, Y_1), \ldots, (X_{t-1}, Y_{t-1})$ and X_t
- Goal: construct $\mathcal{C}_t(X_t)$ with coverage

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

Algorithm

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

- Training data: $(X_1, Y_1), \ldots, (X_{t-1}, Y_{t-1})$
- Test point: X_t
- Score function: s (symmetric)

Miscoverage Events

Definition

Define the miscoverage indicator at time t:

$$\mathrm{err}_t := \mathbb{1}\{Y_t \notin \mathcal{C}_t(X_t)\}$$

From Theorem 3.2(Marginal Coverage), we know:

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

Question

Does marginal coverage imply good **average coverage**?

$$\frac{1}{T}\sum_{t=1}^{T} \mathrm{err}_t \approx \alpha \; ?$$

Online Prediction Guarantee

Proposition 8.1: Independence of Errors

Suppose $(X_1, Y_1), (X_2, Y_2), \ldots, (X_T, Y_T)$ are exchangeable, the score function s is symmetric, and the scores are distinct almost surely at each time t. Then the miscoverage events err_t are **mutually independent**.

- Each data point reused in constructing $C_{t'}$ for all $t' > t$
- Despite complex data reuse, errors are independent!

Online Prediciton Guarantee : Conformal p-values

Definition (8.3)

The conformal p-value at time t is:

$$p_t = \frac{\sum_{i=1}^{t} \mathbb{1}\{s((X_i, Y_i); \mathcal{D}_t) \geq s((X_t, Y_t); \mathcal{D}_t)\}}{t}$$

where $\mathcal{D}_t = ((X_1, Y_1), \ldots, (X_t, Y_t))$.

Theorem 8.2: Independence of Online Conformal p-values

Under the same assumptions, p_t is distributed as a discrete random variable on $\{1/t, 2/t, \ldots, 1\}$, and $p_1, \ldots, p_T$ are **mutually independent**.

We can easily check that $\mathrm{err}_t = \mathbb{1}\{p_t \leq \alpha\}$ (Proposition 3.9), so Proposition 8.1 follows from Theorem 8.2.

Proof Intuition (Theorem 8.2)

Key Idea:

- By symmetry of score function, p_{t+1} does not depend on *ordering* of $Z_1, \ldots, Z_t$
- $Z_1, \ldots, Z_t$ still exchangeable when conditioning on p_{t+1}
- But p_t depends *only* on ordering of first t points
- This leads to independence!

Proof Strategy:

1. Show: $p_t | (\hat{P}_t, Z_{t+1}, \ldots, Z_T) \sim \text{Unif}(\{1/t, \ldots, 1\})$
2. Show: For $t' > t$, $p_{t'}$ is a function of $(\hat{P}_t, Z_{t+1}, \ldots, Z_T)$
3. Combine: $p_t | (p_{t+1}, \ldots, p_T) \sim \text{Unif}(\{1/t, \ldots, 1\})$
4. Therefore: $p_1, \ldots, p_T$ are independent

where $\hat{P}_t = \frac{1}{t} \sum_{i=1}^{t} \delta_{Z_i}$ is empirical distribution.

Average Coverage Guarantee

Corollary 8.3: Average Coverage of Online Conformal

Suppose $(X_1, Y_1), (X_2, Y_2), \ldots$ are exchangeable, the score function is symmetric, and the scores are distinct almost surely at each time t. Then:

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{Y_t \in C_t(X_t)\} \to 1 - \alpha \text{ as } T \to \infty \text{ a.s.}$$

Proof Sketch

- By Proposition 8.1: $\mathrm{err}_1, \mathrm{err}_2, \ldots$ are independent
- Each $\mathbb{E}[\mathrm{err}_t] \in (\alpha - 1/t, \alpha]$ (minor variation)
- Apply Law of Large Numbers
- Can refine with Hoeffding's inequality for finite-sample bounds

Without Distinct Scores Assumption

If we don't assume distinct scores:

- p-values p_t may no longer be independent
- Conformal prediction becomes more conservative with ties
- Conservative version of Corollary 8.3 still holds

Corollary 8.3 doesn't hold

If $(X_1, Y_1), (X_2, Y_2), \ldots$ are exchangeable ONLY, then:

$$\liminf_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\{Y_t \in C_t(X_t)\} \geq 1 - \alpha \text{ a.s.}$$

Online Conformal Without Online Training?

Note

Throughout Section 8.1, we assumed:

- At time t: (X_t, Y_t) is the test point
- At time $t+1$ (and beyond): (X_t, Y_t) is part of training set

Alternative Setting

Sometimes not possible/desirable to add test data to training set:

- Fix training set: $\{(X_i, Y_i)\}_{i \in [n]}$
- Test points $t = n+1, n+2, \ldots$ all compared to same training set
- Conformal p-values $p_{n+1}, p_{n+2}, \ldots$ will be **dependent**
- See Chapter 10.2 for analysis of this dependence

Contents

Motivation: Detecting Distribution Shift

Why Test Exchangeability?

- Monitor online deployment of learning algorithms
- Identify presence of distribution shift
- Especially harmful shifts that affect algorithm errors
- Early detection of changepoints

Main Idea

- Use conformal p-values from definition (8.3)
- Combine them into a **supermartingale**
- Under exchangeability: statistic stays small
- Large values = evidence against exchangeability

Supermartingales

Definition 8.4: Supermartingale

A sequence $M_1, M_2, \ldots$ is a **supermartingale** if for all t:

1. $\mathbb{E}[|M_t|] < \infty$
2. $\mathbb{E}[M_t | M_1, \ldots, M_{t-1}] \leq M_{t-1}$ for all $t \geq 2$

- Sequence whose *conditional expectation* is getting no larger over time
- If equality in inequality 2: **martingale**
- Supermartingales tend to take small values

Q3: Examples of Supermartingales

Question

Supermartingale의 예시를 몇 가지만 더 들어주실 수 있나요?

Example 1: Fair Game [2]

- Gambler starts with wealth $M_0 = w$
- At each time t: bet \$1, win with prob $1/2$, lose with prob $1/2$
- Wealth: $M_t = M_{t-1} + X_t$ where $X_t \in \{-1, +1\}$ with equal probability
- This is a **martingale** : $\mathbb{E}[M_t | M_1, \ldots, M_{t-1}] = M_{t-1} + \mathbb{E}[X_t] = M_{t-1}$

Q3: Examples of Supermartingales (cont.)

Example 2: Unfavorable Game [2]

- Same setup, but now win with prob $p < 1/2$, lose with prob $1 - p$
- $\mathbb{E}[M_t | M_1, \ldots, M_{t-1}] = M_{t-1} + p - (1-p) = M_{t-1} + (2p-1) < M_{t-1}$
- This is a **supermartingale** (unfavorable game)
- Wealth tends to decrease over time

Example 3: Stopped Martingale [2]

- Let $M_1, M_2, \ldots$ be a martingale
- Define stopping time T (e.g., first time $M_t \geq$ threshold)
- Stopped process: $M_t^T = M_{\min(t,T)}$
- Then M_t^T is a **supermartingale**

Q3: Examples of Supermartingales (cont.)

Example 4: Non-negative Submartingale Reciprocal

- If M_t is a non-negative submartingale: $\mathbb{E}[M_t|\mathcal{F}_{t-1}] \geq M_{t-1}$
- Then $N_t = 1/M_t$ is a supermartingale (by Jensen's inequality)

Example 5: Product of Independent Superuniform RVs

- Let $U_1, U_2, \ldots$ be independent, superuniform (i.e., $\mathbb{P}(U_t \leq u) \leq u$)
- Define $M_t = \prod_{i=1}^{t}(1/U_i)$
- Then M_t is a supermartingale
- **This is the type used in Proposition 8.5!**

Q4: Understanding Supermartingales

Question

Supermartingale은 교환가능성이 성립할 때 작아진다는 것과, Supermartingale의 평균이 시간이 흐름에 따라 더 커지지 않는다는 게 이해가 잘 안 갑니다.

Answer: Supermartingale is defined on conditional expectation

Supermartingale does NOT mean:

- $M_t < M_{t-1}$ always (deterministically decreasing)
- M_t cannot increase

Supermartingale DOES mean:

- $\mathbb{E}[M_t|\text{past}] \leq M_{t-1}$ (expected to not increase)
- Individual realizations can go up or down
- But the "drift" is downward (or flat for martingales)

Contents

Conclusion

- 안녕

Thank you!

References I

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

[2] Rick Durrett.
Probability: Theory and examples.
Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.