

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 Testing Exchangeability Online

- Supermartingale
- Construct super MG to test exchangeability
- Conformal p-values and E-values

4 Prediction sets for adversarial sequences

Contents

1 Introduction

2 Online CP with Exchangeable Data

3 Testing Exchangeability Online

- Supermartingale
- Construct super MG to test exchangeability
- Conformal p-values and E-values

4 Prediction sets for adversarial sequences

From Batch to Online Setting

Batch Setting (Previous Chapters)

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

Online Setting (This Chapter)

- Observe data points **sequentially**
- At time t : observed $(X_1, Y_1), \dots, (X_{t-1}, Y_{t-1})$ 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

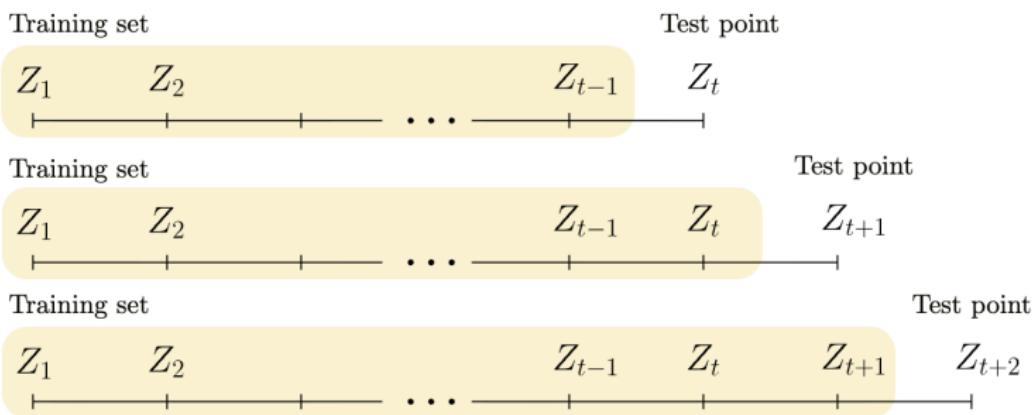


Figure 1: Online prediction loop — observe X_t , produce $\mathcal{C}_t(X_t)$, reveal Y_t , and update the dataset. [1]

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

$X_1, X_2, X_3, \dots, X_t$ 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: $C_1(X_1), C_2(X_2), \dots, C_{24}(X_{24})$
- 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

1 Introduction

2 Online CP with Exchangeable Data

3 Testing Exchangeability Online

- Supermartingale
- Construct super MG to test exchangeability
- Conformal p-values and E-values

4 Prediction sets for adversarial sequences

Setup: Exchangeable Sequences

Assumptions

- Sequence $(X_1, Y_1), (X_2, Y_2), \dots, (X_T, Y_T)$ is **exchangeable**
- At time t : observed $(X_1, Y_1), \dots, (X_{t-1}, Y_{t-1})$ 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: $(X_1, Y_1), \dots, (X_{t-1}, Y_{t-1})$
- Test point: X_t
- Score function: s (symmetric)

Miscoverage Events

Definition

Define the miscoverage indicator at time t :

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

From Theorem 3.2(Marginal Coverage), we know:

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

Question

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

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

Online Prediction Guarantee

Proposition 8.1: Independence of Errors

Suppose $(X_1, Y_1), (X_2, Y_2), \dots, (X_T, Y_T)$ 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 Prediction 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 $\mathcal{D}_t = ((X_1, Y_1), \dots, (X_t, Y_t))$.

Theorem 8.2: Independence of Online Conformal p-values

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

We can easily check that $\text{err}_t = \mathbb{1}\{p_t \leq \alpha\}$ (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, \dots, Z_t
- Z_1, \dots, 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: $p_t | (\hat{P}_t, Z_{t+1}, \dots, Z_T) \sim \text{Unif}(\{1/t, \dots, 1\})$
- 2 Show: For $t' > t$, $p_{t'}$ is a function of $(\hat{P}_t, Z_{t+1}, \dots, Z_T)$
- 3 Combine: $p_t | (p_{t+1}, \dots, p_T) \sim \text{Unif}(\{1/t, \dots, 1\})$
- 4 Therefore: p_1, \dots, 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 $(X_1, Y_1), (X_2, Y_2), \dots$ are exchangeable, the score function is symmetric, and the scores are distinct almost surely at each time t . Then:

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

Proof Sketch

- By Proposition 8.1: $\text{err}_1, \text{err}_2, \dots$ are independent
- Each $\mathbb{E}[\text{err}_t] \in (\alpha - 1/t, \alpha]$ (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 $(X_1, Y_1), (X_2, Y_2), \dots$ are exchangeable ONLY, then:

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

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 $t = n + 1, n + 2, \dots$ all compared to same training set
- Conformal p-values p_{n+1}, p_{n+2}, \dots will be **dependent**
- See Chapter 10.2 for analysis of this dependence

Contents

1 Introduction

2 Online CP with Exchangeable Data

3 Testing Exchangeability Online

- Supermartingale
- Construct super MG to test exchangeability
- Conformal p-values and E-values

4 Prediction sets for adversarial sequences

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, \dots is a **supermartingale** if for all t :

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

- Sequence whose *conditional expectation* is getting no larger over time
- If equality in (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** :

$$\mathbb{E}[M_t | M_1, \dots, M_{t-1}] = M_{t-1} + \mathbb{E}[X_t] = M_{t-1}$$

Q3: Examples of Supermartingales (cont.)

Example 2: Unfavorable Game [2]

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

Example 3: Stopped Martingale [2]

- Let M_1, M_2, \dots be a martingale
- Define stopping time T (e.g., first time $M_t \geq \text{threshold}$)
- 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, \dots 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

- Exchangeability \Rightarrow Making supermartingale (Prop 8.5)
- Supermartingale is defined on conditional expectation

Q4: Understanding Supermartingales (cont.)

Recall: Definition of Supermartingale

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

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

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)

Ville's Inequality

Ville's Inequality

If M_1, M_2, \dots is a nonnegative supermartingale, then for any $a > 0$:

$$\mathbb{P}\left(\sup_{t \geq 1} M_t \geq a\right) \leq \frac{\mathbb{E}[M_1]}{a}$$

- Stronger than Markov's inequality (which only applies at single time)
- Holds uniformly over all time

A Simple Supermartingale from p-values

Proposition 8.5

Consider conformal p-values p_1, p_2, \dots defined at (8.3). Let $\lambda \in [0, 1]$ be fixed. If scores are distinct a.s. at each time t , then:

$$M_t = \prod_{t'=1}^t \frac{1 - \lambda p_{t'}}{1 - \lambda/2}$$

is a supermartingale.

Proof Sketch

- $M_t = M_{t-1} \cdot \frac{1 - \lambda p_t}{1 - \lambda/2}$
- By Theorem 8.2: p_t is superuniform and independent of M_1, \dots, M_{t-1}
- Therefore: $\mathbb{E}[1 - \lambda p_t | M_1, \dots, M_{t-1}] \leq 1 - \lambda/2$

General Recipe for Supermartingales

Theorem 8.6: Online Test for Exchangeability

Let $f_t : [0, 1] \rightarrow [0, \infty)$ be nonincreasing functions with $\int_0^1 f_t(r) dr \leq 1$. Define:

$$M_t = \prod_{t'=1}^t f_{t'}(p_{t'})$$

Under exchangeability, symmetric score, and distinct scores:

- 1 M_t is a supermartingale
- 2 For any $\alpha \in [0, 1]$: $\mathbb{P}(\sup_t M_t \geq 1/\alpha) \leq \alpha$

Proof Sketch (Theorem 8.6)

Part 1: M_t is a supermartingale

- Using similar flow in Proposition 8.5:
- Since f_t is nonincreasing: $\mathbb{E}[f_t(p_t)] \leq \int_0^1 f_t(r)dr \leq 1$
- Therefore: $\mathbb{E}[M_t|M_1, \dots, M_{t-1}] = M_{t-1} \cdot \mathbb{E}[f_t(p_t)] \leq M_{t-1}$

Part 2: Probability bound

- Apply Ville's inequality with $a = 1/\alpha$ and $\mathbb{E}[M_1] = \mathbb{E}[f_1(p_1)] \leq 1$

Online Test Algorithm

- 1 Choose function f_t as specified in Theorem 8.6
 - $f_t(r) = \frac{1-\lambda r}{1-\lambda/2}$ (Proposition 8.5)
 - $f_t(r) = 2(1-r)$ (simple choice)
- 2 For each time $t = 1, 2, \dots$:
 - 1 Observe (X_t, Y_t) and compute conformal p-value and M_t
 - 2 If $M_t \geq 1/\alpha$, reject exchangeability at significance level α . Otherwise, continue.

Visualization of Online Test

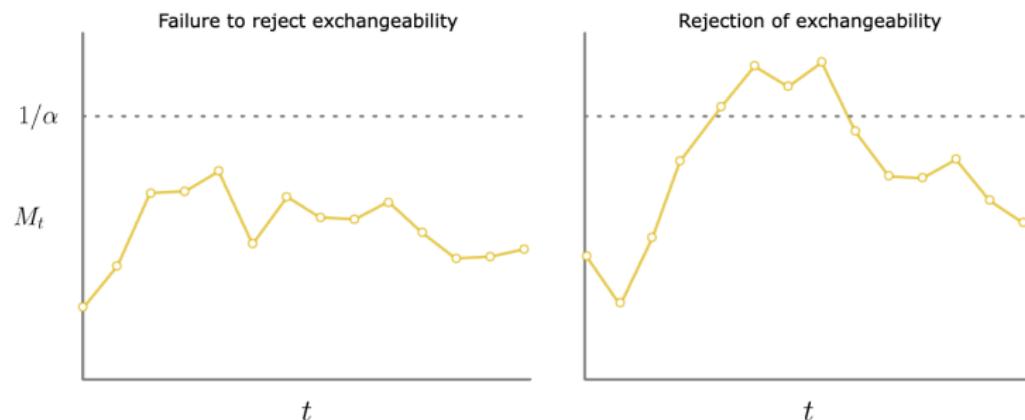


Figure 2: The online test for exchangeability — Under null (exchangeability): crossing happens with probability $\leq \alpha$ [1]

Why is Conformal Prediction Critical?

Key Feature: No Multiplicity Penalty

- Testing exchangeability at *every* time step $t = 1, 2, 3, \dots$
- Reusing the same data over and over
- Naively: would need Bonferroni correction \rightarrow no power
- **But:** Independence of conformal p-values (Theorem 8.2)!
- Can test at all times with *single* α level

Example Application

- Score: $s((x, y); \mathcal{D}) = y$
- p-value $p_t =$ rescaled rank of Y_t relative to Y_1, \dots, Y_{t-1}
- M_t grows if we frequently observe small p_t
- Detects: values of Y_t trending downward
- Application: changepoint detection in error rates

Q4: Relationship to E-values

Question

현재 online CP의 setting과, proposition 8.5. (or Thm 8.6.) 이 제가 생각하기로는 E-value 와 상당히 유사해 보입니다. 둘 사이의 관계에 대해 자세한 설명 부탁드립니다.

Answer @ Bibliographic notes

In particular, we note that the exchangeability supermartingale in Proposition 8.5 is a special case of an e-value, and these are useful more broadly. [3]

E-values: A Better Alternative of P-values

Definition: E-value (Evidence Value)

A random variable $E \geq 0$ is an **e-value** for testing H_0 if:

$$\mathbb{E}_{H_0}[E] \leq 1$$

Interpretation:

- Under null hypothesis, expected value ≤ 1
- Large values of E provide evidence against H_0

Testing with E-values

For any $\alpha \in (0, 1)$, by Markov's inequality:

$$\mathbb{P}_{H_0}(E \geq 1/\alpha) \leq \frac{\mathbb{E}_{H_0}[E]}{1/\alpha} \leq \alpha$$

Decision rule: Reject H_0 if $E \geq 1/\alpha$

E-values vs P-values

Property	P-value	E-value
Range	$[0, 1]$	$[0, \infty)$
Under H_0	Uniform $[0, 1]$	$\mathbb{E}[E] \leq 1$
Evidence direction	Small = evidence	Large = evidence
Testing threshold	$p \leq \alpha$	$E \geq 1/\alpha$
Sequential testing	Problematic	Natural
Combination	Complex	Multiply

Converting between them

If p is a valid p-value, then $E = 1/p$ is **NOT** necessarily an e-value.
 But if p is *superuniform*: $\mathbb{P}(p \leq u) \leq u$, then for nonincreasing f with $\int_0^1 f(p)dp \leq 1$, we have $E = f(p)$ is an e-value. (proved [3])

Contents

1 Introduction

2 Online CP with Exchangeable Data

3 Testing Exchangeability Online

- Supermartingale
- Construct super MG to test exchangeability
- Conformal p-values and E-values

4 Prediction sets for adversarial sequences

Beyond Exchangeability

Limitations of Exchangeability Assumption

- Real data often not exchangeable
- Distributions change over time (concept drift)
- Data might even be deterministic (not random)

New Goal: Long-Run Coverage

Without distributional assumptions, aim for:

$$\left| \frac{1}{T} \sum_{t=1}^T \text{err}_t - \alpha \right| \rightarrow 0$$

- Average miscoverage rate $\approx \alpha$
- No assumption of exchangeability or even randomness!
- Works for **adversarial sequences**



Prediction Set Construction

General Form

Prediction sets of the form:

$$\mathcal{C}_t(X_t) = \{y : s_t(X_t, y) \leq q_t\}$$

where:

- s_t : score function (may depend on previous data)
- q_t : threshold (updated online)

Connection to Conformal Prediction

- **Split conformal:** $s_t(x, y) = s(x, y)$ (fixed), q_t = quantile
- **Full conformal:** $s_t(x, y) = s((x, y); \mathcal{D}_{t-1})$, q_t = quantile
- **Adversarial setting:** Allow arbitrary s_t , update q_t online

Quantile Tracking Algorithm

Update Rule

Initialize $q_1 \in [0, B]$ arbitrarily. For $t \geq 2$:

$$q_{t+1} = q_t + \eta_t (\text{err}_t - \alpha)$$

where $\eta_t > 0$ is the step size.

- If $\text{err}_t = 1$ (miscoverage): *increase* $q_{t+1} \rightarrow$ more conservative
- If $\text{err}_t = 0$ (coverage): *decrease* $q_{t+1} \rightarrow$ less conservative
- Adaptive feedback mechanism
- No distributional assumptions needed!

Q6: Quantile Tracking Algorithm as Optimization

Question

소단원 8.3에서 제시하는 quantile tracking 알고리즘의 update rule (식 8.8)을 보면 형태가 quantile loss에 대한 gradient descent (혹은 gradient ascent)와 닮아 보입니다. 혹시 식 8.8을 어떤 loss function의 최적화 알고리즘으로 생각할 수 있을까요?

Answer

We can interpret the quantile tracking update as a form of stochastic gradient descent on the quantile loss function.

Stochastic Gradient Descent Interpretation

Quantile Loss: The α -quantile can be found by minimizing:

$$L_\alpha(q) = \mathbb{E}[\rho_\alpha(Y - q)]$$

where $\rho_\alpha(u) = u(\alpha - \mathbb{1}\{u < 0\})$ is the "check function" or "pinball loss". Also, we can find **subgradient** as:

$$\partial_q \rho_\alpha(Y_t - q_t) = \begin{cases} -\alpha & \text{if } Y_t > q_t \text{ (coverage, } \text{err}_t = 0\text{)} \\ 1 - \alpha & \text{if } Y_t < q_t \text{ (miscoverage, } \text{err}_t = 1\text{)} \end{cases}$$

Stochastic Gradient Descent Interpretation (cont.)

Connection to SGD

The update rule $q_{t+1} = q_t + \eta_t(\text{err}_t - \alpha)$ is exactly:

$$q_{t+1} = q_t - \eta_t \cdot \partial_q \rho_\alpha(Y_t - q_t)$$

This is stochastic gradient descent on quantile loss!

- Objective: track the α -quantile of score distribution
- Update: use subgradient based on current observation
- Step size η_t : learning rate

Main Result: Deterministic Coverage Guarantee

Theorem 8.7

Consider an **arbitrary** sequence of data points $(X_1, Y_1), (X_2, Y_2), \dots$ and **arbitrary** score functions s_1, s_2, \dots with $s_t(x, y) \in [0, B]$. Let step sizes η_t be any nonincreasing positive sequence. Then for all $T \geq 1$:

$$\frac{1}{T} \sum_{t=1}^T \text{err}_t \in \left[\alpha \pm \frac{B + \eta_1}{\eta_T T} \right]$$

- **Deterministic** guarantee (no probability!)
- For constant step size $\eta_t = \eta$: error concentrates at rate $O(1/T)$
- Holds for *any* data sequence and *any* choice of scores
- Works under distribution shift

Proof Sketch (Theorem 8.7)

Part 1: The iterates are bounded.

We can check $q_t \in [-\eta_1\alpha, B + \eta_1(1 - \alpha)]$ by contradiction.

Part 2: Simultaneous bounds on the weighted coverage gap.

$$\left| \sum_{t=T_0}^{T_1} \eta_t (\text{err}_t - \alpha) \right| = |q_{T_1+1} - q_{T_0}| \leq B + \eta_1$$

Part 3: Bounding the long-run coverage gap.

Visual intuition of boundness

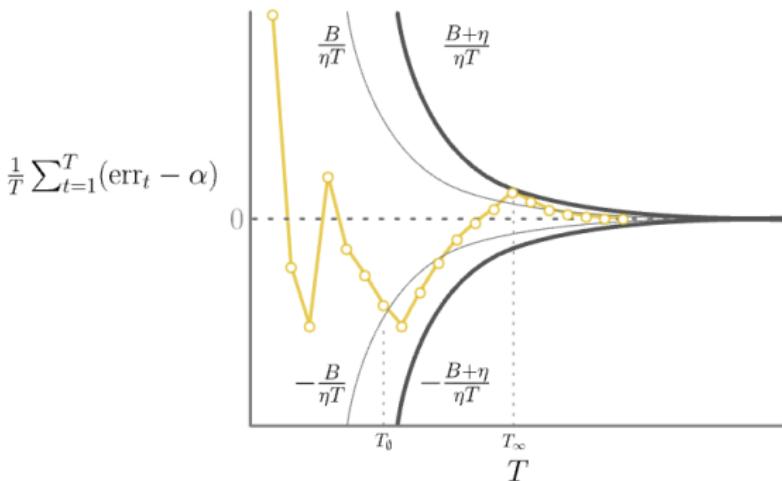


Figure 3: The long-run coverage gap — Once trajectory leaves thin envelope (where q_T would be outside $[0, B]$), next prediction forces it back toward α . [1]

Choice of Step Sizes

Constant Step Size: $\eta_t = \eta$

Pros:

- Fast adaptation to changing score sequences
- Good for highly dynamic systems

Cons:

- Quantiles q_t fluctuate wildly
- Many infinite-size sets ($q_t > B$) and empty sets ($q_t < 0$)
- Never converges even in i.i.d. setting

Choice of Step Sizes

Decaying Step Size: $\eta_t \propto t^{-(1/2+\epsilon)}$

Pros:

- Stabilizes quantile over time
- Can converge in stationary settings (by previous theorem)

Cons:

- Slower adaptation: average error $\sim T^{-(1/2-\epsilon)}$ at time T

I.I.D. Setting: Convergence Results

Theorem 8.8

Suppose $(X_t, Y_t) \stackrel{\text{i.i.d.}}{\sim} P$ and scores s_t trained online and F_s is CDF of $s(X, Y)$ under $(X, Y) \sim P$.

Constant step size $\eta_t = \eta$:

$$\liminf_{t \rightarrow \infty} F_{s_t}(q_t) = 0, \quad \limsup_{t \rightarrow \infty} F_{s_t}(q_t) = 1$$

(infinitely many oscillations)

Decaying step size η_t with $\sum_t \eta_t = \infty$, $\sum_t \eta_t^2 < \infty$:

If $F_{s_t} \rightarrow F$ then $F_{s_t}(q_t) \rightarrow 1 - \alpha$

(convergence to desired coverage)

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.
- [3] Johannes Ruf, Martin Larsson, Wouter M Koolen, and Aaditya Ramdas.
A composite generalization of ville's martingale theorem using e-processes.
28.