DCP 1206: Probability

Lecture 23 — Law of Large Numbers

Ping-Chun Hsieh

December 11, 2019

Announcements

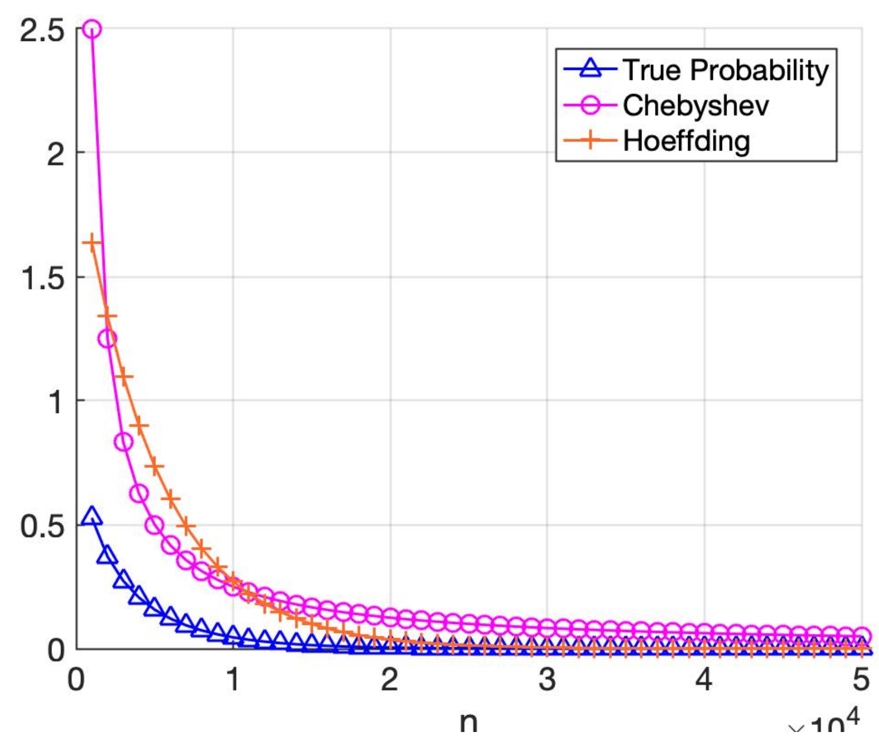
- Class on 12/25 (Wed): 10:10am-**11:40**am
- Class on 12/13 (Fri): 3:30pm-**4:30**pm (10min extension)
- Class on 12/20 (Fri): 3:30pm-**4:30**pm (10min extension)

Review: Hoeffding's Inequality (Formally)

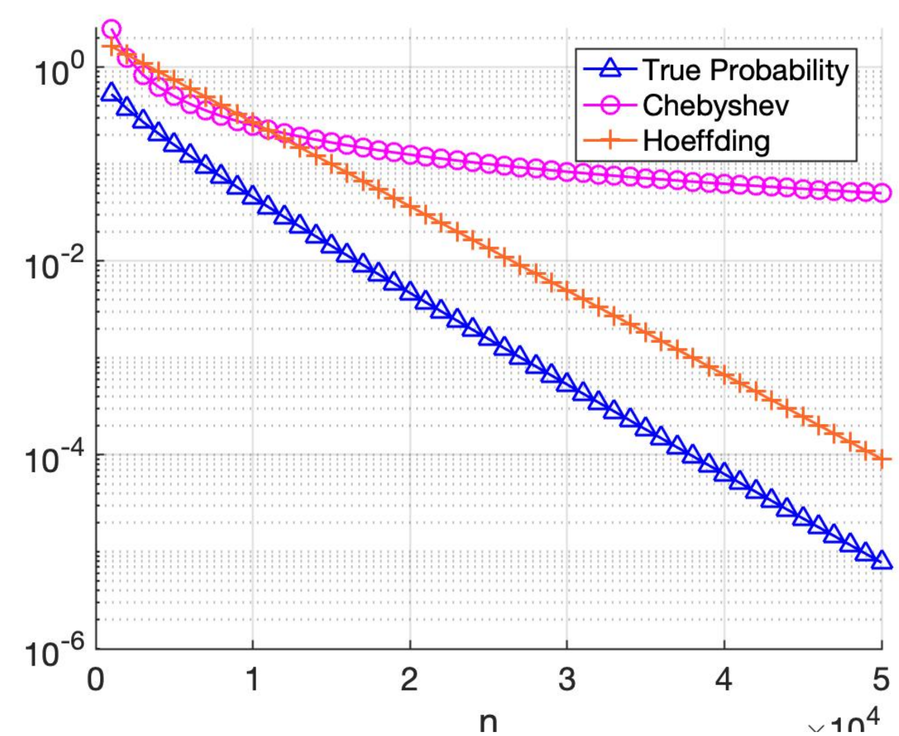
- **Hoeffding's Inequality (For Bernoulli)**: Let $X_1, \cdots, X_n$ be a sequence of i.i.d. Bernoulli random variables with parameter p. Define $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Then, for any $\varepsilon > 0$, we have

$$P(|\bar{X} - p| \geq \varepsilon) \leq 2\exp(-2n\varepsilon^2)$$

$\underline{\varepsilon = 0.01}$



log scale →



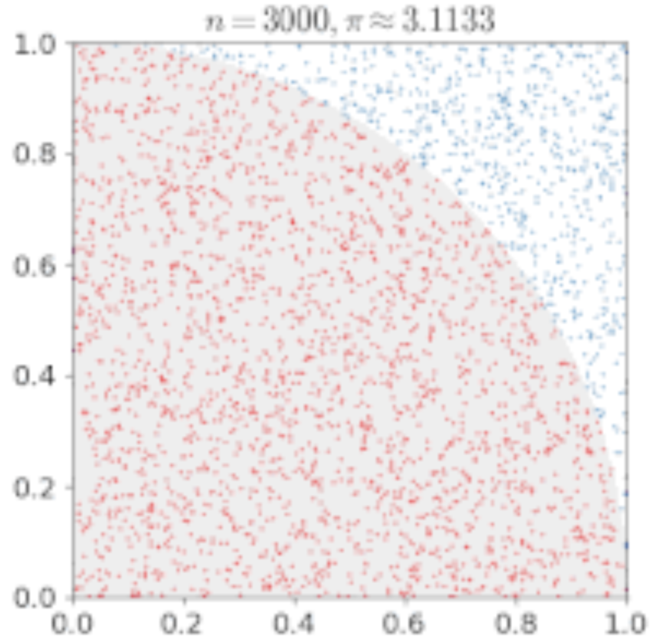
Monte-Carlo Method

(Rafael Nadal: 11 titles at Monte Carlo Masters)

- What is "Monte-Carlo method"?

"… computational algorithms that rely on repeated random sampling to obtain numerical results… use randomness to solve problems that might be deterministic in principle." (by Wikipedia)

- Math principle behind Monte-Carlo?
- Use Monte-Carlo to estimate π



This Lecture

1. Weak Law of Large Numbers (WLLN) and Convergence in Probability

2. Strong Law of Large Numbers (SLLN) and Almost-Sure Convergence

- Reading material: Chapter 11.4

1. Weak Law of Large Numbers (WLLN)

Review: Chebyshev's and Sample Mean: $n \to \infty$

- **Chebyshev's and Sample Mean**: Let $X_1, \cdots, X_n$ be a sequence of <u>independent and identically distributed</u> (i.i.d.) random variables with mean μ and variance σ^2. Define $S_n = (X_1 + \cdots + X_n)$. Then, for any $\varepsilon > 0$, we have

$$P\left(\left| \frac{S_n}{n} - \mu \right| \geq \varepsilon \right) \leq \frac{\sigma^2}{\varepsilon^2 n}$$

- What if we let $n \to \infty$?

$$P\left(\left| \frac{S_n}{n} - \mu \right| \geq \varepsilon \right) \underset{=}{\cancel{\leq}} 0 \quad \text{as } n \to \infty$$

The Weak Law of Large Numbers (WLLN)

- **The Weak Law of Large Numbers (Khinchin's Law)**: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ. Define $S_n = (X_1 + \cdots + X_n)$. Then, for every $\varepsilon > 0$, we have

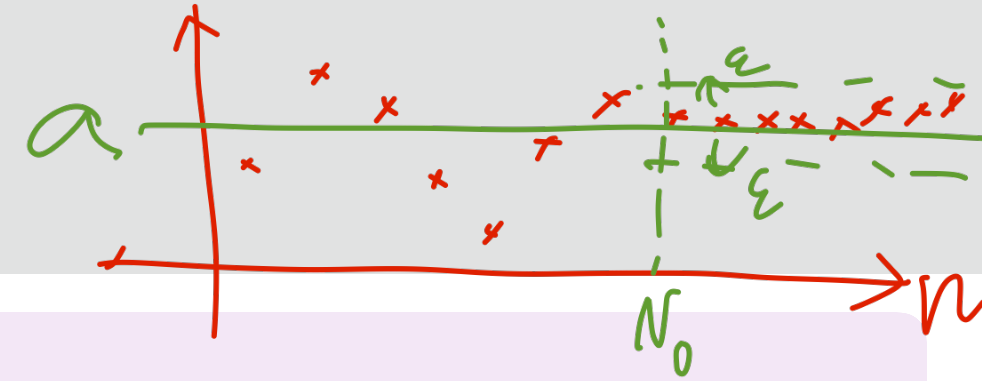
$$P\left(\left| \frac{S_n}{n} - \mu \right| \geq \varepsilon \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

- Question: Any change in technical conditions (cf: Chebyshev's)?

No assumption on "finite variance"

- Question: What does "convergence" mean here?

Convergence in Probability



- **Convergence of a Deterministic Sequence**: Let $a_1, a_2 \cdots$ be a sequence of real numbers. We say that a_n converges to a if for every $\varepsilon > 0$, there exists N_0 such that

$$|a_n - a| \leq \varepsilon \qquad \text{for all } n \geq N_0$$

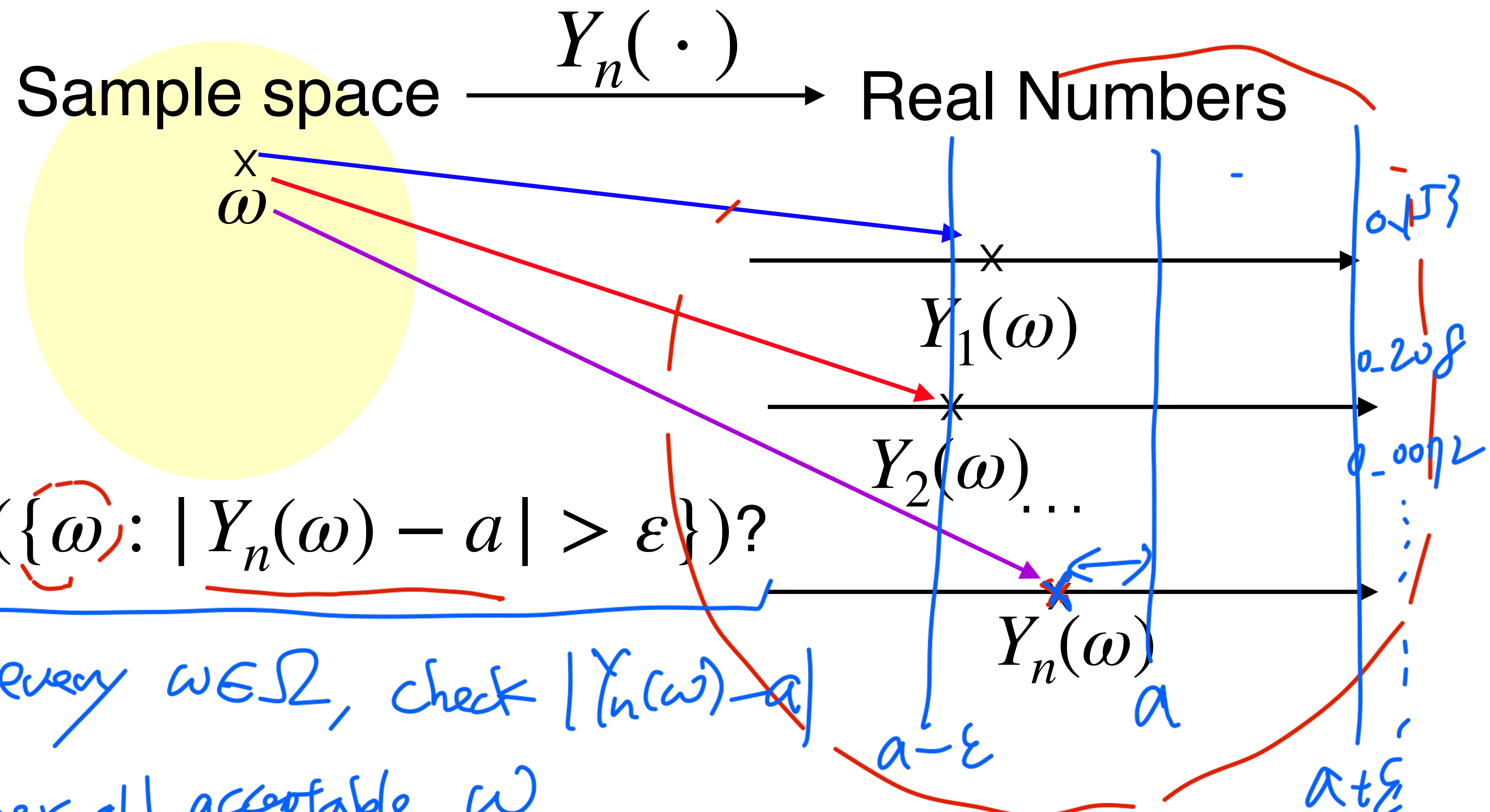
- **Convergence to a Scalar in Probability**: Let $Y_1, Y_2 \cdots$ be a sequence of random variables, and let a be a real number. We say that Y_n converges to a in probability if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P(\{\omega : |Y_n(\omega) - a| > \varepsilon) = 0$$

- Question: How to interpret this definition?

Recall: Random Variables Defined on Ω

- $Y_1, Y_2, \cdots, Y_n, \cdots$ are defined on the same sample space Ω



- How to interpret $P(\{\omega : |Y_n(\omega) - a| > \varepsilon\})$?

step 1: for every $\omega \in \Omega$, check $|Y_n(\omega) - a|$

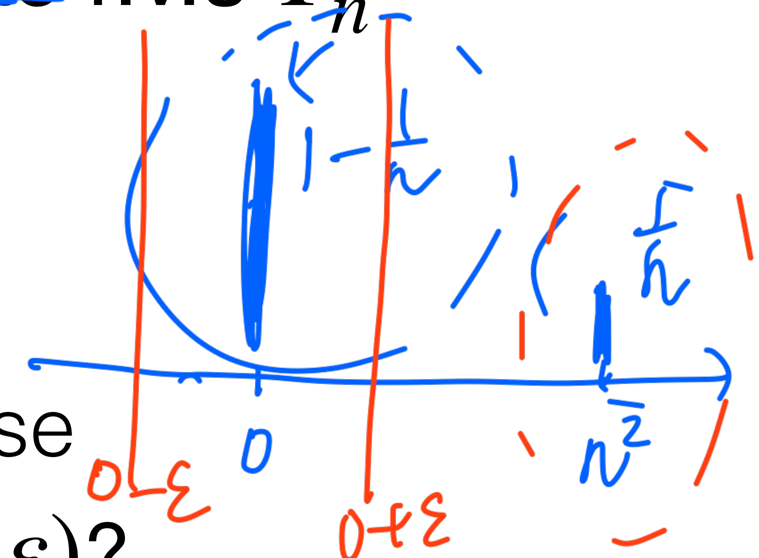
step 2: Gather all acceptable ω

- How about $\boxed{\lim_{n\to\infty} P(\{\omega : |Y_n(\omega) - a| > \varepsilon\}) = 0?}$

Example: Convergence in Probability

- Example: Consider a sequence of ~~discrete~~ r.v.s Y_n

$$P(Y_n = y) = \begin{cases} 1 - \frac{1}{n} & , \text{ if } y = 0 \\ \frac{1}{n} & , \text{ if } y = n^2 \\ 0 & , \text{ otherwise} \end{cases}$$



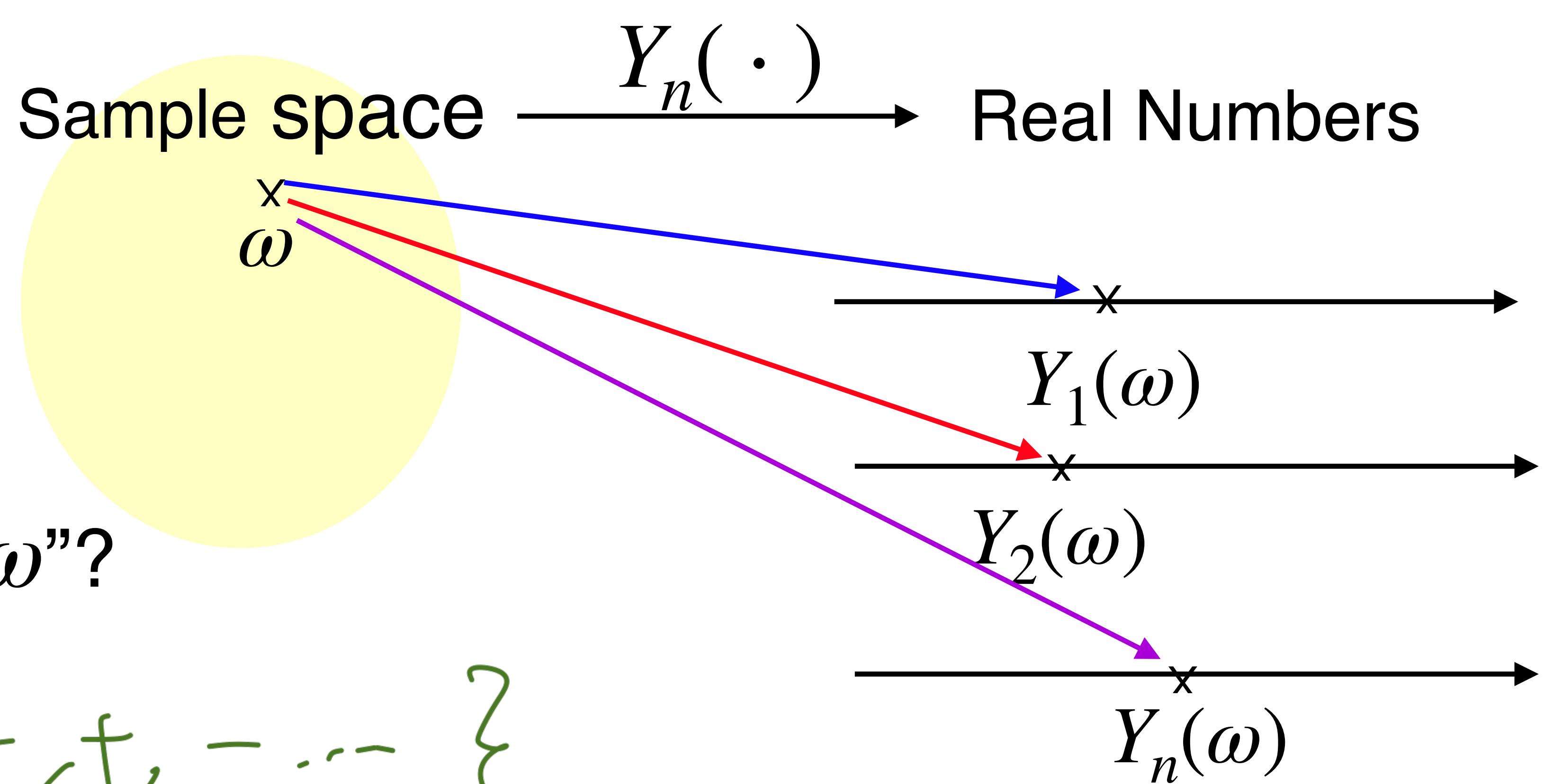
- For every $\varepsilon > 0$, can we find $\underline{P(|Y_n - 0| > \varepsilon)}$?
- How about $\lim_{n\to\infty} P(|Y_n - 0| > \varepsilon)$?

$$P(\{\omega : |Y_n(\omega) - 0| > \varepsilon\}) \le \frac{1}{n}, \quad \forall \varepsilon > 0, \ \forall n \ge 1$$

$$\Rightarrow \lim_{n\to\infty} P(\{\omega : |Y_n(\omega) - 0| > \varepsilon\}) = 0, \quad \forall \varepsilon > 0$$

How to Interpret WLLN?

- Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with mean μ
- Define $Y_n = (X_1 + X_2 \cdots, + X_n)/n$
- WLLN: $\lim_{n\to\infty} P(\{\omega : |Y_n(\omega) - \mu| > \varepsilon\}) = 0$, $\forall \varepsilon > 0$



- Question: What is an "ω"?

$\omega \equiv \{+, +, -, +, -, +, - \ldots - \}$

(a series of infinite outcomes)

Rewriting WLLN (More Formally)

- **The Weak Law of Large Numbers (Khinchin's Law)**: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ. Define $S_n = (X_1 + \cdots + X_n)$. Then, for every $\varepsilon > 0$, we have

$$\lim_{n \to \infty} P\Big(\{\omega : |\frac{S_n(\omega)}{n} - \mu| \geq \varepsilon\} \Big) = 0$$

In short, we have $\frac{S_n}{n} \xrightarrow{p} \mu$

$\frac{S_n}{n}$ converges to μ in probability

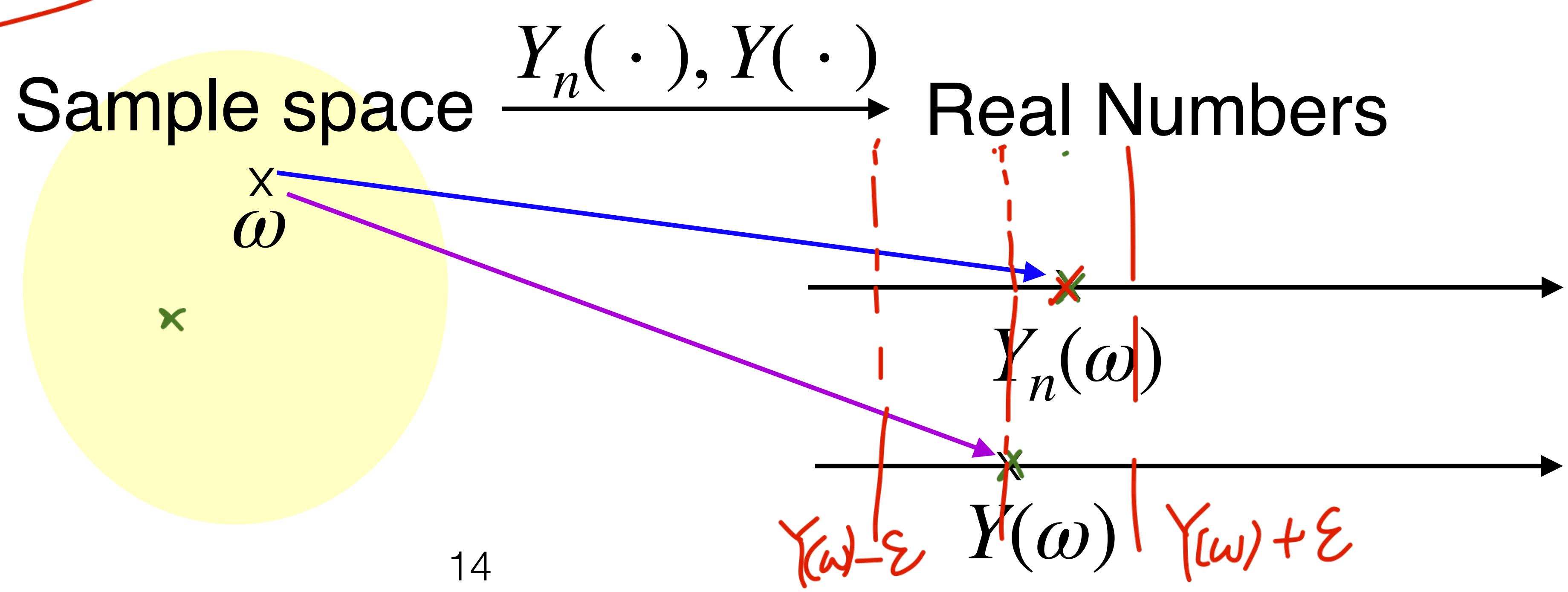
Convergence in Probability (Cont.)

- **Convergence to a Random Variable in Probability**: Let $Y_1, Y_2 \cdots$ be a sequence of random variables defined on a sample space. We say that Y_n converges to a random variable Y in probability if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P(\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}) = 0$$

先找機率. 後取limit

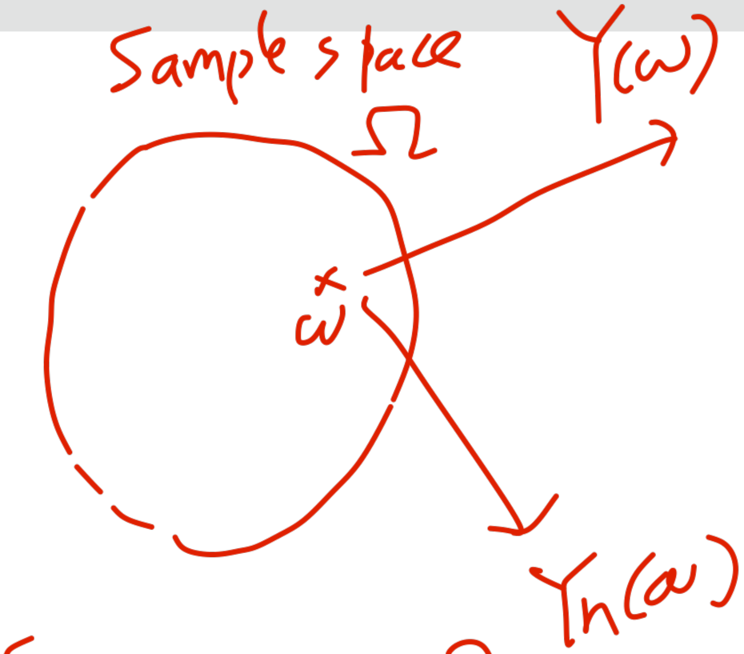
- Notation: $Y_n \xrightarrow{P} Y$
- Interpretation:



Example: Convergence in Probability

- Example: Consider a random variable Y

$$P(Y=y)=\begin{cases} 1/2 & \text{, if } y=0 \\ 1/2 & \text{, if } y=1 \\ 0 & \text{, otherwise} \end{cases}$$



- For every $n \in \mathbb{N}$, define $\underbrace{Y_n = (1+\frac{1}{n})Y}$ $\iff$ for every $\omega \in \Omega$, $Y_n(\omega) = (1+\frac{1}{n})Y(\omega)$
- Do we have $Y_n \xrightarrow{p} Y$ (i.e. $\underbrace{\lim_{n\to\infty} P(|Y_n - Y| > \varepsilon) = 0}$)?

$$P(|Y_n - Y| > \varepsilon) = P(\{\omega : |\underbrace{Y_n(\omega) - Y(\omega)}| > \varepsilon\})$$

$$= P(\{\omega : |\tfrac{1}{n}Y(\omega)| > \varepsilon\}) = 0 \text{ for all } n > \tfrac{1}{\varepsilon}$$

$$\lim_{n\to\infty} P(|Y_n - Y| > \varepsilon) = 0, \ \forall \varepsilon > 0$$

Any Stronger Notion of Convergence?

Example: Convergence With Probability 1

- Example: A sequence of i.i.d. continuous r.v.s $X_n \sim$ Unif$(0,1)$
  - For every n, define $Y_n = \min\{X_1, X_2, \cdots, X_n\}$
  - Question: Can we find $P(\{\omega : Y_n(\omega) \geq \varepsilon\}) = ?$ $(\varepsilon \in (0,1))$

$$P(\{\omega : Y_n(\omega) \geq \varepsilon\}) = P(\{\omega : X_1(\omega) \geq \varepsilon, X_2(\omega) \geq \varepsilon, \cdots, X_n(\omega) \geq \varepsilon\})$$

$$= P(X_1 \geq \varepsilon) \cdot P(X_2 \geq \varepsilon) \cdots \cdot P(X_n \geq \varepsilon)$$

$$= (1-\varepsilon)^n.$$

Example: Convergence With Probability 1 (Cont.)

- Example: A sequence of i.i.d. continuous r.v.s $X_n \sim$ Unif$(0,1)$
  - Define $Y_n = \min\{X_1, X_2, \cdots, X_n\}$
  - Question: How about $P(\{\omega : \lim_{n\to\infty} Y_n(\omega) \geq \varepsilon\}) = ?$ $(\varepsilon \in (0,1))$

$$P(\{\omega : \lim_{n\to\infty} Y_n(\omega) \geq \varepsilon\})$$

$$= P(\{\omega : X_1(\omega) \geq \varepsilon, X_2(\omega) \geq \varepsilon, \cdots, X_n(\omega) \geq \varepsilon, \cdots\})$$

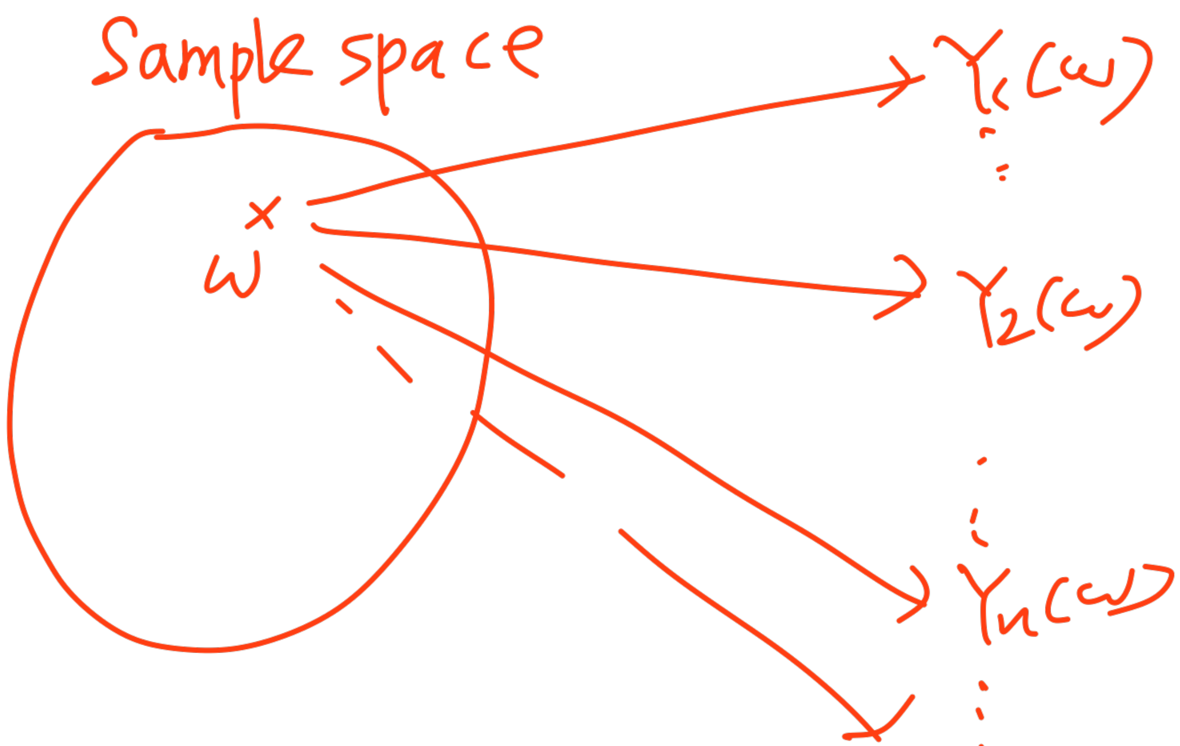
$$= P(X_1 \geq \varepsilon) \cdot P(X_2 \geq \varepsilon) \cdots P(X_n \geq \varepsilon) \cdots$$

$$= \prod_{i=1}^{\infty} \underbrace{P(X_i \geq \varepsilon)}_{1-\varepsilon} = 0, \quad \cancel{\forall \varepsilon \in (0,1)} \quad \forall \varepsilon > 0$$

Example: Convergence With Probability 1 (Cont.)

- Example: A sequence of i.i.d. continuous r.v.s $X_n \sim$ Unif$(0,1)$
  - Define $Y_n = \min\{X_1, X_2, \cdots, X_n\}$
  - Question: How about $\boxed{P(\{\omega : \lim_{n\to\infty} Y_n(\omega) = 0\})}$ = ? 1

Last page: $\boxed{P(\{\omega : \lim_{n\to\infty} Y_n(\omega) \geq \varepsilon\}) = 0}$ $\forall \varepsilon > 0$



Almost-Sure Convergence / Convergence With Probability 1

- **Convergence to a Random Variable in Probability**: Let $Y_1, Y_2 \cdots$ be a sequence of random variables. We say that Y_n converges to a random variable Y in probability if $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}) = 0$$

先找机率，後取limit

- **Convergence to a Random Variable Almost Surely**: Let $Y_1, Y_2 \cdots$ be a sequence of random variables. We say that Y_n converges to a random variable Y almost surely if,

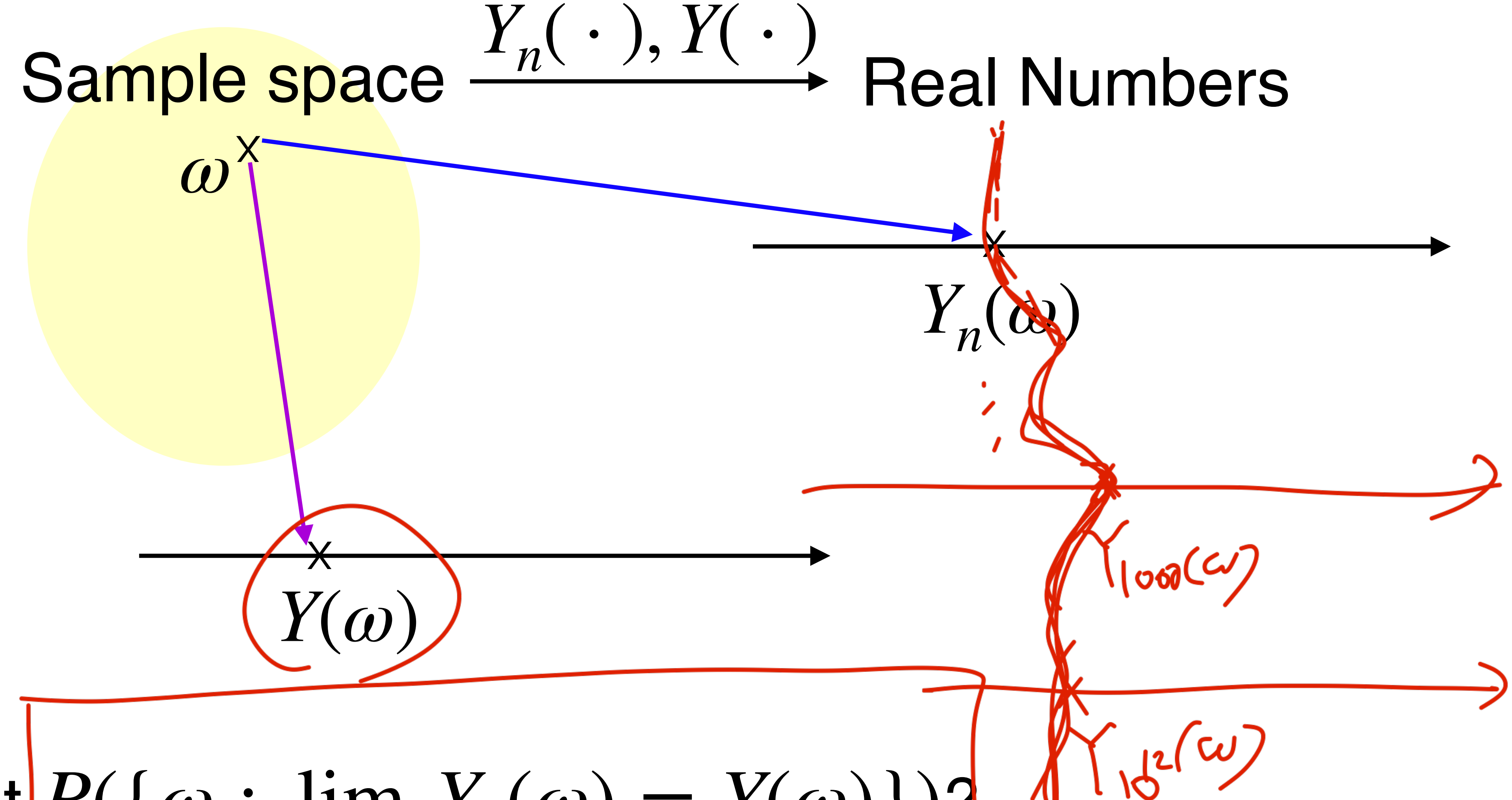
$$P\left(\left\{\omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\right\}\right) = 1$$

先取limit，後找机率

- **Notation:** $Y_n \xrightarrow{\text{a.s.}} Y$ or $\boxed{Y_n \to Y, \text{ w.p.1}}$

Interpretation of Almost-Sure Convergence

- $Y_1, Y_2, \cdots, Y_n, \cdots$ are defined on the <u>same sample space Ω</u>



- How to interpret $\boxed{P(\{\omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\})?}$

Equivalent Definition of Almost-Sure Convergence

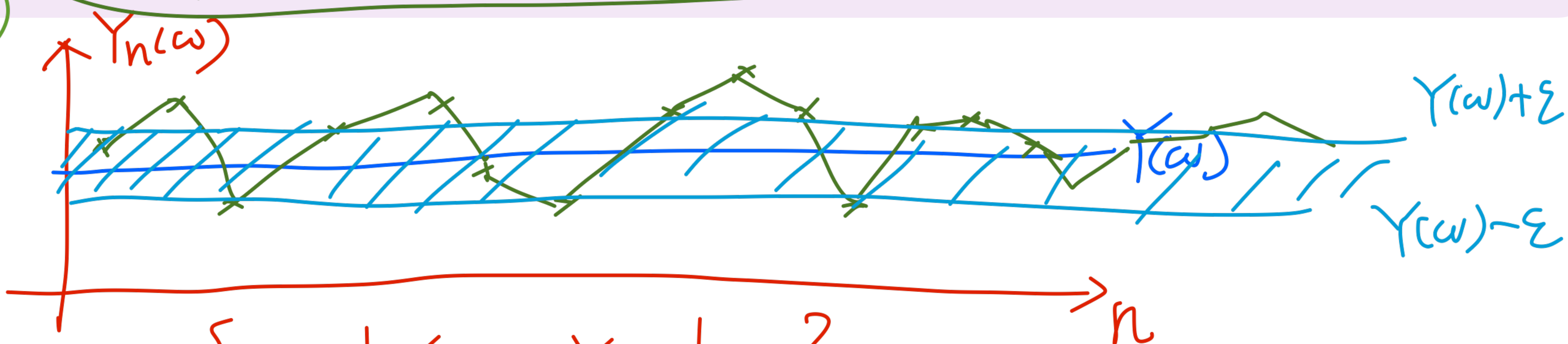
- Almost-Sure Convergence: $P(\{\omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\}) = 1$

$$P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n\right) = 0$$

- **Equivalent Definition of Almost-Sure Convergence**: Let $Y_1, Y_2 \cdots$ be a sequence of random variables. We say that Y_n converges to a random variable Y almost surely if $\forall \varepsilon > 0$,

$$P\Big(\big\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\big\} \text{ infinitely often}\Big) = 0$$

Fix an ω



Define $A_n = \{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}$

Almost-Sure Convergence ⇒ Convergence in Probability

- **Question**: Why "almost-sure convergence" is stronger?

- **Almost-Sure Convergence ⇒ Convergence in Probability**: Let $Y_1, Y_2 \cdots$ be a sequence of random variables. If Y_n converges to Y almost surely, then Y_n converges to Y in probability.

- **Proof:** Please see the supplementary material on E3

- **Question:** How about the converse?

Convergence in Probability, But Not Almost Surely

indicator function $X_1 = \begin{cases} 1, & \text{event occurs} \\ 0, & \text{otherwise} \end{cases}$

- Example: Let X be a continuous uniform r.v. on $(0,1)$
  - Consider a sequence of r.v.s $X_1, X_2, \cdots$ as follows:

$$X_1 = \mathbb{I}\{X \in [0,1]\}$$

$$X_2 = \mathbb{I}\{X \in [0,\frac{1}{2}]\} \quad X_3 = \mathbb{I}\{X \in [\frac{1}{2},1]\}$$

0 1

$$X_4 = \mathbb{I}\{X \in [0,\frac{1}{3}]\} \quad X_5 = \mathbb{I}\{X \in [\frac{1}{3},\frac{2}{3}]\} \quad X_6 = \mathbb{I}\{X \in [\frac{2}{3},1]\}$$

X_7 $\cdots$ X_8 X_9 $\cdots$ X_{10} $\cdots$

- Question: Do we have $\lim_{n\to\infty} P(\{\omega : |X_n(\omega) - 0| > \varepsilon\}) = 0$?

0

Convergence in Probability, But Not Almost Surely (Cont.)

- Example: Let X be a continuous uniform r.v. on $(0,1)$
    - Consider a sequence of r.v.s $X_1, X_2, \cdots$ as follows:

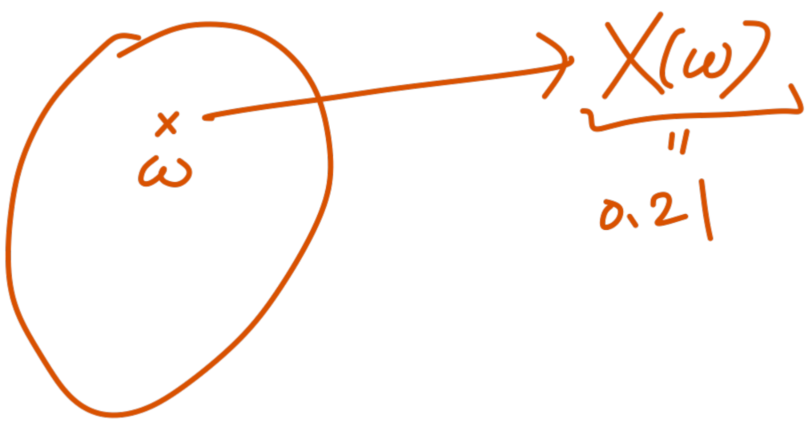
$X_1 = \mathbb{I}\{X \in [0,1]\}$ $\to 1$

$X_2 = \mathbb{I}\{X \in [0,\frac{1}{2}]\}$ $\to 1$ $\quad X_3 = \mathbb{I}\{X \in [\frac{1}{2},1]\}$ $\to 0$

$X_4 = \mathbb{I}\{X \in [0,\frac{1}{3}]\}$ $\to 1$ $\quad X_5 = \mathbb{I}\{X \in [\frac{1}{3},\frac{2}{3}]\}$ $\to 0$ $\quad X_6 = \mathbb{I}\{X \in [\frac{2}{3},1]\}$ $\to 0$

$X_n = \mathbb{I}\{X \in [0,\frac{1}{4}]\} \to 1$ $\quad \cdots \quad \to 0 \quad \to 0 \quad \cdots \quad \to 0$

- Question: Do we have $P(\{\omega : \lim_{n\to\infty} X_n(\omega) = 0\}) = 1$?



2. Strong Law of Large Numbers (SLLN)

WLLN vs SLLN

- **The Weak Law of Large Numbers (Khinchin's Law)**: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ. Define $S_n = (X_1 + \cdots + X_n)$. Then, for every $\varepsilon > 0$, we have

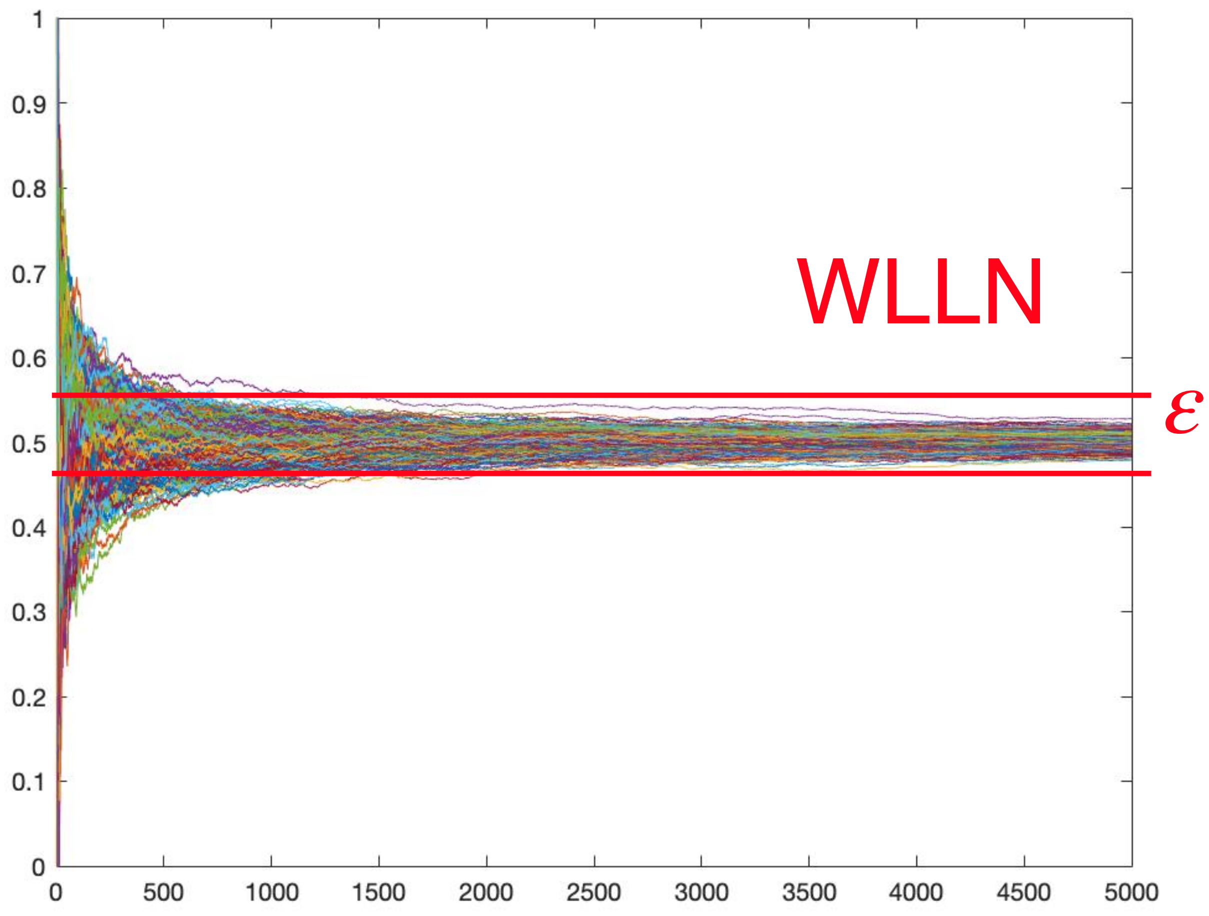
$$\lim_{n\to\infty} P\left(\{\omega : |\frac{S_n(\omega)}{n} - \mu| \geq \varepsilon \right) = 0$$

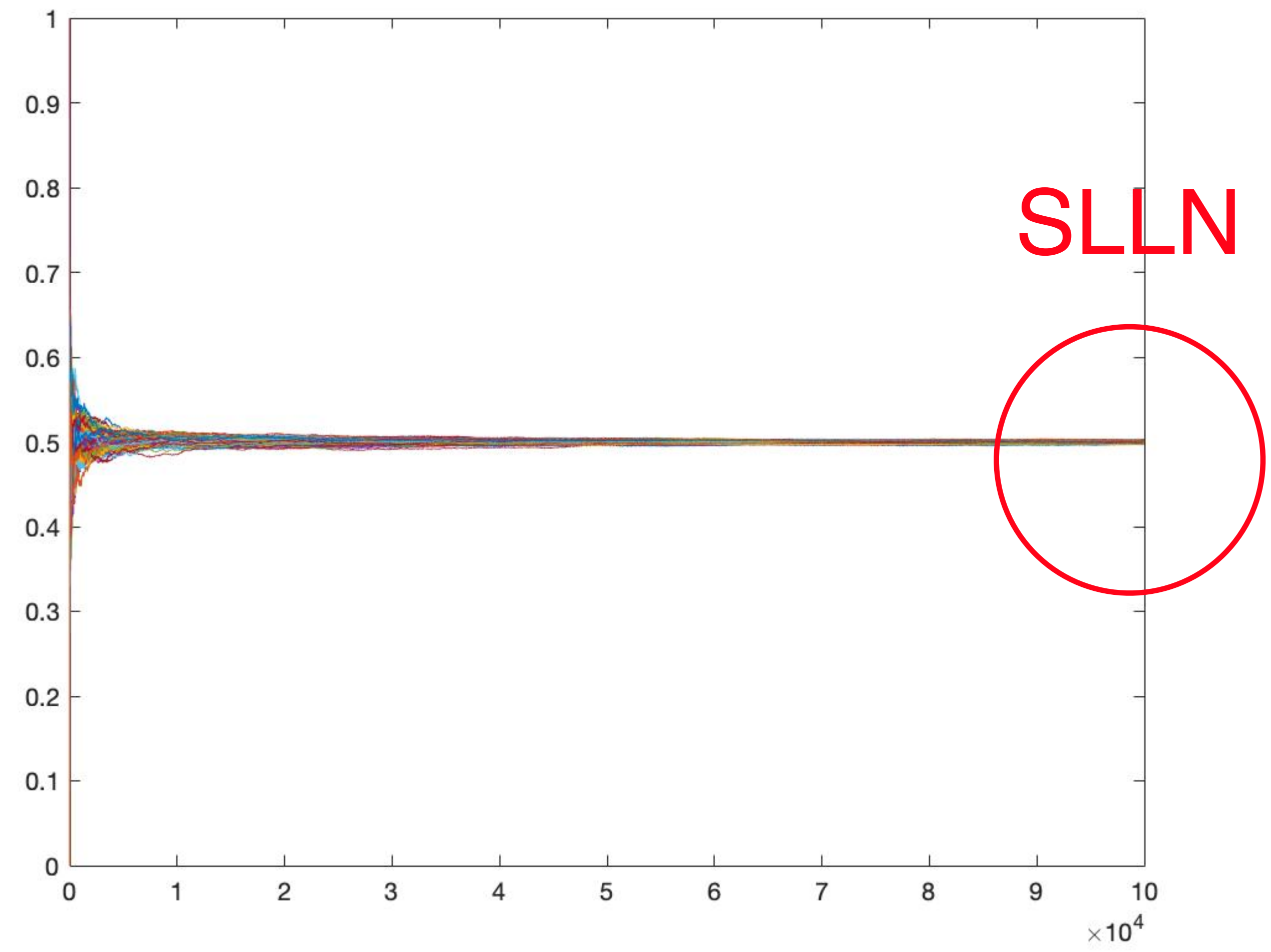
- **The Strong Law of Large Numbers**: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ. Define $S_n = (X_1 + \cdots + X_n)$. Then, we have

$$P\left(\left\{\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu \right\} \right) = 1$$

Visualization of WLLN and SLLN

- Example: $X_i \sim$ Bernoulli(0.5) and $S_n = X_1 + \cdots + X_n$





$$\lim_{n\to\infty} P\Big(\{\omega : |\frac{S_n(\omega)}{n} - \mu| \geq \varepsilon\Big) = 0$$

$$P\Big(\{\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu\}\Big) = 1$$

How to Prove SLLN?

1. Borel-Cantelli Lemma

2. A Bound for the 4-th Moment Condition

3. Markov's Inequality

1. Borel-Cantelli Lemma

- Recall: HW1, Problem 4

Problem 4 (Continuity of Probability Function and Probability Axioms) (8+8=16 points)

Consider an infinite sequence of coin tosses. The probability of having a head at the i-th toss is p_i, with $p_i \in [0, 1]$ (Note: different tosses may not be independent and can potentially have different head probabilities). We use I to denote the event of having infinite number of heads.

(a) Suppose $\sum_{i=1}^{\infty} p_i$ is finite. Show that $P(I) = 0$. (Hint: Define $A_n := \{\text{the } n\text{-th toss is a head}\}$. Then, $B_k := \bigcup_{n=k}^{\infty} A_n$ is the event that there is at least one head after the k-th toss (including the k-th toss). The event I (i.e. we observe infinitely many heads) is equivalent to saying that B_k occurs, for every $k = 1, 2, 3, \cdots$. Therefore, $I = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$. Consider the continuity of probability function for $\bigcap_{k=1}^{\infty} \bigcup_{i=k}^{\infty} A_n$ and then apply the union bound)

- **Borel-Cantelli Lemma**: Let $\{A_n\}$ be any sequence of events. If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then we have

$$P\Big(A_n \text{ occurs infinitely often}\Big) = P\Big(\bigcap_{n\geq 1} \bigcup_{m\geq n} A_n\Big) = 0$$

Review: Proof of Borel-Cantelli Lemma

- **Borel-Cantelli Lemma**: Let $\{A_n\}$ be any sequence of events. If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then we have

$$P\Big(A_n \text{ occurs infinitely often}\Big) = P(\bigcap_{n\geq 1} \bigcup_{m\geq n} A_n) = 0$$

- Proof:

2. A Bound on 4-th Moment

- **A Bound on 4-th Moment**: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ and $E[X_1^4] < \infty$. Define $S_n = (X_1 + \cdots + X_n)$. Then, there exists a constant $K < \infty$ such that

$$E[(S_n - n\mu)^4] \leq Kn^2$$

- Question: How about $E[(\frac{S_n}{n} - \mu)^4] \leq$?

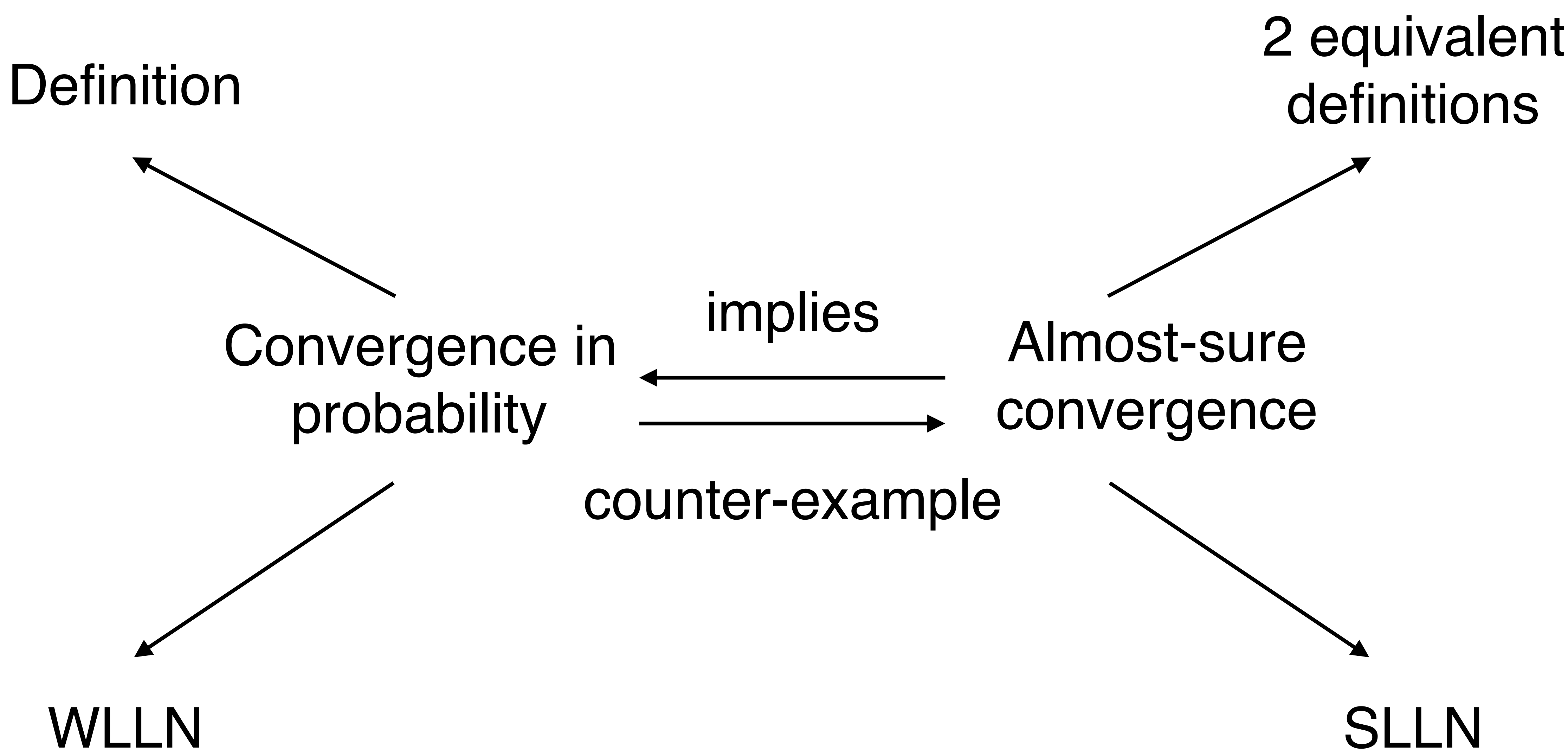
Proof: A Bound on 4-th Moment

- Given: $S_n = (X_1 + \cdots + X_n)$ and $E[X_1^4] < \infty$
- Want: $E[(S_n - n\mu)^4] \leq Kn^2$
- Proof: For simplicity, let $Z_i = X_i - \mu$

Put Everything Together: Proof of SLLN

$$P\left(\{ |\frac{S_n}{n} - \mu| \geq n^{-\gamma} \} \right) = P\left(\{ |\frac{S_n}{n} - \mu|^4 \geq n^{-4\gamma} \} \right) \leq$$

Summary


Definition
2 equivalent definitions
Convergence in probability
implies
Almost-sure convergence
counter-example
WLLN
SLLN

Next Lecture

- Central Limit Theorem (CLT)

1-Minute Summary

1. Weak Law of Large Numbers (WLLN) and Convergence in Probability

- Convergence in probability: $\lim_{n\to\infty} P(\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}) = 0$
- WLLN: $\lim_{n\to\infty} P\left(\left\{\omega : \left|\frac{S_n(\omega)}{n} - \mu\right| \geq \varepsilon\right\}\right) = 0$

2. Strong Law of Large Numbers (SLLN) and Almost-Sure Convergence

- Almost-sure convergence: $P(\{\omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\}) = 1$
- SLLN: $P\left(\left\{\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu\right\}\right) = 1$