1179: Probability
Lecture 27 — Law of Large Numbers

Ping-Chun Hsieh (謝秉均)

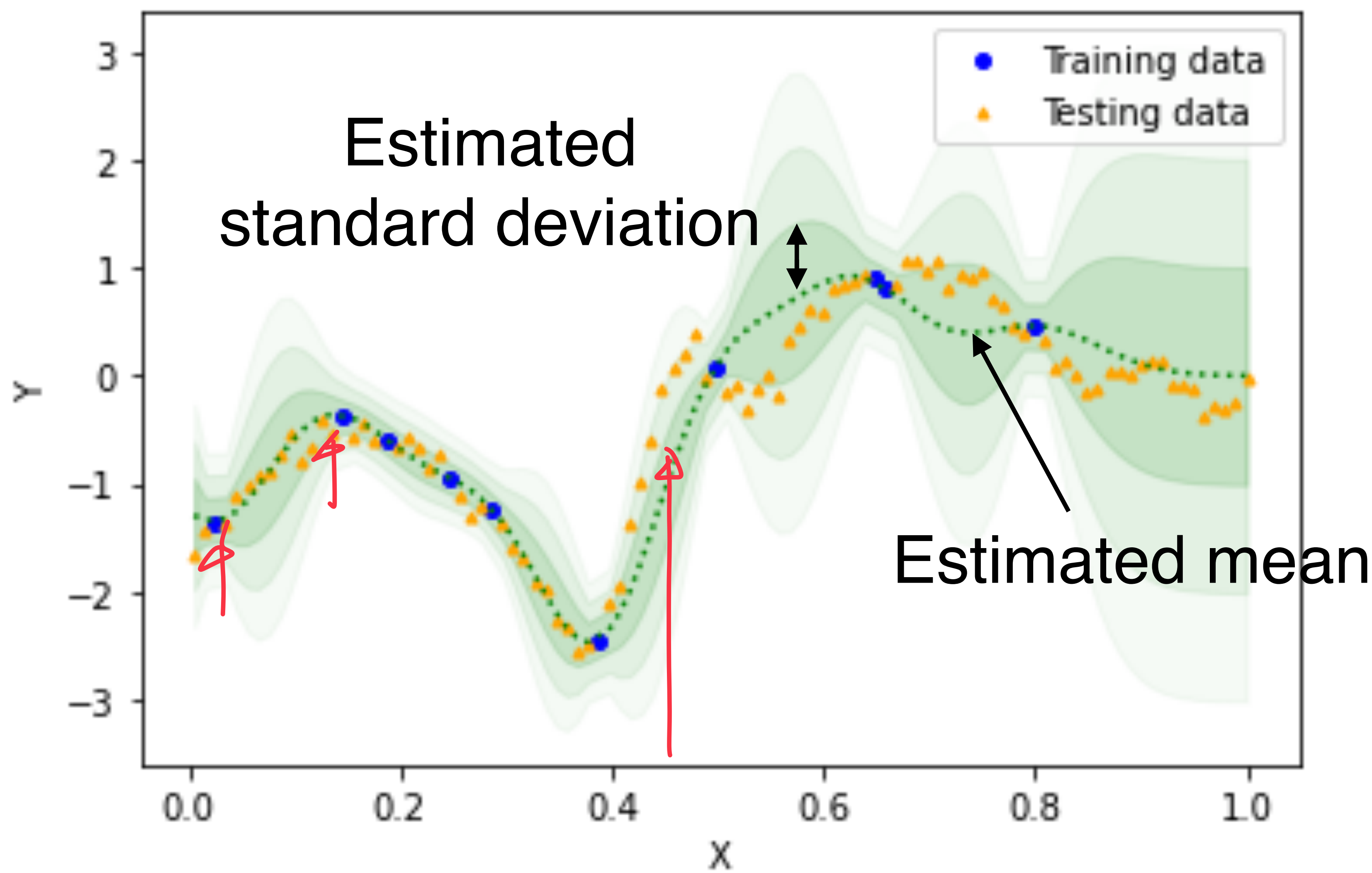
December 22, 2021

Announcements

- HW4 is now available on E3!
  - HW4-Part I will be due on 12/30 (Thursday), 9pm
  - HW4-Part II will be due on 1/3 (Monday), 9pm

- Final exam on 1/5 (on Wednesday, in class)
  - 10:10am - 12:10pm
  - Coverage: Lec 1 - Lec 29
  - You are allowed to bring a cheat sheet (A4 size, 2-sided, without any attachments)
  - Locations: EC015 and EC022

HW4: Multivariate Normal for Regression



- **Task:** Given the training data and the X values of the testing data, estimate the Y values of the testing data points
- **Model:** The Y values form a <u>multivariate normal</u> random variable
- **Result:** Conditional distribution of the Y value of any testing sample is normal with some mean and variance

Recall: Multivariate Normal R.V.

- **Multivariate Normal Random Variables:** $Y_1, \cdots, Y_n$ are said to be <u>multivariate normal</u> random variables if the joint PDF of $Y_1, \cdots, Y_n$ is

$$f_{Y_1 Y_2 \cdots Y_n}(y_1, y_2, \cdots, y_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\det(\Sigma)|}} \exp\left[-\frac{1}{2}(y - \mu)^T \Sigma^{-1} (y - \mu) \right]$$

vector

covariance matrix

where

$$\Sigma = \begin{bmatrix} \text{Cov}(Y_1, Y_1) \cdots & \text{Cov}(Y_1, Y_n) \\ \text{Cov}(Y_2, Y_1) \cdots & \text{Cov}(Y_2, Y_n) \\ \cdots & \\ \text{Cov}(Y_n, Y_1) \cdots & \text{Cov}(Y_n, Y_n) \end{bmatrix}$$

(i, j)

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_n \end{bmatrix}$$

- Question: How to configure the covariance $\text{Cov}(Y_i, Y_j)$?

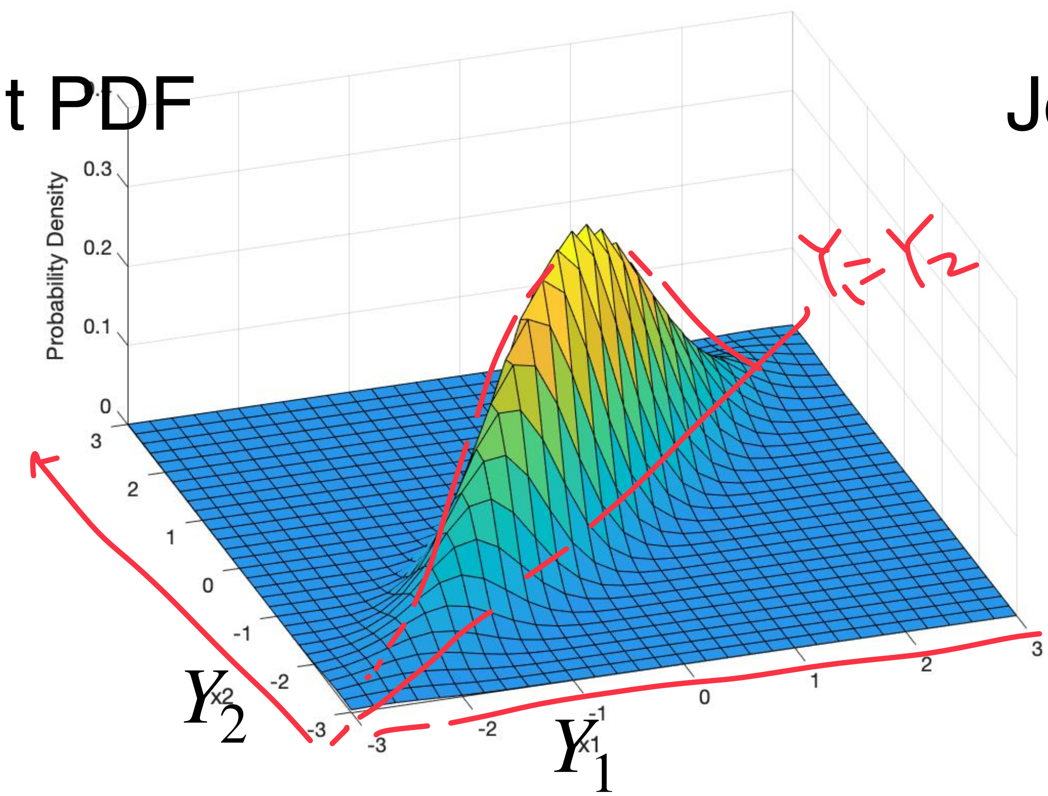
Covariance Captures "Smoothness"

$\sigma_1^2 = 1$ $\rho\sigma_1\sigma_2$

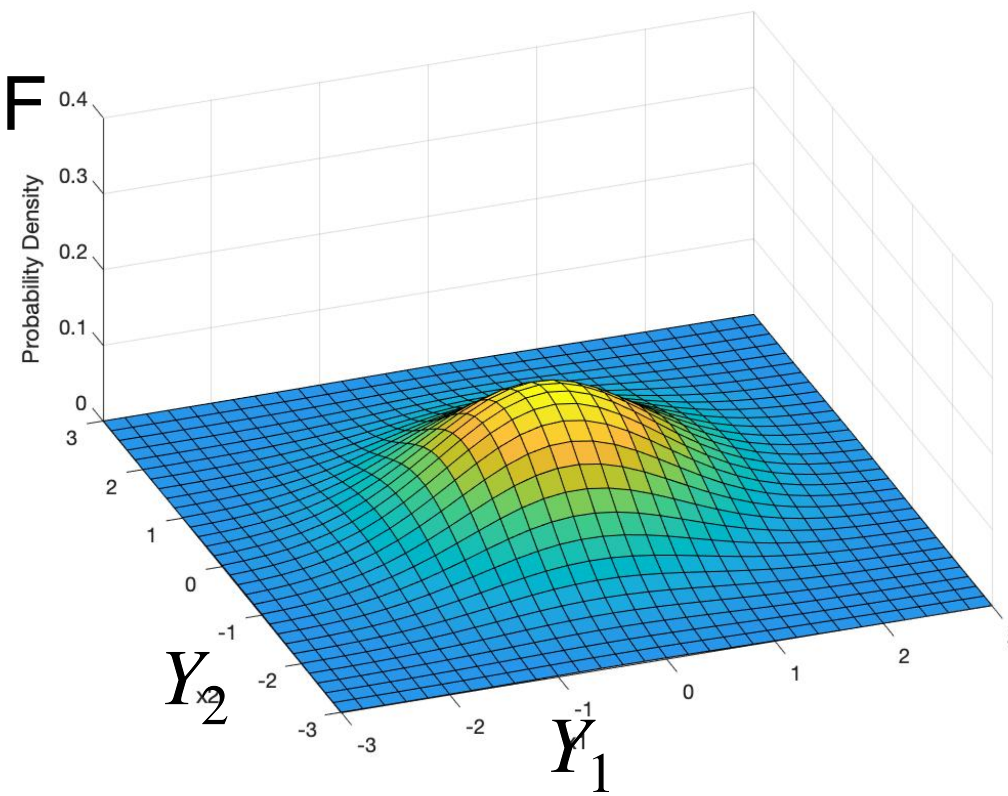
$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & k \\ k & 1 \end{bmatrix} \right)$$

bivariate normal distribution $\sigma_2^2 = 1$

Joint PDF



Joint PDF



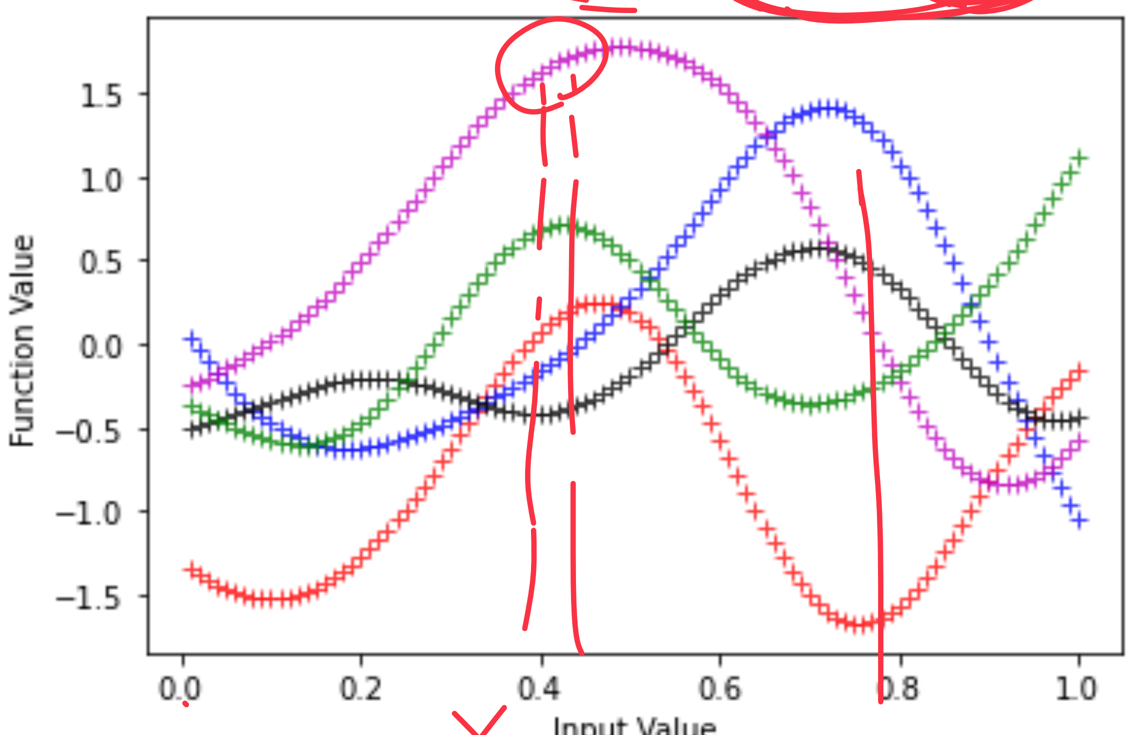
- $k = 0.9$: Y_1, Y_2 are close with high probability
- $k = 0.1$: Y_1, Y_2 are far away with high probability

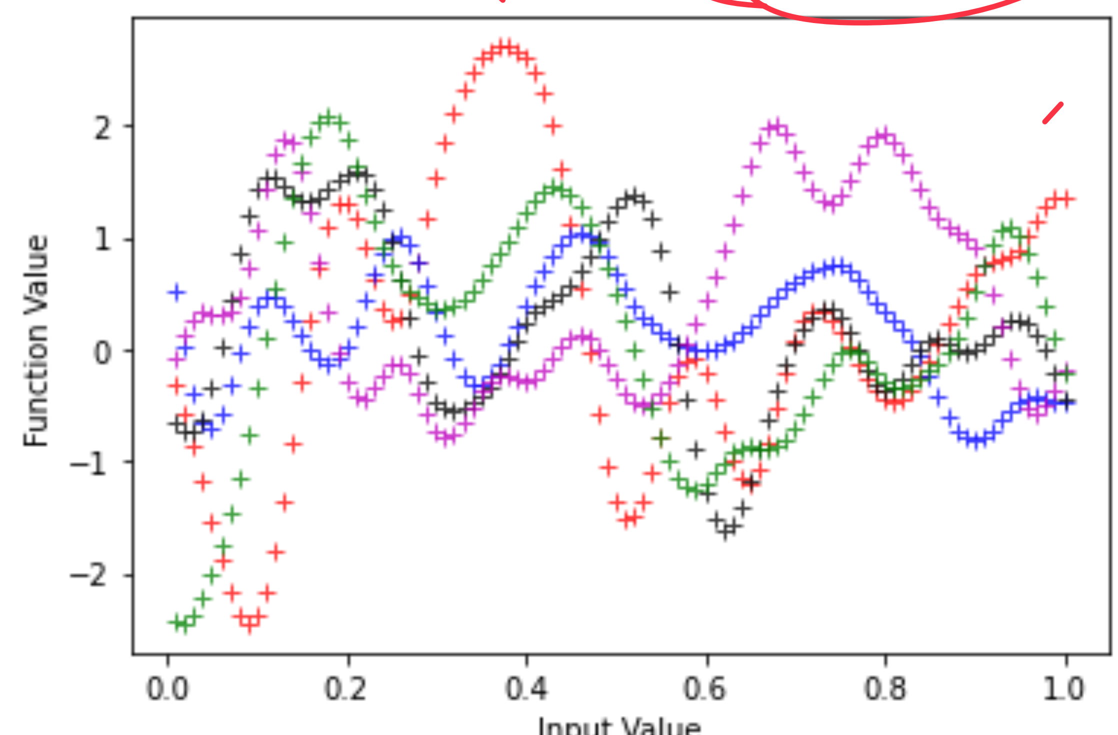
Use “Kernel Functions” to Configure “Smoothness”

kernel function

$$\begin{bmatrix} Y_i \\ Y_j \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & k(x_i, x_j) \\ k(x_i, x_j) & 1 \end{bmatrix} \right)$$

- Example: Consider $x_k = k/100$, for $k \in \{1, \cdots, 100\}$
  - $k(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^2}{2 \times 0.2^2}\right)$
  - $k(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^2}{2 \times 0.05^2}\right)$





This Lecture

1. Weak Law of Large Numbers (WLLN)

2. Strong Law of Large Numbers (WLLN)

- Reading material: Chapter 11.3-11.4

Weak Law of Large Numbers (WLLN)

Review: 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

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

empirical mean

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

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

- Question: What does "convergence" mean here?

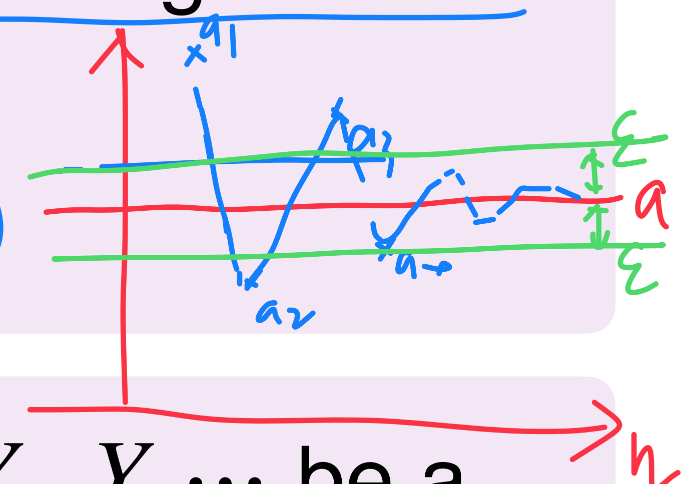
Convergence in Probability

$a_1, a_2, a_3, \ldots$

$\lim_{n\to\infty} a_n = a$

- **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

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



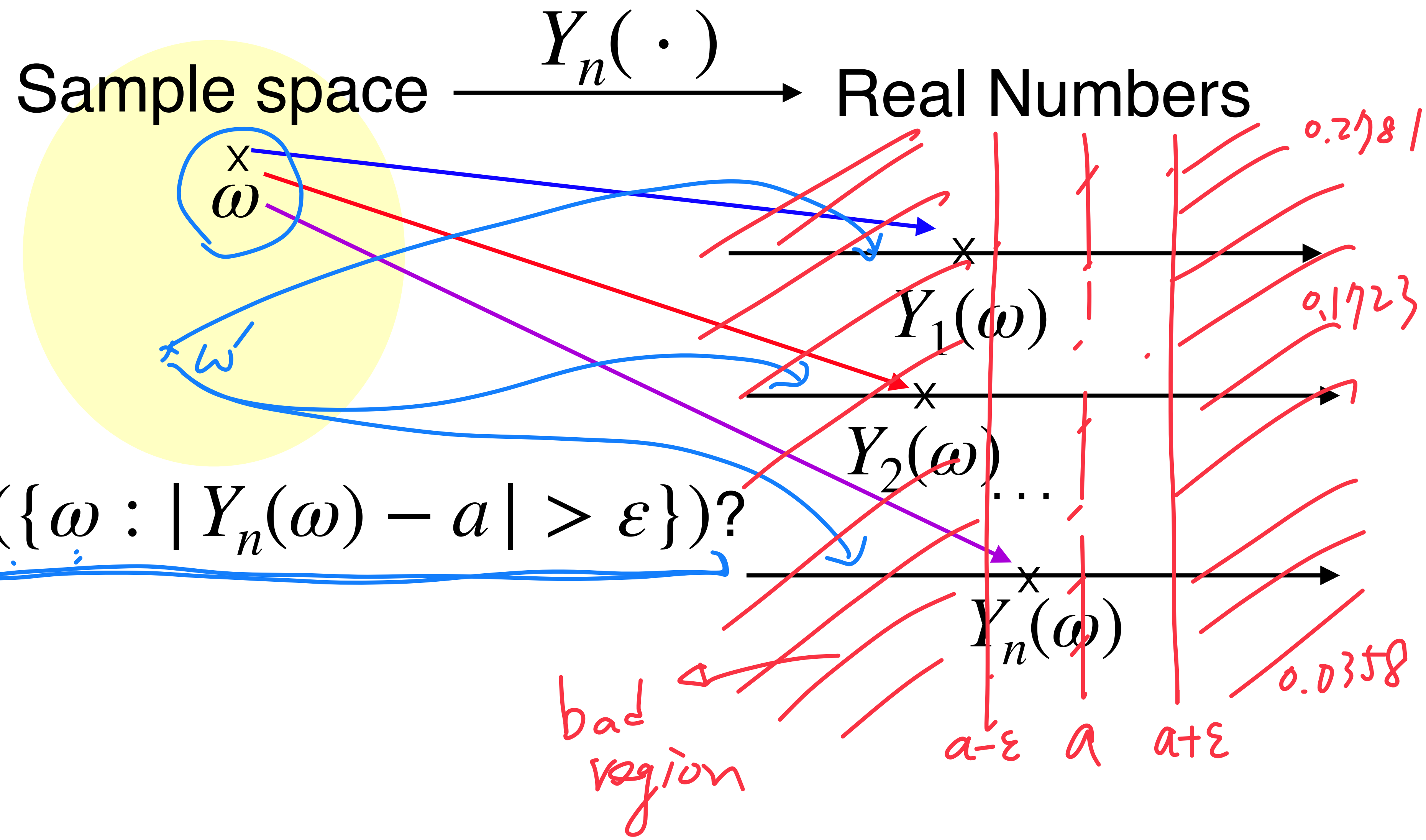
- **Convergence to a Scalar in Probability**: Let $Y_1, Y_2 \cdots$ be a sequence of random variables, and let a be a real number. We say that Y_n converges to a in probability if for every $\varepsilon > 0$,

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

- Question: How to interpret this definition?

Recall: Random Variables Defined on Ω

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

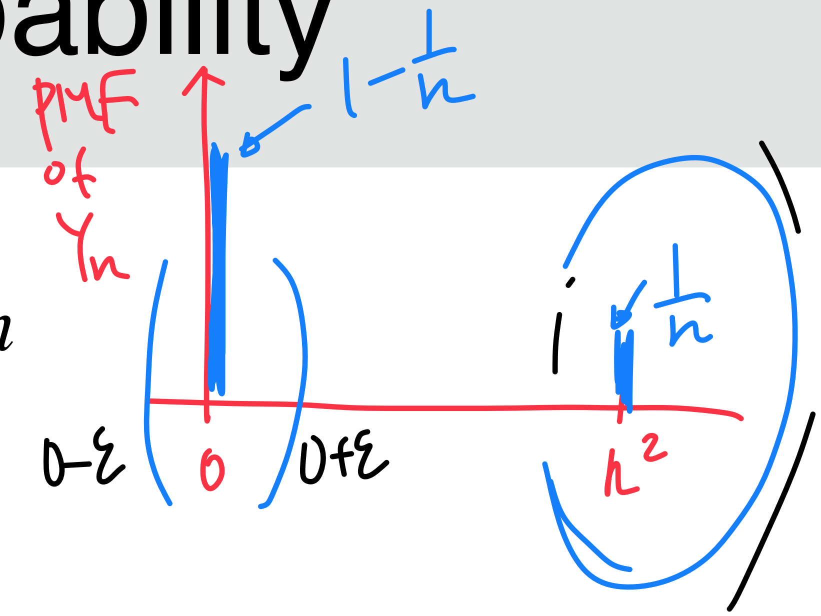


- How to interpret $P(\{\omega : |Y_n(\omega) - a| > \varepsilon\})$?

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

Y_{n+10^8} $\quad 10^{-5}$

Example: Convergence in Probability



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

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

$\varepsilon = 0.1$

- For every $\varepsilon > 0$, can we find $P(|Y_n - 0| > \varepsilon)$?

bad event

$P(|Y_1 - 0| > 0.1) = 1$

$P(|Y_2 - 0| > 0.1) = \frac{1}{2}$

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

$0 \le P(|Y_n - 0| > \varepsilon) \le \frac{1}{n}$, for every $\varepsilon > 0$

$\underbrace{\lim_{n\to\infty} 0}_{=0} \le \lim_{n\to\infty} P(|Y_n - 0| > \varepsilon) \le \lim_{n\to\infty} \frac{1}{n} = 0$, for every $\varepsilon > 0$

$\Rightarrow$ Y_n converges to 0 in probability $\left(Y_n \xrightarrow{P} 0\right)$

How to Interpret WLLN?

$Y_n(\omega) = \dfrac{X_1(\omega) + \cdots + X_n(\omega)}{n}$ for every ω

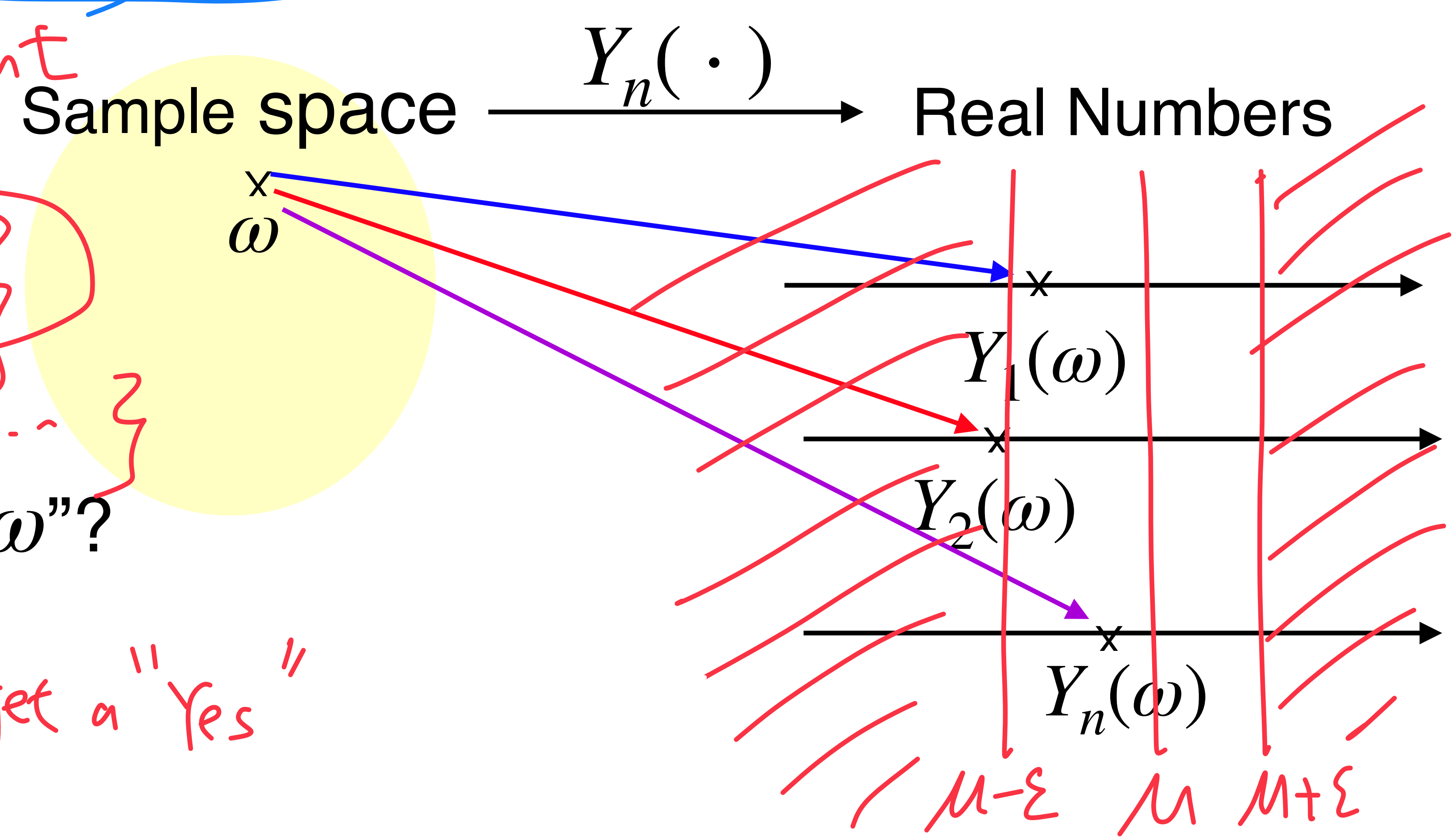
- Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with mean μ
- Define $Y_n = (X_1 + X_2 \cdots, + X_n)/n$ empirical mean
- WLLN: $\lim_{n\to\infty} P(\{\omega : |Y_n(\omega) - \mu| > \varepsilon\}) = 0, \forall \varepsilon > 0$

bad event

$Y_1(\omega) = 1$, $Y_2(\omega) = 1$, $Y_3(\omega) = 2/3$

$\omega = \{$ Yes, Yes, No, Laugh, Yes, $\}$

$\omega' = \{$ No, No, M, Yes, Yes, No, $\cdots \}$



- Question: What is an "ω"?

Moon blocks

$X_n = \begin{cases} 1, & \text{if you get a "Yes"} \\ 0, & \text{otherwise} \end{cases}$

$$\omega = \{ \text{Yes}, \text{Yes}, \text{Laugh}, \text{No}, \text{Yes} \ldots \}$$

$$Y_n = \frac{X_1 + \cdots + X_n}{n}$$

$$Y = \frac{X_1 + X_3 + X_5}{3}$$

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

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

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

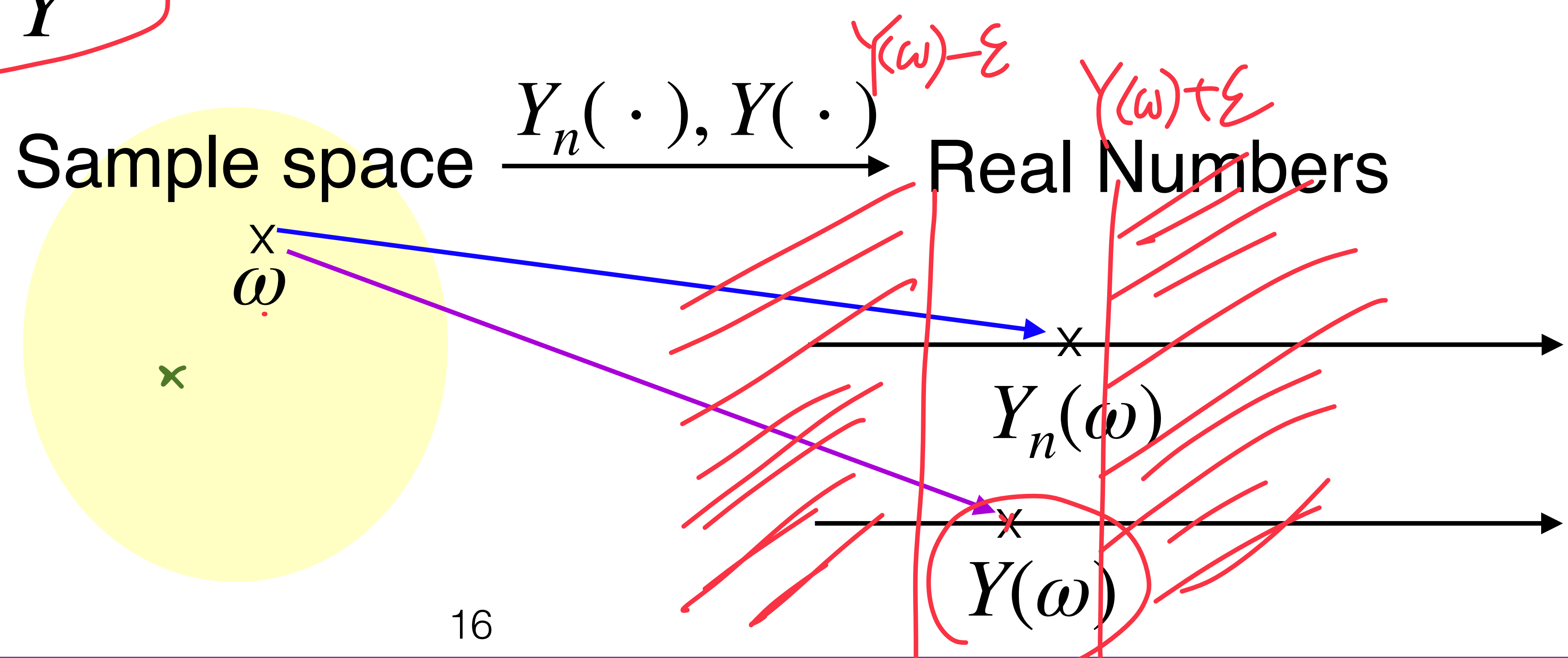
Convergence in Probability (Cont.)

- **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 $\varepsilon > 0$,

a

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

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



Example: Convergence in Probability

- Example: Consider a random variable Y

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

$Y_n(\omega)=(1+\frac{1}{n})\cdot Y(\omega)$ for every $\omega\in\Omega$

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

$$P(\{\omega:\underbrace{|Y_n(\omega)-Y(\omega)|}>\varepsilon\})=P(\{\omega:\underbrace{|((1+\frac{1}{n})Y(\omega)-Y(\omega)|}_{=\,|\frac{1}{n}\cdot Y(\omega)|}>\varepsilon\})$$

$$\lim_{n\to\infty}P(\underbrace{\{\omega:|\frac{1}{n}Y(\omega)|>\varepsilon\}}_{\text{this event is empty for all } n>\frac{1}{\varepsilon}})=0\implies\boxed{Y_n\xrightarrow{P}Y}$$

Any Stronger Notion of Convergence?

Example: Convergence With Probability 1

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

$Y_n(\omega) = \min\{X_1(\omega), \cdots, X_n(\omega)\}$ for every ω

① $\varepsilon > 1$: $P(\{\omega : Y_n(\omega) \geq \varepsilon\}) = 0$

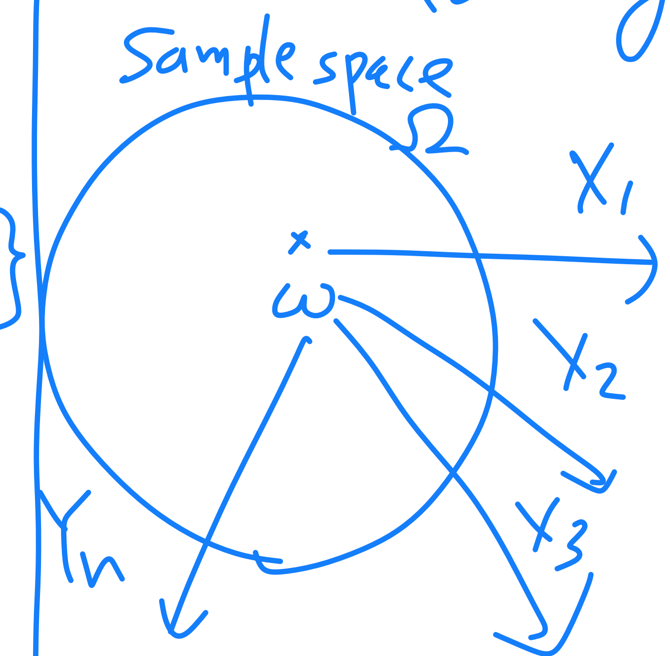
② $0 < \varepsilon \leq 1$:

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

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

$$= \underbrace{P(\{\omega : X_1(\omega) \geq \varepsilon\})}_{1-\varepsilon} \cdot \underbrace{P(\{\omega : X_2(\omega) \geq \varepsilon\})}_{1-\varepsilon} \cdots \underbrace{P(\{\omega : X_n(\omega) \geq \varepsilon\})}_{1-\varepsilon}$$

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



Example: Convergence With Probability 1 (Cont.)

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

Previous page: $P(\{\omega : Y_n(\omega) \geq \varepsilon\}) = (1-\varepsilon)^n \quad (0 < \varepsilon \leq 1)$

$\{\omega : \lim_{n\to\infty} Y_n(\omega) \geq \varepsilon\} \subseteq \{\omega : Y_n(\omega) \geq \varepsilon\}$, for every n

$\Rightarrow P(\{\omega : \lim_{n\to\infty} Y_n(\omega) \geq \varepsilon\}) = 0, \quad (0 < \varepsilon \quad)$

$\Leftrightarrow P(\{\omega : \lim_{n\to\infty} Y_n(\omega) = 0\}) = 1$

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 $P(\{\omega : \lim_{n\to\infty} Y_n(\omega) = 0\})$ = ?

By the previous page, we know

$$P\left(\left\{\omega: \lim_{n\to\infty} Y_n(\omega) \geq \varepsilon\right\}\right) = 0, \text{ for all } \varepsilon > 0.$$

This is equivalent to

$$P\left(\left\{\omega: \lim_{n\to\infty} Y_n(\omega) = 0\right\}\right) = 1$$

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 $\forall \varepsilon > 0$,

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

先取機率，再取limit

"(with probability)"

- **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 (with probability) if,

$P(A) = 1$

$A = \Omega$?

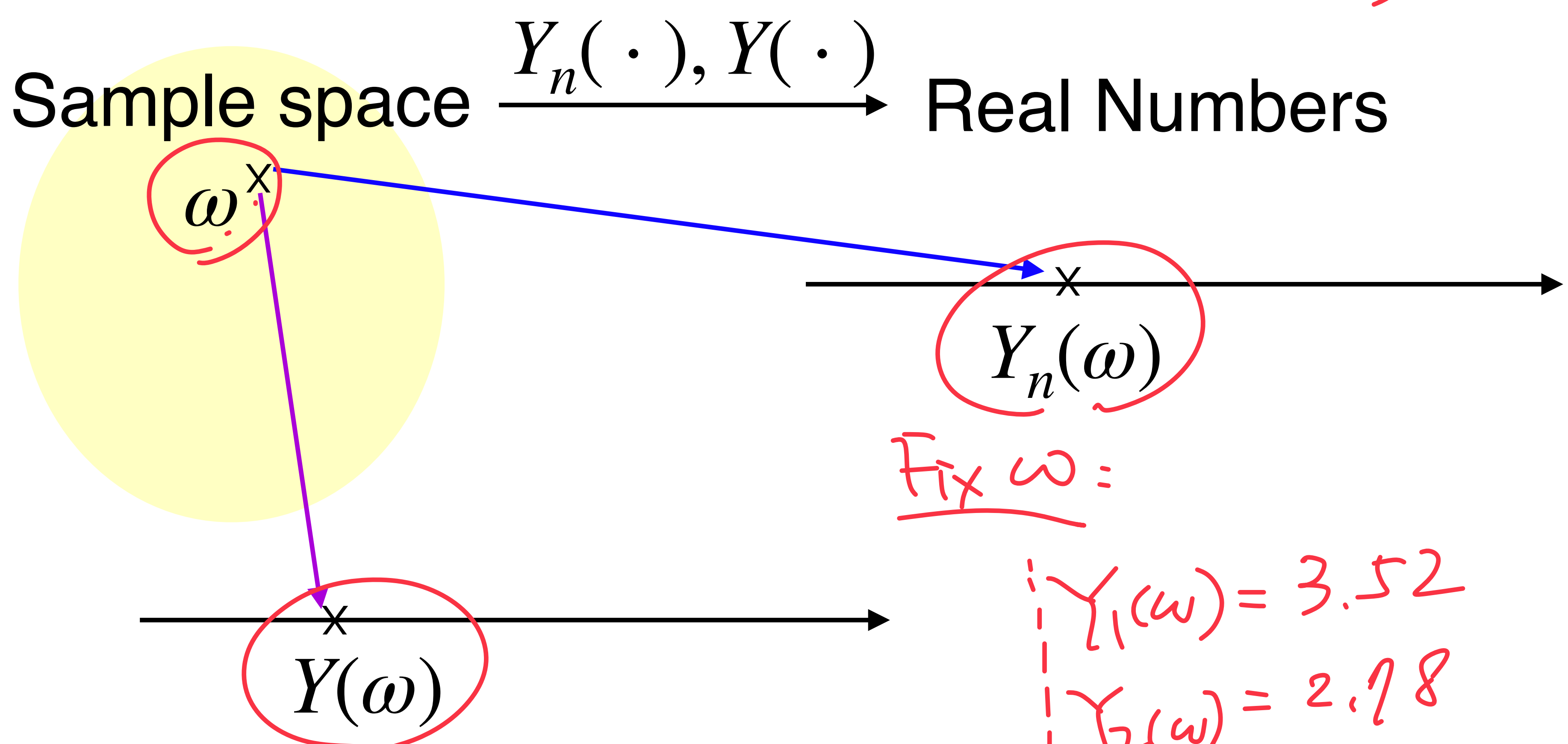
$$P\left(\boxed{\left\{\omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\right\}}\right) = 1$$

先取limit 再取機率

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

Interpretation of Almost-Sure Convergence

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



Fix ω:

$Y_1(\omega) = 3.52$

$Y_2(\omega) = 2.78$

$\vdots$

$Y_{100}(\omega) = 1.25$

$\vdots$

$Y_{10^{10}}(\omega) = 0.035$

$Y_{10^{11}}(\omega) = 10^{-7}$

$Y(\omega) = 0$

- How to interpret $P(\{\omega : \underbrace{\lim_{n\to\infty} Y_n(\omega) = Y(\omega)}_{\text{condition for a "valid" } \omega}\})$?

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? "Counterexample"

Convergence in Probability, But Not Almost Surely

- 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:

indicator

$\mathbb{I}\{\text{event}\} = \begin{cases} 1, & \text{if event happens} \\ 0, & \text{else} \end{cases}$

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

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

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

$X_7 = \mathbb{I}\{X \in [0,\frac{1}{4}]\}$ ... $X_8 = \mathbb{I}\{X \in [\frac{1}{4},\frac{2}{4}]\}$ ... $X_9 = \cdots$ $X_{10} = \cdots$ ...

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

$X_n \xrightarrow{P} 0$?

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:

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

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

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

$$\cdots \qquad \cdots \qquad \cdots$$

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

$X_n \not\xrightarrow{a.s.} 0$

Equivalent Definition of Almost-Sure Convergence

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

- **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 $\forall \varepsilon > 0$,

$$P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\underbrace{\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}}_{A_n}\right) = 0$$

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

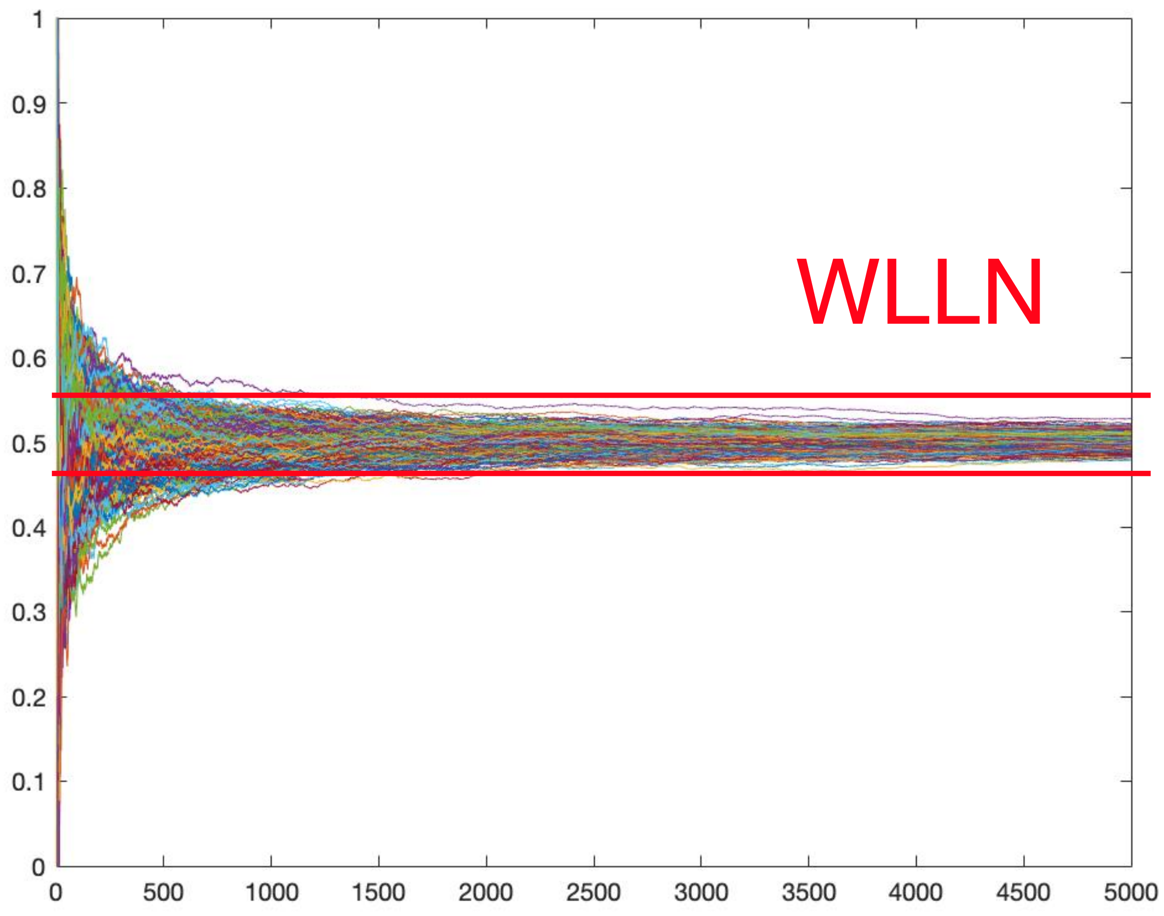
$$\lim_{n\to\infty} P\left(\{\omega : \left| \frac{S_n(\omega)}{n} - \mu \right| \geq \varepsilon \right) = 0$$

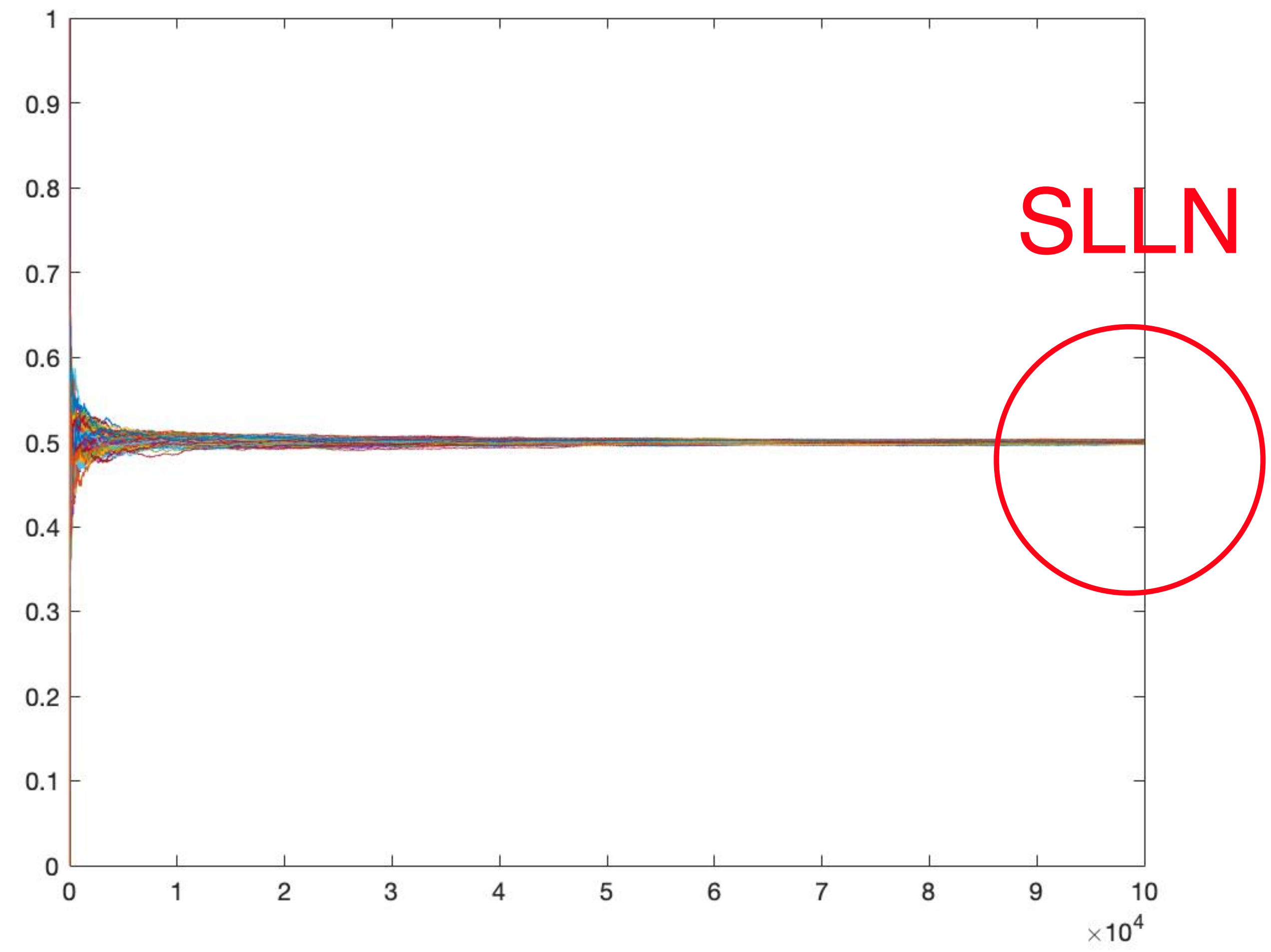
- **The Strong Law of Large Numbers**: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ. Define $S_n = (X_1 + \cdots + X_n)$. Then, we have

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

Visualization of WLLN and SLLN

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



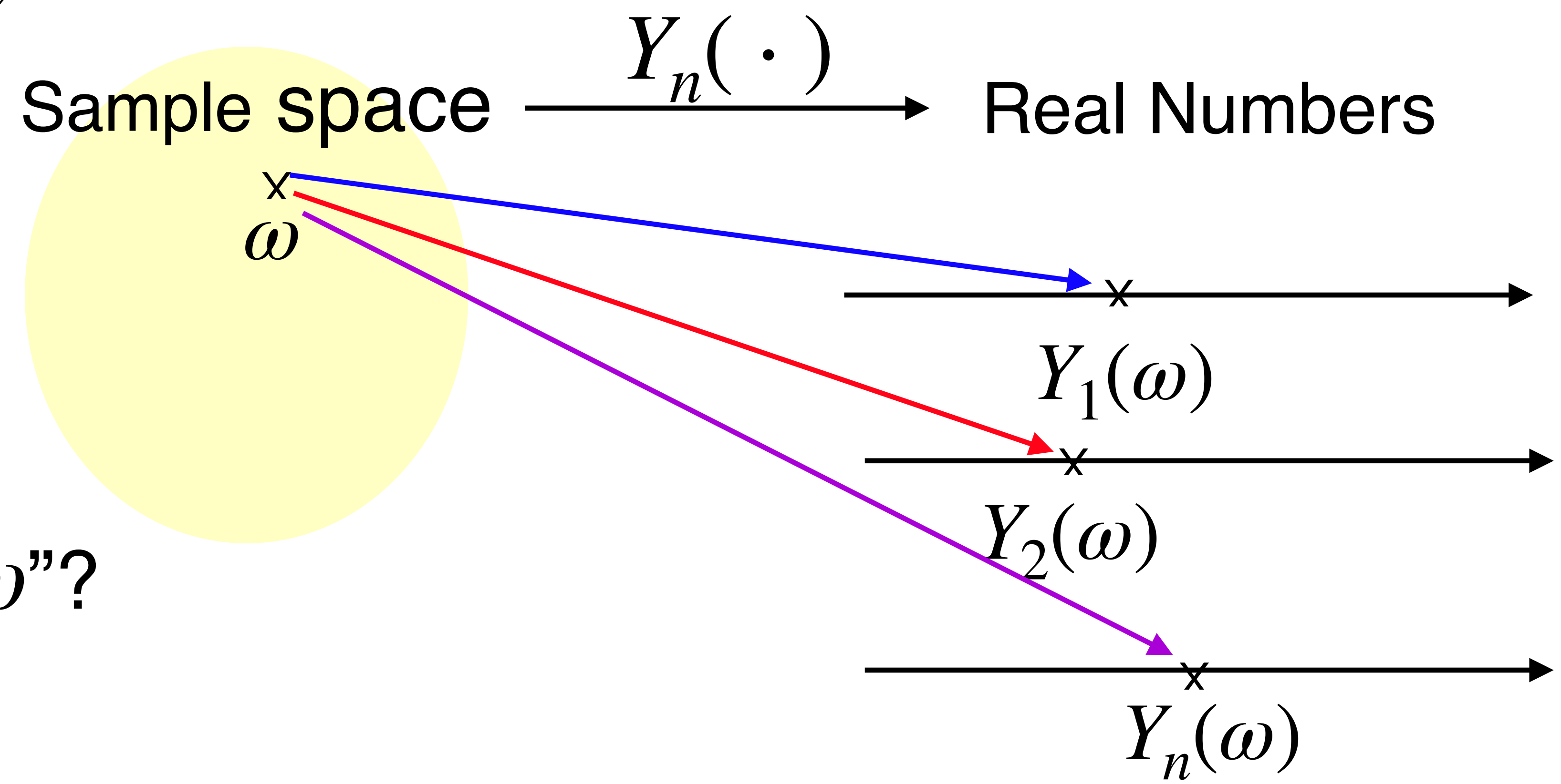


$$\lim_{n\to\infty} P\Big(\{\omega : |\frac{S_n(\omega)}{n} - \mu| \geq \varepsilon\Big) = 0$$

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

How to Interpret SLLN?

- Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with mean μ
- Define $Y_n = (X_1 + X_2 \cdots, + X_n)/n$
- SLLN: $P\left(\left\{\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu\right\}\right) = 1$



- Question: What is an "ω"?

How to Prove SLLN (Under a Mild Condition)?

1. Borel-Cantelli Lemma

2. A Bound for the 4-th Moment Condition

3. Markov’s Inequality

1. Borel-Cantelli Lemma

- **Recall:** HW1, Problem 3

Problem 3 (Continuity of Probability Functions) (12+12=24 points)

(a) Let $A_1, A_2, A_3, \cdots$ be a countably infinite sequence of events. Prove that if $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n) = 0$. This property is known as the *Borel-Cantelli Lemma.* (Hint: Consider the continuity of probability function for $\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$ and then apply the union bound)

(b) Consider a countably infinite sequence of coin tosses. The probability of having a head at the k-th toss is p_k, with $p_k = 100 \cdot k^{-N}$ (Note: different tosses are NOT necessarily independent). We use I to denote the event

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

$$P\Big(\{\omega : \omega \in A_n \text{ for infinitely many } n\}\Big) = P\Big(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\Big) = 0$$

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

$$P\Big(\{\omega : \omega \in A_n \text{ for infinitely many } n\}\Big) = P\Big(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n\Big) = 0$$

- Proof:

2. A Bound For 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

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

- Proof: Please see the supplemental on E3
- Question: How about $E[(\frac{S_n}{n} - \mu)^4] \leq$?

Put Everything Together: Proof of SLLN

- SLLN: $P\Big(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\{\omega : \Big|\frac{S_n(\omega)}{n} - \mu\Big| > \varepsilon\}\Big) = 0, \forall \varepsilon > 0$
- Proof:

$$P\Big(\underbrace{\{\Big|\frac{S_n}{n} - \mu\Big| \geq n^{-\gamma}\}}_{A_n}\Big) = P\Big(\Big|\frac{S_n}{n} - \mu\Big|^4 \geq n^{-4\gamma}\}\Big) \leq$$

A Quick Summary