1179: Probability
Lecture 29 — Monte Carlo Simulation and Central Limit Theorem

Ping-Chun Hsieh (謝秉均)

December 29, 2021

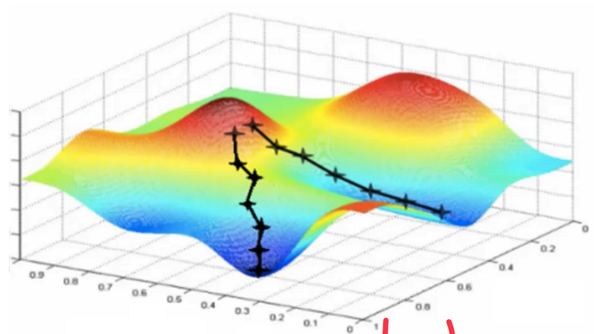
Announcements

- Final exam on 1/5 (next 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

“Monte Carlo Simulation” is a Building Block of Many Practical Algorithms…

- "Monte-Carlo Policy Gradient" in Reinforcement Learning
  - https://www.youtube.com/watch?v=KHZVXao4qXs (by David Silver)

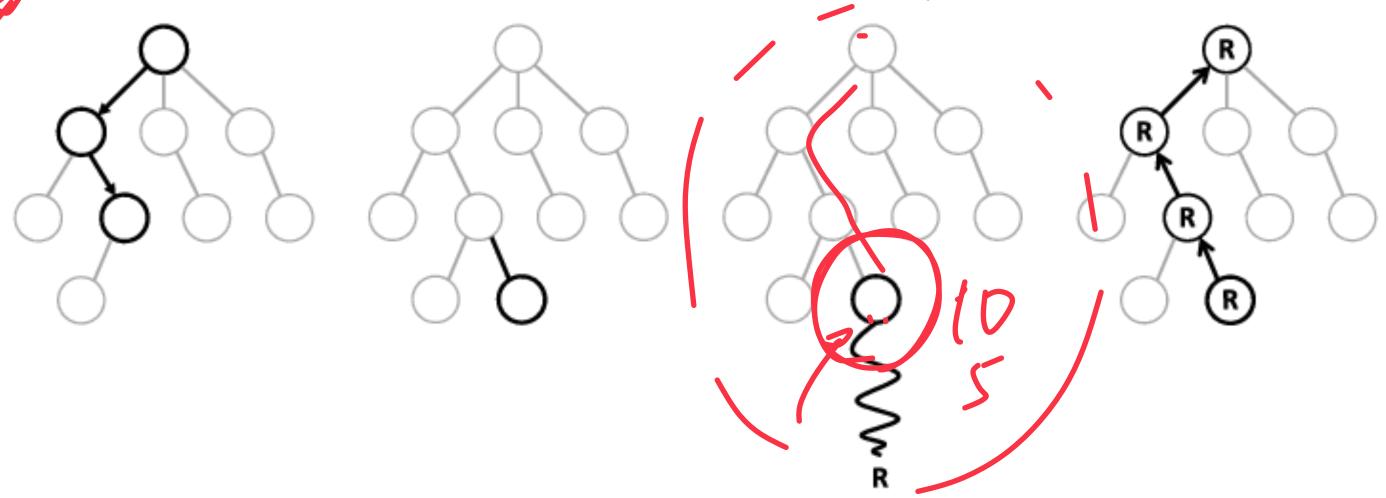
$$\nabla_\theta V^{\pi_\theta}(\mu) = \mathbb{E}_{\tau\sim P_\mu^{\pi_\theta}}\left[R(\tau)\sum_{t=0}^{\infty}\nabla_\theta \log \pi_\theta(a_t \mid s_t)\right]$$



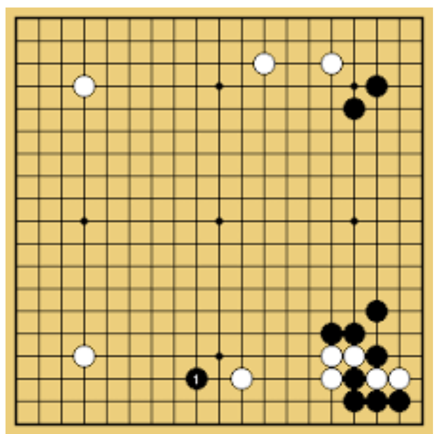
- Monte-Carlo Tree Search (MCTS) in Computer Go

Alphago

  - https://www.youtube.com/watch?v=UXW2yZndl7U (by John Levine)



(a) Selection (b) Expansion (c) Simulation (d) Backpropagation



- Markov-Chain Monte-Carlo (MCMC)
  - https://www.youtube.com/watch?v=TNZk8lo4e-Q (by Nando de Freitas)

Quick Review

$$P\Big(\underbrace{\{\omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\}}_{\text{Good event}}\Big) = 1$$

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

for every $\varepsilon > 0$

Definition

2 equivalent definitions

for every $\varepsilon > 0$

$$P(\cap \cup A_n) = 0$$

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

Convergence in probability

implies

counter-example

Almost-sure convergence

(stronger)

$\frac{S_n}{n}$: empirical mean

WLLN

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

for every ε

SLLN

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

This Lecture

1. SLLN and Monte Carlo Simulation

2. Central Limit Theorem (CLT)

- Reading material: Chapter 11.5

SLLN: Two Equivalent Statements

- **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 $$P\Big(\underbrace{\Big\{\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu\Big\}}_{\text{Good event}}\Big) = 1$$

- **The Strong Law of Large Numbers (Equivalent Statement)**: Let $X_1, \cdots, X_n \cdots$ 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 $$P\Big(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\underbrace{\Big\{\omega : \Big|\frac{S_n(\omega)}{n} - \mu\Big| > \varepsilon\Big\}}_{\text{bad event} \quad A_n}\Big) = 0, \forall \varepsilon > 0$$

$P(\bigcap \bigcup A_n) = 0$

infinitely many

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

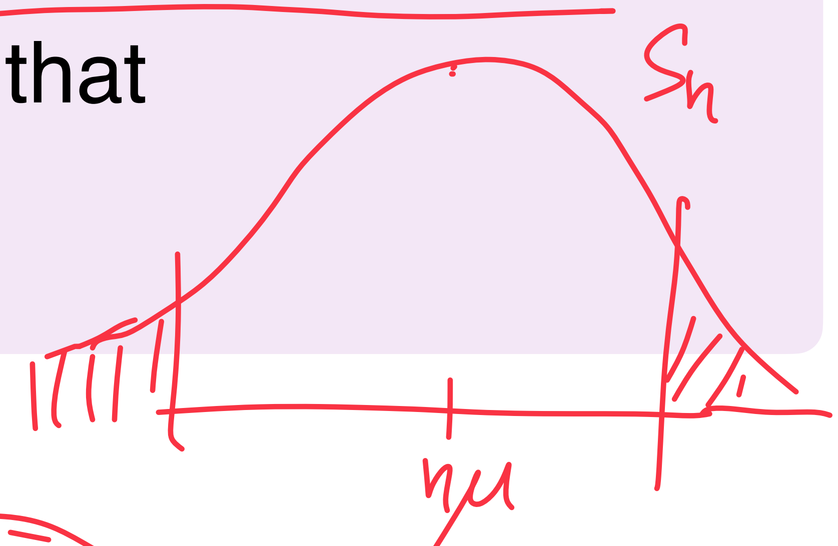
$$P\Big(\{\omega : \omega \in A_n \text{ for infinitely many } n\}\Big) = P\Big(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\Big) = 0$$

2. A Bound For the 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

4-th central moment of S_n

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



- **Proof:** Please see the supplemental on E3

- **Question:** How about $E[(\frac{S_n}{n} - \mu)^4] \leq$? $\frac{Kn^2}{n^4} = \frac{K}{n^2}$

Put Everything Together: Proof of SLLN

$B_{n,\varepsilon}$ $\quad \gamma = \frac{1}{6} \quad n^{-\gamma} < 0.1$

- SLLN: $P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} \left\{\omega : \left|\frac{S_n(\omega)}{n} - \mu\right| > \varepsilon\right\}\right) = 0, \forall \varepsilon > 0$ $\quad$ (ε: 0.1)

- Proof:

$\gamma > 0$ $\quad$ $\le \varepsilon$ $\quad$ Markov's inequality

Step 1: $\underbrace{P\left(\left\{\left|\frac{S_n}{n} - \mu\right| \ge n^{-\gamma}\right\}\right)}_{A_n \text{ (bad event)}} = P\left(\underbrace{\left\{\left|\frac{S_n}{n} - \mu\right|^4 \ge n^{-4\gamma}\right\}}_{\tilde{A}_n}\right) \le \dfrac{E\left[\left|\frac{S_n}{n} - \mu\right|^4\right]}{n^{-4\gamma}}$

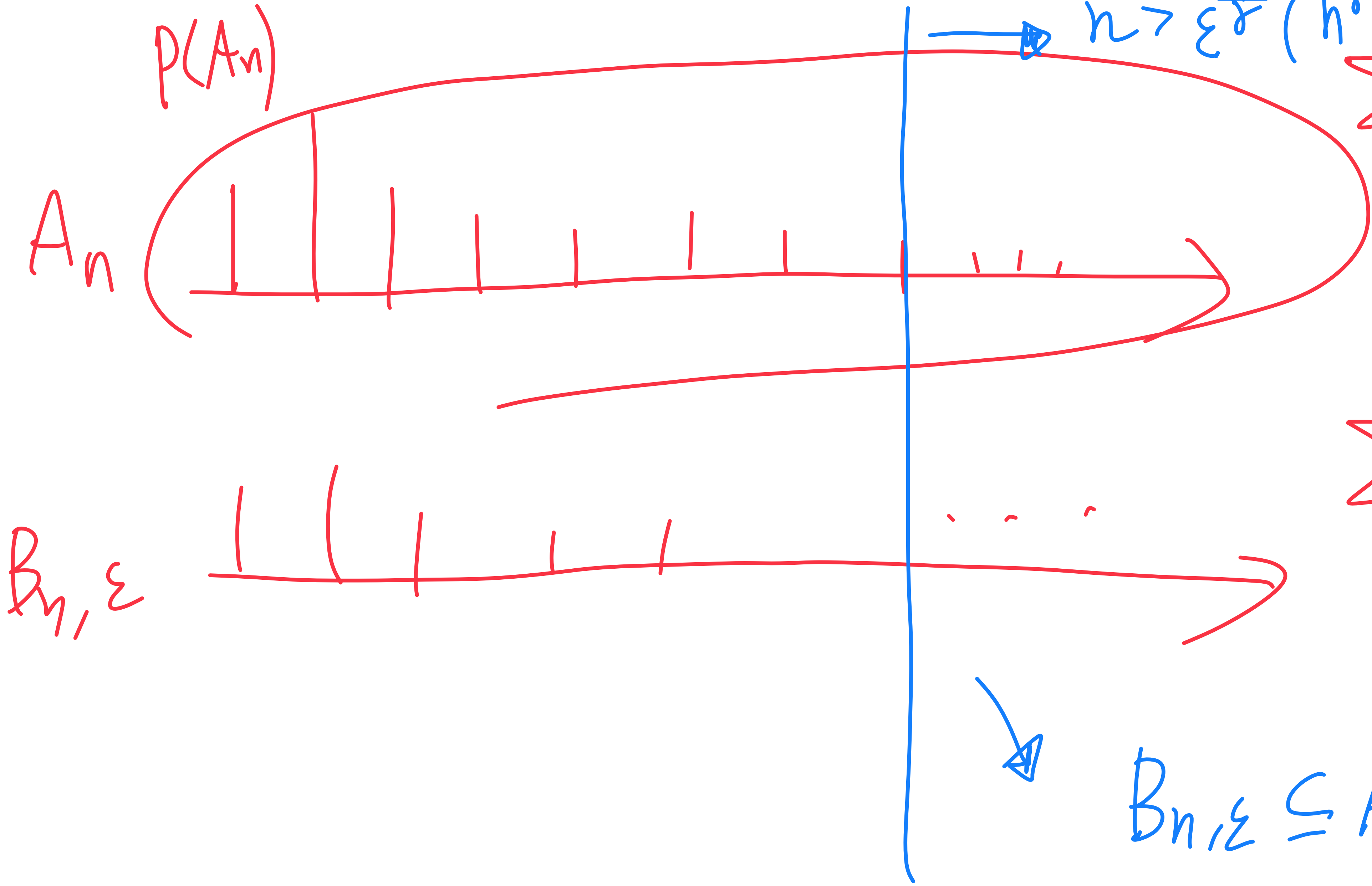
$\le \dfrac{K}{n^2 \cdot n^{-4\gamma}} = \dfrac{K}{n^{2-4\gamma}}$

Step 2: $\sum_{n=1}^{\infty} P(A_n) \le \sum_{n=1}^{\infty} \frac{K}{n^{2-4\gamma}} = K \cdot \sum_{n=1}^{\infty} \frac{1}{n^{2-4\gamma}} < \infty$

If we choose $\gamma \in (0, \frac{1}{4})$

Step 3: $\boxed{\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow \sum_{n=1}^{\infty} P(B_{n,\varepsilon}) < \infty}$

By Borel-Cantelli Lemma, $P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} B_{n,\varepsilon}\right) = 0$ (SLLN)



$n > \varepsilon^{-\frac{1}{\gamma}}$ $(n^{-\gamma} < \varepsilon)$

$\sum P(A_n) < \infty$

$\sum P(B_{n,\varepsilon}) < \infty$

$B_{n,\varepsilon} \subseteq A_n$, if $n > \varepsilon^{-\frac{1}{\gamma}}$

$\Downarrow$

$P(B_{n,\varepsilon}) \leq P(A_n)$

if $n > \varepsilon^{-\frac{1}{\gamma}}$

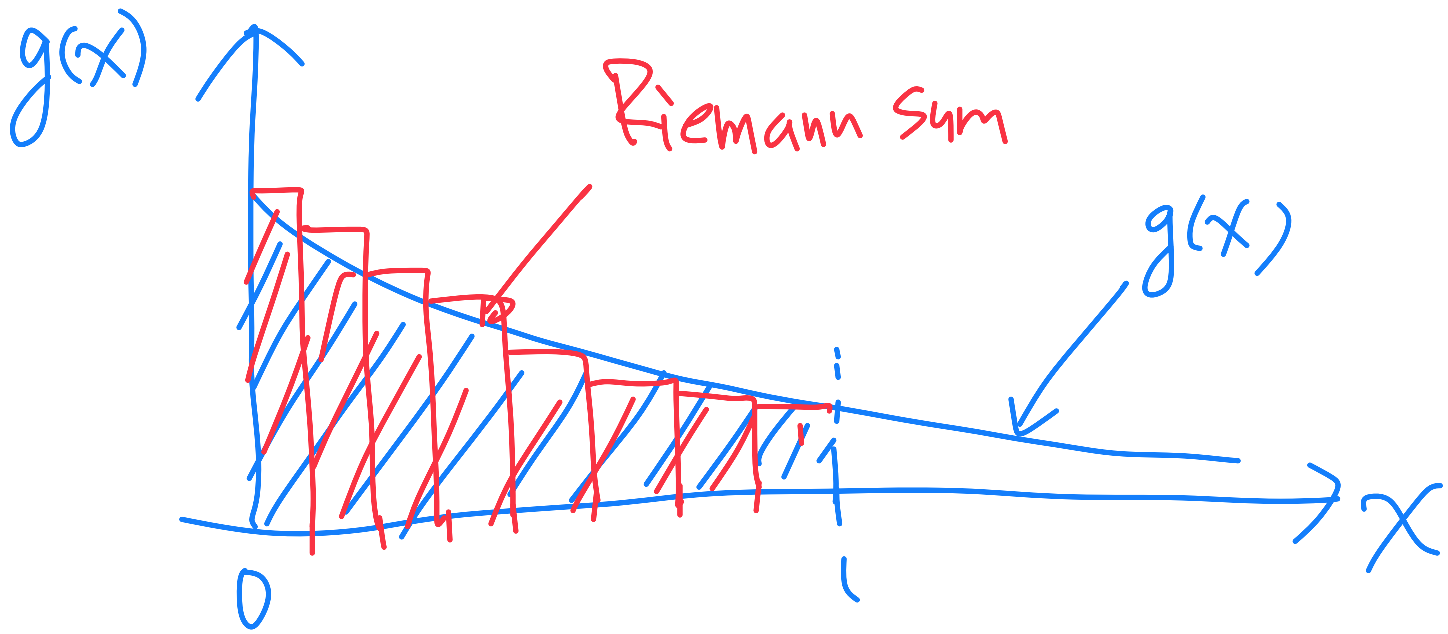
Application of SLLN: Monte Carlo Simulation

- A Motivating Example: Find the following integration

$$I = \int_0^1 \underbrace{e^{-x^3}}_{g(x)} dx$$

$$\int_0^1 e^{-x^3} dx$$

- Question: Is there a closed-form expression for I?
- Question: Could we approximate I by taking the Riemann sum?



 http://faculty.washington.edu/yenchic/18A_stat516/Lec7_MCMC.pdf

Application of SLLN: Monte Carlo Simulation (Cont.)

- **A Motivating Example:** Find the following integration

$$I = \int_0^1 e^{-x^3} dx = \int_0^1 \underbrace{1}_{f(x)} \cdot \underbrace{e^{-x^3}}_{g(x)} dx \overset{\text{LOTUS}}{=} E[e^{-X^3}]$$

$X \sim \mathrm{Unif}(0,1)$

- **Monte-Carlo method:** Let $U \sim \mathrm{Unif}(0,1)$

1. Let $U \sim \mathrm{Unif}(0,1)$. Rewrite $\int_0^1 e^{-x^3} dx = E\left[e^{-U^3}\right]$

2. Draw K i.i.d. random variables $U_1, \cdots, U_K \sim \mathrm{Unif}(0,1)$

$$E\left[e^{-U^3}\right] \approx \frac{1}{K}\sum_{i=1}^{K} e^{-U_i^3} \quad \text{(Why?)}$$

empirical mean

SLLN

Concentration inequalities

 http://faculty.washington.edu/yenchic/18A_stat516/Lec7_MCMC.pdf

Monte Carlo Simulation (Formally)

- **Objective:** Find the integration $I = \int g(x)dx$
- **Monte Carlo Simulation:**

1. Let X be a random variable with PDF $p(x)$. Rewrite I as

$$I = \underbrace{\int \frac{g(x)}{p(x)} \overset{\text{PDF}}{p(x)} dx}_{} = E_{X \sim p(x)}\left[\frac{g(X)}{p(X)}\right] \quad \text{(LOTUS)}$$

2. Draw K i.i.d. random variables $X_1, \cdots, X_K$ with PDF $p(x)$

Construct
$$\hat{I}_K = \frac{1}{K}\sum_{i=1}^{K}\frac{g(X_i)}{p(X_i)} \approx I$$

- **Question:** How to choose K?

Variance Issue in Monte-Carlo Simulation

$$I = \int g(x)dx = \int \frac{g(x)}{p(x)} p(x)dx = E_{p(x)}\left[\frac{g(X)}{p(X)}\right]$$

$$\hat{I}_K = \frac{1}{K}\sum_{i=1}^{K} \frac{g(X_i)}{p(X_i)} \approx I$$

- Question: $E[\hat{I}_K]$ = ? Var$[\hat{I}_K]$ = ?

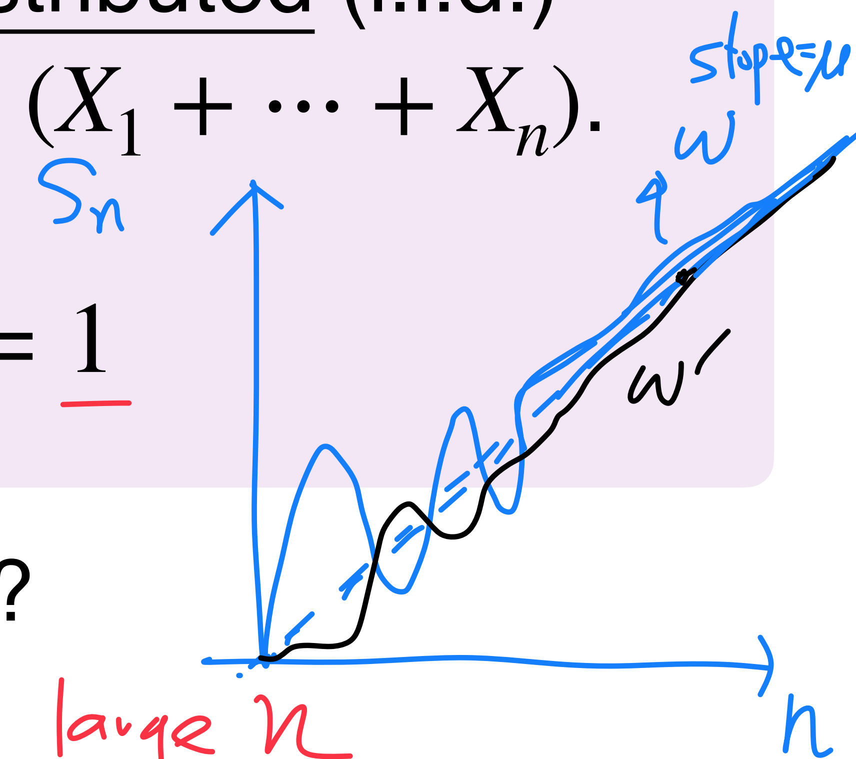
 http://faculty.washington.edu/yenchic/18A_stat516/Lec7_MCMC.pdf

Central Limit Theorem

Beyond SLLN

- **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

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



- Question: What does SLLN say about $S_n(\omega)$?

$S_n(\omega) = n \cdot \mu$ for all ω, large n

(scale linearly with n)

sublinear in n

- Question: Do we have $S_n(\omega) = n\mu + o(n)$?

$S_n(\omega) - n\mu$

$$\begin{cases} S_n(\omega) = n\mu + \sqrt{n} \\ S_n(\omega) = n\mu + \log n \end{cases} \Rightarrow \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu$$

Recall (Lecture 14): Binomial and Normal

- **Example:** $X_1, X_2, \cdots$ are i.i.d. Bernoulli r.v.s with mean μ and variance $\sigma^2 = \mu(1-\mu)$
  - Define $S_n = X_1 + X_2 + \cdots + X_n$
  - **Question**: What type of r.v. is S_n? $E[S_n]$ = ? Var$[S_n]$ = ?

$S_n \sim \text{Binomial}(n, \mu)$

$E[S_n] = n \cdot \mu$

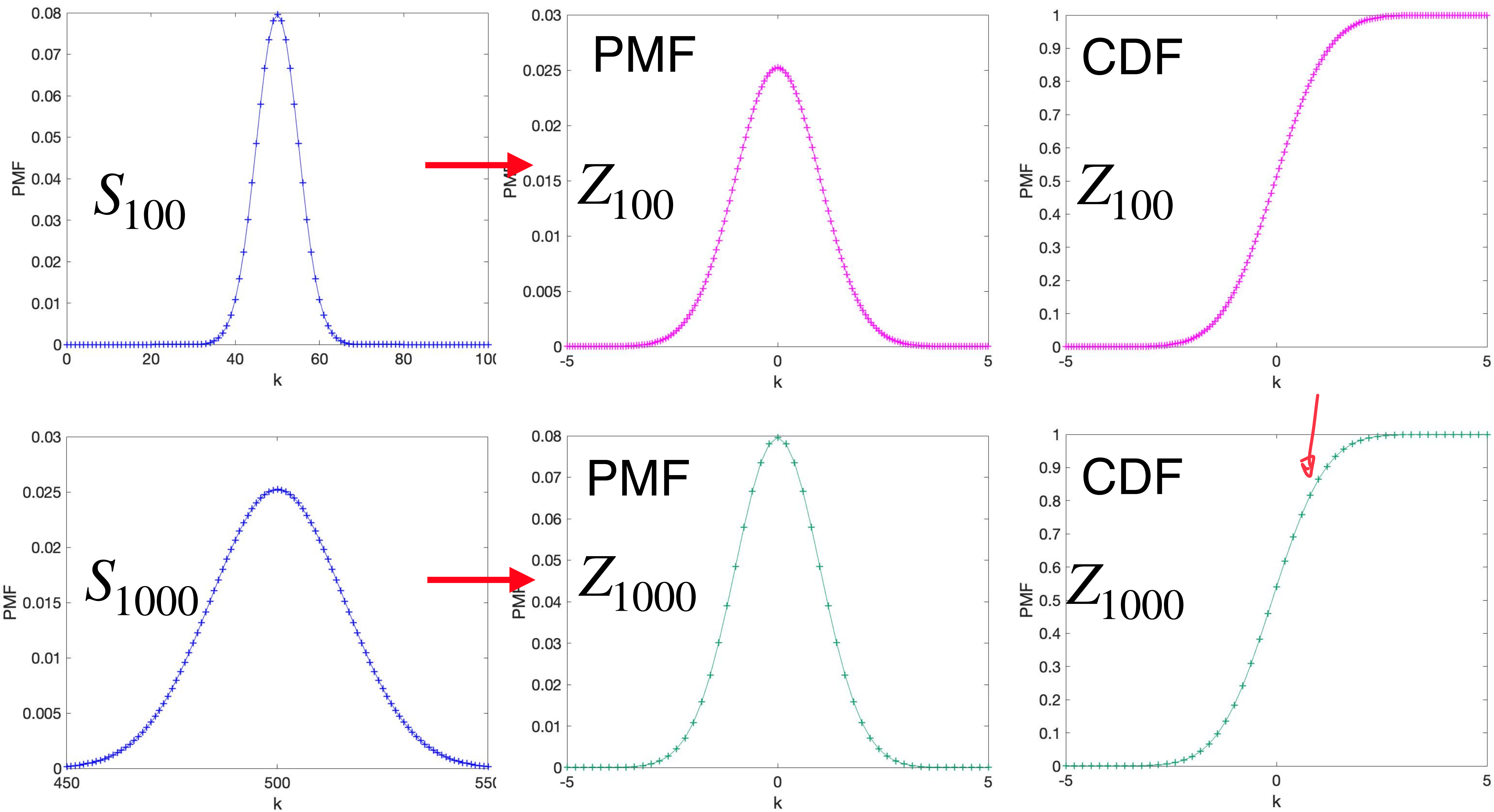
$\text{Var}[S_n] = n \cdot (\mu \cdot (1-\mu)) = n \cdot \sigma^2$

- **Question:** How to find the distribution of $\dfrac{S_n - \overset{E[S_n]}{n\mu}}{\sigma\sqrt{n}}$? $\sigma\sqrt{n} = \sqrt{\text{Var}[S_n]}$

Recall: Plotting $Z_n = (S_n - n\mu)/(\sigma\sqrt{n})$

- Example: $\mu = 0.5$

$\approx$ CDF of a standard normal!



Central Limit Theorem (Formally)

$P(Z \le z)$

- **Central Limit Theorem (CLT)**: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ and variance σ^2. Define $S_n = (X_1 + \cdots + X_n)$. Then, we have

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le z\right) = \Phi(z), \forall z \in \mathbb{R}$$

Z $\quad E[S_n] \quad \sqrt{Var[S_n]}$ $\quad$ CDF of standard normal

where $\Phi(z)$ is the CDF of standard normal

- Question: How to interpret such convergence?

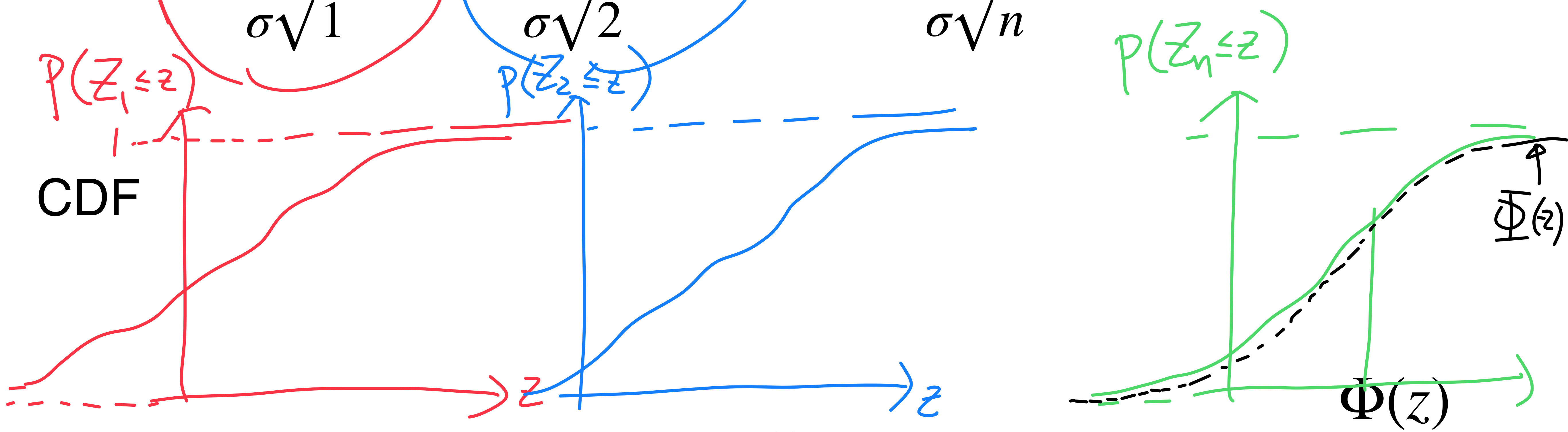
$S_n = n\mu + \sigma\sqrt{n} \cdot Z$

How to Interpret CLT?

- Let $X_1, X_2, \cdots$ be i.i.d. random variables with mean μ and variance σ^2
- Define $S_n = (X_1 + X_2 \cdots + X_n)$
- CLT: $\lim_{n\to\infty} P\left(\underbrace{\frac{S_n - n\mu}{\sigma\sqrt{n}}}_{Z_n =} \leq z\right) = \Phi(z)$

$$\underbrace{\frac{S_1 - 1\cdot\mu}{\sigma\sqrt{1}}}_{Z_1 =} \quad \underbrace{\frac{S_2 - 2\cdot\mu}{\sigma\sqrt{2}}}_{Z_2 =} \quad \cdots \quad \frac{S_n - n\cdot\mu}{\sigma\sqrt{n}} \quad \cdots$$

CDF



Convergence in Distribution (Formally)

- **Convergence in Distribution**: Let $F_1, F_2 \cdots, F_n, \cdots$ be a sequence of CDFs of random variables $Y_1, Y_2, \cdots, Y_n, \cdots$. We say that $\{Y_n\}$ converges in distribution to a random variable Y with CDF F if $\forall t \in \mathbb{R}$,

$$\lim_{n\to\infty} F_n(t) = F(t)$$

(pointwise)

- Remark: Convergence in distribution is also called "weak convergence"
- Notation: $Y_n \Rightarrow Y$ or $F_n \Rightarrow F$
- Question: Why is such convergence regarded "weak"?

“Convergence in Distribution” does not imply “Convergence in Probability” or “Almost-Sure Convergence”

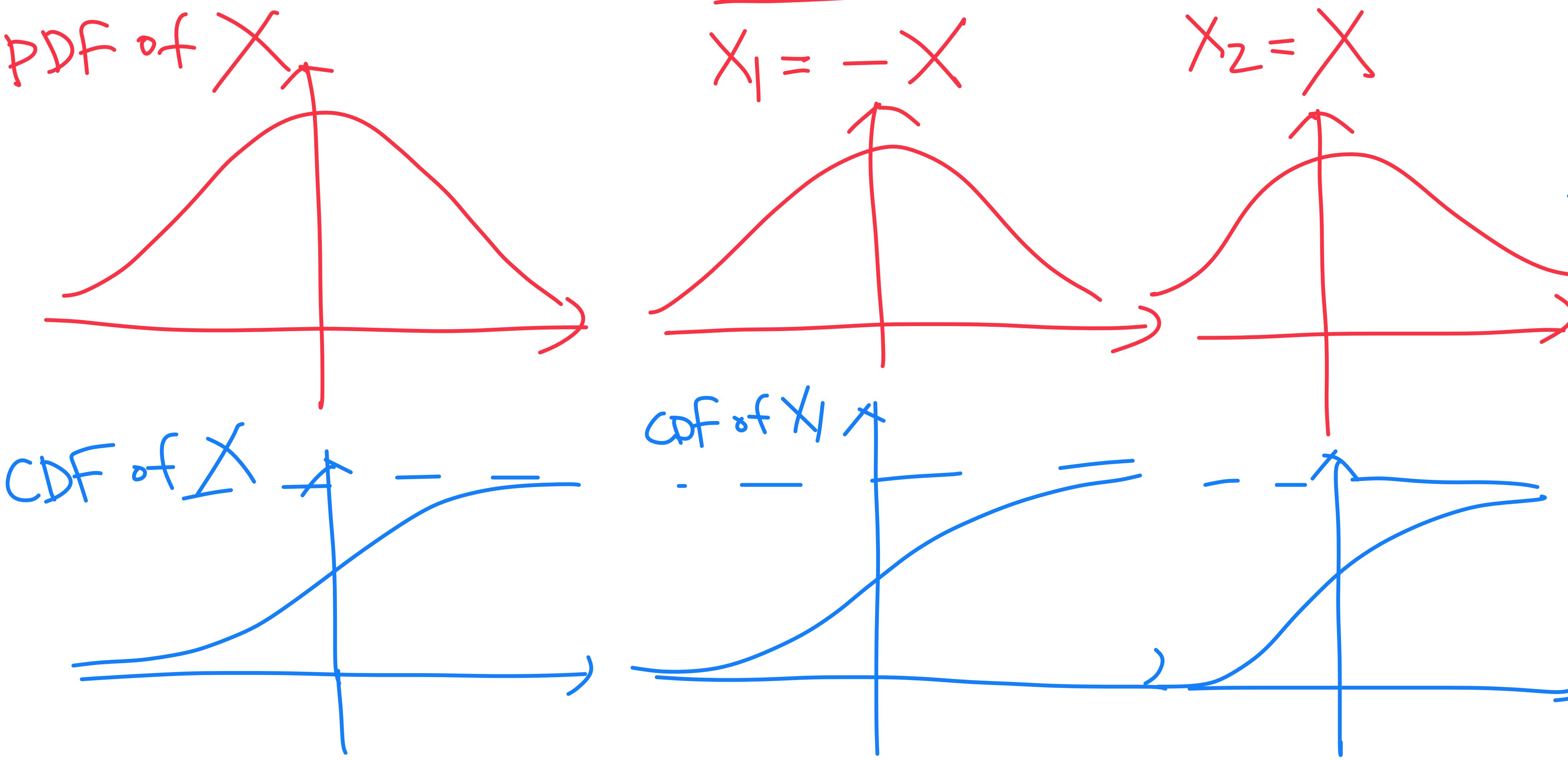
$X(\omega) = 1$

$\lim_{n\to\infty} P(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \neq 0$

$X_1(\omega) = -1, \quad X_3(\omega) = -1, \ \ldots$

$X_2(\omega) = 1 \quad X_4(\omega) = 1$

- **Example:** Let $X \sim N(0,1)$
  - For every $n \in \mathbb{N}$, let $X_n = (-1)^n X$, i.e., $X_n(\omega) = (-1)^n X(\omega)$
  - **Question**: Do we have $X_n \Rightarrow X$? How about $X_n \xrightarrow{\text{p}} X$?



$X_n \Rightarrow X$

$X_n \not\xrightarrow{p} X$

Why is CLT Useful? Approximation!

- Recall that $S_n = (X_1 + X_2 \cdots + X_n)$, where $X_1, \cdots, X_n$ are i.i.d.
- CLT: $\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le z\right) = \Phi(z)$ ← CDF of standard normal
- Idea: For large n, consider $\boxed{P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le z\right) \approx \Phi(z)}$ to find $P(S_n \le c)$ for any c

$$\underbrace{P(S_n \le c)}_{\text{CDF of } S_n} = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le \underbrace{\frac{c - n\mu}{\sigma\sqrt{n}}}_{= z}\right) \approx \Phi\left(\frac{c-n\mu}{\sigma\sqrt{n}}\right)$$

Example: Approximation via CLT

- Example: $X_1, \cdots, X_{20}$ are 20 i.i.d. continuous uniform r.v.s on (0,1)
  - Question: Find $P(\sum_{i=1}^{20} X_i \leq 8)$ using approximation?
  - Hint: $P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right) \approx \Phi(z)$

S_{20}

$E[X_i] = \frac{1}{2}$

$Var[X_i] = \frac{1}{12}$

$\sigma = \sqrt{\frac{1}{12}}$

$$P\left(\sum_{i=1}^{20} X_i \leq 8\right) = P\left(\frac{S_{20} - 20\times\frac{1}{2}}{\sqrt{\frac{1}{12}}\cdot\sqrt{20}} \leq \frac{8 - 20\times\frac{1}{2}}{\sqrt{\frac{1}{12}}\cdot\sqrt{20}}\right)$$

$$\approx \Phi\left(\frac{8 - 20\times\frac{1}{2}}{\sqrt{\frac{1}{12}}\cdot\sqrt{20}}\right)$$

Now let's prove CLT!

Review: From MGF to Distributions

- Recall: Lecture 23

- **MGF Uniqueness Theorem**: Let X_1 and X_2 be two random variables with MGFs $M_{X_1}(t)$ and $M_{X_2}(t)$, respectively. If $M_{X_1}(t) = M_{X_2}(t)$ for all t in some interval $(-\alpha, \alpha)$, then X_1 and X_2 follow the same distribution, i.e.

$$P(X_1 \leq u) = P(X_2 \leq u), \text{ for all } u \in \mathbb{R}$$

Use MGF to Show CLT

- Idea: Suppose we find the MGF of $\dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ for $n \to \infty$
  - Question: Can we find its distribution?

- **Levy Continuity Theorem**: Let $V_1, V_2 \cdots$ be a sequence of random variables with CDFs $F_1, F_2, \cdots$ and MGFs $M_{V_1}(t), M_{V_2}(t) \cdots$. Let V be a random variable with CDF F and MGF $M_V(t)$. If for every $t \in \mathbb{R}$, $\lim_{n \to \infty} M_{V_n}(t) = M_V(t)$, then the CDFs F_n converge to F.

- Remark: MGF of $\mathcal{N}(\mu, \sigma^2)$ is $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

Use MGF to Show CLT (Cont.)

- Example: $X_1, X_2, \cdots$ are i.i.d. r.v.s with mean μ and variance σ^2
  - Define $S_n = X_1 + X_2 + \cdots + X_n$ and $Y_i = X_i - \mu$
  - Question: $E[Y_i]$ = ________? Var$[Y_i]$ = ________?
  - Question: What is the MGF of $\dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ (in terms of MGF of Y_i)?

Use MGF to Show CLT (Cont.)

- Question: When $n \to \infty$, what is the MGF of $\dfrac{S_n - n\mu}{\sigma\sqrt{n}}$?

Most likely we will forget about all the details in about 2 months…

Despite this, we will still know how to think probabilistically

Final Takeaway

1. Sample space and random variables

2. Leverage independence or normal structure

3. Limits could be very useful (e.g. LLN and CLT)

4. Bayesian view and conditioning could help simplify the problem