Lecture 6: Law of Large Numbers and Central Limit Theorem

Yi, Yung (이융)

EE210: Probability and Introductory Random Processes
KAIST EE

June 2, 2021

Roadmap

(1) Weak Law of Large Numbers: Result and Meaning

(2) Central Limit Theorem: Result and Meaning

(3) Weak Law of Large Numbers: Proof
- Inequalities: Markov and Chebyshev

(4) Central Limit Theorem: Proof
- Moment Generating Function (MGF)

◦ Two most remarkable findings in probability theory

Roadmap

(1) Weak Law of Large Numbers: Result and Meaning

(2) Central Limit Theorem: Result and Meaning

(3) Weak Law of Large Numbers: Proof
- Inequalities: Markov and Chebyshev

(4) Central Limit Theorem: Proof
- Moment Generating Function (MGF)

- Example 1. n students who decides their presence, depending on their feeling. Each student is happy or sad at random, and only happy students will show their presence. How many students will show their presence?

Our interest: Sum of Random Variables

- Example 1. n students who decides their presence, depending on their feeling. Each student is happy or sad at random, and only happy students will show their presence. How many students will show their presence?

- Example 2. I am hearing some sound. There are n noisy sources from outside.

Our interest: Sum of Random Variables

- Example 1. n students who decides their presence, depending on their feeling. Each student is happy or sad at random, and only happy students will show their presence. How many students will show their presence?
- Example 2. I am hearing some sound. There are n noisy sources from outside.
- $X_1, X_2, \ldots, X_n$: i.i.d (independent and identically distributed) random variables

Our interest: Sum of Random Variables

- Example 1. n students who decides their presence, depending on their feeling. Each student is happy or sad at random, and only happy students will show their presence. How many students will show their presence?
- Example 2. I am hearing some sound. There are n noisy sources from outside.

- $X_1, X_2, \dots, X_n$: i.i.d (independent and identically distributed) random variables
- $\mathbb{E}[X_i] = \mu$, $\text{var}[X_i] = \sigma^2$

Our interest: Sum of Random Variables

- Example 1. n students who decides their presence, depending on their feeling. Each student is happy or sad at random, and only happy students will show their presence. How many students will show their presence?
- Example 2. I am hearing some sound. There are n noisy sources from outside.

- $X_1, X_2, \ldots, X_n$: i.i.d (independent and identically distributed) random variables
- $\mathbb{E}[X_i] = \mu$, $\text{var}[X_i] = \sigma^2$
- Our interest is to understand how the following sum behaves:

$$S_n = X_1 + X_2 + \ldots + X_n$$

What About Distribution of S_n?

$$S_n = X_1 + X_2 + \ldots + X_n$$

What About Distribution of S_n?

$$S_n = X_1 + X_2 + \ldots + X_n$$

- Figure out the distribution of S_n. Very challenging. Even just for $Z = X + Y$, finding the distribution, for example, requires the complex convolution.

$$p_Z(z) = \mathbb{P}(X + Y = z) = \sum_x p_X(x) p_Y(z - x)$$

What About Distribution of S_n?

$$S_n = X_1 + X_2 + \ldots + X_n$$

- Figure out the distribution of S_n. Very challenging. Even just for $Z = X + Y$, finding the distribution, for example, requires the complex convolution.

$$p_Z(z) = \mathbb{P}(X + Y = z) = \sum_x p_X(x) p_Y(z - x)$$

- Easy case: Sum of normal rvs = a normal rv, however, generally very challenging.

What About Distribution of S_n?

$$S_n = X_1 + X_2 + \ldots + X_n$$

- Figure out the distribution of S_n. Very challenging. Even just for $Z = X + Y$, finding the distribution, for example, requires the complex convolution.

  $$p_Z(z) = \mathbb{P}(X + Y = z) = \sum_x p_X(x) p_Y(z - x)$$

- Easy case: Sum of normal rvs = a normal rv, however, generally very challenging.
- Possible appoach. Take a certain scaling with respect to n that corresponds to a new glass, and investigate the system for large n

Sample Mean

- Consider the sample mean, and try to understand how S_n behaves:

$$\boxed{M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}}$$

Sample Mean

- Consider the sample mean, and try to understand how S_n behaves:

$$\boxed{M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}}$$

- $\mathbb{E}(M_n) = \mu$, $\text{var}(M_n) = \sigma^2/n$

Sample Mean

- Consider the sample mean, and try to understand how S_n behaves:

$$\boxed{M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \dots X_n}{n}}$$

- $\mathbb{E}(M_n) = \mu$, $\text{var}(M_n) = \sigma^2/n$
- For large n, the variance $\text{var}(M_n)$ decays. We expect that, for large n, M_n looses its randomness and concentrates around μ.

Sample Mean

- Consider the sample mean, and try to understand how S_n behaves:

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \dots X_n}{n}$$

- $\mathbb{E}(M_n) = \mu$, $\text{var}(M_n) = \sigma^2/n$
- For large n, the variance $\text{var}(M_n)$ decays. We expect that, for large n, M_n looses its randomness and concentrates around μ.
- We call this law of large numbers (LLN).

Let's Establish Mathematically

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

Let's Establish Mathematically

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

- What about this? What's wrong?

$$M_n \xrightarrow{n \to \infty} \mu$$

Let's Establish Mathematically

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

- What about this? What's wrong?

$$M_n \xrightarrow{n \to \infty} \mu$$

- Ordinary convergence for the sequence of real numbers: $a_n \to L$

Let's Establish Mathematically

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \dots X_n}{n}$$

- What about this? What's wrong?

$$M_n \xrightarrow{n \to \infty} \mu$$

- Ordinary convergence for the sequence of real numbers: $a_n \to L$

For every $\epsilon > 0$, there exists $N = N(\epsilon)$, such that for every $n \geq N$, $|a_n - L| \leq \epsilon$.

Let's Establish Mathematically

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \dots X_n}{n}$$

- What about this? What's wrong?

$$M_n \xrightarrow{n \to \infty} \mu$$

- Ordinary convergence for the sequence of real numbers: $a_n \to L$

For every $\epsilon > 0$, there exists $N = N(\epsilon)$, such that for every $n \geq N$, $|a_n - L| \leq \epsilon$.

https://www.youtube.com/watch?v=4nBmsRA6eVw

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

- What about this? What's wrong?

$$M_n \xrightarrow{n \to \infty} \mu$$

- Ordinary convergence for the sequence of real numbers: $a_n \to L$

For every $\epsilon > 0$, there exists $N = N(\epsilon)$, such that for every $n \geq N$, $|a_n - L| \leq \epsilon$.

`https://www.youtube.com/watch?v=4nBmsRA6eVw`

- However, M_n is a random variable, which is a function from Ω to $\mathbb{R}$.

Let's Establish Mathematically

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

- What about this? What's wrong?

$$M_n \xrightarrow{n \to \infty} \mu$$

- Ordinary convergence for the sequence of real numbers: $a_n \to L$

For every $\epsilon > 0$, there exists $N = N(\epsilon)$, such that for every $n \geq N$, $|a_n - L| \leq \epsilon$.

`https://www.youtube.com/watch?v=4nBmsRA6eVw`

- However, M_n is a random variable, which is a function from Ω to $\mathbb{R}$.
- Need to build up the new concept of convergence for the sequence of rvs.

Convergence in Probability (1)

- What we want: a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ converges to a rv Y in some sense

Convergence in Probability (1)

- What we want: a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ converges to a rv Y in some sense
- For any given $\epsilon > 0$, consider the sequence of events $A_n = \{|Y_n - Y| \geq \epsilon\}$, and compute its sequence of probabilities $a_n = \mathbb{P}(A_n) = \mathbb{P}(|Y_n - Y| \geq \epsilon)$.

Convergence in Probability (1)

- What we want: a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ converges to a rv Y in some sense
- For any given $\epsilon > 0$, consider the sequence of events $A_n = \{|Y_n - Y| \geq \epsilon\}$, and compute its sequence of probabilities $a_n = \mathbb{P}(A_n) = \mathbb{P}(|Y_n - Y| \geq \epsilon)$.
- Now, $\{a_n\}$ are just the real numbers, and show that $a_n \to 0$ as $n \to \infty$.

Convergence in Probability (1)

- What we want: a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ converges to a rv Y in some sense
- For any given $\epsilon > 0$, consider the sequence of events $A_n = \{|Y_n - Y| \geq \epsilon\}$, and compute its sequence of probabilities $a_n = \mathbb{P}(A_n) = \mathbb{P}(|Y_n - Y| \geq \epsilon)$.
- Now, $\{a_n\}$ are just the real numbers, and show that $a_n \to 0$ as $n \to \infty$.
- To show that $a_n \to 0$ as $n \to \infty$, which is just the ordinary convergence, we show:

Convergence in Probability (1)

- What we want: a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ converges to a rv Y in some sense
- For any given $\epsilon > 0$, consider the sequence of events $A_n = \{|Y_n - Y| \geq \epsilon\}$, and compute its sequence of probabilities $a_n = \mathbb{P}(A_n) = \mathbb{P}(|Y_n - Y| \geq \epsilon)$.
- Now, $\{a_n\}$ are just the real numbers, and show that $a_n \to 0$ as $n \to \infty$.
- To show that $a_n \to 0$ as $n \to \infty$, which is just the ordinary convergence, we show:
  - For any $\delta > 0$, there exists $N = N(\delta)$, such that for all $n \geq N$, $|a_n - 0| \leq \delta$

KAIST EE

- What we want: a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ converges to a rv Y in some sense
- For any given $\epsilon > 0$, consider the sequence of events $A_n = \{|Y_n - Y| \geq \epsilon\}$, and compute its sequence of probabilities $a_n = \mathbb{P}(A_n) = \mathbb{P}(|Y_n - Y| \geq \epsilon)$.
- Now, $\{a_n\}$ are just the real numbers, and show that $a_n \to 0$ as $n \to \infty$.
- To show that $a_n \to 0$ as $n \to \infty$, which is just the ordinary convergence, we show:
  - For any $\delta > 0$, there exists $N = N(\delta)$, such that for all $n \geq N$, $|a_n - 0| \leq \delta$
- Convergence in probability: $Y_n \xrightarrow{\text{in prob.}} Y$
  - For any $\epsilon > 0$ and for any $\delta > 0$, there exists $N = N(\delta)$, such that for all $n \geq N$, $\mathbb{P}(|Y_n - Y| \geq \epsilon) \leq \delta$.

Convergence in Probability (1)

- What we want: a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ converges to a rv Y in some sense
- For any given $\epsilon > 0$, consider the sequence of events $A_n = \{|Y_n - Y| \geq \epsilon\}$, and compute its sequence of probabilities $a_n = \mathbb{P}(A_n) = \mathbb{P}(|Y_n - Y| \geq \epsilon)$.
- Now, $\{a_n\}$ are just the real numbers, and show that $a_n \to 0$ as $n \to \infty$.
- To show that $a_n \to 0$ as $n \to \infty$, which is just the ordinary convergence, we show:
  - For any $\delta > 0$, there exists $N = N(\delta)$, such that for all $n \geq N$, $|a_n - 0| \leq \delta$
- Convergence in probability: $Y_n \xrightarrow{\text{in prob.}} Y$
  - For any $\epsilon > 0$ and for any $\delta > 0$, there exists $N = N(\delta)$, such that for all $n \geq N$, $\mathbb{P}(|Y_n - Y| \geq \epsilon) \leq \delta$.
  - For any $\epsilon > 0$, $\mathbb{P}\left(\{|Y_n - Y| \geq \epsilon\}\right) \xrightarrow{n \to \infty} 0$.

Convergence in Probability (2)

KAIST EE

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - Y| \geq \epsilon\}) \xrightarrow{n\to\infty} 0.$
- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0.$

- A special case: when $Y = a$ for some constant a: $Y_n \xrightarrow{\text{in prob.}} a$
- `https://youtu.be/Ajar_6MAOLw?t=248`

Example 1: Convergence in Probability

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0$.

Example 1: Convergence in Probability

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0$.

- A sequence of iid rvs $X_n \sim \mathcal{U}[0,1]$, and let

$$Y_n = \min\{X_1, X_2, \ldots, X_n\}$$

Example 1: Convergence in Probability

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0$.

- A sequence of iid rvs $X_n \sim \mathcal{U}[0, 1]$, and let
$$Y_n = \min\{X_1, X_2, \ldots, X_n\}$$
- Our intuition: Y_n converges to 0, as $n \to 0$. Why?

KAIST EE

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0$.

- A sequence of iid rvs $X_n \sim \mathcal{U}[0,1]$, and let

$$Y_n = \min\{X_1, X_2, \ldots, X_n\}$$

- Our intuition: Y_n converges to 0, as $n \to 0$. Why?

- Proof. For any $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}(|Y_n - 0| \geq \epsilon) &= \mathbb{P}(X_1 \geq \epsilon, \ldots, X_n \geq \epsilon) = \mathbb{P}(X_1 \geq \epsilon) \times \cdots \times \mathbb{P}(X_n \geq \epsilon) \\ &= (1-\epsilon)^n \xrightarrow{n\to\infty} 0\end{aligned}$$

Example 2: Convergence in Probability

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0$.

Example 2: Convergence in Probability

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0.$

- Y: exponential rv with the parameter $\lambda = 1$ (Remind: $\mathbb{P}(Y > y) = e^{-\lambda y}$)
- a sequence of rvs $Y_n = Y/n$ (note that these are dependent)

Example 2: Convergence in Probability

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0$.

- Y: exponential rv with the parameter $\lambda = 1$ (Remind: $\mathbb{P}(Y > y) = e^{-\lambda y}$)
- a sequence of rvs $Y_n = Y/n$ (note that these are dependent)
- Our intuition: Y_n converges to 0

Example 2: Convergence in Probability

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0$.

- Y: exponential rv with the parameter $\lambda = 1$ (Remind: $\mathbb{P}(Y > y) = e^{-\lambda y}$)
- a sequence of rvs $Y_n = Y/n$ (note that these are dependent)
- Our intuition: Y_n converges to 0
- Proof. For any $\epsilon > 0$,

  $\mathbb{P}(|Y_n - 0| \geq \epsilon) = \mathbb{P}(Y \geq n\epsilon) = e^{-n\epsilon} \xrightarrow{n\to\infty} 0$

Example 3: Convergence in Probability

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0$.

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0$.

- Consider a sequence of rvs Y_n with the following distribution:

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

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0$.

- Consider a sequence of rvs Y_n with the following distribution:

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

- For any $\epsilon > 0$,

$$\mathbb{P}(|Y_n| \geq \epsilon) = \frac{1}{n} \xrightarrow{n\to\infty} 0$$

Example 3: Convergence in Probability

- For any $\epsilon > 0$, $\mathbb{P}(\{|Y_n - a| \geq \epsilon\}) \xrightarrow{n\to\infty} 0$.

- Consider a sequence of rvs Y_n with the following distribution:

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

- For any $\epsilon > 0$,

$$\mathbb{P}(|Y_n| \geq \epsilon) = \frac{1}{n} \xrightarrow{n\to\infty} 0$$

- Thus, Y_n converges to 0 in probability.

Weak Law of Large Numbers (WLLN)

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \dots X_n}{n}$$

- Roughly, M_n concetrates around μ

Weak Law of Large Numbers (WLLN)

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \dots X_n}{n}$$

- Roughly, M_n concetrates around μ

Weak law of large numbers

M_n converges to μ in probability, i.e., $M_n \xrightarrow{\text{in prob.}} \mu$

Weak Law of Large Numbers (WLLN)

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

- Roughly, M_n concetrates around μ

Weak law of large numbers

M_n converges to μ in probability, i.e., $M_n \xrightarrow{\text{in prob.}} \mu$

- Why "Weak"? There exists a stronger version, which we call "strong" law of large numbers. We will not cover the strong law of large numbers in this class.

KAIST EE

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \dots X_n}{n}$$

- Roughly, M_n concetrates around μ

Weak law of large numbers

M_n converges to μ in probability, i.e., $M_n \xrightarrow{\text{in prob.}} \mu$

- Why "Weak"? There exists a stronger version, which we call "strong" law of large numbers. We will not cover the strong law of large numbers in this class.
- The proof requires some knowledge about useful inequalities, which we will cover later.

WLLN: Why Useful?

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

WLLN: Why Useful?

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \dots X_n}{n}$$

- If we take the scaling of S_n by $1/n$, it behaves like a deterministic number. This significantly simplifies how we understand the world.

WLLN: Why Useful?

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

- If we take the scaling of S_n by $1/n$, it behaves like a deterministic number. This significantly simplifies how we understand the world.
- For example, assume that a large number of identically distributed noises come to you. Then, you can roughly approximate it as ($n \times$ average noise)

WLLN: Why Useful?

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

- If we take the scaling of S_n by $1/n$, it behaves like a deterministic number. This significantly simplifies how we understand the world.
- For example, assume that a large number of identically distributed noises come to you. Then, you can roughly approximate it as $(n \times \text{average noise})$
- Provides an interpretation of expectations (as well as probabilities) in terms of a long sequence of identical independent experiments. For example, what is the probability of head of a coin? Toss 1000 times, and count the number of heads.

(1) Weak Law of Large Numbers: Result and Meaning

(2) Central Limit Theorem: Result and Meaning

(3) Weak Law of Large Numbers: Proof
- Inequalities: Markov and Chebyshev

(4) Central Limit Theorem: Proof
- Moment Generating Function (MGF)

- Loosely speaking, WLLG says:

  $$(M_n - \mu) \xrightarrow{n \to \infty} 0$$

Central Limit Theorem: Start with Scaling (1)

- Loosely speaking, WLLG says:

  $(M_n - \mu) \xrightarrow{n\to\infty} 0$

- However, we don't know how $M_n - \mu$ converges to 0. For example, what's the speed of convergence?

Central Limit Theorem: Start with Scaling (1)

- Loosely speaking, WLLG says:

  $$(M_n - \mu) \xrightarrow{n \to \infty} 0$$

- However, we don't know how $M_n - \mu$ converges to 0. For example, what's the speed of convergence?

- Question. What should be "something"? Something should blow up for large n.

  $$\boxed{(\text{something})} \times (M_n - \mu) \xrightarrow{n \to \infty} \text{meaningful thing}$$

Central Limit Theorem: Start with Scaling (1)

- Loosely speaking, WLLG says:

$$(M_n - \mu) \xrightarrow{n \to \infty} 0$$

- However, we don't know how $M_n - \mu$ converges to 0. For example, what's the speed of convergence?

- Question. What should be "something"? Something should blow up for large n.

$$\boxed{\text{(something)}} \times (M_n - \mu) \xrightarrow{n \to \infty} \text{meaningful thing}$$

$$\boxed{n^{\alpha}} \times (M_n - \mu) \xrightarrow{n \to \infty} \text{meaningful thing}$$

KAIST EE

- Loosely speaking, WLLG says:

  $(M_n - \mu) \xrightarrow{n \to \infty} 0$

- However, we don't know how $M_n - \mu$ converges to 0. For example, what's the speed of convergence?

- Question. What should be "something"? Something should blow up for large n.

  $$\boxed{\text{(something)}} \times (M_n - \mu) \xrightarrow{n \to \infty} \text{meaningful thing}$$

  $$\boxed{n^{\alpha}} \times (M_n - \mu) \xrightarrow{n \to \infty} \text{meaningful thing}$$

- What's α for our magic?

Central Limit Theorem: Start with Scaling (1)

- Loosely speaking, WLLG says:

$$(M_n - \mu) \xrightarrow{n\to\infty} 0$$

- However, we don't know how $M_n - \mu$ converges to 0. For example, what's the speed of convergence?

- Question. What should be "something"? Something should blow up for large n.

$$\boxed{\text{(something)}} \times (M_n - \mu) \xrightarrow{n\to\infty} \text{meaningful thing}$$

$$\boxed{n^{\alpha}} \times (M_n - \mu) \xrightarrow{n\to\infty} \text{meaningful thing}$$

- What's α for our magic?

- The answer is $\frac{1}{2}$

Central Limit Theorem: Start with Scaling (2)

- Reshaping the equation:

$$\frac{\sqrt{n}}{\sigma} \times (M_n - \mu) =$$

Central Limit Theorem: Start with Scaling (2)

- Reshaping the equation:

$$\frac{\sqrt{n}}{\sigma} \times (M_n - \mu) = \sqrt{n}\left(\frac{S_n - n\mu}{\sigma n}\right) = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Central Limit Theorem: Start with Scaling (2)

- Reshaping the equation:

$$\frac{\sqrt{n}}{\sigma} \times (M_n - \mu) = \sqrt{n}\left(\frac{S_n - n\mu}{\sigma n}\right) = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

- Let $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then, $\mathbb{E}[Z_n] = 0$ and $\text{var}(Z_n) = 1$.

Central Limit Theorem: Start with Scaling (2)

- Reshaping the equation:

$$\frac{\sqrt{n}}{\sigma} \times (M_n - \mu) = \sqrt{n}\left(\frac{S_n - n\mu}{\sigma n}\right) = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

- Let $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then, $\mathbb{E}[Z_n] = 0$ and $\mathrm{var}(Z_n) = 1$.
- Z_n is well-centered with a variance irrespective of n.

Central Limit Theorem: Start with Scaling (2)

- Reshaping the equation:

$$\frac{\sqrt{n}}{\sigma} \times (M_n - \mu) = \sqrt{n}\left(\frac{S_n - n\mu}{\sigma n}\right) = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

- Let $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then, $\mathbb{E}[Z_n] = 0$ and $\text{var}(Z_n) = 1$.
- Z_n is well-centered with a variance irrespective of n.
- We expect that Z_n converges to something meaningful as $n \to \infty$, but what?

Central Limit Theorem: Start with Scaling (2)

- Reshaping the equation:

$$\frac{\sqrt{n}}{\sigma} \times (M_n - \mu) = \sqrt{n}\left(\frac{S_n - n\mu}{\sigma n}\right) = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

- Let $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then, $\mathbb{E}[Z_n] = 0$ and $\mathrm{var}(Z_n) = 1$.
- Z_n is well-centered with a variance irrespective of n.
- We expect that Z_n converges to something meaningful as $n \to \infty$, but what?
- Some deterministic number just like WLLG?

Central Limit Theorem: Start with Scaling (2)

- Reshaping the equation:

$$\frac{\sqrt{n}}{\sigma} \times (M_n - \mu) = \sqrt{n}\left(\frac{S_n - n\mu}{\sigma n}\right) = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

- Let $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then, $\mathbb{E}[Z_n] = 0$ and $\mathrm{var}(Z_n) = 1$.
- Z_n is well-centered with a variance irrespective of n.
- We expect that Z_n converges to something meaningful as $n \to \infty$, but what?
- Some deterministic number just like WLLG?
- Interestingly, it converges to some well-known random variable.

Central Limit Theorem: Start with Scaling (2)

- Reshaping the equation:

$$\frac{\sqrt{n}}{\sigma} \times (M_n - \mu) = \sqrt{n}\left(\frac{S_n - n\mu}{\sigma n}\right) = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

- Let $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then, $\mathbb{E}[Z_n] = 0$ and $\mathrm{var}(Z_n) = 1$.
- Z_n is well-centered with a variance irrespective of n.
- We expect that Z_n converges to something meaningful as $n \to \infty$, but what?
- Some deterministic number just like WLLG?
- Interestingly, it converges to some well-known random variable.
  - Need a new concept of convergence: "convergence in distribution"

Convergence in Distribution

- Consider a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ and a rv Y.

Convergence in Distribution: $Y_n \xrightarrow{\text{in dist.}} Y$

- Consider a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ and a rv Y.

Convergence in Distribution: $Y_n \xrightarrow{\text{in dist.}} Y$

For every y,

$$\mathbb{P}(Y_n \le y) \xrightarrow{n\to\infty} \mathbb{P}(Y \le y)$$

- Consider a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ and a rv Y.

Convergence in Distribution: $Y_n \xrightarrow{\text{in dist.}} Y$

For every y,

$$\mathbb{P}(Y_n \le y) \xrightarrow{n\to\infty} \mathbb{P}(Y \le y)$$

- Another type of convergence of rvs

KAIST EE

- Consider a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ and a rv Y.

Convergence in Distribution: $Y_n \xrightarrow{\text{in dist.}} Y$

For every y,

$$\mathbb{P}(Y_n \leq y) \xrightarrow{n\to\infty} \mathbb{P}(Y \leq y)$$

- Another type of convergence of rvs
- Comparison with convergence in probability?

- Consider a sequence of rvs $(Y_n)_{n=1,2,\ldots}$ and a rv Y.

Convergence in Distribution: $Y_n \xrightarrow{\text{in dist.}} Y$

For every y,

$$\mathbb{P}(Y_n \le y) \xrightarrow{n\to\infty} \mathbb{P}(Y \le y)$$

- Another type of convergence of rvs
- Comparison with convergence in probability?
  - Convergence in probability $\implies$ Convergence in distribution, but the reverse is not true.
  - The proof is beyond what this class covers, but it will be interesting to find an example that shows convergence in distribution, which is not convergence in probability.

Example: in Distribution, but not in Probability

- $X_n \sim \text{Bernoulli}(1/2)$, for all $n \geq 1$.

Example: in Distribution, but not in Probability

- $X_n \sim \text{Bernoulli}(1/2)$, for all $n \geq 1$.
- $X = 1 - X_n$.

Example: in Distribution, but not in Probability

- $X_n \sim$ Bernoulli$(1/2)$, for all $n \geq 1$.
- $X = 1 - X_n$.
- Note that $X \sim$ Bernoulli$(1/2)$. It means that the distributions of X_n and X are equal. It is trivial that X_n converges to X in distribution.

Example: in Distribution, but not in Probability

- $X_n \sim \text{Bernoulli}(1/2)$, for all $n \geq 1$.
- $X = 1 - X_n$.
- Note that $X \sim \text{Bernoulli}(1/2)$. It means that the distributions of X_n and X are equal. It is trivial that X_n converges to X in distribution.
- What about convergence in probability?

$$\begin{aligned}\mathbb{P}(|X_n - X| \geq \epsilon) &= \mathbb{P}(|X_n - 1 + X_n| \geq \epsilon) = \mathbb{P}(|2X_n - 1| \geq \epsilon)\\ &= \mathbb{P}(1 \geq \epsilon) \qquad (\text{because } |2X_n - 1| = 1)\end{aligned}$$

Example: in Distribution, but not in Probability

- $X_n \sim \text{Bernoulli}(1/2)$, for all $n \geq 1$.
- $X = 1 - X_n$.
- Note that $X \sim \text{Bernoulli}(1/2)$. It means that the distributions of X_n and X are equal. It is trivial that X_n converges to X in distribution.
- What about convergence in probability?

$$\begin{aligned}\mathbb{P}(|X_n - X| \geq \epsilon) &= \mathbb{P}(|X_n - 1 + X_n| \geq \epsilon) = \mathbb{P}(|2X_n - 1| \geq \epsilon) \\ &= \mathbb{P}(1 \geq \epsilon) \qquad \text{(because } |2X_n - 1| = 1)\end{aligned}$$

- We can find ϵ small enough so that the above does not go to zero.

- $S_n = X_1 + X_2 + \cdots + X_n, \qquad Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$

Central Limit Theorem

Z_n convergens to Z in distribution, where $Z \sim \mathcal{N}(0, 1)$.

KAIST EE

- $S_n = X_1 + X_2 + \cdots + X_n, \qquad Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$

Central Limit Theorem

Z_n convergens to Z in distribution, where $Z \sim \mathcal{N}(0,1)$.

- Very surprising!
- Irrespecitive of the distribution of X_i, Z is normal.

- For simplicity, assume that $\mathbb{E}(X_i) = 0$ and $\text{var}(X_i) = 1$, $i = 1, 2, \ldots, n$

LLG vs. CLT: Different Scaling Glasses

- For simplicity, assume that $\mathbb{E}(X_i) = 0$ and $\text{var}(X_i) = 1$, $i = 1, 2, \ldots, n$
- Law of Large Numbers

> Scaling S_n by $1/n$, you go to a deterministic world.

LLG vs. CLT: Different Scaling Glasses

- For simplicity, assume that $\mathbb{E}(X_i) = 0$ and $\mathrm{var}(X_i) = 1$, $i = 1, 2, \ldots, n$
- Law of Large Numbers

> Scaling S_n by $1/n$, you go to a deterministic world.

- Central Limit Theorem

> Scaling S_n by $1/\sqrt{n}$, you still stay at the random world, but not an arbitrary random world. That's the normal random world, not depending on the distribution of each X_i.

Practical Use of CLT

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}, \qquad \mathbb{P}(Z_n \le z) \xrightarrow{n\to\infty} \mathbb{P}(Z \le z),\ Z \sim \mathcal{N}(0,1)$$

[1]Only unique mode. A single maximum or minimum.

Practical Use of CLT

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}, \qquad \mathbb{P}(Z_n \le z) \xrightarrow{n\to\infty} \mathbb{P}(Z \le z),\ Z \sim \mathcal{N}(0,1)$$

- Can approximate Z_n with a standard normal rv

[1] Only unique mode. A single maximum or minimum.

Practical Use of CLT

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}, \qquad \mathbb{P}(Z_n \le z) \xrightarrow{n\to\infty} \mathbb{P}(Z \le z),\ Z \sim \mathcal{N}(0,1)$$

- Can approximate Z_n with a standard normal rv

[1] Only unique mode. A single maximum or minimum.

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}, \qquad \mathbb{P}(Z_n \le z) \xrightarrow{n\to\infty} \mathbb{P}(Z \le z),\ Z \sim \mathcal{N}(0,1)$$

- Can approximate Z_n with a standard normal rv
- Can approximate S_n with a normal rv $\sim (n\mu, n\sigma^2)$
  - $S_n = n\mu + Z_n\sigma\sqrt{n}$

[1] Only unique mode. A single maximum or minimum.

Practical Use of CLT

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}, \qquad \mathbb{P}(Z_n \le z) \xrightarrow{n\to\infty} \mathbb{P}(Z \le z),\ Z \sim \mathcal{N}(0,1)$$

- Can approximate Z_n with a standard normal rv
- Can approximate S_n with a normal rv $\sim (n\mu, n\sigma^2)$
  - $S_n = n\mu + Z_n\sigma\sqrt{n}$
- How large should n be?

[1] Only unique mode. A single maximum or minimum.

Practical Use of CLT

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}, \qquad \mathbb{P}(Z_n \le z) \xrightarrow{n\to\infty} \mathbb{P}(Z \le z),\ Z \sim \mathcal{N}(0,1)$$

- Can approximate Z_n with a standard normal rv
- Can approximate S_n with a normal rv $\sim (n\mu, n\sigma^2)$
  - $S_n = n\mu + Z_n\sigma\sqrt{n}$
- How large should n be?
  - A moderate n (20 or 30) usually works, which is the power of CLT.

[1]Only unique mode. A single maximum or minimum.

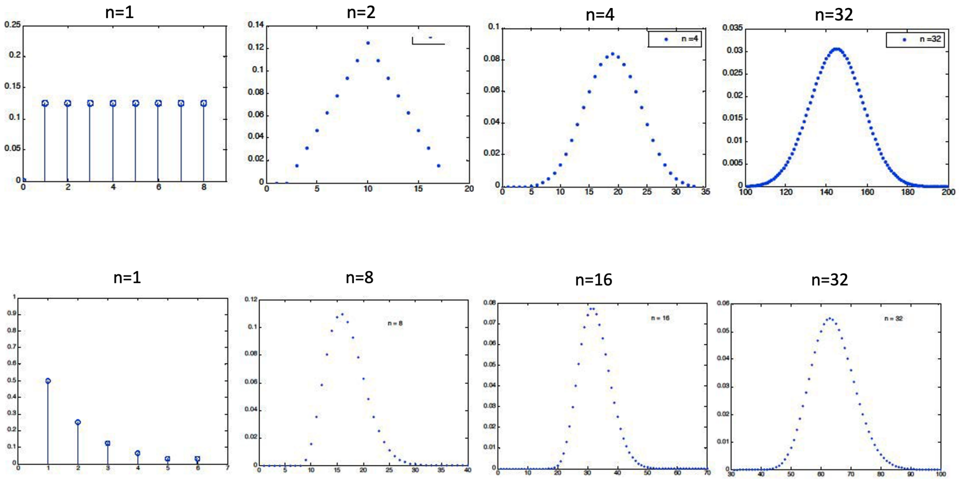
Practical Use of CLT

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}, \qquad \mathbb{P}(Z_n \le z) \xrightarrow{n\to\infty} \mathbb{P}(Z \le z),\ Z \sim \mathcal{N}(0,1)$$

- Can approximate Z_n with a standard normal rv
- Can approximate S_n with a normal rv $\sim (n\mu, n\sigma^2)$
  - $S_n = n\mu + Z_n\sigma\sqrt{n}$
- How large should n be?
  - A moderate n (20 or 30) usually works, which is the power of CLT.
  - If X_i resembles a normal rv more, smaller n works: symmetry and unimodality[1]

[1] Only unique mode. A single maximum or minimum.

CLT: Examples of Required n



$\mathbb{P}(S_n \leq a) \approx b$: Given two parameters, find the third

- Package weights X_i: iid exponential $\lambda = 1/2$ ($\mu = 1/\lambda = 2$ and $\sigma^2 = 1/\lambda^2 = 4$)

Examples of CLT (1)

$\mathbb{P}(S_n \leq a) \approx b$: Given two parameters, find the third

- Package weights X_i: iid exponential $\lambda = 1/2$ ($\mu = 1/\lambda = 2$ and $\sigma^2 = 1/\lambda^2 = 4$)
- Load container with $n = 100$ packages

$$\mathbb{P}(S_{100} \geq 210) = \mathbb{P}\left[\frac{S_{100} - 100 \cdot 2}{2\sqrt{100}} \geq \frac{210 - 200}{20}\right] = \mathbb{P}(Z_{100} \geq 0.5)$$

Examples of CLT (1)

$\mathbb{P}(S_n \le a) \approx b$: Given two parameters, find the third

- Package weights X_i: iid exponential $\lambda = 1/2$ ($\mu = 1/\lambda = 2$ and $\sigma^2 = 1/\lambda^2 = 4$)
- Load container with $n = 100$ packages

$$\mathbb{P}(S_{100} \ge 210) = \mathbb{P}\left[\frac{S_{100} - 100 \cdot 2}{2\sqrt{100}} \ge \frac{210 - 200}{20}\right] = \mathbb{P}(Z_{100} \ge 0.5)$$
$$\approx \mathbb{P}(Z \ge 0.5) = 1 - \mathbb{P}(Z \le 0.5) = 1 - \Phi(0.5)$$

Examples of CLT (2)

$\mathbb{P}(S_n \le a) \approx b$: Given two parameters, find the third

- Package weights X_i: iid exponential $\lambda = 1/2$ ($\mu = 1/\lambda = 2$ and $\sigma^2 = 1/\lambda^2 = 4$)

Examples of CLT (2)

$\mathbb{P}(S_n \leq a) \approx b$: Given two parameters, find the third

- Package weights X_i: iid exponential $\lambda = 1/2$ ($\mu = 1/\lambda = 2$ and $\sigma^2 = 1/\lambda^2 = 4$)
- $n = 100$ packages, and choose the "capacity" a, so that $\mathbb{P}(S_n \geq a) \approx 0.05$

$$\mathbb{P}(S_{100} \geq a) = \mathbb{P}\left[\frac{S_{100} - 100 \cdot 2}{2\sqrt{100}} \geq \frac{a - 200}{20}\right] = \mathbb{P}(Z_{100} \geq \frac{a-200}{20})$$

KAIST EE

$\mathbb{P}(S_n \le a) \approx b$: Given two parameters, find the third

- Package weights X_i: iid exponential $\lambda = 1/2$ ($\mu = 1/\lambda = 2$ and $\sigma^2 = 1/\lambda^2 = 4$)
- $n = 100$ packages, and choose the "capacity" a, so that $\mathbb{P}(S_n \ge a) \approx 0.05$

$$\mathbb{P}(S_{100} \ge a) = \mathbb{P}\left[\frac{S_{100} - 100 \cdot 2}{2\sqrt{100}} \ge \frac{a-200}{20}\right] = \mathbb{P}(Z_{100} \ge \frac{a-200}{20})$$

$$\approx \mathbb{P}(Z \ge \frac{a-200}{20}) = 1 - \mathbb{P}(Z \le \frac{a-200}{20}) = 1 - \Phi(\frac{a-200}{20}) = 0.05$$

Examples of CLT (2)

$\mathbb{P}(S_n \leq a) \approx b$: Given two parameters, find the third

- Package weights X_i: iid exponential $\lambda = 1/2$ ($\mu = 1/\lambda = 2$ and $\sigma^2 = 1/\lambda^2 = 4$)
- $n = 100$ packages, and choose the "capacity" a, so that $\mathbb{P}(S_n \geq a) \approx 0.05$

$$\mathbb{P}(S_{100} \geq a) = \mathbb{P}\left[\frac{S_{100} - 100 \cdot 2}{2\sqrt{100}} \geq \frac{a-200}{20}\right] = \mathbb{P}(Z_{100} \geq \frac{a-200}{20})$$

$$\approx \mathbb{P}(Z \geq \frac{a-200}{20}) = 1 - \mathbb{P}(Z \leq \frac{a-200}{20}) = 1 - \Phi(\frac{a-200}{20}) = 0.05$$

- The value of a such that $\Phi(\frac{a-200}{20}) = 0.95$? $\frac{a-200}{20} = 1.645$ and $a = 232.9$

Examples of CLT (3)

$\mathbb{P}(S_n \leq a) \approx b$: Given two parameters, find the third

- Package weights X_i: iid exponential $\lambda = 1/2$ ($\mu = 1/\lambda = 2$ and $\sigma^2 = 1/\lambda^2 = 4$)

Examples of CLT (3)

$\mathbb{P}(S_n \leq a) \approx b$: Given two parameters, find the third

- Package weights X_i: iid exponential $\lambda = 1/2$ ($\mu = 1/\lambda = 2$ and $\sigma^2 = 1/\lambda^2 = 4$)
- How large n, so that $\mathbb{P}(S_n \geq 210) \approx 0.05$?

$$\mathbb{P}(S_n \geq 210) = \mathbb{P}\left[\frac{S_n - 2n}{2\sqrt{n}} \geq \frac{210 - 2n}{2\sqrt{n}}\right]$$

$\mathbb{P}(S_n \leq a) \approx b$: Given two parameters, find the third

- Package weights X_i: iid exponential $\lambda = 1/2$ ($\mu = 1/\lambda = 2$ and $\sigma^2 = 1/\lambda^2 = 4$)
- How large n, so that $\mathbb{P}(S_n \geq 210) \approx 0.05$?

$$\mathbb{P}(S_n \geq 210) = \mathbb{P}\left[\frac{S_n - 2n}{2\sqrt{n}} \geq \frac{210 - 2n}{2\sqrt{n}}\right] \approx 1 - \Phi(\frac{210 - 2n}{2\sqrt{n}}) = 0.05$$

Examples of CLT (3)

$\mathbb{P}(S_n \le a) \approx b$: Given two parameters, find the third

- Package weights X_i: iid exponential $\lambda = 1/2$ ($\mu = 1/\lambda = 2$ and $\sigma^2 = 1/\lambda^2 = 4$)
- How large n, so that $\mathbb{P}(S_n \ge 210) \approx 0.05$?

$$\mathbb{P}(S_n \ge 210) = \mathbb{P}\Big[\frac{S_n - 2n}{2\sqrt{n}} \ge \frac{210 - 2n}{2\sqrt{n}}\Big] \approx 1 - \Phi(\frac{210 - 2n}{2\sqrt{n}}) = 0.05$$

- The value of n such that $\frac{210-2n}{2\sqrt{n}} = 1.645$? $n = 89$

(1) Weak Law of Large Numbers: Result and Meaning

(2) Central Limit Theorem: Result and Meaning

(3) Weak Law of Large Numbers: Proof
- Inequalities: Markov and Chebyshev

(4) Central Limit Theorem: Proof
- Moment Generating Function (MGF)

- (Q) Knowing $\mathbb{E}(X)$, can we say something about the distribution of X?

Markov Inequality

- (Q) Knowing $\mathbb{E}(X)$, can we say something about the distribution of X?
- Intuition: small $\mathbb{E}(X) \Longrightarrow$ small $\mathbb{P}(X \geq a)$

$\square$

- (Q) Knowing $\mathbb{E}(X)$, can we say something about the distribution of X?
- Intuition: small $\mathbb{E}(X) \implies$ small $\mathbb{P}(X \geq a)$

Markov Inequality

If $X \geq 0$ and $a > 0$, then $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$.

$\square$

KAIST EE

- (Q) Knowing $\mathbb{E}(X)$, can we say something about the distribution of X?
- Intuition: small $\mathbb{E}(X) \implies$ small $\mathbb{P}(X \geq a)$

Markov Inequality

If $X \geq 0$ and $a > 0$, then $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$.

Proof. For any $a > 0$, define Y_a as:

$$Y_a \triangleq \begin{cases} 0, & \text{if } X < a, \\ a, & \text{if } X \geq a \end{cases}$$

$\square$

- (Q) Knowing $\mathbb{E}(X)$, can we say something about the distribution of X?
- Intuition: small $\mathbb{E}(X) \implies$ small $\mathbb{P}(X \geq a)$

Markov Inequality

If $X \geq 0$ and $a > 0$, then $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$.

Proof. For any $a > 0$, define Y_a as:

$$Y_a \triangleq \begin{cases} 0, & \text{if } X < a, \\ a, & \text{if } X \geq a \end{cases}$$

Then, using non-negativity of X, $Y_a \leq X$, which leads to $\mathbb{E}[Y_a] \leq \mathbb{E}[X]$.

$\square$

- (Q) Knowing $\mathbb{E}(X)$, can we say something about the distribution of X?
- Intuition: small $\mathbb{E}(X) \implies$ small $\mathbb{P}(X \geq a)$

Markov Inequality

If $X \geq 0$ and $a > 0$, then $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$.

Proof. For any $a > 0$, define Y_a as:

$$Y_a \triangleq \begin{cases} 0, & \text{if } X < a, \\ a, & \text{if } X \geq a \end{cases}$$

Then, using non-negativity of X, $Y_a \leq X$, which leads to $\mathbb{E}[Y_a] \leq \mathbb{E}[X]$.

Note that we have:

$$\mathbb{E}[Y_a] = a\mathbb{P}(Y_a = a) = a\mathbb{P}(X \geq a).$$

$\square$

KAIST EE

- (Q) Knowing $\mathbb{E}(X)$, can we say something about the distribution of X?
- Intuition: small $\mathbb{E}(X) \implies$ small $\mathbb{P}(X \geq a)$

Markov Inequality

If $X \geq 0$ and $a > 0$, then $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$.

Proof. For any $a > 0$, define Y_a as:

$$Y_a \triangleq \begin{cases} 0, & \text{if } X < a, \\ a, & \text{if } X \geq a \end{cases}$$

Then, using non-negativity of X, $Y_a \leq X$, which leads to $\mathbb{E}[Y_a] \leq \mathbb{E}[X]$.

Note that we have:

$$\mathbb{E}[Y_a] = a\mathbb{P}(Y_a = a) = a\mathbb{P}(X \geq a).$$

Thus, $a \cdot \mathbb{P}(X \geq a) \leq \mathbb{E}[X]$. $\square$

Chebyshev Inequality

Chebyshev Inequality

- (Q) Knowing both $\mathbb{E}(X)$ and $\text{var}(X)$, can we say something about the distribution of X?

Chebyshev Inequality

- (Q) Knowing both $\mathbb{E}(X)$ and $\text{var}(X)$, can we say something about the distribution of X?
- Intuition: small $\text{var}(X) \Longrightarrow X$ is unlikely to be too far away from its mean.

Chebyshev Inequality

- (Q) Knowing both $\mathbb{E}(X)$ and $\text{var}(X)$, can we say something about the distribution of X?
- Intuition: small $\text{var}(X) \Longrightarrow X$ is unlikely to be too far away from its mean.
- $\mathbb{E}(X) = \mu$, $\text{var}(X) = \sigma^2$.

KAIST EE

- (Q) Knowing both $\mathbb{E}(X)$ and $\text{var}(X)$, can we say something about the distribution of X?
- Intuition: small $\text{var}(X) \implies X$ is unlikely to be too far away from its mean.
- $\mathbb{E}(X) = \mu$, $\text{var}(X) = \sigma^2$.

Chebyshev Inequality

$$\mathbb{P}\left(|X - \mu| \geq c\right) \leq \frac{\sigma^2}{c^2}, \quad \text{for all } c > 0$$

- (Q) Knowing both $\mathbb{E}(X)$ and $\text{var}(X)$, can we say something about the distribution of X?
- Intuition: small $\text{var}(X) \implies X$ is unlikely to be too far away from its mean.
- $\mathbb{E}(X) = \mu$, $\text{var}(X) = \sigma^2$.

Chebyshev Inequality

$$\mathbb{P}\left(|X-\mu| \geq c\right) \leq \frac{\sigma^2}{c^2}, \quad \text{for all } c > 0$$

- Proof.

$$\mathbb{P}\left(|X-\mu| \geq c\right) = \mathbb{P}\left((X-\mu)^2 \geq c^2\right) \leq$$

- (Q) Knowing both $\mathbb{E}(X)$ and $\text{var}(X)$, can we say something about the distribution of X?
- Intuition: small $\text{var}(X) \implies X$ is unlikely to be too far away from its mean.
- $\mathbb{E}(X) = \mu$, $\text{var}(X) = \sigma^2$.

Chebyshev Inequality

$$\mathbb{P}\left(|X - \mu| \geq c\right) \leq \frac{\sigma^2}{c^2}, \quad \text{for all } c > 0$$

- Proof.

$$\mathbb{P}\left(|X - \mu| \geq c\right) = \mathbb{P}\left((X - \mu)^2 \geq c^2\right) \leq \frac{\mathbb{E}\left[(X-\mu)^2\right]}{c^2} = \frac{\text{var}(X)}{c^2}$$

Example: MI vs. CI

- $X \sim \exp(1)$. Then, $\mathbb{E}[X] = 1/\lambda = 1$ and $\text{var}[X] = 1/\lambda^2 = 1$.

Example: MI vs. CI

- $X \sim \exp(1)$. Then, $\mathbb{E}[X] = 1/\lambda = 1$ and $\mathrm{var}[X] = 1/\lambda^2 = 1$.

- Exact CCDF: $\mathbb{P}(X \geq a) = e^{-a}$

Example: MI vs. CI

- $X \sim \exp(1)$. Then, $\mathbb{E}[X] = 1/\lambda = 1$ and $\mathrm{var}[X] = 1/\lambda^2 = 1$.

- Exact CCDF: $\mathbb{P}(X \geq a) = e^{-a}$

- Markov inequality

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a} = \frac{1}{a}$$

- $X \sim \exp(1)$. Then, $\mathbb{E}[X] = 1/\lambda = 1$ and $\text{var}[X] = 1/\lambda^2 = 1$.

- Exact CCDF: $\mathbb{P}(X \geq a) = e^{-a}$

- Markov inequality

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a} = \frac{1}{a}$$

- Chebyshev inequality

$$\mathbb{P}(X \geq a) = \mathbb{P}(X - 1 \geq a - 1)$$
$$\leq \mathbb{P}(|X - 1| \geq a - 1) \leq \frac{1}{(a-1)^2}$$

- $X \sim \exp(1)$. Then, $\mathbb{E}[X] = 1/\lambda = 1$ and $\mathrm{var}[X] = 1/\lambda^2 = 1$.

- Exact CCDF: $\mathbb{P}(X \geq a) = e^{-a}$

- Markov inequality

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a} = \frac{1}{a}$$

- Chebyshev inequality

$$\begin{aligned}\mathbb{P}(X \geq a) &= \mathbb{P}(X - 1 \geq a - 1) \\ &\leq \mathbb{P}(|X - 1| \geq a - 1) \leq \frac{1}{(a-1)^2}\end{aligned}$$

- For reasonably large a, CI provides much better bound.

Example: MI vs. CI

- $X \sim \exp(1)$. Then, $\mathbb{E}[X] = 1/\lambda = 1$ and $\mathrm{var}[X] = 1/\lambda^2 = 1$.

- Exact CCDF: $\mathbb{P}(X \geq a) = e^{-a}$

- **Markov inequality**

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a} = \frac{1}{a}$$

- **Chebyshev inequality**

$$\begin{aligned}\mathbb{P}(X \geq a) &= \mathbb{P}(X - 1 \geq a - 1) \\ &\leq \mathbb{P}(|X - 1| \geq a - 1) \leq \frac{1}{(a-1)^2}\end{aligned}$$

- For reasonably large a, CI provides much better bound.

- Knowing the variance helps

Example: MI vs. CI

- $X \sim \exp(1)$. Then, $\mathbb{E}[X] = 1/\lambda = 1$ and $\text{var}[X] = 1/\lambda^2 = 1$.

- Exact CCDF: $\mathbb{P}(X \geq a) = e^{-a}$

- Markov inequality

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a} = \frac{1}{a}$$

- Chebyshev inequality

$$\mathbb{P}(X \geq a) = \mathbb{P}(X - 1 \geq a - 1)$$
$$\leq \mathbb{P}(|X - 1| \geq a - 1) \leq \frac{1}{(a-1)^2}$$

- For reasonably large a, CI provides much better bound.

- Knowing the variance helps

- Both bounds are the ones that bound the probability of rare events.

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

Weak law of large numbers

M_n converges to μ in probability.

Proof. For any given $\epsilon > 0$,

$$M_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \dots X_n}{n}$$

Weak law of large numbers

M_n converges to μ in probability.

Proof. For any given $\epsilon > 0$,

$$\mathbb{P}\left(|M_n - \mu| \geq \epsilon\right) \leq \frac{\text{var}(M_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n\to\infty} 0$$

Comparison: WLLN vs. CLT

We ask the same question, and try to answer it, using WLLN or CLT.

See how the answers becomes different.

Example: Polling using WLLN

- p: fraction of voters who support "Yung".

Example: Polling using WLLN

KAIST EE

- p: fraction of voters who support "Yung".
- Interview n randomly selected voters and record the result in $M_n = \frac{X_1+\ldots+X_n}{n}$ which is an estimate of p, where the Bernoulli rv $X_i = 1$ if i-th interviewee answers "yes", and 0 otherwise.

Example: Polling using WLLN

- p: fraction of voters who support "Yung".
- Interview n randomly selected voters and record the result in $M_n = \frac{X_1 + \ldots + X_n}{n}$ which is an estimate of p, where the Bernoulli rv $X_i = 1$ if i-th interviewee answers "yes", and 0 otherwise.
- $\mathbb{P}(|M_n - p| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$ (because $p(1-p) \leq 1/4$)

Example: Polling using WLLN

- p: fraction of voters who support "Yung".
- Interview n randomly selected voters and record the result in $M_n = \frac{X_1+\ldots+X_n}{n}$ which is an estimate of p, where the Bernoulli rv $X_i = 1$ if i-th interviewee answers "yes", and 0 otherwise.
- $\mathbb{P}(|M_n - p| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$ (because $p(1-p) \leq 1/4$)
- Question. What is n so that the probability that our estimate is incorrect by more than 0.1 is no larger than 0.25?

Example: Polling using WLLN

- p: fraction of voters who support "Yung".
- Interview n randomly selected voters and record the result in $M_n = \frac{X_1+\ldots+X_n}{n}$ which is an estimate of p, where the Bernoulli rv $X_i = 1$ if i-th interviewee answers "yes", and 0 otherwise.
- $\mathbb{P}(|M_n - p| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$ (because $p(1-p) \leq 1/4$)
- Question. What is n so that the probability that our estimate is incorrect by more than 0.1 is no larger than 0.25?
  - $\epsilon = 0.1$ and $\frac{1}{4n\epsilon^2} \leq 0.25 \implies n \geq 100$

Example: Polling using WLLN

- p: fraction of voters who support "Yung".
- Interview n randomly selected voters and record the result in $M_n = \frac{X_1 + \ldots + X_n}{n}$ which is an estimate of p, where the Bernoulli rv $X_i = 1$ if i-th interviewee answers "yes", and 0 otherwise.
- $\mathbb{P}(|M_n - p| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$ (because $p(1-p) \leq 1/4$)
- Question. What is n so that the probability that our estimate is incorrect by more than 0.1 is no larger than 0.25?
  - $\epsilon = 0.1$ and $\frac{1}{4n\epsilon^2} \leq 0.25 \implies n \geq 100$
- Question. What is n so that the probability that our estimate is incorrect by more than 0.01 is no larger than 0.05?

Example: Polling using WLLN

- p: fraction of voters who support "Yung".
- Interview n randomly selected voters and record the result in $M_n = \frac{X_1 + \ldots + X_n}{n}$ which is an estimate of p, where the Bernoulli rv $X_i = 1$ if i-th interviewee answers "yes", and 0 otherwise.
- $\mathbb{P}(|M_n - p| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$ (because $p(1-p) \leq 1/4$)
- Question. What is n so that the probability that our estimate is incorrect by more than 0.1 is no larger than 0.25?
  - $\epsilon = 0.1$ and $\frac{1}{4n\epsilon^2} \leq 0.25 \implies n \geq 100$
- Question. What is n so that the probability that our estimate is incorrect by more than 0.01 is no larger than 0.05?
  - $\epsilon = 0.01$ and $\frac{1}{4n\epsilon^2} \leq 0.05 \implies n \geq 50000$

Example: Polling using CLT

$$\mathbb{P}(|M_n - p| \geq \epsilon) = \qquad\qquad = $$

$$\leq \qquad\qquad = $$

Example: Polling using CLT

$$\mathbb{P}(|M_n - p| \geq \epsilon) = \mathbb{P}\left[\left|\frac{S_n - np}{n}\right| \geq \epsilon\right] =$$

$$\leq \qquad\qquad\qquad =$$

Example: Polling using CLT

$$\mathbb{P}(|M_n - p| \geq \epsilon) = \mathbb{P}\left[\left|\frac{S_n - np}{n}\right| \geq \epsilon\right] = \mathbb{P}\left[\left|\frac{S_n - np}{\sigma\sqrt{n}}\right| \geq \frac{\epsilon\sqrt{n}}{\sigma}\right]$$

$$\leq \qquad\qquad\qquad =$$

$$\mathbb{P}(|M_n - p| \geq \epsilon) = \mathbb{P}\left[\left|\frac{S_n - np}{n}\right| \geq \epsilon\right] = \mathbb{P}\left[\left|\frac{S_n - np}{\sigma\sqrt{n}}\right| \geq \frac{\epsilon\sqrt{n}}{\sigma}\right]$$

$$\leq \mathbb{P}\left[\left|\frac{S_n - np}{\sigma\sqrt{n}}\right| \geq 2\epsilon\sqrt{n}\right] = \qquad \text{(because } \sigma = \sqrt{p(1-p)} \leq 1/2\text{)}$$

Example: Polling using CLT

$$\mathbb{P}(|M_n - p| \geq \epsilon) = \mathbb{P}\left[\left|\frac{S_n - np}{n}\right| \geq \epsilon\right] = \mathbb{P}\left[\left|\frac{S_n - np}{\sigma\sqrt{n}}\right| \geq \frac{\epsilon\sqrt{n}}{\sigma}\right]$$

$$\leq \mathbb{P}\left[\left|\frac{S_n - np}{\sigma\sqrt{n}}\right| \geq 2\epsilon\sqrt{n}\right] = 2\Big(1 - \Phi(2\epsilon\sqrt{n})\Big) \text{ (because } \sigma = \sqrt{p(1-p)} \leq 1/2\text{)}$$

KAIST EE

$$\mathbb{P}(|M_n - p| \geq \epsilon) = \mathbb{P}\left[\left|\frac{S_n - np}{n}\right| \geq \epsilon\right] = \mathbb{P}\left[\left|\frac{S_n - np}{\sigma\sqrt{n}}\right| \geq \frac{\epsilon\sqrt{n}}{\sigma}\right]$$

$$\leq \mathbb{P}\left[\left|\frac{S_n - np}{\sigma\sqrt{n}}\right| \geq 2\epsilon\sqrt{n}\right] = 2\Big(1 - \Phi(2\epsilon\sqrt{n})\Big) \text{ (because } \sigma = \sqrt{p(1-p)} \leq 1/2)$$

- Question. What is n so that the probability that our estimate is incorrect by more than 0.01 is no larger than 0.05?

Example: Polling using CLT

$$\mathbb{P}(|M_n - p| \geq \epsilon) = \mathbb{P}\left[\left|\frac{S_n - np}{n}\right| \geq \epsilon\right] = \mathbb{P}\left[\left|\frac{S_n - np}{\sigma\sqrt{n}}\right| \geq \frac{\epsilon\sqrt{n}}{\sigma}\right]$$

$$\leq \mathbb{P}\left[\left|\frac{S_n - np}{\sigma\sqrt{n}}\right| \geq 2\epsilon\sqrt{n}\right] = 2\Big(1 - \Phi(2\epsilon\sqrt{n})\Big) \text{ (because } \sigma = \sqrt{p(1-p)} \leq 1/2)$$

- Question. What is n so that the probability that our estimate is incorrect by more than 0.01 is no larger than 0.05?
  - $\epsilon = 0.01$ and $2\Big(1 - \Phi(2\epsilon\sqrt{n})\Big) = 0.05$, i.e., $\Phi(2\epsilon\sqrt{n}) = 0.975 \implies 2 \times 0.01 \times \sqrt{n} = 1.96$ and thus $n = 9604$

Example: Polling using CLT

$$\mathbb{P}(|M_n - p| \geq \epsilon) = \mathbb{P}\left[\left|\frac{S_n - np}{n}\right| \geq \epsilon\right] = \mathbb{P}\left[\left|\frac{S_n - np}{\sigma\sqrt{n}}\right| \geq \frac{\epsilon\sqrt{n}}{\sigma}\right]$$

$$\leq \mathbb{P}\left[\left|\frac{S_n - np}{\sigma\sqrt{n}}\right| \geq 2\epsilon\sqrt{n}\right] = 2\Big(1 - \Phi(2\epsilon\sqrt{n})\Big) \text{ (because } \sigma = \sqrt{p(1-p)} \leq 1/2)$$

- Question. What is n so that the probability that our estimate is incorrect by more than 0.01 is no larger than 0.05?
  - $\epsilon = 0.01$ and $2\Big(1 - \Phi(2\epsilon\sqrt{n})\Big) = 0.05$, i.e., $\Phi(2\epsilon\sqrt{n}) = 0.975 \implies$ $2 \times 0.01 \times \sqrt{n} = 1.96$ and thus $n = 9604$
- Compare: 50,000 from LLN vs. 9604 from CLT

Roadmap

(1) Weak Law of Large Numbers: Result and Meaning

(2) Central Limit Theorem: Result and Meaning

(3) Weak Law of Large Numbers: Proof
- Inequalities: Markov and Chebyshev

(4) Central Limit Theorem: Proof
- Moment Generating Function (MGF)

Moment Generating Function (MGF)

- For a rv X, it is a kind of transform

Moment Generating Function (MGF)

- For a rv X, it is a kind of transform
- The moment generating function (MGF) $M_X(s)$ of a rv X is a function of a scalar parameter s, defined by:

$$\boxed{M_X(s) = \mathbb{E}[e^{sX}]}$$

Moment Generating Function (MGF)

- For a rv X, it is a kind of transform
- The moment generating function (MGF) $M_X(s)$ of a rv X is a function of a scalar parameter s, defined by:

$$\boxed{M_X(s) = \mathbb{E}[e^{sX}]}$$

$$M(s) = \sum_x e^{sx} p_X(x) \quad \text{(discrete)}$$

$$M(s) = \int e^{sx} f_X(x) dx \quad \text{(continuous)}$$

Moment Generating Function (MGF)

- For a rv X, it is a kind of transform
- The moment generating function (MGF) $M_X(s)$ of a rv X is a function of a scalar parameter s, defined by:

$$\boxed{M_X(s) = \mathbb{E}[e^{sX}]}$$

$$M(s) = \sum_x e^{sx} p_X(x) \quad \text{(discrete)}$$

$$M(s) = \int e^{sx} f_X(x) dx \quad \text{(continuous)}$$

- If the context is clear, we omit X and use just $M(s)$.

Ex1) Let $p_X(x)$ is given as:

$$p_X(x) = \begin{cases} 1/2, & \text{if } x = 2 \\ 1/6, & \text{if } x = 3 \\ 1/3, & \text{if } x = 5 \end{cases}$$

$$M(s) = \mathbb{E}(e^{sX}) = \frac{1}{2}e^{2s} + \frac{1}{6}e^{3s} + \frac{1}{3}e^{5s}$$

Ex1) Let $p_X(x)$ is given as:

$$p_X(x) = \begin{cases} 1/2, & \text{if } x = 2 \\ 1/6, & \text{if } x = 3 \\ 1/3, & \text{if } x = 5 \end{cases}$$

$$M(s) = \mathbb{E}(e^{sX}) = \frac{1}{2}e^{2s} + \frac{1}{6}e^{3s} + \frac{1}{3}e^{5s}$$

Ex2) $X \sim \exp(\lambda),\ f_X(x) = \lambda e^{-\lambda x}, x \geq 0$

$$M(s) = \lambda \int_0^\infty e^{sx} e^{-\lambda x} dx$$

$$= \lambda \frac{e^{(s-\lambda)x}}{s-\lambda}\bigg|_0^\infty \quad (\text{if } s < \lambda) = \frac{\lambda}{\lambda - s}$$

Examples

Ex1) Let $p_X(x)$ is given as:

$$p_X(x) = \begin{cases} 1/2, & \text{if } x = 2 \\ 1/6, & \text{if } x = 3 \\ 1/3, & \text{if } x = 5 \end{cases}$$

$$M(s) = \mathbb{E}(e^{sX}) = \tfrac{1}{2}e^{2s} + \tfrac{1}{6}e^{3s} + \tfrac{1}{3}e^{5s}$$

Ex2) $X \sim \exp(\lambda),\ f_X(x) = \lambda e^{-\lambda x}, x \geq 0$

$$M(s) = \lambda \int_0^\infty e^{sx} e^{-\lambda x} dx$$

$$= \lambda \frac{e^{(s-\lambda)x}}{s-\lambda}\Big|_0^\infty \quad (\text{if } s < \lambda) = \frac{\lambda}{\lambda - s}$$

Ex3) Let a rv $Y = aX + b$.

$$M_Y(s) = \mathbb{E}(e^{sY}) = \mathbb{E}(e^{s(aX+b)})$$

$$= e^{sb}\mathbb{E}(e^{saX}) = e^{sb} M_X(sa)$$

Ex1) Let $p_X(x)$ is given as:

$$p_X(x) = \begin{cases} 1/2, & \text{if } x = 2 \\ 1/6, & \text{if } x = 3 \\ 1/3, & \text{if } x = 5 \end{cases}$$

$$M(s) = \mathbb{E}(e^{sX}) = \tfrac{1}{2}e^{2s} + \tfrac{1}{6}e^{3s} + \tfrac{1}{3}e^{5s}$$

Ex2) $X \sim \exp(\lambda)$, $f_X(x) = \lambda e^{-\lambda x}, x \geq 0$

$$\begin{aligned} M(s) &= \lambda \int_0^\infty e^{sx} e^{-\lambda x} dx \\ &= \lambda \frac{e^{(s-\lambda)x}}{s-\lambda}\Bigg|_0^\infty \quad (\text{if } s < \lambda) = \frac{\lambda}{\lambda - s} \end{aligned}$$

Ex3) Let a rv $Y = aX + b$.

$$\begin{aligned} M_Y(s) &= \mathbb{E}(e^{sY}) = \mathbb{E}(e^{s(aX+b)}) \\ &= e^{sb}\mathbb{E}(e^{saX}) = e^{sb} M_X(sa) \end{aligned}$$

Ex4) $X \sim \mathcal{N}(0,1)$

$$\begin{aligned} M(s) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} e^{sy} dy + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}+sy} dy \\ &= e^{\frac{s^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y-s)^2} dy \\ &= e^{s^2/2} \ (\text{because it is the pdf of } \mathcal{N}(s,1)) \end{aligned}$$

Ex1) Let $p_X(x)$ is given as:

$$p_X(x) = \begin{cases} 1/2, & \text{if } x = 2 \\ 1/6, & \text{if } x = 3 \\ 1/3, & \text{if } x = 5 \end{cases}$$

$$M(s) = \mathbb{E}(e^{sX}) = \tfrac{1}{2}e^{2s} + \tfrac{1}{6}e^{3s} + \tfrac{1}{3}e^{5s}$$

Ex2) $X \sim \exp(\lambda),\ f_X(x) = \lambda e^{-\lambda x}, x \geq 0$

$$M(s) = \lambda \int_0^{\infty} e^{sx} e^{-\lambda x} dx$$
$$= \lambda \frac{e^{(s-\lambda)x}}{s-\lambda}\Bigg|_0^{\infty} \quad (\text{if } s < \lambda) = \frac{\lambda}{\lambda - s}$$

Ex3) Let a rv $Y = aX + b$.

$$M_Y(s) = \mathbb{E}(e^{sY}) = \mathbb{E}(e^{s(aX+b)})$$
$$= e^{sb}\mathbb{E}(e^{saX}) = e^{sb}M_X(sa)$$

Ex4) $X \sim \mathcal{N}(0,1)$

$$M(s) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} e^{sy} dy + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}+sy} dy$$
$$= e^{\frac{s^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y-s)^2} dy$$
$$= e^{s^2/2} \text{ (because it is the pdf of } \mathcal{N}(s,1))$$

- Question. MGF of $\mathcal{N}(\mu, \sigma^2)$?

Useful Properties of MGF

1. $M'(0) = \mathbb{E}[X]$

KAIST EE

1. $M'(0) = \mathbb{E}[X]$

$$\frac{d}{ds}M(s) = \frac{d}{ds}\int_{-\infty}^{\infty} e^{sx} f_X(x)dx = \int_{-\infty}^{\infty} \frac{d}{ds}e^{sx} f_X(x)dx = \int_{-\infty}^{\infty} xe^{sx} f_X(x)dx$$

$$= \frac{d}{ds}M(s)\bigg|_{s=0} = \mathbb{E}[X]$$

Useful Properties of MGF

1. $M'(0) = \mathbb{E}[X]$

$$\frac{d}{ds}M(s) = \frac{d}{ds}\int_{-\infty}^{\infty} e^{sx} f_X(x)dx = \int_{-\infty}^{\infty} \frac{d}{ds}e^{sx} f_X(x)dx = \int_{-\infty}^{\infty} xe^{sx} f_X(x)dx$$

$$= \frac{d}{ds}M(s)\bigg|_{s=0} = \mathbb{E}[X]$$

2. Similarly, $M''(0) = \mathbb{E}[X^2]$

1. $M'(0) = \mathbb{E}[X]$

$$\frac{d}{ds}M(s) = \frac{d}{ds}\int_{-\infty}^{\infty} e^{sx} f_X(x)dx = \int_{-\infty}^{\infty} \frac{d}{ds} e^{sx} f_X(x)dx = \int_{-\infty}^{\infty} xe^{sx} f_X(x)dx$$

$$= \frac{d}{ds}M(s)\bigg|_{s=0} = \mathbb{E}[X]$$

2. Similarly, $M''(0) = \mathbb{E}[X^2]$

3. $\frac{d^n}{ds^n}M(s)\bigg|_{s=0} = \mathbb{E}[X^n]$

Useful Properties of MGF

1. $M'(0) = \mathbb{E}[X]$

$$\frac{d}{ds}M(s) = \frac{d}{ds}\int_{-\infty}^{\infty} e^{sx} f_X(x)dx = \int_{-\infty}^{\infty} \frac{d}{ds} e^{sx} f_X(x)dx = \int_{-\infty}^{\infty} x e^{sx} f_X(x)dx$$

$$= \frac{d}{ds}M(s)\Big|_{s=0} = \mathbb{E}[X]$$

2. Similarly, $M''(0) = \mathbb{E}[X^2]$

3. $\frac{d^n}{ds^n}M(s)\Big|_{s=0} = \mathbb{E}[X^n]$

4. MGF provides a convenient way of generating moments. That's why it is called moment generating function.

Example

- Exponential rv with parameter λ. We know that $\mathbb{E}(X) = 1/\lambda$ and $\mathsf{var}(X) = 1/\lambda^2$, which we will compute using the MGF.

Example

- Exponential rv with parameter λ. We know that $\mathbb{E}(X) = 1/\lambda$ and $\text{var}(X) = 1/\lambda^2$, which we will compute using the MGF.
- Remind: $M(s) = \frac{\lambda}{\lambda - s}$

Example

- Exponential rv with parameter λ. We know that $\mathbb{E}(X) = 1/\lambda$ and $\text{var}(X) = 1/\lambda^2$, which we will compute using the MGF.
- Remind: $M(s) = \frac{\lambda}{\lambda - s}$
- The first and the second moments are:

$$M'(s) = \frac{\lambda}{(\lambda - s)^2} \quad \rightarrow \quad \mathbb{E}(X) = M'(0) = 1/\lambda$$

$$M''(s) = \frac{2\lambda}{(\lambda - s)^3} \quad \rightarrow \quad \mathbb{E}(X^2) = M''(0) = 2/\lambda^2$$

Example

- Exponential rv with parameter λ. We know that $\mathbb{E}(X) = 1/\lambda$ and $\mathsf{var}(X) = 1/\lambda^2$, which we will compute using the MGF.
- Remind: $M(s) = \frac{\lambda}{\lambda - s}$
- The first and the second moments are:

$$M'(s) = \frac{\lambda}{(\lambda - s)^2} \quad \rightarrow \quad \mathbb{E}(X) = M'(0) = 1/\lambda$$
$$M''(s) = \frac{2\lambda}{(\lambda - s)^3} \quad \rightarrow \quad \mathbb{E}(X^2) = M''(0) = 2/\lambda^2$$

- Thus, $\mathsf{var}(X) = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$

Inversion Property

Inversion Property

The transform $M_X(s)$ associated with a random variable X uniquely determines the CDF of X, assuming that $M_X(s)$ is finite for all s in some interval $[-a, a]$, where a is a positive number.

- In easy words, we can recover the distribution if we know the MGF.
- Thus, each rv has its own unique MGF.

Example 1

- Given the following MGF of rv X, what is the distribution of X?

$$M(s) = \frac{1}{4}e^{-s} + \frac{1}{2} + \frac{1}{8}e^{4s} + \frac{1}{8}e^{5s}$$

Example 1

- Given the following MGF of rv X, what is the distribution of X?

$$M(s) = \frac{1}{4}e^{-s} + \frac{1}{2} + \frac{1}{8}e^{4s} + \frac{1}{8}e^{5s}$$

- Note that $M(s) = \sum_x e^{sx} p_X(x)$

Example 1

- Given the following MGF of rv X, what is the distribution of X?

  $M(s) = \frac{1}{4}e^{-s} + \frac{1}{2} + \frac{1}{8}e^{4s} + \frac{1}{8}e^{5s}$

- Note that $M(s) = \sum_x e^{sx} p_X(x)$

- We can see that

$$p_X(-1) = \frac{1}{4},\ p_X(0) = \frac{1}{2},\ p_X(4) = \frac{1}{8},\ p_X(5) = \frac{1}{8}$$

Example 2

- Given the following MGF of rv X, what is the distribution of X?

$$M(s) = \frac{pe^s}{1-(1-p)e^s}$$

Example 2

- Given the following MGF of rv X, what is the distribution of X?

$$M(s) = \frac{pe^s}{1-(1-p)e^s}$$

- Note that $M(s) = \sum_x e^{sx} p_X(x)$

Example 2

- Given the following MGF of rv X, what is the distribution of X?

  $$M(s) = \frac{pe^s}{1-(1-p)e^s}$$

- Note that $M(s) = \sum_x e^{sx} p_X(x)$
- $M(s)$ can be reexpressed as the following geometric sum: when $(1-p)e^s < 1$,

  $$M(s) = pe^s\left(1 + (1-p)e^s + (1-p)^2 e^{2s} + (1-p)^3 e^{3s} + \cdots\right)$$

Example 2

- Given the following MGF of rv X, what is the distribution of X?

  $$M(s) = \frac{pe^s}{1-(1-p)e^s}$$

- Note that $M(s) = \sum_x e^{sx} p_X(x)$
- $M(s)$ can be reexpressed as the following geometric sum: when $(1-p)e^s < 1$,

  $$M(s) = pe^s\left(1 + (1-p)e^s + (1-p)^2 e^{2s} + (1-p)^3 e^{3s} + \cdots\right)$$

- $p_X(k)$: coefficient of the term e^{ks}, which means:

  $$p_X(1) = p,\ p_X(2) = p(1-p),\ p_X(3) = p(1-p)^2,\ p_X(4) = p(1-p)^3, \ldots$$

Example 2

- Given the following MGF of rv X, what is the distribution of X?

  $$M(s) = \frac{pe^s}{1-(1-p)e^s}$$

- Note that $M(s) = \sum_x e^{sx} p_X(x)$
- $M(s)$ can be reexpressed as the following geometric sum: when $(1-p)e^s < 1$,

  $$M(s) = pe^s\left(1 + (1-p)e^s + (1-p)^2 e^{2s} + (1-p)^3 e^{3s} + \cdots\right)$$

- $p_X(k)$: coefficient of the term e^{ks}, which means:

  $$p_X(1) = p,\ p_X(2) = p(1-p),\ p_X(3) = p(1-p)^2,\ p_X(4) = p(1-p)^3, \ldots$$

- X is a geometric rv with parameter p

Back to CLT Proof (1)

- Without loss of generality, assume $\mathbb{E}(X_i) = 0$ and $\text{var}(X_i) = 1$

Back to CLT Proof (1)

- Without loss of generality, assume $\mathbb{E}(X_i) = 0$ and $\mathrm{var}(X_i) = 1$
- $Z_n = \dfrac{S_n}{\sqrt{n}} = \dfrac{X_1 + X_2 + \dots X_n}{\sqrt{n}}$

Back to CLT Proof (1)

- Without loss of generality, assume $\mathbb{E}(X_i) = 0$ and $\text{var}(X_i) = 1$
- $Z_n = \frac{S_n}{\sqrt{n}} = \frac{X_1 + X_2 + \ldots X_n}{\sqrt{n}}$
- We will show: MGF of Z_n converges to MFG of $\mathcal{N}(0, 1)$ (using inversion property)

Back to CLT Proof (1)

- Without loss of generality, assume $\mathbb{E}(X_i) = 0$ and $\mathrm{var}(X_i) = 1$
- $Z_n = \frac{S_n}{\sqrt{n}} = \frac{X_1 + X_2 + \ldots X_n}{\sqrt{n}}$
- We will show: MGF of Z_n converges to MFG of $\mathcal{N}(0,1)$ (using inversion property)
- Proof.

$$\mathbb{E}\left[e^{sS_n/\sqrt{n}}\right] = \mathbb{E}\left[e^{sX_1/\sqrt{n}}\right] \times \cdots \times \mathbb{E}\left[e^{sX_n/\sqrt{n}}\right]$$

$$= \qquad\qquad =$$

Back to CLT Proof (1)

- Without loss of generality, assume $\mathbb{E}(X_i) = 0$ and $\text{var}(X_i) = 1$
- $Z_n = \dfrac{S_n}{\sqrt{n}} = \dfrac{X_1 + X_2 + \dots X_n}{\sqrt{n}}$
- We will show: MGF of Z_n converges to MFG of $\mathcal{N}(0,1)$ (using inversion property)
- Proof.

$$\mathbb{E}\left[e^{sS_n/\sqrt{n}}\right] = \mathbb{E}\left[e^{sX_1/\sqrt{n}}\right] \times \cdots \times \mathbb{E}\left[e^{sX_n/\sqrt{n}}\right]$$
$$= \left(\mathbb{E}\left[e^{sX_1/\sqrt{n}}\right]\right)^n =$$

Back to CLT Proof (1)

- Without loss of generality, assume $\mathbb{E}(X_i) = 0$ and $\text{var}(X_i) = 1$
- $Z_n = \frac{S_n}{\sqrt{n}} = \frac{X_1 + X_2 + \ldots X_n}{\sqrt{n}}$
- We will show: MGF of Z_n converges to MFG of $\mathcal{N}(0,1)$ (using inversion property)
- Proof.

$$\begin{aligned}\mathbb{E}\left[e^{sS_n/\sqrt{n}}\right] &= \mathbb{E}\left[e^{sX_1/\sqrt{n}}\right] \times \cdots \times \mathbb{E}\left[e^{sX_n/\sqrt{n}}\right] \\ &= \left(\mathbb{E}\left[e^{sX_1/\sqrt{n}}\right]\right)^n = \left(M_{X_1}\left(\frac{s}{\sqrt{n}}\right)\right)^n\end{aligned}$$

Back to CLT Proof (1)

- Without loss of generality, assume $\mathbb{E}(X_i) = 0$ and $\text{var}(X_i) = 1$
- $Z_n = \frac{S_n}{\sqrt{n}} = \frac{X_1 + X_2 + \ldots X_n}{\sqrt{n}}$
- We will show: MGF of Z_n converges to MFG of $\mathcal{N}(0, 1)$ (using inversion property)
- Proof.

$$\begin{aligned}\mathbb{E}\left[e^{sS_n/\sqrt{n}}\right] &= \mathbb{E}\left[e^{sX_1/\sqrt{n}}\right] \times \cdots \times \mathbb{E}\left[e^{sX_n/\sqrt{n}}\right] \\ &= \left(\mathbb{E}\left[e^{sX_1/\sqrt{n}}\right]\right)^n = \left(M_{X_1}\left(\frac{s}{\sqrt{n}}\right)\right)^n\end{aligned}$$

- For simplicity, let $M(\cdot) = M_{X_1}(\cdot)$

Back to CLT Proof (2)

- $M(0) = 1,\ M'(0) = 0,\ M''(0) = 1$

$\square$

Back to CLT Proof (2)

- $M(0) = 1,\ M'(0) = 0,\ M''(0) = 1$
- $\left(M\left(\frac{s}{\sqrt{n}}\right)\right)^n \xrightarrow{n\to\infty}$ what???

$\square$

Back to CLT Proof (2)

- $M(0) = 1,\ M'(0) = 0,\ M''(0) = 1$
- $\left(M\left(\frac{s}{\sqrt{n}}\right)\right)^n \xrightarrow{n\to\infty}$ what???
- Taking log, $n \log M\left(\frac{s}{\sqrt{n}}\right) \xrightarrow{n\to\infty}$ what???

$\square$

Back to CLT Proof (2)

- $M(0) = 1,\ M'(0) = 0,\ M''(0) = 1$
- $\left(M\left(\frac{s}{\sqrt{n}}\right)\right)^n \xrightarrow{n\to\infty}$ what???
- Taking log, $n \log M\left(\frac{s}{\sqrt{n}}\right) \xrightarrow{n\to\infty}$ what???
- For convenience, do the change of variable $y = \frac{1}{\sqrt{n}}$. Then, $\displaystyle\lim_{y\to 0} \frac{\log M(ys)}{y^2}$

$\square$

Back to CLT Proof (2)

- $M(0) = 1,\ M'(0) = 0,\ M''(0) = 1$
- $\left(M\left(\frac{s}{\sqrt{n}}\right)\right)^n \xrightarrow{n\to\infty}$ what???
- Taking log, $n \log M\left(\frac{s}{\sqrt{n}}\right) \xrightarrow{n\to\infty}$ what???
- For convenience, do the change of variable $y = \frac{1}{\sqrt{n}}$. Then, $\displaystyle\lim_{y\to 0} \frac{\log M(ys)}{y^2}$
- If we apply l'hopital's rule twice (please check), we get

$$\lim_{y\to 0} \frac{\log M(ys)}{y^2} = \frac{s^2}{2}$$

□

Questions?

Review Questions

1) What's the practical value of LLN and CLT?
2) Explain LLN and CLT from the scaling perspective.
3) Why are LLN and CLT great?
4) Why do we need different concepts of convergence for random variables?
5) Explain what is convergence in probability.
6) Explain what is convergence in distribution.
7) Why is MGF useful?