CS 107, Probability, Spring 2020
Lecture 37

Michael Poghosyan
mpoghosyan@aua.am

AUA

08 May 2020

Content

- Simple Concentration Inequalities
  - Markov's Inequality
  - Chebyshev's Inequality
- Limit Theorems
  - The Law of Large Numbers
  - The Central Limit Theorem

Simple Concentration Inequalities: Markov and Chebyshev inequalities

Markov and Chebyshev Inequalities: Intro

We have learned so far how to work with r.v.s using their CDFs and PD(M)Fs.

Markov and Chebyshev Inequalities: Intro

We have learned so far how to work with r.v.s using their CDFs and PD(M)Fs. Say, we know how to calculate Probabilities like

$$\mathbb{P}(X \in [a, b]),$$

using the CDF or PD(M)F of a r.v. X.

Markov and Chebyshev Inequalities: Intro

We have learned so far how to work with r.v.s using their CDFs and PD(M)Fs. Say, we know how to calculate Probabilities like

$$\mathbb{P}(X \in [a, b]),$$

using the CDF or PD(M)F of a r.v. X.

During our last few lectures we introduces some important partial characteristics for a r.v. X, its Expectation and Variance, and talked that usually, in practice, we do not have the complete information about the r.v., we do not have its CDF or PD(M)F, rather, we have (or can estimate) its Mean and Variance.

Markov and Chebyshev Inequalities: Intro

We have learned so far how to work with r.v.s using their CDFs and PD(M)Fs. Say, we know how to calculate Probabilities like

$$\mathbb{P}(X \in [a, b]),$$

using the CDF or PD(M)F of a r.v. X.

During our last few lectures we introduces some important partial characteristics for a r.v. X, its Expectation and Variance, and talked that usually, in practice, we do not have the complete information about the r.v., we do not have its CDF or PD(M)F, rather, we have (or can estimate) its Mean and Variance. Then a problem is arising - can we calculate probabilities concerning X, if we have only this partial information about X?

Markov and Chebyshev Inequalities: Intro

Of course, no.

[1] See, e.g., http://www.econ.upf.edu/~lugosi/mlss_conc.pdf

Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not enough to calculate Probabilities like $\mathbb{P}(X \in [a, b])$.

[1] See, e.g., `http://www.econ.upf.edu/~lugosi/mlss_conc.pdf`

Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not enough to calculate Probabilities like $\mathbb{P}(X \in [a, b])$. This is because we can have different (infinitely many) r.v.s with the same $\mathbb{E}(X)$ and $Var(X)$, and, hence, different values for $\mathbb{P}(X \in [a, b])$.

[1] See, e.g., http://www.econ.upf.edu/~lugosi/mlss_conc.pdf

Markov and Chebyshev Inequalities: Intro

Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not enough to calculate Probabilities like $\mathbb{P}(X \in [a, b])$. This is because we can have different (infinitely many) r.v.s with the same $\mathbb{E}(X)$ and $Var(X)$, and, hence, different values for $\mathbb{P}(X \in [a, b])$.

Fortunately, it turns out that having only Mean and Variance, we can estimate some Probabilities.

[1]See, e.g., http://www.econ.upf.edu/~lugosi/mlss_conc.pdf

Markov and Chebyshev Inequalities: Intro

Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not enough to calculate Probabilities like $\mathbb{P}(X \in [a, b])$. This is because we can have different (infinitely many) r.v.s with the same $\mathbb{E}(X)$ and $Var(X)$, and, hence, different values for $\mathbb{P}(X \in [a, b])$.

Fortunately, it turns out that having only Mean and Variance, we can estimate some Probabilities. For example, Concentration Inequalities[1] estimate how concentrated are the values of a r.v. around its mean.

[1] See, e.g., http://www.econ.upf.edu/~lugosi/mlss_conc.pdf

Markov and Chebyshev Inequalities: Intro

Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not enough to calculate Probabilities like $\mathbb{P}(X \in [a, b])$. This is because we can have different (infinitely many) r.v.s with the same $\mathbb{E}(X)$ and $Var(X)$, and, hence, different values for $\mathbb{P}(X \in [a, b])$.

Fortunately, it turns out that having only Mean and Variance, we can estimate some Probabilities. For example, Concentration Inequalities[1] estimate how concentrated are the values of a r.v. around its mean.Chabyshev Inequality is a basic example of a Concentration Inequality.

[1] See, e.g., `http://www.econ.upf.edu/~lugosi/mlss_conc.pdf`

Markov and Chebyshev Inequalities: Intro

Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not enough to calculate Probabilities like $\mathbb{P}(X \in [a, b])$. This is because we can have different (infinitely many) r.v.s with the same $\mathbb{E}(X)$ and $Var(X)$, and, hence, different values for $\mathbb{P}(X \in [a, b])$.

Fortunately, it turns out that having only Mean and Variance, we can estimate some Probabilities. For example, Concentration Inequalities[1] estimate how concentrated are the values of a r.v. around its mean.Chabyshev Inequality is a basic example of a Concentration Inequality. And Markov Inequality is estimating the tail Probability of the r.v..

[1] See, e.g., http://www.econ.upf.edu/~lugosi/mlss_conc.pdf

Markov and Chebyshev Inequalities: Intro

Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not enough to calculate Probabilities like $\mathbb{P}(X \in [a, b])$. This is because we can have different (infinitely many) r.v.s with the same $\mathbb{E}(X)$ and $Var(X)$, and, hence, different values for $\mathbb{P}(X \in [a, b])$.

Fortunately, it turns out that having only Mean and Variance, we can estimate some Probabilities. For example, Concentration Inequalities[1] estimate how concentrated are the values of a r.v. around its mean.Chabyshev Inequality is a basic example of a Concentration Inequality. And Markov Inequality is estimating the tail Probability of the r.v..

First, we will consider the Markov's Inequality, assuming we know the Expectation $\mathbb{E}(X)$.

[1] See, e.g., http://www.econ.upf.edu/~lugosi/mlss_conc.pdf

Markov and Chebyshev Inequalities

First, we state the Markov's Inequality:

Markov and Chebyshev Inequalities

First, we state the Markov's Inequality:

Markov Inequality

Assume X is a r.v. with $X \geq 0$, and let $a > 0$. Then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Markov and Chebyshev Inequalities

First, we state the Markov's Inequality:

Markov Inequality

Assume X is a r.v. with $X \geq 0$, and let $a > 0$. Then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Proof: We introduce a new r.v.

$$Y = \begin{cases} a, & X \geq a \\ 0, & \text{otherwise} \end{cases}$$

Markov and Chebyshev Inequalities

First, we state the Markov's Inequality:

Markov Inequality

Assume X is a r.v. with $X \ge 0$, and let $a > 0$. Then

$$\mathbb{P}(X \ge a) \le \frac{\mathbb{E}(X)}{a}.$$

Proof: We introduce a new r.v.

$$Y = \begin{cases} a, & X \ge a \\ 0, & \text{otherwise} \end{cases}$$

Now, it is easy to see that

$$Y \le X$$

Markov and Chebyshev Inequalities

Hence,

$$\mathbb{E}(Y) \leq \mathbb{E}(X).$$

Markov and Chebyshev Inequalities

Hence,

$$\mathbb{E}(Y) \leq \mathbb{E}(X).$$

But

$$\mathbb{E}(Y) = a \cdot \mathbb{P}(X \geq a),$$

yielding to the Markov Inequality.

Markov and Chebyshev Inequalities

Hence,

$$\mathbb{E}(Y) \le \mathbb{E}(X).$$

But

$$\mathbb{E}(Y) = a \cdot \mathbb{P}(X \ge a),$$

yielding to the Markov Inequality.

Note: In Markov Inequality, it is important that $X \ge 0$.

Markov and Chebyshev Inequalities

Hence,

$$\mathbb{E}(Y) \leq \mathbb{E}(X).$$

But

$$\mathbb{E}(Y) = a \cdot \mathbb{P}(X \geq a),$$

yielding to the Markov Inequality.

Note: In Markov Inequality, it is important that $X \geq 0$. For example, if X is a r.v. taking both negative and positive values, with $\mathbb{E}(X) = 0$, then the Markov Inequality will not be correct for X.

Hence,

$$\mathbb{E}(Y) \leq \mathbb{E}(X).$$

But

$$\mathbb{E}(Y) = a \cdot \mathbb{P}(X \geq a),$$

yielding to the Markov Inequality.

Note: In Markov Inequality, it is important that $X \geq 0$. For example, if X is a r.v. taking both negative and positive values, with $\mathbb{E}(X) = 0$, then the Markov Inequality will not be correct for X.

In the general case, if X is any r.v., we can apply the Markov's Inequality to $|X|$ and obtain

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}.$$

Markov and Chebyshev Inequalities

Markov and Chebyshev Inequalities

Note: The Markov Inequality,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a},$$

is informative if $a > \mathbb{E}(X)$. Otherwise, the inequality is trivial.

Markov and Chebyshev Inequalities

Note: The Markov Inequality,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a},$$

is informative if $a > \mathbb{E}(X)$. Otherwise, the inequality is trivial.

Note: We can rewrite the Markov's Inequality as

$$\mathbb{P}\Big(X \geq a \cdot \mathbb{E}(X)\Big) \leq \frac{1}{a}.$$

Markov and Chebyshev Inequalities

Note: The Markov Inequality,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a},$$

is informative if $a > \mathbb{E}(X)$. Otherwise, the inequality is trivial.

Note: We can rewrite the Markov's Inequality as

$$\mathbb{P}\Big(X \geq a \cdot \mathbb{E}(X)\Big) \leq \frac{1}{a}.$$

We interpret this as for a non-negative r.v., we have a small probability that it takes values much larger than the Expected Value.

Markov and Chebyshev Inequalities

Note: The Markov Inequality,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a},$$

is informative if $a > \mathbb{E}(X)$. Otherwise, the inequality is trivial.

Note: We can rewrite the Markov's Inequality as

$$\mathbb{P}\Big(X \geq a \cdot \mathbb{E}(X)\Big) \leq \frac{1}{a}.$$

We interpret this as for a non-negative r.v., we have a small probability that it takes values much larger than the Expected Value.

Note: Also, we can rewrite the Markov's Inequality as: for any r.v. $X \geq 0$, and for any $a > 0$,

$$\mathbb{P}(X < a) \geq 1 - \frac{\mathbb{E}(X)}{a}.$$

Markov and Chebyshev Inequalities

So the Markov's Inequality gives us a tool to estimate the probability $\mathbb{P}(X \geq a)$ if we know only the Expected Value $\mathbb{E}(X)$.

Markov and Chebyshev Inequalities

So the Markov's Inequality gives us a tool to estimate the probability $\mathbb{P}(X \geq a)$ if we know only the Expected Value $\mathbb{E}(X)$.

Example: Let X be the wage of a (randomly chosen working) person in Armenia. We know that $\mathbb{E}(X) = 190,136$AMD in 2020 first quarter.

So the Markov's Inequality gives us a tool to estimate the probability $\mathbb{P}(X \geq a)$ if we know only the Expected Value $\mathbb{E}(X)$.

Example: Let X be the wage of a (randomly chosen working) person in Armenia. We know that $\mathbb{E}(X) = 190,136$AMD in 2020 first quarter. Then the proportion of persons receiving more than 10 times the average wage is not more than

Markov and Chebyshev Inequalities

So the Markov's Inequality gives us a tool to estimate the probability $\mathbb{P}(X \geq a)$ if we know only the Expected Value $\mathbb{E}(X)$.

Example: Let X be the wage of a (randomly chosen working) person in Armenia. We know that $\mathbb{E}(X) = 190,136$AMD in 2020 first quarter. Then the proportion of persons receiving more than 10 times the average wage is not more than $1/10$.

Markov and Chebyshev Inequalities

So the Markov's Inequality gives us a tool to estimate the probability $\mathbb{P}(X \geq a)$ if we know only the Expected Value $\mathbb{E}(X)$.

Example: Let X be the wage of a (randomly chosen working) person in Armenia. We know that $\mathbb{E}(X) = 190,136$AMD in 2020 first quarter. Then the proportion of persons receiving more than 10 times the average wage is not more than $1/10$.
Proof: Indeed, by the Markov's Inequality,

$$\mathbb{P}\Big(X \geq 10\mathbb{E}(X)\Big) \leq \frac{\mathbb{E}(X)}{10\mathbb{E}(X)} = \frac{1}{10}.$$

Markov and Chebyshev Inequalities

So the Markov's Inequality gives us a tool to estimate the probability $\mathbb{P}(X \geq a)$ if we know only the Expected Value $\mathbb{E}(X)$.

Example: Let X be the wage of a (randomly chosen working) person in Armenia. We know that $\mathbb{E}(X) = 190,136$AMD in 2020 first quarter. Then the proportion of persons receiving more than 10 times the average wage is not more than $1/10$.
Proof: Indeed, by the Markov's Inequality,

$$\mathbb{P}\Big(X \geq 10\mathbb{E}(X)\Big) \leq \frac{\mathbb{E}(X)}{10\mathbb{E}(X)} = \frac{1}{10}.$$

From Wikipedia: by the Markov Inequality, no more than 1/5 of the population can have more than 5 times the average income.

Markov and Chebyshev Inequalities

Now, we give the Chebyshev's Inequality.

Markov and Chebyshev Inequalities

Now, we give the Chebyshev's Inequality. We assume that we know about X not only its Expectation $\mathbb{E}(X)$, but also the Variance $Var(X)$.

Markov and Chebyshev Inequalities

Now, we give the Chebyshev's Inequality. We assume that we know about X not only its Expectation $\mathbb{E}(X)$, but also the Variance $Var(X)$. So, having more information, we can estimate Probabilities more precisely:

Now, we give the Chebyshev's Inequality. We assume that we know about X not only its Expectation $\mathbb{E}(X)$, but also the Variance $Var(X)$. So, having more information, we can estimate Probabilities more precisely:

Chebyshev's Inequality

Assume X is a r.v., and let $a > 0$. Then

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq a\Big) \leq \frac{Var(X)}{a^2}.$$

Markov and Chebyshev Inequalities

Now, we give the Chebyshev's Inequality. We assume that we know about X not only its Expectation $\mathbb{E}(X)$, but also the Variance $Var(X)$. So, having more information, we can estimate Probabilities more precisely:

Chebyshev's Inequality

Assume X is a r.v., and let $a > 0$. Then

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq a\Big) \leq \frac{Var(X)}{a^2}.$$

Note: We can rewrite the Chebyshev's Inequality as

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq a \cdot SD(X)\Big) \leq \frac{1}{a^2}.$$

Markov and Chebyshev Inequalities

Note: Another way to represent the Chabyshev's Inequality is

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| < a\Big) \geq 1 - \frac{Var(X)}{a^2}.$$

Markov and Chebyshev Inequalities

Note: Another way to represent the Chabyshev's Inequality is

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| < a\Big) \geq 1 - \frac{Var(X)}{a^2}.$$

Note: Chebyshev's Inequality says that we can estimate the probability that X takes values away from its Expected Value by its Variance.

Markov and Chebyshev Inequalities

Note: Another way to represent the Chabyshev's Inequality is

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| < a\Big) \geq 1 - \frac{Var(X)}{a^2}.$$

Note: Chebyshev's Inequality says that we can estimate the probability that X takes values away from its Expected Value by its Variance.
Say, if the Variance of X is small, then the probability that X will be far from $\mathbb{E}(X)$ is small.

Markov and Chebyshev Inequalities

Note: Another way to represent the Chabyshev's Inequality is

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| < a\Big) \geq 1 - \frac{Var(X)}{a^2}.$$

Note: Chebyshev's Inequality says that we can estimate the probability that X takes values away from its Expected Value by its Variance.
Say, if the Variance of X is small, then the probability that X will be far from $\mathbb{E}(X)$ is small.
E.g., the probability that X will be more than $3SD(X)$-away from $\mathbb{E}(X)$ is:

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq 3 \cdot SD(X)\Big) \leq \frac{1}{3^2} = \frac{1}{9}.$$

Markov and Chebyshev Inequalities

Note: Another way to represent the Chabyshev's Inequality is

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| < a\Big) \geq 1 - \frac{Var(X)}{a^2}.$$

Note: Chebyshev's Inequality says that we can estimate the probability that X takes values away from its Expected Value by its Variance.
Say, if the Variance of X is small, then the probability that X will be far from $\mathbb{E}(X)$ is small.
E.g., the probability that X will be more than $3SD(X)$-away from $\mathbb{E}(X)$ is:

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq 3 \cdot SD(X)\Big) \leq \frac{1}{3^2} = \frac{1}{9}.$$

And this works for ANY r.v. X!

Markov and Chebyshev Inequalities

Now, let us compare:

Markov and Chebyshev Inequalities

Now, let us compare:

- if $X \sim \mathcal{N}(\mu, \sigma^2)$, then, by the 3σ rule,

$$\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.997;$$

Markov and Chebyshev Inequalities

Now, let us compare:

- if $X \sim \mathcal{N}(\mu, \sigma^2)$, then, by the 3σ rule,

$$\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.997;$$

We can write this as

$$\mathbb{P}(|X - \mu| < 3\sigma) \approx 0.997$$

Markov and Chebyshev Inequalities

Now, let us compare:

- if $X \sim \mathcal{N}(\mu, \sigma^2)$, then, by the 3σ rule,

$$\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.997;$$

  We can write this as

$$\mathbb{P}(|X - \mu| < 3\sigma) \approx 0.997$$

- If X is ANY r.v. with $\mathbb{E}(X) = \mu$ and $SD(X) = \sigma$, then, by Chebyshev Inequality,

$$\mathbb{P}\Big(|X - \mu| < 3\sigma\Big) \geq 1 - \frac{1}{9} = \frac{8}{9}.$$

Chebyshev Inequality, Example

Example 37.1: Assume X is a r.v. with $\mathbb{E}(X) = 9$ and $Var(X) = 0.2$. Estimate

a. $\mathbb{P}(8 < X < 10)$;

b. $\mathbb{P}(8.1 < X < 10)$.

Chebyshev Inequality, Example

Example 37.2: Assume $X_1, X_2, ..., X_n$ are IID r.vs with finite $\mu = \mathbb{E}(X_k)$ and $\sigma^2 = Var(X_k)$, and let

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Prove that

$$\mathbb{P}\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n \cdot \varepsilon^2}.$$

Chebyshev Inequality, Example

Example 37.3: Assume X is a r.v. and $g(t)$ is a nonnegative strictly increasing function.

a. Prove that
$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(g(X))}{g(a)};$$

b. Prove that
$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}\left(e^X\right)}{e^a}.$$

Limit Theorems

General Facts about the Sum and Average of IID r.v.s

IID Sequence of r.v.s

Recall that we defined $X_1, X_2, ..., X_n$ are IID if

- X_1, ... ,X_n are Identically Distributed, i.e., they have the same Distribution (the same CDFs, say);
- $X_1, ..., X_n$ are Independent.

IID Sequence of r.v.s

Recall that we defined $X_1, X_2, ..., X_n$ are IID if

- X_1, ... ,X_n are Identically Distributed, i.e., they have the same Distribution (the same CDFs, say);
- $X_1, ..., X_n$ are Independent.

Note: Please note that since all X_k-s have the same Distribution, then they have the same Expected values, the same Variances and all other characteristics, i.e.

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) = ... = \mathbb{E}(X_n),$$

$$Var(X_1) = Var(X_2) = ... = Var(X_n).$$

IID Sequence of r.v.s

Recall that we defined $X_1, X_2, ..., X_n$ are IID if

- X_1, ... ,X_n are Identically Distributed, i.e., they have the same Distribution (the same CDFs, say);
- $X_1, ..., X_n$ are Independent.

Note: Please note that since all X_k-s have the same Distribution, then they have the same Expected values, the same Variances and all other characteristics, i.e.

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) = ... = \mathbb{E}(X_n),$$

$$Var(X_1) = Var(X_2) = ... = Var(X_n).$$

In the rest, we will consider an infinite sequence of IID r.v.s $X_1, X_2, X_3, ...$

The Question

The Questions we consider here are:

Questions

Assume $X_1, X_2, ..., X_n$ are IID r.v.s.

Q1 What is the Distribution of

$$S_n = X_1 + X_2 + ... + X_n?$$

Q2 What is the Distribution of

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}?$$

The Question

The Questions we consider here are:

Questions

Assume $X_1, X_2, ..., X_n$ are IID r.v.s.

Q1 What is the Distribution of

$$S_n = X_1 + X_2 + ... + X_n?$$

Q2 What is the Distribution of

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}?$$

As you remember, calculation of the Distribution of the sum $X + Y$ is not an easy job (one needs to calculate Convolutions), so calculation of the exact Distribution of S_n and $\overline{X}_n$ is not an easy job, in general.

The Sum and Sample Mean Distribution

It turns out that for some particular cases, we can exactly describe the Distribution of S_n and $\overline{X}_n$:

The Sum and Sample Mean Distribution

It turns out that for some particular cases, we can exactly describe the Distribution of S_n and $\overline{X}_n$:

- If $X_k \sim Bernoulli(p)$, $k = 1, ..., n$ are independent, then

$$S_n = X_1 + ... + X_n \sim Binom(n, p);$$

The Sum and Sample Mean Distribution

It turns out that for some particular cases, we can exactly describe the Distribution of S_n and $\overline{X}_n$:

- If $X_k \sim Bernoulli(p)$, $k = 1, ..., n$ are independent, then
$$S_n = X_1 + ... + X_n \sim Binom(n, p);$$
- If $X_k \sim Binom(m, p)$, $k = 1, ..., n$, are independent, then
$$S_n = X_1 + ... + X_n \sim Binom(n \cdot m, p).$$

The Sum and Sample Mean Distribution

It turns out that for some particular cases, we can exactly describe the Distribution of S_n and $\overline{X}_n$:

- If $X_k \sim Bernoulli(p)$, $k = 1, ..., n$ are independent, then
$$S_n = X_1 + ... + X_n \sim Binom(n, p);$$
- If $X_k \sim Binom(m, p)$, $k = 1, ..., n$, are independent, then
$$S_n = X_1 + ... + X_n \sim Binom(n \cdot m, p).$$
- If $X_k \sim \mathcal{N}(\mu, \sigma^2)$, $k = 1, ..., n$, are Independent, then
$$S_n = X_1 + ... + X_n \sim \mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$$
$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

The Sum and Sample Mean Distribution

- If $X_k \sim Pois(\lambda)$, $k = 1, ..., n$, are independent, then

$$S_n = X_1 + ... + X_n \sim Pois(n \cdot \lambda).$$

The Sum and Sample Mean Distribution

- If $X_k \sim Pois(\lambda)$, $k = 1, ..., n$, are independent, then
$$S_n = X_1 + ... + X_n \sim Pois(n \cdot \lambda).$$

- If $X_k \sim Geom(p)$, $k = 1, ..., n$, are independent, then
$$S_n = X_1 + ... + X_n \sim NBinom(n, p).$$

Some Partial Information

Now, going back to the general case.

Some Partial Information

Now, going back to the general case. The general problem of finding the Distribution of S_n and $\overline{X}_n$ is a very hard one, but we can get some partial information about S_n and $\overline{X}_n$:

Some Partial Information

Now, going back to the general case. The general problem of finding the Distribution of S_n and $\overline{X}_n$ is a very hard one, but we can get some partial information about S_n and $\overline{X}_n$:

$$\mathbb{E}(S_n) = \mathbb{E}(X_1 + ... + X_n) =$$

Some Partial Information

Now, going back to the general case. The general problem of finding the Distribution of S_n and $\overline{X}_n$ is a very hard one, but we can get some partial information about S_n and $\overline{X}_n$:

$$\mathbb{E}(S_n) = \mathbb{E}(X_1 + ... + X_n) = n \cdot \mathbb{E}(X_1);$$

Some Partial Information

Now, going back to the general case. The general problem of finding the Distribution of S_n and $\overline{X}_n$ is a very hard one, but we can get some partial information about S_n and $\overline{X}_n$:

$$\mathbb{E}(S_n) = \mathbb{E}(X_1 + ... + X_n) = n \cdot \mathbb{E}(X_1);$$

$$Var(S_n) = Var(X_1 + ... + X_n) =$$

Some Partial Information

Now, going back to the general case. The general problem of finding the Distribution of S_n and $\overline{X}_n$ is a very hard one, but we can get some partial information about S_n and $\overline{X}_n$:

$$\mathbb{E}(S_n) = \mathbb{E}(X_1 + ... + X_n) = n \cdot \mathbb{E}(X_1);$$

$$Var(S_n) = Var(X_1 + ... + X_n) = n \cdot Var(X_1);$$

and

Some Partial Information

Now, going back to the general case. The general problem of finding the Distribution of S_n and $\overline{X}_n$ is a very hard one, but we can get some partial information about S_n and $\overline{X}_n$:

$$\mathbb{E}(S_n) = \mathbb{E}(X_1 + ... + X_n) = n \cdot \mathbb{E}(X_1);$$

$$Var(S_n) = Var(X_1 + ... + X_n) = n \cdot Var(X_1);$$

and

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{X_1 + ... + X_n}{n}\right) =$$

Some Partial Information

Now, going back to the general case. The general problem of finding the Distribution of S_n and $\overline{X}_n$ is a very hard one, but we can get some partial information about S_n and $\overline{X}_n$:

$$\mathbb{E}(S_n) = \mathbb{E}(X_1 + ... + X_n) = n \cdot \mathbb{E}(X_1);$$

$$Var(S_n) = Var(X_1 + ... + X_n) = n \cdot Var(X_1);$$

and

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{X_1 + ... + X_n}{n}\right) = \mathbb{E}(X_1);$$

Some Partial Information

Now, going back to the general case. The general problem of finding the Distribution of S_n and $\overline{X}_n$ is a very hard one, but we can get some partial information about S_n and $\overline{X}_n$:

$$\mathbb{E}(S_n) = \mathbb{E}(X_1 + ... + X_n) = n \cdot \mathbb{E}(X_1);$$

$$Var(S_n) = Var(X_1 + ... + X_n) = n \cdot Var(X_1);$$

and

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{X_1 + ... + X_n}{n}\right) = \mathbb{E}(X_1);$$

This means that the mean of the means is the mean $\ddot\smile$

Some Partial Information

Now, going back to the general case. The general problem of finding the Distribution of S_n and $\overline{X}_n$ is a very hard one, but we can get some partial information about S_n and $\overline{X}_n$:

$$\mathbb{E}(S_n) = \mathbb{E}(X_1 + \ldots + X_n) = n \cdot \mathbb{E}(X_1);$$

$$Var(S_n) = Var(X_1 + \ldots + X_n) = n \cdot Var(X_1);$$

and

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{X_1 + \ldots + X_n}{n}\right) = \mathbb{E}(X_1);$$

This means that the mean of the means is the mean $\ddot\smile$

$$Var(\overline{X}_n) = Var\left(\frac{X_1 + \ldots + X_n}{n}\right) =$$

Some Partial Information

Now, going back to the general case. The general problem of finding the Distribution of S_n and $\overline{X}_n$ is a very hard one, but we can get some partial information about S_n and $\overline{X}_n$:

$$\mathbb{E}(S_n) = \mathbb{E}(X_1 + ... + X_n) = n \cdot \mathbb{E}(X_1);$$

$$Var(S_n) = Var(X_1 + ... + X_n) = n \cdot Var(X_1);$$

and

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{X_1 + ... + X_n}{n}\right) = \mathbb{E}(X_1);$$

This means that the mean of the means is the mean $\ddot\smile$

$$Var(\overline{X}_n) = Var\left(\frac{X_1 + ... + X_n}{n}\right) = \frac{Var(X_1)}{n}.$$

Some Partial Information

Note: Surely you know the well-known saying: *"7 angam chaphir, mek angam ktrir"*.

Some Partial Information

Note: Surely you know the well-known saying: *"7 angam chaphir, mek angam ktrir"*. The above relationships are the mathematical statement of this proverb $\ddot\smile$

Some Partial Information

Note: Surely you know the well-known saying: *"7 angam chaphir, mek angam ktrir"*. The above relationships are the mathematical statement of this proverb $\ddot\smile$

Note: The interpretation of

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}(X_1) \qquad \text{and} \qquad Var(\overline{X}_n) = \frac{Var(X_1)}{n}$$

is that the values of $\overline{X}_n$ are centered at $\mathbb{E}(X_1)$ and are becoming more and more concentrated around that number as n increases.

Some Partial Information

Note: Surely you know the well-known saying: *"7 angam chaphir, mek angam ktrir"*. The above relationships are the mathematical statement of this proverb $\ddot\smile$

Note: The interpretation of

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}(X_1) \qquad \text{and} \qquad Var(\overline{X}_n) = \frac{Var(X_1)}{n}$$

is that the values of $\overline{X}_n$ are centered at $\mathbb{E}(X_1)$ and are becoming more and more concentrated around that number as n increases.

Note: These relations are important in the Financial Risk and Investment Theory: they are confirming the importance of the **Diversification**.

Some Partial Information

Note: Surely you know the well-known saying: *"7 angam chaphir, mek angam ktrir"*. The above relationships are the mathematical statement of this proverb $\ddot\smile$

Note: The interpretation of

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}(X_1) \qquad \text{and} \qquad Var(\overline{X}_n) = \frac{Var(X_1)}{n}$$

is that the values of $\overline{X}_n$ are centered at $\mathbb{E}(X_1)$ and are becoming more and more concentrated around that number as n increases.

Note: These relations are important in the Financial Risk and Investment Theory: they are confirming the importance of the **Diversification**.

Folklore: Diversification in one sentence: Do not put all your eggs into one basket!

The Law of Large Numbers

Intro to LLN and CLT

Now, we want to study some more properties of Distributions of $\overline{X}_n$ and/or S_n in the general case.

Intro to LLN and CLT

Now, we want to study some more properties of Distributions of $\overline{X}_n$ and/or S_n in the general case.

The two famous Limit Theorems in Probability Theory,

- The Law of the Large Numbers (LLN)
- The Central Limit Theorem (CLT)

help us to get an information about the **asymptotic** (i.e., limiting, or, for large n) properties of $\overline{X}_n$ and S_n, in the general case.

The Weak Law of Large Numbers, WLLN

If $X_1, X_2, ..., X_n, ...$ are IID, with finite $\mathbb{E}(X_1)$ and Variance $Var(X_1)$, then

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}(X_1), \qquad n \to +\infty,$$

The Weak LLN

The Weak Law of Large Numbers, WLLN

If $X_1, X_2, ..., X_n, ...$ are IID, with finite $\mathbb{E}(X_1)$ and Variance $Var(X_1)$, then

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}(X_1), \qquad n \to +\infty,$$

where $\mathbb{P}$ over the arrow sign means the convergence is in Probability:

The Weak LLN

The Weak Law of Large Numbers, WLLN

If $X_1, X_2, ..., X_n, ...$ are IID, with finite $\mathbb{E}(X_1)$ and Variance $Var(X_1)$, then

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}(X_1), \qquad n \to +\infty,$$

where $\mathbb{P}$ over the arrow sign means the convergence is in Probability: for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{X_1 + X_2 + ... + X_n}{n} - \mathbb{E}(X_1)\right| \geq \varepsilon\right) \to 0, \qquad n \to +\infty.$$

The Weak LLN

The Weak Law of Large Numbers, WLLN

If $X_1, X_2, ..., X_n, ...$ are IID, with finite $\mathbb{E}(X_1)$ and Variance $Var(X_1)$, then

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}(X_1), \qquad n \to +\infty,$$

where $\mathbb{P}$ over the arrow sign means the convergence is in Probability: for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left| \frac{X_1 + X_2 + ... + X_n}{n} - \mathbb{E}(X_1) \right| \geq \varepsilon \right) \to 0, \qquad n \to +\infty.$$

Note: This means that for any $\varepsilon > 0$, the chances that $\overline{X}_n$ is far from $\mathbb{E}(X_1)$ more than ε, is very small, if n is large.

The Weak LLN: The Idea

The idea of the LLN is the following:

The Weak LLN: The Idea

The idea of the LLN is the following: let us denote $\mu = \mathbb{E}(X_1)$.

The Weak LLN: The Idea

The idea of the LLN is the following: let us denote $\mu = \mathbb{E}(X_1)$. We know that the Mean of $\overline{X}_n$ is

The Weak LLN: The Idea

The idea of the LLN is the following: let us denote $\mu = \mathbb{E}(X_1)$. We know that the Mean of $\overline{X}_n$ is μ: $\mathbb{E}(\overline{X}_n) = \mu$.

The Weak LLN: The Idea

The idea of the LLN is the following: let us denote $\mu = \mathbb{E}(X_1)$. We know that the Mean of $\overline{X}_n$ is μ: $\mathbb{E}(\overline{X}_n) = \mu$. Is this enough to say that $\overline{X}_n$ is close to μ?

The Weak LLN: The Idea

The idea of the LLN is the following: let us denote $\mu = \mathbb{E}(X_1)$. We know that the Mean of $\overline{X}_n$ is μ: $\mathbb{E}(\overline{X}_n) = \mu$. Is this enough to say that $\overline{X}_n$ is close to μ? Of course, no.

The Weak LLN: The Idea

The idea of the LLN is the following: let us denote $\mu = \mathbb{E}(X_1)$. We know that the Mean of $\overline{X}_n$ is μ: $\mathbb{E}(\overline{X}_n) = \mu$. Is this enough to say that $\overline{X}_n$ is close to μ? Of course, no.
And what we need to have to say that $\overline{X}_n$ is close to μ?

The Weak LLN: The Idea

The idea of the LLN is the following: let us denote $\mu = \mathbb{E}(X_1)$. We know that the Mean of $\overline{X}_n$ is μ: $\mathbb{E}(\overline{X}_n) = \mu$. Is this enough to say that $\overline{X}_n$ is close to μ? Of course, no.

And what we need to have to say that $\overline{X}_n$ is close to μ? That the Variance of $\overline{X}_n$ is close to 0!

The Weak LLN: The Idea

The idea of the LLN is the following: let us denote $\mu = \mathbb{E}(X_1)$. We know that the Mean of $\overline{X}_n$ is μ: $\mathbb{E}(\overline{X}_n) = \mu$. Is this enough to say that $\overline{X}_n$ is close to μ? Of course, no.
And what we need to have to say that $\overline{X}_n$ is close to μ? That the Variance of $\overline{X}_n$ is close to 0! And we know that

$$Var(\overline{X}_n) = \frac{Var(X_1)}{n} \to 0, \qquad n \to +\infty.$$

The Weak LLN: The Idea

The idea of the LLN is the following: let us denote $\mu = \mathbb{E}(X_1)$. We know that the Mean of $\overline{X}_n$ is μ: $\mathbb{E}(\overline{X}_n) = \mu$. Is this enough to say that $\overline{X}_n$ is close to μ? Of course, no.
And what we need to have to say that $\overline{X}_n$ is close to μ? That the Variance of $\overline{X}_n$ is close to 0! And we know that

$$Var(\overline{X}_n) = \frac{Var(X_1)}{n} \to 0, \qquad n \to +\infty.$$

Ura!

The Weak LLN: The Idea

The idea of the LLN is the following: let us denote $\mu = \mathbb{E}(X_1)$. We know that the Mean of $\overline{X}_n$ is μ: $\mathbb{E}(\overline{X}_n) = \mu$. Is this enough to say that $\overline{X}_n$ is close to μ? Of course, no.

And what we need to have to say that $\overline{X}_n$ is close to μ? That the Variance of $\overline{X}_n$ is close to 0! And we know that

$$Var(\overline{X}_n) = \frac{Var(X_1)}{n} \to 0, \qquad n \to +\infty.$$

Ura!

The rigorous proof is by using the Chebyshev Inequality: OTB

The Strong LLN

The Strong LLN Says that the above convergence holds also in the Strong Sense, under less restrictive settings:

The Strong Law of Large Numbers, SLLN, Kolmogorov

If $X_1, X_2, ..., X_n, ...$ are IID, with finite $\mathbb{E}(X_1)$, then

$$\frac{X_1 + X_2 + ... + X_n}{n} \xrightarrow{a.s.} \mathbb{E}(X_1), \qquad n \to +\infty,$$

The Strong LLN

The Strong LLN Says that the above convergence holds also in the Strong Sense, under less restrictive settings:

The Strong Law of Large Numbers, SLLN, Kolmogorov

If $X_1, X_2, ..., X_n, ...$ are IID, with finite $\mathbb{E}(X_1)$, then

$$\frac{X_1 + X_2 + ... + X_n}{n} \xrightarrow{a.s.} \mathbb{E}(X_1), \qquad n \to +\infty,$$

where $a.s.$ over the arrow sign means the convergence is in Almost Sure sense:

The Strong LLN

The Strong LLN Says that the above convergence holds also in the Strong Sense, under less restrictive settings:

The Strong Law of Large Numbers, SLLN, Kolmogorov

If $X_1, X_2, ..., X_n, ...$ are IID, with finite $\mathbb{E}(X_1)$, then

$$\frac{X_1 + X_2 + ... + X_n}{n} \stackrel{a.s.}{\to} \mathbb{E}(X_1), \qquad n \to +\infty,$$

where $a.s.$ over the arrow sign means the convergence is in Almost Sure sense:

$$\mathbb{P}\left(\lim_{n\to+\infty} \frac{X_1 + X_2 + ... + X_n}{n} = \mathbb{E}(X_1) \right) = 1.$$

The Strong LLN

Note: In the advanced theory of Probability, one is proving that the convergence in a.s. sense implies convergence in Probability, and the inverse implication is not true, in general.

The Strong LLN

Note: In the advanced theory of Probability, one is proving that the convergence in a.s. sense implies convergence in Probability, and the inverse implication is not true, in general. So the Strong LLN implies the Weak LLN.

The Strong LLN

Note: In the advanced theory of Probability, one is proving that the convergence in a.s. sense implies convergence in Probability, and the inverse implication is not true, in general. So the Strong LLN implies the Weak LLN.

Note: It is not so easy to explain the difference between the convergence in Probability and a.s. convergence quickly. But, in fact, it can happen that some sequence Y_n of r.v.s tends to r.v. Y in Probability, but, at no point $Y_n \to Y$ in the ordinary sense.

The Strong LLN

Note: In the advanced theory of Probability, one is proving that the convergence in a.s. sense implies convergence in Probability, and the inverse implication is not true, in general. So the Strong LLN implies the Weak LLN.

Note: It is not so easy to explain the difference between the convergence in Probability and a.s. convergence quickly. But, in fact, it can happen that some sequence Y_n of r.v.s tends to r.v. Y in Probability, but, at no point $Y_n \to Y$ in the ordinary sense.

Note: Assume $g : \mathbb{R} \to \mathbb{R}$ is a continuous function. Then one can prove that, under the SLLN(WLLN) conditions,

$$g\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) \xrightarrow{a.s. \ \ (\mathbb{P})} g\left(\mathbb{E}(X_1)\right), \qquad n \to +\infty.$$

The Strong LLN

Note: Sometimes we are required to calculate limits of the form:

$$\lim_{n\to+\infty} \frac{g(X_1) + g(X_2) + ... + g(X_n)}{n}$$

in the Probability or a.s. sense, for some nice function g.

The Strong LLN

Note: Sometimes we are required to calculate limits of the form:

$$\lim_{n\to+\infty} \frac{g(X_1)+g(X_2)+...+g(X_n)}{n}$$

in the Probability or a.s. sense, for some nice function g. Clearly, under the condition that $\mathbb{E}(g(X_1))$ and $Var(g(X_1))$ are finite (for the WLLN), or just $\mathbb{E}(g(X_1))$ is finite (for the SLLN), we will have, for $n \to +\infty$,

$$\frac{g(X_1)+g(X_2)+...+g(X_n)}{n} \xrightarrow{a.s.\ (\mathbb{P})} \mathbb{E}(g(X_1)).$$

W(S)LLN, Example

Example 37.4: Assume $X_1, X_2, ..., X_n \sim Unif[-1, 2]$ are IID. Calculate, in the $\mathbb{P}$ and a.s. sense,

a. $\lim\limits_{n\to+\infty} \dfrac{X_1 + X_2 + ... + X_n}{n}$;

b. $\lim\limits_{n\to+\infty} \dfrac{X_1^2 + X_2^2 + ... + X_n^2}{n}$;

c. $\lim\limits_{n\to+\infty} \dfrac{e^{X_1} + e^{X_2} + ... + e^{X_n}}{n}$.

W(S)LLN, Example

Example 37.5: Assume we have a coin, for which the probability of having Heads is $p \in (0, 1)$. We are tossing that coin many times, independently. We calculate the proportion of the Heads for that tosses. What is the limit of that proportion, almost surely, if we repeat tossing infinitely many times?

W(S)LLN, Example

Example 37.5: Assume we have a coin, for which the probability of having Heads is $p \in (0, 1)$. We are tossing that coin many times, independently. We calculate the proportion of the Heads for that tosses. What is the limit of that proportion, almost surely, if we repeat tossing infinitely many times?

Note: This example is the base of Frequentists interpretation of the Probability: "the Probability of an Event is p" can be understood as the following: we repeat the Experiment many times, independently, and calculate the proportion of times we will have that Event appearing. Then the limit of that proportion is exactly p.

W(S)LLN, Example

Example 37.6: Assume $g : [a, b] \to \mathbb{R}$ is a Continuous function, and we want to calculate the integral

$$\int_a^b g(x)\,dx.$$

a. Prove that if $X_1, X_2, \dots$ are IID with $X_k \sim Unif[a, b]$, then

$$\frac{g(X_1) + g(X_2) + \dots + g(X_n)}{n} \to \frac{1}{b-a} \cdot \int_a^b g(x)\,dx;$$

b. Calculate, in **R**, the integral

$$\int_0^2 \sin(x) \cdot e^{-x^2}\,dx.$$

Example:

Question: Why the idea of a Casino works?

Example:

Question: Why the idea of a Casino works? And why the idea of the Insurance works?

Example:

Question: Why the idea of a Casino works? And why the idea of the Insurance works? Sometimes Casino and Insurance companies pay a lot of money. So why their jobs are profitable?

Example:

Question: Why the idea of a Casino works? And why the idea of the Insurance works? Sometimes Casino and Insurance companies pay a lot of money. So why their jobs are profitable?
Idea: Because of the LLN!

Example:

Question: Why the idea of a Casino works? And why the idea of the Insurance works? Sometimes Casino and Insurance companies pay a lot of money. So why their jobs are profitable?
Idea: Because of the LLN! You cannot run away from the fortune, I mean, from Math. $\ddot\smile$.

Example:

Question: Why the idea of a Casino works? And why the idea of the Insurance works? Sometimes Casino and Insurance companies pay a lot of money. So why their jobs are profitable?
Idea: Because of the LLN! You cannot run away from the fortune, I mean, from Math. $\ddot\smile$. In particular, from LLN.

Example:

Question: Why the idea of a Casino works? And why the idea of the Insurance works? Sometimes Casino and Insurance companies pay a lot of money. So why their jobs are profitable?
Idea: Because of the LLN! You cannot run away from the fortune, I mean, from Math. $\ddot\smile$. In particular, from LLN. **In the long run, Casinos and Insurance Companies will win**!

Example:

Question: Why the idea of a Casino works? And why the idea of the Insurance works? Sometimes Casino and Insurance companies pay a lot of money. So why their jobs are profitable?
Idea: Because of the LLN! You cannot run away from the fortune, I mean, from Math. $\ddot\smile$. In particular, from LLN. **In the long run, Casinos and Insurance Companies will win**!

Consider, say, the American Roulette game: we have 38 pockets (numbers from 0 to 36, and 00), 0 and 00 are in green, 18 numbers are in red, and 18 are in black.

Example:

Question: Why the idea of a Casino works? And why the idea of the Insurance works? Sometimes Casino and Insurance companies pay a lot of money. So why their jobs are profitable?
Idea: Because of the LLN! You cannot run away from the fortune, I mean, from Math. $\ddot\smile$. In particular, from LLN. **In the long run, Casinos and Insurance Companies will win**!

Consider, say, the American Roulette game: we have 38 pockets (numbers from 0 to 36, and 00), 0 and 00 are in green, 18 numbers are in red, and 18 are in black. One of the betting methods is to bet on the color - either black or red.

Example:

Question: Why the idea of a Casino works? And why the idea of the Insurance works? Sometimes Casino and Insurance companies pay a lot of money. So why their jobs are profitable?
Idea: Because of the LLN! You cannot run away from the fortune, I mean, from Math. $\ddot\smile$. In particular, from LLN. **In the long run, Casinos and Insurance Companies will win**!

Consider, say, the American Roulette game: we have 38 pockets (numbers from 0 to 36, and 00), 0 and 00 are in green, 18 numbers are in red, and 18 are in black. One of the betting methods is to bet on the color - either black or red. Say, a person is winning $1\$$ if he/she correctly guessed the color, and, otherwise, is losing $1\$$.

Example:

We have calculated that the Expected winning of that person is $\mathbb{E}(W) = -\frac{1}{19}$, or, which is the same, from the point of view of our Casino, the Expected Gain of Casino will be $\mathbb{E}(G) = \frac{1}{19}$.

Example:

We have calculated that the Expected winning of that person is $\mathbb{E}(W) = -\dfrac{1}{19}$, or, which is the same, from the point of view of our Casino, the Expected Gain of Casino will be $\mathbb{E}(G) = \dfrac{1}{19}$.

OK, but what this means?

Example:

We have calculated that the Expected winning of that person is $\mathbb{E}(W) = -\dfrac{1}{19}$, or, which is the same, from the point of view of our Casino, the Expected Gain of Casino will be $\mathbb{E}(G) = \dfrac{1}{19}$. OK, but what this means?

Assume G_1 is the gain of Casino from the 1st person, G_2 - from the second one etc. So G_k-s are random, IID.

Example:

We have calculated that the Expected winning of that person is $\mathbb{E}(W) = -\frac{1}{19}$, or, which is the same, from the point of view of our Casino, the Expected Gain of Casino will be $\mathbb{E}(G) = \frac{1}{19}$. OK, but what this means?

Assume G_1 is the gain of Casino from the 1st person, G_2 - from the second one etc. So G_k-s are random, IID. It can happen, of course that, in some scenario, $G_1 = G_2 = ... = G_{50} = -1$.

Example:

Now, if we will consider n players, then the total gain of Casino will be $G_1 + G_2 + ... + G_n$.

Example:

Now, if we will consider n players, then the total gain of Casino will be $G_1 + G_2 + ... + G_n$. And the LLN says that

$$\frac{G_1 + G_2 + ... + G_n}{n} \to \mathbb{E}(G_1) = \frac{1}{19},$$

Example:

Now, if we will consider n players, then the total gain of Casino will be $G_1 + G_2 + ... + G_n$. And the LLN says that

$$\frac{G_1 + G_2 + ... + G_n}{n} \to \mathbb{E}(G_1) = \frac{1}{19},$$

so, for large n,

$$\text{Total Gain} = G_1 + G_2 + ... + G_n \approx n \cdot \mathbb{E}(G_1) = \frac{1}{19} \cdot n.$$

Example:

Now, if we will consider n players, then the total gain of Casino will be $G_1 + G_2 + \ldots + G_n$. And the LLN says that

$$\frac{G_1 + G_2 + \ldots + G_n}{n} \to \mathbb{E}(G_1) = \frac{1}{19},$$

so, for large n,

$$\text{Total Gain} = G_1 + G_2 + \ldots + G_n \approx n \cdot \mathbb{E}(G_1) = \frac{1}{19} \cdot n.$$

So, even sometimes Casino will pay to players, in the long run, for many players, Casino will win: it will have almost $\frac{n}{19}\$$ from n persons.

The Central Limit Theorem

Reminder about Standard Normal r.vs

Let me first recall that if Z is Standard Normal, i.e.

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

then we denote by $\Phi(x)$ its CDF and $\varphi(x)$ its PDF:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \qquad x \in \mathbb{R}$$

and

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)\, dt, \qquad x \in \mathbb{R}.$$

Reminder about Standard Normal r.vs

Let me first recall that if Z is Standard Normal, i.e.

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

then we denote by $\Phi(x)$ its CDF and $\varphi(x)$ its PDF:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \qquad x \in \mathbb{R}$$

and

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,dt, \qquad x \in \mathbb{R}.$$

And we know that

$$\mathbb{P}(a \le Z \le b) = \Phi(b) - \Phi(a),$$

and we can calculate the values of Φ either by some mathematical software, or by using Standard Normal tables.

The basic Idea of the CLT

Let's consider $\overline{X}_n$.

The basic Idea of the CLT

Let's consider $\overline{X}_n$. The LLN says that the values of $\overline{X}_n$ are concentrated around $\mathbb{E}(X_1)$.

The basic Idea of the CLT

Let's consider $\overline{X}_n$. The LLN says that the values of $\overline{X}_n$ are concentrated around $\mathbb{E}(X_1)$. But it is not giving us an idea about how the values are distributed around that mean value.

The basic Idea of the CLT

Let's consider $\overline{X}_n$. The LLN says that the values of $\overline{X}_n$ are concentrated around $\mathbb{E}(X_1)$. But it is not giving us an idea about how the values are distributed around that mean value. CLT helps us in this regard.

The basic Idea of the CLT

Let's consider $\overline{X}_n$. The LLN says that the values of $\overline{X}_n$ are concentrated around $\mathbb{E}(X_1)$. But it is not giving us an idea about how the values are distributed around that mean value. CLT helps us in this regard. So CLT is giving the asymptotic distribution of $\overline{X}_n$.

The basic Idea of the CLT

Let's consider $\overline{X}_n$. The LLN says that the values of $\overline{X}_n$ are concentrated around $\mathbb{E}(X_1)$. But it is not giving us an idea about how the values are distributed around that mean value. CLT helps us in this regard. So CLT is giving the asymptotic distribution of $\overline{X}_n$. Recall the Galton board!

The basic Idea of the CLT

Let's consider $\overline{X}_n$. The LLN says that the values of $\overline{X}_n$ are concentrated around $\mathbb{E}(X_1)$. But it is not giving us an idea about how the values are distributed around that mean value. CLT helps us in this regard. So CLT is giving the asymptotic distribution of $\overline{X}_n$. Recall the Galton board!

To give the general idea of the CLT, we use the Standardization of r.v.s: recall that, for a r.v. X, the following r.v. is its Standardization:

$$Y = \frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}} = \frac{X - \mathbb{E}(X)}{SD(X)},$$

The basic Idea of the CLT

Let's consider $\overline{X}_n$. The LLN says that the values of $\overline{X}_n$ are concentrated around $\mathbb{E}(X_1)$. But it is not giving us an idea about how the values are distributed around that mean value. CLT helps us in this regard. So CLT is giving the asymptotic distribution of $\overline{X}_n$. Recall the Galton board!

To give the general idea of the CLT, we use the Standardization of r.v.s: recall that, for a r.v. X, the following r.v. is its Standardization:

$$Y = \frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}} = \frac{X - \mathbb{E}(X)}{SD(X)},$$

And, we have

$$\mathbb{E}(Y) = 0 \qquad \text{and} \qquad Var(Y) = 1.$$

The basic Idea of the CLT

Let's consider $\overline{X}_n$. The LLN says that the values of $\overline{X}_n$ are concentrated around $\mathbb{E}(X_1)$. But it is not giving us an idea about how the values are distributed around that mean value. CLT helps us in this regard. So CLT is giving the asymptotic distribution of $\overline{X}_n$. Recall the Galton board!

To give the general idea of the CLT, we use the Standardization of r.v.s: recall that, for a r.v. X, the following r.v. is its Standardization:

$$Y = \frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}} = \frac{X - \mathbb{E}(X)}{SD(X)},$$

And, we have

$$\mathbb{E}(Y) = 0 \qquad \text{and} \qquad Var(Y) = 1.$$

Note: But Y is Normal only if X is Normal r.v.

The basic Idea of the CLT

The basic idea of the CLT is the following: if we have a sequence of IID r.vs X_n, with finite Expectation and Variance, and we consider their sum S_n or their average $\overline{X}_n$, then, for large n,

The basic Idea of the CLT

The basic idea of the CLT is the following: if we have a sequence of IID r.vs X_n, with finite Expectation and Variance, and we consider their sum S_n or their average $\overline{X}_n$, then, for large n,

> The Standardized S_n or $\overline{X}_n$ are approximately $\mathcal{N}(0,1)$ distributed,

The basic Idea of the CLT

The basic idea of the CLT is the following: if we have a sequence of IID r.vs X_n, with finite Expectation and Variance, and we consider their sum S_n or their average $\overline{X}_n$, then, for large n,

The Standardized S_n or $\overline{X}_n$ are approximately $\mathcal{N}(0,1)$ distributed,

i.e.,

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \text{ is approximately } \mathcal{N}(0,1) \text{ distributed,}$$

$$\frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \text{ is approximately } \mathcal{N}(0,1) \text{ distributed.}$$

The basic Idea of the CLT

The basic idea of the CLT is the following: if we have a sequence of IID r.vs X_n, with finite Expectation and Variance, and we consider their sum S_n or their average $\overline{X}_n$, then, for large n,

The Standardized S_n or $\overline{X}_n$ are approximately $\mathcal{N}(0,1)$ distributed,

i.e.,

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \text{ is approximately } \mathcal{N}(0,1) \text{ distributed,}$$

$$\frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \text{ is approximately } \mathcal{N}(0,1) \text{ distributed.}$$

And these hold **for any IID sequence X_k, from any Distribution**.

CLT, in the Sums form

Now, lets consider the rigorous statement of CLT, in the sums form.

CLT, in the Sums form

Now, lets consider the rigorous statement of CLT, in the sums form.

Let X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$.

CLT, in the Sums form

Now, lets consider the rigorous statement of CLT, in the sums form.

Let X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$.
We consider sums

$$S_n = X_1 + X_2 + ... + X_n.$$

CLT, in the Sums form

Now, lets consider the rigorous statement of CLT, in the sums form.

Let X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$.
We consider sums

$$S_n = X_1 + X_2 + ... + X_n.$$

We Standardize S_n: we denote

$$Z_n =$$

CLT, in the Sums form

Now, lets consider the rigorous statement of CLT, in the sums form.

Let X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$.
We consider sums

$$S_n = X_1 + X_2 + ... + X_n.$$

We Standardize S_n: we denote

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}.$$

CLT, in the Sums form

Now, lets consider the rigorous statement of CLT, in the sums form.

Let X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$.
We consider sums

$$S_n = X_1 + X_2 + ... + X_n.$$

We Standardize S_n: we denote

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}.$$

Now we use

$$\mathbb{E}(S_n) =$$

CLT, in the Sums form

Now, lets consider the rigorous statement of CLT, in the sums form.

Let X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$.
We consider sums

$$S_n = X_1 + X_2 + ... + X_n.$$

We Standardize S_n: we denote

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}.$$

Now we use

$$\mathbb{E}(S_n) = n \cdot \mu, \qquad Var(S_n) =$$

CLT, in the Sums form

Now, lets consider the rigorous statement of CLT, in the sums form.

Let X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$.
We consider sums

$$S_n = X_1 + X_2 + ... + X_n.$$

We Standardize S_n: we denote

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}.$$

Now we use

$$\mathbb{E}(S_n) = n \cdot \mu, \qquad Var(S_n) = n \cdot \sigma^2.$$

CLT, in the Sums form

Then,

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma}.$$

CLT, in the Sums form

Then,

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma}.$$

The CLT states:

CLT, in the Sums form

Then,

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma}.$$

The CLT states:

CLT, in the Sums form

For any $a < b$, as $n \to \infty$,

$$\mathbb{P}(a \leq Z_n \leq b) \to \mathbb{P}(a \leq Z \leq b),$$

where $Z \sim \mathcal{N}(0, 1)$,

CLT, in the Sums form

Then,

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma}.$$

The CLT states:

CLT, in the Sums form

For any $a < b$, as $n \to \infty$,

$$\mathbb{P}(a \le Z_n \le b) \to \mathbb{P}(a \le Z \le b),$$

where $Z \sim \mathcal{N}(0, 1)$, i.e.,

$$\mathbb{P}\left(a \le \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \le b\right) \to \mathbb{P}(a \le Z \le b).$$

CLT, in the Averages form

Now, let us give CLT in the Averages form.

CLT, in the Averages form

Now, let us give CLT in the Averages form. Again, assume X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$.

CLT, in the Averages form

Now, let us give CLT in the Averages form. Again, assume X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$. We consider the Averages

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

CLT, in the Averages form

Now, let us give CLT in the Averages form. Again, assume X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$. We consider the Averages

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

We Standardize $\overline{X}_n$: we denote

$$Z_n =$$

CLT, in the Averages form

Now, let us give CLT in the Averages form. Again, assume X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$. We consider the Averages

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

We Standardize $\overline{X}_n$: we denote

$$Z_n = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}.$$

CLT, in the Averages form

Now, let us give CLT in the Averages form. Again, assume X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$. We consider the Averages

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

We Standardize $\overline{X}_n$: we denote

$$Z_n = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}.$$

Now, we use

$$\mathbb{E}(\overline{X}_n) =$$

CLT, in the Averages form

Now, let us give CLT in the Averages form. Again, assume X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$. We consider the Averages

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

We Standardize $\overline{X}_n$: we denote

$$Z_n = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}.$$

Now, we use

$$\mathbb{E}(\overline{X}_n) = \mu, \qquad Var(\overline{X}_n) =$$

CLT, in the Averages form

Now, let us give CLT in the Averages form. Again, assume X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$. We consider the Averages

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

We Standardize $\overline{X}_n$: we denote

$$Z_n = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}.$$

Now, we use

$$\mathbb{E}(\overline{X}_n) = \mu, \qquad Var(\overline{X}_n) = \frac{\sigma^2}{n}.$$

CLT, in the Averages form

Then,

$$Z_n = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

CLT, in the Averages form

Then,

$$Z_n = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

The CLT states:

CLT, in the Averages form

For any $a < b$, as $n \to \infty$,

$$\mathbb{P}(a \le Z_n \le b) \to \mathbb{P}(a \le Z \le b),$$

where $Z \sim \mathcal{N}(0,1)$,

CLT, in the Averages form

Then,

$$Z_n = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

The CLT states:

CLT, in the Averages form

For any $a < b$, as $n \to \infty$,

$$\mathbb{P}(a \leq Z_n \leq b) \to \mathbb{P}(a \leq Z \leq b),$$

where $Z \sim \mathcal{N}(0, 1)$, i.e.,

$$\mathbb{P}\left(a \leq \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \leq b\right) \to \mathbb{P}(a \leq Z \leq b).$$

Two forms of the CLT

Now, it is easy to see that the Standardization of S_n and $\overline{X}_n$ yields to the same r.v.:

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}$$

that is,

$$\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} = \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}}.$$

Two forms of the CLT

Now, it is easy to see that the Standardization of S_n and $\overline{X}_n$ yields to the same r.v.:

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}$$

that is,

$$\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} = \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}}.$$

This is because

$$\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} = \frac{n \cdot \left(\frac{S_n}{n} - \mu\right)}{\sqrt{n} \cdot \sigma} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}}.$$

Two forms of the CLT

Now, it is easy to see that the Standardization of S_n and $\overline{X}_n$ yields to the same r.v.:

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}$$

that is,

$$\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

This is because

$$\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} = \frac{n \cdot \left(\frac{S_n}{n} - \mu\right)}{\sqrt{n} \cdot \sigma} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Hence, the above two versions of CLT are the same, just one is in terms of S_n, the other one is in terms of $\overline{X}_n$.

Usage of CLT

Now, about how we are using the CLT:

Usage of CLT

Now, about how we are using the CLT: CLT says that, for the Standardized r.vs Z_n (standardized from the sum S_n or the average $\overline{X}_n$), we will have

$$\mathbb{P}(a \leq Z_n \leq b) \to \mathbb{P}(a \leq Z \leq b).$$

where $Z \sim \mathcal{N}(0,1)$.
We use this in the following form: for large n (in most of the cases, $n \geq 30$ will suffice)

$$\mathbb{P}(a \leq Z_n \leq b) \approx \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

Usage of CLT

Now, about how we are using the CLT: CLT says that, for the Standardized r.vs Z_n (standardized from the sum S_n or the average $\overline{X}_n$), we will have

$$\mathbb{P}(a \leq Z_n \leq b) \to \mathbb{P}(a \leq Z \leq b).$$

where $Z \sim \mathcal{N}(0, 1)$.
We use this in the following form: for large n (in most of the cases, $n \geq 30$ will suffice)

$$\mathbb{P}(a \leq Z_n \leq b) \approx \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

And we know how to calculate the values of Φ!

Usage of CLT

Now, about how we are using the CLT: CLT says that, for the Standardized r.vs Z_n (standardized from the sum S_n or the average $\overline{X}_n$), we will have

$$\mathbb{P}(a \leq Z_n \leq b) \to \mathbb{P}(a \leq Z \leq b).$$

where $Z \sim \mathcal{N}(0,1)$.
We use this in the following form: for large n (in most of the cases, $n \geq 30$ will suffice)

$$\mathbb{P}(a \leq Z_n \leq b) \approx \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

And we know how to calculate the values of Φ!

Now, let me give more precisely how to calculate Probabilities using the CLT.

Usage of the CLT

Example: Assume, for IID r.v.s $X_1, ..., X_n$, we want to calculate, approximately,

$$\mathbb{P}(a \leq X_1 + X_2 + ... + X_n \leq b).$$

Usage of the CLT

Example: Assume, for IID r.v.s $X_1, ..., X_n$, we want to calculate, approximately,

$$\mathbb{P}(a \leq X_1 + X_2 + ... + X_n \leq b).$$

Assume we only know $\mathbb{E}(X_k)$ and $Var(X_k)$.

Usage of the CLT

Example: Assume, for IID r.v.s $X_1, ..., X_n$, we want to calculate, approximately,

$$\mathbb{P}(a \leq X_1 + X_2 + ... + X_n \leq b).$$

Assume we only know $\mathbb{E}(X_k)$ and $Var(X_k)$. So finding the distribution of $X_1 + ... + X_n$ is impossible.

Usage of the CLT

Example: Assume, for IID r.v.s $X_1, ..., X_n$, we want to calculate, approximately,

$$\mathbb{P}(a \leq X_1 + X_2 + ... + X_n \leq b).$$

Assume we only know $\mathbb{E}(X_k)$ and $Var(X_k)$. So finding the distribution of $X_1 + ... + X_n$ is impossible. And we need to rely on the CLT.

Usage of the CLT

Example: Assume, for IID r.v.s $X_1, ..., X_n$, we want to calculate, approximately,

$$\mathbb{P}(a \le X_1 + X_2 + ... + X_n \le b).$$

Assume we only know $\mathbb{E}(X_k)$ and $Var(X_k)$. So finding the distribution of $X_1 + ... + X_n$ is impossible. And we need to rely on the CLT.

We do the following steps: let $S_n = X_1 + ... + X_n$, so we need to calculate

$$\mathbb{P}(a \le S_n \le b).$$

Usage of the CLT

Example: Assume, for IID r.v.s $X_1, ..., X_n$, we want to calculate, approximately,

$$\mathbb{P}(a \leq X_1 + X_2 + ... + X_n \leq b).$$

Assume we only know $\mathbb{E}(X_k)$ and $Var(X_k)$. So finding the distribution of $X_1 + ... + X_n$ is impossible. And we need to rely on the CLT.

We do the following steps: let $S_n = X_1 + ... + X_n$, so we need to calculate

$$\mathbb{P}(a \leq S_n \leq b).$$

The CLT is about Standardized Sum, so instead of S_n we want to have its standardized version:

Usage of the CLT

we write

$$\mathbb{P}(a \leq S_n \leq b) = \mathbb{P}\left(\frac{a - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \leq \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \leq \frac{b - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}\right);$$

Usage of the CLT

we write

$$\mathbb{P}(a \le S_n \le b) = \mathbb{P}\left(\frac{a-\mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \le \frac{S_n-\mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \le \frac{b-\mathbb{E}(S_n)}{\sqrt{Var(S_n)}}\right);$$

Now, it is the CLT turn: we approximate the last Probability by:

$$\mathbb{P}\left(\frac{a-\mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \le \frac{S_n-\mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \le \frac{b-\mathbb{E}(S_n)}{\sqrt{Var(S_n)}}\right) \approx$$

$$\approx \mathbb{P}\left(\frac{a-\mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \le Z \le \frac{b-\mathbb{E}(S_n)}{\sqrt{Var(S_n)}}\right)$$

where $Z \sim \mathcal{N}(0,1)$.

Usage of the CLT

we write

$$\mathbb{P}(a \leq S_n \leq b) = \mathbb{P}\left(\frac{a - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \leq \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \leq \frac{b - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}\right);$$

Now, it is the CLT turn: we approximate the last Probability by:

$$\mathbb{P}\left(\frac{a - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \leq \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \leq \frac{b - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}\right) \approx$$

$$\approx \mathbb{P}\left(\frac{a - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \leq Z \leq \frac{b - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}\right)$$

where $Z \sim \mathcal{N}(0,1)$. Now, we can calculate the last Probability.

Usage of the CLT

Example: Now, another example. Assume we have IID r.v.s $X_1, ..., X_n$, we have $\mathbb{E}(X_1)$ and $Var(X_1)$, and we want calculate, approximately,

$$\mathbb{P}\left(\frac{X_1 + X_2 + ... + X_n}{n} \leq b\right).$$

Usage of the CLT

Example: Now, another example. Assume we have IID r.v.s $X_1, ..., X_n$, we have $\mathbb{E}(X_1)$ and $Var(X_1)$, and we want calculate, approximately,

$$\mathbb{P}\left(\frac{X_1 + X_2 + ... + X_n}{n} \leq b\right).$$

We do not know the exact distribution of the average, so we cannot calculate by CDF, PDF or PMF of that average.

Usage of the CLT

Example: Now, another example. Assume we have IID r.v.s $X_1, ..., X_n$, we have $\mathbb{E}(X_1)$ and $Var(X_1)$, and we want calculate, approximately,

$$\mathbb{P}\left(\frac{X_1 + X_2 + ... + X_n}{n} \leq b\right).$$

We do not know the exact distribution of the average, so we cannot calculate by CDF, PDF or PMF of that average. Instead, we can use the CLT to approximate.

Usage of the CLT

Example: Now, another example. Assume we have IID r.v.s $X_1, ..., X_n$, we have $\mathbb{E}(X_1)$ and $Var(X_1)$, and we want calculate, approximately,

$$\mathbb{P}\left(\frac{X_1 + X_2 + ... + X_n}{n} \le b\right).$$

We do not know the exact distribution of the average, so we cannot calculate by CDF, PDF or PMF of that average. Instead, we can use the CLT to approximate.The steps are as follows.

Usage of the CLT

Example: Now, another example. Assume we have IID r.v.s $X_1, ..., X_n$, we have $\mathbb{E}(X_1)$ and $Var(X_1)$, and we want calculate, approximately,

$$\mathbb{P}\left(\frac{X_1 + X_2 + ... + X_n}{n} \leq b\right).$$

We do not know the exact distribution of the average, so we cannot calculate by CDF, PDF or PMF of that average. Instead, we can use the CLT to approximate.The steps are as follows.

First, we denote

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Usage of the CLT

Example: Now, another example. Assume we have IID r.v.s $X_1, ..., X_n$, we have $\mathbb{E}(X_1)$ and $Var(X_1)$, and we want calculate, approximately,

$$\mathbb{P}\left(\frac{X_1 + X_2 + ... + X_n}{n} \leq b\right).$$

We do not know the exact distribution of the average, so we cannot calculate by CDF, PDF or PMF of that average. Instead, we can use the CLT to approximate.The steps are as follows.

First, we denote

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Then we need to calculate

$$\mathbb{P}(\overline{X}_n \leq b).$$

Usage of the CLT

We need to have Standardized r.v. for the CLT, so first we write the above inequality in the following form:

$$\mathbb{P}(\overline{X}_n \leq b) = \mathbb{P}\left(\frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \leq \frac{b - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}\right).$$

Usage of the CLT

We need to have Standardized r.v. for the CLT, so first we write the above inequality in the following form:

$$\mathbb{P}(\overline{X}_n \leq b) = \mathbb{P}\left(\frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \leq \frac{b - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}\right).$$

Now, using CLT, we can write

$$\mathbb{P}\left(\frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \leq \frac{b - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}\right) \approx \mathbb{P}\left(Z \leq \frac{b - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}\right).$$

Usage of the CLT

We need to have Standardized r.v. for the CLT, so first we write the above inequality in the following form:

$$\mathbb{P}(\overline{X}_n \leq b) = \mathbb{P}\left(\frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \leq \frac{b - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}\right).$$

Now, using CLT, we can write

$$\mathbb{P}\left(\frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \leq \frac{b - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}\right) \approx \mathbb{P}\left(Z \leq \frac{b - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}\right).$$

And we can calculate $\mathbb{E}(\overline{X}_n)$, $Var(\overline{X}_n)$, and, finally, the last Probability.

Example 37.7: Assume I have a Piggy Bank. I am collecting coins worth 500 AMD in my Piggy. The mean weight of a 500 AMD coin is 5 gr with Standard Deviation 0.1 gr. I have collected 250 coins. Assume W is the weight, in grams, of all that 250 coins.

a. Calculate, approximately, $\mathbb{P}(W > 1248)$;
b. Calculate, approximately, $\mathbb{P}(1240 \leq W < 1260)$.

CLT, Example

Example 37.8: The lengths of rods produced by a machine have a mean 100 cm and standard deviation 5 cm. Find the probability that if 60 rods are randomly chosen from the machine, the mean length of the sample will be at most 101 cm.

CLT, Roughly

Let us go back again to our CLT statement. In a non-rigorous way, we can write, for large n (here $\approx$ means approximately distributed as):

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \approx \mathcal{N}(0,1) \qquad and \qquad \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \approx \mathcal{N}(0,1).$$

CLT, Roughly

Let us go back again to our CLT statement. In a non-rigorous way, we can write, for large n (here $\approx$ means approximately distributed as):

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \approx \mathcal{N}(0,1) \qquad and \qquad \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \approx \mathcal{N}(0,1).$$

Using $\mu = \mathbb{E}(X_1)$ and $\sigma^2 = Var(X_1)$, we can write

$$\frac{S_n - n\cdot\mu}{\sqrt{n}\cdot\sigma} \approx \mathcal{N}(0,1) \qquad and \qquad \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0,1).$$

CLT, Roughly

Let us go back again to our CLT statement. In a non-rigorous way, we can write, for large n (here $\approx$ means approximately distributed as):

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \approx \mathcal{N}(0,1) \qquad and \qquad \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \approx \mathcal{N}(0,1).$$

Using $\mu = \mathbb{E}(X_1)$ and $\sigma^2 = Var(X_1)$, we can write

$$\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} \approx \mathcal{N}(0,1) \qquad and \qquad \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \approx \mathcal{N}(0,1).$$

or

$$S_n \approx \mathcal{N}(n\mu, n\sigma^2) \qquad and \qquad \overline{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

CLT, Roughly -2

Let us summarize:

CLT, Roughly -2

Let us summarize:

- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and **are Normally Distributed**, i.e., $X_k \sim \mathcal{N}(\mu, \sigma^2)$, then

CLT, Roughly -2

Let us summarize:

- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and **are Normally Distributed**, i.e., $X_k \sim \mathcal{N}(\mu, \sigma^2)$, then

$$S_n \sim \mathcal{N}(n\mu, n\sigma^2) \qquad and \qquad \overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right);$$

so we know the **exact Distributions** of S_n and $\overline{X}_n$.

CLT, Roughly -2

Let us summarize:

- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and **are Normally Distributed**, i.e., $X_k \sim \mathcal{N}(\mu, \sigma^2)$, then

$$S_n \sim \mathcal{N}(n\mu, n\sigma^2) \qquad and \qquad \overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right);$$

  so we know the **exact Distributions** of S_n and $\overline{X}_n$.
- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and are **from any Distribution**, then

CLT, Roughly -2

Let us summarize:

- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and **are Normally Distributed**, i.e., $X_k \sim \mathcal{N}(\mu, \sigma^2)$, then
$$S_n \sim \mathcal{N}(n\mu, n\sigma^2) \qquad and \qquad \overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right);$$
so we know the **exact Distributions** of S_n and $\overline{X}_n$.
- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and are **from any Distribution**, then
$$S_n \approx \mathcal{N}(n\mu, n\sigma^2) \qquad and \qquad \overline{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right);$$
and we know the **asymptotic Distributions** (approximate Distributions for large n) of S_n and $\overline{X}_n$.