Probability and Statistics: Lecture-39

Monsoon-2020

by Dr. Pawan Kumar (IIIT, Hyderabad)
on November 13, 2020

» Convex Functions and Jensen's Inequality...

Definition of convex function

A function $g : I \to \mathbb{R}$ is convex if for any two points x and y in I and any $\alpha \in [0, 1]$, we have

Definition of convex function

A function $g : I \to \mathbb{R}$ is convex if for any two points x and y in I and any $\alpha \in [0, 1]$, we have

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha) g(y)$$

Definition of convex function

A function $g : I \to \mathbb{R}$ is convex if for any two points x and y in I and any $\alpha \in [0, 1]$, we have

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y)$$

If the above inequality is $\geq$, then the function g is concave.

- Here $\alpha x + (1-\alpha)y$ is the weighted average of x and y

$X \in \{x, y\}$

$g(X) \in \{g(x), g(y)\}$

$g[E[X]] \leq E[g(X)]$

Definition of convex function

A function $g : I \to \mathbb{R}$ is convex if for any two points x and y in I and any $\alpha \in [0, 1]$, we have

$$g(\underbrace{\alpha x + (1 - \alpha)y}) \leq \alpha g(x) + (1 - \alpha)g(y)$$

If the above inequality is $\geq$, then the function g is concave.

- Here $\alpha x + (1 - \alpha)y$ is the weighted average of x and y
- Here $\alpha g(x) + (1 - \alpha)g(y)$ is the weighted average of $g(x)$ and $g(y)$

$z(\alpha) = \alpha x + (1-\alpha) y \qquad \alpha \in [0,1]$

if $\alpha = 0 \Rightarrow z(0) = y$

if $\alpha = 1 \Rightarrow z(1) = x$

if $\alpha = 1/2 \Rightarrow z(1/2) = \frac{x+y}{2} \in$ mid-pt. of x & y.

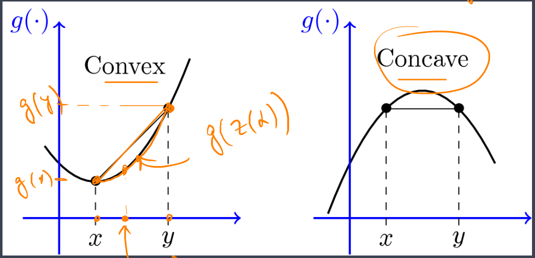
Definition of convex function

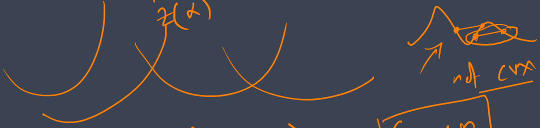
A function $g : I \to \mathbb{R}$ is convex if for any two points x and y in I and any $\alpha \in [0,1]$, we have

$$g(\alpha x + (1-\alpha) y) \leq \alpha g(x) + (1-\alpha) g(y)$$

If the above inequality is $\geq$, then the function g is concave.

- Here $\alpha x + (1-\alpha) y$ is the weighted average of x and y
- Here $\alpha g(x) + (1-\alpha) g(y)$ is the weighted average of x and y





$-(\text{Concave}) = \boxed{\text{Convex}}$

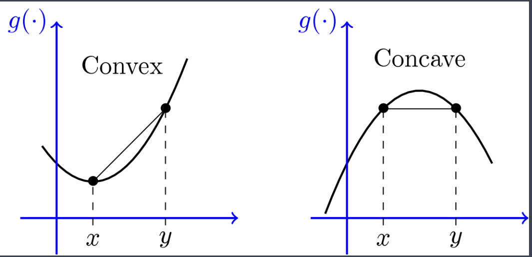
Definition of convex function

A function $g : I \to \mathbb{R}$ is convex if for any two points x and y in I and any $\alpha \in [0,1]$, we have

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha) g(y)$$

If the above inequality is $\geq$, then the function g is concave.

* Here $\alpha x + (1-\alpha)y$ is the weighted average of x and y
* Here $\alpha g(x) + (1-\alpha) g(y)$ is the weighted average of x and y



* From the definition of convexity on left, we conclude

$$E[g(X)] \geq g(E[X])$$

» Jensen's Inequality...

Jensen's inequality

If $g(x)$ is a convex function on R_X, and $E[g(X)]$ and $g(E[X])$ are finite, then

Jensen's inequality

If $g(x)$ is a convex function on R_X, and $E[g(X)]$ and $g(E[X])$ are finite, then

$$E[g(X)] \geq g(E[X]).$$

Jensen's inequality

If $g(x)$ is a convex function on R_X, and $E[g(X)]$ and $g(E[X])$ are finite, then

$$E[g(X)] \geq g(E[X]).$$

* To know whether a function is convex,

Jensen's inequality

If $g(x)$ is a convex function on R_X, and $E[g(X)]$ and $g(E[X])$ are finite, then

$$E[g(X)] \geq g(E[X]).$$

* To know whether a function is convex, a useful method for differentiable function is second derivative test:

Jensen's inequality

If $g(x)$ is a convex function on R_X, and $E[g(X)]$ and $g(E[X])$ are finite, then

$$E[g(X)] \geq g(E[X]).$$

* To know whether a function is convex, a useful method for differentiable function is second derivative test: A twice differentiable function $g : I \to \mathbb{R}$ is convex if and only if $g''(x) \geq 0$ for all $x \in I$

Jensen's inequality

If $g(x)$ is a convex function on R_X, and $E[g(X)]$ and $g(E[X])$ are finite, then

$$E[g(X)] \geq g(E[X]).$$

* To know whether a function is convex, a useful method for differentiable function is second derivative test: A twice differentiable function $g : I \to \mathbb{R}$ is convex if and only if $g''(x) \geq 0$ for all $x \in I$
* For example, $g(x) = x^2$ is convex in $\mathbb{R}$

$g'(x) = 2x$, $\rightarrow g''(x) = 2 > 0 \quad \forall x \in \mathbb{R}$

$\Rightarrow g$ is cvx

» Application of Jensen's Inequality...

Example (Jensen's Inequality)

Consider a random variable X with $E[X] = 10$, and X being positive.

Example (Jensen's Inequality)

Consider a random variable X with $E[X] = 10$, and X being positive. Estimate the following quantities

Example (Jensen's Inequality)

Consider a random variable X with $E[X] = 10$, and X being positive. Estimate the following quantities

1. $E[\frac{1}{X+1}]$

Example (Jensen's Inequality)

Consider a random variable X with $E[X] = 10$, and X being positive. Estimate the following quantities

1. $E[\frac{1}{X+1}]$
2. $E[e^{\frac{1}{X+1}}]$

Example (Jensen's Inequality)

Consider a random variable X with $E[X] = 10$, and X being positive. Estimate the following quantities

1. $E[\frac{1}{X+1}]$
2. $E[e^{\frac{1}{X+1}}]$
3. $E[\ln \sqrt{X}]$

$g(x) = \frac{1}{x+1}$

» Answer to previous problem...

a) $g(x) = \frac{1}{x+1}$, $g'(x) = \frac{-1}{(x+1)^2}$

$g''(x) = \frac{2}{(1+x)^3} > 0$ for $x > 0$

$\Rightarrow g$ cvx on $(0, \infty)$.

Using Jensen's ineq

$E\left[\frac{1}{X+1}\right] \geq g(E[X]) = \frac{1}{E[X]+1}$

$= \frac{1}{10+1} = \frac{1}{11}$

$E[g(X)] \geq g(E[X])$

(b) $E\left[e^{\frac{1}{X+1}}\right]$, $g(x) = e^{\frac{1}{x+1}}$

$g'(x) = e^{\frac{1}{x+1}} \cdot \frac{-1}{(x+1)^2}$. tedious

try something else

Let, $h(x) = e^{x}$
$g(x) = \frac{1}{x+1}$

Observation: (1) $h(x)$ is cvx, non-decreasing

(2) $g(x)$ is convex

$f(x) = h(g(x)) = e^{\frac{1}{x+1}}$

» Answer to previous problem...

$f'(x) = h' \cdot g'$ (Using chain rule of diff)

$= \frac{dh}{dg} \cdot \frac{dg}{dx}$

$f''(x) = \frac{dh}{dg} \cdot \frac{d^2g}{dx^2} + \frac{d^2h}{dg^2} \cdot \left[\frac{dg}{dx}\right]^2$

$\frac{dh}{dg} \geq 0$ (Since h non decr.), $\frac{d^2g}{dx^2} \geq 0$ (Since g is cvx), $\frac{d^2h}{dg^2} \geq 0$ (Since h is cvx), $\left[\frac{dg}{dx}\right]^2 \geq 0$ (obv.)

≥ 0

$E\left[e^{\frac{1}{x+1}}\right] \geq e^{\frac{1}{E[x]+1}}$

$= \frac{1}{e^{11}}$

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be n i.i.d. random variables,

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be n i.i.d. random variables, then the sample mean $\overline{X}$ is defined as follows

Pick n sample

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be n i.i.d. random variables, then the sample mean $\overline{X}$ is defined as follows

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

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be n i.i.d. random variables, then the sample mean $\overline{X}$ is defined as follows

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

It is easy to establish the following:

$E[X_1] = E[X_2] = \cdots = E[X_n] = E[X]$

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be n i.i.d. random variables, then the sample mean $\overline{X}$ is defined as follows

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

$E[\bar{X}] = \frac{1}{n}(E[X_1] + \cdots + E[X_n])$

$= \frac{\cancel{n} E[X]}{\cancel{n}} = E[X]$

It is easy to establish the following:

1. $E[\overline{X}] = E[X]$

$Var(\underline{a}X) = a^2 \, Var(X)$

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be n i.i.d. random variables, then the sample mean $\overline{X}$ is defined as follows

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

It is easy to establish the following:

1. $E[\overline{X}] = E[X]$
2. $\mathrm{Var}(\overline{X}) = \dfrac{\mathrm{Var}(X)}{n}$

» Answer to previous problem...

» Weak Law of Large Numbers...

Weak Law of Large Numbers

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean $E[X_i] = \mu < \infty$. Then for any $\epsilon > 0$,

Weak Law of Large Numbers

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean $E[X_i] = \mu < \infty$. Then for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

Proof. Assume $Var(X) = \sigma^2$ is finite $(< \infty)$

$$P(|\bar{x} - \mu| \geq \epsilon) \overset{\text{Chebyshev}}{\leq} \frac{Var(\bar{X})}{\epsilon^2} = \frac{Var(X)}{n\epsilon^2}$$

$\left(\text{Since } Var(\bar{X}) = \frac{Var(X)}{n}\right)$

$$\Rightarrow \lim_{n\to\infty} P(|\bar{x} - \mu| \geq \epsilon) = 0$$

» Central Limit Theorem...

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with expected value $E[X_i] = \mu < \infty$ and variance $0 < \text{Var}(X_i) = \sigma^2 < \infty$. Then, the random variable

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with expected value $E[X_i] = \mu < \infty$ and variance $0 < \text{Var}(X_i) = \sigma^2 < \infty$. Then, the random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with expected value $E[X_i] = \mu < \infty$ and variance $0 < \text{Var}(X_i) = \sigma^2 < \infty$. Then, the random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal random variable as n goes to infinity, that is

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with expected value $E[X_i] = \mu < \infty$ and variance $0 < \text{Var}(X_i) = \sigma^2 < \infty$. Then, the random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal random variable as n goes to infinity, that is

$$\lim_{n\to\infty} P(X_n \le x) = \Phi(x), \quad \text{for all } x \in \mathbb{R},$$

CDF of Std. Normal

where $\Phi(x)$ is the standard normal CDF.

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with expected value $E[X_i] = \mu < \infty$ and variance $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$. Then, the random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal random variable as n goes to infinity, that is

$$\lim_{n\to\infty} P(X_n \leq x) = \Phi(x), \quad \text{for all } x \in \mathbb{R},$$

where $\Phi(x)$ is the standard normal CDF.

* It does not matter what the distribution of X_i is

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with expected value $E[X_i] = \mu < \infty$ and variance $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$. Then, the random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal random variable as n goes to infinity, that is

$$\lim_{n\to\infty} P(X_n \leq x) = \Phi(x), \quad \text{for all } x \in \mathbb{R},$$

where $\Phi(x)$ is the standard normal CDF.

- It does not matter what the distribution of X_i is
- The X_i can be discrete, continuous, or mixed random variables

1. Let X_i be Bernoulli(p)
2. Then $E[X_i] = p, \text{Var}(X_i) = p(1-p)$
3. $Y_n = X_1 + X_2 + \cdots + X_n$ has Binomial((n, p))
4. Hence,

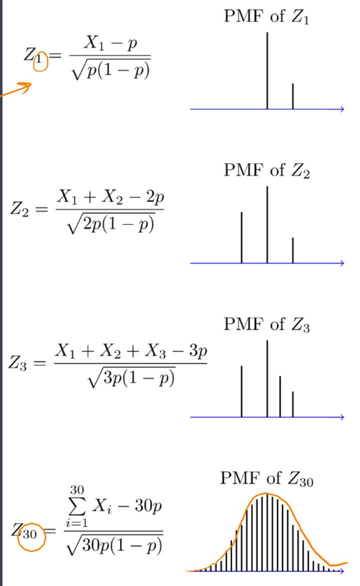
$$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}}$$

5. The figure on the right shows PMF of Z_n for different values of n
6. As we observe, the shape of PMF gets closer to a normal PDF

1. Let X_i be Bernoulli(p)
2. Then $E[X_i] = p, \mathrm{Var}(X_i) = p(1-p)$
3. $Y_n = X_1 + X_2 + \cdots + X_n$ has Binomial((n, p))
4. Hence,

$$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}}$$

5. The figure on the right shows PMF of Z_n for different values of n
6. As we observe, the shape of PMF gets closer to a normal PDF



1. Let X_i be $\text{Uniform}(0, 1)$

1. Let X_i be Uniform$(0, 1)$
2. Then $E[X_i] = 1/2, Var(X_i) = 1/12$

1. Let X_i be Uniform$(0,1)$
2. Then $E[X_i] = 1/2, \mathit{Var}(X_i) = 1/12$
3. Hence,

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n - n/2}{\sqrt{n/12}}$$

1. Let X_i be Uniform$(0, 1)$
2. Then $E[X_i] = 1/2, \mathit{Var}(X_i) = 1/12$
3. Hence,

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n - n/2}{\sqrt{n/12}}$$

4. The figure on the right shows PMF of Z_n for different values of n

1. Let X_i be Uniform$(0, 1)$
2. Then $E[X_i] = 1/2, Var(X_i) = 1/12$
3. Hence,

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n - n/2}{\sqrt{n/12}}$$

4. The figure on the right shows PMF of Z_n for different values of n
5. As we observe, the shape of PMF gets closer to a normal PDF

1. Let X_i be Uniform$(0, 1)$
2. Then $E[X_i] = 1/2, \mathit{Var}(X_i) = 1/12$
3. Hence,

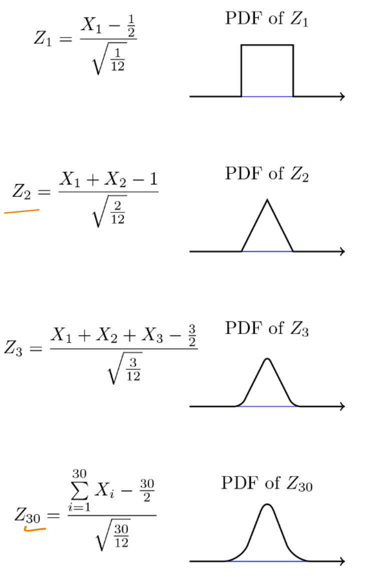
$$Z_n = \frac{X_1 + X_2 + \cdots + X_n - n/2}{\sqrt{n/12}}$$

4. The figure on the right shows PMF of Z_n for different values of n
5. As we observe, the shape of PMF gets closer to a normal PDF

1. Let X_i be Uniform$(0,1)$
2. Then $E[X_i] = 1/2, \text{Var}(X_i) = 1/12$
3. Hence,

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n - n/2}{\sqrt{n/12}}$$

4. The figure on the right shows PMF of Z_n for different values of n
5. As we observe, the shape of PMF gets closer to a normal PDF



How to Apply The Central Limit Theorem (CLT)

1. Write the random variable of interest, Y, as the sum of n i.i.d. random variables
$$Y = X_1 + X_2 + \cdots + X_n$$

» How to Apply The Central Limit Theorem (CLT)?

How to Apply The Central Limit Theorem (CLT)

1. Write the random variable of interest, Y, as the sum of n i.i.d. random variables
$$Y = X_1 + X_2 + \cdots + X_n$$

2. Find $E[Y]$ and $\mathrm{Var}(Y)$ by noting that
$$E[Y] = n\mu, \quad \mathrm{Var}(Y) = n\sigma^2, \quad \text{where } \mu = E[X_i],\ \sigma^2 = \mathrm{Var}(X_i)$$

How to Apply The Central Limit Theorem (CLT)

1. Write the random variable of interest, Y, as the sum of n i.i.d. random variables
$$Y = X_1 + X_2 + \cdots + X_n$$
2. Find $E[Y]$ and $\mathrm{Var}(Y)$ by noting that
$$E[Y] = n\mu, \quad \mathrm{Var}(Y) = n\sigma^2, \quad \text{where } \mu = E[X_i],\ \sigma^2 = \mathrm{Var}(X_i)$$
3. From CLT, we conclude that

» How to Apply The Central Limit Theorem (CLT)?

How to Apply The Central Limit Theorem (CLT)

1. Write the random variable of interest, Y, as the sum of n i.i.d. random variables
$$Y = X_1 + X_2 + \cdots + X_n$$
2. Find $E[Y]$ and Var(Y) by noting that
$$E[Y] = n\mu, \quad \mathrm{Var}(Y) = n\sigma^2, \quad \text{where } \mu = E[X_i],\ \sigma^2 = \mathrm{Var}(X_i)$$
3. From CLT, we conclude that
$$\frac{Y - E[Y]}{\sqrt{\mathrm{Var}(Y)}} = \frac{Y - n\mu}{\sqrt{n}\sigma}$$
is approximately standard normal, hence, we have

How to Apply The Central Limit Theorem (CLT)

1. Write the random variable of interest, Y, as the sum of n i.i.d. random variables

$$Y = X_1 + X_2 + \cdots + X_n$$

2. Find $E[Y]$ and $\mathrm{Var}(Y)$ by noting that

$$E[Y] = n\mu, \quad \mathrm{Var}(Y) = n\sigma^2, \quad \text{where } \mu = E[X_i],\ \sigma^2 = \mathrm{Var}(X_i)$$

3. From CLT, we conclude that

$$\frac{Y - E[Y]}{\sqrt{\mathrm{Var}(Y)}} = \frac{Y - n\mu}{\sqrt{n}\sigma}$$

is approximately standard normal, hence, we have

$$P(y_1 \le Y \le y_2) = P\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma} \le \underbrace{\frac{Y - n\mu}{\sqrt{n}\sigma}}_{\text{Std. Normal}} \le \frac{y_2 - n\mu}{\sqrt{n}\sigma}\right) \approx \Phi\left(\frac{y_2 - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma}\right)$$

(approx)

» Example of Application of CLT...

» Example of Application of CLT...

Example (Applications of CLT)

A bank teller serves customers standing in the queue one by one.

Example (Applications of CLT)

A bank teller serves customers standing in the queue one by one. Suppose that the service time X_i for customer i has mean $E[X_i] = 2$ (minutes) and $Var(X_i) = 1$.

Example (Applications of CLT)

A bank teller serves customers standing in the queue one by one. Suppose that the service time X_i for customer i has mean $E[X_i] = 2$ (minutes) and $Var(X_i) = 1$. We assume that service times for different bank customers are independent.

Example (Applications of CLT)

A bank teller serves customers standing in the queue one by one. Suppose that the service time X_i for customer i has mean $E[X_i] = 2$ (minutes) and $Var(X_i) = 1$. We assume that service times for different bank customers are independent. Let Y be the total time the bank teller spends serving 50 customers.

Example (Applications of CLT)

A bank teller serves customers standing in the queue one by one. Suppose that the service time X_i for customer i has mean $E[X_i] = 2$ (minutes) and $Var(X_i) = 1$. We assume that service times for different bank customers are independent. Let Y be the total time the bank teller spends serving 50 customers. Find $P(90 < Y < 110)$.

» Answer to previous problem...

We have

$$Y = X_1 + X_2 + \cdots + X_n,$$

where $n = 50$, $E[X_i] = \mu = 2$ (given)

and $Var(X_i) = \sigma^2 = 1$ (given)

$$P(90 < Y < 110)$$

$$= P\left(\frac{90 - 50\cdot 2}{\sqrt{50}\cdot 1} < \frac{Y - n\mu}{\sqrt{n}\,\sigma} < \frac{110 - 50\cdot 2}{\sqrt{50}\cdot 1}\right)$$

$$= P\left(\frac{-10}{\sqrt{50}} < \frac{Y - n\mu}{\sqrt{n}\,\sigma} < \frac{10}{\sqrt{50}}\right)$$

$$= \underline{\Phi}\left(\frac{10}{\sqrt{50}}\right) - \underline{\Phi}\left(\frac{-10}{\sqrt{50}}\right)$$

» Example of Application of CLT...

Example (Applications of CLT)

In a communication system each data packet consists of 1000 bits.

Example (Applications of CLT)

In a communication system each data packet consists of 1000 bits. Due to the noise, each bit may be received in error with probability 0.1.

Example (Applications of CLT)

In a communication system each data packet consists of 1000 bits. Due to the noise, each bit may be received in error with probability 0.1. It is assumed bit errors occur independently.

X_i : R.V. for the ith bit

$X_i = 1$ if the ith bit is received

$X_i = 0$ otherwise.

Example (Applications of CLT)

In a communication system each data packet consists of 1000 bits. Due to the noise, each bit may be received in error with probability 0.1. It is assumed bit errors occur independently. Find the probability that there are more than 120 errors in a certain data packet.

X_i's are i.i.d $\quad X_i \sim$ Bernoulli $(p = 0.1)$

$E[X_i] = \mu = p = 0.1$, $\quad Var(X_i) = \sigma^2 = p(1-p) = 0.09$

$Y = X_1 + X_2 + \dots + X_n$,

Using CLT: $\rightarrow P(Y > 120) = P\left(\frac{Y - n\mu}{\sqrt{n}\,\sigma} > \frac{120 - n\mu}{\sqrt{n}\,\sigma}\right)$

$= 1 - P\left(\underbrace{\frac{Y - n\mu}{\sqrt{n}\,\sigma}}_{\text{std norm}} \le \frac{120 - n\mu}{\sqrt{n}\,\sigma}\right) = 1 - \left[\Phi\left(\frac{120 - 1000 \cdot 0.1}{\sqrt{1000} \cdot \sqrt{0.09}}\right)\right]$