Probability Distribution

F. M. Arifur Rahman

Senior Lecturer, Department of Mathematical & Physical Sciences

Contents

Random Variable

A random variable is a variable that takes on **numerical values** as a result of a random experiment or measurement; associates a numerical value with each possible outcome.

The differences between variable and random variable are-

- Random variable always takes **numerical values**
- There is a **probability associated** with each possible values

Random variable is denoted by capital letters such as X, Y, Z etc.

And the possible outcomes are denoted by small letters such as x, y, z etc.

Random Variable

Example 1:

A coin is tossed. It has two possible outcomes- Head and Tail.

Consider a variable, X= outcome of a coin toss= $\begin{cases} H, & if Head\ appears \\ T, & if\ Tail\ appears \end{cases}$

Here, S= {H, T}.

But, these are not numerical values.

Random Variable

Example 1(contd.):

Consider a variable, X= Number of heads obtained in a trial

Then, $X = \begin{cases} 1, & if\ Head\ appears \\ 0, & if\ Tail\ appears \end{cases}$

For a fair coin, we can write, P(X=1) = ½ and P(X=0) = ½

So, X is a random variable.

Random Variable

Types of random variable:

Random Variable

- **Discrete Random Variable**

  A random variable defined over a discrete sample space

- **Continuous Random Variable**

  A random variable defined over a continuous sample space

Random Variable

Examples:

Discrete Random Variable:

1. X= Number of correct answers in a 100-MCQ test= $0, 1, 2, \ldots, 100$
2. X= Number of cars passing a toll both in a day= $0, 1, 2, \ldots, \infty$
3. X= Number of balls required to take the first wicket = $1, 2, 3, \ldots, \infty$

Continuous Random Variable:

1. X= Weight of a person. $0<X<\infty$
2. X= Monthly Profit. $-\infty<X<\infty$

Probability Distributions

Distribution of the probabilities among the different values of a random variable.

Discrete probability distribution- probability distribution of a discrete random variable

Continuous probability distribution- probability distribution of a continuous random variable

Probability Distributions

Examples:

Discrete probability distribution-

- Tossing a coin 2 times.

X= Number of Heads appeared

S= {HH, HT, TH, TT}

x	0	1	2
P(x)	¼	2/4	¼

Probability Distributions

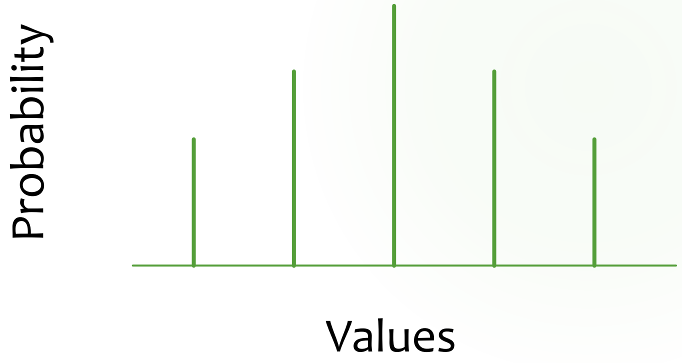
Different types of probability distributions:

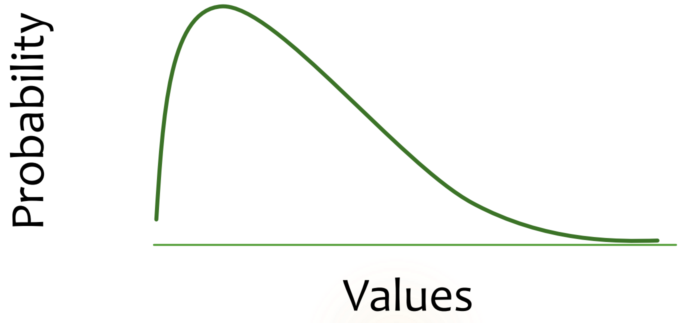
Discrete probability distribution-

1. Bernoulli Distribution
2. *Binomial Distribution*
3. Poisson Distribution etc.

Continuous probability distribution-

1. Uniform Distribution
2. *Normal Distribution*
3. Exponential Distribution
4. t-distribution etc.





PMF and PDF

Probability Mass Function (pmf)- the probability distribution function of a discrete random variable X is called a pmf and is denoted by p(x)

Probability Density Function (pdf)- the probability distribution function of a continuous random variable X is called a pdf and is denoted by f(x)

Mathematical Expectations

- For a discrete random variable X with pmf p(x), the mathematical expectation of X is-

$$\mu = E(X) = \sum_{x} x\, p(x)$$

- For a continuous random variable X with pdf f(x), the mathematical expectation of X is-

$$\mu = E(X) = \int_{x} x\, f(x)$$

Mathematical expectation is also known as population mean or expected value.

Mathematical Expectations

$$E(X^2) = \begin{cases} \sum_x x^2\, p(x) & , if\ x\ is\ a\ discrete\ r.v. \\ \int_x x^2\, f(x) & , if\ x\ is\ a\ continuous\ r.v. \end{cases}$$

Variance:

$$\sigma^2 = Var(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$$

Standard deviation: $\sigma = \sqrt{Var(X)}$

Properties of Mathematical Expectations

Let, c is a constant number

X and Y are two independent random variables

1. $E(c) = c$
2. $E(c\,X) = c\,E(x)$
3. $E(X + c) = E(x) + c$
4. $E(X+Y) = E(X) + E(Y)$
5. $E(X-Y) = E(X) - E(Y)$
6. $E(XY) = E(X)\,.\,E(Y)$

1. $Var(c) = 0$
2. $Var(c\,X) = c^2\,Var(x)$
3. $Var(X + c) = Var(x)$
4. $Var(X+Y) = Var(X) + Var(Y)$
5. $Var(X-Y) = Var(X) + Var(Y)$

Mathematical Expectation

Example 2-

A company estimates the net profit on a new product, it is launching, to be Rs. 3 million during first year, if it is 'successful', Rs. 1 million if it is 'moderately successful', and a loss of Rs. 1 million if it is 'unsuccessful'.

The company assigns the following probabilities to first year prospects for the product-

Successful: 0.25, Moderately successful: 0.40, and Unsuccessful: 0.35

What are the **expected value** and **standard deviation** of the first year net profit for the product? Also, find the expected value of net profit if there is a fixed cost of Rs. 0.2 million, whatever the success status is.

Mathematical Expectation

Solution-

Let,

X= Net profit on the new product in the 1^{st} year (Rs. Million)

Given that,

x	3	1	-1
P(x)	0.25	0.4	0.35

Expected net profit, $E(X) = \sum x\, p(x) = (3 * 0.25) + (1 * 0.4) + (-1 * 0.35)$
$= 0.8\ million$

Mathematical Expectation

Solution **(contd.)-**

$$E(X^2) = \sum x^2\, p(x) = (3^2 * 0.25) + (1^2 * 0.4) + \left((-1)^2 * 0.35\right)$$
$$= (9 * 0.25) + (1 * 0.4) + (1 * 0.35) = 3$$

$$Var(X) = E(X^2)\ - [E(X)]^2 = 3 - 0.8^2 = 2.36$$
$$\therefore SD(X) = \sqrt{Var(X)} = \sqrt{2.36} = 1.54\, million$$

If there is a fixed cost of Rs. 0.2 million, then expected net profit-

$$E(X - 0.2) = E(X) - 0.2 = 0.8 - 0.2 = 0.6\, million$$

Binomial Distribution

Bernoulli trial:

A trial that has only two possible outcomes (often called 'Success' and 'Failure')

Outcome	Success	failure
Probability	p	1-p

Let,

- n independent Bernoulli trials are performed
- Each trial has the same probability of success, p

Binomial Distribution

Let,

X= number of success in n trials

Then, X is a binomial random variable with distribution function (pmf),

$$p(x) = {}^{n}C_{x}\, p^{x}\, (1-p)^{n-x} \qquad ;\ x = 0, 1, 2, \dots, n$$
$$= \frac{n!}{(n-x)!\ x!}\ p^{x}\, (1-p)^{n-x}$$

Here, n!= n(n-1)(n-2) … 1 0!=1 1!=1 2!= 2*1= 2 3!= 3*2*1=6

We write it as, $X \sim binomial\ (n, p)$

Binomial Distribution

Mean of the binomial distribution, $\mu = E(X) = \sum x\, p(x) = np$

Variance of binomial distribution, $\sigma^2 = E(X^2) - \mu^2 = np(1-p) = npq$

Standard deviation of binomial distribution, $\sigma = \sqrt{npq}$

Binomial Distribution

Example 3:

There are 3 multiple choice questions in a MCQ test. Each MCQ consists of four possible choices and only one of them is correct. If an examinee answers those MCQ randomly (without knowing the correct answers)

a. What is the probability that exactly any two of the answers will be correct?

b. What is the probability that at least two of the answers will be correct?

c. What is the probability that at most two of the answers will be correct?

d. What will be the average or expected number of correct answers?

e. Also, find the standard deviation of number of correct answers.

Binomial Distribution

Solution:

Let,

X= number of correct answers selected in 3 MCQs

Here, p = probability of selecting correct answer per question = ¼ = 0.25

$$\therefore X \sim binomial\ (n = 2, p = 0.25)$$

$$p(x) = {}^{3}C_{x}\ (0.25)^{x}\ (1 - 0.25)^{3-x} \qquad ;\ x = 0, 1, 2, 3$$

$$= \frac{3!}{(3 - x)!\ x!} (0.25)^{x}\ (0.75)^{n-x}$$

Binomial Distribution

Solution (contd.)**:**

a. probability that exactly any two of the answers will be correct-

$$P(X = 2) = \frac{3!}{(3-2)!\,2!}(0.25)^2\,(0.75)^{3-2}$$

$$= \frac{3!}{1!\,2!}(0.25)^2\,(0.75)^1 = \frac{3*2*1}{1*(2*1)} * 0.0625 * 0.75 = 0.141$$

b. probability that at least two of the answers will be correct-

$$P(X \geq 2) = P(X = 2) + P(X = 3)$$

$$= \frac{3!}{(3-2)!\,2!}(0.25)^2\,(0.75)^{3-2} + \frac{3!}{(3-3)!\,3!}(0.25)^3\,(0.75)^{3-3}$$

$$= \frac{3!}{1!\,2!}(0.25)^2\,(0.75)^1 + \frac{3!}{0!\,3!}(0.25)^3\,(0.75)^0 = 0.141 + 0.016 = 0.157$$

Binomial Distribution

Solution **(contd.):**

c. probability that at most two of the answers will be correct-

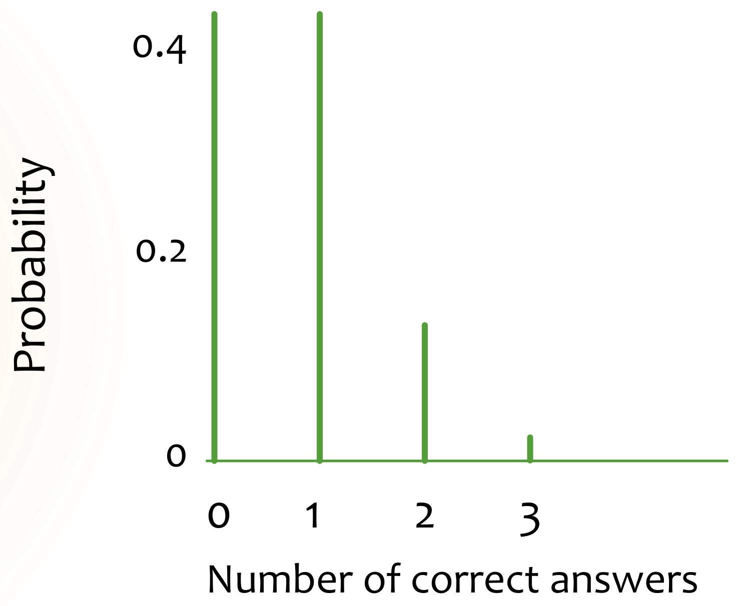
$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$
$$= \frac{3!}{(3-0)!\,0!}(0.25)^0\,(0.75)^{3-0} + \frac{3!}{(3-1)!\,1!}(0.25)^1\,(0.75)^{3-1}$$
$$+ \frac{3!}{(3-2)!\,2!}(0.25)^2\,(0.75)^{3-2}$$
$$= \frac{3!}{3!\,0!}(0.25)^0\,(0.75)^3 + \frac{3!}{2!\,1!}(0.25)^1\,(0.75)^2 + \frac{3!}{1!\,2!}(0.25)^2\,(0.75)^1$$
$$= 0.422 + 0.422 + 0.141 = 0.985$$

d. $E(X) = np = 3 * .25 = 0.75$

e. $SD(X) = \sqrt{npq} = \sqrt{3 * 0.25 * 0.75} = \ 0.75$

Binomial Distribution

X	0	1	2	3
P(x)	0.422	0.422	0.141	0.016



Poisson Distribution

Let,

X= a random variable usually counts or number of occurrences

Then, X is a Poisson random variable with distribution function (pmf),

$$p(x) = \frac{e^{-\lambda}\,\lambda^{x}}{x!} \qquad ;\ x = 0, 1, 2, \ldots$$

We write it as, $X \sim Poisson\ (\lambda)$

Poisson Distribution

Mean of the Poisson distribution, $\mu = E(X) = \sum x\, p(x) = \lambda$

Variance of Poisson distribution, $\sigma^2 = E(X^2) - \mu^2 = \lambda$

Standard deviation of Poisson distribution, $\sigma = \sqrt{\lambda}$

Poisson Distribution

Example 4:

The average number of errors on a page of a certain magazine is 0.2. What is the probability that the next page (or a randomly selected page) you read contains

i. 0 (zero) error?

ii. 2 or more errors?

iii. What is the average error per page?

iv. Also, find standard deviation of the number of errors.

Poisson Distribution

Solution:

Let,

X= number of errors in a page

Here, λ = average number of errors per page= 0.2

$$\therefore X \sim Poisson\ (\lambda = 0.2)$$

$$p(x) = \frac{e^{-\lambda}\ \lambda^{x}}{x!} \quad ;\ x = 0, 1, 2, \ldots$$

$$= \frac{e^{-0.2}\ 0.2^{x}}{x!}$$

Poisson Distribution

Solution:

i. $P(X = 0) = \frac{e^{-0.2}\ 0.2^{x}}{x!} = \frac{e^{-0.2}\ 0.2^{0}}{0!} = \frac{e^{-0.2}\ *1}{1} = 0.8187$

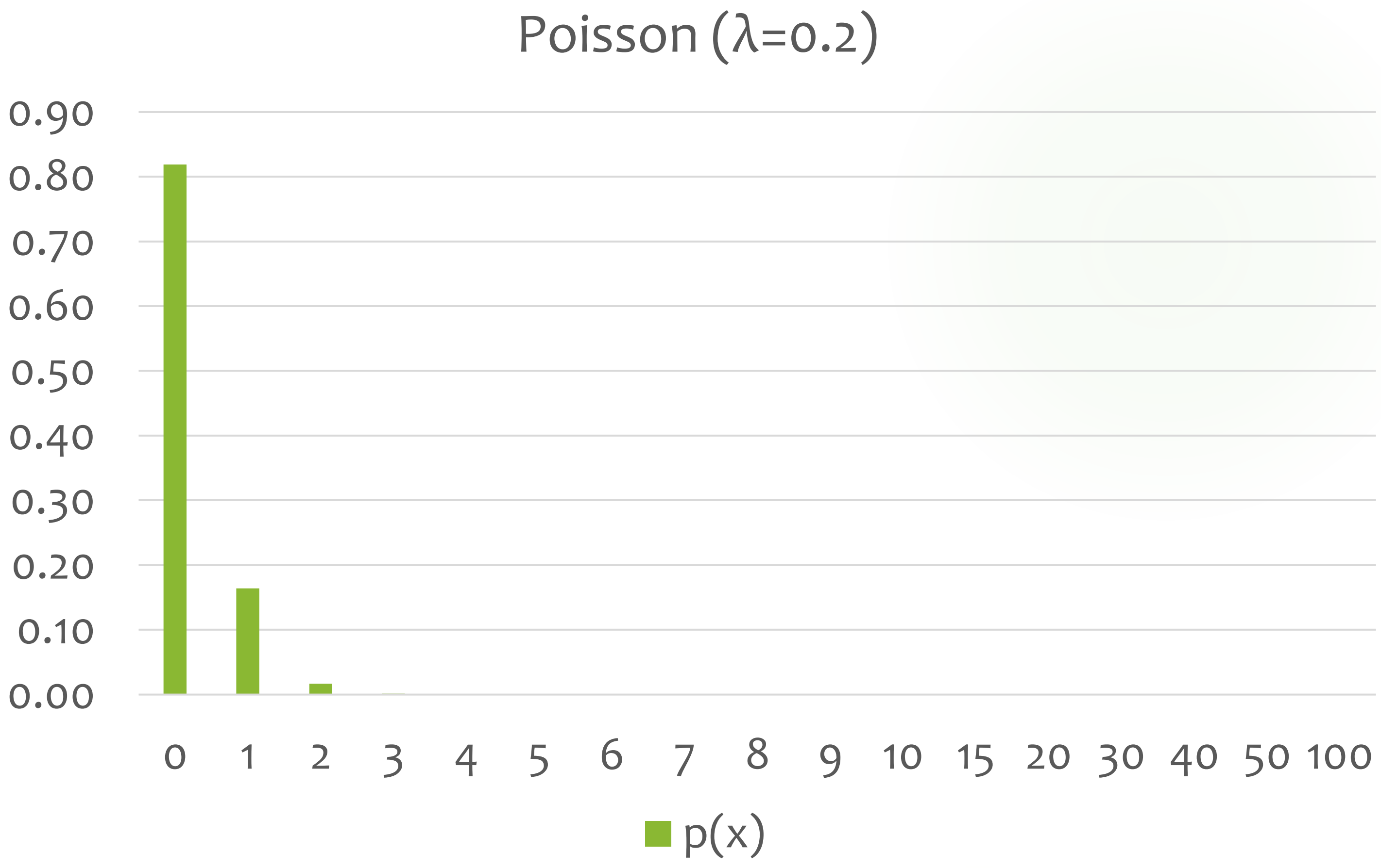
ii. $P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)]$
$= \frac{e^{-0.2}\ 0.2^{0}}{0!} + \frac{e^{-0.2}\ 0.2^{1}}{1!} = 1 - [e^{-0.2} + e^{-0.2}\ * 0.2] = 0.01756$

iii. Average number of errors, $E(X) = \lambda = 0.2$

iv. Standard deviation, $SD(X) = \sqrt{\lambda} = \sqrt{0.2} = 0.45$

Poisson Distribution

x	p(x)
0	0.82
1	0.16
2	0.02
3	0.00
4	0.00
5	0.00
6	0.00
7	0.00
8	0.00
9	0.00
10	0.00
15	0.00
20	0.00
30	0.00
40	0.00
50	0.00
100	0.00



Uniform Distribution

Let,

X is a continuous random variable

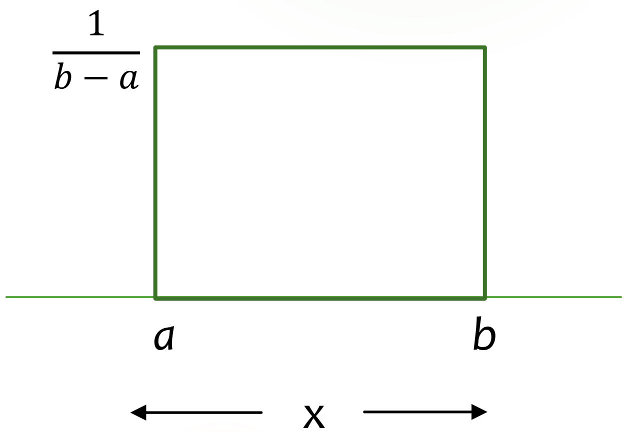
Then, if X has a probability density function (pdf),

$$f(x) = \frac{1}{b-a} \quad ; \quad a < x < b$$

We write it as, $X \sim uniform\,(a, b)$

Mean, $E(X) = \frac{a+b}{2}$

Variance, $V(X) = \frac{(b-a)^2}{12}$



Uniform Distribution

Example 5:

The waiting time (in minutes) for train is uniform (10, 50).

Find-

a. The probability that you have to wait at least 20 minutes.
b. Average waiting time.
c. Standard deviation of waiting time.

Uniform Distribution

Solution:

Let,

X= waiting time (in minutes)

$$\therefore X \sim uniform\ (10, 50)$$

$$f(x) = \frac{1}{b-a} \quad ; \ a < x < b$$

$$= \frac{1}{50-10}; 10 < x < 50$$

$$= \frac{1}{40}$$

Uniform Distribution

Solution:

a. Probability that you have to wait at least 20 minutes-

$$P[X \geq 20] = \int_{20}^{50} f(x)\, dx = \int_{20}^{50} \frac{1}{40}\, dx = \frac{1}{40}\int_{20}^{50} 1\, dx$$

$$= \frac{1}{40}\, [x]_{20}^{50} = \frac{1}{40}\, [50 - 20] = \frac{30}{40} = \frac{3}{4} = 0.75$$

b. Average waiting time-

$$E(X) = \frac{a+b}{2} = \frac{10+50}{2} = \frac{60}{2} = 30\, minutes$$

Uniform Distribution

Solution:

c. Standard deviation of waiting time-

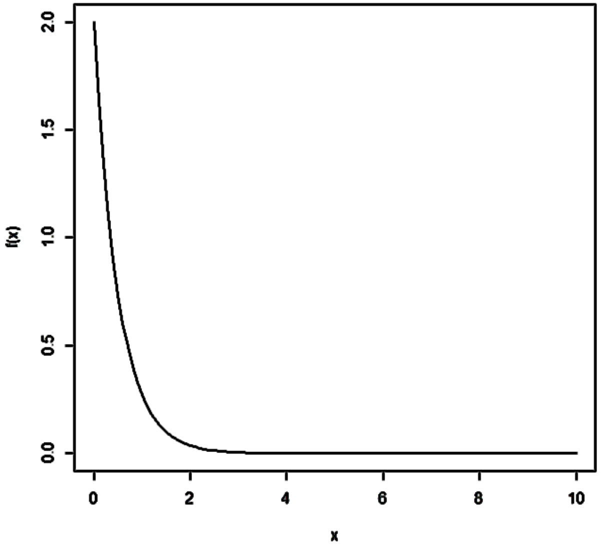
$$SD(X) = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(50-10)^2}{12}} = \sqrt{\frac{(40)^2}{12}} = 3.65\ minutes$$

Exponential Distribution

A continuous random variable X is said to follow exponential distribution, if its pdf is,

$$f(x) = \lambda e^{-\lambda x}; \qquad x > 0, where\ \lambda > 0$$

We write as, $X \sim \exp(\lambda)$



Here, X is usually time until certain event occurs.

Mean, $E(X) = \frac{1}{\lambda}$

Variance, $V(X) = \frac{1}{\lambda^2}$

Exponential Distribution

Example 6:

Average time required to repair a machine is 0.5 hours. What is the probability that the next repair will take more than 2 hours?

Solution:

Let, X= time required to repair the machine

$$\therefore X \sim \exp(\lambda = 2) \qquad \left[since, E(X) = \frac{1}{\lambda} = 0.5 \Rightarrow \lambda = \frac{1}{0.5} = 2\right]$$

$$\Pr[X > 2] = \int_2^\infty \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_2^\infty = \left[-0 + e^{-2\lambda}\right] = e^{-2\lambda} = e^{-2\times 2}$$

$$= e^{-4} = 0.0183$$

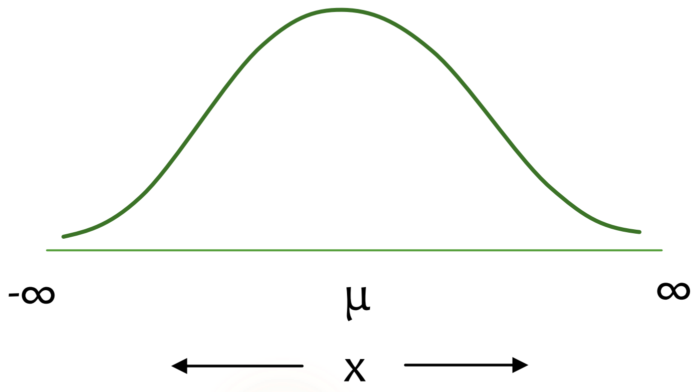
Normal Distribution

Let,

X is a continuous random variable

Then, if X has a probability density function (pdf),

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad ; \quad -\infty < x < \infty$$



We write it as, $X \sim N\,(\mu, \sigma^2)$

Mean, $E(X) = \mu$

Variance, $V(X) = \sigma^2$

Standard Normal Distribution

Let,

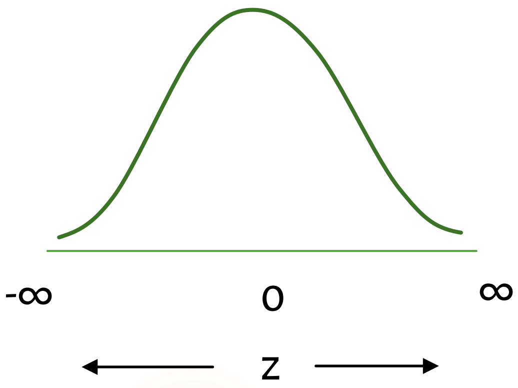
$$Z = \frac{X - \mu}{\sigma}$$

Then, Mean, $E(Z) = 0$

Variance, $V(Z) = 1$

And, if Z has a probability density function (pdf),

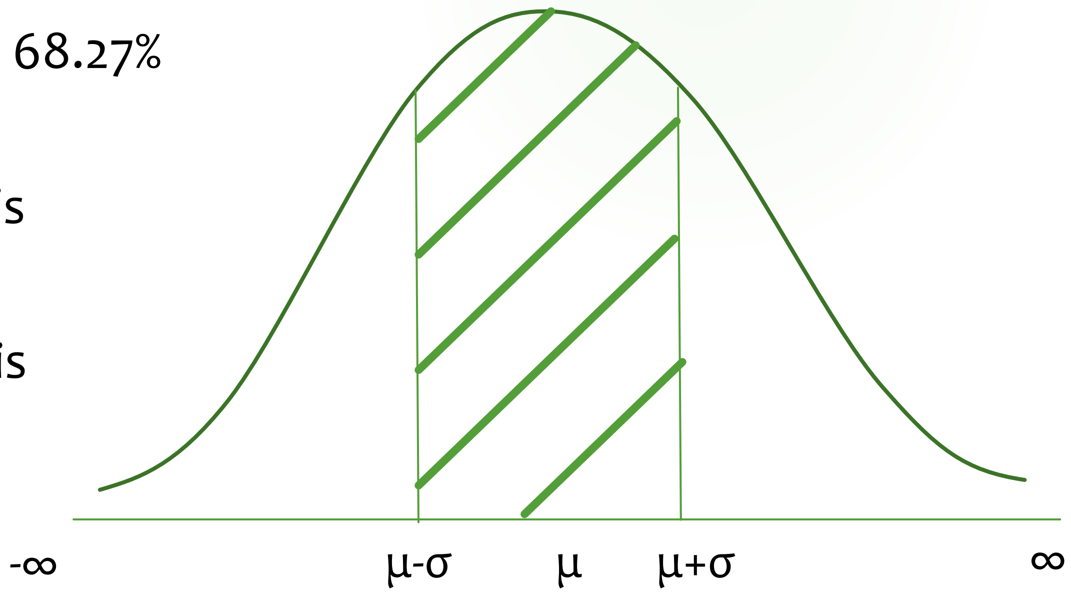
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad ; \quad -\infty < z < \infty$$



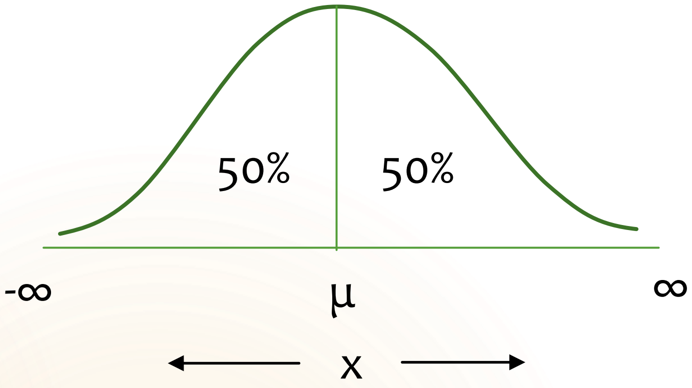
We write it as, $Z \sim N\,(0, 1)$

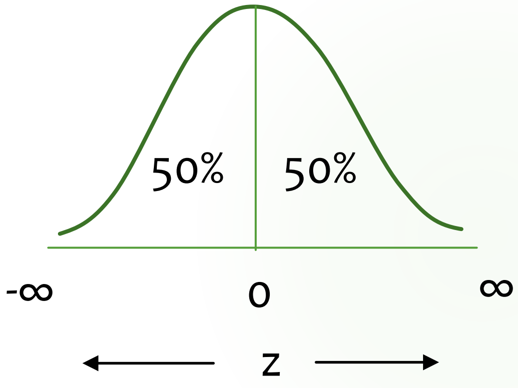
Characteristics of a Normal Distribution

1. Mean= Median = Mode
2. Symmetric and Mesokurtic
3. Bell-shaped curve
4. The area under the curve lying between $\mu \pm \sigma$ is 68.27% of the total area
5. The area under the curve lying between $\mu \pm 2\sigma$ is 95.45% of the total area
6. The area under the curve lying between $\mu \pm 3\sigma$ is 99.73% of the total area



Characteristics of a Normal Distribution





$P[X<\mu] = P[X>\mu] = 0.5$

$P[X<-x] = P[X>x]$

$P[Z<0] = P[Z>0] = 0.5$

$P[Z<-z] = P[Z>z]$

Normal Distribution Table

Z-table

- Normal distribution table provides probabilities for N(0,1) i.e. for standard normal distribution
- Usually, normal table gives P[0 < Z < z] for positive values of Z.
- For other values, we can use the property of symmetry with median 0 of standard normal distribution
- To find probabilities for a normal random variable X, we can transform the probability statement about X in terms of probability statement for Z and then calculate the probability using the standard normal distribution table or Z-table

$$P[X < a] = P\left[\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right] = P\left[Z < \frac{a-\mu}{\sigma}\right]$$

Finding Area Under the Normal Curve using Z-table

Example 7:

The number of viewers of a TV show per week has a mean of 29 million with a standard deviation of 5 million. Assume that, the number of viewers of that show follows a normal distribution.

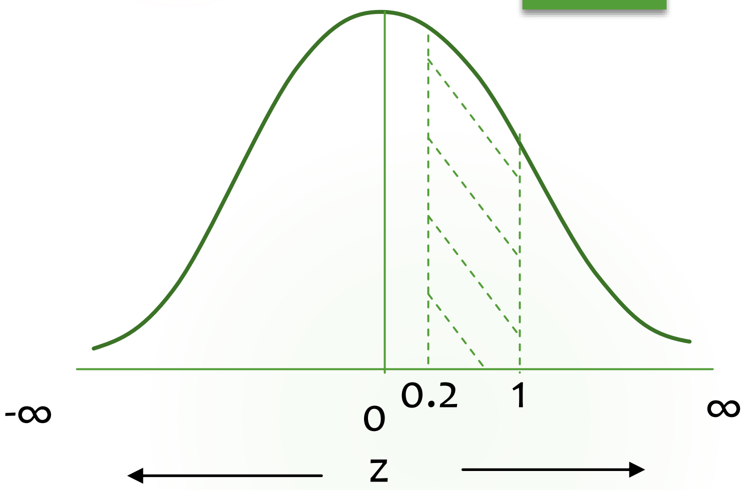
What is the probability that, next week's show will-

a. Have between 30 and 34 million viewers?

b. Have at least 23 million viewers?

c. Exceed 40 million viewers?

Finding Area Under the Normal Curve using Z-table



Solution:

Let, X= Number of viewers of the show per week (in million)

$$\therefore X \sim N\,(\mu, \sigma^2)$$

a. the probability that, next week's show will have between 30 and 34 million viewers-

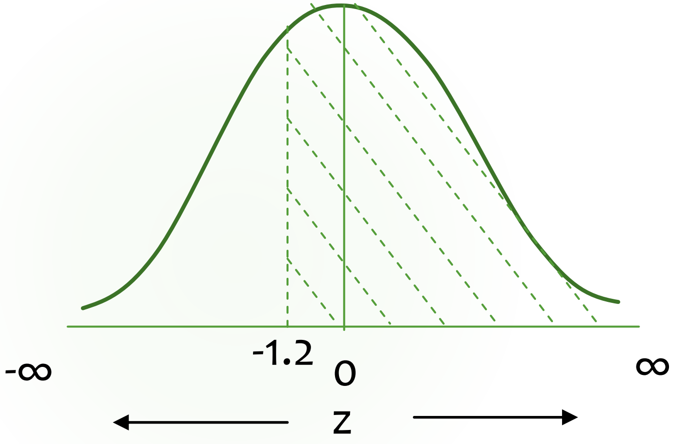
$$P[30 \le X \le 34] = P\left[\frac{30-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{34-\mu}{\sigma}\right] = P\left[\frac{30-29}{5} \le \frac{X-\mu}{\sigma} \le \frac{34-29}{5}\right]$$
$$= P[0.20 \le Z \le 1] = P[0 \le Z \le 1] - P[0 \le Z \le 0.2] = 0.3413 - 0.0793$$
$$= 0.262$$

Finding Area Under the Normal Curve using Z-table

Solution (contd.)**:**

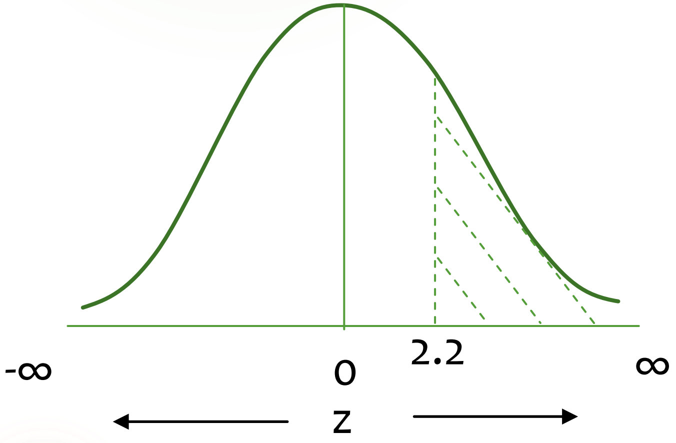
b. the probability that, next week's show will have at least 23 million viewers-

$$P[X \geq 23] = P\left[\frac{X-\mu}{\sigma} \geq \frac{23-\mu}{\sigma}\right] = P\left[\frac{X-\mu}{\sigma} \geq \frac{23-29}{5}\right]$$
$$= P[Z \geq -1.2] = P[-1.2 \leq Z \leq 0] + P[Z \geq 0] = 0.3849 + 0.5$$
$$= 0.8849$$



c. the probability that, next week's show will exceed 40 million viewers-

$$P[X > 40] = P\left[\frac{X-\mu}{\sigma} > \frac{40-\mu}{\sigma}\right] = P\left[\frac{X-\mu}{\sigma} > \frac{40-29}{5}\right]$$
$$= P[Z > 2.2] = P[Z \geq 0] - P[0 \leq Z \leq 2.2] = 0.5 - 0.4861 = 0.0139$$



Finding Area Under the Normal Curve using Z-table

Example 8:

a. For what value of 'a', $P[Z \le a] = 0.95$?

b. For what value of 'a', $P[Z \ge a] = 0.05$?

c. For what value of 'a', $P[Z \le a] = 0.975$?

Solution:

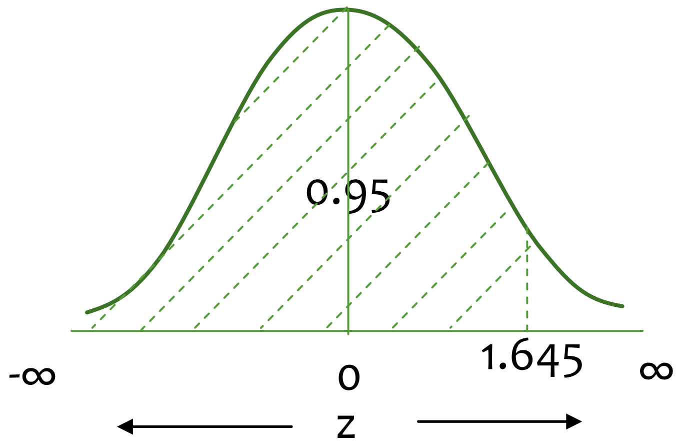
a. $P[Z \le a] = 0.95$

Or, $P[Z \le 0] + P[0 < Z \le a] = 0.95$

Or, $0.5 + P[0 < Z \le a] = 0.95$

Or, $P[0 < Z \le a] = 0.95 - 0.5 = 0.45$

For $a = 1.645$, $P[0 < Z \le a] = 0.45$



Finding Area Under the Normal Curve using Z-table

Solution **(contd.):**

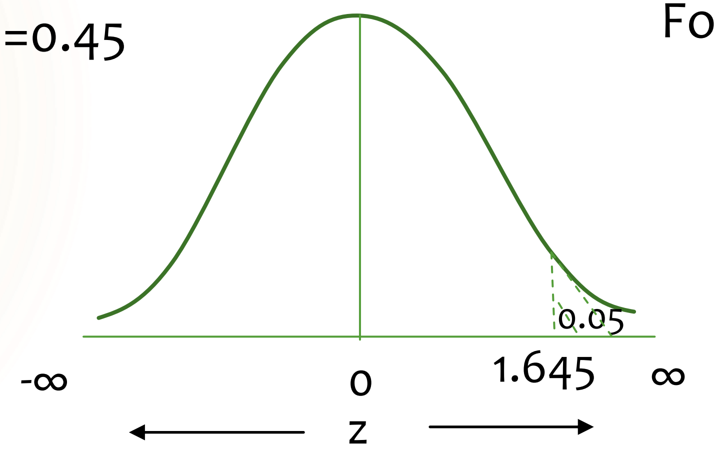
b. $P[Z \geq a] = 0.05$

Or, $P[Z \geq 0] - P[0<Z \leq a] = 0.05$

Or, $0.5 - P[0<Z \leq a] = 0.05$

Or, $P[0<Z \leq a] = 0.5 - 0.05 = 0.45$

For $a = 1.645$, $P[0<Z \leq a] = 0.45$



Solution **(contd.):**

c. $P[Z \leq a] = 0.975$

Or, $P[Z \leq 0] + P[0<Z \leq a] = 0.975$

Or, $0.5 + P[0<Z \leq a] = 0.975$

Or, $P[0<Z \leq a] = 0.975 - 0.5 = 0.475$

For $a = 1.96$, $P[0<Z \leq a] = 0.475$