

Tour de Distributions!

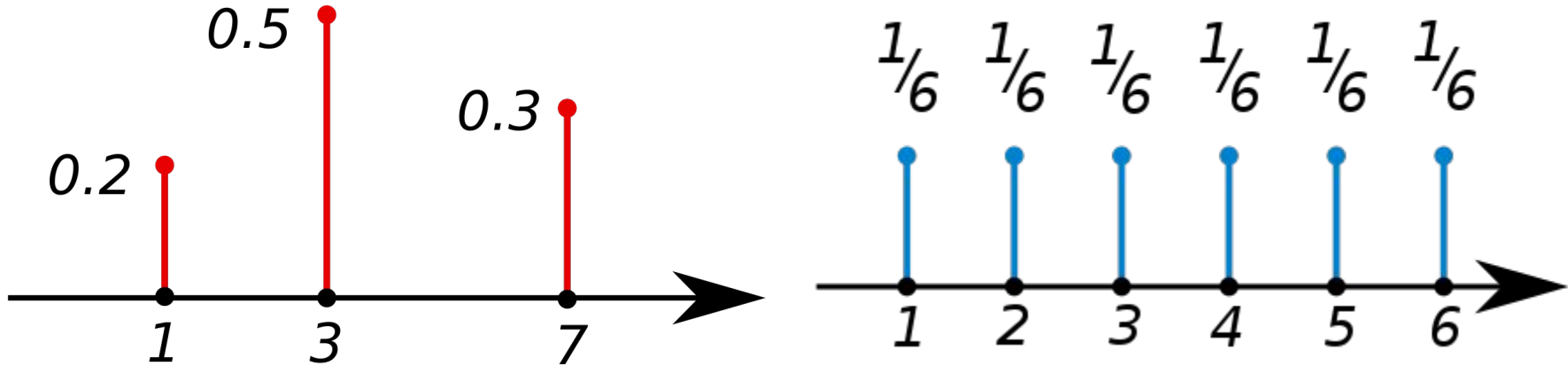
ATL-DS-0624

Goals

- Understand PMF and PDF.
- CDF for discrete and continuous space.
- How to Calculate and Interpret Z-Score.

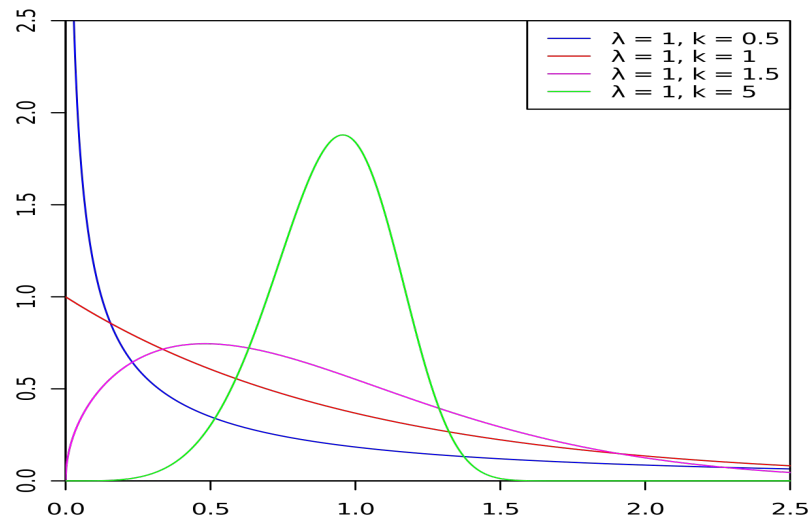
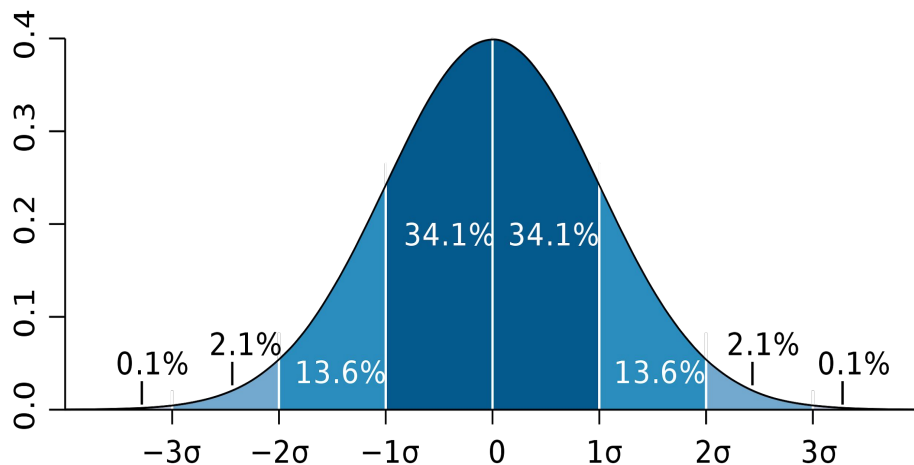
Probability Mass Function

- a function that gives the probability that a **discrete random variable** is exactly equal to some value.



Probability Density Function

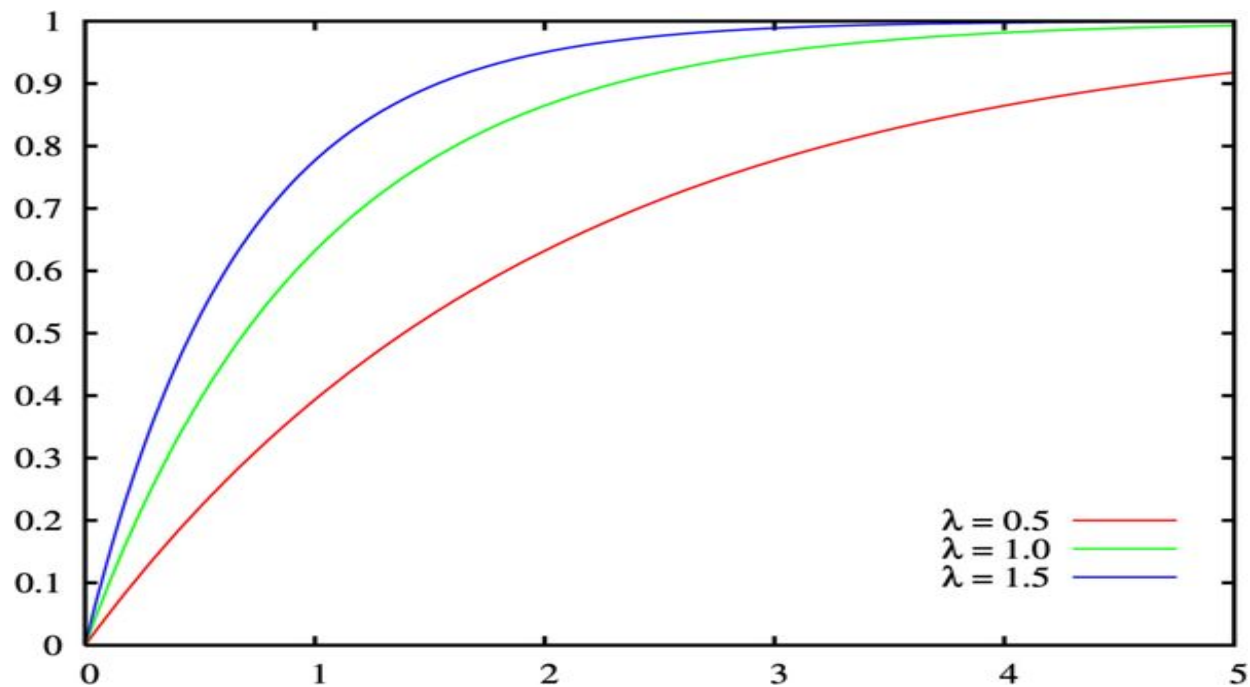
- Specifies the probability of the random variable falling within a particular range of values, as opposed to taking on any one value.
- This probability is given by the integral of this variable's PDF over that range.



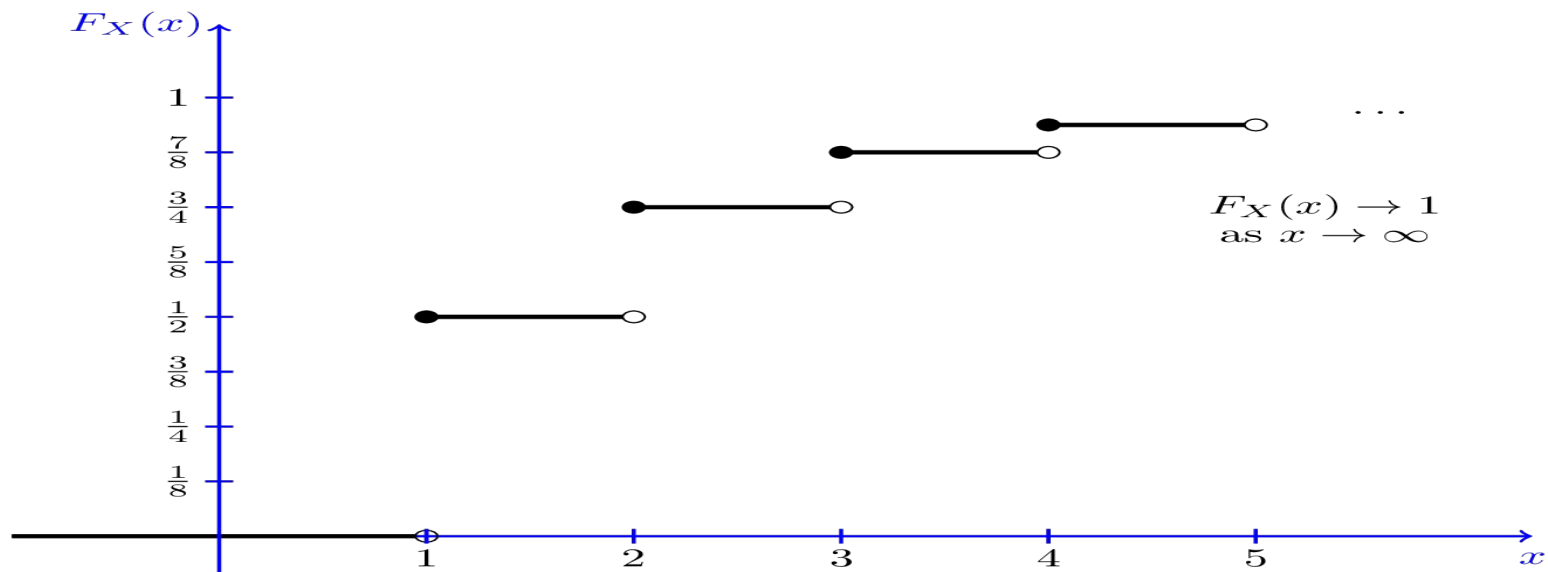
Cumulative Distribution Function

- The probability that X will take a value less than or equal to x .
- In the case of a continuous distribution, it gives the area under the probability density function from minus infinity to x .

CDF Continuous Case



CDF Discrete Case



Example Of PMF: Sum of Two Dice

Example Distribution

Set of possible values: $\mathbf{X} = \{2, 3 \dots 12\}$

Specific value of the random variable: $x \in \mathbf{X}$

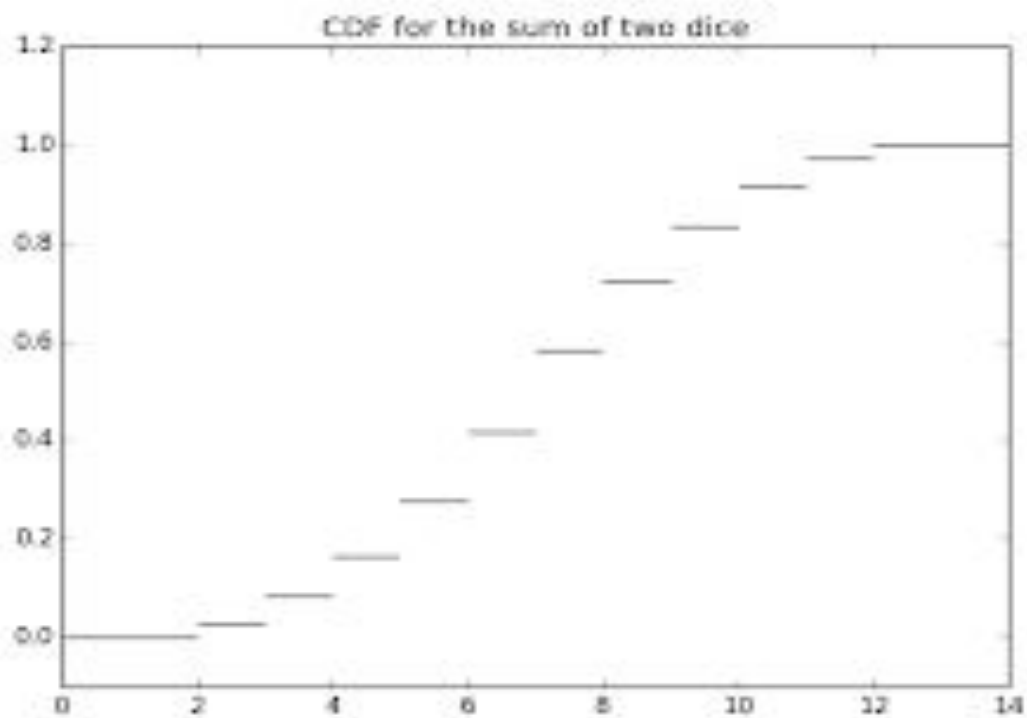
Probability of the value x : $P(x)$

x	2	3	4	5	6	7	8	9	10	11	12
$P(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$


PMF



CDF



Normal Distribution

 The Normal Distribution:
as mathematical function (pdf)

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Note constants:

$\pi=3.14159$

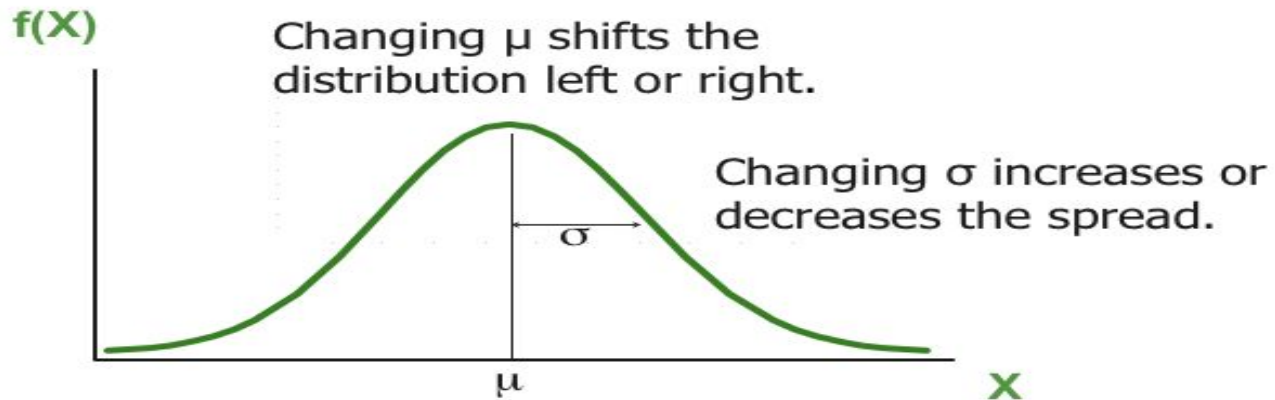
$e=2.71828$

This is a bell shaped curve with different centers and spreads depending on μ and σ

Parameters



The Normal Distribution





Normal distribution is defined
by its mean and standard dev.

$$E(X) = \mu$$

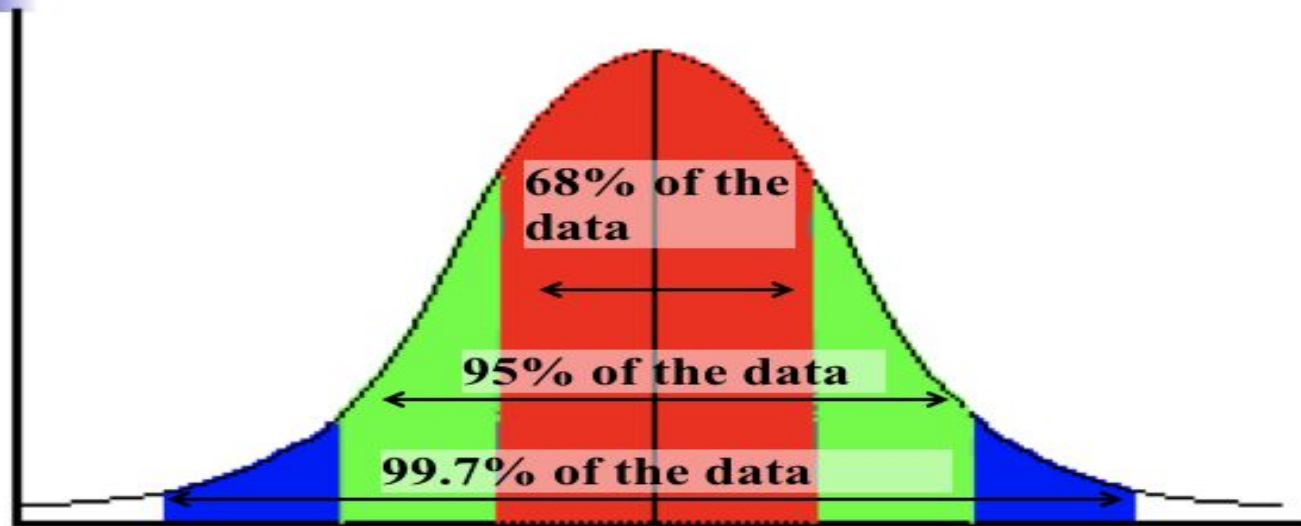
$$\text{Var}(X) = \sigma^2$$

$$\text{Standard Deviation}(X) = \sigma$$

Useful Facts

- No matter what μ and σ are,
- The area between $\mu - \sigma$ and $\mu + \sigma$ is about 68%.
- The area between $\mu - 2\sigma$ and $\mu + 2\sigma$ is about 95%.
- The area between $\mu - 3\sigma$ and $\mu + 3\sigma$ is about 99.7%. Almost all values fall within 3 standard deviations.

68-95-99.7 Rule



Standard Normal Distribution (Z)

All normal distributions can be converted into the standard normal curve by subtracting the mean and dividing by the standard deviation:

$$z = \frac{x - \mu}{\sigma}$$

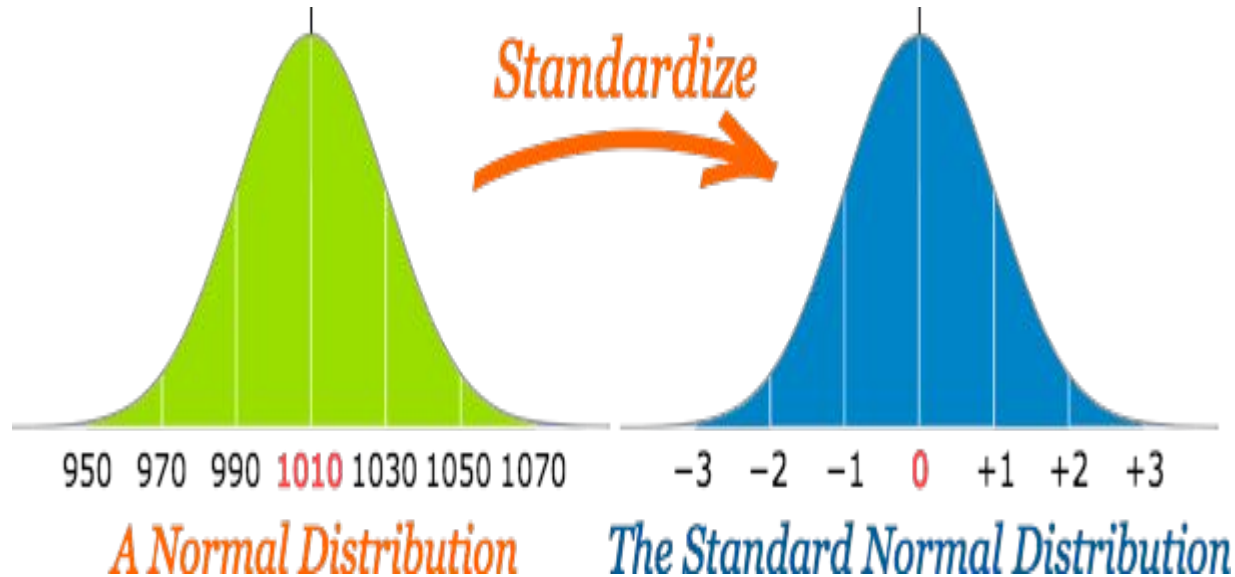
μ = Mean

σ = Standard Deviation

Somebody calculated all the integrals for the standard normal and put them in a table! So we never have to integrate!

Even better, computers now do all the integration.

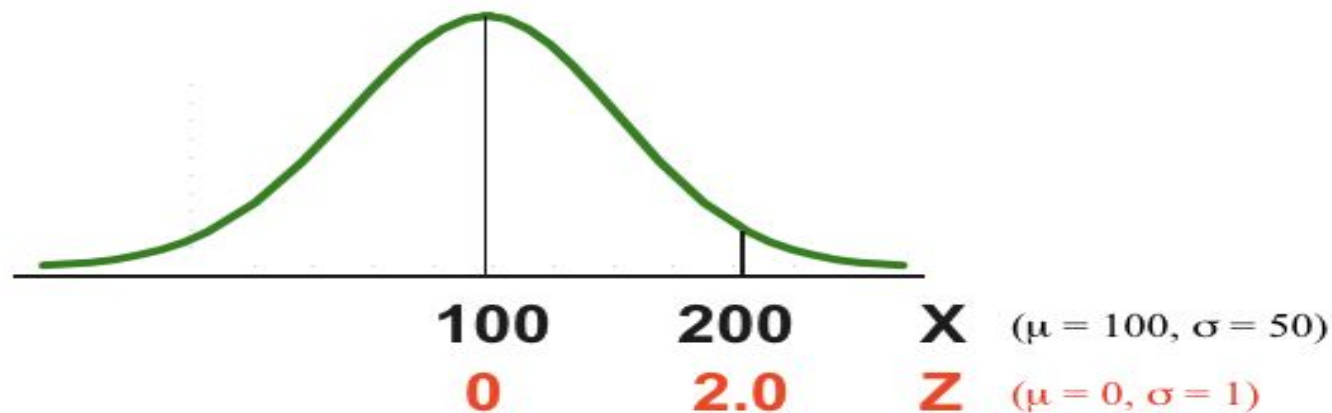
Standard Normal Distribution



A Standard Normal Distribution is a Normal Distribution with a mean of 0 and a standard deviation of 1



Comparing X and Z units



Why Standardizing?

- Gives us a good idea the relative location of raw values
- Allows us to compare different values in a more informative way
- Scaling for features if we conduct algorithms that rely on distance metrics

Example 1

Assume snowfall follows a normal distribution over time and the mean snowfall in New York City is 25 inches with a variance of 16 inches.

What is:

- 1) $P(X < 25) = 0.5$
- 2) $P(17 < X < 32) = 0.93$
- 3) $P(X = 25) = \text{Not possible!!!!}$



```
1 z_first = (17 - 25)/4
2 z_second = (32-25)/4
3 print('z score of 17 is : ',z_first)
4 print('z score of 33 is : ',z_second)
5 stats.norm.cdf(1.75) - stats.norm.cdf(-2)
```

z score of 17 is : -2.0
z score of 33 is : 1.75

0.9371907111880037

Distribution Deep Dives

- Patrick
 - Poisson Distribution
- Maia
 - Exponential Distribution
- Princess
 - Uniform Distribution (Discrete)
- Gary
 - Binomial Distribution
- Christiaan
 - Bernoulli Distribution
- Thoa
 - Uniform distribution (continuous)

Deep Dive

- 5 to 10 minute presentation on the assigned distribution.
- Presentation needs to include a description of the distribution.
- Graphical representation of PMF/PDF and CDF.
- Mean and variance of the distribution.
- Application of distribution.