



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY
BY EVERYONE'S SMART UNIVERSITY

Mathematical Foundations For Data Science and AI

Lecture 09

AI102

Probability; Distributions

By
Dr. Khaled ELMenshawy

www.midocean.ae

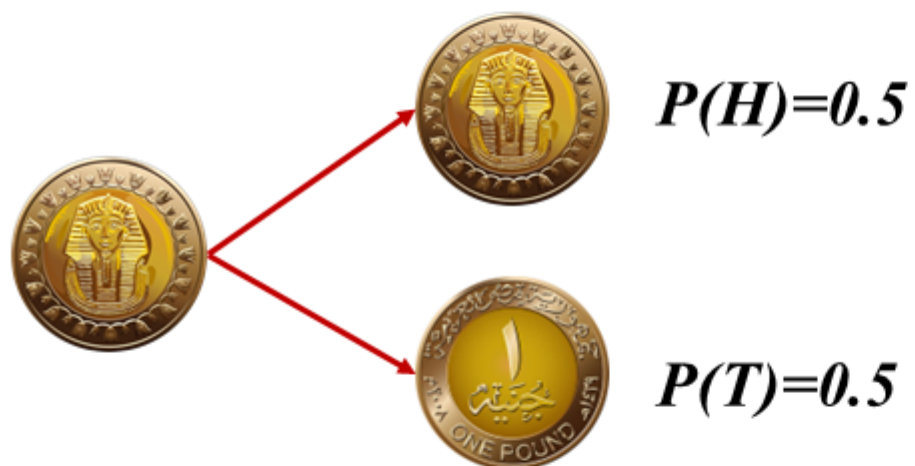
Lecture

09

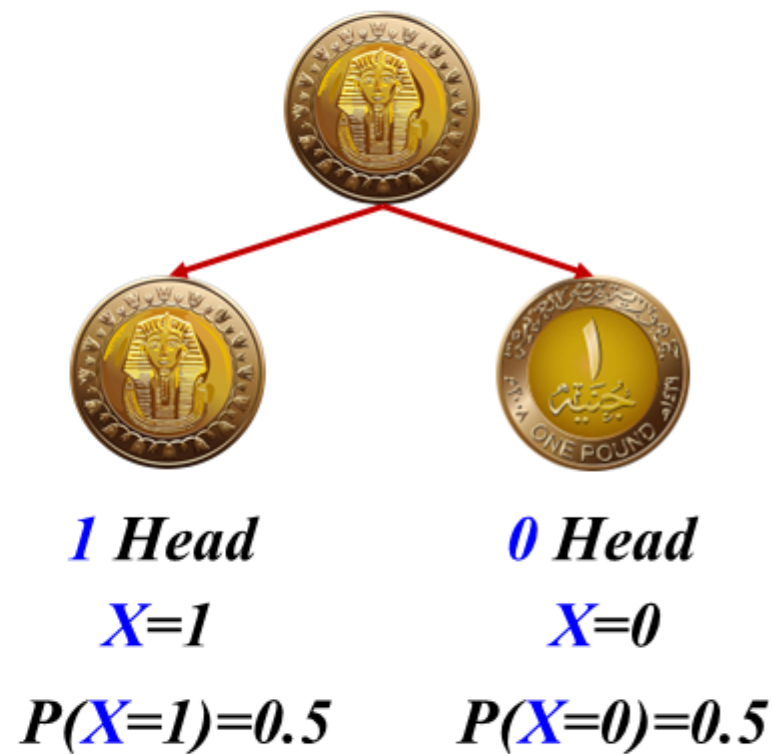
LEARNING OUTCOMES

- Understanding Random Variable.
- Understanding Probability Distributions.
- Understanding Bernoulli Distribution.
- Understanding Binomial Distribution.
- Understanding Poisson Distribution.
- Understanding Probability Density Function.
- Understanding Cumulative Distribution Function.
- Understanding Normal Distribution.
- Understanding the Mean, Median, Mode, Variance, and Standard deviation

Random Variable



$X = \text{No. of Heads}$



X is a random variable

Random variable mapping all the outcomes of a random experiment, allow to model the whole experiment at once.



$X = \text{No. of Heads}$



$X = \text{No. of 1's}$



$X = \text{No. Sick patients}$



$X = \text{the waiting time until the bus arrive}$



$X = \text{the Hight of a gymnast's jump}$

Random Variable

Discrete



A discrete random variable can take only a *finite number of distinct values* such as 0, 1, 2, 3,.. (*countable number of values*).

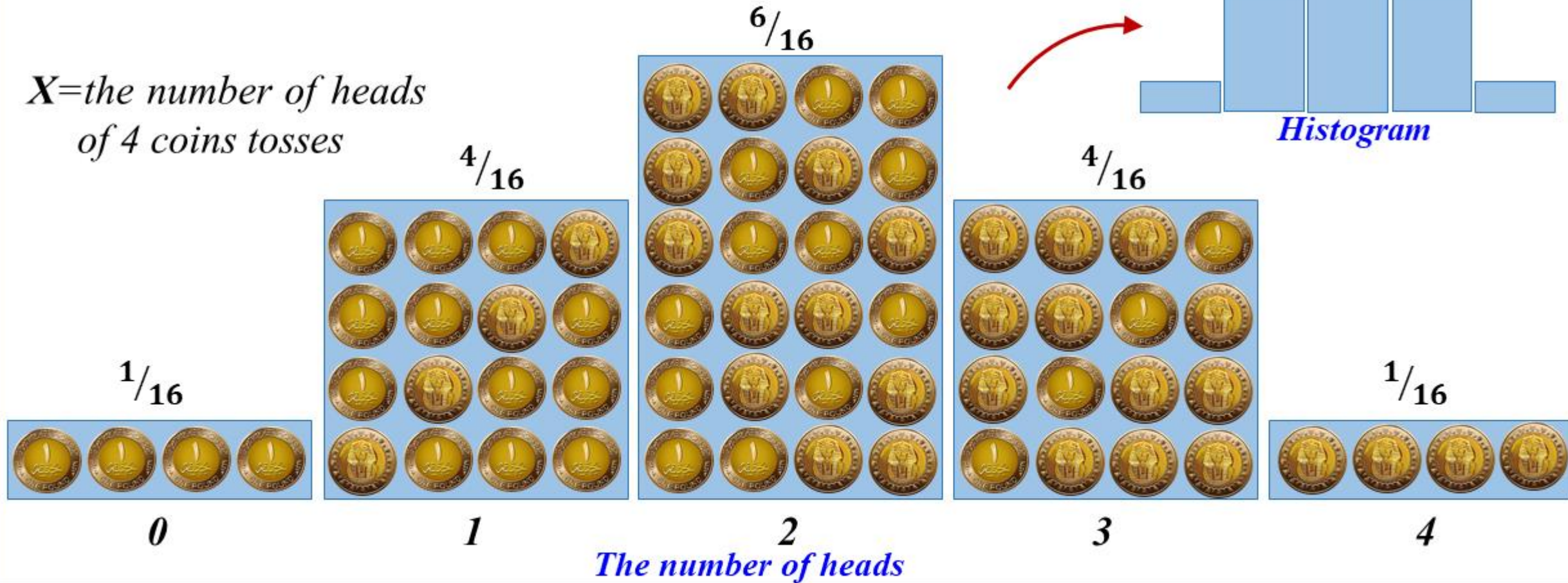
Continuous



A continuous random variable is an *infinite and uncountable set of values*, it is any value in a given interval (a,b) .

*A **probability distribution** is a mathematical function that describes the possible values a variable can take and how frequently they occur.*

*X = the number of heads
of 4 coins tosses*

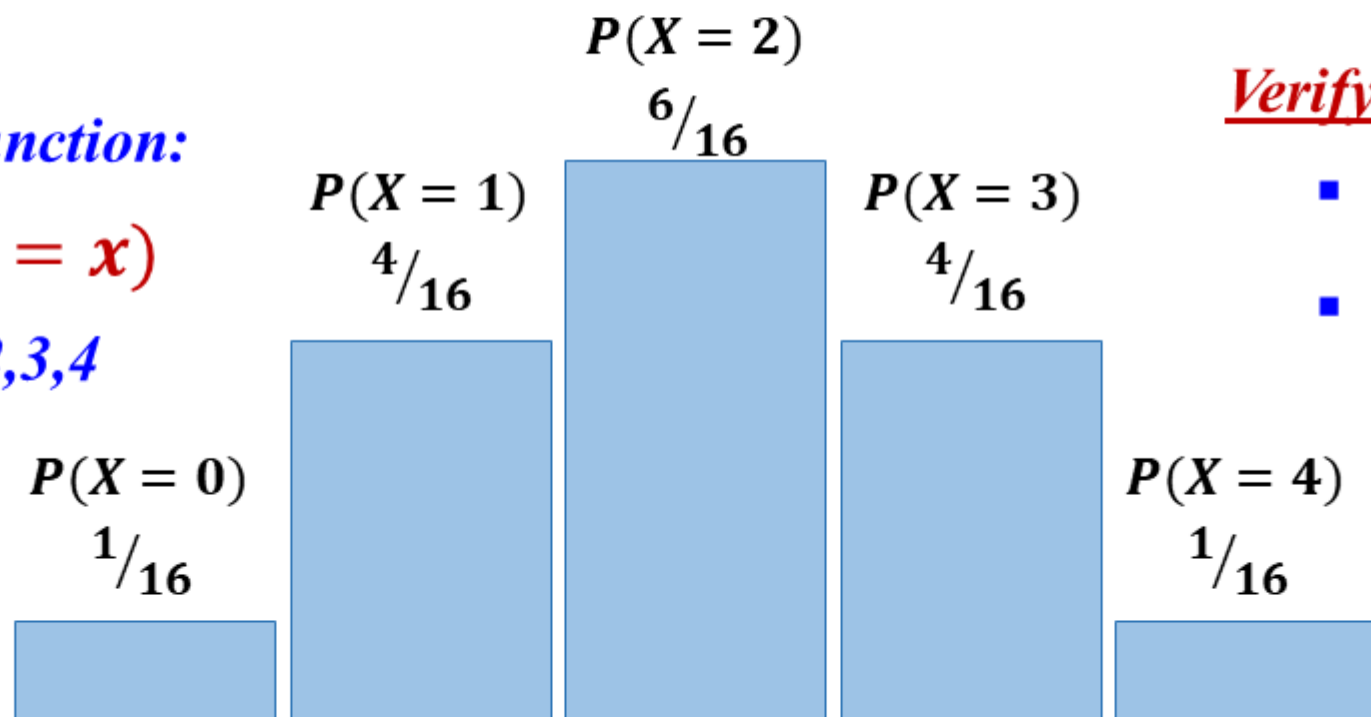


X =the actual outcome of an event (e.g. the number of heads obtained from 4 coins tosses),
 and x =one of the possible outcomes(e.g. specific number of heads such as 1,2,3,or 4)

Probability Mass Function:

$$p_X(x) = P(X = x)$$

where $x=0,1,2,3,4$



Verify that:

- $p_X(x) \geq 0$
- $\sum p_X(x) = 1$

$$p_X(0) + p_X(1) + p_X(2) + p_X(3) + p_X(4) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16} = 1$$

Bernoulli distribution is a discrete probability distribution where the Bernoulli random variable can have only 0 or 1 as the outcome.

X is number of heads of flip coin

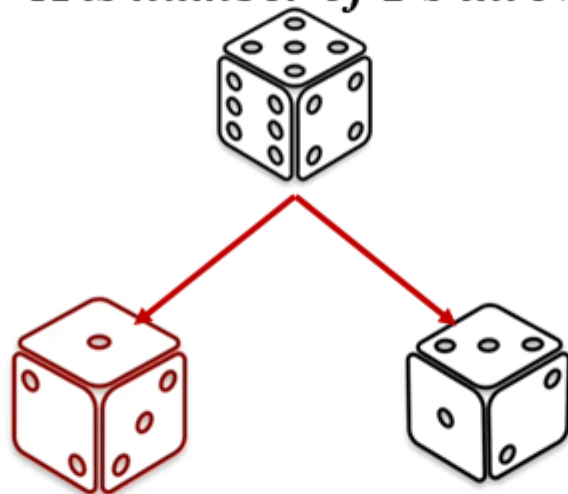


$$P(X = 1) = 0.5 \quad P(X = 0) = 0.5$$

Success

Failure

X is number of 1's throw a dice

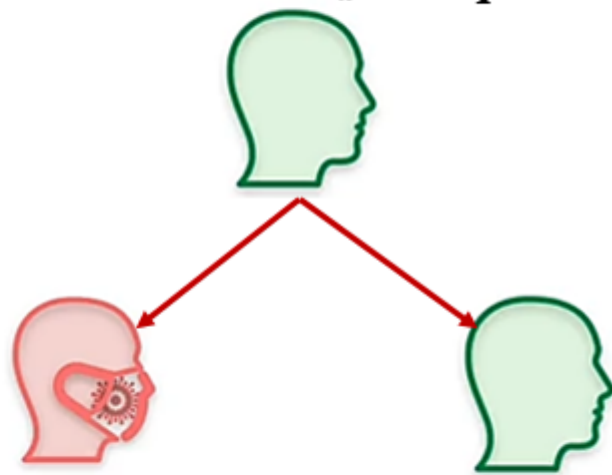


$$P(X = 1) = \frac{1}{6} \quad P(X = 0) = \frac{5}{6}$$

Success

Failure

X is number of sick patients



$$P(X = 1) = p$$

Success

$$P(X = 0) = 1 - p$$

Failure

$X \sim \text{Bernoulli "Bern"} (p)$

Probability of Success

Binomial Distribution a type of discrete distribution that summarizes the probability that a value will take one of two independent values under a given set of parameters or assumptions.

What is the probability to have 2 heads through flipping 5 coins?



10 outcomes

Binomial Coefficient

$$10 = \frac{5!}{2! (5-2)!} = \binom{5}{2} = \binom{n}{k} \text{ or } \binom{n}{n-k}$$

120 ← 5! 2 ← 2! 6 ← (5-2)!

Number of ways for getting 2 heads in 5 coins tosses.

General Probability Mass Function for X when the coin has $P(H)=p$ and event $X=x$ heads in n tosses:

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ where } x=0,1,2,3,\dots,n$$

Diagram: A red arrow points from the binomial coefficient $\binom{n}{x}$ to the expression $\frac{n!}{x!(n-x)!}$ above it.

Diagram: Three red arrows point from the parameters of the binomial distribution to their descriptions.

$X \sim \text{Binomial "B" } (n, p)$

- No. of trials* (points to n)
- Probability of Success for each trial* (points to p)
- Parameters of Binomial distribution* (points to the pair (n, p))

What is the probability of having 3 Ones on rolling 5 dices no matter on which dice ?



One

$$p = \frac{1}{6}$$



One

$$p = \frac{1}{6}$$



One

$$p = \frac{1}{6}$$



Not One

$$p = \frac{5}{6}$$



Not One

$$p = \frac{5}{6}$$

➤ $n = 5$

➤ $p = \frac{1}{6} = 0.1666$

$$X \sim B(5, 0.1666)$$

Quiz

Guessing entire Quiz is Binomial event

The number of times expected to get a specific outcome.



Guessing one question is Bernoulli event

Which outcome expected for a single trial

***A Poisson distribution** is a discrete probability distribution, It gives the probability of how often an event occurs over a specified period.*

***For example**, a firefly might light up three times in 10 seconds on average. We should use a **Poisson distribution** if we want to determine the likelihood of it lighting up eight times in 20 seconds.*

***Another example**, a teacher creates an online course on probability. Usually, his students asked around **four questions per day**, but **yesterday they asked seven**. How do we use **Poisson distribution** in this example.*

Usually, $\rightarrow 4$. ($\lambda = 4$)

Yesterday, $\rightarrow 7$. ($y=7$)

The probability that the students asked seven questions per day ($P(y)$)

$$P(y) = \frac{\lambda^y e^{-\lambda}}{y!} \quad e \approx 2.72$$

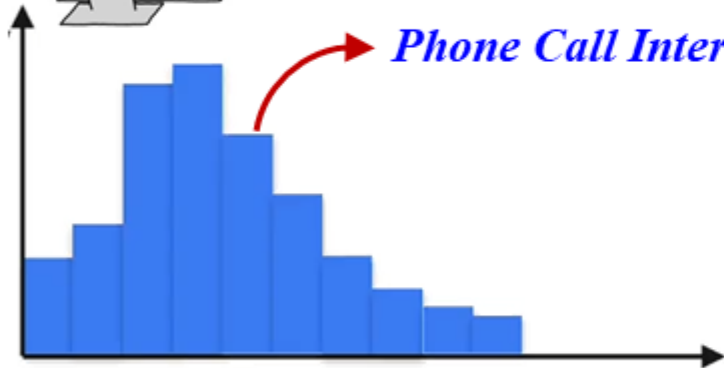
$$P(7) = \frac{4^7 e^{-4}}{7!} = \frac{16384 \times 0.0183}{5040} \approx 0.06$$

Only 6% chance of asked seven questions per day.

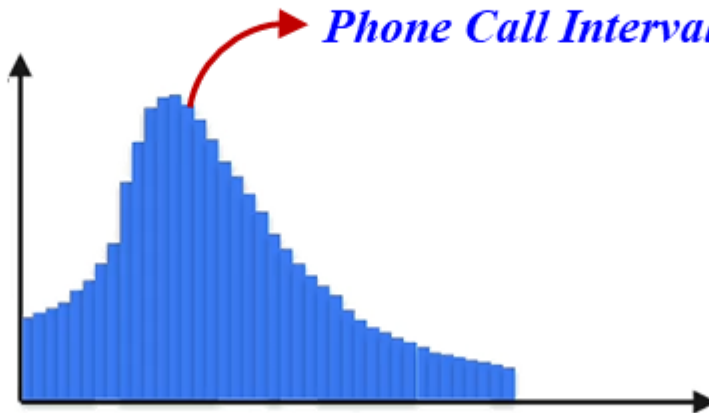
*Some events have infinitely many consecutive outcomes. We call such distributions continuous, Thus we can no longer represent these distributions with a table, instead, we represent them with a graph of the **probability density function (PDF)**.*



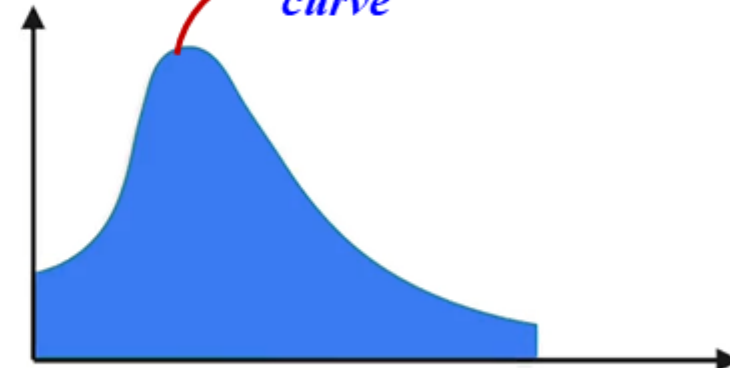
Phone Call Interval



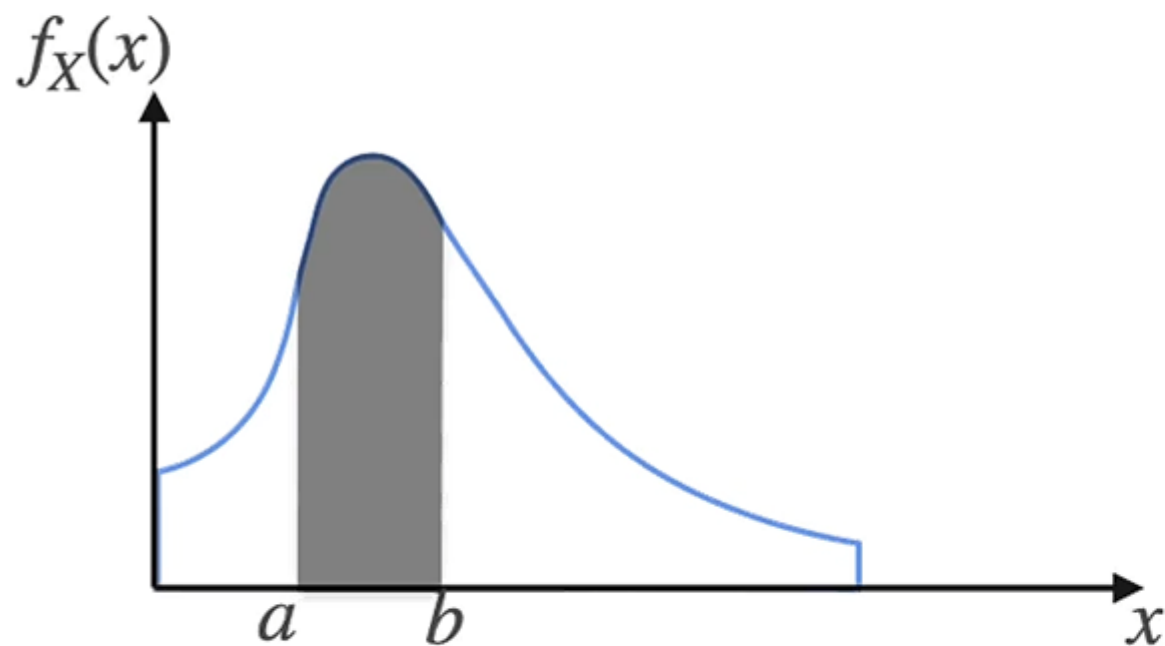
Phone Call Interval



Probabilty distribution curve



Probability density function (PDF); $f_X(x)$, defined only for continuous variable.

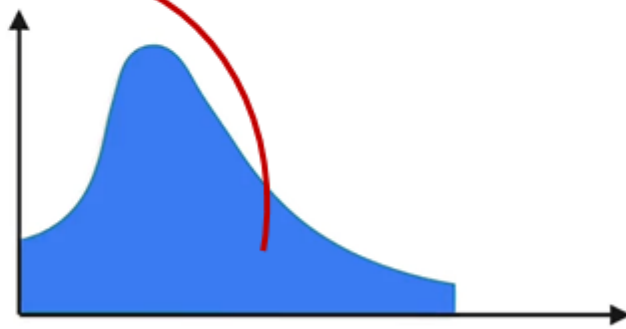


$P(a < X < b)$ Area under the curve

$f_X(x)$ Defined for all numbers

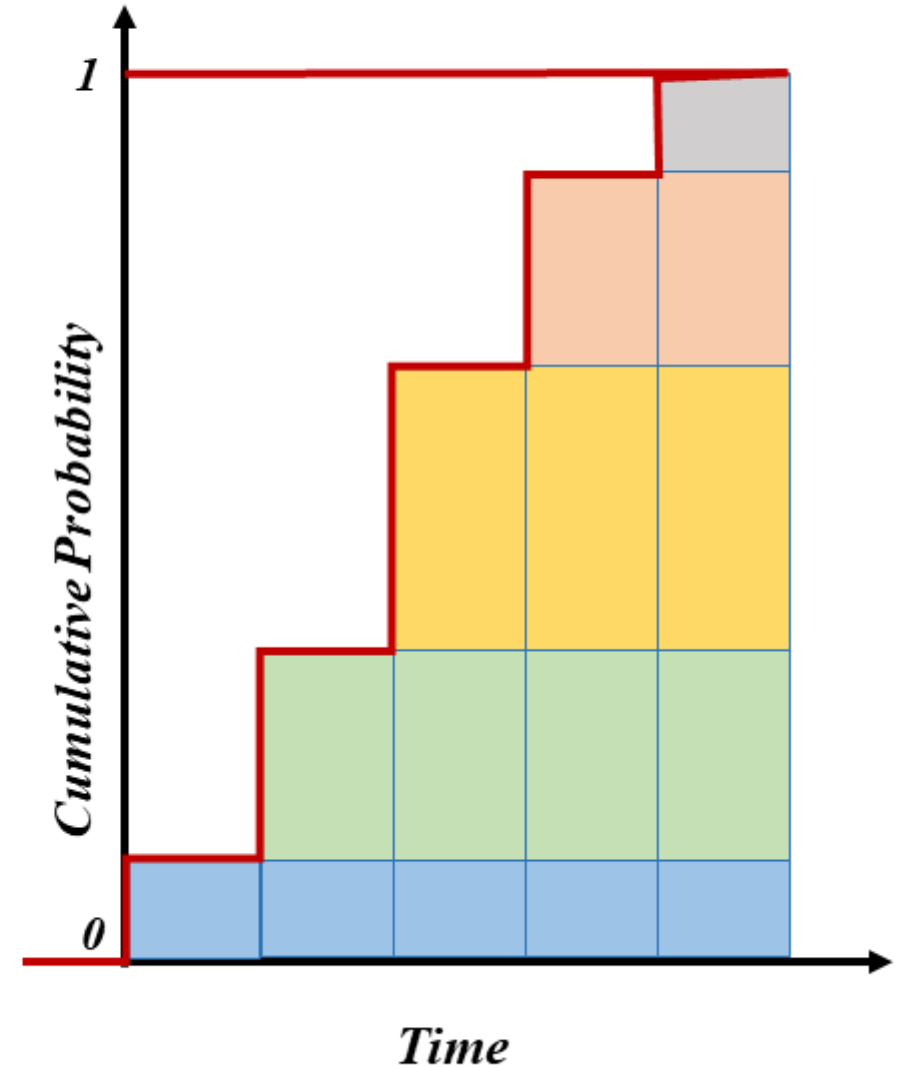
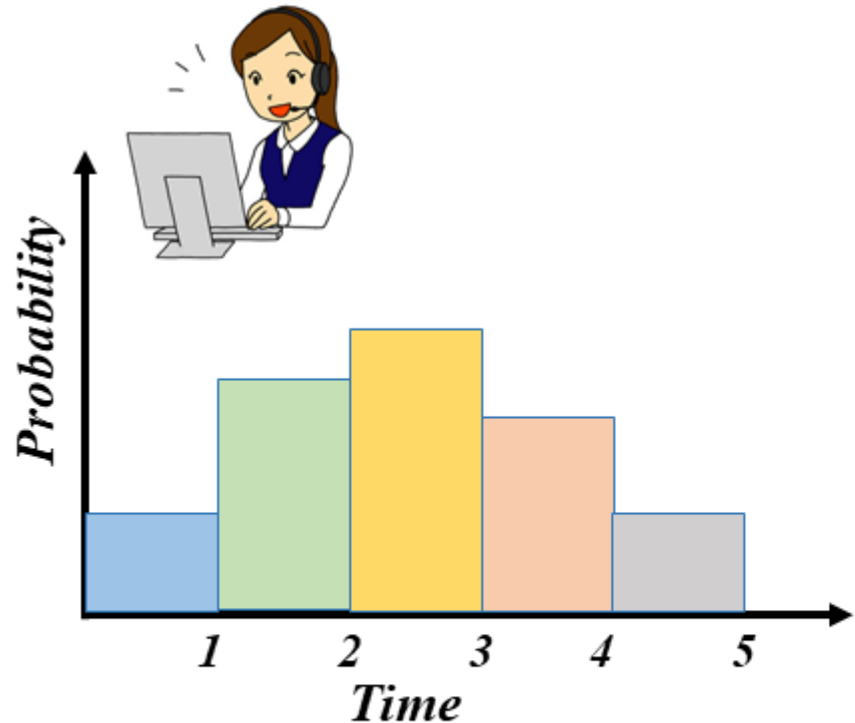
$$f_X(x) \geq 0$$

Area under $f_X(x) = 1$



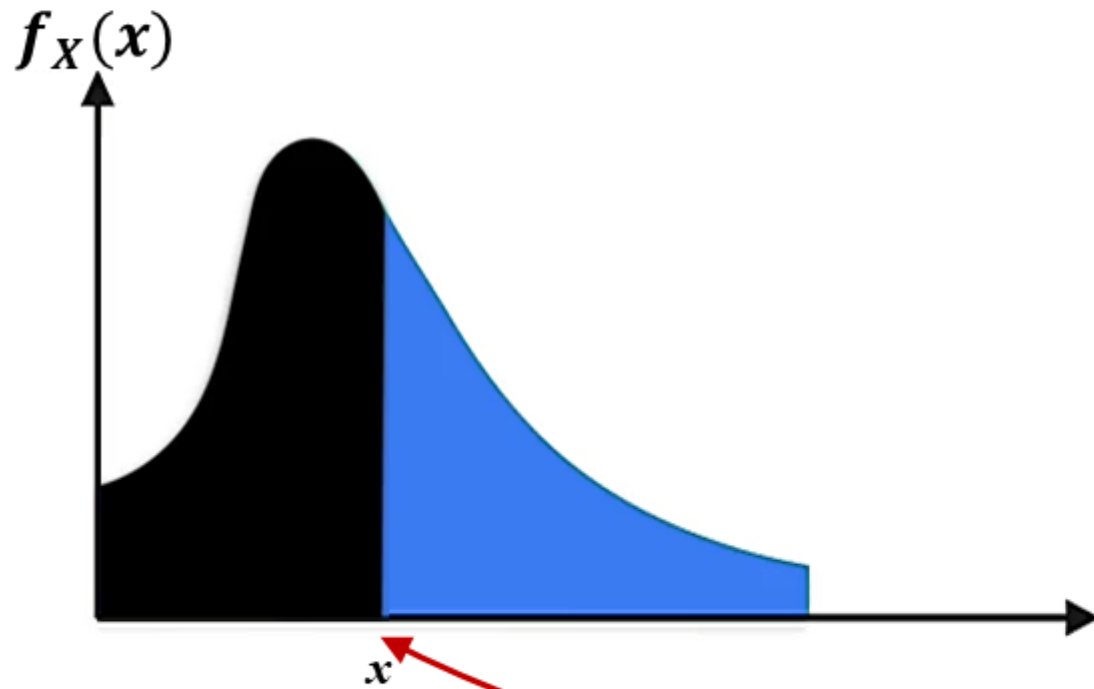
Cumulative Distribution Function

The cumulative distribution function (CDF) gives the probability that the random variable X is less than or equal to a certain value x

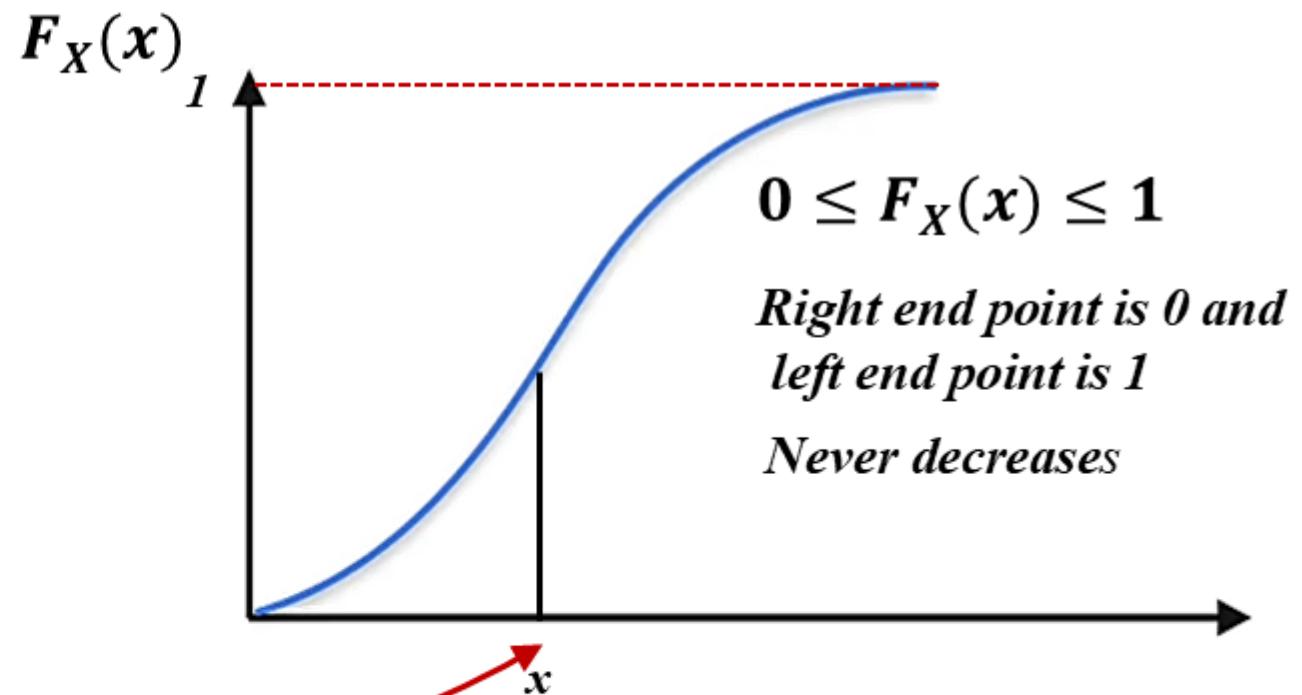


Cumulative Distribution Function

Probability Density Function



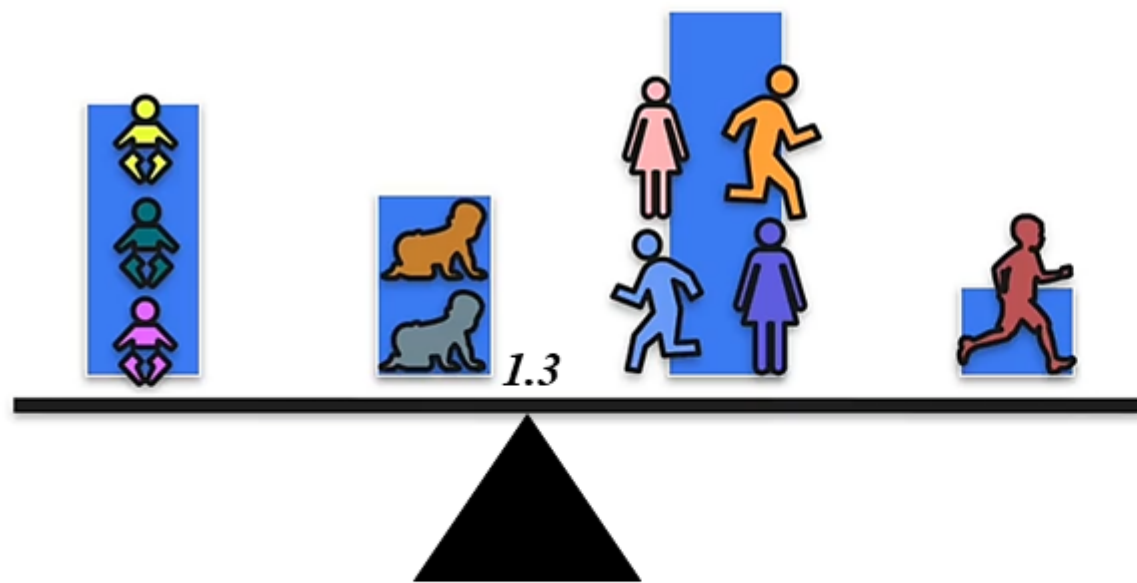
Cumulative Distribution Function



Less than or equal 2 minutes

***The Mean** is adding all numbers in the data set and then dividing by the number of values in the set.*

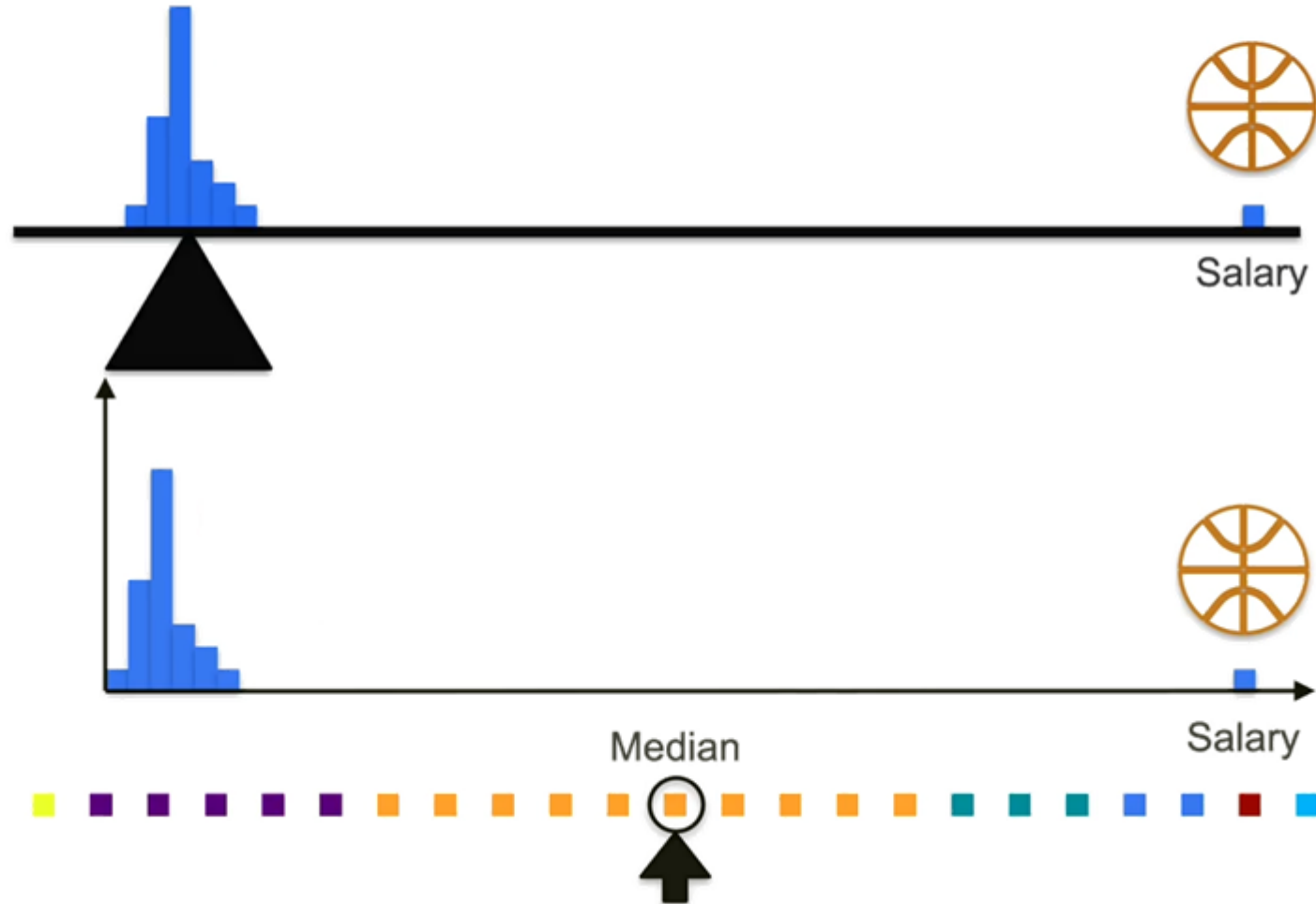
Age: 0 1 2 3



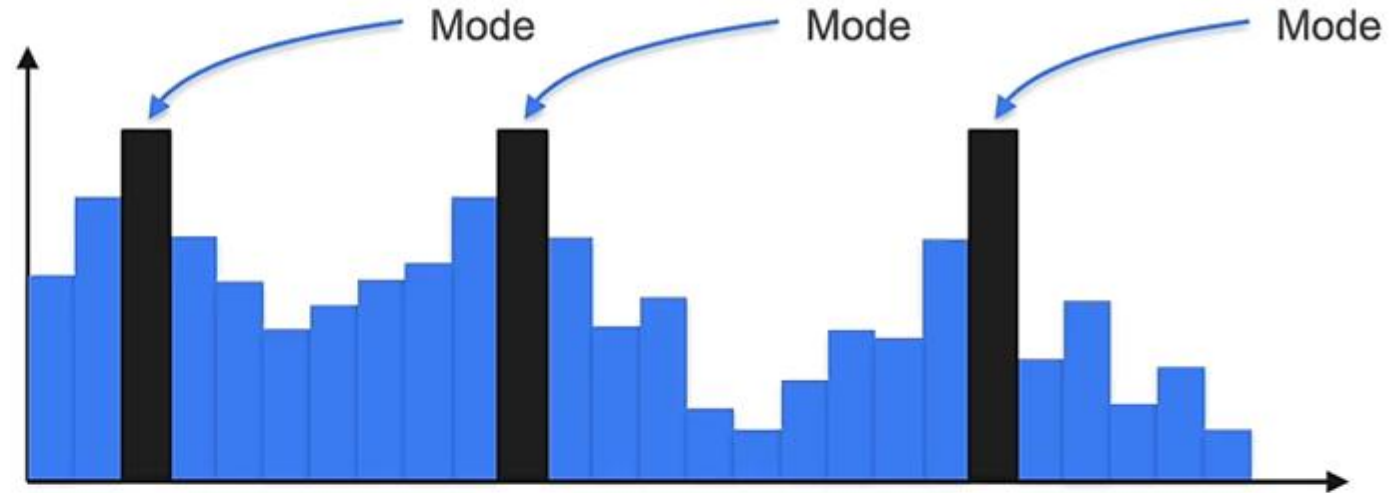
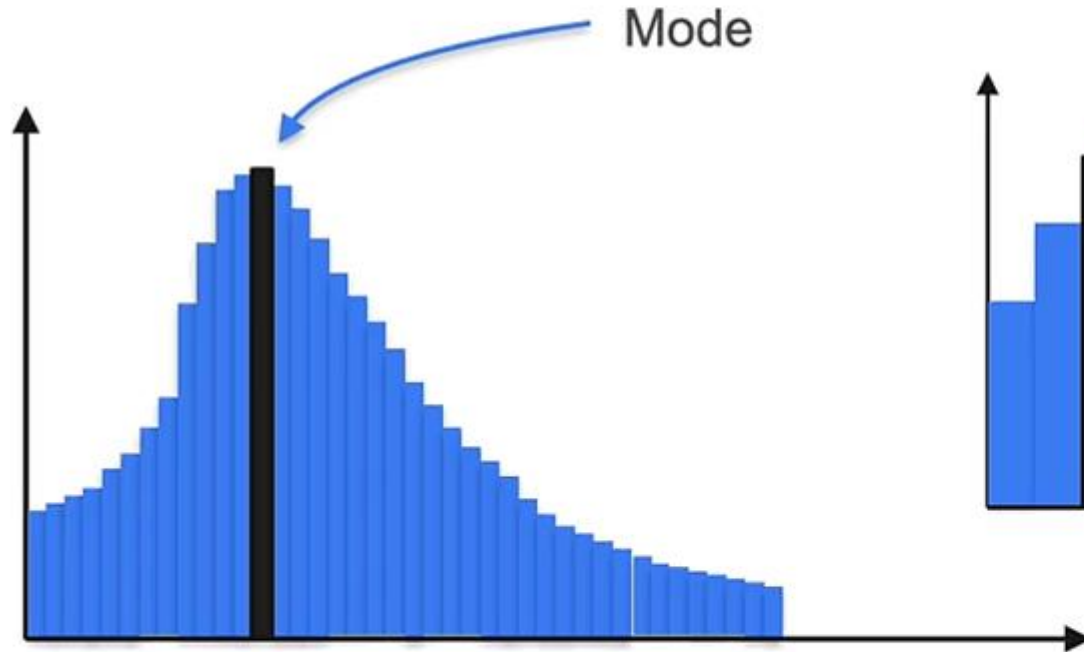
$$\frac{0 + 0 + 0 + 1 + 1 + 2 + 2 + 2 + 2 + 3}{10}$$

$$= \frac{13}{10} = 1.3$$

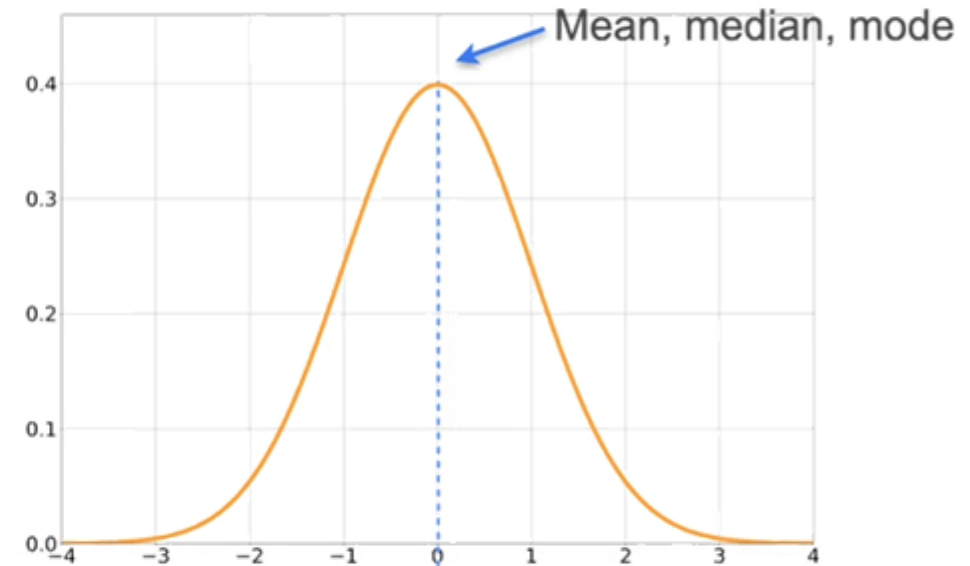
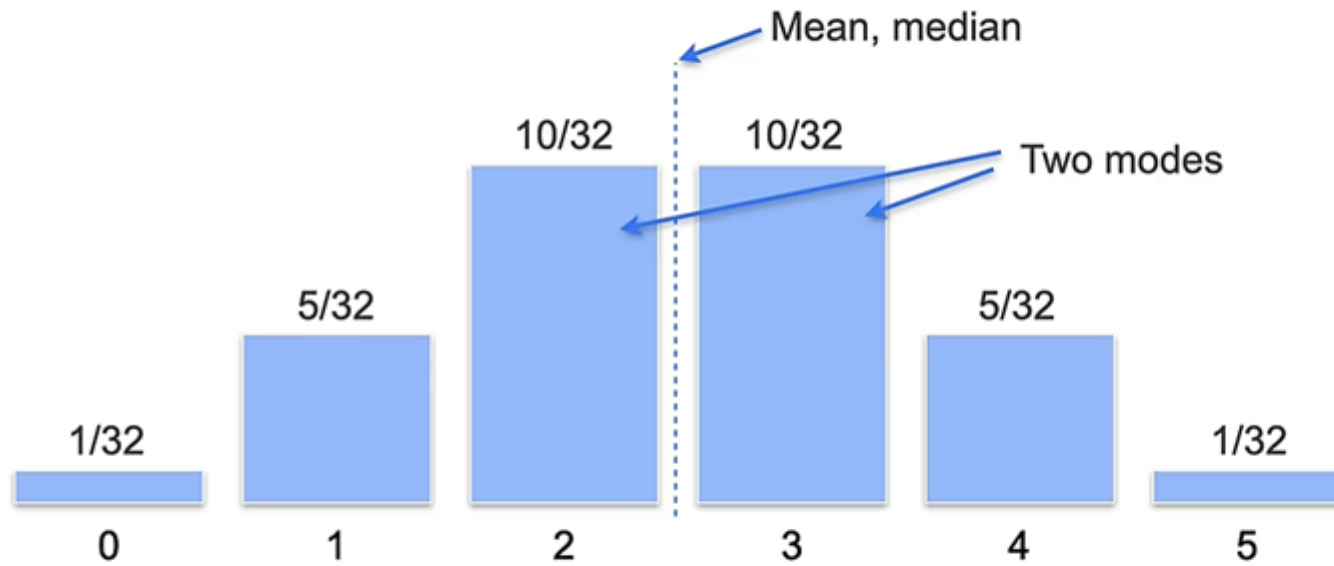
The median is the value in the middle of a data set.



The mode is a number in a set of numbers that appears the most often.



The Mean, Median, and Mode



*If you have a any probability distribution discrete or continuous, and you are to draw a random sample from it, what do you expect this sample to be on average? This is called **the expected value**.*

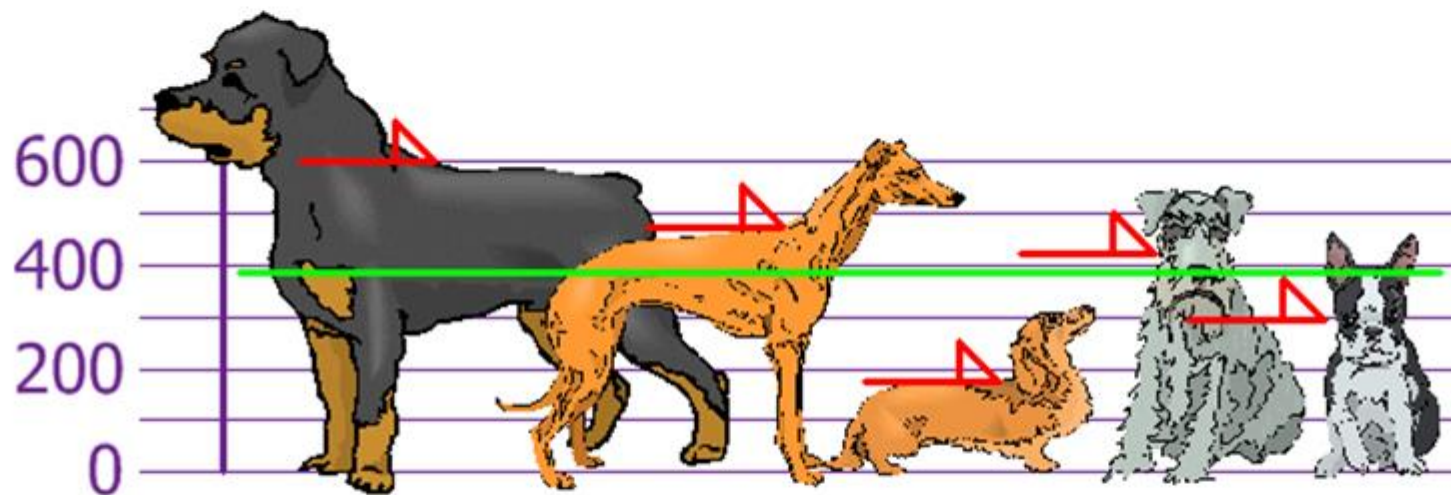


What is the fair amount of money to play the game?

$$0.5 \times 100 + 0.5 \times 0 = 50$$

*Expect to win 50 pound on average, **$E[X]=50$***

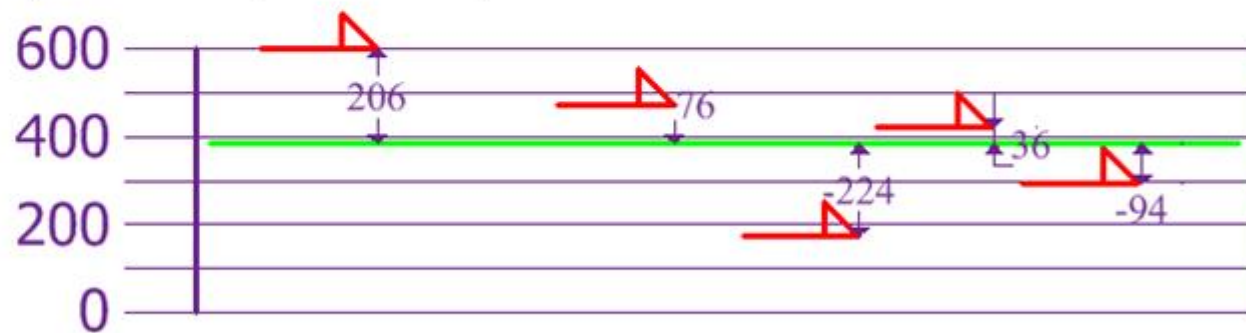
Variance σ^2 is The average of the squared differences from the Mean.



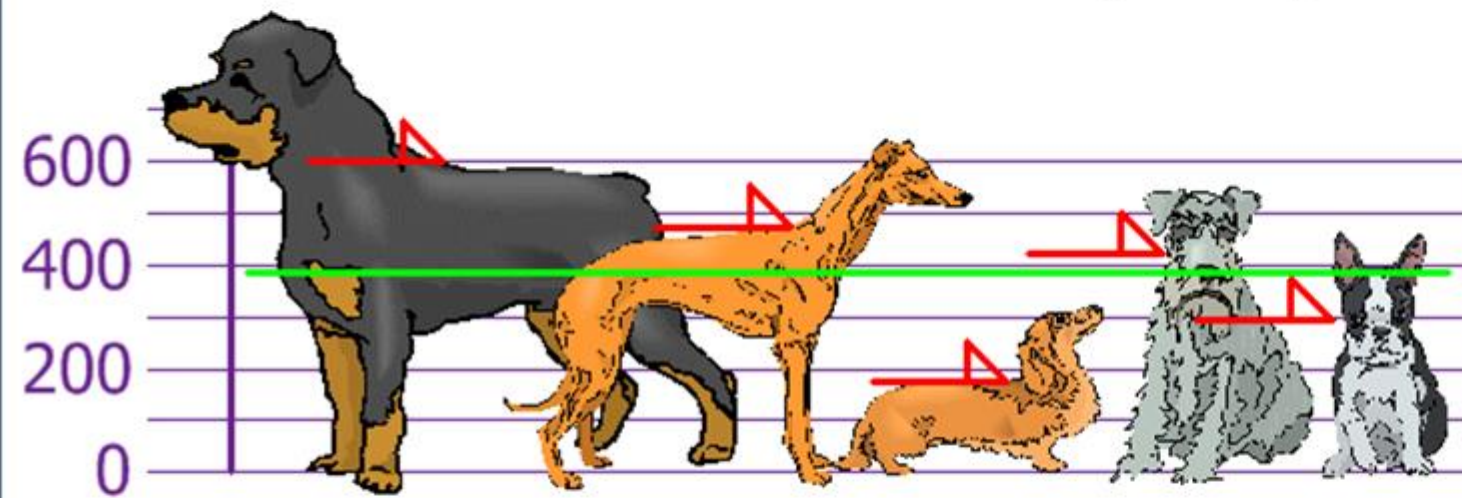
The heights (at the shoulders) are: 600mm, 470mm, 170mm, 430mm and 300mm.

The mean = 394mm

The variance = 21704



The Standard Deviation σ is a measure of how spread-out numbers are.



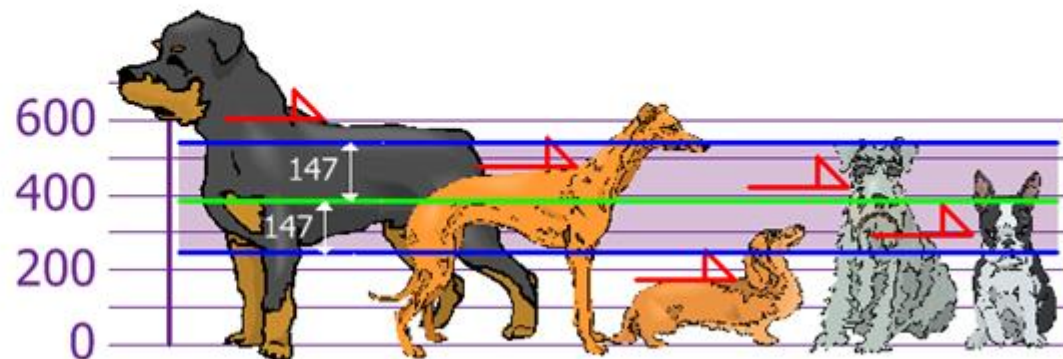
A standard deviation close to zero indicates that data points are very close to the mean, whereas a larger standard deviation indicates data points are spread further away from the mean

The heights (at the shoulders) are: 600mm, 470mm, 170mm, 430mm and 300mm.

The mean = 394mm

The variance= 21704

The standard deviation= 147



***The normal distribution** or Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.*

Mean ← *Variance*

$$X \sim \text{Normal "N"} (\mu, \sigma^2)$$

The graph of a normal distribution is bell shaped. Thus, the majority of data is centered around the mean.

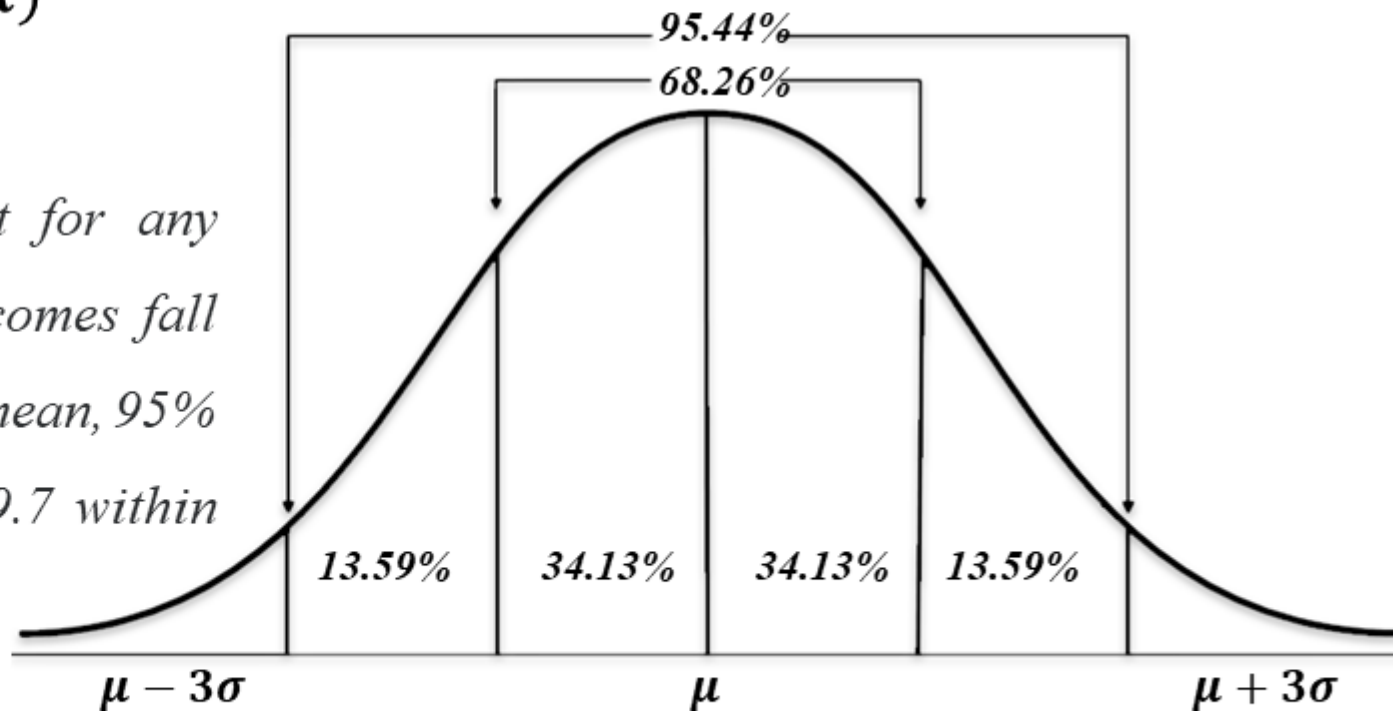


Expected value $E(x) = \mu$

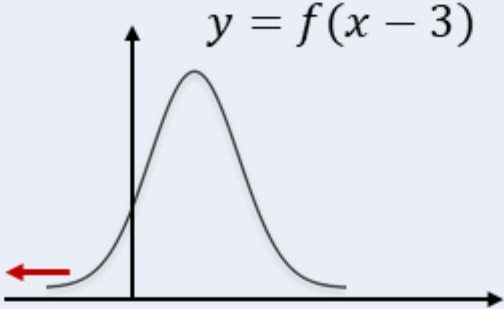
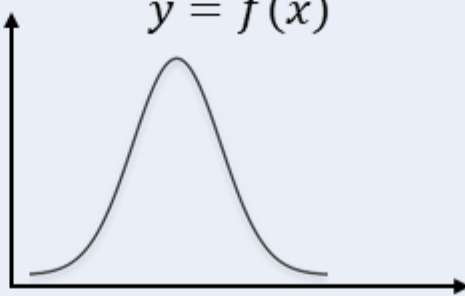
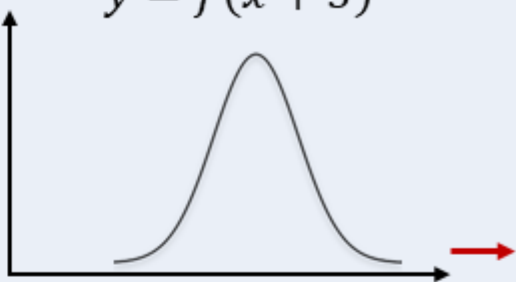
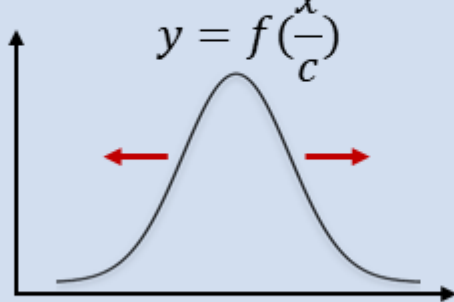
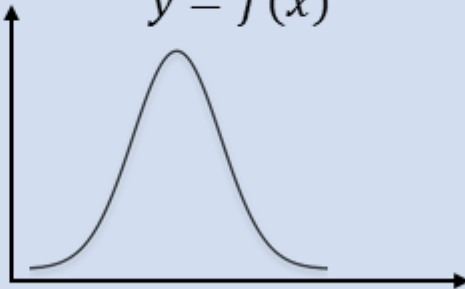
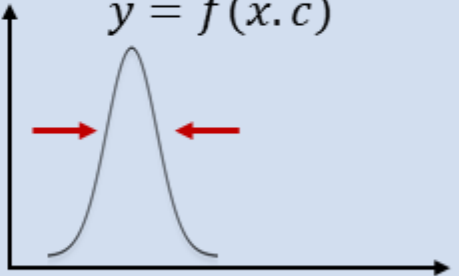
$$\text{Var}(x) = \sigma^2$$

$$\text{Var}(x) = E(x^2) - E(x)^2$$

68, 95, 99.7 law; This law suggests that for any normally distributed event, 68% of all outcomes fall within one standard deviation away from the mean, 95% fall within two standard deviations, and 99.7 within three.



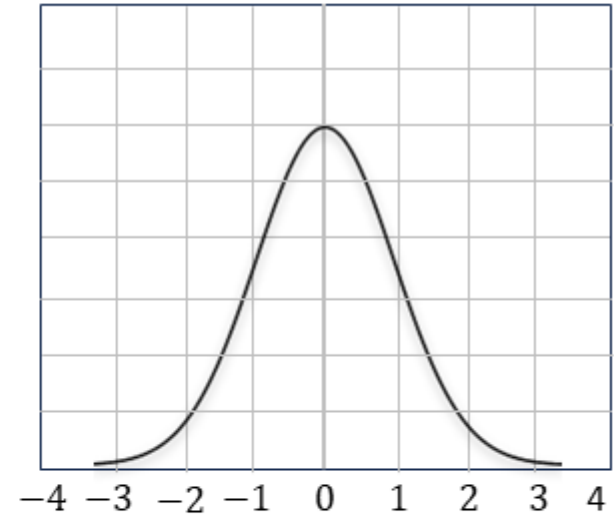
The transformation is a way in which we can alter every element of a distribution to get a new distribution with similar characteristics.

Summation and Subtraction	 $y = f(x - 3)$	 $y = f(x)$	 $y = f(x + 3)$
Multiplication and Division	 $y = f\left(\frac{x}{c}\right)$	 $y = f(x)$	 $y = f(x.c)$

Standardization is a special kind of transformation in which we make the expected value equal to zero and the variance equal to one.

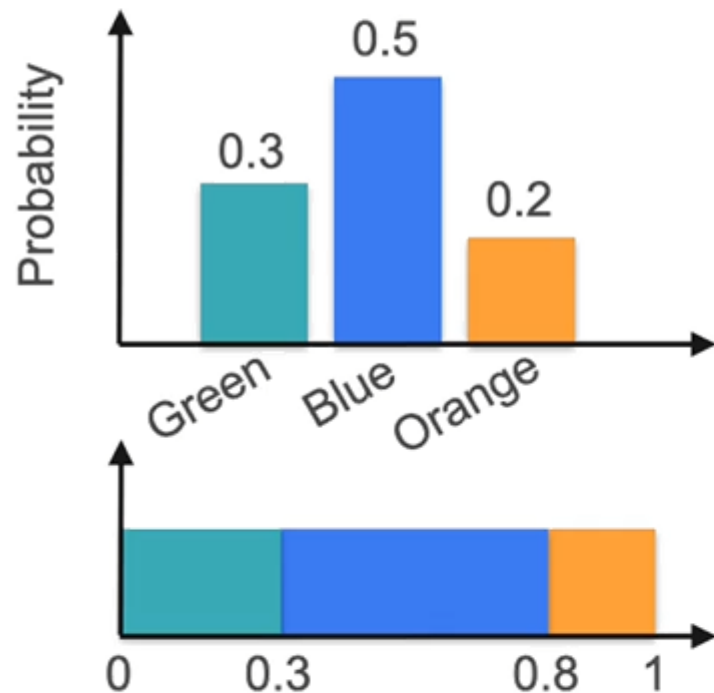
First, we wish to move the graph either to the left or to the right until its mean equals zero by subtracting the mean μ , from every element.

Second, to make the standardization complete we need to make sure the standard deviation is one by divide every element of the newly obtained distribution by the value of the standard deviation, σ .



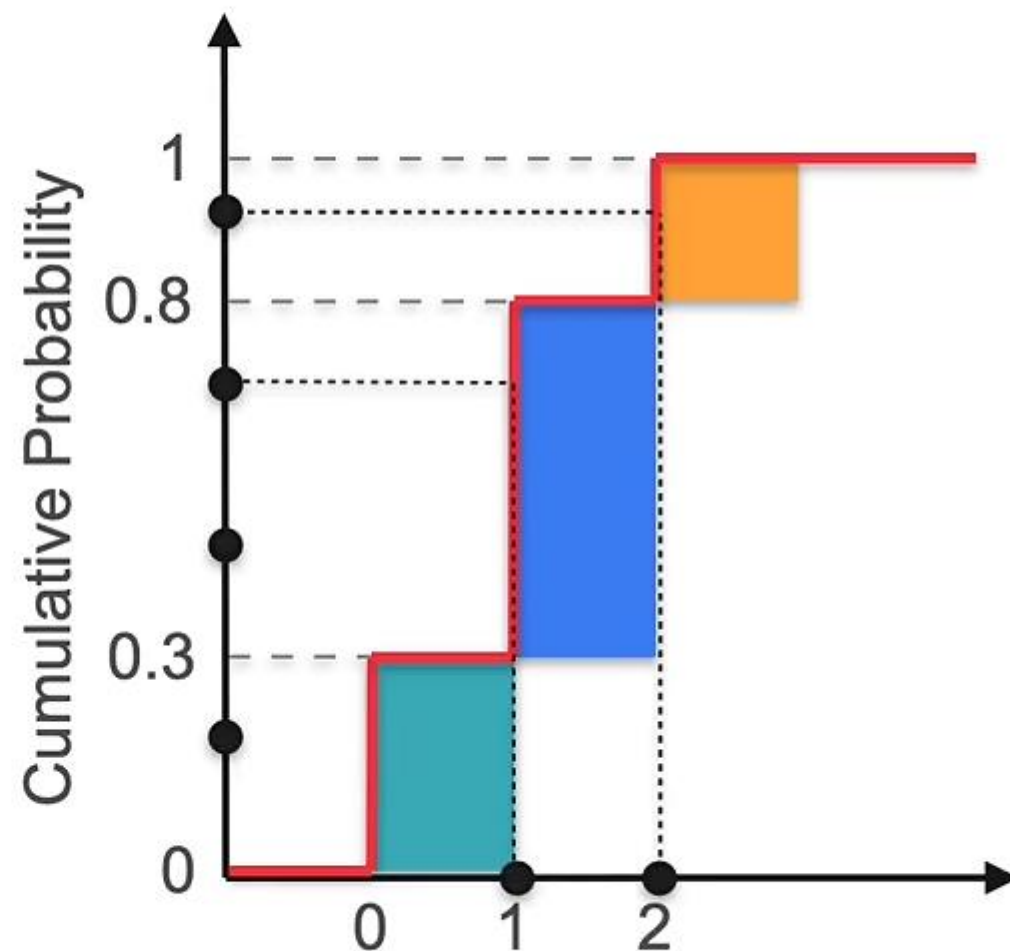
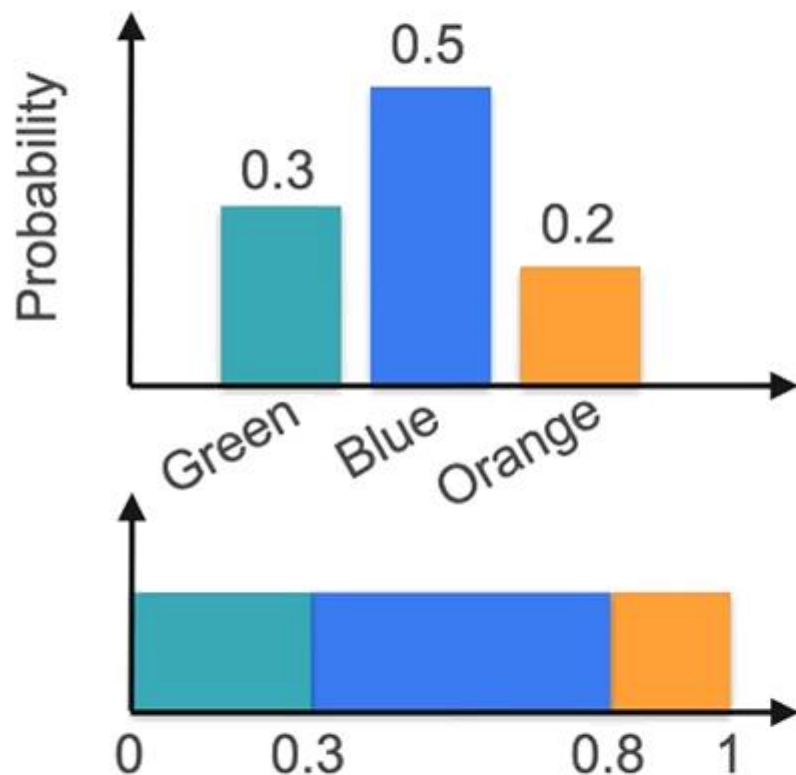
$$Z = \frac{X - \mu}{\sigma}$$

Sampling from distribution is create some synthetic data that looks a lot like the original one by construct a distribution out of this data.



1. Generate random number between Zero and One.
2. Find out which interval that number is belong to:
 - $0 \rightarrow 0.3$
 - $0.3 \rightarrow 0.8$
 - $0.8 \rightarrow 1$
3. Assign an outcome based on the interval.

Sampling from Distribution



Sampling from Distribution

