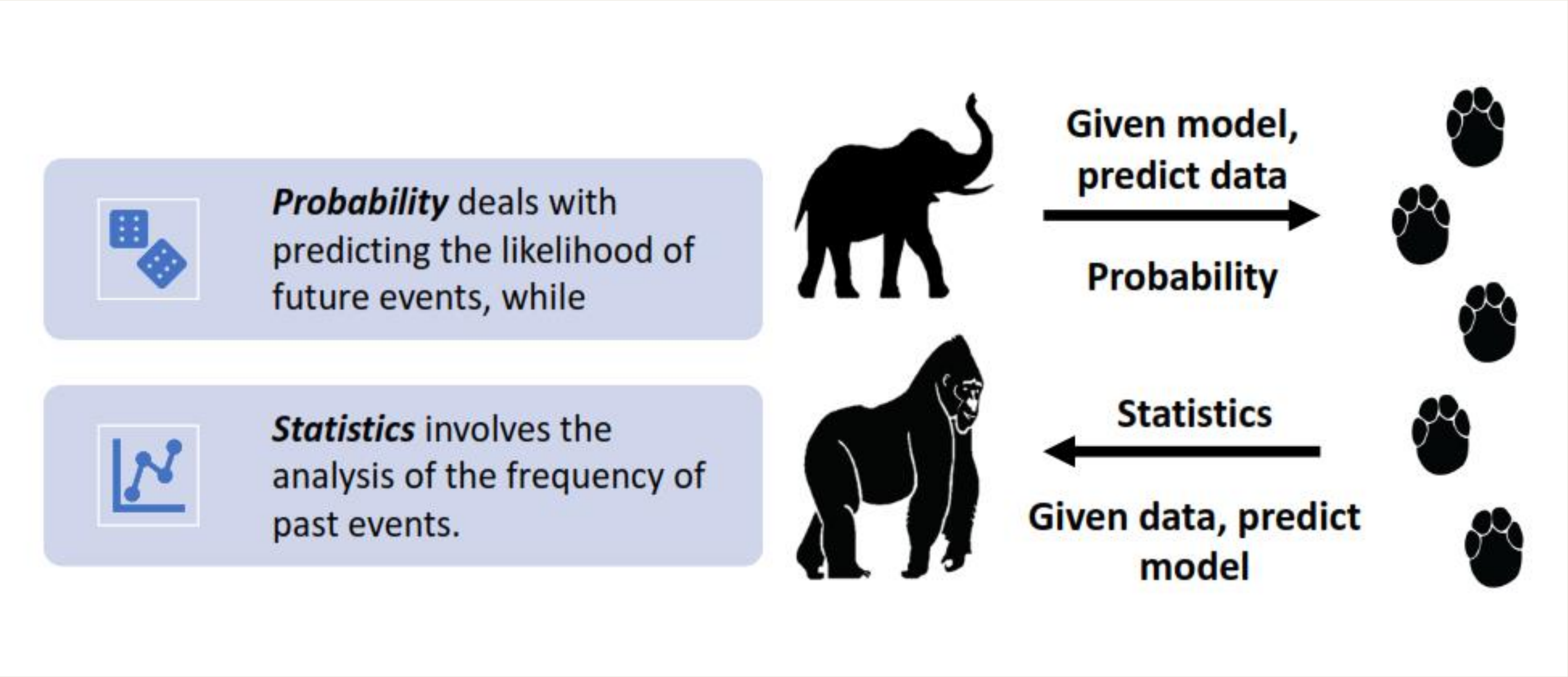
PROBABILITY AND STATISTICS FOR MACHINE LEARNING

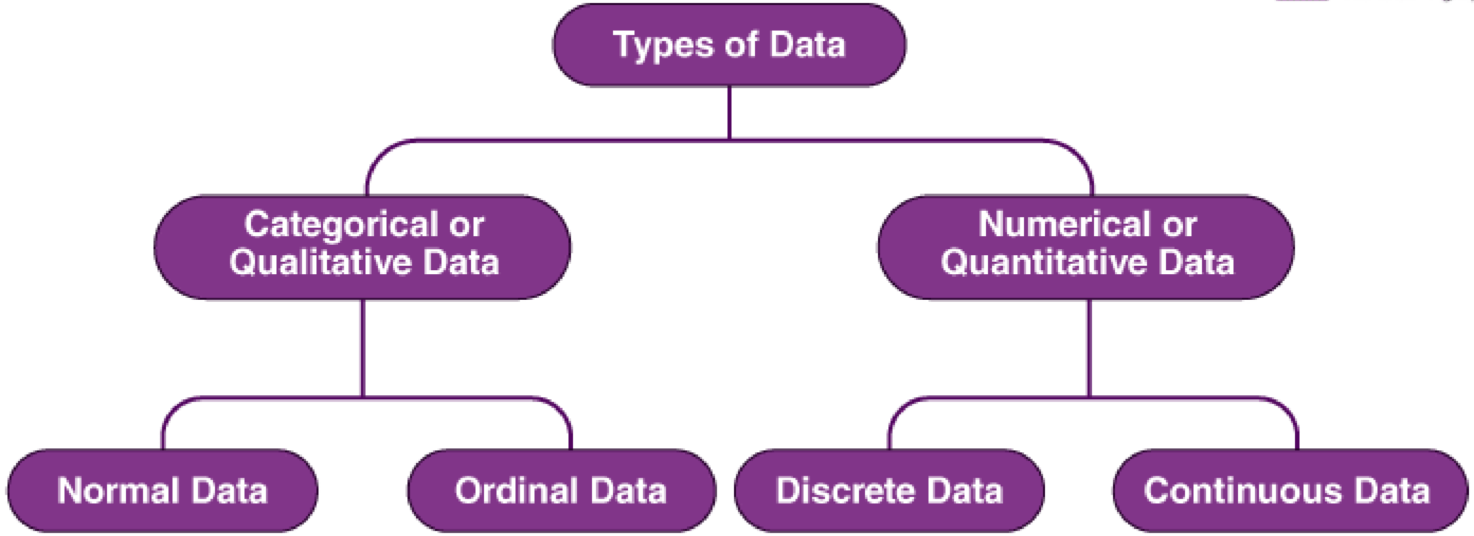
By/ Sarah Abdelmoaty

Outline

- Introduction to Statistics
- Probability theory
- Probability Distributions

Probability vs Statistics





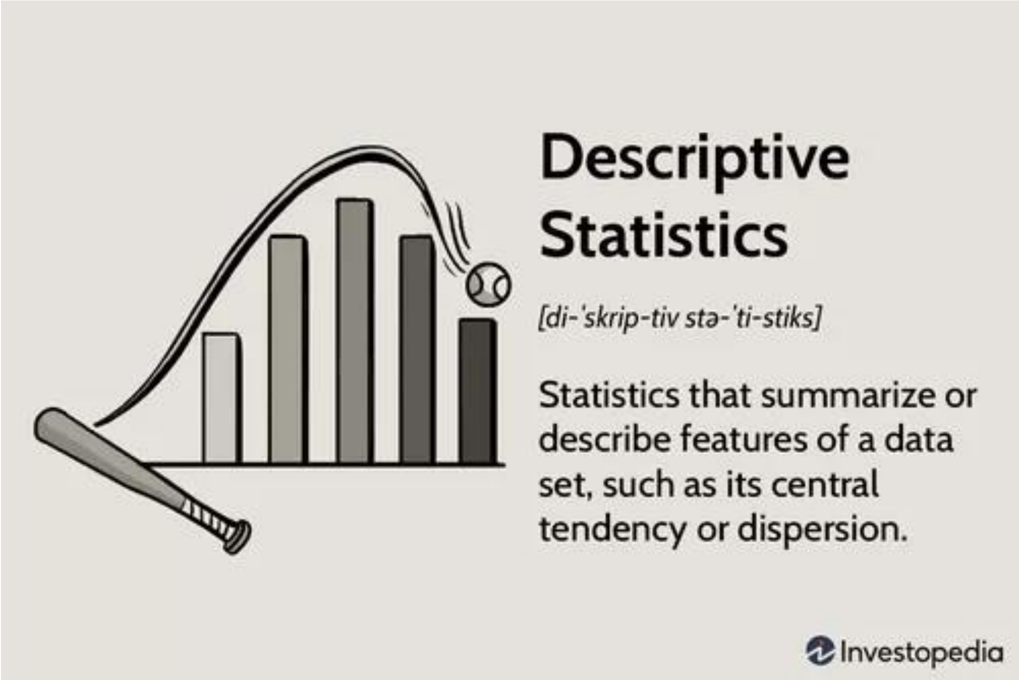
By/ Sarah Abdelmoaty

Types of Data in Statistics

- The data are the individual pieces of factual information recorded, and it is used for the purpose of the analysis process.
- The two processes of data analysis are interpretation and presentation.
- Statistics are the result of data analysis.

Name	Age	Gender	level	Courses
Ahmed	20	M	1	5
Ali	22	M	1	6
Sarah	23	F	2	4
Hamza	21	M	2	3
Radwa	19	F	1	5

EXAMPLE



Descriptive Statistics

- Use a little data to stand in for a lot of data.
- Descriptive statistics summarizes or describes the characteristics of a data set.
- Descriptive statistics consists of three basic categories of measures: measures of central tendency, measures of variability (or spread), and frequency distribution.
- Measures of central tendency describe the center of the data set (mean, median, mode).
- Measures of variability describe the dispersion of the data set (variance, standard deviation).
- Measures of frequency distribution describe the occurrence of data within the data set (count).

Statistical Language - Measures of Central Tendency

Mode

- **The mode is the *most commonly occurring value* in a distribution**
- Consider this dataset showing the retirement age of 11 people, in whole years:
- 54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60
- The most commonly occurring value is 54, therefore the mode of this distribution is 54 years.

Median

- **The median is the *middle value* in distribution when the values are arranged in ascending or descending order.**
- Looking at the retirement age distribution (which has 11 observations), the median is the middle value, which is 57 years:

  54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

  When the distribution has an even number of observations, the median value is the mean of the two middle values. In the following distribution, the two middle values are 56 and 57, therefore the median equals 56.5 years:

  52, 54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

Mean

- **. The mean is the sum of the value of each observation in a dataset divided by the number of observations. This is also known as the arithmetic average.**
- Looking at the retirement age distribution again:

  54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

  The mean is calculated by adding together all the values (54+54+54+55+56+57+57+58+58+60+60 = 623) and dividing by the number of observations (11) which equals 56.6 years.

5 Number Summary

Min	=	Smallest number
Q1	=	Median of the first half of the data
Q2	=	Median
Q3	=	Median of the second half of the data
Max	=	Largest number

Range = max - min
Interquartile range (IQR) = Q3 - Q1

Five Number Summary For Data Set:
1,2,3,4,5,11,11,12,14,20,20

1,2.5,8,13,20

Minimum
1st Quartile
Median
3rd Quartile
Maximum

boxplot

minimum

Lower quartile

Median

Upper quartile

maximum

22, 30, 32, 35, 35, 39, 41, 45, 51, 60

Q1 Q2 Q3

37

$\left(\frac{35+39}{2}\right)$

IQR

Middle of the bottom half is 32

Middle of the top half is 45

Statistical Language - Measures of Spread

- Variance and Standard Deviation are essentially a measure of the spread of the data in the data set.
- Variance is the average of the squared differences from the mean.

- For Further information https://www.abs.gov.au/websitedbs/D3310114.nsf/home/statistical+language+-+measures+of+spread

$$Variance, \sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

$$Standard\ Deviation, \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

Where $x_i = data\ set\ values$

$\bar{x} = mean\ of\ the\ data\ set$

Table 1: Mean and Standard Deviation				
Measure Name	Symbol for Population	Symbol for Sample	Computation for Population	Computation for Sample
Mean	μ	$\overline{x}$	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
Standard Deviation	σ	s_x	$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$ or the equivalent form $\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x^2) - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N}}$	$s_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$ or the equivalent form $s_x = \sqrt{\frac{\sum_{i=1}^{n} (x^2) - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$

$$Cov(X,Y) = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})$$

where

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

are the means of X, Y

Statistical Tools

- Covariance and correlation are the tools that we must measure if the two attributes are related to each other or not.
- Covariance measures how two variables vary in tandem to their means .

Statistical Tools

$$r = r_{xy} = \frac{\mathrm{Cov}(x,y)}{S_x \times S_y}$$

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Correlation also measures how two variables move with respect to each other.

A perfect positive correlation means that the correlation coefficient is 1.

A perfect negative correlation means that the correlation coefficient is -1.

A correlation coefficient of 0 means that the two variables are uncorrelated.

Measures of Shape

What is a measure of shape?

Measures of shape describe the distribution (or pattern) of the data within a dataset.

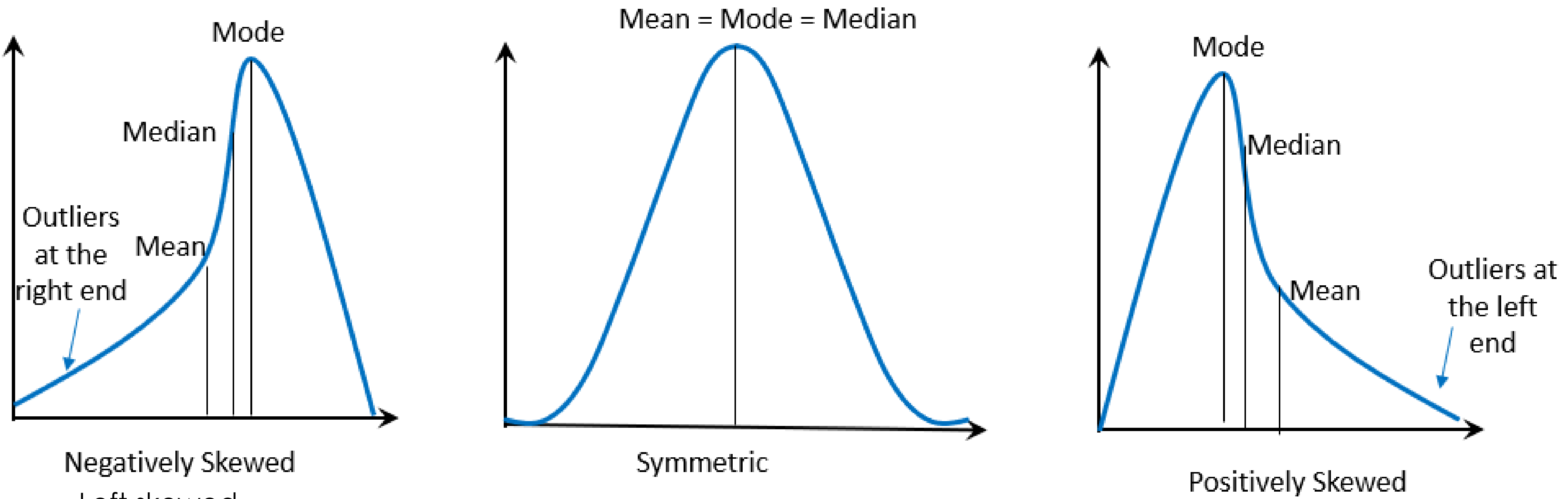
The distribution shape of quantitative data can be described as there is a logical order to the values, and the 'low' and 'high' end values on the x-axis of the histogram are able to be identified.

The distribution shape of a qualitative data cannot be described as the data are not numeric.

For further information: https://www.abs.gov.au/websitedbs/D3310114.nsf/home/statistical+language+-+measures+of+shape

STATISTICAL LANGUAGE - MEASURES OF SHAPE

STATISTICAL LANGUAGE - MEASURE OF SHAPE



Left skewed
Median > Mean

Right skewed
Median <Mean

By/ Sarah Abdelmoaty

Probability

Number used to quantify the likelihood, or chance, that an outcome of a random experiment will occur.

We can obtain probabilities from:

- Subjective experience
- Observations and Data collection
- Computations using mathematical rules

Random Experiment

• An experiment that can result in

Different outcomes, even though it is repeated in the same manner every

Time.

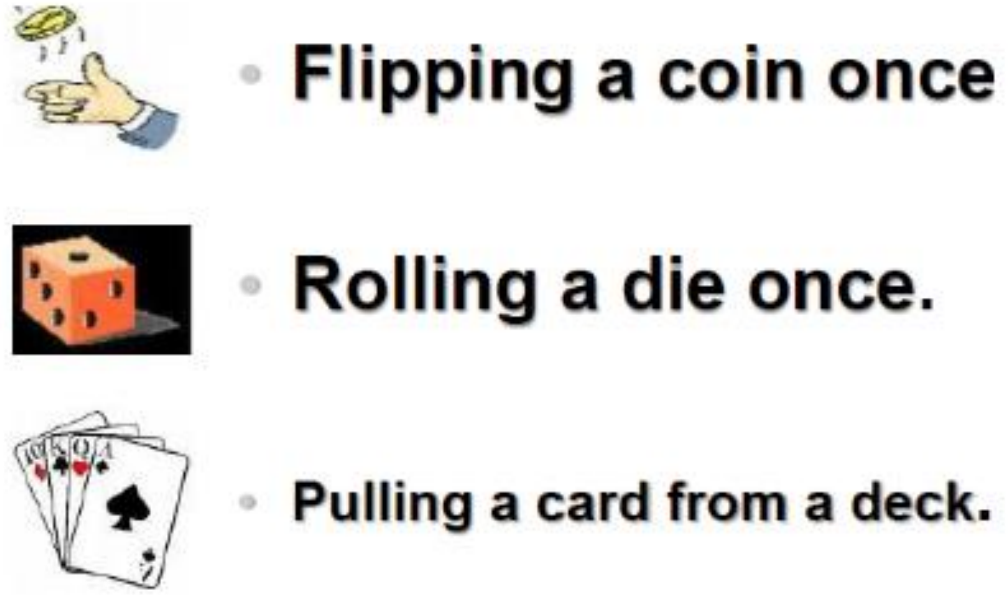
• An outcome is the result of a single

Trial.

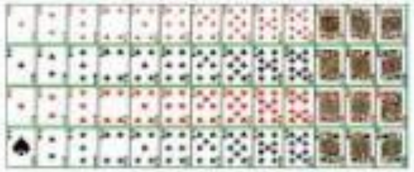
• Sample space 'S': is the set of all

Possible outcomes

• Event: a subset of S.



SAMPLE SPACES, OUTCOMES, AND EVENTS

Experiment	Sample space
Tossing coin once	Head, Tail 
Roll a die 	1,2,3,4,5,6 
Answer a true or false question	True, False
Draw a card from a regular deck	52 cases 

CLASSICAL PROBABILITY

The (classical) probability, $P(E)$, of an event E is defined as:

$$\boxed{P(E) = \frac{N(E)}{N(S)} = \frac{|E|}{|S|}}$$

where $N(E)$ or $|E|$ is the number of elements in the set E.

(assuming all events are equally likely)

EMPIRICAL VS THEORETICAL PROBABILITY

Experimental Probability is found by repeating an experiment and observing the outcomes.

$$P(\text{event}) = \frac{\text{number of times event occurs}}{\text{total number of trials}}$$

Example:

A coin is tossed 10 times:

A head is recorded 7 times and a tail 3 times.

$$P(\text{head}) = \frac{7}{10}$$

$$P(\text{tail}) = \frac{3}{10}$$

Theoretical Probability is what is expected to happen based on mathematics

$$P(\text{event}) = \frac{\text{number of favorable outcomes}}{\text{total number of possible outcomes}}$$

Example:

A coin is tossed.

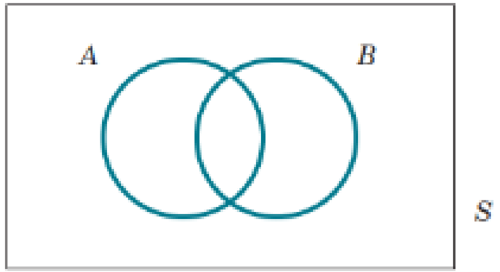
$$P(\text{head}) = \frac{1}{2}$$

$$P(\text{tail}) = \frac{1}{2}$$

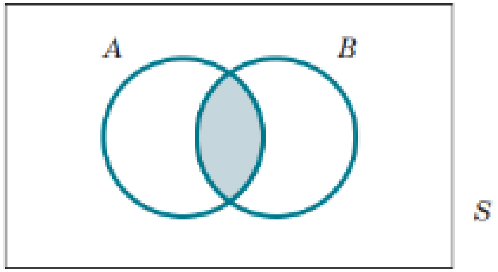
Basic Probability Concepts

Event
An **event** is a subset of the sample space of a random experiment.

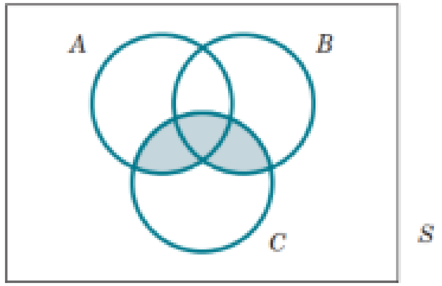
Sample space S with events A and B



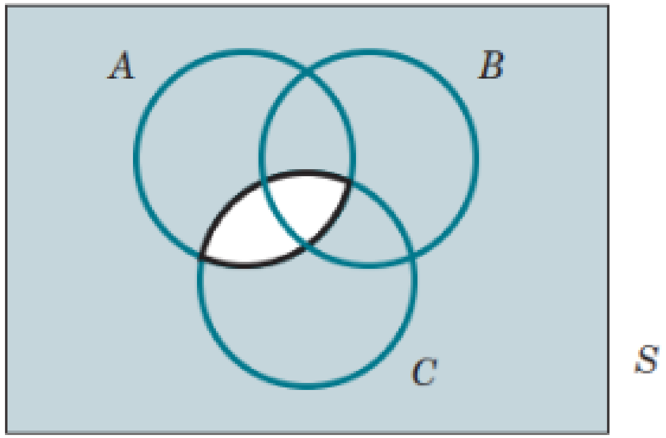
$A \cap B$



$(A \cup B) \cap C$



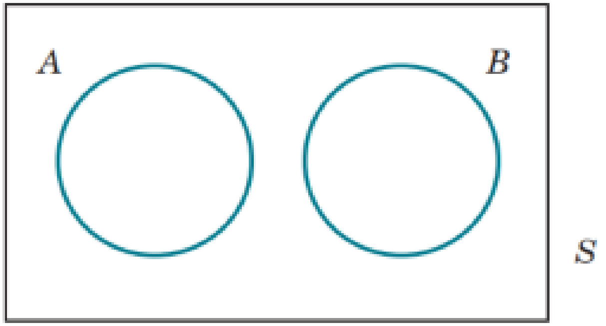
$(A \cap C)'$



Mutually Exclusive Events
Two events, denoted as E_1 and E_2, such that

$$E_1 \cap E_2 = \emptyset$$

are said to be **mutually exclusive**.



Fundamentals of Probability

Axioms of Probability

Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties:

(1) $P(S) = 1$ where S is the sample space
(2) $0 \le P(E) \le 1$ for any event E
(3) For two events E_1 and E_2 with $E_1 \cap E_2 = \emptyset$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

E1 and E2 are disjoint (mutually exclusive)

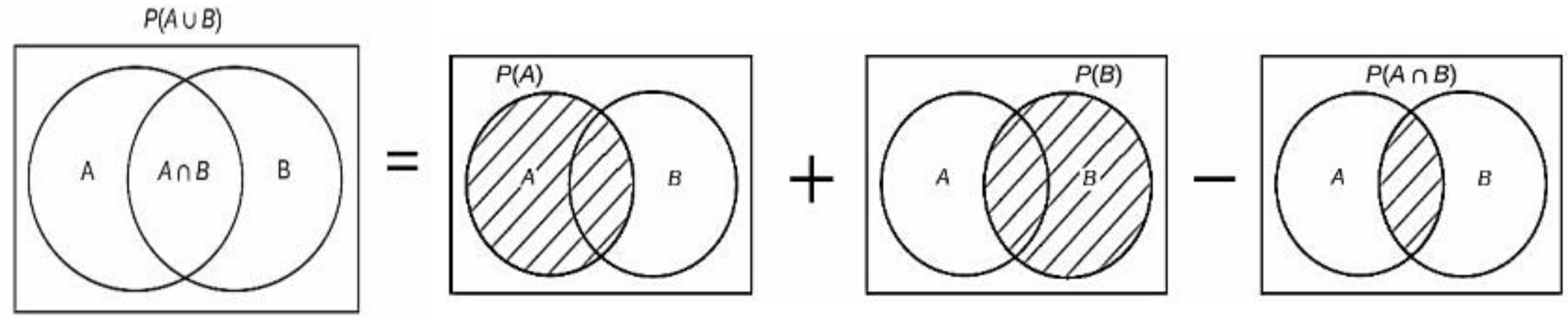
- Probability of impossible\Empty event: $P(\emptyset) = 0$

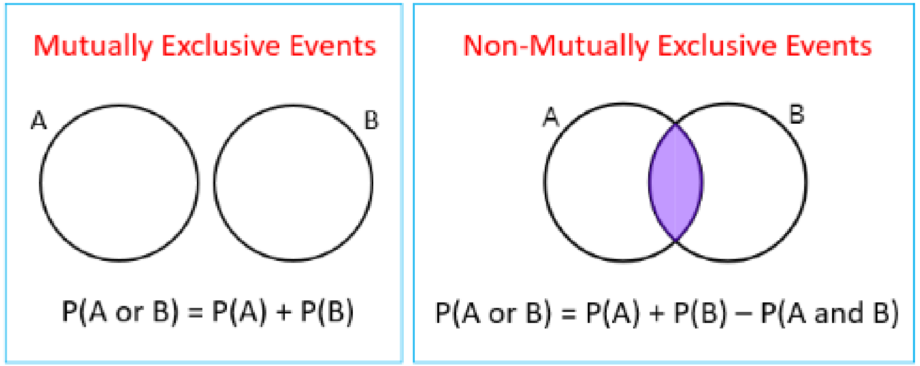
- Probability of complementary event: $P(E') = 1 - P(E)$

Union

Probability of a Union

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$





Example

- S={1,2,3,4,5,6}
- A={1,3,5} B={2,4} C={4,5,6}
- P(S) =
- P(A) =
- P(B) =
- P(A ∪ B) =
- P(A ∪ C) =

Examples

- If you flipped a coin 1 time, find the probability of getting **a head**.

- If you flipped a coin 2 times, find the probability of getting **two heads**.

- If you flipped a coin 3 times, find the probability of getting **two heads** and **one tail**.

- Two dice are rolled. Find the probabilities of the following events.
  1. Getting two even numbers.
  2. Getting two numbers whose sum is 7

If A and B are independent, then

1. $P(A|B) = P(A)$
2. $P(B|A) = P(B)$
3. $P(A \cap B) = P(A)P(B)$

Independence of random events

- If the occurrence or nonoccurrence of either events A and B do not affect the other, then A and B are independent.

Independence of random events

Independent Events

The outcome of one event does not affect the outcome of the other.

If A and B are independent events then the probability of both occurring is

$$P(A \text{ and } B) = P(A) \times P(B)$$

Dependent Events

The outcome of one event affects the outcome of the other.

If A and B are dependent events then the probability of both occurring is

$$P(A \text{ and } B) = P(A) \times P(B|A)$$

Probability of B given A

Bayes Theorem

- Principled way of calculating a conditional probability without the joint probability.
- It is often the case that we do not have access to the denominator directly, e.g., P(B).
- P(B) = P(B|A) * P(A) + P(B|not A) * P(not A)

- This gives a formulation of Bayes Theorem that we can use that uses the alternate calculation of P(B), described below:
- P(A|B) = P(B|A) * P(A) / P(B|A) * P(A) + P(B|not A) * P(not A)
- Or with brackets around the denominator for clarity:
- P(A|B) = P(B|A) * P(A) / (P(B|A) * P(A) + P(B|not A) * P(not A))
- **Note**: the denominator is simply the expansion we gave above.

- As such, if we have P(A), then we can calculate P(not A) as its complement; for example:

P(not A) = 1 – P(A)

- Additionally, if we have P(not B|not A), then we can calculate P(B|not A) as its complement; for example:

P(B|not A) = 1 – P(not B|not A)

Notations of Bayes Theorem

- Firstly, in general, the result P(A|B) is referred to as the **posterior probability** and P(A) is referred to as the **prior probability**.
- P(A|B): Posterior probability.
- P(A): Prior probability.

- Sometimes P(B|A) is referred to as the **likelihood** and P(B) is referred to as the **evidence**.
- P(B|A): Likelihood.
- P(B): Evidence.

- This allows Bayes Theorem to be restated as:
- Posterior = Likelihood * Prior / Evidence
- We can make this clear with a smoke and fire case.

Some illustrations will improve the understanding of the concept.

Example 1:

A bag I contains 4 white and 6 black balls while another Bag II contains 4 white and 3 black balls. One ball is drawn at random from one of the bags, and it is found to be black. Find the probability that it was drawn from Bag I.

Solution:

Let E_1 be the event of choosing bag I, E_2 the event of choosing bag II, and A be the event of drawing a black ball.

Then,

$$P(E_1) = P(E_2) = \frac{1}{2}$$

Also, $P(A|E_1)$ = P(drawing a black ball from Bag I) = 6/10 = 3/5

$P(A|E_2)$ = P(drawing a black ball from Bag II) = 3/7

By using Bayes' theorem, the probability of drawing a black ball from bag I out of two bags,

$$P(E_1|A) = \frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1)+P(E_2)P(A|E_2)}$$

$$= \frac{\frac{1}{2} \times \frac{3}{5}}{\frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{3}{7}}$$

$$= \frac{7}{12}$$

Example 2:

A man is known to speak the truth 2 out of 3 times. He throws a die and reports that the number obtained is a four. Find the probability that the number obtained is actually a four.

Solution:

Let A be the event that the man reports that number four is obtained.

Let E_1 be the event that four is obtained and E_2 be its complementary event.

Then, $P(E_1)$ = Probability that four occurs = 1/6.

$P(E_2)$ = Probability that four does not occur = 1- $P(E_1)$ = 1 – (1/6) = 5/6.

Also, $P(A|E_1)$= Probability that man reports four and it is actually a four = 2/3

$P(A|E_2)$ = Probability that man reports four and it is not a four = 1/3.

By using Bayes' theorem, probability that number obtained is actually a four, $P(E_1|A)$

$$= \frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1) + P(E_2)P(A|E_2)} = \frac{\frac{1}{6} \times \frac{2}{3}}{\frac{1}{6} \times \frac{2}{3} + \frac{5}{6} \times \frac{1}{3}}$$

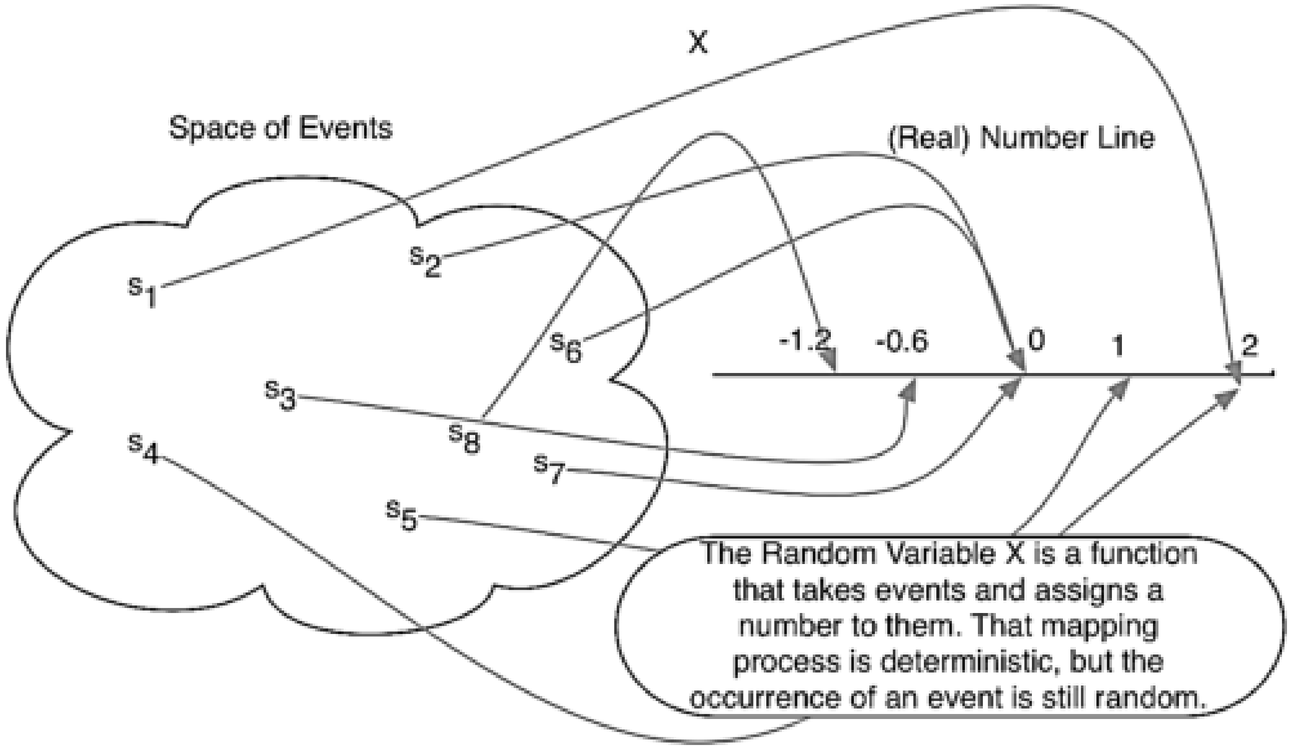
$$= \frac{2}{7}$$

Random variable

Random variable : denotes a value that depends on the result of some random experiment.

A Random Variable is a set of **possible values** from a random experiment.

A Random Variable has a whole **set of values**, and it could take on **any** of those values, randomly.



Types of random variables

Discrete random variables

- Countable (finite). It can take only distinct, separate values.
    - (e.g., number of cars in a street)

Continuous random variables

- Measurable, uncountable. It can take on any value in some interval (low, high).
    - Continuous Data can take any value within a range (such as a person's height, weight, temperature, ...)

What Is a Probability Distribution

A probability distribution is a function that describes all the possible values and likelihoods that a random variable can take within a given range.

This range will be bounded between the minimum and maximum possible values, but precisely where the possible value is likely to be plotted on the probability distribution depends on a number of factors that include the distribution's mean (average), and the standard deviation

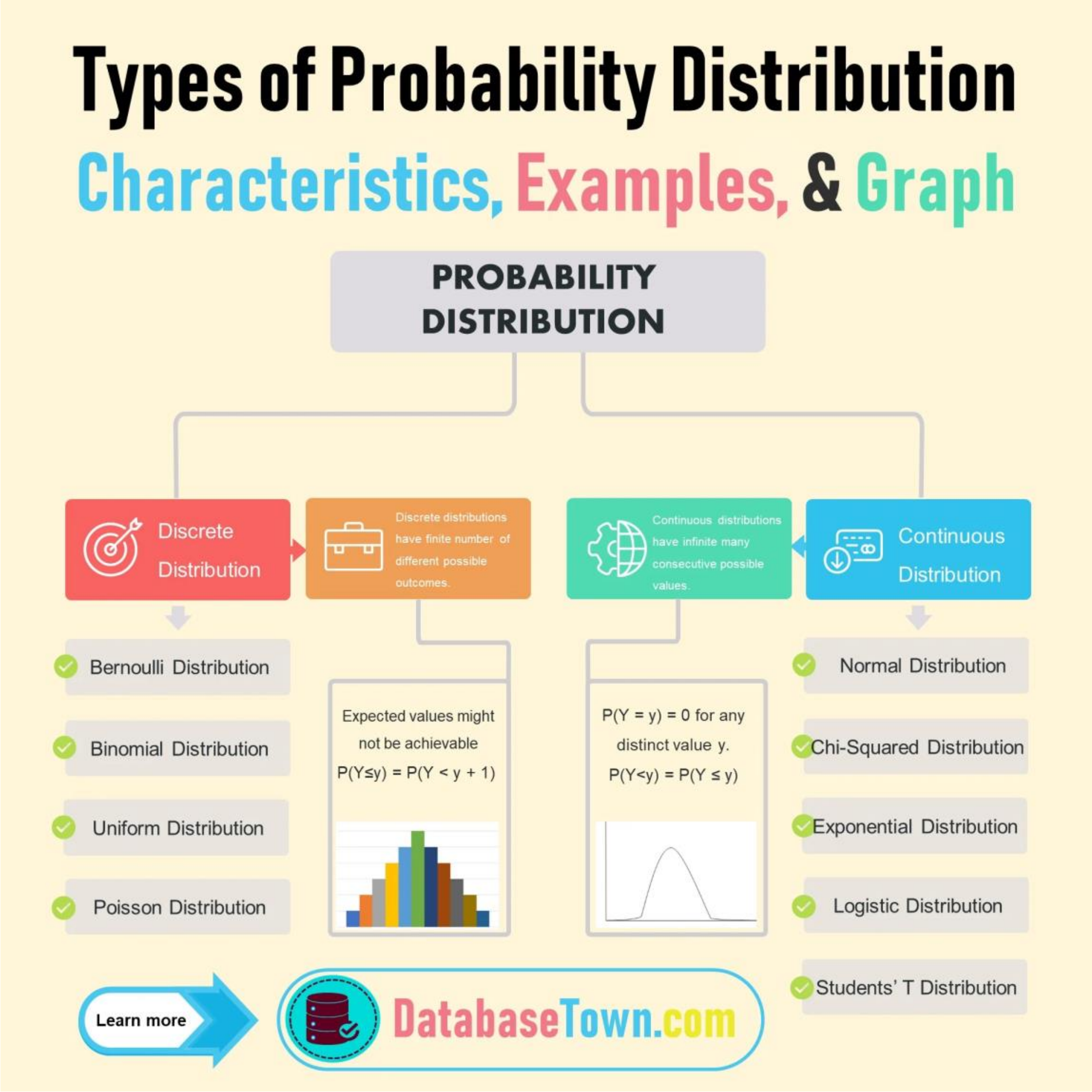
For further information:
https://www.scribbr.com/statistics/probability-distributions/

Probability Distributions

A **probability distribution**, like a frequency distribution, can be only partially described by a graph. To aid in a decision situation, you may need to calculate the distribution's mean and standard deviation. These values measure the central location and spread, respectively, of the probability distribution.

Two classes of random variables

Discrete Random Variables | Continuous Random Variables.



Types of Probability Distribution
Characteristics, Examples, & Graph
PROBABILITY DISTRIBUTION
Discrete Distribution
Discrete distributions have finite number of different possible outcomes.
Continuous distributions have infinite many consecutive possible values.
Continuous Distribution
Bernoulli Distribution
Binomial Distribution
Uniform Distribution
Poisson Distribution
Expected values might not be achievable
P(Y≤y) = P(Y < y + 1)
P(Y = y) = 0 for any distinct value y.
P(Y<y) = P(Y ≤ y)
Normal Distribution
Chi-Squared Distribution
Exponential Distribution
Logistic Distribution
Students' T Distribution
Learn more
DatabaseTown.com

THANK YOU

sarahdreny11@gmail.com

LinkedIn : Sarah Abdelmoaty