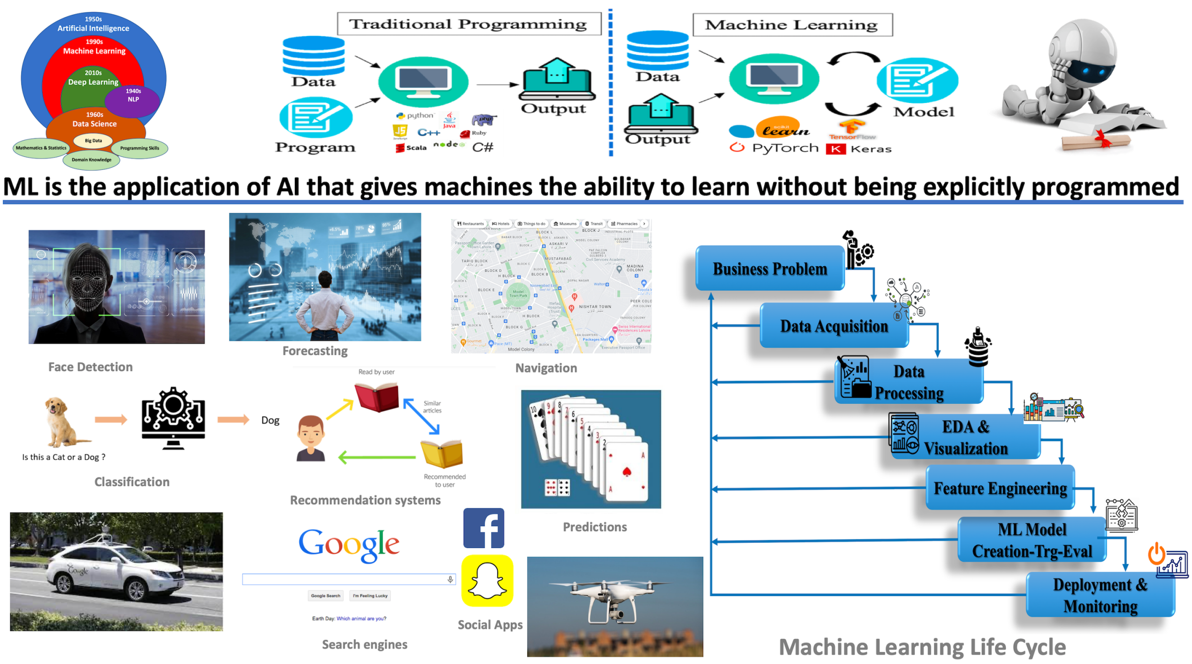
Department of Data Science

Course: Tools and Techniques for Data Science

Instructor: Muhammad Arif Butt, Ph.D.

Lecture 6.25 (A Comprehensive Recap of Probability for Bayes' Theorem)

Open in Colab
(https://colab.research.google.com/github/arifpucit/data-science/blob/master/Section-4-Mathematics-for-Data-Science/Lec-4.1(Descriptive-Statistics).ipynb)



Learning agenda of this notebook

A Comprehensive Recap of Probability and Bayes' Theorem

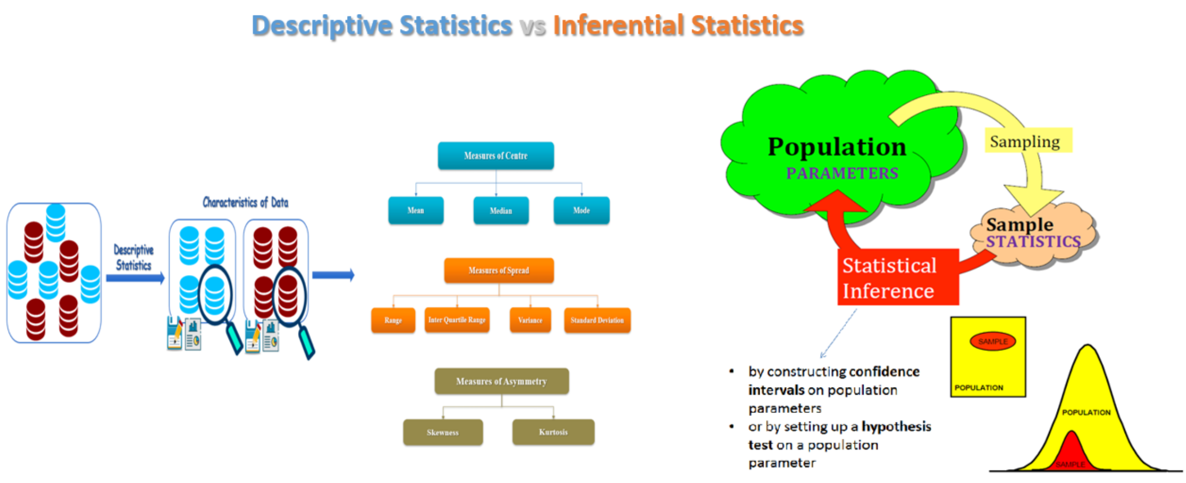
- Overview of Probability and Statistics
  - Classical vs Emperical Probability
  - Some Important Terms related to Probability
- Marginal/Simple/Unconditional Probability (with examples)
- Types of Events
  - Mutually Exclusive vs Not Mutually Exclusive (with examples)

- Mutually Exclusive vs Not Mutually Exclusive (with examples)
    - Independent and Dependent Events (with examples)
- Joint Probability
    - Reducing the Sample Space
    - Use of Venn Diagrams and Contigency Tables
    - Examples
- Conditional Probability
    - Intuition
    - Examples

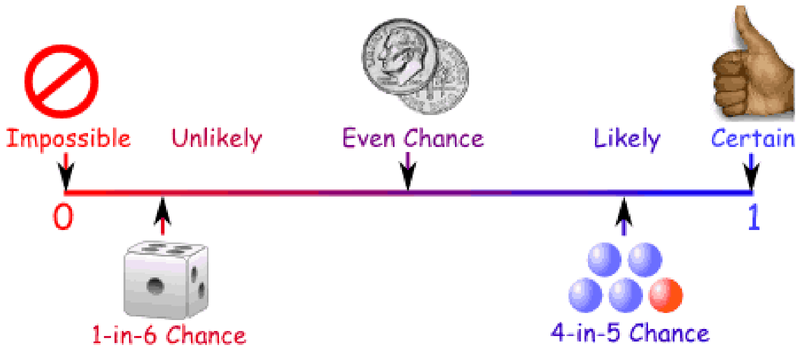
1. Overview of Probability and Statistics

> - Descriptive Statistics: https://www.youtube.com/watch?v=Q6EZ-KGm3UE&list=PL7B2bn3G_wfDTEU20jwGcUKlVqdAyQMax&index=1 (https://www.youtube.com/watch?v=Q6EZ-KGm3UE&list=PL7B2bn3G_wfDTEU20jwGcUKlVqdAyQMax&index=1)
> - Inferential Statistics: https://www.youtube.com/watch?v=y4t9ELO3Nxc&list=PL7B2bn3G_wfDTEU20jwGcUKlVqdAyQMax&index=2 (https://www.youtube.com/watch?v=y4t9ELO3Nxc&list=PL7B2bn3G_wfDTEU20jwGcUKlVqdAyQMax&index=2)

- **Statistics** is the area of applied mathematics that deals with data collection & organization, data analysis, data interpretation and data representation.
    - `Descriptive Statistics` is used to describe, summarize, and present different datasets through numerical calculations and graphs.
    - `Inferential Statistics` is used to make prediction and inferences for an entire population based on sample data from that population



a. Classical vs Emperical Probability?



- Probability theory is a branch of mathematics concerned with the analysis of random phenomena, and defines the `likelihood of occurrence of an event`.
- The value of the probability of an event to happen can lie between 0 and 1 because the favorable number of outcomes can never cross the total number of outcomes.
- When the probability of something approaches 1, then it means it is very likely, and when the probability of something approaches 0, then it means that it is very unlikely.
- To understand Probability, we normally start to predict the outcomes for the `tossing of coins`, `rolling of dice`, or `drawing a card from a pack of playing cards`.

- **Classical/Theoratical Probability:**
  - It is a probability measure that is used when you `DO NOT have past data` and is based on theoratical assumptions

$$P(\text{event}) = \frac{\text{Number of favourable outcomes of event}}{\text{Total outcomes in sample space}}$$

  - Consider a scenario of tossing a coin hundred times (without performing the experiment). The theoratical probability of getting a head is : $P(H) = \frac{50}{100} = 0.5$

- **Emperical/Experimental Probability:**
  - It is a probability measure that is `based on observed/past data`. It donot use sample space rather use frequency distributions

$$P(\text{event}) = \frac{\text{Frequency of event}}{\text{Sum of the frequencies}}$$

  - Suppose we toss a coin hundred times and get heads 60 times and tails 40 times. The emperical probability of getting a head is: $P(H) = \frac{60}{100} = 0.6$

- When we `increase the number of trials/tosses` the result get close and close to 0.5. This is called the `Law of Large Numbers` , which states that the more experiments we run, the closer we will tend to get to the Classical probability.

b. Important Terms Related to Probability

- **Random Experiment:** An experiment is called random experiment if it satisfies following two conditions:
  - It has more than two possible outcomes
  - It is not possible to predict the outcome in advance
- **Trial:** It refers to a single execution of a random experiment. Each trial produces an outcome.
- **Sample Space:** It is the `set of all possible outcomes` that can occur. Generally one random experiment will have one set of sample space. Sample space for tossing a coin is {H, T}, sample space for rolling a die is {1,2,3,4,5,6}, sample space for tossing two coins is {HH,HT,TH,TT}
- **Outcome:** It is the `result of a single trial of a random experiment`
- **Event:** Event is a `subset of sample space` and can include a single outcome or multiple outcomes. One random experiment can have multiple events.

Tossing a Coin Twice:

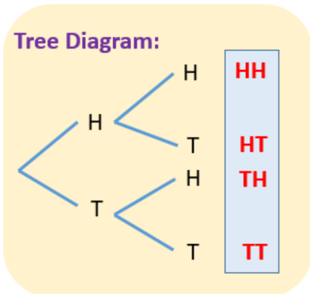
- **Random Experiment:** Tossing a Coin
- **Trial:** Tossing a coin (twice)
- **Sample Space:** {(HH), (TH), (HT), (TT)}
- **Outcome:** {H,T}
- **Event:**
  - Getting two heads {(HH)}
  - Getting at least one head{((HH), (TH), (HT)}

List:

HH HT TH TT

Table:

	H	T
H	HH	HT
T	TH	TT



Rolling two Dice:

- **Random Experiment:** Rolling two dice
- **Trial:** Rolling two dice (once)
- **Sample Space:** {(1,1), (1,2), ... (6,6)}
- **Outcome:** {(3,4}
- **Event:**
  - Getting a sum of 12 {(6,6)}
  - Getting a sum of 4 {(3,1), (2,2), (1,3)}
  - Getting numbers less than 3 on both dice {(1,1), (2,1), (1,2)}

(a,b)	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Titanic Dataset:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True	NaN	Southampton	no	True
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False	B	Southampton	yes	True
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False	NaN	Southampton	no	False
889	1	1	male	26.0	0	0	30.0000	C	First	man	True	C	Cherbourg	yes	True
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True	NaN	Queenstown	no	True

891 rows × 15 columns

- **Random Experiment:** Randomly selecting a passenger and finding his/her pclass
- **Trial:** Randomly selecting a passenger and finding its pclass (once)
- **Sample Space:** {1, 2, 3}
- **Outcome:** {1}
- **Event:**
  - Getting a passenger belonging to pclass 2: {2}
  - Getting a passenger NOT belonging to pclass 3: {1, 2}

2. Marginal/Simple/Unconditional Probability

Marginal Probability is the probability of a single event occurring, independent of the outcome of other events.

Example 1: Consider a scenario of tossing a fair coin three times.

- Event A: Getting exactly two heads:

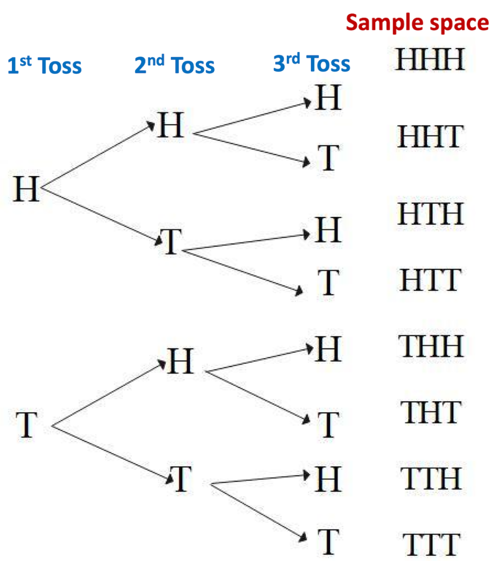
$$P(A) = \frac{3}{8} = 0.375$$

- Event B: Getting three consecutive heads:

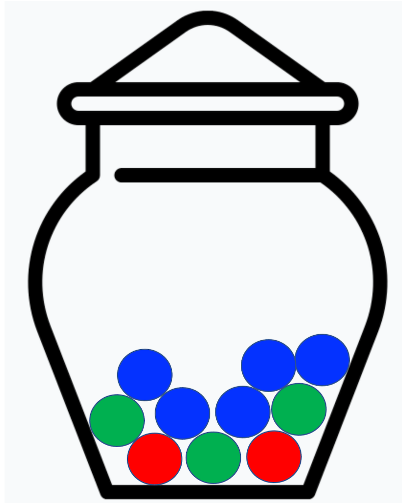
$$P(B) = \frac{1}{8} = 0.125$$

- Event C: Getting at least two heads:

$$P(C) = \frac{4}{8} = 0.5$$



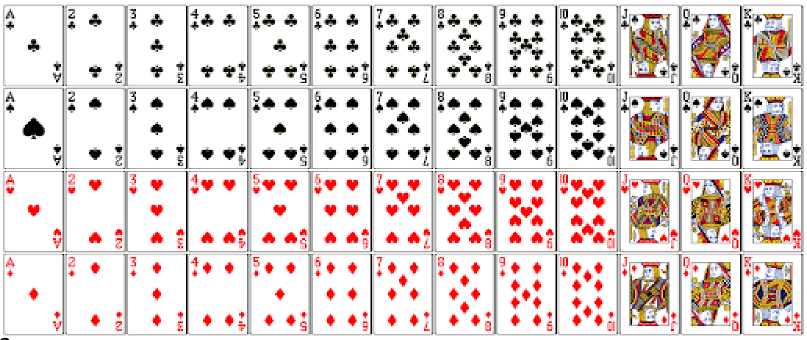
Example 2: An urn contains 2 red balls, 3 green balls, and 5 blue balls. A single ball is selected at random `with replacement`



- Event A: Getting a red ball:
$$P(A) = \frac{2}{10} = 0.2$$
- Event B: Getting a green ball:
$$P(B) = \frac{3}{10} = 0.3$$
- Event C: Getting a blue ball:
$$P(C) = \frac{5}{10} = 0.5$$

Example 3: Consider drawing a single card from a standard deck of 52 playing cards (Clubs, Spades, Hearts, Diamonds). In this case, the number of possible outcomes in the sample space Ω is 52.

- Event A: Getting a `spades` card:



$$P(A) = \frac{13}{52} = 0.25$$

- Event B: Getting a `ace` card: $P(B) = \frac{4}{52} \approx 0.0769$
- Event C: Getting a `Queen of Hearts` card: $P(C) = \frac{1}{52} \approx 0.019$
- Event D: Getting a card: $P(D) = \frac{52}{52} = 1.0$
- Event E: Getting an `Apple` : $P(E) = \frac{0}{52} \approx 0.0$

Example 4: Consider a scenario of rolling two fair 6-sided dice.

Sample Space of rolling two dice

Die 2

	(a,b)	1	2	3	4	5	6
	1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
D	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
i	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
e	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
1	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Output: Sum of two dice

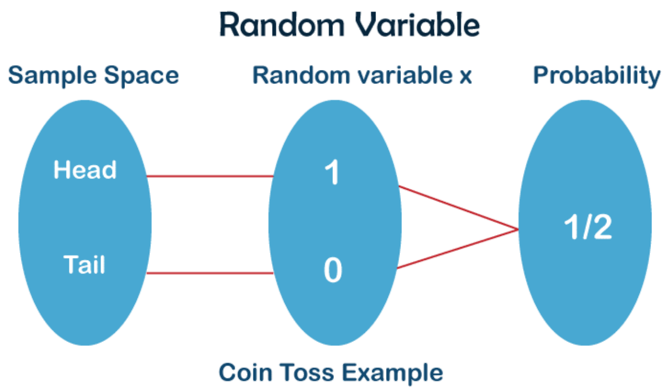
Unique o/p={2,3,4,5,6,7,8,9,10,11,12}

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

3. The Concept of Random Variables and Probability Distributions

a. Random Variable

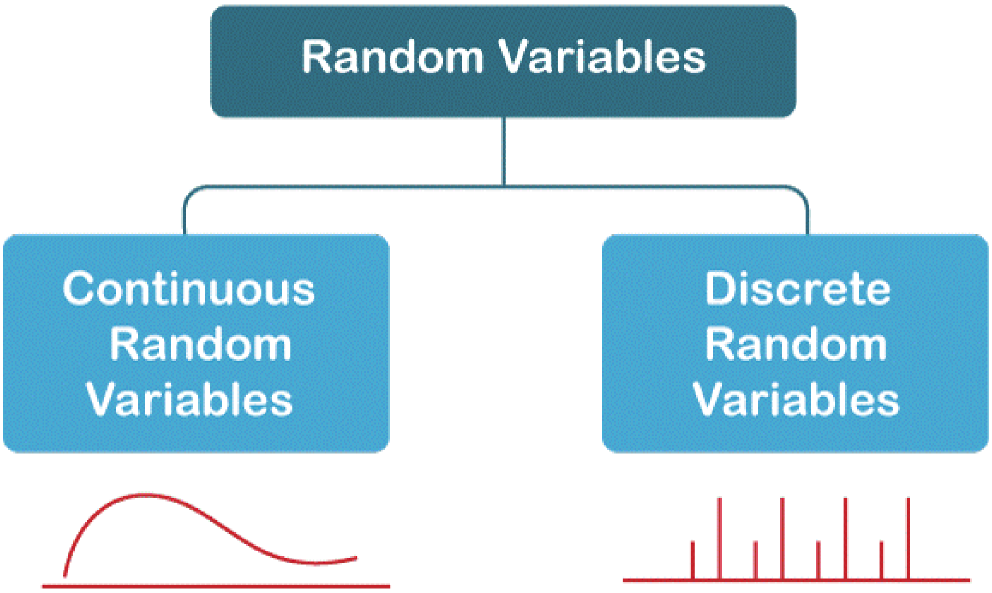
A random variable is a mathematical function that assigns a numerical value to every possible outcome of a random experiment



- Probability Distribution for the random experiment of tossing a coin is as follows:

$$P(X = 0) = 1/2$$
$$P(X = 1) = 1/2$$

b. Types of Random Variables



Discrete Random Variable:

- A `discrete random variable` is a variable which can take on discrete values as outcome, which can be counted not measured, e.g., $0, 1, 2, 3, 4, 5, \cdots$
- Examples:
  - The number of times a coin lands on tails after being flipped 20 times.
  - The number of times a dice lands on the number 4 after being rolled 100 times.
  - Number of items sold at a store on a certain day. Using historical sales data, a store could create a probability distribution that shows how likely it is that they sell a certain number of items in a day.
  - Number of customers that enter a shop on a given day. Using historical data, a shop could create a probability distribution that shows how likely it is that a certain number of customers enter the store.
  - Number of defective products produced per batch by a certain manufacturing plant. Using historical data on defective products, a plant could create a probability distribution that shows how likely it is that a certain number of products will be defective in a given batch.
  - Number of traffic accidents that occur in a specific city on a given day. Using historical data, a police department could create a probability distribution that shows how likely it is that a certain number of accidents occur on a given day.

Continouous Random Variable:

- A `continuous random variable` is a variable which can take on an infinite number of possible values as outcome, which can be measured not counted, e.g., $0,0.1,0.2,3.5,4.2,4.3,\cdots.
- Examples:
  - Weight of a certain animal like a dog is a continuous random variable because it can take on an infinite number of values (e.g., 30.333 pounds, 50.340999 pounds, 60.5 pounds). In this case, we could collect data on the weight of dogs and create a probability distribution that tells us the probability that a randomly selected dog weighs between two different amounts.

- Height of a certain species of plant is a continuous random variable because it can take on an infinite number of values (e.g., 6.5555 inches, 8.95 inches, 12.32426 inches). In this case, we could collect data on the height of this species of plant and create a probability distribution that tells us the probability that a randomly selected plant has a height between two different values.
- Distance traveled by a certain wolf during migration season is a continuous random variable because it can take on an infinite number of values (e.g., a wolf may travel 40.335 miles, 80.5322 miles, 105.59 miles). In this scenario, we could collect data on the distance traveled by wolves and create a probability distribution that tells us the probability that a randomly selected wolf will travel within a certain distance interval.

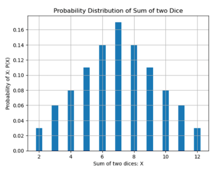
c. Probability Distribution

> A probability distribution tells us the probability of all the possible values of a random variable, and sums up to 1

Types of Probability Distributions

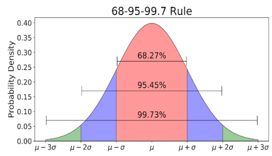
Discrete Probability Distributions

- A discrete probability distribution, model the probabilities of discrete random variables that can have discrete values as outcome, which are counted, not measured
- Described by a formula called Probability Mass Function (pmf)
- The graph of a discrete probability distribution is a bar graph or histogram and the probability is represented by the height of the bar
- Types:
  - Bernoulli Distribution
  - Binomial Distribution
  - Multinomial Distribution
  - Uniform Distribution
  - Poisson Distribution



Continuous Probability Distributions

- A continuous probability distribution, model the probabilities of continuous random variables that can have continuous values as outcome, which are measured, not counted
- Described by a formula called Probability Density Function (pdf)
- The graph of a continuous probability distribution is a curve and the probability is represented by area under the curve
- Types:
  - Normal/Gaussian distribution
  - Log Normal distribution
  - Exponential distribution
  - Student's T distribution
  - Chi-square distribution



Example of Probability Distribution for Discrete Random Variable

Sample Space of rolling two dice

Die 2

Die 1	(a,b)	1	2	3	4	5	6
	1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

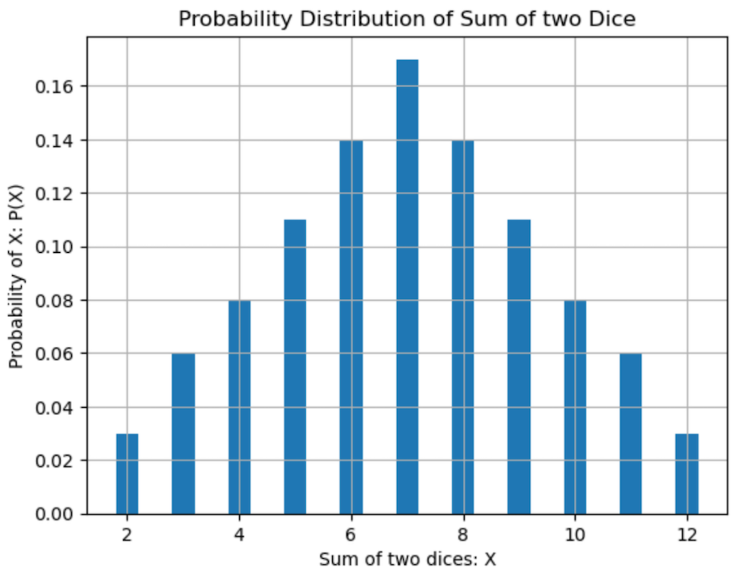
Output: Sum of two dice

Unique o/p={2,3,4,5,6,7,8,9,10,11,12}

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Probability Distribution

X	P(X)
2	0.03
3	0.06
4	0.08
5	0.11
6	0.14
7	0.17
8	0.14
9	0.11
10	0.08
11	0.06
12	0.03



- The probability that the sum is exactly 2: $P(X = 2) = \frac{1}{36} = 0.03$

4. Joint Probability

Joint Probability is the probability of two events occurring simultaneously

$$P(X = x, Y = y)$$

a. Mutually Exclusive vs Not Mutually Exclusive Events

(i) Mutually Exclusive Events:

- Mutually Exclusive events are events that `CANNOT` occur at the same time.
- **Examples:**
  - If a die is rolled the two events getting a `3` and getting a `5` are mutually exclusive.
  - If a card is selected at random from a deck of 52 playing cards, the probability that the card is an `Ace` or a `King` are mutually exclusive.

Addition rule: $$P(A \cup B) \quad = \quad P(A) + P(B)$$

(ii) Not Mutually Exclusive Events:

- Not Mutually Exclusive events are events that `CAN` occur at the same time.
- **Example:**
  - If a card is selected at random from a deck of 52 playing cards, the probability that the card is a `Queen` or a card of `Hearts` are not mutually exclusive.

Addition rule:

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

Note: `Addition rule is used to find the probability of at least one of two events occurring`

b. Independent vs Dependent Events

(i) Independent Events:

- Two events are said to be `independent`, when the occurrence of one event `DOES NOT` affect the probability of occurrence of other event.
- **Examples:**
  - If a coin is tossed and then a die is rolled, the outcome of the coin in no way affects or changes the probability of the outcome of the die.
  - Selecting a card from a deck, replacing it, and then selecting a second card from a deck are two independent events.
  - Drawing a ball from an urn, replacing it, and then drawing a second ball.

Multiplication Rule: $$P(A \cap B) \quad = \quad P(A) \times P(B)$$

(ii) Dependent Events:

- Two events are said to be `dependent`, if they happen one after another, and the occurrence of one event `DOES` affect the probability of occurrence of other event.

- **Examples:**
  - Drawing a ball from an urn, not replacing it, and then drawing a second ball.
  - Drawing two cards without replacement from a deck of cards.

Multiplication Rule:

$$P(A \cap B) \quad = \quad P(A) \times P(B \mid A)$$

Note: `Multiplication rule is used to find the probability of two events occurring in sequence`

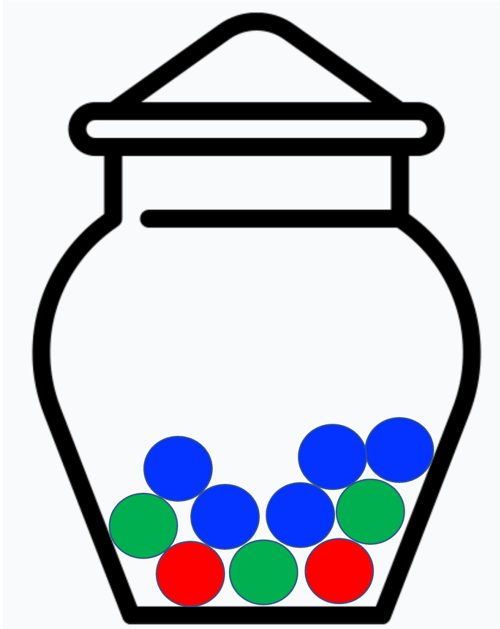
Example 1: An urn contains 2 red balls, 3 green balls, and 5 blue balls.

- Find the `joint probability` if two balls are drawn one after another `with replacement` (independent events):
  - Selecting a blue ball and then a red ball:

    $$P(B \cap R) = P(B) \times P(R) = \frac{5}{10} \times \frac{2}{10} = \frac{1}{10}$$
  - Selecting a green ball and then a blue ball:

    $$P(G \cap B) = P(G) \times P(B) = \frac{3}{10} \times \frac{5}{10} = \frac{3}{20}$$

        - Selecting a blue ball and then another blue ball:



$$P(B \cap B) = P(B) \times P(B) = \frac{5}{10} \times \frac{5}{10} = \frac{1}{4}$$

- Find the `joint probability` if two balls are drawn one after another `without replacement` (dependent events):
    - Selecting a blue ball and then a red ball:
    $$P(B \cap R) = P(B) \times P(R \mid B) = \frac{5}{10} \times \frac{2}{9} = \frac{1}{9}$$
    - Selecting a green ball and then a blue ball:

Example 2 (Dependent Events): Two cards are drawn `without` replacement from a deck of 52 cards. Find the probability that both are queens.

- Event A: First card is Queen
- Event B: Second card is Queen

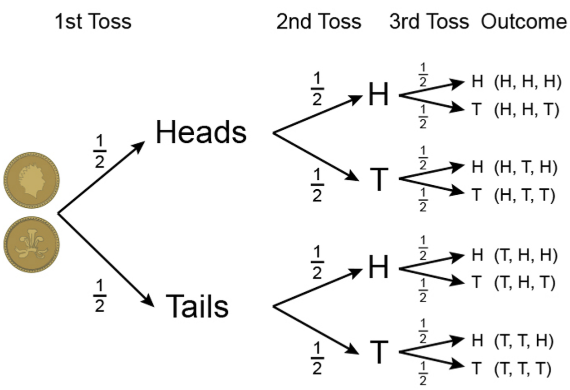
$$P(A \cap B) \quad = \quad P(A) \times P(B \mid A) \quad = \quad \frac{4}{52} \times \frac{3}{51} \quad = \quad \frac{12}{2652} \quad =$$

Example 3 (Dependent Events): A box contains 24 toasters, 3 of which are defective. If two toasters are selected and tested, find the probability that both are defective (assume toasters are `not replaced`).

- Event A: First toaster is defective
- Event B: Second toaster is defective

$$P(A \cap B) \quad = \quad P(A) \times P(B \mid A) \quad = \quad \frac{3}{24} \times \frac{2}{23} \quad = \quad \frac{1}{92} \quad = \quad 0.$$

Example 4 (Independent Events): A coin is tossed three times and a die is rolled twice. Find the probability of getting exactly two heads on the coin and a sum of 5 on the dice.



(a,b)	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

- Event A: Getting exactly two heads on the coin: $P(A) = \frac{3}{8}$

Example 5: Consider a random experiment of rolling two fair six-sided dice. Also consider the random variable is the sum of the two numbers that appeared on the two dice.

Sample Space of rolling two dice

Die 2

	(a,b)	1	2	3	4	5	6
	1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
D	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
i	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
e	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
1	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Output: Sum of two dice

Unique o/p={2,3,4,5,6,7,8,9,10,11,12}

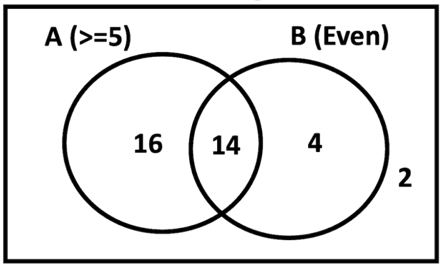
+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

- Event A: Getting a sum of greater than or equal to 5, i.e., {5,6,7,8,9,10,11,12}
- Event B: Getting an even number, i.e., {2,4,6,8,10,12}
- Calculate the probability of the two events occurring in sequence (one after another): $P(A \cap B)$
- Calculate the probability of at least one of the two events occurring: $P(A \cup B)$

Contingency Table

		B		
		Even	**Not Even**	**All**
A	**>=5**	14	16	30
	<5	4	2	6
	All	18	18	36

Venn Diagram



Calculating Probabilities using Venn-Diagram and Contingency Table:

$$P(A) = \frac{30}{36} = 0.8333, \qquad P(B) = \frac{18}{36} = 0.5$$

Calculating Probabilities using Addition and Multiplication Rules:

$$P(A \cup B) \quad = \quad P(A) + P(B) - P(A \cap B) \quad = \quad \frac{30}{36} + \frac{18}{36} - \frac{14}{36} \quad =$$

$$P(A \cap B) \quad = \quad P(A) \times P(B \mid A) \quad = \quad \frac{30}{\ } \times \frac{14}{\ } \quad =$$

		B		
		Even	Not Even	All
A	>=5	0.389	0.444	0.833
	<5	0.111	0.056	0.167
	All	0.5	0.5	1.0

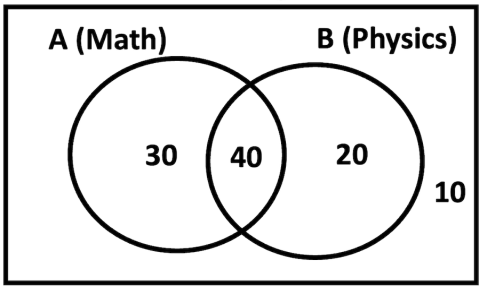
Example 6:

- A class contains 100 students; 70 of them like mathematics, 60 like physics, and 40 like both.
  - Event A: Student like mathematics
  - Event B: Student like physics
- Calculate the probability of the two events occurring in sequence (one after another): $P(A \cap B)$
- Calculate the probability of at least one of the two events occurring: $P(A \cup B)$

Contingency Table

		B		
		physics	not physics	Total
A	math	40	30	70
	not math	20	10	30
	Total	60	40	100

Venn Diagram



- If a student is chosen at random, find the probability that he/she
  - like mathematics $= P(A) = \frac{70}{100}$
  - like physics $= P(B) = \frac{60}{100}$
  - like both mathematics and physics $= P(A \cap B) = \frac{40}{100}$
  - like mathematics or physics $= P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{70}{100} + \frac{60}{100} - \frac{40}{100} = \frac{90}{100}$
  - like mathematics but not physics $= \frac{30}{100}$
  - like physics but not mathematics $= \frac{20}{100}$
  - like neither physics nor mathematics $= \frac{10}{100}$

		B		
		physics	**not physics**	**Total**
A	**math**	0.4	0.3	0.7
	not math	0.2	0.1	0.3

Example 7: Suppose `Netflix` took a survey of their subscribers to determine people favourite movies. Load the following toy dataset, create contingency table using the Panda's `crosstab` method, and calculate different joint and marginal probabilities.

- Let X be a random variable associated with the `movie` and can take three different values
- Let Y be a random variable associated with the `gender` and can take two values

In [1]:
```
import numpy as np
import pandas as pd
```

In [2]:
```
df = pd.read_csv('datasets/sample1.csv')
df
```

Out[2]:

	movie	gender
0	Titanic	female
1	Titanic	male
2	Titanic	female
3	Titanic	male
4	Titanic	female
5	Brave Heart	male
6	Brave Heart	male
7	Brave Heart	female
8	The Terminator	male
9	The Terminator	female
10	The Terminator	male
11	The Terminator	male

- From above dataset, construct contingency table by keeping `movie` along horizontal rows, while `gender` along vertical columns

In [3]:
```
pd.crosstab(index=df['movie'], columns=df['gender'])
```

Out[3]:

gender	female	male
movie		
Brave Heart	1	2
The Terminator	1	3
Titanic	3	2

In [4]:
```
pd.crosstab(index=df['movie'], columns=df['gender'], margins='all')
```

Out[4]:

gender	female	male	All
movie			
Brave Heart	1	2	3
The Terminator	1	3	4
Titanic	3	2	5
All	5	7	12

In [5]:
```
# normalize = {'all', 'index', columns'}
pd.crosstab(index=df['movie'], columns=df['gender'], margins=True, norma
```

Out[5]:

gender	female	male	All
movie			
Brave Heart	0.083333	0.166667	0.250000
The Terminator	0.083333	0.250000	0.333333
Titanic	0.250000	0.166667	0.416667
All	0.416667	0.583333	1.000000

Example 8: Load the Titanic dataset, create contingency table using the Panda's `crosstab` method (for the survived and passenger class columns), and calculate different joint and marginal probabilities.

- Let X be a random variable associated with the `pclass` of a Titanic passenger and can take the values {1,2,3}
- Let Y be a random variable associated with the `survived` status of a Titanic passenger and can take the values {0, 1}

In [6]:
```
import numpy as np
import pandas as pd
df = pd.read_csv('datasets/titanic2.csv')
df
```

Out[6]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_ma
0	0	3	male	22.0	1	0	7.2500	S	Third	man	Tr
1	1	1	female	38.0	1	0	71.2833	C	First	woman	Fal
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	Fal
3	1	1	female	35.0	1	0	53.1000	S	First	woman	Fal
4	0	3	male	35.0	0	0	8.0500	S	Third	man	Tr
...	...	...	...	...	...	...	...	...	...	...	
886	0	2	male	27.0	0	0	13.0000	S	Second	man	Tr
887	1	1	female	19.0	0	0	30.0000	S	First	woman	Fal
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	Fal
889	1	1	male	26.0	0	0	30.0000	C	First	man	Tr
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	Tr

891 rows × 15 columns

In [7]:
```
pd.crosstab(index=df['pclass'], columns=df['survived'], margins=True)
```

Out[7]:

survived	0	1	All
pclass			
1	80	136	216
2	97	87	184
3	372	119	491
All	549	342	891

In [8]:
```
pd.crosstab(index=df['survived'], columns=df['pclass'], margins=True)
```

Out[8]:

pclass	1	2	3	All
survived				
0	80	97	372	549
1	136	87	119	342
All	216	184	491	891

- The cell value of 80 tells us that there are 80 passengers who belong to pclass=1 and do not survived.
- The cell value of 136 tells us that there are 136 passengers who belong to pclass=1 and survived.
- and so on....
- Now this tells us the count, and to calculate the probability we simply has to divide this count with the total number of passenger onboard and i.e., 891 and you get 0.08978. which is the probability of a passenger belonging to class 1 and who do not survived. This is what joint probability is, where you take more than one event and you calculate the probability of happening of both the events

In [9]:
```
pd.crosstab(index=df['survived'], columns=df['pclass'], margins=True, no
```

Out[9]:

pclass survived	1	2	3	All
0	0.089787	0.108866	0.417508	0.616162
1	0.152637	0.097643	0.133558	0.383838
All	0.242424	0.206510	0.551066	1.000000

If a passenger is chosen at random, find the `joint probability` that he/she:

- belongs to `pclass 1 and not survived` : $P(\text{pclass} = 1, \text{survived} = 0) = 0.089787$
- belongs to `pclass 2 and not survived` : $P(\text{pclass} = 2, \text{survived} = 0) = 0.108866$
- belongs to `pclass 3 and not survived` : $P(\text{pclass} = 3, \text{survived} = 0) = 0.417508$
- belongs to `pclass 1 and survived` : $P(\text{pclass} = 1, \text{survived} = 1) \quad = 0.152637$
- belongs to `pclass 2 and survived` : $P(\text{pclass} = 2, \text{survived} = 1) \quad = 0.097643$
- belongs to `pclass 3 and survived` : $P(\text{pclass} = 3, \text{survived} = 1) \quad = 0.133558$

- Each of above six individual numbers (probabilities) is called joint probability, and all the six probabilities together is known as `Joint Probability Distribution` , and note the sum of these six probability values sums upto one.

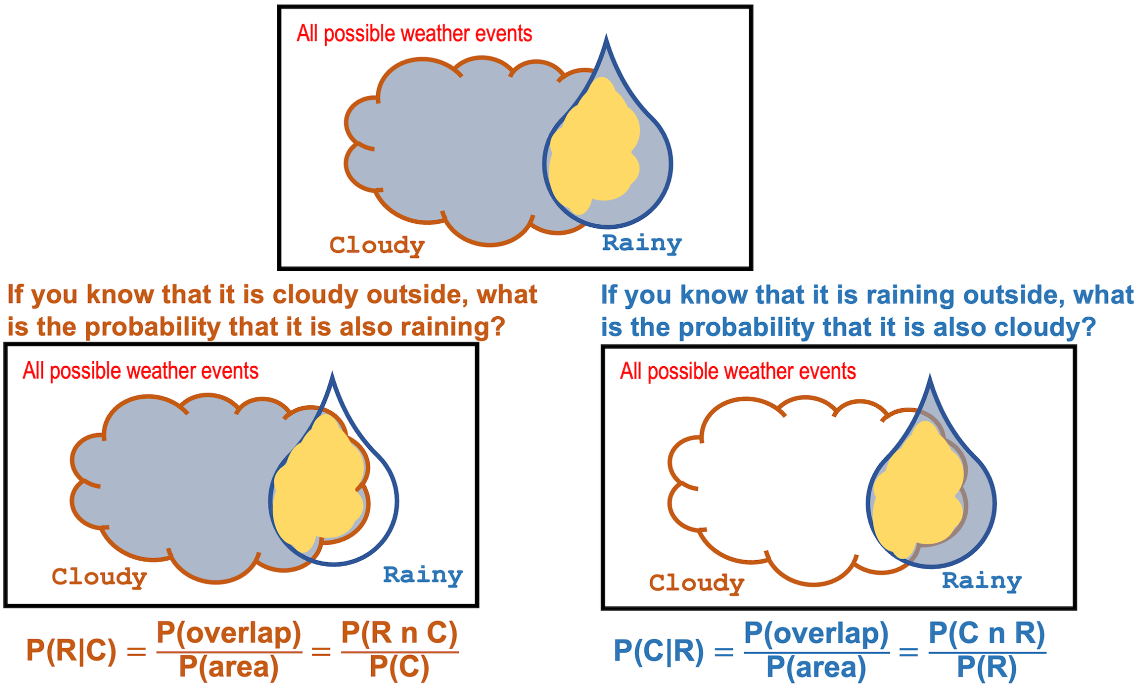
If a passenger is chosen at random, find the `marginal probability` that he/she:

- died, irrespective of pclass he/she is travelling: $P(\text{survived=0}) = 0.616162$
- survived, irrespective of pclass he/she is travelling: $P(\text{survived=1}) = 0.383838$
- travels in pclass 1, irrespective he/she died or survived: $P(\text{pclass=1}) = 0.242424$
- travels in pclass 2, irrespective he/she died or survived: $P(\text{pclass=2}) = 0.206510$
- travels in pclass 3, irrespective he/she died or survived: $P(\text{pclass=3}) = 0.551066$

- 0.616162 and 0.383838 is called the `Marginal Probability Distribution` of random variable `survived` and note the sum of these two probability values is 1.
- 0.242424, 0.206510 and 0.551066 is called the `Marginal Probability Distribution` of random variable `pclass` and the sum of these three

6. Conditional Probability

Conditional Probability is the probability of an event occurring given that another event has already occurred.



> In `joint probability` the two events A and B were occurring together, while in `conditional probability` we normally say that event B has already occurred and we see its impact on the probability of occurring of event A

Example 1: Three unbiased coins are tossed. What is the conditional probability that at least two coins show heads, given that at least one coin shows heads?

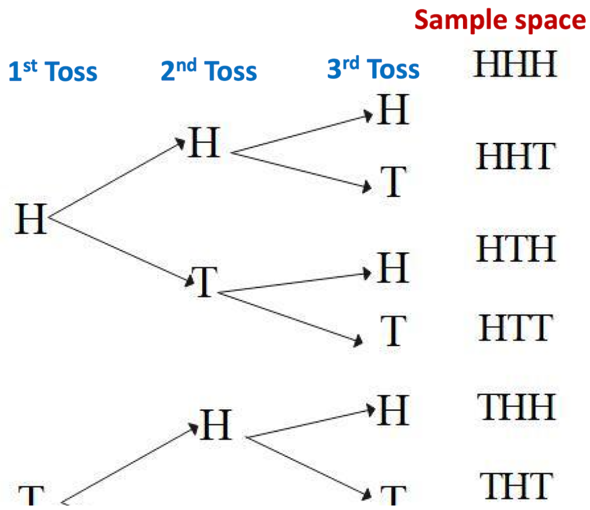
- Event-A: Atleast two heads {HHH, HHT, HTH, THH}
- Event-B: At least one heads (Event has already happened): {HHH, HHT, HTH, THH, HTT, THT, TTH}

Calculate probability of event A given that event B has already occurred:

$$P(A \mid B) = \frac{4}{7} = 0.5714$$

Calculate probability of event B given that event A has already occurred:

$$P(B \mid A) = \frac{4}{4} = 1.0$$



Example 2: Two fair six-sided dice are rolled. What is the conditional probability that the die-1 equals 2, given that the sum of numbers on two dice is less than or equal to 5?

- Event-A: Number on first die is 2: $\{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)\}$
- Event-B: Sum of numbers on two dice is less than or equal to 5: $\{(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1)\}$

Sample Space of rolling two dice

Die 2 (columns), Die 1 (rows)

(a,b)	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Output: Sum of two dice

Unique o/p={2,3,4,5,6,7,8,9,10,11,12}

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Calculate probability of event A given that event B has already occurred:

$$P(A \mid B) = \frac{3}{10} = 0.3$$

Calculate probability of event B given that event A has already occurred:

$$P(B \mid A) = \frac{3}{6} = 0.5$$

Calculate Conditional Probability using Contingency Table:

A \ B	sum<= 5	sum>5	All
D1=2	3	3	6
D1<> 2	7	23	30
All	10	26	36

Example 3: Two fair six-sided dice are rolled. What is the conditional probability that the sum of the numbers rolled is 8, given that the sum is an even number.

- Event-A: Sum is 8 $(2,6),(3,5),(4,4),(5,3),(6,2)$
- Event-B: Sum is an even number (Eighteen possible outcomes)

Sample Space of rolling two dice

Die 2

Die 1 (a,b)	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Output: Sum of two dice

Unique o/p={2,3,4,5,6,7,8,9,10,11,12}

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Calculating $P(A \mid B)$ using the reduced Sample Space:

- Since event B has already occurred, so the new reduced sample space has 18 total outcomes instead of 36

$$P(A \mid B) = \frac{5}{18} = 0.2778$$

Calculating $P(B \mid A)$ using the reduced Sample Space:

- Since event A has already occurred, so the new reduced sample space has 5 total outcomes instead of 36

$$P(A \mid B) = \frac{5}{5} = 1.0$$

Calculate Conditional Probability using Contingency Table:

A \ B	sum is even	sum is odd	All
sum=8	5	0	5
sum<> 8	13	18	31
All	18	18	36

Example 4: Consider the given contingency table containing data of 200 male and female birds in the zoo having brown and blue eyes.

	Brown	Blue	Total
Male Birds:	70	20	90
Female Birds:	100	10	110
Total:	170	30	200

Compute different conditional probabilities of a randomly selected bird being male/female given that it has brown/blue eyes:

- $P(\text{X=male} \mid \text{Y=brown}) = \dfrac{P(\text{X=male} \cap \text{Y=brown})}{P(\text{Y=brown})} = \dfrac{70}{170} \approx 0.41$

- $P(\text{X=female} \mid \text{Y=brown}) = \frac{P(\text{X=female} \cap \text{Y=brown})}{P(\text{Y=brown})} = \frac{100}{170} \approx 0.58$
- $P(\text{X=male} \mid \text{Y=blue}) = \frac{P(\text{X=male} \cap \text{Y=blue})}{P(\text{Y=blue})} = \frac{20}{30} \approx 0.66$
- $P(\text{X=female} \mid \text{Y=blue}) = \frac{P(\text{X=female} \cap \text{Y=blue})}{P(\text{Y=blue})} = \frac{10}{30} \approx 0.33$

Compute different conditional probabilities of a randomly selected bird having brown/blue eyes given that it is male/female:

- $P(\text{Y=brown} \mid \text{X=male}) = \frac{P(\text{Y=brown} \cap \text{X=male})}{P(\text{X=male})} = \frac{70}{90} \approx 0.$
- $P(\text{Y=blue} \mid \text{X=male}) = \frac{P(\text{Y=blue} \cap \text{X=male})}{P(\text{X=male})} = \frac{20}{90} \approx 0.$
- $P(\text{Y=brown} \mid \text{X=female}) = \frac{P(\text{Y=brown} \cap \text{X=female})}{P(\text{X=female})} = \frac{100}{110} \approx 0$
- $P(\text{Y=blue} \mid \text{X=female}) = \frac{P(\text{Y=blue} \cap \text{X=female})}{P(\text{X=female})} = \frac{10}{110} \approx 0$

Example 5: Given the Titanic Dataset (pclass, survived columns), given that a person belong to a specific class, determine the conditional probability of his dieing

`Load the Dataset:`

In [10]:
```python
import numpy as np
import pandas as pd
df = pd.read_csv('datasets/titanic2.csv')
df
```

Out[10]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_ma
0	0	3	male	22.0	1	0	7.2500	S	Third	man	Tru
1	1	1	female	38.0	1	0	71.2833	C	First	woman	Fal
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	Fal
3	1	1	female	35.0	1	0	53.1000	S	First	woman	Fal
4	0	3	male	35.0	0	0	8.0500	S	Third	man	Tru
...	...	...	...	...	...	...	...	...	...	...	
886	0	2	male	27.0	0	0	13.0000	S	Second	man	Tru
887	1	1	female	19.0	0	0	30.0000	S	First	woman	Fal
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	Fal
889	1	1	male	26.0	0	0	30.0000	C	First	man	Tru
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	Tru

891 rows × 15 columns

- Let X be a random variable associated with the `pclass` of a Titanic passenger and can take the values {1,2,3}
- Let Y be a random variable associated with the `survived` status of a Titanic passenger and can take the values {0, 1}

`Conditional Probabilities of survival given the pclass info of passenger:`

In [11]:
```python
pd.crosstab(index=df['survived'], columns=df['pclass'], margins=True)
```

Out[11]:

pclass	1	2	3	All
survived				
0	80	97	372	549
1	136	87	119	342
All	216	184	491	891

- Find the conditional probability of a person `not surviving` given that he/she is travelling in `pclass 1`:

$$P(\text{survived} = 0 \mid \text{pclass} = 1) \quad = \quad \frac{P(\text{survived} = 0 \ \cap \ \text{pclass} = 1)}{P(\text{pclass} = 1)} \quad = \quad \frac{80}{216}$$

- Find the conditional probability of a person `surviving` given that he/she is travelling in `pclass 1`:

$$P(\text{survived} = 1 \mid \text{pclass} = 1) \quad = \quad \frac{P(\text{survived} = 1 \;\cap\; \text{pclass} = 1)}{P(\text{pclass} = 1)} \quad = \quad \frac{136}{216}$$

In [12]:
```
# by setting the normalize='columns', you will get the conditional proba
pd.crosstab(index=df['survived'], columns=df['pclass'], normalize='colum
```

Out[12]:

pclass	1	2	3
survived			
0	0.37037	0.527174	0.757637
1	0.62963	0.472826	0.242363

> The above table shows the conditional probabilities
>
> - 0.37037 is the probability of a person `not surviving` given that he/she belong to pclass 1
> - 0.62963 is the probability of a person `surviving`, given that he/she is traveling in pclass 1.
> - The sum of the two probabilities in that column is 1.

`Conditional Probabilities of pclass given the survival info of passenger:`

In [13]:
```
pd.crosstab(index=df['survived'], columns=df['pclass'], margins=True)
```

Out[13]:

pclass	1	2	3	All
survived				
0	80	97	372	549
1	136	87	119	342
All	216	184	491	891

- Find the conditional probability of a person traveling in `pclass 1` given that he/she `died`:

$$P(\text{pclass} = 1 \mid \text{survived} = 0) \quad = \quad \frac{P(\text{pclass} = 1 \;\cap\; \text{survived} = 0)}{P(\text{survived} = 0)} \quad = \quad \frac{80}{549}$$

- Find the conditional probability of a person traveling in `pclass 2` given that he/she `died`:

$$P(\text{pclass} = 2 \mid \text{survived} = 0) \quad = \quad \frac{P(\text{pclass} = 2 \;\cap\; \text{survived} = 0)}{P(\text{survived} = 0)} \quad = \quad \frac{97}{549}$$

- Find the conditional probability of a person traveling in `pclass 3` given that he/she `died`:

$$P(\text{pclass} = 3 \mid \text{survived} = 0) \quad = \quad \frac{P(\text{pclass} = 3 \quad \cap \quad \text{survived} = 0)}{P(\text{survived} = 0)} \quad = \quad \frac{372}{549}$$

In [14]:

```
# you can set the normalize argument to index instead of columns
pd.crosstab(index=df['survived'], columns=df['pclass'], normalize='index
```

Out[14]:

pclass survived	1	2	3
0	0.145719	0.176685	0.677596
1	0.397661	0.254386	0.347953

The above table shows the conditional probabilities

- 0.145719 is the probability of a person traveling in `pclass 1` given that he/she `died`.
- 0.176685 is the probability of a person traveling in `pclass 2` given that he/she `died`.
- 0.677596 is the probability of a person traveling in `pclass 3` given that he/she `died`.
- The sum of the probabilities in that row is 1.