Probability

Topics Covered:

1. Probability - Meaning and Concepts

2. Rules for Computing Probability

3. Marginal Probability

4. Bayes' Theorem

Meteor showers are rare, but the probability of them occurring can be

In [ ]:

```

```

1. Probability - Meaning and Concepts

Probability – Meaning & Concepts

- **Probability** refers to chance or likelihood of a particular event-taking place.
- An **event** is an outcome of an experiment.
- An **experiment** is a process that is performed to understand and observe possible outcomes.
- Set of all outcomes of an experiment is called the **sample space.**

Example

In a manufacturing unit 3 parts from the assembley are selected. You are observing whether they are Defective or Non-Defective. Determine:

- a) The sample space
- b) The event of getting atleast two defective parts

**Note- You can use "G" to represnt Non-Defectives & "D" for Defectives

Answers:

a)

- Total events are 8
- {[GGG][GDD][DDD][DGG][DDG][DGD][GDG][GGD]}

b)

- Total 4 events

In [ ]:

```

```

Definition of Probability

Probability of an event A is defined as the ratio of two numbers m and n. In symbols

$$P(A) = \frac{m}{n}$$

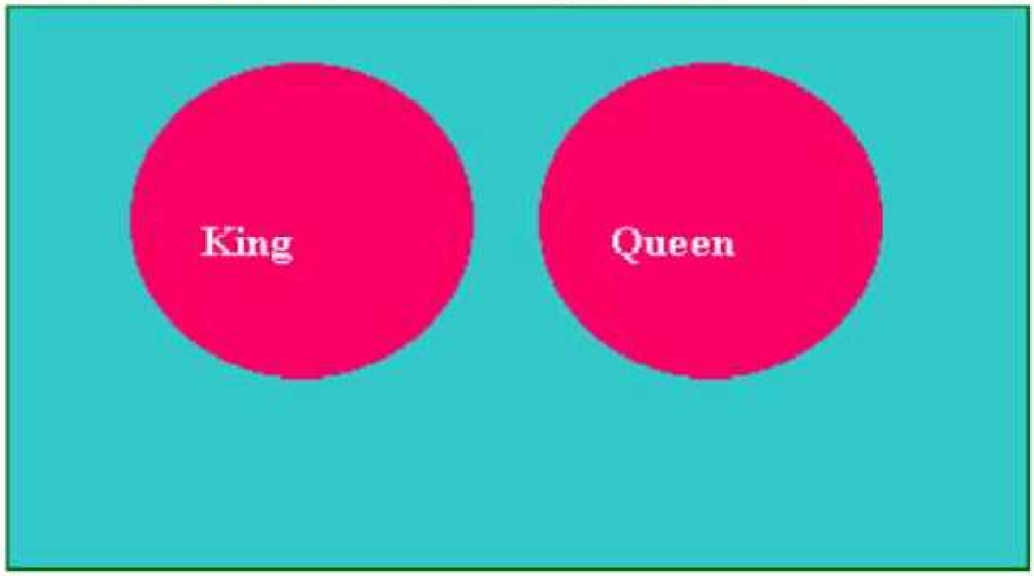
where m= number of ways that are favorable to the occurrence of A and n= the total number of outcomes of the experiment (all possible outcomes)

Please note that P (A) is always > = 0 and always < = 1.
P (A) is a pure number.

In [ ]:

Mutually Exclusive Events

Two events A and B are said to be mutually exclusive if the occurrence of A precludes the occurrence of B. For example, from a well shuffled pack of cards, if you pick up one card at random and would like to know whether it is a King or a Queen. The selected card will be either a King or a Queen. It cannot be both a King and a Queen. If King occurs, Queen will not occur and Queen occurs, King will not occur.



- A and B are mutually exclusive events if they cannot occur at the same time.
- $P(A \cap B)=0$

NOTE- If it is not known whether A and B are mutually exclusive, assume they are not until you can show otherwise.

In [ ]:

Independent Events

- Two events A and B are said to be independent if the occurrence of A is in no way influenced by the occurrence of B. Likewise occurrence of B is in no way influenced by the occurrence of A.

- Two events are independent if one of the following are true:
  - P(A|B)=P(A)
  - P(B|A)=P(B)
  - P(A∩B)=P(A)P(B)
- Two events A and B are independent if the knowledge that one occurred does not affect the chance the other occurs.

NOTE- If it is not known whether A and B are independent or dependent, assume they are dependent until you can show otherwise.

Example

- Consider a standard six-sided die. We'll define two events related to rolling the die once:
- Event A: Rolling an even number (2, 4, 6).
- Event B: Rolling a 1.

Check if these events are mutually exclusive OR Independent

```
In [ ]:
```

Solution

- These events are mutually exclusive because you cannot roll a 1 and an even number at the same time. If you roll a 1, you cannot roll a 2, 4, or 6, and vice versa.
- For these events to be independent:
  - P(A∩B)=P(A)×P(B)
  - First, let's calculate the individual probabilities:
  - P(A)= 3/6
  - $P(B)$ = 1/6
  - P(A∩B)=0 (since they are mutually exclusive)
  - P(A)×P(B)= 1/12
  - But, P(A∩B)=0
- Hence the events are not independent.

Problem1

Flip two fair coins. Find the probabilities of the events.

- Let F = the event of getting at most one tail (zero or one tail).
- Let G = the event of getting two faces that are the same.
- Let H = the event of getting a head on the first flip followed by a head or tail on the second flip.
- Are F and G mutually exclusive?
- Let J = the event of getting all tails. Are J and H mutually exclusive?

```
In [ ]:
```

Solution

- The sample space is {HH, HT, TH, TT}
- Zero (0) or one (1) tails occur when the outcomes HH, TH, HT show up. P(F) = 3/4
- Two faces are the same if HH or TT show up. P(G) = 2/4
- A head on the first flip followed by a head or tail on the second flip occurs when HH or HT show up. P(H) = 24

- F and G share HH so $P(F \cap G)$ is not equal to zero (0). F and G are not mutually exclusive.
- Getting all tails occurs when tails shows up on both coins (TT). H's outcomes are HH and HT.
  - J and H have nothing in common so $P(J \cap H) = 0$. J and H are mutually exclusive.

In [ ]:

Problem2

- Let event C = taking an English class. Let event D = taking a speech class.
- Suppose
  - $P(C)=0.75$
  - $P(D)=0.3$
  - $P(C|D)=0.75$
  - and $P(C \cap D)=0.225$
- Justify your answers to the following questions numerically.
  - a) Are C and D independent?
  - b) Are C and D mutually exclusive?

In [ ]:

Solution

- Yes, because $P(C|D)=P(C)$
- No, because $P(C \cap D)$ is not equal to zero.

In [ ]:

Steps to Analyze Events

- Check for Mutual Exclusivity:
  - Determine if the events can occur simultaneously.
  - Two events A and B are mutually exclusive if $P(A \cap B)=0$.

- Check for Independence:
  - Determine if the occurrence of one event affects the probability of the other event.
  - Two events A and B are independent if P(A∩B)=P(A)×P(B).

		Mutually Exclusive	
		YES	NO
Independent	YES	* Event A: Rolling a 7 * Event B: Rolling an odd number (1, 3, 5)	* Event E: Rolling an even number (2, 4, 6) * Event F: Rolling a number less than 5 (1, 2, 3, 4).
	NO	no valid example	* Event G: Rolling an even number (2, 4, 6) * Event H: Rolling a number less than 5 (1, 2, 3, 4).

Events which are Mutually exclusive and Independent as well

- Event A: Rolling a 7.
- Event B: Rolling an odd number (1, 3, 5).
- Note: Rolling a 7 is an impossible event with a six-sided die, so P(A)=0.

Checking Conditions

- Mutual Exclusivity:
  - Since rolling a 7 is impossible and rolling an odd number is possible, these events are mutually exclusive.
  - Therefore, P(A∩B)=0.
- Independence:
  - P(A)=0
  - P(B)= 3/6
- check if the independence condition holds:
  - P(A∩B)=P(A)×P(B)
  - 0=0
- Since both sides of the equation are equal, the independence condition is satisfied.

In [ ]:

2. Rules for Computing Probability

- There are several rules for computing probabilities:
- Addition Rule: The probability of either of two mutually exclusive events occurring is the sum of their individual probabilities.
- Multiplication Rule: The probability of two independent events both occurring is the product of their individual probabilities.

1) Addition Rule -Mutually Exclusive Events

$$P(A \bigcup B) = P(A) + P(B)$$

This rule says that the probability of the union of A and B is determined by adding the probability of the events A and B.

Here the symbol $A \bigcup B$ is called A union B meaning A occurs, or B occurs or both A and B simultaneously occur. When A and B are mutually exclusive, A and B cannot simultaneously occur.

In [ ]:

2) Addition Rule –Events are not Mutually Exclusive

$$P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$$

This rule says that the probability of the union of A and B is determined by adding the probability of the events A and B and then subtracting the probability of the intersection of the events A and B.

The symbol $A \bigcap B$ is called A intersection B meaning

both A and B simultaneously occur.

In [ ]:

```

```

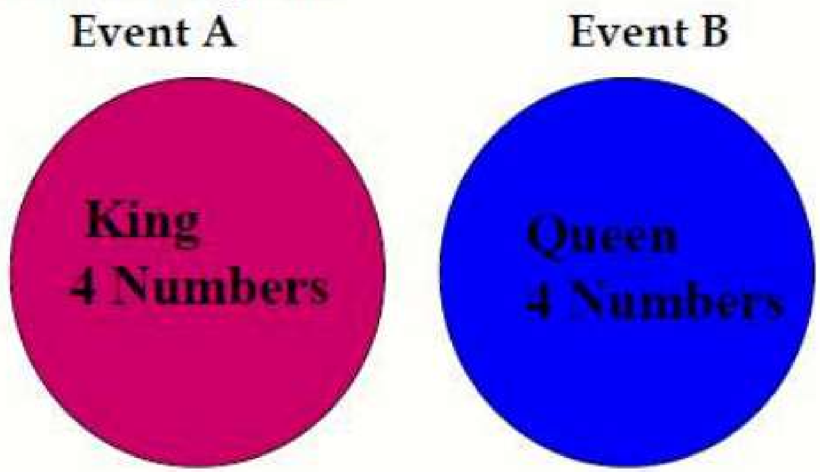
Example for Addition Rules

- From a pack of well-shuffled cards, a card is picked up at random.

1) What is the probability that the selected card is a King or a Queen?
2) What is the probability that the selected card is a King or a Diamond?

```
In [ ]:
```

Look at the Diagram:



Let A = getting a King
Let B = getting a Queen

There are 4 kings and there are 4 Queens. The events are clearly mutually exclusive. Applying the formula $P(A \bigcup B) = P(A) + P(B)$
$= 4/52 + 4/52 = 8/52 = 2/13$

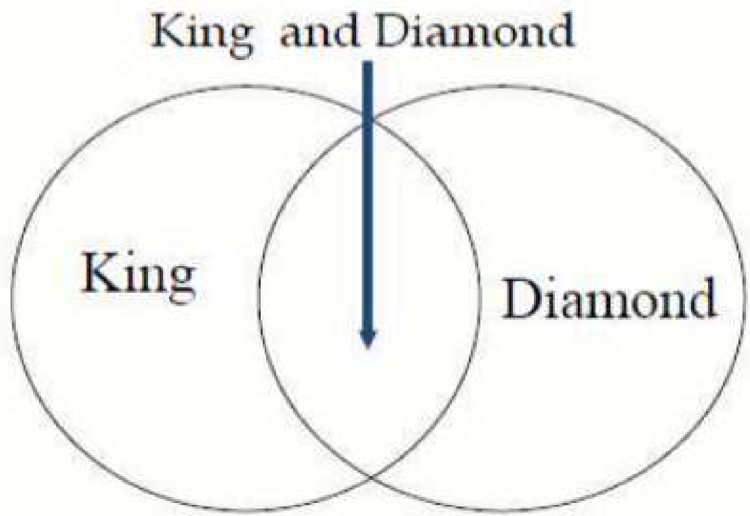
In [ ]:

Look at the Diagram:

There are totally 52 cards in a pack out of which 4 are Kings and 13 are Diamonds. Let A= getting a King and B= getting a Diamond. The two events here are not mutually exclusive because you can have a card, which is both a King and a Diamond called King Diamond.

$$P(K \bigcup D) = P(K) + P(D) - P(K \bigcap D)$$

= 4/52+13/52-1/52= 16/52=4/13



In [ ]:

Multiplication Rule

Independent Events

$$P(A \bigcap B) = P(A).P(B)$$

This rule says when the two events A and B are independent, the probability of the simultaneous occurrence of A and B (also known as probability of intersection of A and B) equals the product of the probability of A and the probability of B. Of course this rule can be extended to more than two events.

```
In [ ]:
```

Independent Events-Example

Example:

The probability that you will get an A grade in Quantitative Methods is 0.7. The probability that you will get an A grade in Marketing is 0.5. Assuming these two courses are independent, compute the probability that you will get an A grade in both these subjects.

Solution:

Let A = getting A grade in Quantitative Methods

Let B =getting A grade in Marketing

It is given that A and B are independent.

$$P(A \cap B) = P(A).P(B) = 0.7.0.5 = 0.35.$$

In [ ]:

Events are not independent

$$P(A \cap B) = P(A).P(B/A)$$

This rule says that the probability of the intersection of the events A and B equals the product of the probability of A and the probability of B given that A has happened or known to you. This is symbolized in the second term of the above expression as P (B/A). P(B/A) is called the conditional probability of B given the fact that A has happened.

We can also write $P(A \cap B) = P(B).P(A/B)$ if B has already happened.

From a pack of cards, 2 cards are drawn in succession one after the other. After every draw, the selected card is not replaced. What is the probability that in both the draws you will get Spades?

Solution:
Let A = getting Spade in the first draw
Let B = getting spade in the second draw.
The cards are not replaced.

This situation requires the use of conditional probability.

P(A) = 13/52 (There are 13 Spades and 52 cards in a pack)
P(B/A)=12/51(There are 12 Spades and 51 cards because the first card selected is not replaced after the first draw)

$$P(A \cap B) = P(A).P(B/A) = (13/52).(12/51)=156/2652= 1/17.$$

In [ ]:

```

```

3. Marginal Probability

- Marginal Probability refers to the probability of an event occurring regardless of the outcome of another event. It is derived by summing the joint probabilities over all possible outcomes of the other variable.
- Example:

- If you have a joint probability table for two variables, the marginal probability of one variable can be obtained by summing the probabilities across the rows or columns.

Marginal Probability

- Contingency table consists of rows and columns of two attributes at different levels with frequencies or numbers in each of the cells. It is a matrix of frequencies assigned to rows and columns.
- The term marginal is used to indicate that the probabilities are calculated using a contingency table (also called joint probability table).

Marginal Probability - Example

A survey involving 200 families was conducted. Information regarding family income per year and whether the family buys a car are given in the following table.

Family	Income below $ 20K	Income of >= $ 20K	Total
Buyer of Car	38	42	80
Non–Buyer	82	38	120
Total	120	80	200

a) What is the probability that a randomly selected family is a buyer of the car?

b) What is the probability that a randomly selected family is both a buyer of car and belonging to income of $ 20000 and above?

c) A family selected at random is found to be belonging to income of $ 20K and above. What is the probability that this family is buyer of car?

In [ ]:

Solution

a) What is the probability that a randomly selected family is a buyer of the Car?

- 80/200 =0.40.

b) What is the probability that a randomly selected family is both a buyer of car and belonging to income of $20 k and above?

- 42/200 =0.21.

c) A family selected at random is found to be belonging to income of $ 20K and above. What is the probability that this family is buyer of car?

- 42/80 =0.525. Note this is a case of conditional probability of buyer given income is $20 k and above.

Problems- Marginal Probability

In [3]:
```
import seaborn as sns
df=sns.load_dataset('titanic')
```

Questions

- What is the probability that a randomly selected passenger survived given that they were in first class (Pclass = 1)?
- What is the probability that a passenger was in third class (Pclass = 3) and did not survive?
- What is the probability that a passenger did not survive given that they were in second class (Pclass = 2)?

In [7]:
```
df.head(1)
```

Out[7]:

	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	7.25	S	Third	man	True	NaN	Southampton	no	False

In [ ]:
```
P(Survived=1|pclass = 1)
```

In [4]:
```
import pandas as pd
pd.crosstab(df['pclass'],df['survived'],margins=True)
```

Out[4]:

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

In [6]:
```
round(136/216*100,3)
```

Out[6]:
```
62.963
```

- Solutions:
- a) 136/216
- b) 372/491
- c) 97/184

In [16]:

```python
import pandas as pd
pd.crosstab(df['pclass'],[df['sex'],df['survived']],margins=True)
```

Out[16]:

sex	female		male		All
survived	0	1	0	1	
pclass					
1	3	91	77	45	216
2	6	70	91	17	184
3	72	72	300	47	491
All	81	233	468	109	891

In [18]:

```python
pd.crosstab(index=[df['survived'], df['pclass'], df['sex']],
            columns=df['embark_town'], margins=True)
```

Out[18]:

		embark_town	Cherbourg	Queenstown	Southampton	All
survived	pclass	sex				
0	1	female	1	0	2	3
		male	25	1	51	77
	2	female	0	0	6	6
		male	8	1	82	91
	3	female	8	9	55	72
		male	33	36	231	300
1	1	female	42	1	46	89
		male	17	0	28	45
	2	female	7	2	61	70
		male	2	0	15	17
	3	female	15	24	33	72
		male	10	3	34	47
All			168	77	644	889

In [ ]:

4. Bayes' Theorem

Bayes' Theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event. It is given by:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

In [ ]:

- Bayes' Theorem is used to revise previously calculated probabilities based on new information.
- Developed by Thomas Bayes in the 18th Century.
- It is an extension of conditional probability.

Conditional Probability

Formula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

where:

- $P(A \cap B)$ is the probability that both events A and B occur.
- $P(B)$ is the probability that event B occurs.

Generalized Baye's Theorem;

$$\boxed{P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_k)P(B_k)}}$$

- where:

  B_i = i^{th} event of k mutually exclusive and collectively exhaustive events

  A = new event that might impact $P(B_i)$

P(A|B)= P(B|A)

In [ ]:

```

```

In [ ]:

```

```

Bayesian Spam Filtering

One clever application of Bayes' Theorem is in spam filtering. We have

- Event A: The message is spam.
- Test X: The message contains certain words (X)

Plugged into a more readable formula (from Wikipedia):

$$\Pr(\text{spam}|\text{words}) = \frac{\Pr(\text{words}|\text{spam})\Pr(\text{spam})}{\Pr(\text{words})}$$

In [ ]:

```

```

Problem1

If 1% of people have a certain disease and a test correctly identifies 99% of positive cases but also gives 5% false positives, Bayes' Theorem can be used to find the probability that a person actually has the disease given a positive test result.Using Bayes' Theorem, find the probability of having a disease given a positive test result with the aforementioned conditions.

In [ ]:
```
P(Disease)= 0.01
P(Positive_result|Disease)=0.99
P(Positive_result|Not_Disease)=0.05
P(Not_Disease)= 1-0.01
```

In [ ]:
```
P(Positive_result|Disease)*P(Disease) + P(Positive_result|Not_Disease)*P(Not_Disease)
```

P(Disease|Positive_result)= P(Positive_result|Disease)* P(Disease)/(Evidences)

Evidences= [P(Positive_result|Disease)* P(Disease)] + []

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In [ ]:
```
```

Problem2

If the probability of rain is 30%, and the probability of carrying an umbrella given rain is 80%, what is the probability of rain given that a person is carrying an umbrella?

In [ ]:
```
```

In [ ]:

```python

```

In [ ]:

```python

```

In [ ]:

```python

```

In [2]:

```python
# Problem 1: Probability of having the disease given a positive test result
P_disease = 0.01
P_positive_given_disease = 0.99
P_positive_given_no_disease = 0.05
P_no_disease = 1 - P_disease

P_positive = (P_positive_given_disease * P_disease) + (P_positive_given_no_disease * P_no_disease)
P_disease_given_positive = (P_positive_given_disease * P_disease) / P_positive
print("Probability of having the disease given a positive test result:", round(P_disease_given_positive,3)
```

```
Probability of having the disease given a positive test result: 0.167
```

In [16]:

```python
# Problem 2: Probability of rain given that a person is carrying an umbrella
P_rain = 0.3
P_umbrella_given_rain = 0.8
P_umbrella_given_no_rain = 0.2
P_no_rain = 1 - P_rain

P_umbrella = (P_umbrella_given_rain * P_rain) + (P_umbrella_given_no_rain * P_no_rain)
P_rain_given_umbrella = (P_umbrella_given_rain * P_rain) / P_umbrella
print("Probability of rain given that a person is carrying an umbrella:", P_rain_given_umbrella)
```

```
Probability of rain given that a person is carrying an umbrella: 0.631578947368421
```

****