Probability

> Probability is the branch of mathematics concerning numerical descriptions of how likely an event is to occur, or how likely it is that a proposition is true. The probability of an event is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility of the event and 1 indicates certainty. The higher the probability of an event, the more likely it is that the event will occur. - Wikipedia

Probability for an event A happening within a set of possible results S is given by:

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

Q. What is the probability of getting 3 when a fair, six-sided die is rolled?

Solution:

In [1]:
```python
# Let,
# A : Getting 3 when a fair, six-sided die is rolled
# S : Sample space generated when a fair, six-sided die is rolled

A = {3}
S = {1, 2, 3, 4, 5, 6}

# Probability of event A (represented here as P_A) is given by:
# P(A) = No. of elements in set A / No. of elements in set S
# i.e.
# P(A) = n(A) / n(S)

P_A = len(A) / len(S)
print(f"The probability of getting 3 on a single roll of a six sided die is {P_A} or {round(P_A, 3) * 100}%")
```

```
The probability of getting 3 on a single roll of a six sided die is 0.16666666666666666 or 16.7%
```

Q. Calculate the probability of atleast getting one head when a coin is tossed thrice.

Solution:

In [2]:
```python
# Let,
# A : Getting atleast one head when a coin is tossed thrice
# S : Sample space generated when a fair coin is tossed thrice

S = {'HHH', 'HHT', 'HTH', 'THH', 'HTT', 'THT', 'TTH', 'TTT'}
A = {'HHH', 'HHT', 'HTH', 'THH', 'HTT', 'THT', 'TTH'}

# Probability of event A (represented here as P_A) is given by:
# P(A) = No. of elements in set A / No. of elements in set S
# i.e.
# P(A) = n(A) / n(S)

P_A = len(A) / len(S)
print(f"The probability of getting atleast one head when a fair coin is tossed thrice is {P_A} or {round(P_A, 3)
```

```
The probability of getting atleast one head when a fair coin is tossed thrice is 0.875 or 87.5%
```

Q. A glass jar contains 5 red, 3 blue and and 2 green jelly beans. If a jelly bean is chosen at random from the jar, what is the probability that it is not blue?

Solution::

In [3]:
```python
# Let,
# A : Not getting a blue jellybean  when a jellybean is chosen from the jar
# S : Sample space generated when a jellybean is chosen from the jar
# and R1, R2, R3, R4, R5, B1, B2, B3, G1 and G2 represent the five red, 3 blue and 2 green jellybeans
# respectively.

A = {'R1', 'R2', 'R3', 'R4', 'R5', 'G1', 'G2'}
S = {'R1', 'R2', 'R3', 'R4', 'R5', 'B1', 'B2', 'B3', 'G1', 'G2'}

# Probability of event A (represented here as P_A) is given by:
# P(A) = No. of elements in set A / No. of elements in set S
# i.e.
# P(A) = n(A) / n(S)
```

```
P_A = len(A) / len(S)
print(f"The probability of not getting a blue jellybean from the jar is {P_A} or {round(P_A, 3) * 100}%")
```

```
The probability of not getting a blue jellybean from the jar is 0.7 or 70.0%
```

Independent and Dependent events

Independence is a fundamental notion in probability theory, as in statistics and the theory of stochastic processes.

Two events are independent, statistically independent, or stochastically independent if the occurrence of one does not affect the probability of occurrence of the other (equivalently, does not affect the odds). Similarly, two random variables are independent if the realization of one does not affect the probability distribution of the other. - Wikipedia)

Q. If the probability that a person A will be alive after 20 years is 0.7 and the probability that person B will be alive after 20 years is 0.5, what is the probability that they will both be alive after 20 years?

Solution:

In [4]:
```
# Let,
# A : Person A is alive after 20 yrs
# B : Person B is alive after 20 yrs

# Given:
# P(A) = 0.7
# P(B) = 0.5
P_A = 0.7
P_B = 0.5

# As the lives of A and B do not depend on each other, we conclude that events A and B are independent
# Required probability is P(A n B) = P(A) * P(B)

P_A_and_B = P_A * P_B
print(f"Probability that both A and B will be alive after 20 years is {P_A_and_B} or {round(P_A_and_B, 3) * 100}%
```

```
Probability that both A and B will be alive after 20 years is 0.35 or 35.0%
```

Note:

1. If we want the probability of event A **and** event B then,

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

1. If we want the probability of event A **or** event B then,

$$P(A \text{ or } B) = P(A) + P(B)$$

Q. A fair die is tossed twice. Find the probability of getting a 4 or 5 on first toss and a 1, 2 or 3 in the second toss.

Solution:

In [5]:
```
# Let,
# A : Getting a 4 or 5 on a roll
# B : Getting a 1, 2 or 3 on a roll
# S : Sample space generated when a die is rolled once.

A = {4, 5}
B = {1, 2, 3}
S = {1, 2, 3, 4, 5, 6}

# Probability of event A (represented here as P_A) is given by:
# P(A) = No. of elements in set A / No. of elements in set S
# i.e.
# P(A) = n(A) / n(S)
P_A = len(A) / len(S)
P_B = len(A) / len(S)

# Required probability is P(A and B)
P_A_and_B = P_A * P_B
print(f"Probability that 4 or 5 will appear on first roll and 1, 2 or 3 will appear on second roll is ", end = "
print(f"{P_A_and_B} or {round(P_A_and_B, 3) * 100}%")
```

```
Probability that 4 or 5 will appear on first roll and 1, 2 or 3 will appear on second roll is  0.1111111111111111
or 11.1%
```

Q. A bag contains 5 white marbles, 3 black marbles and 2 green marbles. In each draw, a marble is drawn from the bag and not replaced. In three draws, find the probability of obtaining white, black and green in that order.

Solution:

```python
# Let,
# A : White marble is drawn
# B : Black marble is drawn after event A
# C : Green marble is drawn after event B
# S : Total marbles after each event (currently 5 + 3 + 2 i.e. 10)

n_A = 5
n_B = 3
n_C = 2
n_S = 10

# Probability of event A (represented here as P_A) is given by:
# P(A) = No. of elements in set A / No. of elements in set S
# i.e.
# P(A) = n(A) / n(S)

P_A = n_A / n_S

# As the drawn marble is not replaced, the number of marbles in the bag reduces by 1
n_S -= 1
P_B = n_B / n_S

# As the drawn marble is not replaced, the number of marbles in the bag reduces by 1
n_S -= 1
P_C = n_C / n_S

# Required probability is P(A and B and C) = P(A) * P(B) * P(C)
P_A_and_B_and_C = P_A * P_B * P_C
print(f"Probability that the marbles will come out in the order white, black and green is ", end = " ")
print(f"{P_A_and_B_and_C} or {round(P_A_and_B_and_C, 3) * 100}%")
```

```
Probability that the marbles will come out in the order white, black and green is  0.041666666666666664 or 4.2%
```

Q. Find the probability of drawing a heart or a club from a shuffled deck of cards.

Solution:

```python
# Let,
# A : Drawing a heart card from the deck
# B : Drawing a club card from the deck
# S : Sample space generated from a deck of 52 cards

# We know that there are exacty 13 heart and 13 club cards in a deck of 52 cards
n_A = 13
n_B = 13
n_S = 52

# Probability of event A (represented here as P_A) is given by:
# P(A) = No. of elements in set A / No. of elements in set S
# i.e.
# P(A) = n(A) / n(S)
P_A = n_A / n_S
P_B = n_B / n_S

# Required probability is P(A or B) = P(A) + P(B)
P_A_or_B = P_A + P_B
print(f"Probability of getting a heart or a club card is {P_A_or_B} or {round(P_A_or_B, 3) * 100}%")
```

```
Probability of getting a heart or a club card is 0.5 or 50.0%
```

Q. Find the probability of drawing an ace, a king or a queen from a deck of cards

Solution:

```python
# Let,
# A : Drawing an ace from the deck
# B : Drawing a king from the deck
# C : Drawing a queen from the deck
# S : Sample space generated from a deck of 52 cards

# We know that there are exacty 4 aces, 4 kings and 4 queen cards in a deck of 52 cards
n_A = 4
```

```
n_B = 4
n_C = 4
n_S = 52

# Probability of event A (represented here as P_A) is given by:
# P(A) = No. of elements in set A / No. of elements in set S
# i.e.
# P(A) = n(A) / n(S)
P_A = n_A / n_S
P_B = n_B / n_S
P_C = n_C / n_S

# Required probability is P(A or B or C) = P(A) + P(B) + P(C)
P_A_or_B_or_C = P_A + P_B + P_C
print(f"Probability of getting an ace, king or queen card is {P_A_or_B_or_C} or {round(P_A_or_B_or_C, 3) * 100}%'
```

```
Probability of getting an ace, king or queen card is 0.23076923076923078 or 23.1%
```

If events A and B are not mutually exclusive then:

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

Q. Find the probability of drawing an heart or an ace from a deck of cards.

Solution:

In [9]:
```
# Let,
# A : Drawing a heart card from the deck
# B : Drawing a ace card from the deck
# S : Sample space generated from a deck of 52 cards

# We know that there are exacty 13 hearts and 4 aces in a deck of 52 cards with one common card (i.e. the ace of
n_A = 13
n_B = 4
n_A_n_B = 1
n_S = 52

# Probability of event A (represented here as P_A) is given by:
# P(A) = No. of elements in set A / No. of elements in set S
# i.e.
# P(A) = n(A) / n(S)
P_A = n_A / n_S
P_B = n_B / n_S
P_A_n_B = n_A_n_B / n_S

# Required probability is P(A U B) = P(A) + P(B) - P(A n B)
P_A_U_B = P_A + P_B - P_A_n_B
print(f"Probability of getting a heart or ace card is {P_A_U_B} or {round(P_A_U_B, 3) * 100}%")
```

```
Probability of getting a heart or ace card is 0.3076923076923077 or 30.8%
```

Complementary Events

> In probability theory, the complement of any event A is the event (not A), i.e. the event that A does not occur.The event A and its complement (not A) are mutually exclusive and exhaustive. Generally, there is only one event B such that A and B are both mutually exclusive and exhaustive; that event is the complement of A. The complement of an event A is usually denoted as A', A^C. - Wikipedia

That is, for an event A the complement A^C (or A') is:

$$P(A^C) = 1 - P(A)$$

Q. What is the probability of not getting 5 when a fair die is thrown?

Solution:

In [10]:
```
# Let,
# A : Getting 5 when a fair, six-sided die is rolled
# S : Sample space generated when a fair, six-sided die is rolled

A = {5}
S = {1, 2, 3, 4, 5, 6}

# Probability of event A (represented here as P_A) is given by:
```

```
# P(A) = No. of elements in set A / No. of elements in set S
# i.e.
# P(A) = n(A) / n(S)

P_A = len(A) / len(S)

# Required probability is P(A') where A' is the complement of event A (represented in code as P_A_c)
P_A_c = 1 - P_A
print(f"The probability of not getting 5 on a single roll of a six sided die is {P_A_c} or {round(P_A_c, 3) * 100
```

```
The probability of not getting 5 on a single roll of a six sided die is 0.8333333333333334 or 83.3%
```

Conditional Probability

> In probability theory, conditional probability is a measure of the probability of an event occurring, given that another event (by assumption, presumption, assertion or evidence) has already occurred. If the event of interest is A and the event B is known or assumed to have occurred, "the conditional probability of A given B", or "the probability of A under the condition B", is usually written as P(A|B) or occasionally P$_B$(A). This can also be understood as the fraction of probability B that intersects with A. - Wikipedia

The conditional probability of A given B is:

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

Q. Determine the probability of a student getting 80% or more marks given that he/she has been absent for more than 10 classes. Use the student-mat.csv file for the data. (Consider subject G3)

Prerequisites:

1. Load the CSV file and convert it into a dataframe. Check whether the dataframe is created by viewing few rows.

In [2]:
```python
import pandas as pd
import numpy as np

data_frame = pd.read_csv(r'C:\Users\acer\Desktop\Sem 1\data science\DataSet\student-mat.csv')
data_frame.head(3)
```

Out[2]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	6	5
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	4	5
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	10	7

3 rows × 33 columns

1. We are concerned with the columns, absences and G3 (final grade from 0 to 20). Let us create a couple new boolean columns based on these columns to make our lives easier.

   i) Add a boolean column called grade_A noting if a student has got more than 80% marks or not.

   ii) Add a boolean column called high_absentee noting if a student has missed 10 or more classes.

In [12]:
```python
data_frame['grade_A'] = np.where(data_frame['G3'] * 5 >= 80, 1, 0)
```

In [13]:
```python
data_frame['high_absentee'] = np.where(data_frame['absences']  >= 10, 1, 0)
```

1. Add an extra column to make building the pivot table easier.

In [14]:
```python
data_frame['count'] = 1
```

1. Drop all columns we don't care about

In [15]:
```python
data_frame = data_frame[['grade_A', 'high_absentee', 'count']]
```

```
data_frame.head()
```

	grade_A	high_absentee	count
0	0	0	1
1	0	0	1
2	0	1	1
3	0	0	1
4	0	0	1

1. Generate the pivot table.

In [16]:
```
pivot_table = pd.pivot_table(data_frame, values = 'count', index = ['grade_A'], columns = ['high_absentee'], aggf
print(pivot_table)
```

```
high_absentee    0    1
grade_A
0              277   78
1               35    5
```

Solution:

In [17]:
```
# Let,
# A : Event that the student gets atleast than 80% marks
# B : Event that the student is absent for 10 or more classes.
# S : Total students

# Given,
n_A = 35 + 5
n_B = 78 + 5
n_A_n_B = 5
n_S = 277 + 78 + 35 + 5

# Using marginal probability,
P_A = n_A / n_S
P_B = n_B / n_S
P_A_n_B = n_A_n_B / n_S

print(f"Proabability of a student getting atleast than 80% marks is {P_A} or {round(P_A, 3) * 100}%")
print(f"Proabability of a student being absent for atleast 10 classes is {P_B} or {round(P_B, 3) * 100}%")
print(f"Proabability of a student getting atleast than 80% marks and being absent for atleast 10 classes is", end
print(f"{P_A} or {round(P_A, 4) * 100}%", end = "\n")

# Required probability is P(A | B) = P(A n B) / P(B)
P_A_given_B = P_A_n_B / P_B
print(f"Probability of a student getting atleast than 80% marks given that he/she was absent for", end = " ")
print(f"atleast 10 classes is {P_A_given_B} or {round(P_A_given_B, 4) * 100}%")
```

```
Proabability of a student getting atleast than 80% marks is 0.10126582278481013 or 10.100000000000001%
Proabability of a student being absent for atleast 10 classes is 0.21012658227848102 or 21.0%
Proabability of a student getting atleast than 80% marks and being absent for atleast 10 classes is 0.10126582278
481013 or 10.13%
Proabability of a student getting atleast than 80% marks given that he/she was absent for atleast 10 classes is 0
.060240963855421686 or 6.02%
```

In [ ]:

Processing math: 100%