

What is Probability?

Probability is a branch of mathematics that measures the likelihood of an event happening. It uses numbers between 0 and 1 to represent the chance of an event occurring, where 0 means impossible and 1 means certain. Probability helps predict outcomes and quantify uncertainty in fields like statistics and economics. It relies on concepts like sample space, events, and different types of probabilities. Probability is expressed as fractions, decimals, or percentages and is used for risk assessment, decision-making, and drawing conclusions from data.



In simplest terms, probability is a measure of the likelihood that a particular event will occur. It is a fundamental concept in statistics and is used to make predictions and informed decisions in a wide range of disciplines, including science, engineering, medicine, economics, and social sciences.

Probability is usually expressed as a number between 0 and 1, inclusive:

- A probability of 0 means that an event will not happen.
- A probability of 1 means that an event will certainly happen.
- A probability of 0.5 means that an event will happen half the time (or that it is as likely to happen as not to happen).

Terminology

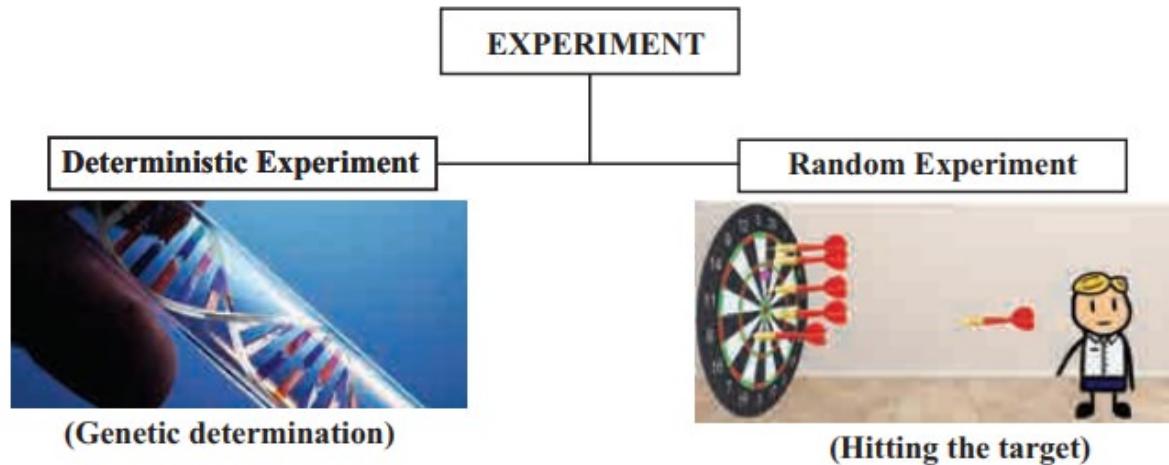
1. Random Experiment

An experiment is called random experiment if it satisfies the following two conditions:

- (i) It has more than one possible outcome.
- (ii) It is not possible to predict the outcome in advance

Random experiments can take various forms depending on the context. Examples of random experiments include flipping a coin, rolling a dice, drawing cards from a deck, conducting scientific experiments with uncertain outcomes, or observing the outcome of a sports game.

The study of random experiments allows us to quantify and analyze uncertainty, make predictions about the likelihood of specific outcomes, and reason about events and their probabilities. Probability theory provides a framework for understanding and analyzing random experiments, enabling us to make informed decisions and draw conclusions based on probabilities.



2. Trial

Trial refers to a single execution of a random experiment. Each trial produces an outcome.

It typically refers to a single occurrence or attempt of an experiment or event. In probability theory, an experiment is often repeated multiple times to observe the outcomes and analyze the probabilities associated with those outcomes.

Each repetition or attempt of the experiment is considered a trial. For example, flipping a fair coin multiple times would involve a series of independent trials. Rolling a fair six-sided die and recording the number that appears can also be considered a trial.

The notion of trials is particularly relevant in scenarios where the outcomes are random and can vary from trial to trial. By conducting multiple trials and observing the results, probabilities can be estimated or calculated based on the frequency or likelihood of specific outcomes occurring over the course of those trials.

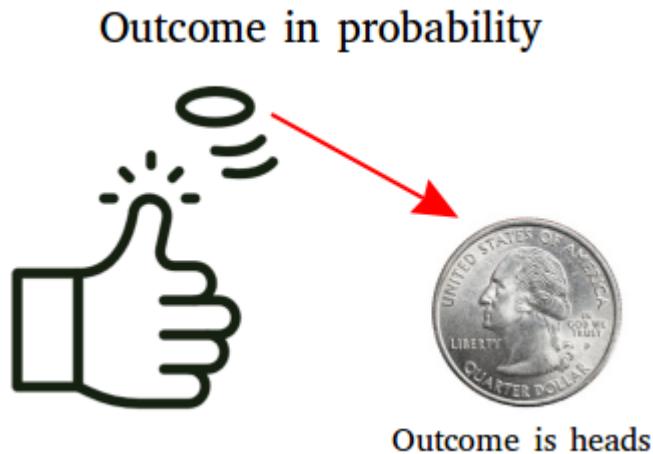
Trials are a fundamental concept in probability, as they allow for the study of probabilities across multiple iterations of an experiment and help build a deeper understanding of the likelihood of various outcomes.

3. Outcome

Outcome refers to a single possible result of a trial.

In probability theory, an outcome refers to a particular result or observation that occurs as a result of a random experiment or event. It represents a specific element or event within the sample space of the experiment.

When conducting a random experiment, the sample space consists of all possible outcomes that could potentially occur. Each outcome within the sample space represents a unique result of the experiment.



4. Sample Space

Sample Space of a random experiment is the set of all possible outcomes that can occur. Generally, one random experiment will have one set of sample space.

In probability theory, the sample space is the set of all possible outcomes of a random experiment or event. It represents the complete range of potential results that could occur. The sample space is denoted by the symbol Ω (Omega) and is a fundamental concept used to analyze probabilities.

The elements or outcomes within the sample space are typically denoted as simple events. Each outcome represents a distinct and unique result that can be observed or measured. The sample space encompasses all possible values or combinations of values that can arise from the experiment.

	1	2	3	4	5	6
H	1H	2H	3H	4H	5H	6H
T	1T	2T	3T	4T	5T	6T

A

$$\frac{2}{20}$$

B

$$\frac{3}{12}$$

C

$$\frac{3}{18}$$

5. Event

Event is a specific set of outcomes from a random experiment or process. Essentially, it's a subset of the sample space. An event can include a single outcome, or it can include multiple outcomes. One random experiments can have multiple events.

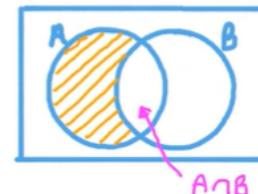
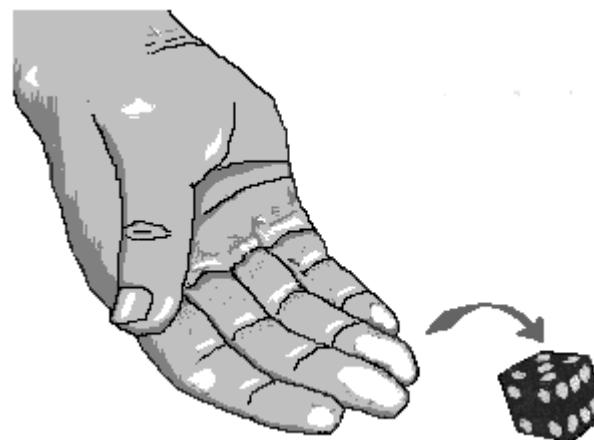
In probability theory, an event refers to a specific subset of the sample space in a random experiment. It represents a particular outcome or combination of outcomes that we are interested in analyzing or considering.

Formally, an event is a collection of outcomes from the sample space. It can consist of a single outcome (a simple event) or multiple outcomes (a compound event).

Events are denoted using capital letters, such as A, B, or C. They are used to describe and analyze the occurrence of specific outcomes or combinations of outcomes.

Probability Rules

- ① $0 \leq P(A) \leq 1$
- ② The sum of the probabilities of all possible outcomes = 1
- ③ $P(\bar{A}) = 1 - P(A)$ $\bar{A} = A'$ complement of A , NOT A
- ④ $P(A \text{ or } B) = P(A \cup B)$
- ⑤ $P(A \text{ and } B) = P(A \cap B)$
- ⑥ $P(A \cap \bar{B})$ the probability that A occurs but B does not
or
 $P(A) - P(A \cap B)$

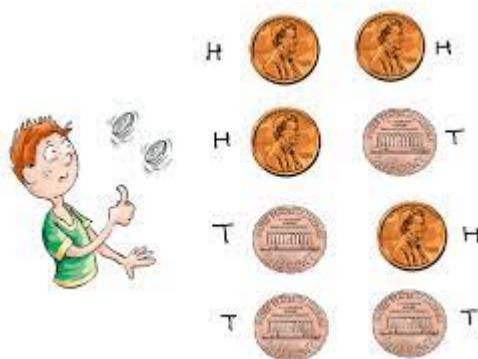
**Examples****1. Rolling a dice**

For Rolling dice: a random experiment, trial, outcome, sample space, and event for rolling a die:

- **Random Experiment:** Rolling a die.
- **Trial:** A single roll of the die.
- **Outcome:** The number that appears on the top face of the die after the roll.
- **Sample Space:** The set of all possible outcomes of the roll, which is $\{1, 2, 3, 4, 5, 6\}$.

- **Event:** The event of rolling a number greater than 3. The possible outcomes for this event are 4, 5, and 6.

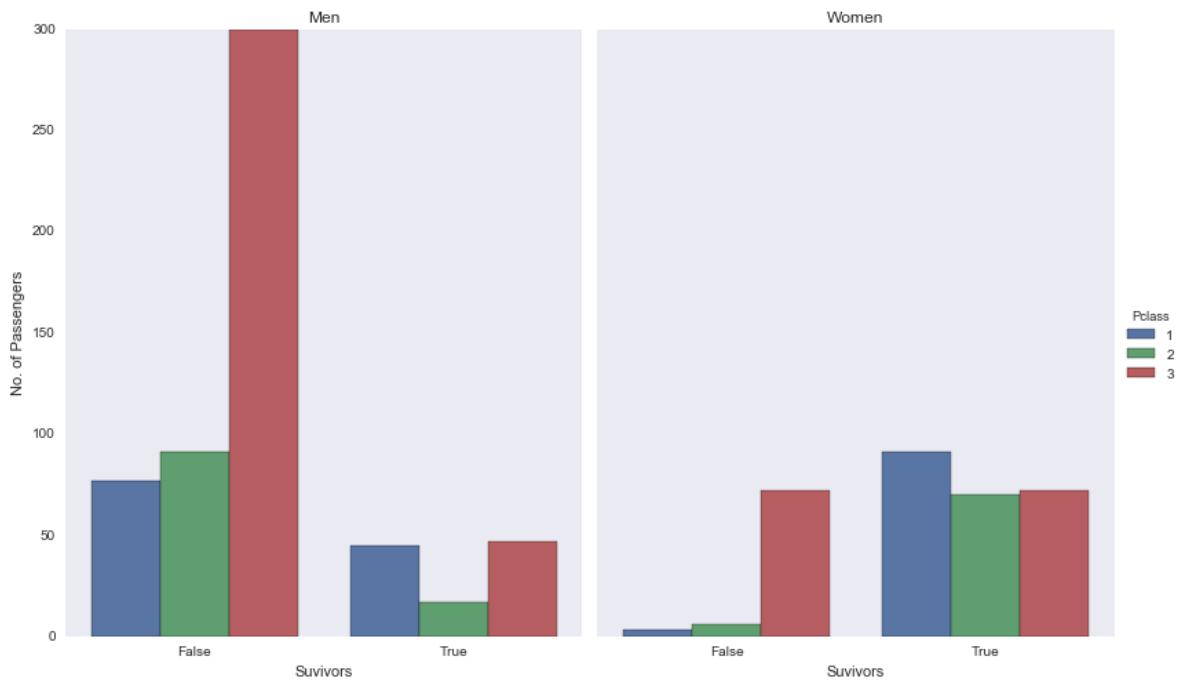
2. Tossing a coin twice



For tossing a coin twice a random experiment, trial, outcome, sample space, and event for tossing a coin twice:

- **Random Experiment:** Tossing a coin twice.
- **Trial:** A single toss of the coin.
- **Outcome:** The result of a single toss of the coin, which is heads or tails.
- **Sample Space:** The set of all possible outcomes of the two tosses, which is {HH, HT, TH, TT}.
- **Event:** The event of getting two heads. The possible outcomes for this event are HH.

Titanic Survived

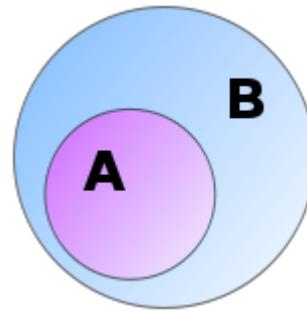


For Titanic dataset of a random experiment, trial, outcome, sample space, and event for the Titanic passengers, given their pclass:

- **Random Experiment:** Selecting a Titanic passenger at random and recording their pclass.
- **Trial:** A single selection of a Titanic passenger.
- **Outcome:** The pclass of the selected passenger.
- **Sample Space:** The set of all possible pclasses of the Titanic passengers, which is {1, 2, 3}.
- **Event:** A subset of the sample space. In this case, an event could be the event of selecting a passenger in pclass 1, or the event of selecting a passenger in pclass 3.

For example, one possible trial would be to select passenger number 320, who was in pclass 1. The outcome of this trial would be 1. The sample space for this experiment would be {1, 2, 3}. An event could be the event of selecting a passenger in pclass 1, which has probability $222/891 = 111/445$.

Types of Events

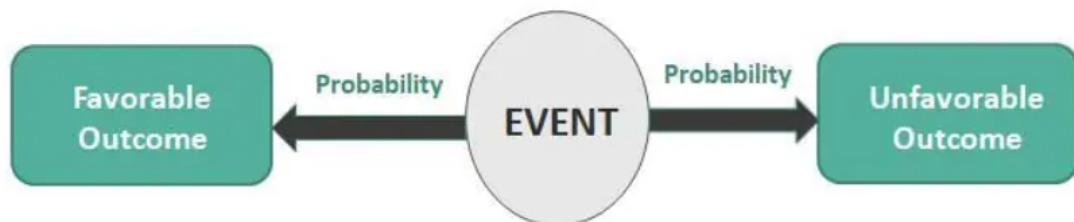


1. **Simple event:** Also known as an elementary event, a simple event is an event that consists of exactly one outcome.
 - For example, when rolling a fair six-sided die, getting a 3 is a simple event.
2. **Compound event:** A compound event consists of two or more simple events.
 - For example, when rolling a die, the event "rolling an odd number" is a compound event because it consists of three simple events: rolling a 1, rolling a 3, or rolling a 5.
3. **Independent events:** Two events are independent if the occurrence of one event does not affect the probability of the occurrence of the other event.
 - For example, if you flip a coin and roll a die, the outcome of the coin flip does not affect the outcome of the die roll.
4. **Dependent events:** Events are dependent if the occurrence of one event does affect the probability of the occurrence of the other event.
 - For example, if you draw two cards from a deck without replacement, the outcome of the first draw affects the outcome of the second draw because there are fewer cards left in the deck.
5. **Mutually exclusive events:** Two events are mutually exclusive (or disjoint) if they cannot both occur at the same time.
 - For example, when rolling a die, the events "roll a 2" and "roll a 4" are mutually exclusive because a single roll of the die cannot result in both a 2 and a 4.
- 6 **Exhaustive events:** A set of events is exhaustive if at least one of the events must occur when the experiment is performed.
 - For example, when rolling a die, the events "roll an even number" and "roll an odd number" are exhaustive because one or the other must occur on any roll.

Empirical Probability Vs Theoretical Probability

1. **Empirical probability:** Empirical probability, also known as **experimental probability**, is a probability measure that is based on observed data, rather than theoretical assumptions. It's calculated as the ratio of the number of times a particular event occurs to the total number of trials.
 - it is the probability of an event based on experimental data. For example, if you flip a coin 100 times and get heads 50 times, then the empirical probability of getting heads is 0.5.

Empirical Probability



2. **Theoretical probability :** Theoretical (or classical) probability is used when each outcome in a sample space is equally likely to occur. If we denote an event of interest as Event A, we calculate the theoretical probability of that event as:

Theoretical Probability of Event A = Number of Favourable Outcomes (that is, outcomes in Event A) / Total Number of Outcomes in the Sample Space

- it is the probability of an event based on mathematical calculations. For example, the theoretical probability of getting heads when flipping a fair coin is 0.5, because there are two equally likely outcomes (heads or tails).

$$\text{Theoretical Probability} = \frac{\text{Number of favorable (desired) outcomes}}{\text{Total number of possible outcomes}}$$

The main difference between empirical probability and theoretical probability is that empirical probability is based on data, while theoretical probability is based on assumptions. Empirical probability is more accurate when there is a large amount of data, but theoretical probability

can be used to calculate the probability of events that have never happened before.

- Here is a table that summarizes the key differences between empirical probability and theoretical probability:

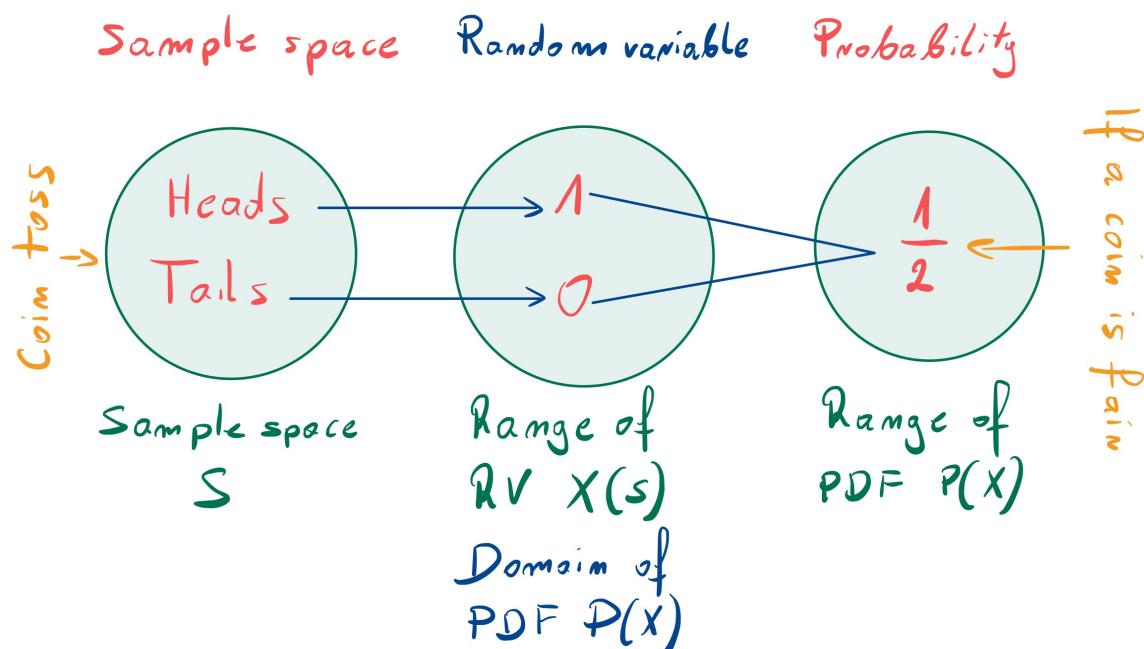
Characteristic	Empirical probability	Theoretical probability
Based on	Experimental data	Mathematical calculations
Accuracy	More accurate with large amounts of data	Less accurate, but can be used for events that have never happened before
Assumptions	None	Based on assumptions about the fairness of the experiment

Random Variable

A random variable is a variable whose value is determined by chance. It is a function that assigns a numerical value to each outcome in a sample space. The sample space is the set of all possible outcomes of an experiment.

There are two types of random variables: discrete and continuous.

- **Discrete random variables** can take on a finite number of values. For example, the number of heads in two coin flips is a discrete random variable. It can take on the values 0, 1, or 2.
- **Continuous random variables** can take on an infinite number of values. For example, the height of a randomly selected person is a continuous random variable. It can take on any value between the minimum and maximum height of a person.



- The probability distribution of a random variable is a function that gives the probability of each possible value of the random variable. For example, the probability distribution of the number of heads in two coin flips is:

$$P(0 \text{ heads}) = 1/4$$

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

$$P(2 \text{ heads}) = 1/4$$

- The expected value of a random variable is the average of its possible values. It is calculated by taking the sum of the possible values of the random variable, weighted by their probabilities. For example, the expected value of the number of heads in two coin flips is:

$$E = (0 \text{ heads} * 1/4) + (1 \text{ head} * 1/2) + (2 \text{ heads} * 1/4) = 1/2$$

- The variance of a random variable is a measure of how spread out its possible values are. It is calculated by taking the average of the squared deviations from the expected value. For example, the variance of the number of heads in two coin flips is:

$$\text{Var} = (0 \text{ heads} - 1/2)^2 * 1/4 + (1 \text{ head} - 1/2)^2 * 1/2 + (2 \text{ heads} - 1/2)^2 * 1/4 = 1/8$$

Random variables are used in many different areas of mathematics and statistics, including probability theory, statistics, and machine learning. They are a powerful tool for quantifying uncertainty and making predictions.

Probability Distribution of a Random Variable

A probability distribution is a list of all of the possible outcomes of a random variable along with their corresponding probability values.

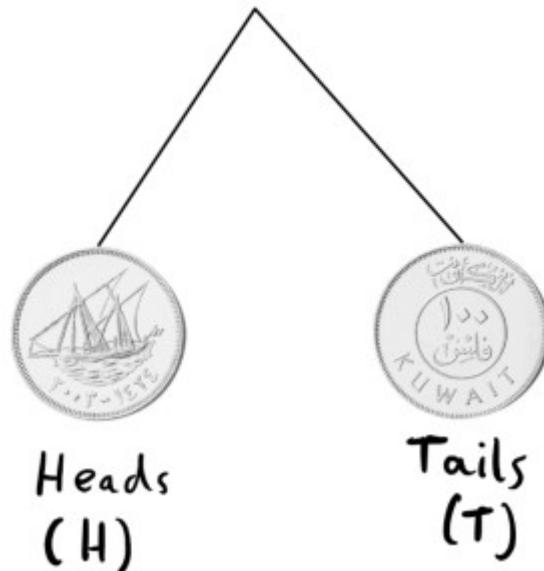
To describe the probability distribution of a random variable associated with different scenarios involving coin tosses and dice rolls, let's examine each case separately:

1. Tossing a Coin:

When tossing a fair coin, there are two possible outcomes: heads (H) or tails (T). We can define a random variable X to represent the outcome of a single coin toss. The probability distribution of X is as follows:

- $P(X = H) = 0.5$ (the probability of getting heads)
- $P(X = T) = 0.5$ (the probability of getting tails)

Only Two Possible Outcomes



2. Rolling a Single Die:

When rolling a fair six-sided die, there are six possible outcomes: 1, 2, 3, 4, 5, or 6. We can define a random variable Y to represent the outcome of a single die roll. The probability distribution of Y is as follows:

- $P(Y = 1) = 1/6$ (the probability of rolling a 1)
- $P(Y = 2) = 1/6$ (the probability of rolling a 2)
- $P(Y = 3) = 1/6$ (the probability of rolling a 3)
- $P(Y = 4) = 1/6$ (the probability of rolling a 4)
- $P(Y = 5) = 1/6$ (the probability of rolling a 5)
- $P(Y = 6) = 1/6$ (the probability of rolling a 6)

Probability of Rolling a Die

3. Rolling Two Dice:

When rolling two fair six-sided dice, the outcomes are determined by the sum of the numbers on the two dice. We can define a random variable Z to represent the sum of the two dice rolls. The probability distribution of Z is as follows:

(a,b)	1	2	3	4	5	6		1	2	3	4	5	6
1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)	+	2	3	4	5	6	7
2	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)		3	4	5	6	7	8
3	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)		4	5	6	7	8	9
4	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)		5	6	7	8	9	10
5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)		6	7	8	9	10	11
6	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)		7	8	9	10	11	12

- $P(Z = 2) = 1/36$ (the probability of rolling a sum of 2: 1+1)
- $P(Z = 3) = 2/36$ (the probability of rolling a sum of 3: 1+2 or 2+1)
- $P(Z = 4) = 3/36$ (the probability of rolling a sum of 4: 1+3, 2+2, or 3+1)
- $P(Z = 5) = 4/36$ (the probability of rolling a sum of 5: 1+4, 2+3, 3+2, or 4+1)
- $P(Z = 6) = 5/36$ (the probability of rolling a sum of 6: 1+5, 2+4, 3+3, 4+2, or 5+1)
- $P(Z = 7) = 6/36$ (the probability of rolling a sum of 7: 1+6, 2+5, 3+4, 4+3, 5+2, or 6+1)
- $P(Z = 8) = 5/36$ (the probability of rolling a sum of 8: 2+6, 3+5, 4+4, 5+3, or 6+2)
- $P(Z = 9) = 4/36$ (the probability of rolling a sum of 9: 3+6, 4+5, 5+4, or 6+3)
- $P(Z = 10) = 3/36$ (the probability of rolling a sum of 10: 4+6, 5+5, or 6+4)
- $P(Z = 11) = 2/36$ (the probability of rolling a sum of 11: 5+6 or 6+5)
- $P(Z = 12) = 1/36$ (the probability of rolling a sum of 12: 6+6)

In each case, the probability distribution assigns a probability to each possible value of the random variable. This allows us to understand the likelihood of different outcomes occurring and perform further calculations and analysis based on the probabilities associated with each value of the random variable.

Mean of a Random Variable

The mean of a random variable, often called the **expected value**, is essentially the average outcome of a random process that is repeated many times. More technically, it's a weighted average of the possible outcomes of the random variable, where each outcome is weighted by its probability of occurrence.

- The mean, or expected value, of a random variable is a measure of its central tendency. It represents the average value that the random variable is expected to take over many

repetitions of the experiment.

Expected Value, E(Y)

- For a continuous random variable Y, define the expected value of Y as

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy, \text{ if it exists.}$$

- Note this parallels our earlier definition for the discrete random variable:

$$E(Y) = \sum_y y p(y)$$

Mathematically, the expected value of a discrete random variable X with probability distribution $P(X = x)$ is calculated as:

$$E(X) = \Sigma(x * P(X = x))$$

where Σ represents the summation over all possible values of x that X can take.

For example,

let's consider the random variable Y representing the outcome of rolling a fair six-sided die. We already know its probability distribution:

- $P(Y = 1) = 1/6$
- $P(Y = 2) = 1/6$
- $P(Y = 3) = 1/6$
- $P(Y = 4) = 1/6$
- $P(Y = 5) = 1/6$
- $P(Y = 6) = 1/6$

To find the expected value of Y, we calculate:

$$E(Y) = (1/6 * 1) + (1/6 * 2) + (1/6 * 3) + (1/6 * 4) + (1/6 * 5) + (1/6 * 6)$$

Simplifying the expression gives:

$$E(Y) = 1/6 * (1 + 2 + 3 + 4 + 5 + 6) = 1/6 * 21 = 3.5$$

Hence, the mean or expected value of rolling a fair six-sided die is 3.5.

The expected value provides a useful summary measure of the central tendency of a random variable and is often used in probability theory and statistics for making predictions, decision-making, and further analysis of random phenomena.

X	1	2	3	4	5	6
P(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X) = \sum x_i p_i$$

$$E(X) = x_1 p_1 + x_2 p_2 + \dots + x_6 p_6$$

$$\begin{aligned} E(X) &= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \dots + 6 \times \frac{1}{6} \\ &= \frac{1}{6}(1+2+3+4+5+6) \\ &= \frac{7}{2} \end{aligned}$$

$$E(X^2) = \sum x_i^2 p_i$$

$$E(X^2) = x_1^2 p_1 + x_2^2 p_2 + \dots + x_6^2 p_6$$

$$\begin{aligned} E(X^2) &= \left(1^2 \times \frac{1}{6}\right) + \left(2^2 \times \frac{1}{6}\right) + \dots + \left(6^2 \times \frac{1}{6}\right) \\ &= \frac{1}{6}(1+4+9+16+25+36) \\ &= \frac{1}{6}(91) = \frac{91}{6} \end{aligned}$$

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

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2$$

$$= \frac{91}{6} - \frac{49}{4}$$

$$\text{Var}(X) = \frac{35}{12}$$

variance of a random variable

- The variance of a random variable is a statistical measurement that describes how much individual observations in a group differ from the mean (expected value).

VARIANCE OF DISCRETE RANDOM VARIABLES

mean
 $E(x)$

For discrete random variables X and Y ,

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

$$\text{Var}(X \pm a) = \text{Var}(X)$$

$$\text{Var}(ax) = a^2 \text{Var}(X)$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

The variance of a random variable quantifies the spread or dispersion of its probability distribution. It measures how much the values of the random variable deviate from its expected value or mean.

Mathematically, the variance of a discrete random variable X with probability distribution $P(X = x)$ and mean μ is calculated as:

$$\text{Var}(X) = \sum((x - \mu)^2 * P(X = x))$$

where Σ represents the summation over all possible values of x that X can take.

The variance provides a measure of the average squared deviation from the mean. Taking the square root of the variance gives the standard deviation, which is another commonly used measure of dispersion.

Continuing with the previous example of the random variable Y representing the outcome of rolling a fair six-sided die, with a mean of 3.5, let's calculate its variance:

$$\text{Var}(Y) = (1/6 * (1 - 3.5)^2) + (1/6 * (2 - 3.5)^2) + (1/6 * (3 - 3.5)^2) + (1/6 * (4 - 3.5)^2) + (1/6 * (5 - 3.5)^2) + (1/6 * (6 - 3.5)^2)$$

Simplifying the expression gives:

$$\text{Var}(Y) = (1/6 * 12.25) + (1/6 * 2.25) + (1/6 * 0.25) + (1/6 * 0.25) + (1/6 * 2.25) + (1/6 * 12.25)$$

$$\text{Var}(Y) = 35/12 \approx 2.9167$$

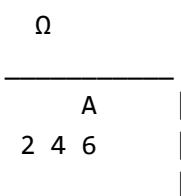
Hence, the variance of rolling a fair six-sided die is approximately 2.9167.

The variance provides valuable information about the spread of the random variable's values around its mean. It is a fundamental concept in probability theory and statistics and is used in various applications, such as assessing the reliability of estimators, analyzing the variability of data, and determining confidence intervals.

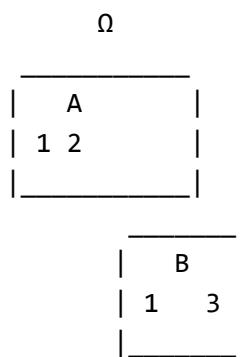
Venn diagrams in Probability

Venn diagrams are useful graphical representations that can be used to illustrate and analyze relationships between different sets or events in probability theory. While they are commonly associated with set theory, they can also be applied to depict various aspects of probability.

- Single Set:** Let's consider a sample space Ω that represents rolling a fair six-sided die. We can represent the event of rolling an even number (A) using a Venn diagram. The set A would contain the elements {2, 4, 6}, while the complement of A, which represents rolling an odd number, would contain the elements {1, 3, 5}. The Venn diagram would consist of a single circle representing the sample space Ω , and within it, a smaller circle representing the event A (even numbers).



- Two Sets:** Consider two events, A and B, representing rolling a number less than 3 and rolling an odd number, respectively. The Venn diagram for these events would have two circles overlapping, with each circle representing one event. The overlapping region represents the intersection of events A and B, which would correspond to rolling a number less than 3 and being odd.



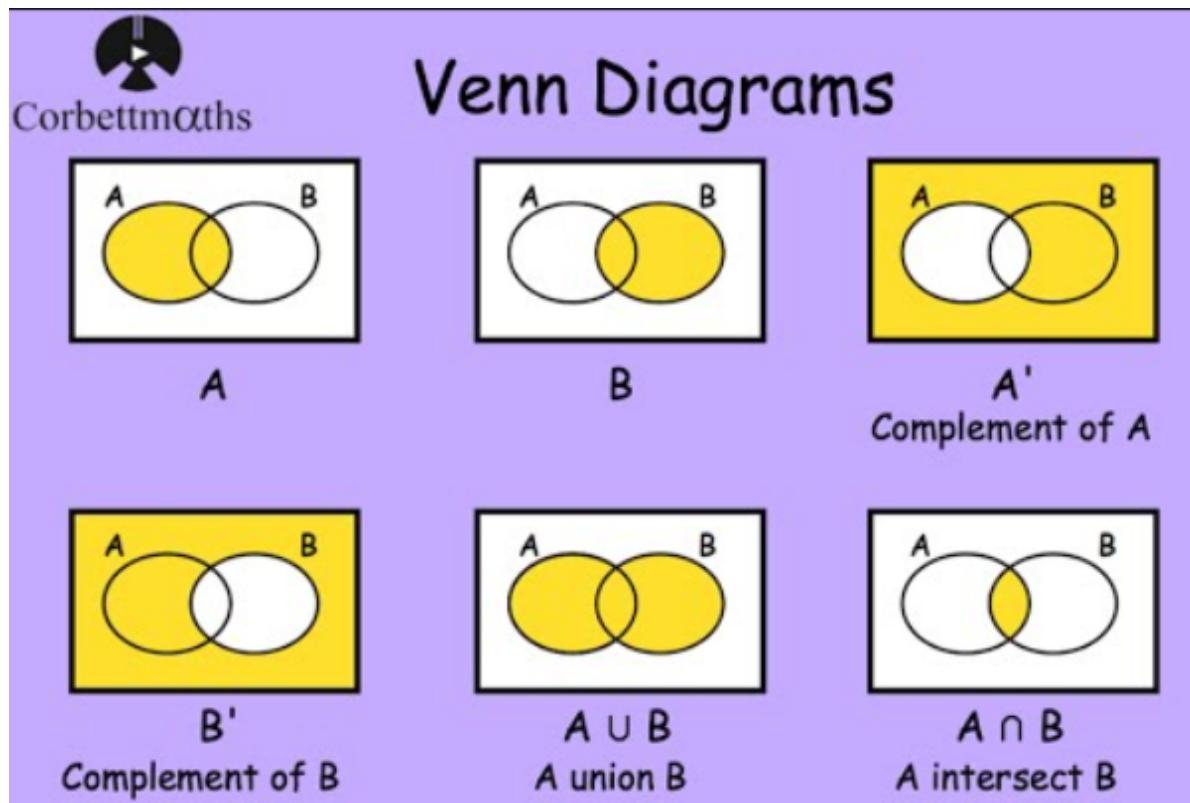
- Three Sets:** Suppose we have three events: A, B, and C, representing rolling an even number, rolling a number greater than 4, and rolling a prime number, respectively. The Venn diagram for these events would consist of three circles that overlap in various combinations to show the relationships between the events.

Ω

A
2 4 6

B
5 6

C
2 3 5



contingency table in probability

A contingency table, also known as a cross-tabulation table or a two-way frequency table, is a tabular representation that displays the joint frequencies or probabilities of two or more categorical variables. It is commonly used in probability and statistics to analyze the relationship between different variables and understand their associations.

- The contingency table organizes the data into rows and columns, with each cell representing the frequency or probability of a specific combination of values from the categorical variables.

	Pass	Fail	Total
Males	46	56	102
Females	68	30	98
Total	114	86	200

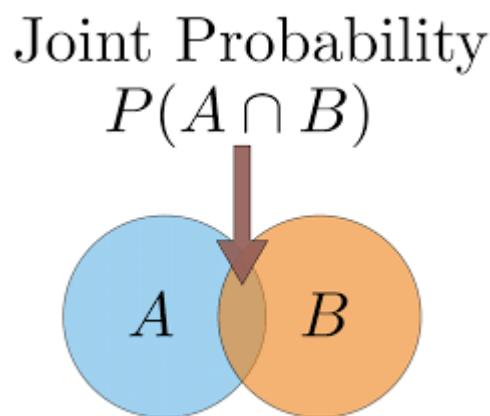
Contingency table

Types of Probability

1. Joint Probability
2. Marginal Probability
3. Conditional Probability

1. Joint Probability

Joint probability is the probability of two or more events occurring simultaneously. It is denoted as $P(A \text{ and } B)$ and measures the likelihood of events A and B happening together. Joint probability is commonly used in probability calculations involving multiple variables or events.



- Let's say we have two random variables X and Y. The joint probability of X and Y, denoted as $P(X = x, Y = y)$, is the probability that X takes the value x and Y takes the value y at the same time.

Example: titanic dataset

1. Let X be a random variable associated with the Pclass of a passenger. Let Y be a random variable associated with the survival status of a passenger

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

In [2]: `df =pd.read_csv("https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv")
df.head()`

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0		1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	T
1		2	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina	female	38.0	1	0	PC 17599	71.2833	S
2		3	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	26.0	0	0	STON/O2. 3101282	7.9250	C
3		4	1	Allen, Mr. William Henry	male	35.0	1	0	113803	53.1000	C
4		5	0				0	0	373450	8.0500	T

In [3]: `# Contingency table`

`pd.crosstab(df['Survived'], df['Pclass'])`

Out[3]:

Pclass	1	2	3
Survived			
0	80	97	372
1	136	87	119

In [4]: `80/891 # Dividing every value by 891 which is the total`

Out[4]: `0.08978675645342311`

In [5]: # Joint Probability Distribution

```
pd.crosstab(df[ 'Survived' ],df[ 'Pclass' ],normalize='all')
```

Out[5]:

Pclass	1	2	3
Survived			
0	0.089787	0.108866	0.417508
1	0.152637	0.097643	0.133558

2. Marginal Probability

The marginal probability of a random variable refers to the probability distribution of that variable without considering the values of other variables in a joint probability distribution. It represents the probability of a single variable independently, without taking into account the values of any other variables.

To calculate the marginal probability of a random variable, you sum (or integrate, in the case of continuous variables) the joint probabilities over all possible values of the other variables, effectively "marginalizing" them out.

Example: titanic dataset

1. Let X be a random variable associated with the Pclass of a passenger 2. Let Y be a random variable associated with the survival status of a passenger

In [6]: # Marginal Probability

```
pd.crosstab(df[ 'Survived' ],df[ 'Pclass' ],normalize='all', margins=True)
```

Out[6]:

Pclass	1	2	3	All
Survived				
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

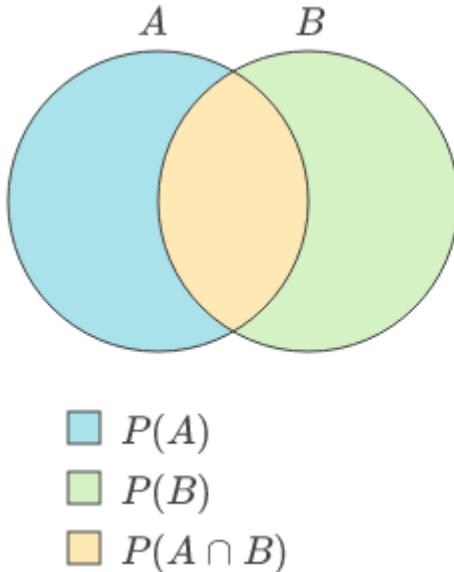
3. Conditional Probability

Conditional probability is the probability of an event occurring given that another event has already occurred. In the context of random variables, conditional probability is the probability that a random variable will take on a certain value given that another random variable has already taken on a certain value.

The conditional probability of event B given event A is denoted by $P(B|A)$. It is calculated as follows:

- $P(A|B) = P(A \cap B) / P(B)$

where $P(A \cap B)$ is the probability that both events A and B occur, and $P(A)$ is the probability that



Conditional Probability Formula

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

Probability that A occurs given that B has already occurred

KEY POINTS

- If the probability of event B is affected by the outcome of event A, then we say the probability of event B is **CONDITIONAL** on event A.

- $P(B|A)$ ~ the probability of B given A

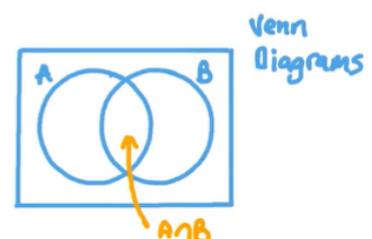
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$ or $P(B|A) = \frac{P(B \cap A)}{P(A)}$

- Events A and B are **INDEPENDENT** if

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

AND

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



- **For example**, let's say we have two random variables, X and Y. X is the number of heads that appear when we flip a coin 10 times, and Y is the number of tails that appear when we flip a coin 10 times.

The probability that X is 5 given that Y is 5 is calculated as follows:

$$P(X = 5 | Y = 5) = P(X = 5 \cap Y = 5) / P(Y = 5)$$

$P(X = 5 \cap Y = 5)$ is the probability that both X is 5 and Y is 5. This can be calculated as follows:

$$P(X = 5 \cap Y = 5) = (1/2)^{10} = 1/1024$$

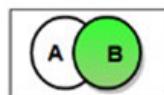
$P(Y = 5)$ is the probability that Y is 5. This can be calculated as follows:

$$P(Y = 5) = (1/2)^{10} = 1/1024$$

Therefore, $P(X = 5 | Y = 5) = 1/1$.

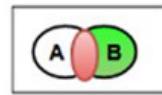
Conditional probability is a useful tool for calculating the probability of events that are related to each other. It can be used to make predictions about the future, and to make decisions under uncertainty.

$$P(A|B) = P(A \text{ given } B \text{ has occurred})$$



If B has already occurred
then our sample space
must be somewhere within B

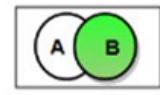
Now A can occur only
within sample space B



$P(A|B)$ is the ratio of Red
space divided by Green space

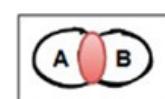
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = P(B \text{ given } A \text{ has occurred})$$



If A has already occurred
then our sample space
must be somewhere within A

Now B can occur only
within sample space A



$P(B|A)$ is the ratio of Red
space divided by White space

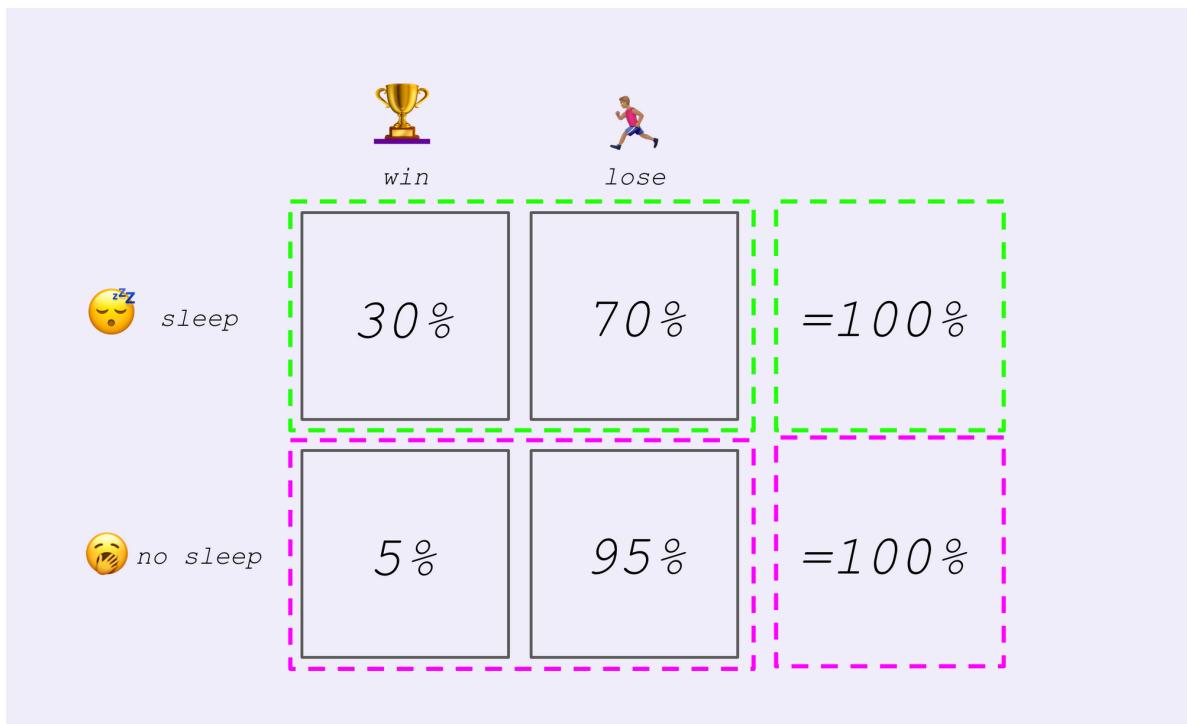
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Here are some of the properties of conditional probability:

- $P(B|A) = P(A|B)$ if A and B are independent events.
- $P(B|A) \leq 1$.
- $P(B|A) = 0$ if A and B are mutually exclusive events.

Conditional probability is a powerful tool that can be used to analyze the relationships between random variables. It is a fundamental concept in probability theory and statistics, and it has many applications in a variety of fields, such as machine learning, data science, and decision making.

For example, consider the probability of winning a race, given the condition you didn't sleep the night before. You might expect this probability to be lower than the probability you'd win if you'd had a full night's sleep.



1. **Prob** 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?

Sol: The sample space for tossing three unbiased coins is $2^3 = 8$. The events that satisfy the condition "at least two coins show heads" are:

- HHH
- HHT
- HTH
- THH

The events that satisfy the condition "at least one coin shows heads" are:

- HHH
- HHT
- HTH
- THH
- TTH
- THT
- HTT
- TTT

The probability that at least two coins show heads is $\frac{4}{8} = \frac{1}{2}$. The probability that at least one coin shows heads is $\frac{7}{8}$.

Therefore, the conditional probability that at least two coins show heads, given that at least one coin shows heads is:

$$P(\text{at least 2 heads} | \text{at least 1 head}) = \frac{P(\text{at least 2 heads} \cap \text{at least 1 head})}{P(\text{at least 1 head})} = \frac{\frac{4}{8}}{\frac{7}{8}} = \boxed{\frac{4}{7}}$$

In other words, the probability of getting at least two heads given that at least one head has already been flipped is $\frac{4}{7}$.

2. **Prob:** Two fair six-sided dice are rolled. What is the conditional probability that the sum of the numbers rolled is 7, given that the first die shows an odd number?

Sol:

		Dice 1					
		1	2	3	4	5	6
Dice 2	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

Ruander.com

To calculate the conditional probability that the sum of the numbers rolled is 7, given that the first die shows an odd number, we need to determine the total number of outcomes where the first die shows an odd number and the sum of the numbers rolled is 7, and the total number of outcomes where the first die shows an odd number.

Let's consider the possible outcomes when rolling two fair six-sided dice. Each die has six possible outcomes: 1, 2, 3, 4, 5, or 6.

Out of these possible outcomes, there are three outcomes where the first die shows an odd number: (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6).

Among these outcomes, there is one outcome where the sum of the numbers rolled is 7: (1, 6).

Therefore, the conditional probability that the sum of the numbers rolled is 7, given that the first die shows an odd number, is:

$$P(\text{sum is 7} | \text{first die is odd}) = (\text{Number of outcomes with sum 7 and first die is odd}) / (\text{Number of outcomes with first die is odd}) = 1 / 6 \approx 0.1667$$

So, the conditional probability is approximately 0.1667 or 16.67%.

3. **Prob:** Two fair six-sided dice are rolled, denoted as D1 and D2. What is the conditional probability that D1 equals 2, given that the sum of D1 and D2 is less than or equal to 5?



Sol : To calculate the conditional probability that D1 equals 2, given that the sum of D1 and D2 is less than or equal to 5, we need to determine the total number of outcomes where D1 equals 2 and the sum of D1 and D2 is less than or equal to 5, and the total number of outcomes where the sum of D1 and D2 is less than or equal to 5.

Let's consider the possible outcomes when rolling two fair six-sided dice. Each die has six possible outcomes: 1, 2, 3, 4, 5, or 6.

Among these outcomes, there are three outcomes where the sum of D1 and D2 is less than or equal to 5: (1, 1), (1, 2), (2, 1).

Out of these outcomes, only one outcome has D1 equals 2: (2, 1).

Therefore, the conditional probability that D1 equals 2, given that the sum of D1 and D2 is less than or equal to 5, is:

$$P(D1 = 2 | D1 + D2 \leq 5) = (\text{Number of outcomes with } D1 = 2 \text{ and } D1 + D2 \leq 5) / (\text{Number of outcomes with } D1 + D2 \leq 5) = 1 / 3$$

So, the conditional probability is 1/3 or approximately 0.3333, which is about 33.33%.

4. Prob Sol

Suppose $P(A) = \frac{2}{5}$ and $P(B) = \frac{3}{7}$. The probability that event A occurs and event B also occurs is $\frac{1}{5}$. Calculate $P(A|B)$, and then evaluate whether events A and B are independent.

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

$$\begin{aligned} P(A|B) &= \frac{\frac{1}{5}}{\frac{3}{7}} \\ &= \frac{1}{5} \div \frac{3}{7} \\ &= \frac{1}{5} \times \frac{7}{3} \end{aligned}$$

$$P(A|B) = \frac{7}{15}$$

If $P(A) = P(A|B)$ and $P(B) = P(B|A)$ then events A and B are independent.

$$P(A) = \frac{2}{5}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{7}{15}$$

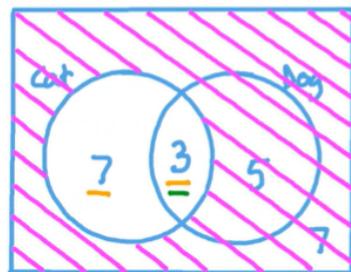
$$\frac{6}{15} \neq \frac{7}{15}$$

$P(A) \neq P(A|B)$ so NOT independent

5. Prob Sol

On the street, 10 houses have a cat (C), 8 houses have a dog (D), 3 houses have both, and 7 houses have neither.

- 1) Find the total number of houses on the street. Hence find the probability that a house chosen at random has both a cat and a dog. Give your answer to 3 decimal places. [22 houses, $P(\text{cat and dog}) = 0.316$ (3dp)]
- 2) Find the probability that a house on the street has either a cat or a dog or both. Give your answer to 3 decimal places. $P(\text{cat or dog or both}) = 0.682$ (3dp)



- 3) If a house on the street has a cat, find the probability that there is also a dog living there.

$$P(\text{dog} | \text{cat}) = \frac{3}{10} = 0.3$$

probability of having a dog GIVEN that the house has a cat

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

In [12]: # code

```
# Conditional Probability

pd.crosstab(df['Survived'],df['Pclass'],normalize='columns')
```

Out[12]:

Pclass	1	2	3
Survived			
0	0.37037	0.527174	0.757637
1	0.62963	0.472826	0.242363

In [13]: ## Reverse

```
pd.crosstab(df['Survived'],df['Pclass'],normalize='index')
```

Out[13]:

Pclass	1	2	3
Survived			
0	0.145719	0.176685	0.677596
1	0.397661	0.254386	0.347953

- **Explanation of code** In the given code, the pd.crosstab() function from the pandas library is used to create a cross-tabulation or contingency table. This table shows the relationship between two variables: 'Survived' and 'Pclass' from the DataFrame df. The purpose of the code is to calculate the conditional probability of survival ('Survived') for each passenger class ('Pclass').

The original code, which is commented out, calculates the conditional probability by normalizing the columns. It means that for each passenger class, the sum of probabilities of survival and non-survival will be equal to 1. This provides insights into the distribution of survivors among different passenger classes.

However, in the focal cell, the code is modified by using normalize='index' instead. This normalizes the rows, so for each survival outcome (0 or 1), the sum of probabilities across different passenger classes will be equal to 1. This gives information about the distribution of passenger classes among survivors and non-survivors.

The output of the code is a table that displays the conditional probabilities. Each row represents a survival outcome (0 or 1), and each column represents a passenger class (1, 2, or 3). The values in the table represent the probabilities of belonging to a specific passenger class given the survival outcome. For example, the value 0.145719 indicates the probability of belonging to passenger class 1 given the survival outcome 0 (non-survivor).

Please note that the STDOUT and result sections in the question are empty, so we don't have any specific output to refer to."""

Intuition behind the conditional Probability formula

The conditional probability formula provides a way to calculate the probability of an event occurring given that another event has already occurred. It is derived from the concept of joint probability and the idea that when one event has occurred, the sample space or set of possible outcomes is reduced to the subset of outcomes where the given event has occurred.

Intuitively, the conditional probability formula can be understood as follows:

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

The numerator $P(A \cap B)$ represents the probability that both events A and B occur together. This can be seen as considering the subset of outcomes where both A and B occur out of the total set of outcomes where event B has occurred.

The denominator $P(B)$ represents the probability of event B occurring. It provides the context or the "condition" for the probability of A given B.

By dividing the probability of A and B occurring together by the probability of B occurring, we obtain the conditional probability $P(A|B)$, which measures the likelihood of event A occurring given that event B has occurred.

The formula intuitively captures the idea that when the condition B has occurred, we focus only on the subset of outcomes where B has occurred, and within that subset, we calculate the proportion of outcomes where A also occurs.

In summary, the conditional probability formula allows us to adjust our probabilities based on new information or conditions, providing a framework to update our understanding of the likelihood of events given certain constraints or previous outcomes.

Independent Events , Dependent Events and Mutually Exclusive Events

- **Independent events** : are events where the occurrence of one event does not affect the occurrence of another.

Independent Events are events where the occurrence of one event does not affect the probability of the other event occurring. For example, flipping a coin twice is a set of independent events. The outcome of the first flip does not affect the outcome of the second flip.

Examples:

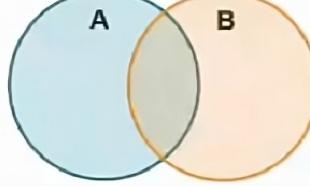
1. Flipping a coin and rolling a die
2. Drawing a card with replacement

- **Dependent events** are events where the occurrence of one event does affect the occurrence of another.

Dependent Events are events where the occurrence of one event does affect the probability of the other event occurring. For example, drawing a card from a deck and then drawing another card without replacing the first card are dependent events. The probability of drawing a heart on the second draw is affected by the fact that you already drew a heart on the first draw.

Examples

1. Drawing a card with replacement

Mutually Exclusive Events	Independent Events
<ul style="list-style-type: none"> ■ Cannot occur at the same time ■ No intersection ■ "OR" events ■ Addition rule 	<ul style="list-style-type: none"> ■ Can occur at the same time ■ Can have intersection ■ "AND" events ■ Multiplication rule 

- **Mutually exclusive events** are events that cannot both occur at the same time. In other words, if one event occurs, the other cannot.

Mutually exclusive events are events where two events cannot occur at the same time. For example, flipping a coin and getting heads and tails are mutually exclusive events. It is not possible to get both heads and tails on the same flip.

Examples

1. Flipping a coin
2. rolling dice

Here is a table that summarizes the differences between independent events, dependent events, and mutually exclusive events:

Property	Independent Events	Dependent Events	Mutually Exclusive Events
Occurrence of one event affects the probability of the other event occurring	No	Yes	No
Can two events occur at the same time?	Yes	No	No
Example	Flipping a coin twice	Drawing a card from a deck and then drawing another card without replacing the first card	Flipping a coin and getting heads and tails

	If statistically independent	If mutually exclusive
$P(A B) =$	$P(A)$	0
$P(B A) =$	$P(B)$	0
$P(A \cap B) =$	$P(A)P(B)$	0

Bayes Theorem

Bayes' theorem is a fundamental concept in probability theory and statistics that allows us to update our beliefs about the probability of an event based on new evidence or information. It relates the conditional probabilities of two events.

Mathematically, Bayes' theorem is expressed as follows:

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

where:

- $P(A|B)$ is the conditional probability of event A given event B.
- $P(B|A)$ is the conditional probability of event B given event A.
- $P(A)$ and $P(B)$ are the individual probabilities of events A and B, respectively.

Bayes' theorem can be derived using the definition of conditional probability:

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

Rearranging the terms and using the commutative property of intersection (\cap):

$$P(A \cap B) = P(B \cap A) = P(B|A) * P(A)$$

Substituting these values back into the conditional probability equation:

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

Bayes' theorem is particularly useful in situations where we have prior knowledge or beliefs about the probabilities of certain events (prior probability), and we want to update those probabilities based on new observations or evidence (posterior probability).

It has applications in various fields, including medical diagnosis, machine learning, spam filtering, and many other areas where we need to update our beliefs based on new information.

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

posterior prior likelihood marginal

PROOF OF BAYES THEOREM

The probability of two events A and B happening, $P(A \cap B)$, is the probability of A, $P(A)$, times the probability of B given that A has occurred, $P(B|A)$.

$$P(A \cap B) = P(A)P(B|A) \quad (1)$$

On the other hand, the probability of A and B is also equal to the probability of B times the probability of A given B.

$$P(A \cap B) = P(B)P(A|B) \quad (2)$$

Equating the two yields:

$$P(B)P(A|B) = P(A)P(B|A) \quad (3)$$

and thus

$$P(A|B) = P(A) \frac{P(B|A)}{P(B)} \quad (4)$$

This equation, known as Bayes Theorem is the basis of statistical inference.

Example problem that involves applying Bayes' theorem.

Problem: A factory produces two types of products, Type A and Type B. The probability of a Type A product being defective is 0.05, while the probability of a Type B product being defective is 0.10. The factory produces 60% Type A products and 40% Type B products. If a randomly chosen product from the factory is defective, what is the probability that it is Type A?

Solution:

Let's define the events: A = Event that the product is Type A B = Event that the product is defective

We are given: $P(B|A) = 0.05$ (Probability of a product being defective given that it is Type A)
 $P(B|B') = 0.10$ (Probability of a product being defective given that it is Type B) $P(A) = 0.60$
(Probability of selecting a Type A product) $P(B') = 1 - P(A) = 0.40$ (Probability of selecting a Type B product)

We need to calculate: $P(A|B)$ (Probability that the product is Type A given that it is defective)

Using Bayes' theorem:

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

We need to calculate $P(B)$, which can be done using the law of total probability:

$$P(B) = P(B|A) * P(A) + P(B|B') * P(B')$$

Substituting the given values:

$$P(B) = (0.05 * 0.60) + (0.10 * 0.40) = 0.03 + 0.04 = 0.07$$

Now, we can calculate $P(A|B)$:

$$P(A|B) = (P(B|A) * P(A)) / P(B) = (0.05 * 0.60) / 0.07 = 0.03 / 0.07 \approx 0.4286$$

Therefore, the probability that the defective product is Type A is approximately 0.4286 or

Let's work through another problem that involves applying Bayes' theorem.

Problem: A factory produces two types of widgets, Type A and Type B. The production rate for Type A widgets is 60%, while for Type B widgets, it is 40%. However, 5% of Type A widgets and 10% of Type B widgets are defective. If a randomly selected widget is defective, what is the probability that it is of Type A?

Solution:

To solve this problem, we can use Bayes' theorem. Let's define the events:

A: Widget is of Type A B: Widget is defective

We are given: $P(A) = 0.6$ (probability of selecting a Type A widget) $P(B|A) = 0.05$ (probability of a widget being defective given it is of Type A)

We need to find: $P(A|B)$ (probability that a defective widget is of Type A)

Using Bayes' theorem: $P(A|B) = (P(B|A) * P(A)) / P(B)$

To calculate $P(B)$, we need to consider the probability of a widget being defective, regardless of its type. We can calculate this using the law of total probability:

$$P(B) = P(B|A) * P(A) + P(B|not A) * P(not A)$$

Since we only have information about defective rates for Type A and Type B widgets, we need to calculate $P(B|not A)$ and $P(not A)$:

$$P(B|not A) = 0.1 \text{ (probability of a widget being defective given it is not of Type A)} \\ P(not A) = 1 - P(A) = 1 - 0.6 = 0.4 \text{ (probability of selecting a widget that is not of Type A)}$$

Substituting the values into the formula for P(B):

$$P(B) = P(B|A) * P(A) + P(B|\text{not } A) * P(\text{not } A) = 0.05 * 0.6 + 0.1 * 0.4 = 0.03 + 0.04 = 0.07$$

Now we can substitute all the values into Bayes' theorem:

$$P(A|B) = (P(B|A) * P(A)) / P(B) = (0.05 * 0.6) / 0.07 = 0.03 / 0.07 \approx 0.4286$$

Therefore, the probability that a defective widget is of Type A is approximately 0.4286 or 42.86%.

Note: It's essential to carefully identify the events, probabilities, and conditional probabilities in the problem and apply Bayes' theorem accordingly.

In []: