



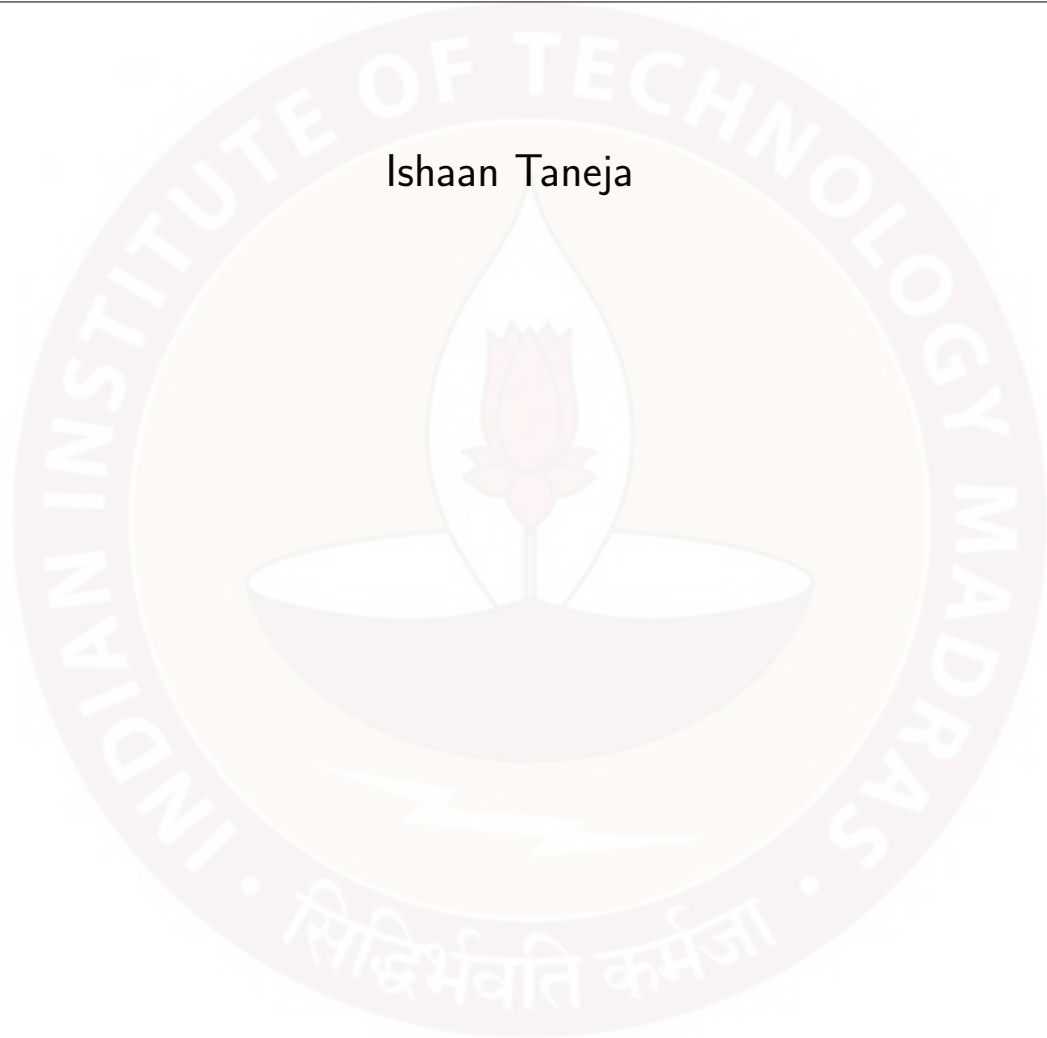
IIT MADRAS BS DEGREE



STATISTICS

PROBABILITY AND PROBABILITY DISTRIBUTIONS

Ishaan Taneja



About the author

The author, Ishaan Taneja, is a highly accomplished and driven professional with a profound expertise in the fields of Statistics and Computer Science. With a solid academic background, including a B.Sc. (Hons.) in Statistics from Ram Lal Anand College, University of Delhi, and an M.Sc. in Statistics from Hindu College, University of Delhi. They have consistently demonstrated exceptional performance and a passion for data analysis.

Currently pursuing an M.S. in Computer Science at IIT Madras, Ishaan Taneja is actively exploring the intersection of Statistics and Machine Learning, further expanding their knowledge and understanding of the subject. With their diverse experiences, including their role as a Course Instructor at IIT Madras BS Degree Programme, where they have contributed to the development of an online Statistics coursework.

Ishaan Taneja possesses the unique ability to bridge theoretical concepts with practical applications. Through their book, "Probability and Probability Distributions," they aim to provide readers with a comprehensive and accessible exploration of these subjects, sharing their expertise and empowering others in the field.

Contents

1	Data, Statistics and Probability	6
1.1	Introduction	6
1.1.1	Deterministic pattern	6
1.1.2	Random-like pattern	6
1.1.3	Statistical study of phenomenon	7
1.2	Basic Concepts	7
1.2.1	Experiment	7
1.2.2	Outcome	7
1.2.3	Sample Space	8
1.2.3.1	Solved Examples:	8
1.3	Events	9
1.3.1	Occurrence of events	10
1.3.1.1	One event can be contained in another:	10
1.3.1.2	Complement of an event:	10
1.3.2	Combining events	11
1.3.2.1	Union of events:	11
1.3.2.2	Intersection of events:	11
1.3.2.3	Solved examples:	12
1.3.3	Disjoint events	12
1.3.3.1	Partition	13
1.3.4	De Morgan's Law	13
1.3.5	Solved Examples:	14
1.4	Venn Diagram	15
2	Introduction to Probability	17
2.1	Events and Chance	17
2.2	Mathematical and Probability theory	18
2.2.1	Mathematical Theory	18
2.2.2	Probability Theory	18
2.2.2.1	Probability	18
2.3	Basic properties of probability	20
2.3.1	Working with probability spaces	22
2.4	Distributions	24
2.4.1	Uniform Distribution on a finite sample space	25
2.4.2	Solved Examples:	25
2.5	Conditional Probability	27
2.5.1	Conditional Probability Space	27
2.5.2	Multiplication rule:	28
2.5.3	Solved Examples:	28
2.6	Bayes' Theorem and Independence	31
2.6.1	Law of Total Probability	31

	2.6.1.1	Solved Examples:	32
2.6.2	Bayes' Theorem		34
	2.6.2.1	Solved Examples:	35
2.6.3	Independence of Two Events		38
	2.6.3.1	Solved Examples:	39
2.6.4	Mutual Independence of Three Events		41
	2.6.4.1	Solved Examples:	41
2.7	Repeated Independent Trials		43
2.7.1	Bernoulli Distribution		43
	2.7.1.1	Single Bernoulli Trial	43
	2.7.1.2	Repeated Bernoulli Trials	44
	2.7.1.3	Solved Examples:	44
2.7.2	Binomial Distribution		47
	2.7.2.1	Visualtion of Binomial Distribution	47
	2.7.2.2	Solved Examples:	49
2.7.3	Geometric Distribution		51
	2.7.3.1	Visualisation of Geometric Distribution	52
	2.7.3.2	Solved Examples:	53
2.8	Discrete Random Variable		56
2.8.1	Random Variable and events		57
2.8.2	Distribution of a Discrete Random Variable		58
	2.8.2.1	Probability Mass Function (PMF)	58
	2.8.2.2	Properties of PMF	59
	2.8.3	Solved Examples:	59
2.9	Common Discrete Distributions		64
2.9.1	Uniform Random Variable		64
2.9.2	Bernoulli Random Variable		64
2.9.3	Binomial Random Variable		65
2.9.4	Geometric Random Variable		65
2.9.5	Negative Binomial Distribution		65
	2.9.5.1	Negative Binomial Random Variable	66
2.9.6	Poisson Distribution		66
	2.9.6.1	Application of Poisson Distribution	66
	2.9.6.2	Visualtion of Poisson Distribution	69
	2.9.6.3	Poisson Random Variable	69
2.9.7	Hypergeometric Distribution		69
	2.9.7.1	Hypergeometric Random Variable	70
2.10	Functions of one discrete random variable		70
2.10.1	Solved Examples:		71

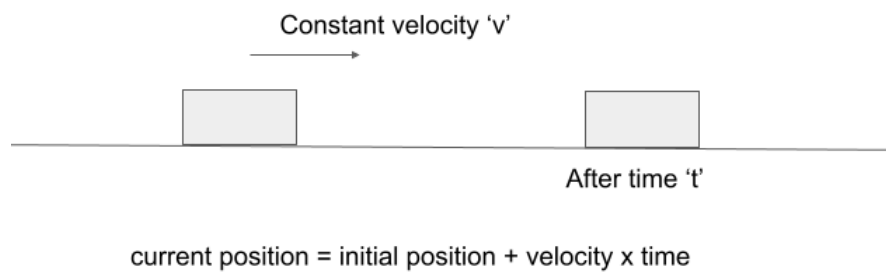
Chapter 1

1 Data, Statistics and Probability

1.1 Introduction

1.1.1 Deterministic pattern

- The result can be predicted with certainty.
- There is a theoretical framework for deterministic phenomenon.
- For example:

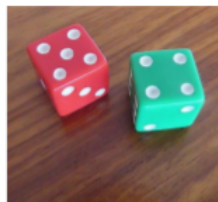


1.1.2 Random-like pattern

- The result can be predicted with chances of it happening but not with certainty.
- We cannot expect exact models but can expect some patterns when the random-phenomenon repeats.
- For example:



Tossing a coin



Rolling a die

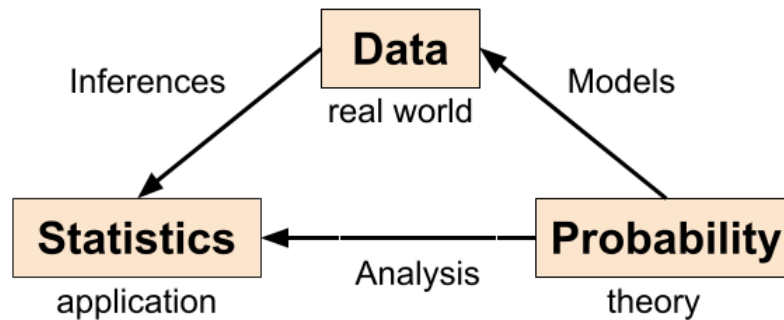


Rainfall quantity



Value of stocks

1.1.3 Statistical study of phenomenon



1.2 Basic Concepts

1.2.1 Experiment

In statistics, any process or phenomenon that we wish to study statistically is called an experiment.

In other words, when we want to study statistically to see if there are statistical patterns in the data or phenomena, we use the word experiment to denote that.

Examples

- (i) Tossing a coin.
- (ii) Rolling a die.
- (iii) Indian Premier League tournament.

1.2.2 Outcome

The result of an experiment (in as much detail as necessary) is referred to as the Outcome of that experiment.

It may also be referred to as data in some sense.

Examples

- (i) When the experiment is 'Tossing a coin', the possible outcomes could be either 'Heads' or 'Tails'.
- (ii) When the experiment is 'Rolling a die', the possible outcomes could be either '1' or '2' or '3' or '4' or '5' or '6'.
- (iii) When the experiment is complex like 'Indian Premier League tournament', we cannot list down all possible outcomes easily. (It could be available in YAML format file)

1.2.3 Sample Space

The sample space is a set that contains all outcomes of an experiment. It is typically denoted by ' S '.

Note:

- In practice, it is enough to imagine a sample space instead of explicitly writing it down.
- In situations where confusion occurs, writing down the sample space can give the clarity.

Examples

- (i) If the experiment is 'Tossing a coin' then, the sample space; $S = \{\text{Heads, Tails}\}$.
- (ii) If the experiment is 'Rolling a die' then, the sample space; $S = \{1, 2, 3, 4, 5, 6\}$.
- (iii) If the experiment is as complex as 'Indian Premier League tournament' then the sample space is difficult to write down. So, we generally break it into small experiments depending on what level of details we want.
 - Runs scored in one delivery: $\{1, 2, 3, 4, \dots\}$
 - Winner in 2022: $\{\text{CSK, MI, DC, } \dots\}$ (10 teams in 2022)

1.2.3.1 Solved Examples:

- Q1. In an urn of marbles, there are three marbles each of colour red, blue and white. If the experiment is drawing a marble from the urn, then write down its sample space.

Solution:

The possible outcomes are that either the marble drawn is of red colour or blue colour or white colour.

Hence, the sample space; $S = \{\text{red, blue, white}\}$

- Q2. In an urn of marbles, there are two red coloured marbles, and each one of blue and white colour. If the experiment is drawing two marbles from the urn with replacement, then write down its sample space.

Solution:

The possible outcomes in the first draw is that the marble drawn is of red colour or blue colour or white colour.

Since, we are drawing the marble with replacement, so for the second draw the possible outcomes are the same as first draw.

Let us denote, Red: R , Blue: B and White: W

Hence, the sample space; $S = \{RR, RB, RW, BR, BW, BB, WR, WB, WW\}$

- Q3. In an urn of marbles, there are two red coloured marbles, and 1 each of blue and white colour. If the experiment is drawing two marbles from the urn without replacement, then write down its sample space.

Solution:

The possible outcomes in the first draw is that the marble drawn is of red colour or blue colour or white colour.

Since, we are drawing the marble without replacement, so for the second draw the possible outcomes depends on the first draw.

Let us denote, Red: R , Blue: B and White: W

Hence, the sample space; $S = \{RR, RB, RW, BR, BW, WR, WB\}$

- Q4. If the experiment is to draw a card from a well shuffled pack of 52 cards, then write down its sample space.

Solution:

The possible outcomes are that either the numbered card or the face card is drawn.

In addition, it could belong to any of the four suits, i.e. Hearts, Diamonds, Clubs and Spades.

So, the sample space can conveniently be written as the Cartesian product of these two sets, i.e. Set of suits and Set of values.

Hence, $S = \{\text{Hearts, Diamonds, Clubs, Spades}\} \times \{2, 3, 4, 5, 6, 7, 8, 9, 10, J, K, Q, A\}$ which is the same as the set of 52 cards in the shuffled pack.

Remark:

In the above example, if we were to draw 13 cards from the well shuffled pack of 52 cards, then the number of possible outcomes are ${}^{52}C_{13}$.

And, writing down the sample space for this experiment is a very tedious task. Hence, it is not useful to write down sample space in such cases, but to imagine it instead.

1.3 Events

An event is a subset of the sample space. There is a technical restriction on what subsets can be events.

(More details on technical restrictions in section 2.8.1)

Note:

- Events are central objects in probability theory.
- All set theory notions apply to events, and that tends to have a natural meaning.
- If the number of elements in sample space = N
Then, the maximum number of events (i.e. number of possible subsets) = 2^N

Examples

- (i) Tossing a coin: $S = \{\text{Heads, Tails}\}$.

- Number of events = $2^2 = 4$

- Events: empty set, {Heads}, {Tails}, {Heads, Tails}

(ii) Rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$.

- Number of events $= 2^6 = 64$
- Events: empty set, {1}, {2}, {3}, {4}, {5}, {6}, {1,2}, {1,3}, ..., {1, 2, 3, 4, 5, 6}
- Some events can also be described in words: Getting an even number i.e. {2,4,6}, Getting a multiple of 3 i.e. {3,6}, etc.

(iii) Fisherman goes out to fish:

- Sample space is complicated as we have to describe everything like catch of fish in kilos, type of fish and so on.
- We can still define events of interest quite easily, even on the complicated sample space.
Events: Catch is more than 100 kg, Pomfret is the catch, etc.

1.3.1 Occurrence of events

An event is said to have “occurred” if the actual outcome of the experiment belongs to the event.

1.3.1.1 One event can be contained in another:

If the set A is a subset of the set B i.e. $A \subseteq B$ implies that:

- If A occurred, B has also occurred.
- If B occurred, A may or may not have occurred.

Examples

Rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$.

(i) $A = \{2, 6\}$ and $B = \{\text{even number}\} \implies A \subseteq B$

(ii) $A = \{1, 3, 5\}$ and $B = \{\text{prime number} > 2\} \implies B \subseteq A$

1.3.1.2 Complement of an event:

The complement of an event A is denoted by A^c and is defined as:
 $A^c = \{\text{outcomes in } S \text{ and not in } A\} = (S \setminus A)$ which implies that:

- If A occurred, A^c did not occur.
- If A^c occurred, A did not occur.

Examples

Rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$.

(i) $A = \{2, 4, 6\} \implies A^c = \{1, 3, 5\}$

(ii) $A = \{\text{not a prime number}\} \implies A^c = \{\text{prime number}\}$

1.3.2 Combining events

We can combine events to create new events, and the standard way to operate on is by using the two operations on the events.

These set operations are very central in set theory, namely, ‘Unions’ and ‘Intersections’.

1.3.2.1 Union of events:

The union of two or more than two events is the collection of all the outcomes that belong to either of the events. It is denoted by ‘ \cup ’.

Note:

- In English terminology, it is referred to as “OR”.
- The union of events is said to be occurred if either of the events have occurred.

Examples

- (i) Rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$.
 $A = \{\text{even number}\} = \{2, 4, 6\}$ and $B = \{\text{multiple of 3}\} = \{3, 6\}$
Implies that, $A \cup B = \{\text{either an even number or a multiple of 3}\} = \{2, 3, 4, 6\}$
- (ii) Fisherman’s Catch: $A = \{\text{more than 200 Kg}\}$ and $B = \{\text{less than 50 Kg}\}$
Implies that, $A \cup B = \{\text{either more than 200 Kg or less than 50 Kg}\}$

1.3.2.2 Intersection of events:

The intersection of two or more than two events is the collection of all the outcomes that are common to all the events. It is denoted by ‘ \cap ’.

Note:

- In English terminology, it is referred to as “AND”.
- The intersection of events is said to be occurred if all the events have occurred.

Examples

- (i) Rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$.
 $A = \{\text{even number}\} = \{2, 4, 6\}$ and $B = \{\text{multiple of 3}\} = \{3, 6\}$
Implies that, $A \cap B = \{\text{even number and multiple of 3}\} = \{6\}$
- (ii) Fisherman’s Catch: $A = \{\text{more than 100 Kg}\}$ and $B = \{\text{less than 150 Kg}\}$
Implies that, $A \cap B = \{\text{more than 100 Kg and less than 150 Kg}\}$

1.3.2.3 Solved examples:

Q1. For an experiment of rolling a die, let us define the events as follows:

$A = \{\text{even number}\}$ and $B = \{\text{prime number}\}$

(i) Find $A \cup B$.

(ii) Find $A \cap B$.

Solution:

$A = \{\text{even number}\} = \{2, 4, 6\}$

$B = \{\text{prime number}\} = \{2, 3, 5\}$

(i) $A \cup B = \{\text{either an even number or a prime number}\} = \{2, 3, 4, 5, 6\}$

(ii) $A \cap B = \{\text{even number and prime number}\} = \{2\}$

Q2. 5 cards are to be drawn without replacement from a well shuffled pack of 52 cards. Can we represent the event that there are no aces as an intersection of 5 events?

Solution:

Yes. Let us define the events as follows:

$E = \{\text{no aces in 5 draws of cards}\}$

$E_1 = \{\text{no ace in the first draw of card}\}$

$E_2 = \{\text{no ace in the second draw of card}\}$

$E_3 = \{\text{no ace in the third draw of card}\}$

$E_4 = \{\text{no ace in the fourth draw of card}\}$

$E_5 = \{\text{no ace in the fifth draw of card}\}$

Then, we can express the event E as $(E_1 \cap E_2 \cap E_3 \cap E_4 \cap E_5)$

Q3. In an IPL experiment, consider an event $E = \{5 \text{ runs being scored in two legal deliveries}\}$. Can we represent this event as the union of events?

Solution:

Yes. The possible number of runs that can be scored off bat on any legal delivery is 0, 1, 2, 3, 4 and 6. (Assuming there are no extras and no overthrows)

Let us define the events as follows:

$E_1 = \{1 \text{ run off the first delivery and 4 runs off the second delivery}\} = \{1+4\}$

$E_2 = \{2 \text{ runs off the first delivery and 3 runs off the second delivery}\} = \{2+3\}$

$E_3 = \{3 \text{ runs off the first delivery and 2 runs off the second delivery}\} = \{3+2\}$

$E_4 = \{4 \text{ runs off the first delivery and 1 run off the second delivery}\} = \{4+1\}$

Then, we can express the event E as $(E_1 \cup E_2 \cup E_3 \cup E_4)$

1.3.3 Disjoint events

Two or more than two events are said to be disjoint events if they have an empty intersection. In general, the events E_1, E_2, E_3, \dots are disjoint if:

$E_i \cap E_j = \text{empty set}$; for any $i \neq j$

If the events A and B are disjoint, then it implies that:

- If A occurred, B did not occur.
- If B occurred, A did not occur.
- A is subset of B^c and B is subset of A^c .

Examples

(i) Rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$.

$A = \{\text{even number}\} = \{2, 4, 6\}$ and $B = \{\text{odd number}\} = \{1, 3, 5\}$

A and B are disjoint events because, $A \cap B = \text{empty set}$

(ii) Fisherman's Catch: $A = \{\text{more than 200 Kg}\}$ and $B = \{\text{less than 50 Kg}\}$

A and B are disjoint events because, $A \cap B = \text{empty set}$

1.3.3.1 Partition

If two or more disjoint events make up the whole sample space together, then it is referred to as the Partition.

We can make a partition of large sample spaces into multiple disjoint events for study.

Note:

The event and its complement are referred to as a "Partition" because:

- The event and its complement are always disjoint events, i.e. $A \cap A^c = \text{empty set}$
- The event and its complement together cover all the outcomes, i.e. $A \cup A^c = S$

Examples

For an experiment of drawing a card from a pack of 52 well shuffled cards, consider the events:

$E_1 = \{\text{The card is a Spade}\}$, $E_2 = \{\text{The card is a Heart}\}$, $E_3 = \{\text{The card is a Club}\}$

and $E_4 = \{\text{The card is a Diamond}\}$

Then, the events E_1 , E_2 , E_3 and E_4 are referred to as the Partition.

1.3.4 De Morgan's Law

De Morgan's Laws are defined as:

- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$

These laws are very useful and will come to a rescue when we want to interpret an event given in English.

1.3.5 Solved Examples:

Going through these examples will enhance your skill of translating English to events.

- Q1. The hats of 5 persons are identical and get mixed up in a box. If each person picks a hat at random from the box, then consider the following events:

$A = \{\text{No person gets their own hat}\}$

$B = \{\text{Every person gets their own hat}\}$

$C = \{\text{At least one person does not get their own hat}\}$

$D = \{\text{At least one person gets their own hat}\}$

Based on the given information, answer the following questions.

- (i) What is A^c ?
- (ii) What is B^c ?
- (iii) Are A and B disjoint events?
- (iv) What is $A \cap B^c$?

Solution:

- (i) As we know that, $A^c = \{\text{outcomes in } S \text{ and not in } A\} = (S \setminus A)$

So, even if any one of the 5 persons get their own hat, that outcome will not belong to the event A but will belong to S .

Hence, $A^c = \{\text{At least one person gets their own hat}\} = D$

(**Note:** Don't take the A^c as "Everyone gets their own hat", it's not an English comprehension. That's where writing down complete sample space is a complicated task, but having an idea of sample space is useful.)

- (ii) Again, as we know that, $B^c = \{\text{outcomes in } S \text{ and not in } B\} = (S \setminus B)$

So, even if any one of the 5 persons does not get their own hat, that outcome will not belong to the event B but will belong to S .

Hence, $B^c = \{\text{At least one person does not get their own hat}\} = C$

- (iii) Since, $A \cap B = \text{empty set} \implies A$ and B are disjoint events.

i.e. the occurrence of these events that no person gets their own hat and every person gets their own hat is not possible together.

- (iv) $A \cap B^c \implies$ All the outcomes that are in the event A but not in B .

Now,

$$A \cap B^c = A \cap C ; \quad (\text{because, } B^c = C)$$

$$= \{\text{No person gets their own hat and Atleast one person does not get their own hat}\}$$

$$= \{\text{No person gets their own hat}\}$$

$$\therefore A \cap B^c = A$$

Another way to approach:

Since A and B are disjoint events $\implies A \subseteq B^c$

Hence, $A \cap B^c = A$

Q2. For the IPL experiment, in one over 6 deliveries are bowled where, in each delivery 0, 1, 2, 3, 4 or 6 runs may be scored. Consider the following events:

$A = \{\text{No 4s in one over}\}$

$B = \{\text{No 6s in one over}\}$

$C = \{\text{Exactly 20 runs scored in one over}\}$

Based on the given information, answer the following questions.

- (i) What is $A \cup B$?
- (ii) What is $A \cap B$?
- (iii) Can $A \cap B \cap C$ occur?

Solution:

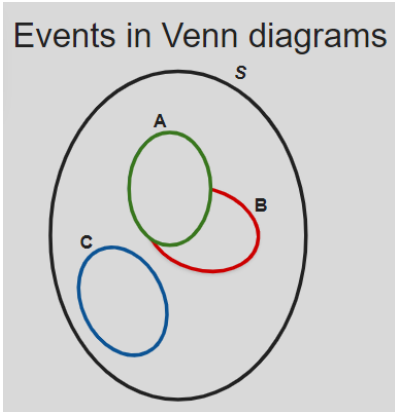
- (i) $A \cup B = \{\text{Either no 4s in one over or no 6s in one over}\}$
i.e. either of the events A or B occur.
Remember that the outcome $\{6,6,6,6,6,6\}$ also belong to $A \cup B$ because event A has occurred even if B didn't occur.
- (ii) $A \cap B = \{\text{No 4s and no 6s in one over}\}$
i.e. both the events A and B must occur.
- (iii) For $A \cap B \cap C$ to occur, one should be able to score 20 runs with no 4s and no 6s. But, the maximum runs scored in one over with no 4s and no 6s is only 18 (i.e. $\{3,3,3,3,3,3\}$).
Hence, $A \cap B \cap C = \text{empty set} \implies$ It can not occur.

1.4 Venn Diagram

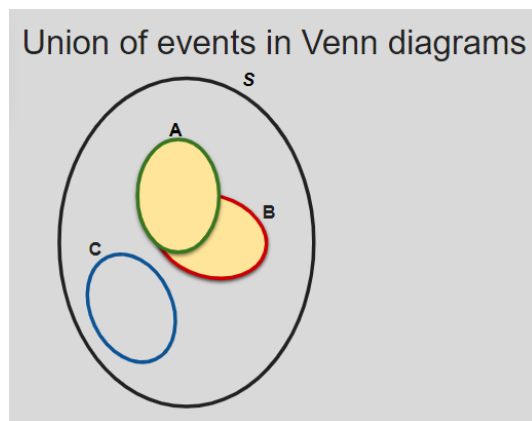
Venn diagram is a very useful visualization for sets.

We consider sample space to be one big ellipse or circle or rectangle, while every subset is represented as a smaller circle inside it.

Let us consider three events, i.e. A , B and C and represent it in Venn diagram as follows:

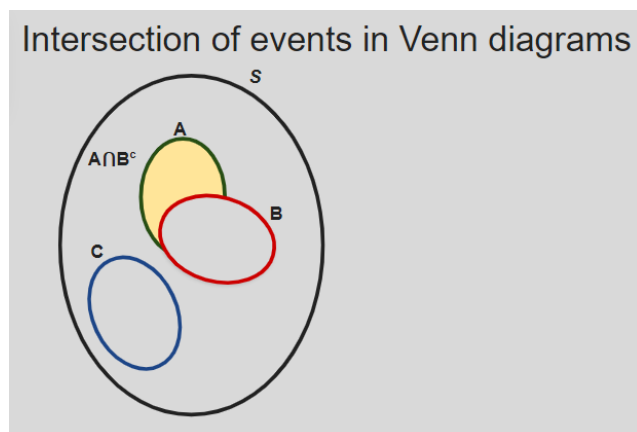


- Union of events:
All covered regions.



The shaded region in the above Venn diagram represents $A \cup B$.

- Intersection of events:
Common region.

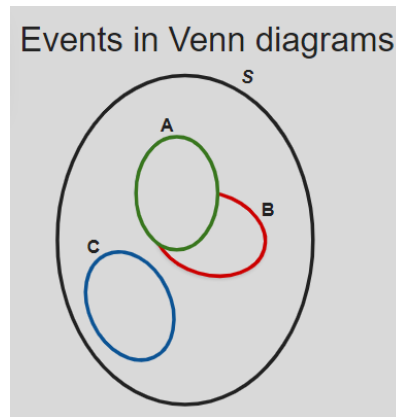


The shaded region in the above Venn diagram represents $A \cap B^c$, i.e. Region of A outside B .

Example:

Rolling a die: Even number, but not a multiple of 3.

- Disjoint events:
No overlap in regions.



In the above Venn diagram, events A and C and events B and C are disjoint events.

Similarly, we can do more complicated operations and will get different looking results.

One can also verify the De Morgan's law using the Venn diagram. **(Do give it a try!)**

2 Introduction to Probability

2.1 Events and Chance

Now that we have defined the sample space, outcomes and events, we are interested in whether the event will occur or not. If yes, then what are the chances of it occurring?

For instance,

- Tossing a coin: What are the chances of getting a head?
- 13 cards drawn randomly from a well shuffled pack of 52 cards: What are the chances of getting all 13 cards of the same suit?
- Football world cup: What are the chances of India winning?
- Cricket world cup: What are the chances of India winning?

Though intuitively, we can say that the chances of occurring of an event is low or high, but still we want to assign the proper values of chance to the events in an experiment.

2.2 Mathematical and Probability theory

The probability theory let us work with numbers, which represent the chance and to associate a mathematical value of chance to the occurrence of an event in the experiment.

The probability theory is a mathematically developed theory.

2.2.1 Mathematical Theory

Let us first discuss the ingredient of mathematical theory as defined:

- Define the basic objects of interest precisely, and this will come from other prior fields. For instance, in probability it comes primarily from the set theory as discussed in the above sections that sample space is a set and events are its subset and so on.
- A few assumptions about the object we have or define the conditions which are always valid i.e., axioms.
- Deduce everything else with logical proof.

2.2.2 Probability Theory

The three important objects in probability theory (together make up the Probability Space) are:

- Sample space
- Events
- Probability function.

We have already discussed sample space and events in detail in the earlier sections. Let us now discuss the probability function in detail.

2.2.2.1 Probability

Probability is a function P that assigns to each event a real number between 0 and 1. The entire probability space (sample space, events and probability function) should satisfy the following two axioms:

(i) $P(S) = 1$ (Probability of entire sample space equals 1)

(ii) If E_1, E_2, E_3, \dots are disjoint events then,

$$P(E_1 \cup E_2 \cup E_3 \cup \dots) = P(E_1) + P(E_2) + P(E_3) + \dots$$

It implies that the function take subsets of sample space i.e. events, as input and gives a real number between 0 and 1 as output (referred to as probability), which is a representative of chance of occurrence of that event.

Note:

- Higher the output value means higher the chance of occurring of that event.
- 0 means the event cannot occur, and 1 means it will always occur.
- Often, probability can be mentioned in percentages.
As an example, if probability of an event occurrence is 0.9 then, we can say there is a 90% chance of it occurring.

Examples

Tossing a coin: $S = \{H, T\}$.

Which of the following probability function(s) satisfy the axioms?

- (i) $P(\text{empty}) = 0$, $P(\{H\}) = 0.5$, $P(\{T\}) = 0.5$ and $P(\{H, T\}) = 1$
- (ii) For $0 \leq p \leq 1$:
 $P(\text{empty}) = 0$, $P(\{H\}) = p$, $P(\{T\}) = 1 - p$ and $P(\{H, T\}) = 1$
- (iii) $P(\text{empty}) = 0$, $P(\{H\}) = 0.5$, $P(\{T\}) = 0.6$ and $P(\{H, T\}) = 1$

Solution:

- (i)
 - $P(S) = P(\{H, T\}) = 1$
 - The events $\{H\}$ and $\{T\}$ are disjoint and,
 $P(\{H\} \cup \{T\}) = P(\{H, T\}) = P(S) = 1$ which is the same as,
 $P(\{H\}) + P(\{T\}) = 0.5 + 0.5 = 1$

Hence, the given probability function satisfy the axioms.

- (ii)
 - $P(S) = P(\{H, T\}) = 1$
 - The events $\{H\}$ and $\{T\}$ are disjoint and,
 $P(\{H\} \cup \{T\}) = P(\{H, T\}) = P(S) = 1$ which is the same as,
 $P(\{H\}) + P(\{T\}) = p + (1 - p) = 1$

Hence, the given probability function satisfy the axioms.

- (iii)
 - $P(S) = P(\{H, T\}) = 1$
 - The events $\{H\}$ and $\{T\}$ are disjoint and,
 $P(\{H\} \cup \{T\}) = P(\{H, T\}) = P(S) = 1$ which is not the same as,
 $P(\{H\}) + P(\{T\}) = 0.5 + 0.6 = 1.1$

Hence, the given probability function does not satisfy the axioms.

2.3 Basic properties of probability

Property 1: Empty Set

The probability of an empty set (denoted by Φ) equals 0.
i.e., $P(\Phi) = 0$

Proof:

We know that an event and its complement are always disjoint.
Since $\Phi^c = S$, S and Φ are disjoint and $\Phi \cup S = S$
By using Axiom (2),

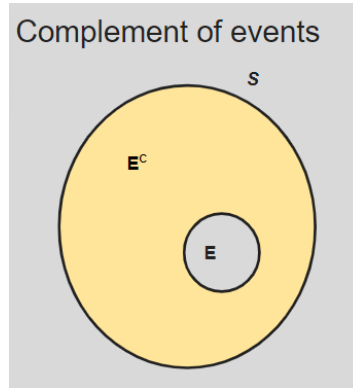
$$\begin{aligned} P(S \cup \Phi) &= P(S) + P(\Phi) \\ \implies P(S) &= P(S) + P(\Phi) \\ \therefore P(\Phi) &= 0 \end{aligned}$$

Property 2: Complement

Let E^c be the complement of an event E . Then,

$$P(E^c) = 1 - P(E)$$

Proof:



We know that an event and its complement are always disjoint. So, E and E^c are disjoint
and $E \cup E^c = S$
By using Axiom (2),

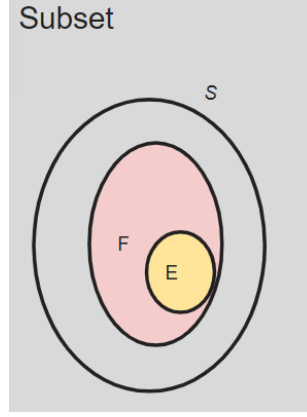
$$\begin{aligned} P(E \cup E^c) &= P(E) + P(E^c) \\ P(S) &= P(E) + P(E^c) \\ \implies 1 &= P(E) + P(E^c) \quad ; \quad (\text{By Axiom 1, } P(S) = 1) \\ \therefore P(E^c) &= 1 - P(E) \end{aligned}$$

Property 3: Subset

If the event E is a subset of the event F , i.e., $E \subseteq F$, then

$$P(F) = P(E) + P(F \setminus E) \implies P(E) \leq P(F)$$

Proof:



$F \setminus E = F \cap E^c$ (Outside E and inside F)

Also, E and $F \setminus E$ are disjoint and $E \cup (F \setminus E) = F$.

By using Axiom (2),

$$\begin{aligned} P(E \cup (F \setminus E)) &= P(E) + P(F \setminus E) \\ \implies P(F) &= P(E) + P(F \setminus E) \end{aligned}$$

Since probability of an event cannot be negative i.e., $P(F \setminus E) \geq 0$; (Always)

$$\begin{aligned} P(F) &= P(E) + (\text{something non-negative}) \\ \therefore P(F) &\geq P(E) \end{aligned}$$

Property 3a: Difference and Intersection

If E and F are events, then

$$\begin{aligned} P(E) &= P(E \cap F) + P(E \setminus F) \\ P(F) &= P(E \cap F) + P(F \setminus E) \end{aligned}$$

Proof:

Since $E \cap F$ is a subset of E , using the subset property

$$P(E) = P(E \cap F) + P(E \setminus (E \cap F))$$

Now, $E \setminus (E \cap F) = E \setminus F$

Hence, $P(E) = P(E \cap F) + P(E \setminus F)$

Similarly,

$$P(F) = P(E \cap F) + P(F \setminus E)$$

Property 4: Union and Intersection

If E and F are events, then

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Proof:

The events $(E \setminus F)$, $(E \cap F)$ and $(F \setminus E)$ are disjoint.

Also, $E \cup F = (E \setminus F) \cup (E \cap F) \cup (F \setminus E)$

By using Axiom (2),

$$P(E \cup F) = P(E \setminus F) + P(E \cap F) + P(F \setminus E)$$

(As we know that for any events A and B , $P(A \setminus B) = P(A) - P(A \cap B)$)

$$\implies P(E \cup F) = [P(E) - P(E \cap F)] + P(E \cap F) + [P(F) - P(E \cap F)]$$

$$\therefore P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

If we consider the events E and F as disjoint, then $P(E \cap F) = 0$. Plugging in, we will get the Axiom (2).

Hence, this property is a generalization of the Axiom (2).

2.3.1 Working with probability spaces

Q1. A waiter and a cashier are to be hired. There are 4 applicants - David and Megha from Delhi, and Rajesh and Veronica from Mumbai. The restaurant hires one person at random as waiter, and another from the remaining as cashier. Write out the following:

- (i) Sample space.
- (ii) Event A : Cashier is from Delhi
- (iii) Event B : Exactly one position is filled by a Delhiite
- (iv) Event C : Neither position is filled by a Delhiite

Solution:

Let us denote, David: D , Megha: M , Rajesh: R and Veronica: V .

- (i) The outcome of this experiment is of the form {Waiter name, Cashier name}
Therefore, the sample space is given by:

$$S = \{(D, M), (D, R), (D, V), (M, D), (M, R), (M, V), \\ (R, D), (R, M), (R, V), (V, D), (V, M), (V, R)\}$$

- (ii) If cashier is from Delhi, then the cashier position can be filled either by David or by Megha.

Therefore, the event A is given by:

$$A = \{(D, M), (M, D), (R, D), (R, M), (V, D), (V, M)\}$$

- (iii) If exactly one position is to be filled by a Delhiite, then either the cashier position or the waiter position, but not both, can be filled either by David or by Megha. Therefore, the event B is given by:

$$B = \{(D, R), (D, V), (M, R), (M, V), (R, D), (V, D), (R, M), (V, M)\}$$

- (iv) Neither position is to be filled by a Delhiite, means that none of the position is to be filled by David and Megha. Therefore, the event C is given by:

$$C = \{(R, V), (V, R)\}$$

Q2. In a town, there are fishing boats that go out to catch fish everyday. Over the years, folks have observed the following:

- Chances of catching more than 400 kg of fish in a day is 35%.
- Chances of catching more than 500 kg of fish in a day is 10%.

What are the chances of catching between 400 and 500 kg of fish in a day?

Solution:

Let us define the events as follows:

A : Catching more than 400 kg of fish in a day.

B : Catching more than 500 kg of fish in a day.

So, $P(A) = 0.35$ and $P(B) = 0.10$

Also, $B \subseteq A$, because all the outcomes of B will be contained in A as well.

Now, the event $E = \{\text{The catch of fish in a day is between 400 and 500}kg\} = A \setminus B$ i.e., the outcome should belong to the event A but not the event B .

In other words, the catch of fish in a day should be more than 400 kg but not more than 500 kg . Therefore,

$$P(E) = P(A \setminus B)$$

(By subset property, if $B \subseteq A$ then, $P(A) = P(B) + P(A \setminus B)$)

$$\implies P(E) = P(A) - P(B)$$

$$= 0.35 - 0.10$$

$$\therefore P(E) = 0.25$$

Q3. Suppose you hear the following forecast for rain and temperature:

- Chance of rain tomorrow is 60%.

- Chance of maximum temperature above $30^{\circ}C$ tomorrow is 70%.
- Chance of rain and maximum temperature above $30^{\circ}C$ tomorrow is 40%.

What are the chances of no rain and maximum temperature below $30^{\circ}C$ tomorrow?

Solution:

Let us define the events as follows:

A : There will be rain ; A^c : There will be no rain
 B : The maximum temperature is above $30^{\circ}C$; B^c : The maximum temperature is below $30^{\circ}C$

So, $P(A) = 0.60$, $P(B) = 0.70$ and $P(A \cap B) = 0.40$

Now, the event

$E = \{\text{There will be no rain and maximum temperature below } 30^{\circ}C\} = A^c \cap B^c$

Therefore,

$$\begin{aligned}
 P(E) &= P(A^c \cap B^c) \\
 &= P(A \cup B)^c \quad ; \quad (\text{Using De Morgan's Law}) \\
 &= 1 - P(A \cup B) \\
 \implies P(E) &= 1 - [P(A) + P(B) - P(A \cap B)] \quad ; \quad (\text{Using Property 4: Union and intersection}) \\
 &= 1 - [0.60 + 0.70 - 0.40] \\
 \therefore P(E) &= 0.10
 \end{aligned}$$

2.4 Distributions

The idea of distributions is to assign probabilities to each of the individual outcomes in the sample space.

It gives us the sense of how the probabilities are distributed over the outcomes.

It is possible when:

- When the outcomes can be enumerated as first, second, third and so on.
- The sample space is countable or finite sample space.

Example:

Rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$

Let us suppose that all the individual outcomes are events and,

$P(\{1\}) = p_1$, $P(\{2\}) = p_2$, $P(\{3\}) = p_3$, $P(\{4\}) = p_4$, $P(\{5\}) = p_5$ and, $P(\{6\}) = p_6$

where, $0 \leq p_i \leq 1$; $i = 1, 2, 3, 4, 5, 6$

Also, the events $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$ and $\{6\}$ are disjoint events.

And, $\{1\} \cup \{2\} \cup \{3\} \cup \{4\} \cup \{5\} \cup \{6\} = S$.

Hence, by Axiom (2), $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = P(S) = 1$

Therefore, the notion of distribution for this case is that:

Each of the p_i ; $i = 1, 2, 3, 4, 5, 6$, have to be between 0 and 1, and they have to add up to 1.

We can also find the probability of any complicated event by splitting it as union of individual outcomes.

$$\begin{aligned} P(\{1, 3, 5\}) &= P(\{1\} \cup \{3\} \cup \{5\}) \\ &= P(\{1\}) + P(\{3\}) + P(\{5\}) \quad ; \quad (\text{because, disjoint events}) \\ &= p_1 + p_3 + p_5 \end{aligned}$$

Note:

If the die is fair, then $p_i = \frac{1}{6}$; $i = 1, 2, 3, 4, 5, 6$

i.e. each individual outcome has the same probability. (Generally, referred to as equally-likely outcomes)

2.4.1 Uniform Distribution on a finite sample space

$S = \{\text{finite number of outcomes}\}$ and the outcomes are equally likely.

i.e. each individual outcome, in sample space S , has the same probability of occurrence.

$$\text{Hence, probability of an outcome} = \frac{1}{\text{Number of outcomes in } S} \implies P(\text{an outcome}) = \frac{1}{|S|}$$

And,

$$P(\text{event}) = \frac{\text{Number of outcomes in the event}}{\text{Number of outcomes in } S}$$

2.4.2 Solved Examples:

- Q1. There are 5 red and 8 blue marbles in an urn. A marble is drawn from the urn at random. What is the probability that a red marble is drawn?

Solution:

The sample space; $S = \{R_1, R_2, R_3, R_4, R_5, B_1, B_2, B_3, B_4, B_5, B_6, B_7, B_8\}$

Since, the marble is drawn at random, the distribution is uniform.

Let us define the event,

$$A = \{\text{The marble drawn is red}\} = \{R_1, R_2, R_3, R_4, R_5\}$$

Therefore,

$$\begin{aligned} P(A) &= \frac{\text{Number of outcomes in the event } A}{\text{Number of outcomes in } S} \quad ; \quad (\text{because, Uniform distribution}) \\ \therefore P(A) &= \frac{5}{13} \end{aligned}$$

- Q2. In a throw of two fair dice, what is the probability that the sum of the two numbers is 8?

Solution:

The sample space consists of 36 outcomes given by:

$$S = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), \\ (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), \\ (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), \\ (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), \\ (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), \\ (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$$

Since, the fair dice are rolled, the distribution is assumed to be uniform.

Let us define the event,

$$A = \{\text{The sum of the two numbers is 8}\} = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$$

Therefore,

$$P(A) = \frac{\text{Number of outcomes in the event A}}{\text{Number of outcomes in S}} \quad ; \quad (\text{because, Uniform distribution}) \\ \therefore P(A) = \frac{5}{36}$$

- Q3. The hats of 3 persons are identical and get mixed up. Each person picks a hat at random. What is the probability that none of the persons gets their own hat?

Solution:

Let the three persons be P_1, P_2 and P_3 , and their hats are H_1, H_2 and H_3 respectively. Then the possible outcomes are listed in a table below, where each column represents a person and each row is the hat picked by them. The sample space is as follows:

	P_1	P_2	P_3
1	H_1	H_2	H_3
2	H_1	H_3	H_2
3	H_2	H_1	H_3
4	H_2	H_3	H_1
5	H_3	H_1	H_2
6	H_3	H_2	H_1

Since, each person picks a hat at random, the distribution is uniform.

Let us define the event,

$$A = \{\text{None of the person gets their own hat}\} = \{(H_2, H_3, H_1), (H_3, H_1, H_2)\}$$

Therefore,

$$P(A) = \frac{\text{Number of outcomes in the event A}}{\text{Number of outcomes in S}} \quad ; \quad (\text{because, Uniform distribution}) \\ \therefore P(A) = \frac{2}{6}$$

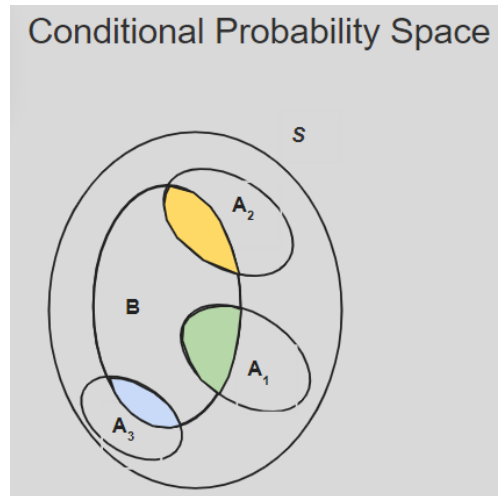
2.5 Conditional Probability

The basic idea is that, we can split the initial probability space into an event that has occurred and some conditional probability space given that event has occurred.

2.5.1 Conditional Probability Space

Consider a probability space:

Sample space S , Collection of events and Probability function P .



Let B be an event with $P(B) > 0$, then the Conditional Probability Space given B is:

- Sample space : B , because the event B has already occurred, and we have moved into this sample space.
- Events : $A \cap B$, for every event A in the original space.
- Probability function : It is given by, $P(A|B) = \frac{P(A \cap B)}{P(B)}$. It is also called conditional probability of A given B or probability of A given B .

Remember that, the new probability function also satisfy both the axioms:

- 1) The probability of entire sample space equals 1, i.e.

$$P(B|B) = \frac{P(B \cap B)}{P(B)} = 1$$

- 2) If A_1, A_2, A_3, \dots are disjoint events then,

$$P(A_1 \cup A_2 \cup A_3 \cup \dots | B) = P(A_1|B) + P(A_2|B) + P(A_3|B) + \dots$$

Note:

- $P(B) > 0$ because, if $P(B) = 0$ then, it has no chance of occurring, and can't be observed.
- One can also check that all the basic properties of probability are satisfied by the probability function $P(A|B)$.

Example:

Toss a coin three times:

Initial sample space ; $S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$

Let us suppose, an event $B = \{\text{First toss result in tails}\}$ has occurred.

Therefore, the reduced probability space to work on for the second and third toss is:

- Sample space ; $B = \{THH, TTH, THT, TTT\}$.
- Events and probability function can be redefined to account for occurrence of B .

2.5.2 Multiplication rule:

In some cases, it is easier to work with conditional probability space than the original probability space. So, we can compute probabilities like:

$$\text{For any event A in original space ; } P(A \cap B) = P(B)P(A|B)$$

This formula is often referred to as the multiplication rule of probability.

It can be generalized to n events by keep on iterating the multiplication rule: **(Do try!)**

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

2.5.3 Solved Examples:

Q1. A fair die is rolled. Consider an event $E = \{2, 4, 6\}$, then compute the following probabilities:

- (i) $P(\{2\}|E)$
- (ii) $P(\{1\}|E)$
- (iii) $P(\{2, 3, 4\}|E)$

Solution:

Sample space: $S = \{1, 2, 3, 4, 5, 6\}$

Since, the fair die is rolled, the distribution is assumed to be uniform, i.e.

$$P(\{i\}) = \frac{1}{6} \quad ; \quad i = 1, 2, 3, 4, 5, 6$$

The event $E = \{2, 4, 6\}$. Therefore,

$$P(E) = \frac{\text{Number of outcomes in the event}}{\text{Number of outcomes in S}} \quad ; \quad (\text{because, Uniform distribution})$$

$$\therefore P(E) = \frac{3}{6}$$

$$\begin{aligned} \text{(i)} \quad P(\{2\}|E) &= \frac{P(\{2\} \cap E)}{P(E)} \\ &= \frac{P(\{2\})}{P(E)} \end{aligned}$$

$$\therefore P(\{2\}|E) = \frac{1/6}{3/6} = \frac{1}{3}$$

$$\begin{aligned} \text{(ii)} \quad P(\{1\}|E) &= \frac{P(\{1\} \cap E)}{P(E)} \\ &= \frac{P(\{\Phi\})}{P(E)} \end{aligned}$$

$$\therefore P(\{1\}|E) = \frac{0}{3/6} = 0$$

$$\begin{aligned} \text{(iii)} \quad P(\{2, 3, 4\}|E) &= \frac{P(\{2, 3, 4\} \cap E)}{P(E)} \\ &= \frac{P(\{2, 4\})}{P(E)} \end{aligned}$$

$$\therefore P(\{2, 3, 4\}|E) = \frac{2/6}{3/6} = \frac{2}{3}$$

Q2. There are two urns 'A' and 'B'. Urn 'A' contains 7 red marbles and 6 blue marbles while, Urn 'B' contains 5 red marbles and 8 blue marbles. If an urn is chosen at random and then the marble is drawn at random from the chosen urn. Based on this information, compute the following probabilities:

- (i) The red ball is drawn given that Urn 'A' is chosen.
- (ii) The blue ball is drawn given that Urn 'B' is chosen.

Solution:

- (i) It is given that the Urn 'A' is chosen. Hence, we have the conditional probability space, where the sample space is given by:

$$S = \{R_1, R_2, R_3, R_4, R_5, R_6, R_7, B_1, B_2, B_3, B_4, B_5, B_6\}$$

Since, we are drawing the marble at random, the distribution is uniform.

$$\therefore P(\text{Red Marble} \mid \text{Urn A}) = \frac{7}{13}$$

- (ii) It is given that the Urn 'B' is chosen. Hence, we have the conditional probability space, where the sample space is given by:

$$S = \{R_1, R_2, R_3, R_4, R_5, B_1, B_2, B_3, B_4, B_5, B_6, B_7, B_8\}$$

Since, we are drawing the marble at random, the distribution is uniform.

$$\therefore P(\text{Blue Marble} \mid \text{Urn } B) = \frac{8}{13}$$

- Q3. Consider a class of 15 students: 4 from State-1, 8 from State-2 and 3 from State-3. Three different students are chosen at random one after another. What is the probability that the selected three are from State-1, State-3 and State-1 again in that order?

Solution:

Let us define the following events:

A_1 : The first student is from State-1

A_2 : The second student is from State-3

A_3 : The third student is from State-1

Since, students are chosen at random. Therefore,

$$P(A_1) = \frac{4}{15}$$

$$P(A_2|A_1) = \frac{3}{14} \quad (\text{Because, it is given that one student of State-1 is already selected})$$

$$\text{And, } P(A_3|A_1 \cap A_2) = \frac{3}{13}$$

Because, it is given that the one from each State-1 and 3 is already selected.

The probability that the selected three are from State-1, State-3 and State-1 again in that order is $P(A_1 \cap A_2 \cap A_3)$.

Now,

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(B \cap A) && (\text{Consider, } A_1 = B \text{ and, } A_2 \cap A_3 = A) \\ &= P(B)P(A|B) && (\text{Using multiplication rule}) \\ &= P(A_1)P(A_2 \cap A_3|A_1) \\ &= \frac{4}{15} \times P(A_2 \cap A_3|A_1) && (\text{Because, students are chosen at random}) \\ &= \frac{4}{15} \times \{P(A_2|A_1)P(A_3|A_1 \cap A_2)\} && (\text{Again, using multiplication rule}) \\ &= \frac{4}{15} \times \frac{3}{14} \times \frac{3}{13} \\ \therefore P(A_1 \cap A_2 \cap A_3) &= \frac{6}{455} \end{aligned}$$

- Q4. A family has 2 children. What is the probability that both are girls, given that at least one is a girl?

Solution:

Let us denote Boy: B and Girl: G

Then the sample space: $S = \{BB, BG, GB, GG\}$ and, the distribution is assumed to be uniform.

Let us define the following events:

$E = \{\text{At least one is a girl}\} = \{GB, BG, GG\}$

$F = \{\text{Both are girls}\} = \{GG\}$

Therefore, the probability that both are girls, given that at least one is a girl is:

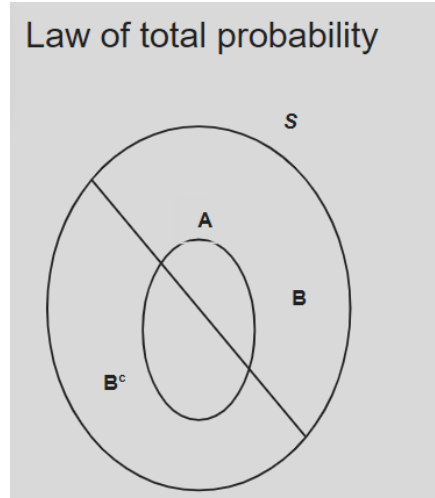
$$\begin{aligned}
 P(F|E) &= \frac{P(F \cap E)}{P(E)} \\
 &= \frac{P(\{GG\} \cap \{GB, BG, GG\})}{P(\{GB, BG, GG\})} \\
 &= \frac{P(\{GG\})}{P(\{GB, BG, GG\})} \\
 &= \frac{1/4}{3/4} \quad ; \quad (\text{Because, Uniform ditribution}) \\
 \therefore P(F|E) &= \frac{1}{3}
 \end{aligned}$$

2.6 Bayes' Theorem and Independence

2.6.1 Law of Total Probability

Consider a probability space:

Sample space S , Collection of events and Probability function P .



Let us consider the events A and B such that event B partitions the sample space into two halves i.e. B and B^c , and A is an event of interest.

Then, the Law of Total Probability states that:

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

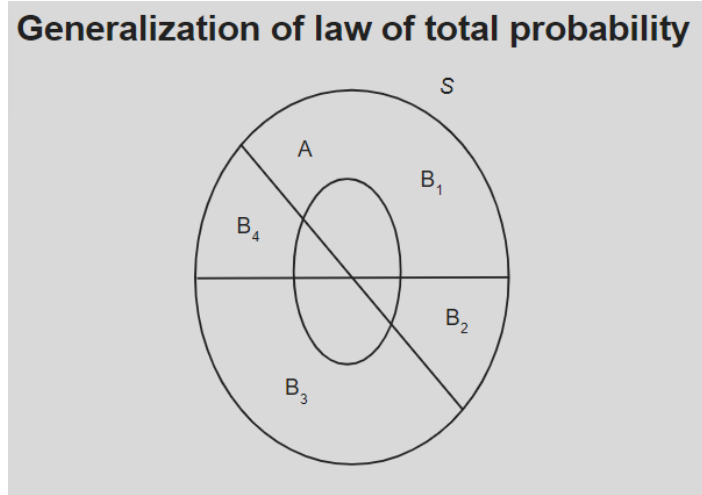
Proof:

Since, $A \cap B$ and $A \cap B^c$ are disjoint events and, $A = (A \cap B) \cup (A \cap B^c)$

By using Axiom (2),

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

$$\implies P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) \quad ; \quad (\text{Using Multiplication rule})$$

Generalization:

If there are n partitions of the sample space, say B_1, B_2, \dots, B_n , then for any event A in the sample space:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) \dots P(A|B_n)P(B_n)$$

2.6.1.1 Solved Examples:

- Q1. There are two urns 'A' and 'B'. Urn 'A' contains 7 red marbles and 6 blue marbles while, Urn 'B' contains 5 red marbles and 8 blue marbles. If an urn is chosen at random and then the marble is drawn at random from the chosen urn, then find the probability that a red ball is drawn.

Solution:

In Q2 of section 2.5.3, we have already computed that:

$$P(\text{Red Marble} \mid \text{Urn } A) = \frac{7}{13} \quad ; \quad P(\text{Red Marble} \mid \text{Urn } B) = \frac{5}{13}$$

$$P(\text{Blue Marble} \mid \text{Urn } A) = \frac{6}{13} \quad ; \quad P(\text{Blue Marble} \mid \text{Urn } B) = \frac{8}{13}$$

Let us define the event,

B_1 : The marble is drawn from Urn 'A'

B_2 : The marble is drawn from Urn 'B'

R : The marble drawn is red

The events B_1 and B_2 partitions the sample space into two halves, i.e. $B_1^c = B_2$.

Therefore, by using the Law of Total Probability, we get:

$$\begin{aligned} P(R) &= P(R|B_1)P(B_1) + P(R|B_2)P(B_2) \\ &= \left(\frac{7}{13} \times \frac{1}{2}\right) + \left(\frac{5}{13} \times \frac{1}{2}\right) \\ \therefore P(R) &= \frac{6}{13} \end{aligned}$$

- Q2. An economic model predicts that if interest rates rise, then there is a 60% chance that unemployment will increase, but that if interest rates do not rise, then there is only a 30% chance that unemployment will increase. If the economist believes that there is a 40% chance that interest rates will rise, what should she calculate is the probability that unemployment will increase?

Solution:

Let us define the events,

B_1 : The interest rates rise

B_2 : The interest rates do not rise

A : There is an increase in unemployment

Then,

$$\begin{aligned} P(B_1) &= \frac{40}{100} & ; & & P(A|B_1) &= \frac{60}{100} \\ P(B_2) &= \frac{60}{100} & ; & & P(A|B_2) &= \frac{30}{100} \end{aligned}$$

Either the interest rates will rise or will not rise. Hence, the events B_1 and B_2 partitions the sample space into two halves. ($B_1^c = B_2$)

Therefore, by using the Law of Total Probability, we get:

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) \\ &= \left(\frac{60}{100} \times \frac{40}{100}\right) + \left(\frac{30}{100} \times \frac{60}{100}\right) \\ \therefore P(A) &= 0.42 \end{aligned}$$

- Q3. A man possesses 5 coins- 2 are double-headed, 1 is double-tailed, and 2 are normal. He picks a coin at random and tosses it. What is the probability that he sees a head?

Solution:

Let us define the events,

B_1 : The double-headed coin is tossed

B_2 : The double-tailed coin is tossed

B_3 : The normal coin is tossed

A : He sees a head

Then,

$$\begin{aligned}
P(B_1) &= \frac{2}{5} & ; & & P(A|B_1) &= 1 \\
P(B_2) &= \frac{1}{5} & ; & & P(A|B_2) &= 0 \\
P(B_3) &= \frac{2}{5} & ; & & P(A|B_3) &= \frac{1}{2}
\end{aligned}$$

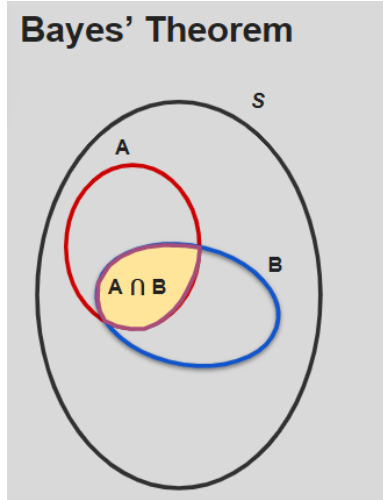
Since, the events B_1 , B_2 and B_3 partitions the sample space into three parts. ($B_1^c = B_2 \cup B_3$)
Therefore, by using the Law of Total Probability, we get:

$$\begin{aligned}
P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) \\
&= \left(1 \times \frac{2}{5}\right) + \left(0 \times \frac{1}{5}\right) + \left(\frac{1}{2} \times \frac{2}{5}\right) \\
\therefore P(A) &= 0.60
\end{aligned}$$

2.6.2 Bayes' Theorem

Consider a probability space:

Sample space S , Collection of events and Probability function P .



The Bayes' theorem states that if A and B are events with $P(A) > 0$, $P(B) > 0$ then:

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

$$\text{And, } P(A \cap B) = P(A)P(B|A) \quad (2)$$

From (1) and (2), we get:

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

Note:

- Bayes' theorem allows us to move across different conditional probability spaces.
- Bayes' theorem along with the Law of Total Probability are at the heart of so many applications in probability.

2.6.2.1 Solved Examples:

Q1. There are two urns 'A' and 'B'. Urn 'A' contains 7 red marbles and 6 blue marbles while, Urn 'B' contains 5 red marbles and 8 blue marbles. If an urn is chosen at random and then the marble is drawn at random from the chosen urn, then find the probability that urn 'A' is chosen given that the red ball is drawn.

Solution:

In Q1 of section 2.6.1.1, we have already computed that:

$$P(A|B_1) = \frac{7}{13} \quad ; \quad P(A|B_2) = \frac{5}{13}$$

$$P(A^c|B_1) = \frac{6}{13} \quad ; \quad P(A^c|B_2) = \frac{8}{13}$$

$$\text{And, } P(A) = \frac{6}{13} \quad ; \quad (\text{By using the Law of Total Probability})$$

Therefore, by using the Bayes' theorem, we get:

$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A)} \\ &= \frac{(7/13) \times (1/2)}{6/13} \\ \therefore P(B_1|A) &= \frac{7}{12} \end{aligned}$$

Q2. In a city, 1% of people have Swine Flu. In the Flu test, 95% of people with Swine flu test positive, while 2% of people without the disease will test positive. A person is randomly chosen from the city and tests positive. What is the probability that the person actually has Swine Flu?

Solution:

Let us define the events,

A: Person tests positive

B: Person has Swine Flu

$\implies B^c$: Person does not have Swine Flu

Then,

$$\begin{aligned} P(B) &= 0.01 & ; & & P(A|B) &= 0.95 \\ P(B^c) &= 0.99 & ; & & P(A|B^c) &= 0.02 \end{aligned}$$

The events B and B^c partition the sample space in two halves. Therefore, by using the Law of Total Probability, we get:

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|B^c)P(B^c) \\ &= (0.95 \times 0.01) + (0.02 \times 0.99) \\ \implies P(A) &= 0.0293 \end{aligned}$$

Now, by using the Bayes' Theorem, we get:

$$\begin{aligned} P(B|A) &= \frac{P(A|B)P(B)}{P(A)} \\ &= \frac{0.01 \times 0.95}{0.0293} \\ \therefore P(B|A) &= 0.3242 \end{aligned}$$

This implies that there is only 32.42% chance that the person has Swine Flu given that the person tests positive. Hence, the test is not much reliable.

- Q3. A student attempting an MCQ with 4 choices (of which one is correct) knows the correct answer with probability $\frac{3}{4}$. If she does not know, she guesses a random choice. Given that a question was answered correctly, what is the conditional probability that she knows the answer?

Solution:

Let us define the events,

A : She answered the question correctly

B : She knows the correct answer of the question

$\implies B^c$: She guesses the answer of the question

Then,

$$\begin{aligned} P(B) &= \frac{3}{4} & ; & & P(A|B) &= 1 \\ P(B^c) &= \frac{1}{4} & ; & & P(A|B^c) &= \frac{1}{4} \end{aligned}$$

The events B and B^c partition the sample space in two halves. Therefore, by using the Law of Total Probability, we get:

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|B^c)P(B^c) \\ &= \left(1 \times \frac{3}{4}\right) + \left(\frac{1}{4} \times \frac{1}{4}\right) \\ \implies P(A) &= \frac{13}{16} \end{aligned}$$

Now, by using the Bayes' Theorem, we get:

$$\begin{aligned}
 P(B|A) &= \frac{P(A|B)P(B)}{P(A)} \\
 &= \frac{1 \times (3/4)}{13/16} \\
 &= \frac{12}{13} \\
 \therefore P(B|A) &= 0.9231
 \end{aligned}$$

This implies that, there is approximately, 92.31% chance that she knew the answer to a question given that the question was answered correctly.

- Q4. You first roll a fair die, then toss as many fair coins as the number that showed on the die. Given that 5 heads are obtained, what is the probability that the die showed 5?

Solution:

Let us define the events,

E_i : The die showed the number i ; $i = 1, 2, 3, 4, 5, 6$

A: 5 Heads are obtained

Then,

$$P(E_i) = \frac{1}{6} ; i = 1, 2, 3, 4, 5, 6 \quad \text{and,} \quad P(A|E_i) = 0 ; i = 1, 2, 3, 4$$

Now, by using Uniform distribution, we get:

$$P(A|E_5) = \frac{1}{32} \quad \text{and} \quad P(A|E_6) = \frac{3}{32}$$

Since, the events E_1, E_2, \dots, E_6 partition the sample space in two halves. Therefore, by using the Law of Total Probability, we get:

$$\begin{aligned}
 P(A) &= P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \dots + P(A|E_6)P(E_6) \\
 &= 0 + \left(\frac{1}{6} \times \frac{1}{32}\right) + \left(\frac{1}{6} \times \frac{3}{32}\right) \\
 \implies P(A) &= \frac{1}{48}
 \end{aligned}$$

Now, by using the Bayes' theorem, we get:

$$\begin{aligned}
 P(E_5|A) &= \frac{P(A|E_5)P(E_5)}{P(A)} \\
 &= \frac{(1/32) \times (1/6)}{1/48} \\
 &= \frac{1}{4} \\
 \therefore P(E_5|A) &= 0.25
 \end{aligned}$$

2.6.3 Independence of Two Events

Two events A and B are independent if:

$$P(A \cap B) = P(A)P(B) \quad ; \quad P(A) > 0, P(B) > 0$$

In other words, the events A and B are independent if the probability of occurrence of A (B) is unaffected by the occurrence of B (A), i.e. $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

If the two events A and B are independent, then it implies that:

- (i) The events A and B^c are also independent.

Proof:

The events $A \cap B$ and $A \cap B^c$ are disjoint and, $(A \cap B) \cup (A \cap B^c) = A$.

By using Axiom (2),

$$\begin{aligned} P((A \cap B) \cup (A \cap B^c)) &= P(A \cap B) + P(A \cap B^c) \\ P(A) &= P(A \cap B) + P(A \cap B^c) \\ \implies P(A \cap B^c) &= P(A \cap B) - P(A) \\ &= P(A)P(B) - P(A) \quad ; \quad (A \text{ and } B \text{ are independent events}) \\ &= P(A)[1 - P(B)] \\ \therefore P(A \cap B^c) &= P(A)P(B^c) \end{aligned}$$

- (ii) The events A^c and B are also independent.

Proof:

The events $A \cap B$ and $A^c \cap B$ are disjoint and, $(A \cap B) \cup (A^c \cap B) = B$.

By using Axiom (2),

$$\begin{aligned} P((A \cap B) \cup (A^c \cap B)) &= P(A \cap B) + P(A^c \cap B) \\ P(B) &= P(A \cap B) + P(A^c \cap B) \\ \implies P(A^c \cap B) &= P(A \cap B) - P(B) \\ &= P(A)P(B) - P(B) \quad ; \quad (A \text{ and } B \text{ are independent events}) \\ &= [1 - P(A)]P(B) \\ \therefore P(A^c \cap B) &= P(A^c)P(B) \end{aligned}$$

- (iii) The events A^c and B^c are also independent.

Proof:

$$\begin{aligned}
P((A^c \cap B^c)) &= P(A \cup B)^c \quad ; \quad (\text{Using De Morgan's Law}) \\
&= 1 - P(A \cup B) \\
&= 1 - [P(A) + P(B) - P(A \cap B)] \\
&= 1 - P(A) - P(B) + P(A)P(B) \quad ; \quad (A \text{ and } B \text{ are independent events}) \\
&= 1 - P(A) - P(B)[1 - P(A)] \\
&= [1 - P(A)][1 - P(B)] \\
\therefore P(A^c \cap B^c) &= P(A^c)P(B^c)
\end{aligned}$$

Note:

- The independence is defined very precisely in probability that one should not use their intuition to conclude that the events are independent.
- If the two events A and B are independent, then the multiplicative rule states that the probability of occurrence of event A and event B is given by, $P(A \cap B) = P(A)P(B)$
- If the two events A and B are disjoint and B occurs then A definitely did not occur. Thus, they can never be independent, as the occurrence of B impacts the conditional probability of A .
 \implies For event to be independent, they should have non-empty intersection

2.6.3.1 Solved Examples:

Q1. If a fair coin is tossed thrice and the events are defined as follows:

A : First toss is heads

B : Second toss is heads

Then, check the following:

- Are events A and B independent?
- Are events A and B^c independent?

Solution:

The sample space; $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

Since, the same fair coin is tossed thrice. Therefore, the distribution is uniform.

$$A = \{HHH, HHT, HTH, HTT\} \quad ; \quad P(A) = \frac{1}{2}$$

$$B = \{HHH, HHT, THH, THT\} \quad ; \quad P(B) = \frac{1}{2}$$

$$\text{and, } A \cap B = \{HHH, HHT\} \quad ; \quad P(A \cap B) = \frac{1}{4}$$

(i). Since,

$$\begin{aligned}P(A \cap B) &= \frac{1}{4} \\&= \frac{1}{2} \times \frac{1}{2} \\ \therefore P(A \cap B) &= P(A)P(B)\end{aligned}$$

Hence, A and B are independent events.

(ii). Since, the events A and B are independent, this implies that the event A and B^c are also independent.

Q2. If a fair dice is rolled and the events are defined as follows:

A : The die showed even number

B : The die showed odd number

Are events A and B independent?

Solution:

The sample space; $S = \{1, 2, 3, 4, 5, 6\}$

$$A = \{2, 4, 6\} \quad ; \quad P(A) = \frac{1}{2}$$

$$B = \{1, 3, 5\} \quad ; \quad P(B) = \frac{1}{2}$$

$$A \cap B = \Phi \quad \implies \quad A \text{ and } B \text{ are disjoint events}$$

Since, A and B are disjoint events, they can not be independent.

Therefore, A and B are dependent events.

Q3. If a card is drawn from a well shuffled pack of 52 cards, and the events are defined as follows:

A : Card is a spade

B : Card is a king

Are events A and B independent?

Solution:

$$A = \{\text{Spade card}\} \quad ; \quad P(A) = \frac{13}{52}$$

$$B = \{\text{King card}\} \quad ; \quad P(B) = \frac{4}{52}$$

$$A \cap B = \{\text{Spade-King card}\} \quad ; \quad P(A \cap B) = \frac{1}{52}$$

Now,

$$\begin{aligned}P(A)P(B) &= \left(\frac{13}{52}\right) \times \left(\frac{4}{52}\right) \\&= \frac{1}{52} \\ \therefore P(A)P(B) &= P(A \cap B)\end{aligned}$$

Hence, the events A and B are independent.

Remark: As there's only one outcome, which can belong to multiple events. That's why even though only one card is drawn, still two or more events can be independent.

2.6.4 Mutual Independence of Three Events

The events A , B and C are mutually independent if:

$$(1) P(A \cap B \cap C) = P(A)P(B)P(C)$$

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

$$(3) P(A \cap C) = P(A)P(C)$$

$$(4) P(B \cap C) = P(B)P(C)$$

Generalization:

The events A_1, A_2, \dots, A_n are mutually independent if, for all i_1, i_2, \dots, i_k :

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$$

Note:

- For mutual independence of n events, there are almost 2^n constraints i.e.
 $P(\{\text{Intersection of any subset of events}\}) = \text{Product of the } P(\{\text{Events}\})$
- If n events are mutually independent, then it implies that any subset of events with or without complementing are independent as well.

2.6.4.1 Solved Examples:

Q1. If a fair coin is tossed twice and the events are defined as follows:

A : Either both toss are heads or tails

B : First toss is heads

C : Second toss is heads

Then, check for independence of the three events.

Solution:

The sample space; $S = \{HH, HT, TH, TT\}$

Since, the same fair coin is tossed twice. Therefore, the distribution is uniform.

$$A = \{HH, TT\} \quad ; \quad P(A) = \frac{2}{4}$$

$$B = \{HH, HT\} \quad ; \quad P(B) = \frac{2}{4}$$

$$C = \{HH, TH\} \quad ; \quad P(C) = \frac{2}{4}$$

Now,

$$A \cap B = \{HH\} ; \quad P(A \cap B) = \frac{1}{4} = P(A)P(B) \implies \text{Events } A \text{ and } B \text{ are independent}$$

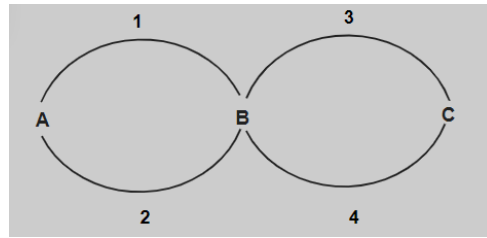
$$A \cap C = \{HH\} ; \quad P(A \cap C) = \frac{1}{4} = P(A)P(C) \implies \text{Events } A \text{ and } C \text{ are independent}$$

$$B \cap C = \{HH\} ; \quad P(B \cap C) = \frac{1}{4} = P(B)P(C) \implies \text{Events } B \text{ and } C \text{ are independent}$$

$$\text{But, } A \cap B \cap C = \{HH\} ; \quad P(A \cap B \cap C) = \frac{1}{4} \neq P(A)P(B)P(C)$$

Hence, the events A , B and C are not mutually independent but are pairwise independent, i.e. each event is independent of every other possible combination of paired events.

- Q2. Two roads each connect A and B and B and C . Each of the four roads get blocked with probability p independent of all other roads. What is the probability that there is an open route from A to B given that there is no open route from A to C ?



Solution:

Let us define the events,

E_1 : First road from A to B i.e. Road 1, gets blocked

E_2 : Second road from A to B i.e. Road 2, gets blocked

E_3 : First road from B to C i.e. Road 3, gets blocked

E_4 : Second road from B to C i.e. Road 4, gets blocked

F : There is no open route from A to B

E : There is no open route from A to C

It is given that each road get blocked with probability p independent of all other roads. This implies that the 4 events are mutually independent and,

$$P(E_i) = \frac{1}{4} \quad ; \quad i = 1, 2, 3, 4$$

Now,

$$\begin{aligned}
P(F^c) &= P(\text{Either road 1 or road 2 is open}) \\
&= P(E_1^c \cup E_2^c) \\
&= P(E_1 \cap E_2)^c \quad ; \quad (\text{Using De Morgan's Law}) \\
&= 1 - P(E_1 \cap E_2) \\
&= 1 - P(E_1)P(E_2) \quad ; \quad (E_1 \text{ and } E_2 \text{ are independent events}) \\
\therefore P(F^c) &= 1 - p^2 \\
\text{and, } P(F) &= 1 - P(F^c) = p^2
\end{aligned}$$

And, $P(E|F) = 1$

Because, If there is no open route from A to B , then certainly there would be no open route from A to C

Also,

$$\begin{aligned}
P(E|F^c) &= P(\text{Road 3 and Road 4 are blocked}) \\
&= P(E_3 \cap E_4) \\
&= P(E_3)P(E_4) \quad ; \quad (E_3 \text{ and } E_4 \text{ are independent events}) \\
\therefore P(E|F^c) &= p^2
\end{aligned}$$

Therefore,

$$\begin{aligned}
P(E) &= P(E|F)P(F) + P(E|F^c)P(F^c) \quad ; \quad (\text{Law of total probability}) \\
&= (1 \times p^2) + (p^2 \times (1 - p^2)) \\
\therefore P(E) &= 2p^2 - p^4
\end{aligned}$$

Hence, probability that there is an open route from A to B given that there is no open route from A to C is:

$$\begin{aligned}
P(F^c|E) &= \frac{P(E \cap F^c)}{P(E)} \\
&= \frac{P(E|F^c)P(F^c)}{P(E)} \quad ; \quad (\text{Using Multiplication rule}) \\
&= \frac{p^2 \times (1 - p^2)}{2p^2 - p^4} \\
\therefore P(F^c|E) &= \frac{1 - p^2}{2 - p^2}
\end{aligned}$$

2.7 Repeated Independent Trials

2.7.1 Bernoulli Distribution

2.7.1.1 Single Bernoulli Trial

Consider a probability space: Sample space S , Collection of events and probability

function P .

Consider an event A , such that $P(A) = p$

If in an experiment, the occurrence of a particular event A is considered to be a success and the non-occurrence of the event A is considered a failure, then we can interpret it as a Bernoulli trial as follows:

- Sample space (with only two possible outcomes); $S = \{\text{Success, Failure}\}$.
- $P(\text{Success}) = p$

Or, if we denote the success by 1 and failure by 0 then,

$S = \{0, 1\}$ with $P(1) = p$ and $P(0) = 1 - p$

This distribution is denoted as Bernoulli(p).

2.7.1.2 Repeated Bernoulli Trials

Consider a Bernoulli(p) trial with $S = \{0, 1\}$.

If this Bernoulli(p) trial is repeated n times independently, then the sample space will consist of 2^n outcomes.

Example:

If $n = 3$, then $S = \{000, 001, 010, 100, 110, 101, 011, 111\}$ with probabilities as follows:

- $P(000) = P(0 \text{ in trial 1 and } 0 \text{ in trial 2 and } 0 \text{ in trial 3})$
 $= P(0 \text{ in trial 1})P(0 \text{ in trial 2})P(0 \text{ in trial 3})$; (Trials are independent)
 $= (1 - p)(1 - p)(1 - p)$
 $\therefore P(000) = (1 - p)^3$
- $P(101) = P(1 \text{ in trial 1 and } 0 \text{ in trial 2 and } 1 \text{ in trial 3})$
 $= P(1 \text{ in trial 1})P(0 \text{ in trial 2})P(1 \text{ in trial 3})$; (Trials are independent)
 $= p(1 - p)p$
 $\therefore P(101) = p^2(1 - p)$

Generalization:

If a Bernoulli(p) trial is repeated n times independently, then

$$P(b_1 b_2 \dots b_n) = p^w (1 - p)^{n-w} \quad , \text{ where } w = \text{Number of 1s in } b_1, b_2, \dots, b_n$$

2.7.1.3 Solved Examples:

Q1. Suppose a fair coin is tossed five times. If in each toss of a coin, getting tails is considered to be as a success, then compute the following:

- (i) Probability of getting 0 tails
- (ii) Probability of getting 2 tails

Solution:

For a single toss of a fair coin, consider an event $A = \{\text{Tail}\}$

Since in a trial, the occurrence of a particular event A is considered to be a success and the non-occurrence of the event A is considered a failure, then we can interpret it as a Bernoulli(p) trial with:

- $S = \{0, 1\}$, where tail is denoted by 1 and head by 0.
- $p = \frac{1}{2}$ (because fair coin)

Now the coin is tossed 5 times, i.e., Bernoulli(0.5) trial is repeated 5 times independently, which implies that the sample space will consist of 2^5 (i.e., 32) equally likely outcomes.

$$\begin{aligned}
 \text{(i)} \quad P(0 \text{ Tails}) &= P(HHHHH) \\
 &= P(0 \text{ in all 5 trials}) \\
 &= P(00000) \\
 &= \left(\frac{1}{2}\right)^0 \times \left(1 - \frac{1}{2}\right)^{5-0} \quad ; \quad (\text{Trials are independent}) \\
 &= \left(\frac{1}{2}\right)^5 \\
 \therefore P(0 \text{ Tails}) &= \frac{1}{32}
 \end{aligned}$$

$$\begin{aligned}
 \text{(ii)} \quad P(2 \text{ Tails}) &= P(\{TTHHH, THTHH, THHTH, THHHT, HTTHH, \\
 &\quad HTHTH, HTHHT, HHTTH, HHTHT, HHHTT\}) \\
 &= P(\{11000, 10100, 10010, 10001, 01100, 01010, 01001, 00110, 00101, 00011\}) \\
 &\quad (\text{Since, the events } \{11000\}, \{10100\}, \dots, \{00011\} \text{ are disjoint}) \\
 \implies P(2 \text{ Tails}) &= P(\{11000\}) + P(\{10100\}) + P(\{10010\}) + P(\{10001\}) + P(\{01100\}) \\
 &\quad + P(\{01010\}) + P(\{01001\}) + P(\{00110\}) + P(\{00101\}) + P(\{00011\}) \\
 &= \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 + \dots + \left(\frac{1}{2}\right)^5 \quad ; \quad (\text{Trials are independent}) \\
 &= 10 \times \left(\frac{1}{2}\right)^5 \\
 \therefore P(2 \text{ Tails}) &= \frac{10}{32}
 \end{aligned}$$

Q2. Suppose a biased coin with $P(T) = \frac{2}{3}$ is tossed five times. If in each toss of a coin, getting tails is considered to be as a success, then compute the following:

- (i) Probability of getting 1 tail
- (ii) Probability of getting 5 tails

Solution:

For a single toss of the given coin, consider an event $A = \{Tail\}$
 Since in a trial, the occurrence of a particular event A is considered to be a success and the non-occurrence of the event A is considered a failure, then we can interpret it as a Bernoulli(p) trial with:

- $S = \{0, 1\}$ where, tail is denoted by 1 and head by 0.
- $p = \frac{2}{3}$ (given)

Now the coin is tossed 5 times, i.e., Bernoulli($\frac{2}{3}$) trial is repeated 5 times independently, which implies that the sample space will consist of 32 outcomes.

$$(i) \quad P(1 \text{ Tail}) = P(\{THHHH, HTHHH, HHTHH, HHHTH, HHHHT\}) \\ = P(\{10000, 01000, 00100, 00010, 00001\})$$

(Since, the events $\{10000\}, \{01000\}, \dots, \{00001\}$ are disjoint)

$$\Rightarrow P(1 \text{ Tail}) = P(\{10000\}) + P(\{01000\}) + P(\{00100\}) + P(\{00010\}) + P(\{00001\}) \\ = \left(\frac{2}{3}\right) \left(\frac{1}{3}\right)^{5-1} + \left(\frac{2}{3}\right) \left(\frac{1}{3}\right)^{5-1} + \dots + \left(\frac{2}{3}\right) \left(\frac{1}{3}\right)^{5-1} ; \text{ (Trials are independent)} \\ = 5 \times \left(\frac{2}{3}\right) \left(\frac{1}{3}\right)^{5-1} \\ \therefore P(1 \text{ Tail}) = \frac{10}{243}$$

$$(ii) \quad P(5 \text{ Tails}) = P(TTTTT) \\ = P(1 \text{ in all the 5 trials}) \\ = P(11111) \\ = \left(\frac{2}{3}\right)^5 \times \left(1 - \frac{2}{3}\right)^{5-5} ; \text{ (Trials are independent)} \\ = \left(\frac{2}{3}\right)^5 \\ \therefore P(5 \text{ Tails}) = \frac{32}{243}$$

2.7.2 Binomial Distribution

We can simplify the repeated Bernoulli trials (section 2.7.1.2) to Binomial distribution.

Consider a Bernoulli(p) trial with $S = \{0, 1\}$.

If this Bernoulli(p) trial is repeated n times independently, then the number of successes in these n trials is given by Binomial distribution as follows:

- Sample space; $S = \{0, 1, 2, \dots, n\}$
Because we can have 0 successes, 1 success, 2 successes, \dots , n successes.
- Distribution function is given by

$$P(B(n, p) = k) = {}^nC_k p^k (1 - p)^{n-k} \quad ; \quad k = 0, 1, 2, \dots, n$$

Proof:

$$\begin{aligned} P(B(n, p) = k) &= P(\text{Out of } n \text{ Bernoulli}(p) \text{ trials exactly } k \text{ are } 1's) \\ &= P(\text{Trials } b_1, b_2, \dots, b_n \text{ result in exactly } k \text{ } 1's) \\ &= (\text{Number of } b_1, b_2, \dots, b_n \text{ with exactly } k \text{ } 1's) \times p^k (1 - p)^{n-k} \end{aligned}$$

Number of ways to get exactly k 1's in n trials = nC_k .

$$\therefore P(B(n, p) = k) = {}^nC_k p^k (1 - p)^{n-k} \quad ; \quad k = 0, 1, 2, \dots, n$$

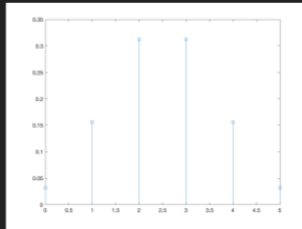
Remark:

$$\begin{aligned} &P(B(n, p) = 0 \text{ or } B(n, p) = 1 \text{ or } \dots \text{ or } B(n, p) = n) = 1 \\ \implies &P(B(n, p) = 0) + P(B(n, p) = 1) + \dots + P(B(n, p) = n) = 1 \quad ; \quad (\text{Because disjoint events}) \\ \implies &(1 - p)^n + {}^nC_1 p (1 - p)^{n-1} + {}^nC_2 p^2 (1 - p)^{n-2} + \dots + p^n = 1 \end{aligned}$$

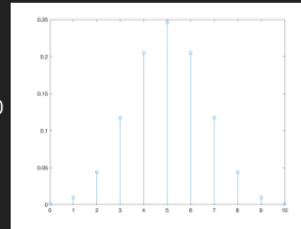
2.7.2.1 Visualtion of Binomial Distribution

Binomial($n, 0.5$)

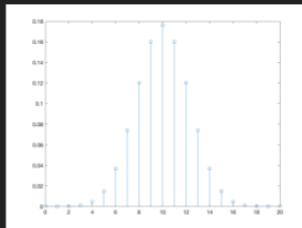
$n = 5$



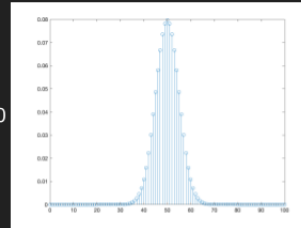
$n = 10$



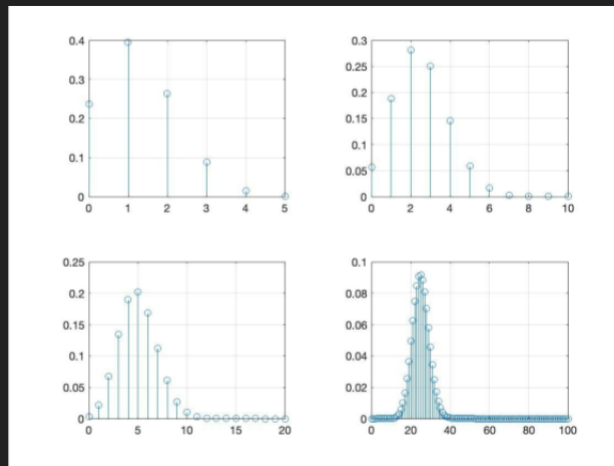
$n = 20$



$n = 100$



$p = 0.25, n = 5, 10, 20, 100$



Note:

- The plot starts at $(1 - p)^n$ and then increases till it reaches the peak and then falls to p^n
- The peak is roughly around np and the exact values are as follows:
 - (1) If ' $(n + 1)p$ ' is an integer, then it is bimodal (i.e. two peaks) and the two peak values are ' $(n + 1)p$ ' and ' $(n + 1)p - 1$ '.
 - (2) If ' $(n + 1)p$ ' is not an integer, then there exists a unique modal value (i.e. unique peak value) and it's the integral part of ' $(n + 1)p$ '.

2.7.2.2 Solved Examples:

Q1. Each person has a disease with probability 0.1 independently. Out of 100 random persons tested for the disease, what is the probability that 20 persons test positive? Assume that the disease can be tested accurately with no false positives.

Solution:

Each test can be considered as a Bernoulli(p) trial, where success is the person tested as positive and $p = 0.1$.

Now, this Bernoulli(0.1) trial is repeated 100 times independently. Hence, the probability of getting 20 successes in these 100 trials is given by:

$$\begin{aligned} P(B(100, 0.1) = 20) &= {}^{100}C_{20} (0.1)^{20} (1 - 0.1)^{100-20} \\ &= {}^{100}C_{20} (0.1)^{20} (0.9)^{80} \\ \therefore P(B(100, 0.1) = 20) &\approx 0.0012 \end{aligned}$$

Therefore, probability that out of 100 random persons tested for the disease, exactly 20 persons test positive is 0.0012 approximately.

Q2. Suppose a fair coin is tossed 10 times, then find the following probabilities:

- (i) The number of heads is a multiple of 3.
- (ii) The number of heads is even.

Solution:

Each toss of a fair coin can be considered as a Bernoulli(p) trial, where success is getting a head and $p = 0.5$.

Now, this Bernoulli(0.5) trial is repeated 10 times independently.

Hence, its distribution is defined as Binomial(10, 0.5), say B with sample space $S = \{0, 1, 2, \dots, 10\}$

Let us define the events

$E : \{\text{Number of heads is a multiple of 3}\} = \{0, 3, 6, 9\}$

$F : \{\text{Number of heads is even}\} = \{0, 2, 4, 6, 8, 10\}$

(i) The probability that the number of heads is a multiple of 3 is given by:

$$\begin{aligned}
P(E) &= P(B = 0 \text{ or } B = 3 \text{ or } B = 6 \text{ or } B = 9) \\
&= P(B = 0) + P(B = 3) + P(B = 6) + P(B = 9) ; \text{ (Because, disjoint events)} \\
&= \left(1 - \frac{1}{2}\right)^{10} + {}^{10}C_3 \left(\frac{1}{2}\right)^3 \left(1 - \frac{1}{2}\right)^{10-3} + {}^{10}C_6 \left(\frac{1}{2}\right)^6 \left(1 - \frac{1}{2}\right)^{10-6} \\
&\quad + {}^{10}C_9 \left(\frac{1}{2}\right)^9 \left(1 - \frac{1}{2}\right)^{10-9} \\
&= \left(\frac{1}{2}\right)^{10} [1 + {}^{10}C_3 + {}^{10}C_6 + {}^{10}C_9] \\
\therefore P(E) &= 0.33
\end{aligned}$$

Hence, probability that the number of heads is a multiple of 3 is 0.33.

(ii) The probability that the number of heads is even is given by:

$$P(F) = P(B = 0 \text{ or } B = 2 \text{ or } B = 4 \text{ or } B = 6 \text{ or } B = 8 \text{ or } B = 10)$$

Since, the events $B = 0, B = 2, \dots, B = 10$ are disjoint events

$$\begin{aligned}
P(F) &= P(B = 0) + P(B = 2) + P(B = 4) + P(B = 6) + P(B = 8) + P(B = 10) \\
&= \left(1 - \frac{1}{2}\right)^{10} + {}^{10}C_2 \left(\frac{1}{2}\right)^2 \left(1 - \frac{1}{2}\right)^{10-2} + {}^{10}C_4 \left(\frac{1}{2}\right)^4 \left(1 - \frac{1}{2}\right)^{10-4} \\
&\quad + {}^{10}C_6 \left(\frac{1}{2}\right)^6 \left(1 - \frac{1}{2}\right)^{10-6} + {}^{10}C_8 \left(\frac{1}{2}\right)^8 \left(1 - \frac{1}{2}\right)^{10-8} + \left(\frac{1}{2}\right)^{10} \\
&= \left(\frac{1}{2}\right)^{10} [1 + {}^{10}C_2 + {}^{10}C_4 + {}^{10}C_6 + {}^{10}C_8 + {}^{10}C_{10}] \\
\therefore P(E) &= 0.50
\end{aligned}$$

Hence, probability that the number of heads is even is 0.50.

Q3. A bit (0 or 1) sent by Alice to Bob gets flipped with probability 0.1.

- (i) If 5 bits are sent by Alice independently, what is the probability that at most 2 bits get flipped?
- (ii) If 10 bits are sent by Alice independently, what is the probability that at most 2 bits get flipped?

Solution:

Each bit sent by Alice can be considered as a Bernoulli(p) trial, where success is a bit getting flipped and $p = 0.1$.

- (i) The Bernoulli(0.1) trial is repeated 5 times independently.

Hence, its distribution is defined as Binomial(5, 0.1), say B with sample space;
 $S = \{0, 1, 2, \dots, 5\}$

Let us define an event $E : \{\text{At most 2 bits gets flipped}\} = \{0, 1, 2\}$

$$\begin{aligned} P(E) &= P(B = 0 \text{ or } B = 1 \text{ or } B = 2) \\ &= P(B = 0) + P(B = 1) + P(B = 2) \quad ; \quad (\text{Because, disjoint events}) \\ &= (1 - 0.1)^5 + {}^5C_1(0.1)(1 - 0.1)^{5-1} + {}^5C_2(0.1)^2(1 - 0.1)^{5-2} \\ &= (0.9)^5 + {}^5C_1(0.1)(0.9)^4 + {}^5C_2(0.1)^2(0.9)^3 \\ \therefore P(E) &= 0.9914 \end{aligned}$$

Hence, the probability that at most 2 of the 5 bits gets flipped is 0.9914.

- (ii) The Bernoulli(0.1) trial is repeated 10 times independently.

Hence, its distribution is defined as Binomial(10, 0.1), say B , with sample space;
 $S = \{0, 1, 2, \dots, 10\}$

Let us define an event $E : \{\text{At most 2 bits gets flipped}\} = \{0, 1, 2\}$

$$\begin{aligned} P(E) &= P(B = 0 \text{ or } B = 1 \text{ or } B = 2) \\ &= P(B = 0) + P(B = 1) + P(B = 2) \quad ; \quad (\text{Because, disjoint events}) \\ &= (1 - 0.1)^{10} + {}^{10}C_1(0.1)(1 - 0.1)^{10-1} + {}^{10}C_2(0.1)^2(1 - 0.1)^{10-2} \\ &= (0.9)^{10} + {}^{10}C_1(0.1)(0.9)^9 + {}^{10}C_2(0.1)^2(0.9)^8 \\ \therefore P(E) &= 0.9298 \end{aligned}$$

Hence, the probability that at most 2 of the 10 bits gets flipped is 0.9298.

2.7.3 Geometric Distribution

We can also simplify the repeated Bernoulli trials (section 2.7.1.2) to Geometric distribution.

Consider a Bernoulli(p) trial with $S = \{0, 1\}$.

If this Bernoulli(p) trial is repeated independently, then the number of trials needed for the first success is given by the Geometric distribution as follows:

- Sample space; $S = \{1, 2, 3, 4, 5, 6, \dots\}$
 This is because we can get the first success in 1st trial, 2nd trial, 3rd trial, ..., and so on.
- $P(G(p) = k) = (1 - p)^{k-1} p \quad ; \quad k = 1, 2, 3, \dots$

Proof:

$$\begin{aligned}
 P(G(p) = k) &= P(\text{Failure in the first } (k-1) \text{ trials and success in the } k\text{-th trial}) \\
 &= P(\text{First } (k-1) \text{ trials result in 0 and } k\text{-th trial is 1}) \\
 &= P(\underbrace{00 \dots 0}_{(k-1) \text{ times}} 1) \\
 \therefore P(G(p) = k) &= (1-p)^{k-1}p \quad ; \quad k = 1, 2, 3 \dots
 \end{aligned}$$

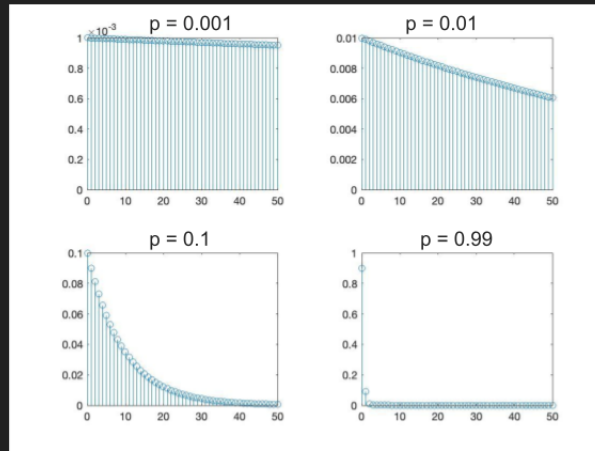
Remark:

- The number of trials is unknown, as we keep on repeating the trials independently till we get the first success.
- The probability of getting the first success within the first k trials is given by:

$$\begin{aligned}
 P(G(p) \leq k) &= P(G(p) = 1 \text{ or } G(p) = 2 \text{ or } \dots \text{ or } G(p) = k) \\
 P(G(p) \leq k) &= P(G(p) = 1) + P(G(p) = 2) + \dots + P(G(p) = k) \\
 P(G(p) \leq k) &= p + (1-p)p + (1-p)^2p + \dots + (1-p)^{k-1}p \\
 P(G(p) \leq k) &= p[1 + (1-p) + (1-p)^2 + \dots + (1-p)^{k-1}] \\
 P(G(p) \leq k) &= p \left[\frac{1 - (1-p)^k}{1 - (1-p)} \right] \quad ; \quad (\text{By using sum of G.P}) \\
 \therefore P(G(p) \leq k) &= 1 - (1-p)^k
 \end{aligned}$$

2.7.3.1 Visualisation of Geometric Distribution

Plot of geometric distribution (shown till $k = 50$)



Note:

- The plot starts at p and then keeps falling.
- Even though the plot keeps on decreasing but, if $p < 1$, it never goes all the way to zero.

2.7.3.2 Solved Examples:

Q1. In Ludo, a player needs to repeatedly throw a fair die till she gets a 1. Find the following:

- (i) The probability that she needs lesser than 6 throws.
- (ii) The probability that she needs lesser than 11 throws.
- (iii) The probability that she needs lesser than 21 throws.

Solution:

Each throw of a fair die can be considered as a Bernoulli(p) trial, where success is getting 1 and $p = \frac{1}{6}$.

Now, this Bernoulli($\frac{1}{6}$) trial is repeated independently till we get the first success.

Hence, its distribution is defined as Geometric($\frac{1}{6}$), say G , with sample space ; $S = \{1, 2, 3, \dots\}$

- (i) The probability that she needs lesser than 6 throws is:

$$\begin{aligned} P(G < 6) &= P(G \leq 5) \\ &= 1 - \left(1 - \frac{1}{6}\right)^5 \quad ; \quad (\text{Remark in section 2.7.3}) \\ &= 1 - \left(\frac{5}{6}\right)^5 \\ \therefore P(G < 6) &= 0.5981 \end{aligned}$$

Hence, the probability that she needs lesser than 6 throws is 0.5981.

- (ii) The probability that she needs lesser than 11 throws is:

$$\begin{aligned} P(G < 11) &= P(G \leq 10) \\ &= 1 - \left(1 - \frac{1}{6}\right)^{10} \quad ; \quad (\text{Remark in section 2.7.3}) \\ &= 1 - \left(\frac{5}{6}\right)^{10} \\ \therefore P(G < 11) &= 0.8385 \end{aligned}$$

Hence, the probability that she needs lesser than 11 throws is 0.8385.

(iii) The probability that she needs lesser than 21 throws is:

$$\begin{aligned}P(G < 21) &= P(G \leq 20) \\&= 1 - \left(1 - \frac{1}{6}\right)^{20} \quad ; \quad (\text{Remark in section } \mathbf{2.7.3}) \\&= 1 - \left(\frac{5}{6}\right)^{20} \\ \therefore P(G < 21) &= 0.9739\end{aligned}$$

Hence, the probability that she needs lesser than 6 throws is 0.9739.

Q2. Player 1 is 40% free throw shooter, while Player 2 is 70% shooter. Each throw is independent of all previous throws. The two players alternate shooting, with Player 1 starting till the basket is scored.

- (i) What is the probability that Player 1 wins before the 3rd round?
- (ii) What is the probability that Player 1 wins?

Solution:

Each throw of basketball by Player 1 and by Player 2 can be considered as Bernoulli(0.4) and Bernoulli(0.7) trial, respectively, where success is getting the basket scored.

- (i) The favourable outcomes for an event that Player 1 wins before the third round is as follows:

- Player 1 wins in the first round, i.e. 1_{P_1}
And, the probability is given by:

$$P(1_{P_1}) = 0.4$$

- Player 1 wins in the second round, which implies that both player loses in the first round i.e. $0_{P_1}0_{P_2}1_{P_1}$
And, the probability is given by:

$$\begin{aligned}P(0_{P_1}0_{P_2}1_{P_1}) &= P(0_{P_1})P(0_{P_2})P(1_{P_1}) \quad ; \quad (\text{because, independent events}) \\&= (1 - 0.4)(1 - 0.7)(0.4) \\&= (0.6)(0.3)(0.4) \\ \therefore P(0_{P_1}0_{P_2}1_{P_1}) &= 0.072\end{aligned}$$

Since, the events $\{1_{P_1}\}$ and $\{0_{P_1}0_{P_2}1_{P_1}\}$ are disjoint. Therefore, the probability that Player 1 wins before the third round is:

$$\begin{aligned}P(\text{Player 1 wins before 3rd round}) &= P(1_{P_1}) + P(0_{P_1}0_{P_2}1_{P_1}) \\&= 0.4 + 0.072 \\&= 0.472\end{aligned}$$

Hence, there is 47.2% chance that Player 1 wins the game before the third round.

(ii) Similarly, the favourable outcomes for the event that Player 1 wins the game are:

$$\{1_{P_1}, 0_{P_1}0_{P_2}1_{P_1}, (0_{P_1}0_{P_2})_21_{P_1}, (0_{P_1}0_{P_2})_31_{P_1}, (0_{P_1}0_{P_2})_41_{P_1}, \dots\}$$

where, $(0_{P_1}0_{P_2})_i$ means that $(0_{P_1}0_{P_2})$ is repeated i times.

Now,

- The probability for $\{1_{P_1}\}$ given by:

$$P(1_{P_1}) = 0.4$$

- The probability for $\{0_{P_1}0_{P_2}1_{P_1}\}$ given by:

$$\begin{aligned} P(0_{P_1}0_{P_2}1_{P_1}) &= P(0_{P_1})P(0_{P_2})P(1_{P_1}) \quad ; \quad (\text{because, independent events}) \\ &= (1 - 0.4)(1 - 0.7)(0.4) \\ \therefore P(0_{P_1}0_{P_2}1_{P_1}) &= (0.6)(0.3)(0.4) \end{aligned}$$

- The probability for $\{(0_{P_1}0_{P_2})_21_{P_1}\}$ given by:

$$\begin{aligned} P((0_{P_1}0_{P_2})_21_{P_1}) &= P(0_{P_1})P(0_{P_2})P(0_{P_1})P(0_{P_2})P(1_{P_1}) \\ &\quad (\text{Because, independent events}) \\ &= (1 - 0.4)(1 - 0.7)(1 - 0.4)(1 - 0.7)(0.4) \\ \therefore P(0_{P_1}0_{P_2}1_{P_1}) &= [(0.6)(0.3)]^2(0.4) \end{aligned}$$

and so on.

Since, the events $\{1_{P_1}\}, \{0_{P_1}0_{P_2}1_{P_1}\}, \{(0_{P_1}0_{P_2})_21_{P_1}\}, \dots$ are disjoint, probability that Player 1 wins the game is:

$$\begin{aligned} P(\text{Player 1 wins}) &= P(1_{P_1}) + P(0_{P_1}0_{P_2}1_{P_1}) + P((0_{P_1}0_{P_2})_21_{P_1}) + \dots \\ &= 0.4 + [(0.6)(0.3)]0.4 + [(0.6)(0.3)]^20.4 + \dots \\ &= \frac{0.4}{1 - (0.6)(0.3)} \quad ; \quad (\text{Using sum of infinite GP series}) \\ &= \frac{0.4}{1 - 0.18} \\ \therefore P(\text{Player 1 wins}) &= 0.4878 \end{aligned}$$

Hence, there is 48.78% chance that Player 1 wins the game.

Remark:

This question is not quite based on geometric distribution because, if it had been geometric, then instead of $(0.6)(0.3)$ it would have been 0.6.

2.8 Discrete Random Variable

Random Variable:

A random variable is a function with domain as the sample space of an experiment and range as the real numbers, i.e. a function from the sample space to the real line. There is a technical condition that functions need to satisfy, which will be discussed in section 2.8.1.

Discrete Random Variable:

A random variable is said to be discrete if its range is a discrete set.

Remarks:

- The range of a random variable is the set of values taken by it and is a subset of the real line.
- The discrete sets are as follows:
 - (i) All finite subsets of the real line are discrete.
 - (ii) All subsets of integers are discrete.
 - (iii) The set of all integer multiples of a real number is discrete.
 - (iv) Any subset of a discrete set is discrete.
- The non-discrete sets are as follows:
 - (i) Any interval is not discrete, i.e. (a, b) is not discrete for $a < b$.
 - (ii) Any subset that contains a non-discrete set is not discrete.

Examples:

(1) Tossing a coin: $S = \{H, T\}$.

A few possible valid random variables are:

- Random variable $X : X(H) = 0, X(T) = 1$
Range $(X) = \{0, 1\}$
- Random variable $Y : Y(H) = -10, Y(T) = 200$
Range $(Y) = \{-10, 200\}$
- Random variable $Z : Z(H) = \sqrt{2}, Z(T) = \pi$
Range $(Z) = \{\sqrt{2}, \pi\}$
- Random variable $U : U(H) = 0, U(T) = 0$
Range $(U) = \{0\}$

(2) Throw a die: $S = \{1, 2, 3, 4, 5, 6\}$

A random variable can be defined as follows:

- Pick any 6 real numbers, say $x_1, x_2, x_3, x_4, x_5, x_6$.
- $X(1) = x_1, X(2) = x_2, X(3) = x_3, X(4) = x_4, X(5) = x_5, X(6) = x_6$
i.e. Range $(X) = \{x_1, x_2, x_3, x_4, x_5, x_6\}$

However, x_i 's need not be distinct.

Suppose $x_2 = x_4 = x_6 = 1$ and $x_1 = x_3 = x_5 = 0$, then

Random variable $E : E(2) = E(4) = E(6) = 1, E(1) = E(3) = E(5) = 0$

But if x_i 's are distinct, then the random variable is a one-to-one function.

Note:

- For an experiment, a lot of valid random variables are possible but usually, the meaningful functions are considered, like random variable X in the first example.
- In practice, the random variables are easier to assign probabilities and easier to work with.

2.8.1 Random Variable and events

The technical conditions for a subset of sample space to be an event is that if X is a random variable associated to the experiment then, the event should be expressible in terms of X .

For any real x , a few examples of expression of events in terms of random variable X are as follows:

- (i) $(X < x) = \{s \in S : X(s) < x\}$
i.e. set of all outcomes for which X takes value less than x .
- (ii) $(X > x) = \{s \in S : X(s) > x\}$
i.e. set of all outcomes for which X takes value greater than x .
- (iii) $(X = x) = \{s \in S : X(s) = x\}$
i.e. set of all outcomes for which X takes value equal to x .
- (iv) $(X \leq x) = \{s \in S : X(s) \leq x\}$
i.e. set of all outcomes for which X takes value less than or equal to x .
- (v) $(X \geq x) = \{s \in S : X(s) \geq x\}$
i.e. set of all outcomes for which X takes value greater than or equal to x .
- (vi) $(x_1 < X < x_2) = \{s \in S : x_1 < X(s) < x_2\} \quad ; \quad x_1, x_2 \in \mathbf{R}$
i.e. set of all outcomes for which X takes value greater than x_1 but less than x_2 .

Examples:

- (1) Rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$
Random variable $X : X(1) = 1, X(2) = 2, X(3) = 3, X(4) = 4, X(5) = 5, X(6) = 6$
Range $(X) = \{1, 2, 3, 4, 5, 6\}$
Any event E can be expressed in terms of X as follows:
 - $E : \{\text{The die shows 1}\} : (X = 1)$

- $E : \{1, 2, 3\} : (X < 4)$
 - $E : \{2, 5\} : (X = 2) \cup (X = 5)$
- (2) Rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$
Random variable $X : X(2) = X(4) = X(6) = 1, X(1) = X(3) = X(5) = 0$
Range $(X) = \{0, 1\}$

Any event E can be expressed in terms of X as follows:

- $E : \{1, 3, 5\} : (X = 0)$
- $E : \{2, 4, 6\} : (X = 1)$
- $E : \{\text{Null Event}\} : (X < 0)$
- $E : \{1, 2, 3, 4, 5, 6\} : (X \leq 1)$

Remark: Not all events can be written in terms of X , like $E : \{2, 5\}$.

Hence, due to the technical restriction, this subset of sample space cannot be considered as an event.

2.8.2 Distribution of a Discrete Random Variable

2.8.2.1 Probability Mass Function (PMF)

The probability mass function (PMF) of a discrete random variable X with range set T is the function $f_X : T \rightarrow [0, 1]$ defined as:

$$f_X(t) = P(X = t) \text{ for } t \in T$$

In general, the probability mass function is the probability that the random variable takes one particular value.

The probability of any event, defined using the random variable X , can be computed by using the PMF as follows:

- Let $(X = t)$ be an event then,

$$P(X = t) = P(\text{all outcomes that result in } X \text{ taking value } t)$$

- Let B be a subset of range of random variable T and $(X \in B)$ be an event, then

$$\begin{aligned} P(X \in B) &= P(\text{all outcomes that result in } X \text{ taking values in } B) \\ &= \sum_{t \in B} P(X = t) \\ &= \sum_{t \in B} f_X(t) \end{aligned}$$

Note: A discrete random variable X with range $T = \{t_1, t_2, \dots, t_k\}$ and PMF f_X can be anticipated in tabular form as follows:

t	t_1	t_2	\dots	t_k
$f_X(t)$	$f_X(t_1)$	$f_X(t_2)$	\dots	$f_X(t_k)$

where, the first row is the value that the random variable takes and the second row is the probability with which the random variable takes that value.

2.8.2.2 Properties of PMF

Consider a random variable X with range $T = \{t_1, t_2, \dots, t_k\}$ and PMF f_X . The defining properties of PMF are:

- (1) $0 \leq f_X(t) \leq 1$, because $f_X(t) = P(X = t)$ is a probability.
- (2) $\sum_{x \in T} f_X(t) = 1$, because $P(X \in T) = 1$

2.8.3 Solved Examples:

Q1. The PMF of a random variable X is given in the following table:

t	-1	1	2	4
$f_X(t)$	0.5	0.25	0.125	

Compute the following:

- (i) Range of X
- (ii) $f_X(4)$
- (iii) $P(X > 3)$
- (iv) $P\left(X < \frac{3}{2}\right)$

Solution:

- (i) The range of a random variable is the set of values taken by it, i.e.
 $T_X = \{-1, 1, 2, 4\}$

- (ii) By the property of PMF, we know that $\sum_{x \in T} f_X(t) = 1$.

$$\implies f_X(-1) + f_X(1) + f_X(2) + f_X(4) = 1$$

$$\implies 0.5 + 0.25 + 0.125 + f_X(4) = 1$$

$$\implies 0.875 + f_X(4) = 1$$

$$\therefore f_X(4) = 0.125$$

(iii) Let $(X \in B)$ be an event such that $B = (X > 3) = \{4\}$ then,

$$\begin{aligned} P(X > 3) &= \sum_{t \in B} P(X = t) \\ &= P(X = 4) \\ \therefore P(X > 3) &= 0.125 \end{aligned}$$

(iv) Let $(X \in B)$ be an event such that $B = \left(X < \frac{3}{2}\right) = \{-1, 1\}$ then,

$$\begin{aligned} P\left(X < \frac{3}{2}\right) &= \sum_{t \in B} P(X = t) \\ &= P(X = -1) + P(X = 1) \\ &= 0.5 + 0.25 \\ \therefore P\left(X < \frac{3}{2}\right) &= 0.75 \end{aligned}$$

Q2. The PMF of a random variable X is given by:

$$f_X(k) = \frac{c}{3^k} \quad ; \quad k = 1, 2, 3, \dots$$

Compute the following:

- (i) c
- (ii) $P(X > 10)$
- (iii) $P(X > 10 | X > 5)$

Solution:

- (i) The range of random variable X is not finite but countable, i.e. $T = \{1, 2, 3, \dots\}$

By property of PMF, we know that $\sum_{x \in T} f_X(t) = 1$

$$\begin{aligned} \implies f_X(1) + f_X(2) + f_X(3) + \dots &= 1 \\ \implies \frac{c}{3} + \frac{c}{3^2} + \frac{c}{3^3} + \dots &= 1 \\ \implies \frac{c}{3} \left(1 + \frac{1}{3} + \frac{1}{3^2} + \dots\right) &= 1 \end{aligned}$$

Using the sum of infinite GP series, we get:

$$\begin{aligned} \frac{c}{3} \left(\frac{1}{1 - \frac{1}{3}}\right) &= 1 \\ \therefore c &= 3 \times \frac{2}{3} = 2 \end{aligned}$$

(ii)

$$\begin{aligned} P(X > 10) &= P(X = 11) + P(X = 12) + P(X = 13) + \dots \\ &= \frac{c}{3^{11}} + \frac{c}{3^{12}} + \frac{c}{3^{13}} + \dots \\ &= \frac{c}{3^{11}} \times \left(1 + \frac{1}{3} + \frac{1}{3^2} + \dots \right) \\ &= \frac{c}{3^{11}} \times \left(\frac{1}{1 - \frac{1}{3}} \right) \quad ; \quad (\text{Sum of infinite GP series}) \\ &= \frac{2}{3^{11}} \times \frac{3}{2} \quad ; \quad (\text{because, } c = 2) \\ \therefore P(X > 10) &= \frac{1}{3^{10}} \end{aligned}$$

(iii) By using the conditional probability, we get:

$$\begin{aligned} P(X > 10 | X > 5) &= \frac{P[(X > 10) \cap (X > 5)]}{P(X > 5)} \\ &= \frac{P(X > 10)}{P(X > 5)} \end{aligned} \tag{1}$$

Now,

$$\begin{aligned} P(X > 5) &= P(X = 6) + P(X = 7) + P(X = 8) + \dots \\ &= \frac{c}{3^6} + \frac{c}{3^7} + \frac{c}{3^8} + \dots \\ &= \frac{c}{3^6} \times \left(1 + \frac{1}{3} + \frac{1}{3^2} + \dots \right) \\ &= \frac{c}{3^6} \times \left(\frac{1}{1 - \frac{1}{3}} \right) \quad ; \quad (\text{Sum of infinite GP series}) \\ &= \frac{2}{3^6} \times \frac{3}{2} \quad ; \quad (\text{because, } c = 2) \\ \therefore P(X > 5) &= \frac{1}{3^5} \end{aligned}$$

Putting the value of $P(X > 5)$ in (1), we get:

$$\begin{aligned} P(X > 10 | X > 5) &= \frac{P(X > 10)}{P(X > 5)} \\ &= \frac{\left(\frac{1}{3^{10}} \right)}{\left(\frac{1}{3^5} \right)} \\ \therefore P(X > 10 | X > 5) &= \frac{1}{3^5} \end{aligned}$$

Q3. A fair coin is flipped three times.

- (i) Associate a random variable to the number of heads and compute its PMF.
- (ii) Associate a random variable to the flip number (if any) that shows the first head and compute its PMF.

Solution:

Let us define the following random variable:

X : Number of heads in three flips of a fair coin

Y : Flip number that shows the first head

Sample space; $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

Since the same fair coin is flipped thrice, the distribution is uniform.

Outcome	P(Outcome)	X	Y
HHH	$1/8$	3	1
HHT	$1/8$	2	1
HTH	$1/8$	2	1
HTT	$1/8$	1	1
THH	$1/8$	2	2
THT	$1/8$	1	2
TTH	$1/8$	1	3
TTT	$1/8$	0	None

- (i) Range (X) ; $T = \{0, 1, 2, 3\}$

Now,

- $f_X(0) = P(\{TTT\}) = \frac{1}{8}$
- $f_X(1) = P(\{HTT, THT, TTH\})$
 $= P(\{HTT\}) + P(\{THT\}) + P(\{TTH\})$; (because, disjoint events)
 $= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$
- $f_X(2) = P(\{HHT, HTH, THH\})$
 $= P(\{HHT\}) + P(\{HTH\}) + P(\{THH\})$; (because, disjoint events)
 $= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$
- $f_X(3) = P(\{HHH\}) = \frac{1}{8}$

Hence,

t	0	1	2	3
$f_X(t)$	$1/8$	$3/8$	$3/8$	$1/8$

(ii) Range (Y) ; $T = \{0, 1, 2, 3\}$ (Denoting 'None' as 0)

Now,

- $f_Y(0) = P(\{TTT\}) = \frac{1}{8}$
- $f_Y(1) = P(\{HHH, HHT, HTH, HTT\})$
 $= P(\{HHH\}) + P(\{HHT\}) + P(\{HTH\}) + P(\{HTT\})$; (because, disjoint events)
 $= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{4}{8}$
- $f_Y(2) = P(\{THH, THT\})$
 $= P(\{THH\}) + P(\{THT\})$; (because, disjoint events)
 $= \frac{1}{8} + \frac{1}{8} = \frac{2}{8}$
- $f_Y(3) = P(\{TTH\}) = \frac{1}{8}$

Hence,

t	0	1	2	3
$f_Y(t)$	$1/8$	$4/8$	$2/8$	$1/8$

Q4. For a certain lottery, a three-digit number is randomly selected (from 000 to 999). If a ticket matches the number exactly, it is worth 2 lakhs. If the ticket matches exactly two of the three digits, it is worth ₹20000. Otherwise, it is worth nothing. Let X be the value of the ticket. Find the distribution of X .

Solution:

There are a total of 1000 tickets numbered from 000 to 999.

Range (X) ; $T = \{0, 20000, 200000\}$

- $f_X(200000) = P(\text{All three digits are matched})$
 $= \frac{1}{1000}$
- $f_X(20000) = P(\text{Any two of three digits matched})$
(Since any two of three digits can be same and third one can be any of remaining 9)
 $f_X(20000) = \frac{{}^3C_2 \times 1^2 \times 9}{1000}$
 $\therefore f_X(20000) = \frac{27}{1000}$

- $f_X(0) = 1 - f_X(20000) - f_X(200000)$
(By using the property of PMF)

$$\implies f_X(0) = 1 - \frac{1}{1000} - \frac{27}{1000}$$

$$f_X(0) = 1 - \frac{28}{1000}$$

$$\therefore f_X(0) = \frac{972}{1000}$$

2.9 Common Discrete Distributions

2.9.1 Uniform Random Variable

A random variable X follows Uniform distribution with parameter T , i.e. $X \sim U(T)$, where T is some finite set, then:

- Range $(X) = T$
- PMF $f_X(t) = \frac{1}{|T|}$ for all $t \in T$ and $|T| = \text{Size of finite set } T$

Note: If all the outcomes of a random variable are equally likely, then it can be assumed to be uniform.

Examples:

- Tossing a fair coin: $S = \{H, T\}$.
Random variable $X : X(H) = 0, X(T) = 1$
 $\implies X \sim U(\{0, 1\})$
- Throw a fair die: $S = \{0, 1, 2, 3, 4, 5, 6\}$.
Random variable $X : X(1) = 1, X(2) = 2, X(3) = 3, X(4) = 4, X(5) = 5, X(6) = 6$
 $\implies X \sim U(\{1, 2, 3, 4, 5, 6\})$

2.9.2 Bernoulli Random Variable

A random variable X follows Bernoulli distribution with parameter p , i.e. $X \sim \text{Bernoulli}(p)$, where $p = \text{probability of success}$ and $0 \leq p \leq 1$, then:

- Range $(X) ; T = \{0, 1\}$
- PMF $f_X(t) = p^t(1 - p)^{1-t}$ for all $t \in T$

It is also referred to as binary in some sense because it takes only two values, 0 and 1. Refer section **2.7.1** for more details.

2.9.3 Binomial Random Variable

A random variable X follows Binomial distribution with parameter n and p , i.e.

$X \sim \text{Binomial}(n, p)$, where

p = Probability of success ; $0 \leq p \leq 1$

n = Number of independent Bernoulli trials ; n : Positive integer

then:

- Range (X) ; $T = \{0, 1, 2, 3, \dots, n\}$
- PMF $f_X(t) = {}^nC_t p^t (1-p)^{n-t}$ for all $t \in T$

Refer section 2.7.2 for more details.

2.9.4 Geometric Random Variable

A random variable X follows Geometric distribution with parameter p , i.e.

$X \sim \text{Geometric}(p)$, where p = probability of success and $0 \leq p \leq 1$ then:

- Range (X) = T
- PMF $f_X(t) = \frac{1}{|T|}$ for all $t \in T$ and $|T|$ = Size of finite set T

Refer section 2.7.3 for more details.

2.9.5 Negative Binomial Distribution

Consider a Bernoulli(p) trial with $S = \{0, 1\}$

If this Bernoulli(p) trial is repeated independently, then the number of trials needed for the first r successes is given by the Negative Binomial distribution as follows:

- Sample space ; $S = \{r, r+1, r+2, r+3, \dots\}$
Because we can get the first r successes in r trials, $r+1$ trials, $r+2$ trials, \dots , so on.
- $P(NB(r, p) = k) = {}^{k-1}C_{r-1} (1-p)^{k-r} p^r$; $k = r, r+1, r+2, \dots$

Note: Negative binomial distribution is a generalization of the Geometric distribution. On plugging $r = 1$ in the PMF of negative binomial distribution, one can obtain the geometric distribution.

2.9.5.1 Negative Binomial Random Variable

A random variable X follows Negative Binomial distribution with parameters r and p , i.e.

$X \sim \text{Negative Binomial}(r, p)$, where

$p = \text{Probability of success ; } 0 \leq p \leq 1$

$r = \text{Number of successes ; } r : \text{Positive integer}$

then:

- Range (X) ; $T = \{r, r + 1, r + 2, r + 3, \dots\}$
- PMF $f_X(t) = {}^{k-1}C_{r-1} p^r (1 - p)^{k-r}$ for all $t \in T$

2.9.6 Poisson Distribution

Consider an event that keeps occurring over a period of time with the following two assumptions:

1. The rate of occurrence (λ) of the event is constant.
2. Given one occurrence, the time for the next occurrence is not affected by it, i.e. independent occurrences of the event.

Then the number of times the event occurred in a fixed interval of time is given by Poisson distribution as follows:

- Sample space ; $S = \{0, 1, 2, 3, \dots\}$
Because, in a given interval of time the event can occur 0 times, 1 time, 2 times, ..., so on.
- $P(P(\lambda) = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad ; \quad k = 0, 1, 2, 3, \dots$

2.9.6.1 Application of Poisson Distribution

- (1) Emission of a particle by radioactive decay.

In 2608 time intervals of 7.5 seconds each, emission of alpha particles is as follows:

No. of particles	0	1	2	3	4	5	6	7	8	9	10
No. of times	57	203	383	525	532	408	273	139	45	27	16
Fraction	0.022	0.078	0.147	0.201	0.204	0.156	0.105	0.053	0.017	0.0104	0.0061

The rate of occurrence of the event, i.e.

$$\begin{aligned}
 \text{Emission rate ; } \lambda &= \frac{\text{Total number of particles}}{\text{Number of intervals}} \\
 &= \frac{(0 \times 57) + (1 \times 203) + (2 \times 383) + \dots + (10 \times 16)}{2608} \\
 &= \frac{10086}{2608} \\
 \therefore \lambda &= 3.8673
 \end{aligned}$$

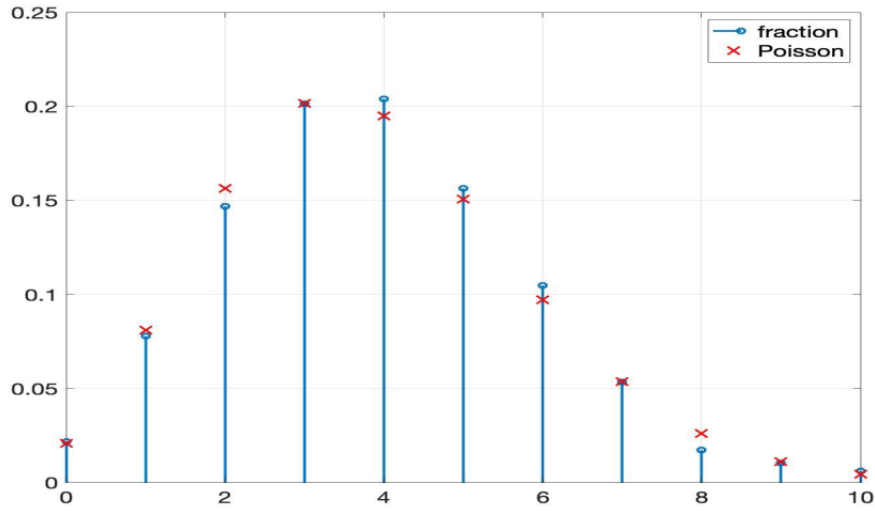
This implies that there are 3.8673 emission of particles per 7.5 seconds.

Here, the emission rate can be assumed to be constant, and also it is reasonable to assume that the time of next emission is independent of the past emissions.

Hence, by using the Poisson distribution, we get:

$$P(\text{No. of Particles} = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

The probabilities computed using the Poisson distribution are close to the fractions given in the Table.



Visually, this implies that the above assumptions holds true and the Poisson distribution fits the data, i.e. the chances that k particles are emitted in a given time interval of 7.5 seconds can be obtained using the Poisson distribution.

(2) A meteorite entering earth's atmosphere

In 276 months from 1994-2016, the number of fireballs entering the atmosphere is as follows:

No. of fireballs	0	1	2	3	4	5	6	7	8	9	10
No. of times	24	54	76	52	37	21	7	3	1	0	1
Fraction	0.087	0.196	0.275	0.188	0.134	0.076	0.025	0.0109	0.0036	0	0.0036

The rate of occurrence of the event, i.e.

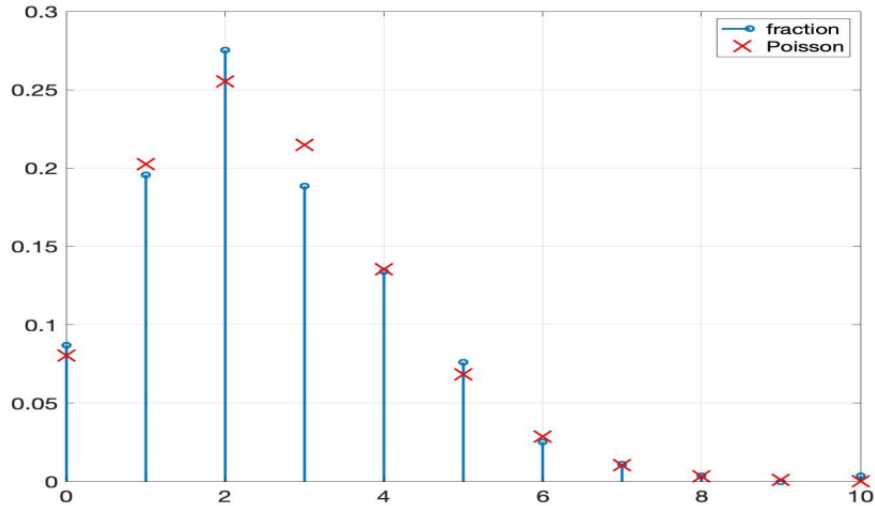
$$\begin{aligned}
 \text{Arrival rate ; } \lambda &= \frac{\text{Total number of fireballs}}{\text{Number of intervals}} \\
 &= \frac{(0 \times 24) + (1 \times 54) + (2 \times 76) + \dots + (10 \times 1)}{276} \\
 &= \frac{696}{276} \\
 \therefore \lambda &= 2.5217
 \end{aligned}$$

This implies that there are 2.5217 fireballs per month.

Here, the arrival rate can assumed to be constant, and also it is reasonable to assume that the time of next arrival is independent of the past arrival of fireballs. Hence, by using the Poisson distribution, we get:

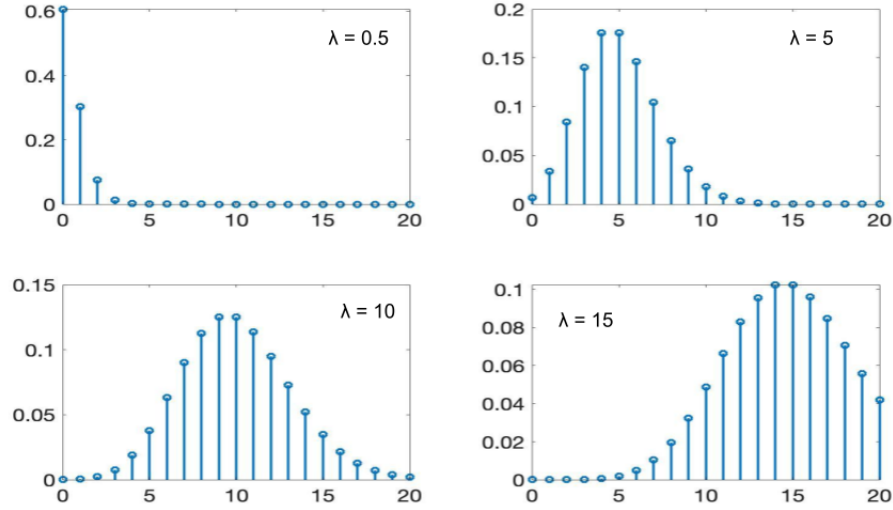
$$P(\text{No. of fireballs} = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

The probabilities computed using the Poisson distribution are close to the fractions given in the Table.



Visually, this implies that the above assumptions holds true and the Poisson distribution fits the data, i.e. the chances that k fireballs arrived in a month can be obtained using the Poisson distribution.

2.9.6.2 Visualtion of Poisson Distribution



Note: As the value of λ increases, the peak is shifted from left to right.

2.9.6.3 Poisson Random Variable

A random variable X follows Poisson distribution with parameter λ , i.e. $X \sim \text{Poisson}(\lambda)$, where $\lambda > 0$ then:

- Range (X) ; $T = \{0, 1, 2, 3, \dots\}$
- PMF $f_x(t) = e^{-\lambda} \times \frac{\lambda^t}{t!}$ for all $t \in T$

2.9.7 Hypergeometric Distribution

Consider a population of N persons of which r are of Type 1 and the remaining $N - r$ are of Type 2. If m persons are selected uniformly at random, without replacement, then the number of Type 1 persons selected is given by Hypergeometric distribution as follows:

- Sample space ; $S = \{\max(0, m - (N - r)), \dots, \min(r, m)\}$

Because, we can select $0, 1, 2, \dots, m$ persons of Type 1, but:

- If the number of persons to be selected is more than the persons of Type 1, i.e. $m > r$, then the maximum number of persons of Type 1 selected is r only.
- If the number of persons to be selected is more than persons of Type 2, i.e. $m > N - r$, then the minimum number of persons of Type 1 selected is $m - (N - r)$ only.

For more details, refer to examples of section **2.9.7.1**

- $P(H(N, r, m) = k) = \frac{{}^r C_k \times {}^{N-r} C_{m-k}}{{}^N C_m} \quad ; \quad k = \max(0, m - (N - r)), \dots, \min(r, m)$

2.9.7.1 Hypergeometric Random Variable

A random variable X follows Hypergeometric distribution with parameter N, r and m , i.e. $X \sim \text{Hypergeometric}(N, r, m)$, where

N = Population size ; N : Positive integer

r = Number of Type 1 entities in population ; r : Positive integer

m = Number of entities to be selected from population ; m : Positive integer

then:

- Range (X) ; $T = \{\max(0, m - (N - r)), \dots, \min(r, m)\}$

- PMF $f_X(t) = \frac{{}^r C_t \times {}^{N-r} C_{m-t}}{{}^N C_m}$ for all $t \in T$

Examples:

- If $N = 100, r = 50, m = 20 \implies \text{Range}(X) = \{0, 1, 2, \dots, 20\}$
- If $N = 100, r = 10, m = 20 \implies \text{Range}(X) = \{0, 1, 2, \dots, 10\}$
- If $N = 100, r = 90, m = 20 \implies \text{Range}(X) = \{10, 11, 12, \dots, 20\}$

2.10 Functions of one discrete random variable

Consider a random variable X with range T and PMF $f_X(t)$.

Let $g(X)$ be a function from \mathbb{R} to real line. Then, $g(X)$ can be seen as a composite of two functions, as:

- X is a function from sample space to T (subset of real line).
- $g(X)$ is a function from T to another subset of real line.

So, $g(X)$ is also a random variable in the same probability space.

Remarks:

If X is a random variable with range T and PMF $f_X(t)$, then the PMF of function of the random variable X , i.e. $g(X)$ can be found using $f_X(t)$ as follows:

$$\begin{aligned} f_{g(X)}(a) &= P(g(X) = a) \\ &= P(X \in \{t : g(t) = a\}) \\ \therefore f_{g(X)}(a) &= \sum_{t: g(t)=a} f_X(t) \end{aligned}$$

2.10.1 Solved Examples:

Q1. Let $X \sim \text{Uniform}(\{-2, -1, 0, 1, 2\})$ and $g(x) = x^2$. Find:

- (i) Range of $g(X)$
- (ii) PMF of $g(X)$

Solution:

Range (X) ; $T = \{-2, -1, 0, 1, 2\}$

$f_X(t) = \frac{1}{5}$ for all $t \in T$; (Since $X \sim \text{Uniform}$)

t	$f_X(t)$	$g(t) = t^2$
-2	$\frac{1}{5}$	4
-1	$\frac{1}{5}$	1
0	$\frac{1}{5}$	0
1	$\frac{1}{5}$	1
2	$\frac{1}{5}$	4

- (i) Range $g(X) = \{0, 1, 4\}$
- (ii) PMF of $g(X) = X^2$ is as follows:

- $$\begin{aligned} f_{g(X)}(0) &= P(X = 0) \\ &= \frac{1}{5} \end{aligned}$$
- $$\begin{aligned} f_{g(X)}(1) &= P(X \in \{-1, 1\}) \\ &= P(X = 1) + P(X = -1) \\ &= \frac{1}{5} + \frac{1}{5} = \frac{2}{5} \end{aligned}$$
- $$\begin{aligned} f_{g(X)}(4) &= P(X \in \{-2, 2\}) \\ &= P(X = 2) + P(X = -2) \\ &= \frac{1}{5} + \frac{1}{5} = \frac{2}{5} \end{aligned}$$

Hence,

a	0	1	4
$f_{g(X)}(a)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{5}$

Q2. Let $X \sim \text{Geometric}\left(\frac{1}{2}\right)$ and $g(X) = \begin{cases} x & \text{for } x < 5 \\ 5 & \text{for } x \geq 5 \end{cases}$

Find:

- (i) Range of $g(X)$
- (ii) PMF of $g(X)$

Solution:

Range (X) ; $T = \{1, 2, 3, 4, 5, 5, 5, \dots\}$

$$f_X(t) = \frac{1}{2} \times \left(\frac{1}{2}\right)^{t-1} \text{ for all } t \in T \quad ; \quad (\text{Because } X \sim \text{Geometric})$$

t	$f_X(t)$	$g(t)$
1	$1/2$	1
2	$1/2^2$	2
3	$1/2^3$	3
4	$1/2^4$	4
5	$1/2^5$	5
6	$1/2^6$	5
7	$1/2^7$	5
\vdots	\vdots	\vdots

- (i) Range $g(X) = \{1, 2, 3, 4, 5\}$
- (ii) The PMF of $g(X)$ is as follows:

- $$\begin{aligned} f_{g(X)}(1) &= P(X = 1) \\ &= \frac{1}{2} \end{aligned}$$

- $$\begin{aligned} f_{g(X)}(2) &= P(X = 2) \\ &= \frac{1}{2^2} \\ \therefore f_{g(X)}(2) &= \frac{1}{4} \end{aligned}$$

- $$\begin{aligned} f_{g(X)}(3) &= P(X = 3) \\ &= \frac{1}{2^3} \\ \therefore f_{g(X)}(3) &= \frac{1}{8} \end{aligned}$$

- $$\begin{aligned}
 f_{g(X)}(4) &= P(X = 4) \\
 &= \frac{1}{2^4} \\
 \therefore f_{g(X)}(4) &= \frac{1}{16}
 \end{aligned}$$
- $$\begin{aligned}
 f_{g(X)}(5) &= P(X \in \{5, 6, 7, \dots\}) \\
 &= P(X \geq 5) \\
 &= 1 - P(X \leq 4) \\
 &= 1 - \frac{15}{16} \\
 \therefore f_{g(X)}(5) &= \frac{1}{16}
 \end{aligned}$$

Hence,

a	1	2	3	4	5
$f_{g(X)}(a)$	$1/2$	$1/4$	$1/8$	$1/16$	$1/16$