

Probability Models

Prof. Abhinava Tripathi

NPTEL



Introduction



Introduction

- Probability models empower statisticians to draw inferences from sample data
- The average salary of a data scientist in India is INR 25 Lakhs plus-minus INR 10 Lakh , with 90% of confidence
- How do we get to these conclusions?
- Do we extensively examine the entire population of data scientists



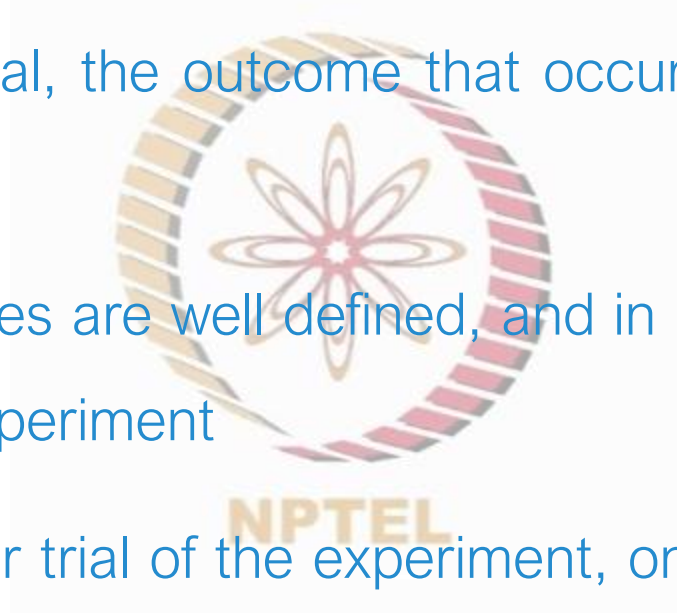
Random Experiments and Sample Space



Random Experiments

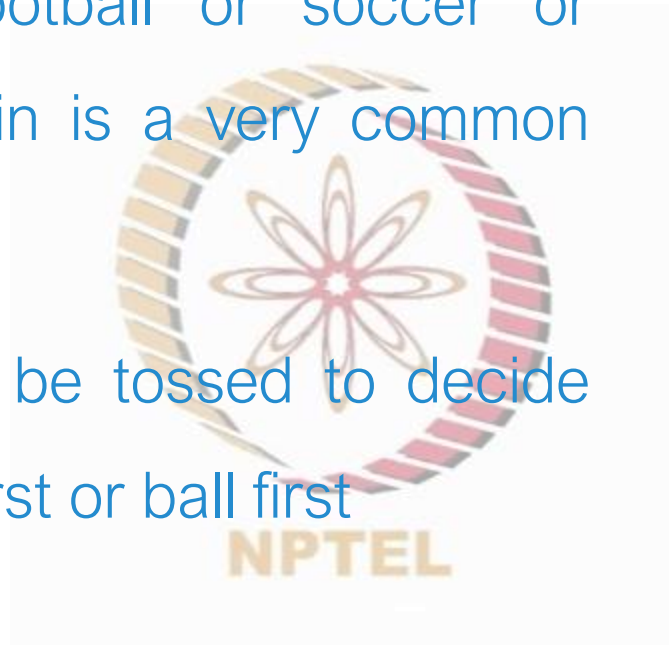
A **random experiment** is a process that generates well-defined experimental outcomes. On any single repetition or trial, the outcome that occurs is determined completely by chance

1. The experimental outcomes are well defined, and in many cases can even be listed prior to conducting the experiment
2. On any single repetition or trial of the experiment, one and only one of the possible experimental outcomes will occur
3. The experimental outcome that occurs on any trial is determined solely by chance



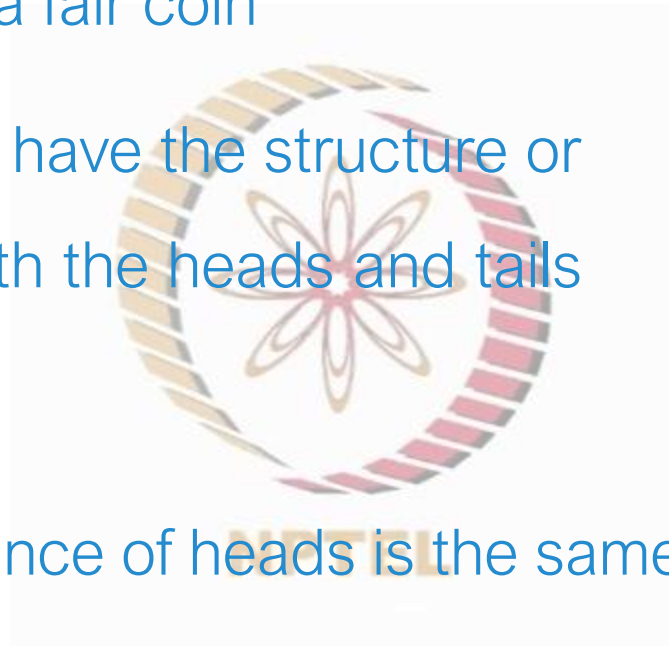
Random Experiments and Sample Space

- Whether it's cricket, football or soccer or basketball, tossing a coin is a very common feature
- In cricket, a coin might be tossed to decide which team gets to bat first or ball first




Random Experiments and Sample Space

- Let's take an example of a fair coin
- A fair coin is a coin that'll have the structure or the make identical for both the heads and tails side of the coin
- When you toss it, the chance of heads is the same as the chance of tails
- The captain of the Indian team calls for heads.
- What do you think will be the chance that the captain wins?



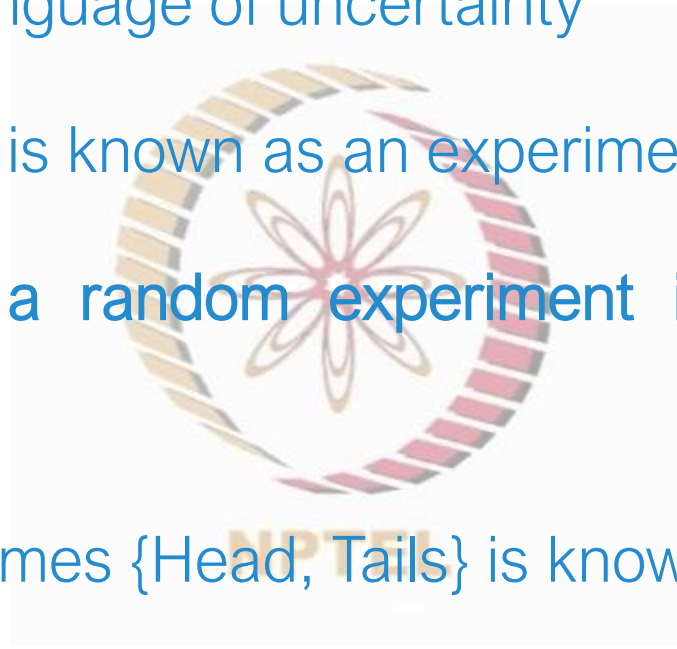
Random Experiments and Sample Space

- You can have only 2 outcomes: heads or tails
- The chance that he wins is computed like this:

$$\frac{1(\text{heads})}{2(\text{heads and tails})} = 0.5$$
- The chances of something happening is known as the probability



Random Experiments and Sample Space

- Probability is really the language of uncertainty
- The coin toss game here is known as an experiment
- The sample space for a random experiment is the set of all experimental outcomes
- The set of possible outcomes {Head, Tails} is known as the sample space
- Sample space can be mathematically represented as Ω (omega) or $S = \{\text{Heads}, \text{Tails}\}$
- In our example, the probability of the captain winning is known as the event

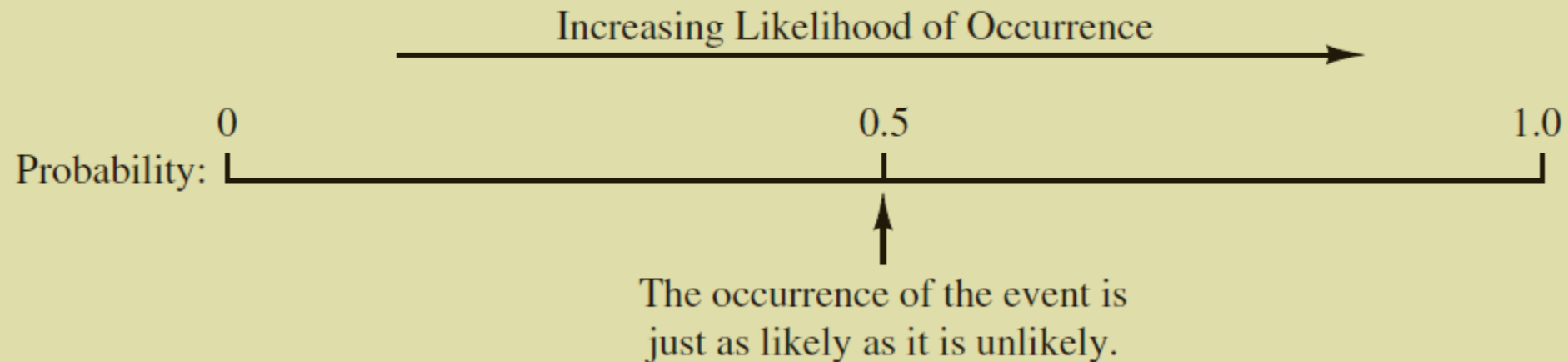


Probability Events



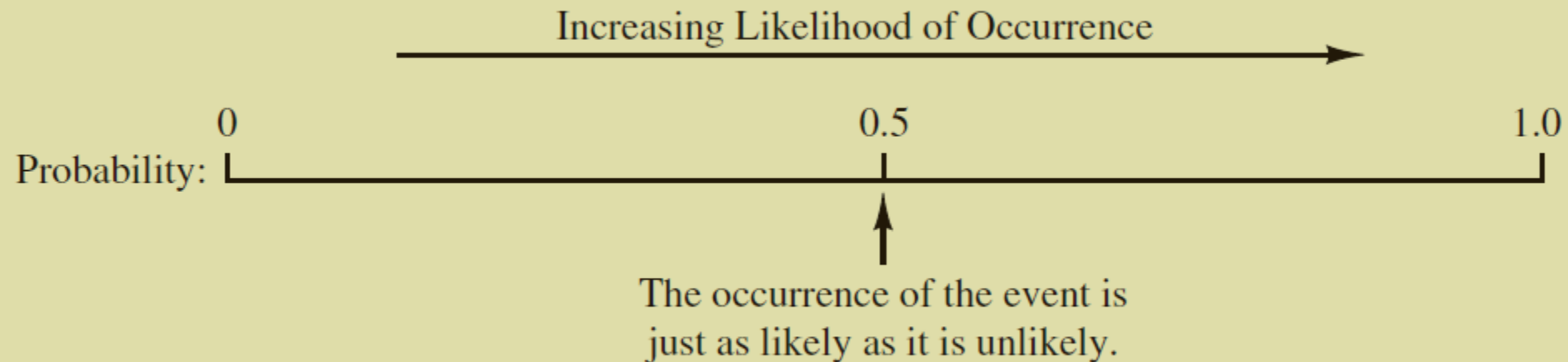
Probability Events

- Probability is a numerical measure of the likelihood that an event will occur
- Thus, probabilities can be used as measures of the degree of uncertainty associated with an event



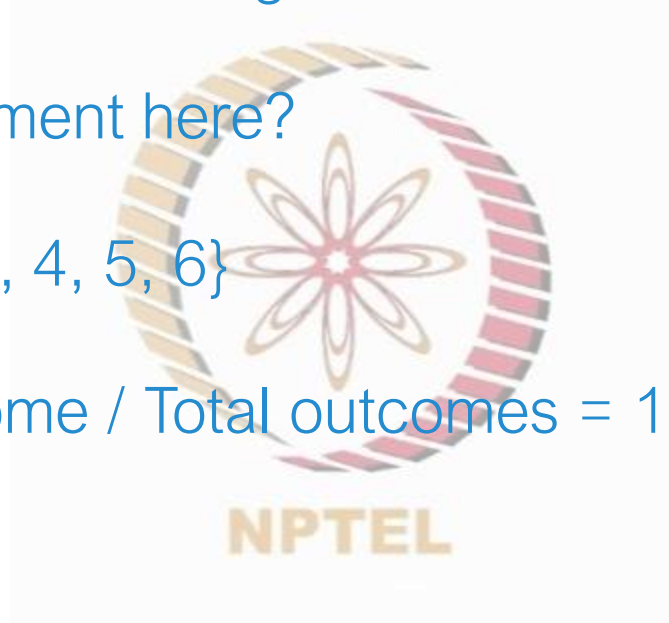
Probability Events

- Probability values are always assigned on a scale from 0 to 1
- A probability near zero indicates an event is unlikely to occur; a probability near 1 indicates an event is almost



Probability Events

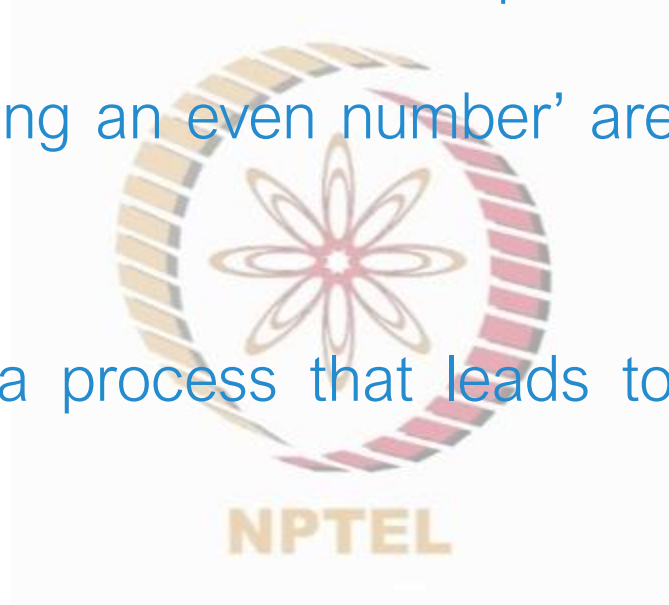
- Let us consider the example of rolling dice with 6 faces
- What is the random experiment here?
- Sample space: $S = \{1, 2, 3, 4, 5, 6\}$
- Number of favorable outcome / Total outcomes = $1(\text{getting a } 1) / 6(1, 2, 3, 4, 5, 6) = 1/6$
- What about the probability of getting an even number, that is, 2, 4, 6: $3/6 = 0.5$.



Probability Events



- In this example, Rolling a dice is a random experiment
- 'Rolling a one' or the 'Rolling an even number' are the events for which we find the probability
- A random experiment is a process that leads to one of several possible outcomes
- The probability formula we discussed may not apply at all the places

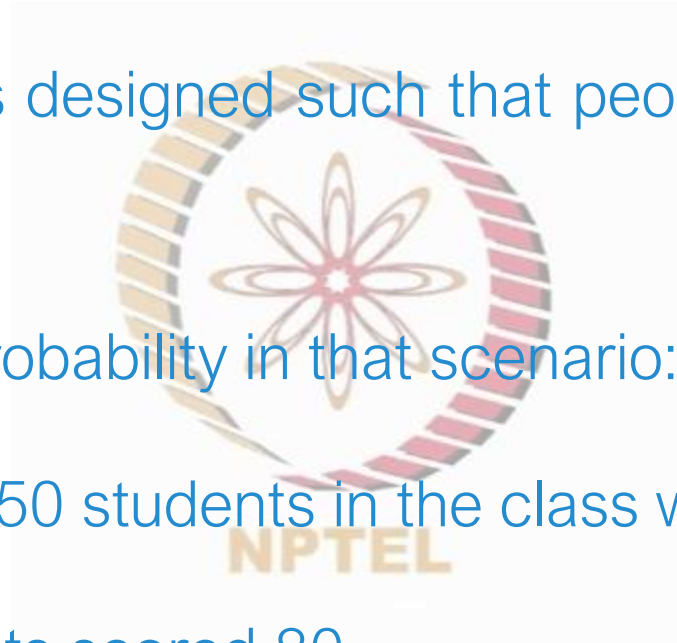


Probability Events

- Let's consider the marks scored out of 100 by a student in a math test
- What can be the outcomes of the test: any number between 0 and 100
- If evaluation is in whole numbers, the sample space becomes the set of whole numbers between 0 and 100, which is $\{0, 1, 2, \dots, 100\}$
- What is the probability of the event of "Student scoring 80 in a math test"
- According to our formula: Number of favorable outcomes / Total outcomes = $1(\text{getting } 80) / 101(0 \text{ to } 100) = 1/101$
- Does it look right?

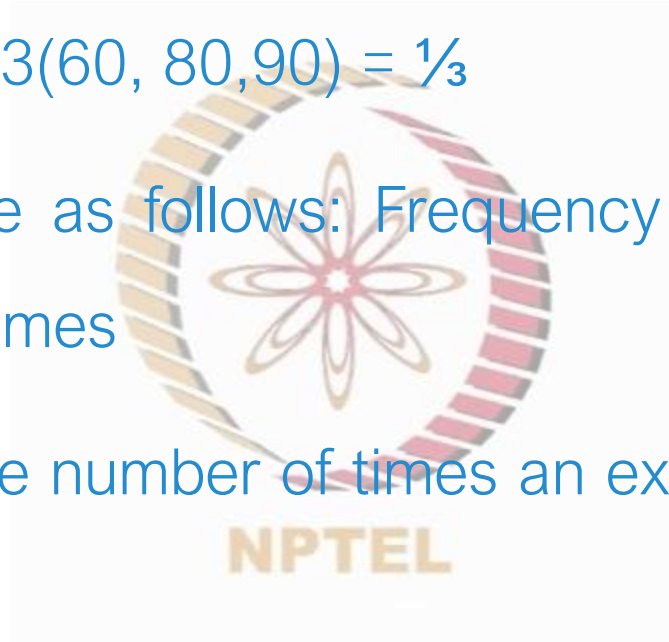
Probability Events

- Not all experiments are designed to make all outcomes equally likely
- You can have a test that is designed such that people get at least a passing grade
- How would we calculate probability in that scenario: look past data
- For example, if there were 50 students in the class who took the math test
 - Out of which 10 students scored 80
 - 10 students scored 90 and the rest scored 60
- Then what is the probability of a student scoring 80 marks?



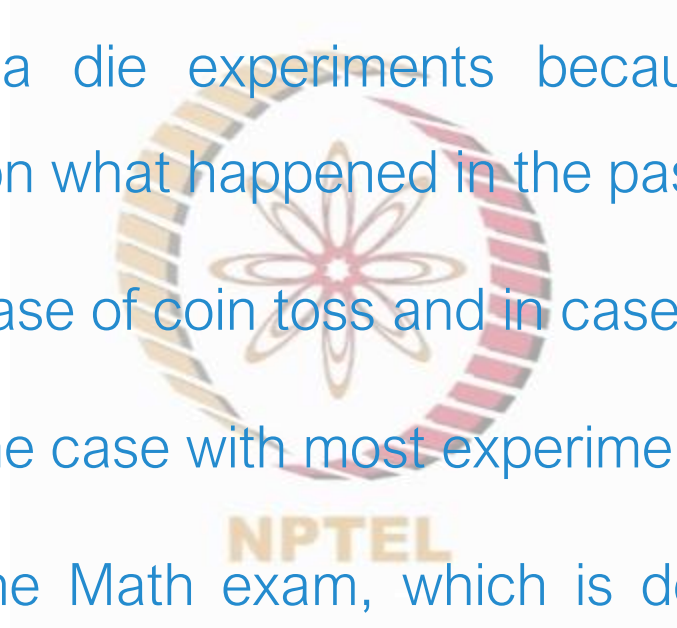
Probability Events

- Incorrect to use the formula number of favorable outcomes divided by Total outcomes = 1(getting 80)/ 3(60, 80,90) = $\frac{1}{3}$
- The formula changes here as follows: Frequency of favourable outcome / Total frequency of all outcomes
- Frequency is essentially the number of times an experiment is repeated with the outcome
- In our case, the experiment is repeated for 50 students and the Frequency of favorable outcomes is 10: probability is $10/50 = 0.20$



Probability Events

- The formula: “Number of favorable outcome divided by Total outcomes” worked for the coin toss and rolling a die experiments because different outcomes of experiments do not depend on what happened in the past
- The different outcomes the case of coin toss and in case of die roll are equally likely
- However, this is usually not the case with most experiments
- That is not the case with the Math exam, which is designed in a way that one outcome, say, scoring 0, is not as likely as scoring a 100, or scoring 60 or 80

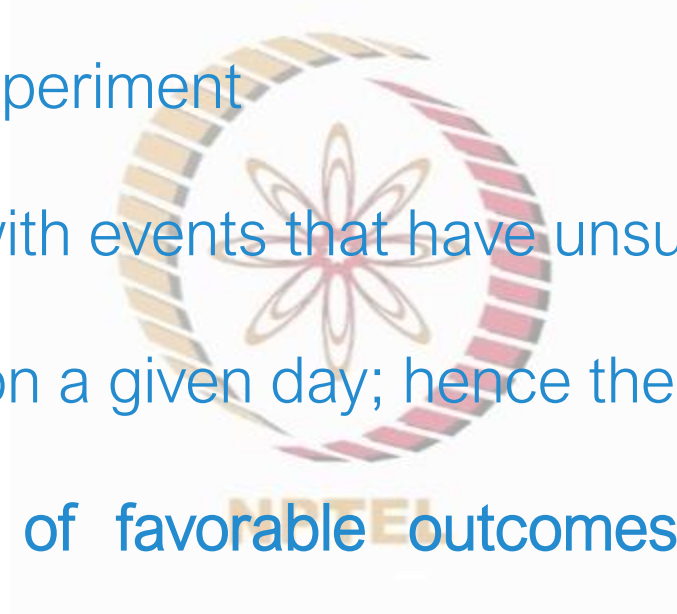


Probability Definition



Probability Definition

- We calculate probability when we want to understand the chances of an event happening in a random experiment
- Probability is associated with events that have unsure outcomes
- It might or might not rain on a given day; hence there are two possible events
- $\text{Probability} = \frac{\text{Frequency of favorable outcomes}}{\text{Total frequency of all outcomes}}$
- Probabilities are always between 0 and 1



Probability Definition

- Probability of getting 7 in a die roll can be zero, that is minimum
- Conversely, if the event is defined as getting 1, 2, 3, 4, 5, or 6, then the probability can be 1, that is maximum
- Hence the probability of an event in a random experiment can only be between 0 and 1
- A probability of 0 means a certainty that an event will not occur
- A probability of 1 means that an event is certain to occur



Probability Definition

- Probability is also represented as percentages: 0 corresponds to 0% and 1 corresponds to 100%
- Often, the news on weather channel tells you that there is a 20% or 30% chance of rainfall on a given day
- An investment in financial markets may have a low or high probability of providing high returns
- Similar probabilities of success and failure exists in medicine, business, and life itself

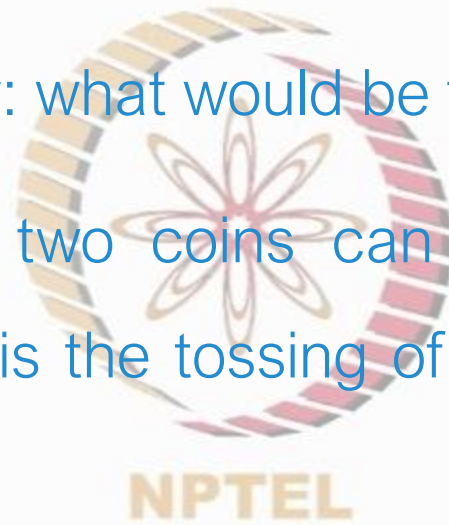


Combining two or more experiments



Combining two or more experiments

- Let's now look at some more complex experiments
- Let us flip two coins together: what would be the sample space here
- The experiment of tossing two coins can be thought of as a two-step experiment in which step 1 is the tossing of the first coin and step 2 is the tossing of the second coin
- If we use H to denote head and T to denote tail, then then our sample space becomes: $S = \{HH, HT, TH, TT\}$



Combining two or more experiments

- In this experiment, the total number of possible outcomes are $2 \times 2 = 4$
- Here 2 (H, T) is the sample space of first and second coins
- Let us generalize this with a random event in k-steps, where the possible outcomes are n_1, n_2, \dots, n_k
- In such multi-step experiments, the total number of experimental outcomes are: $S = n_1 \times n_2 \times \dots \times n_k$
- Such events can be shown in the form of tree diagram

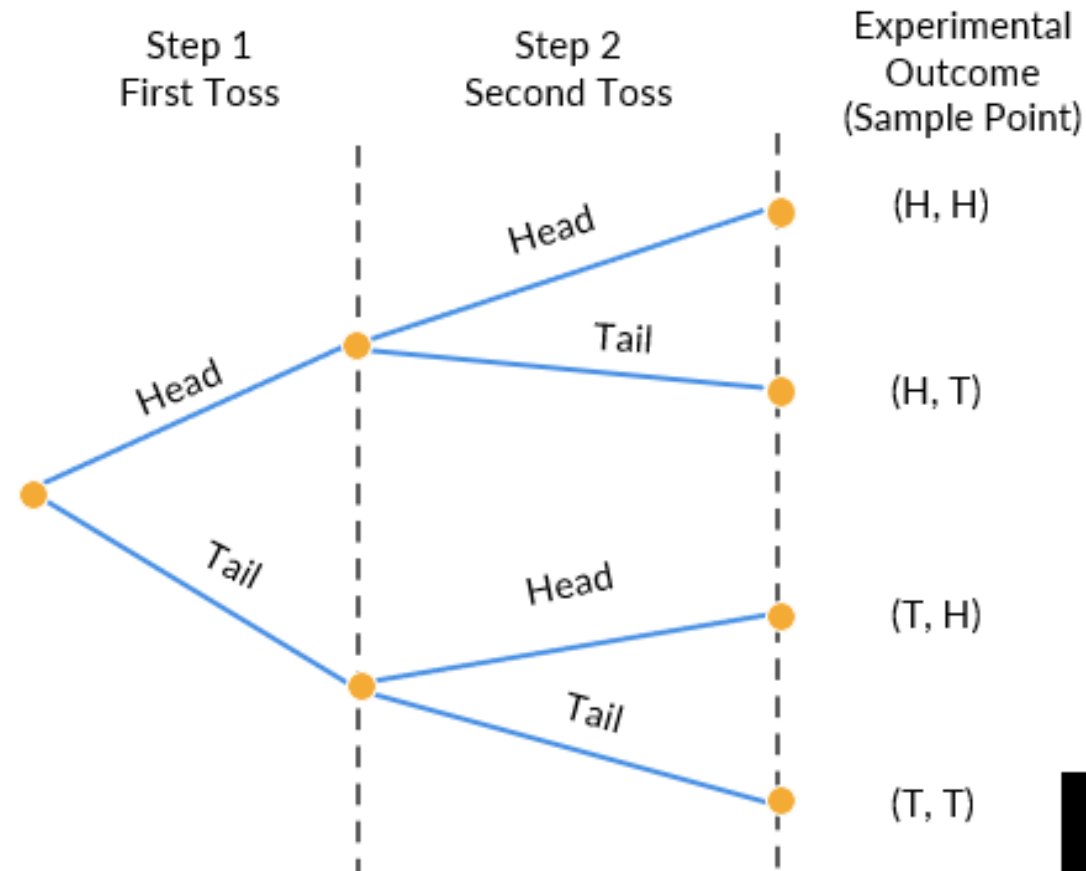


Combining two or more experiments

- Tree diagram for a simple coin toss game is provided here

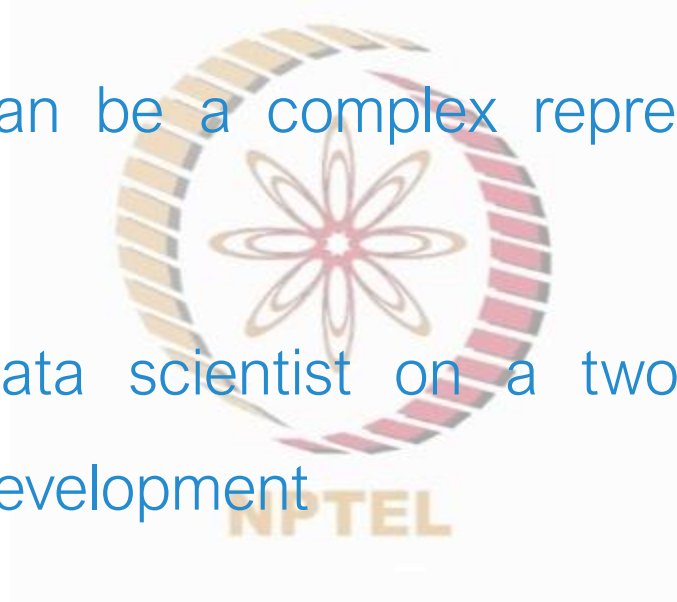
Tree-Diagram example

- Consider tossing a balanced coin twice



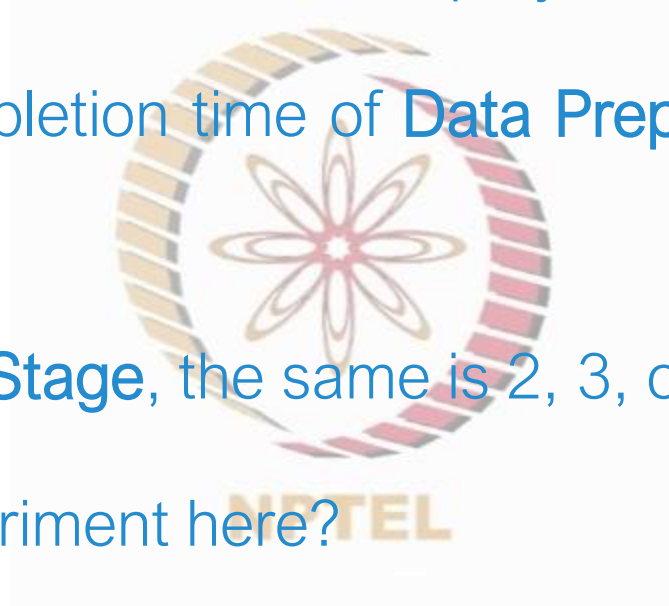
Combining two or more experiments

- The experiments need not be only a single step, they can be multistep
- The sample outcome can be a complex representation of the multistep experiment
- Consider yourself as data scientist on a two stage project: (1) Data Preparation; (2) Model Development
- How to estimate the possible completion time for that project?



Combining two or more experiments

- Let us make use of past data from similar projects
- Past data suggests completion time of **Data Preparation** stage as 5, 6, or 7 months, equally likely
- For model **Development Stage**, the same is 2, 3, or 4 months, equally likely
- What is the random experiment here?
- What are the possible different events here?
- Let us try to find the probabilities of some of these different events

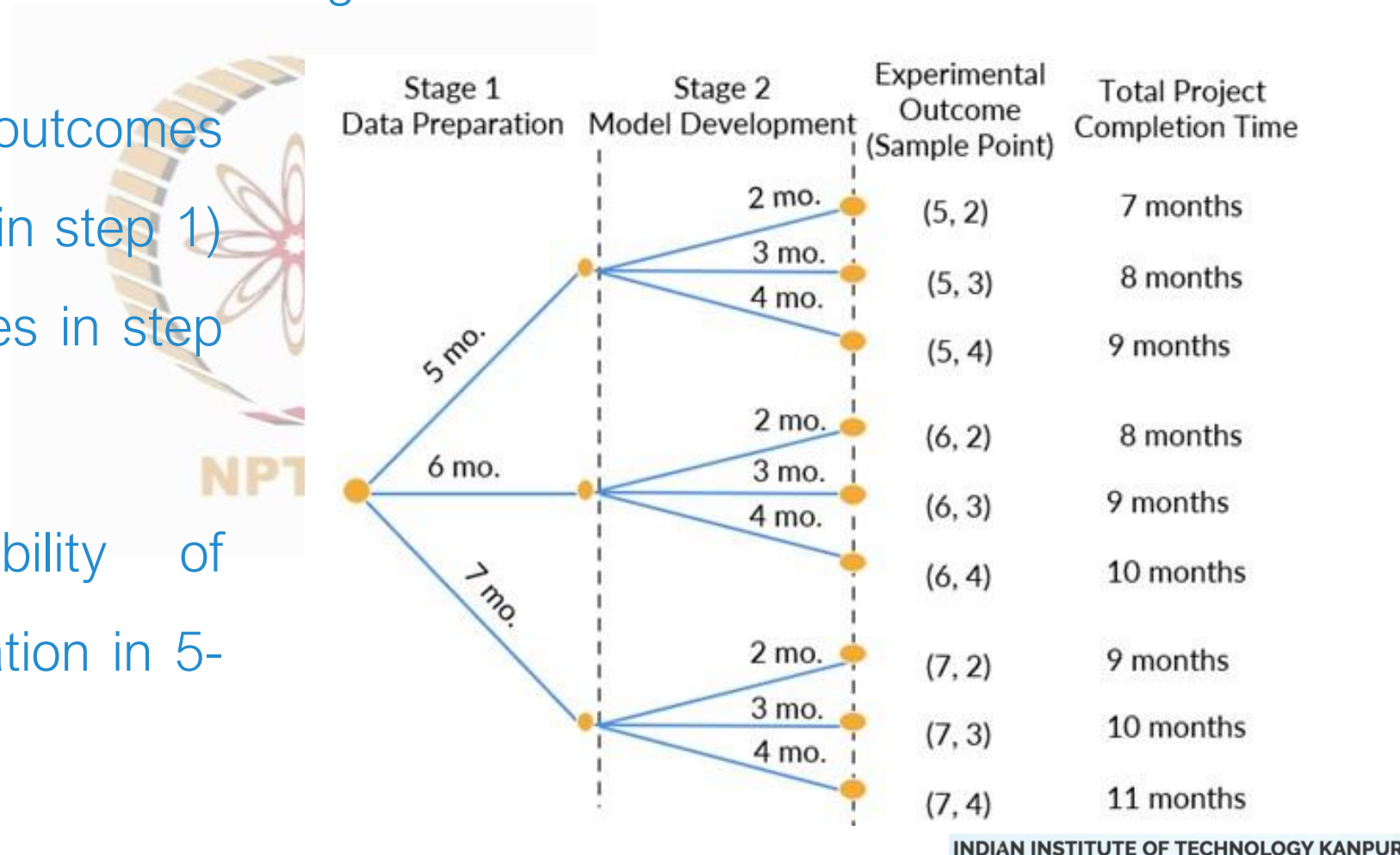


Combining two or more experiments

- Let us understand this with the tree diagram

- The number of sample outcomes = (number of outcomes in step 1) into (number of outcomes in step 2) = $3 * 3 = 9$

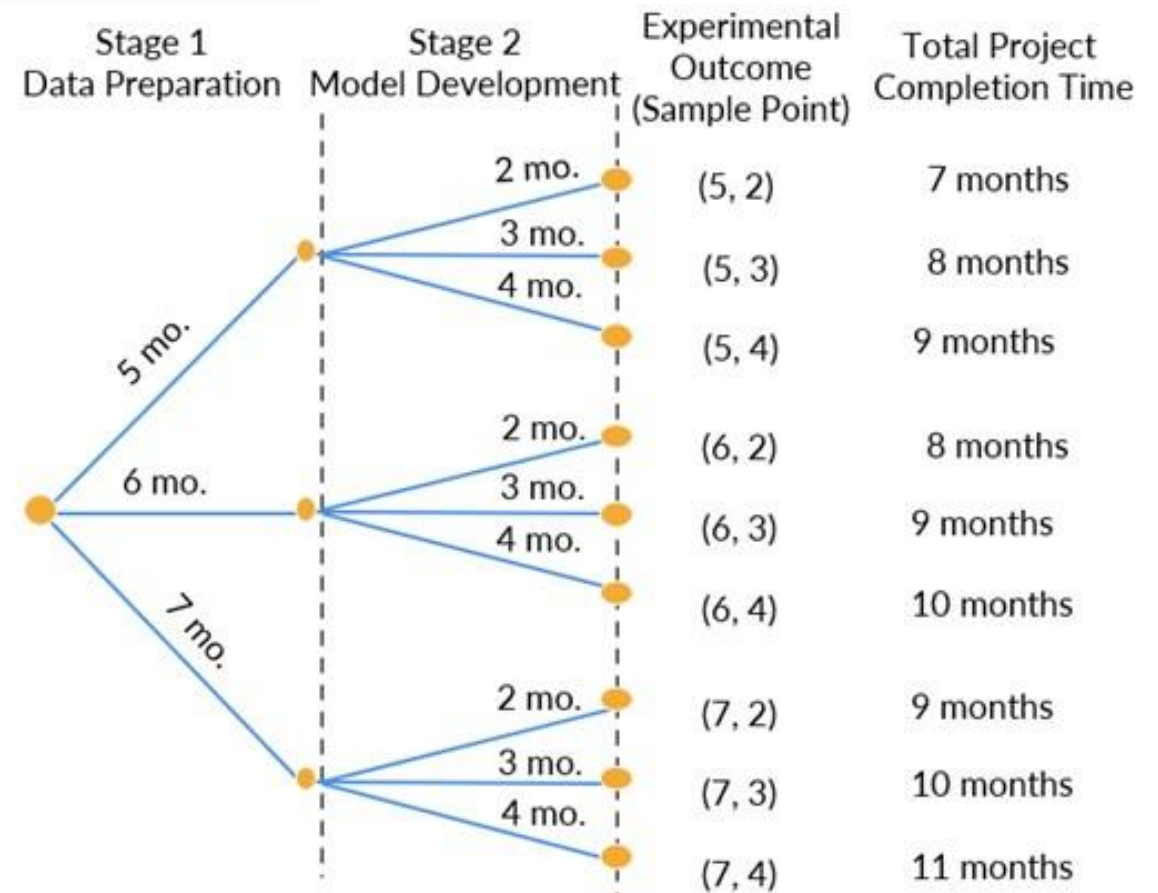
- What is the probability of completing data preparation in 5-Months: $3/9=1/3$



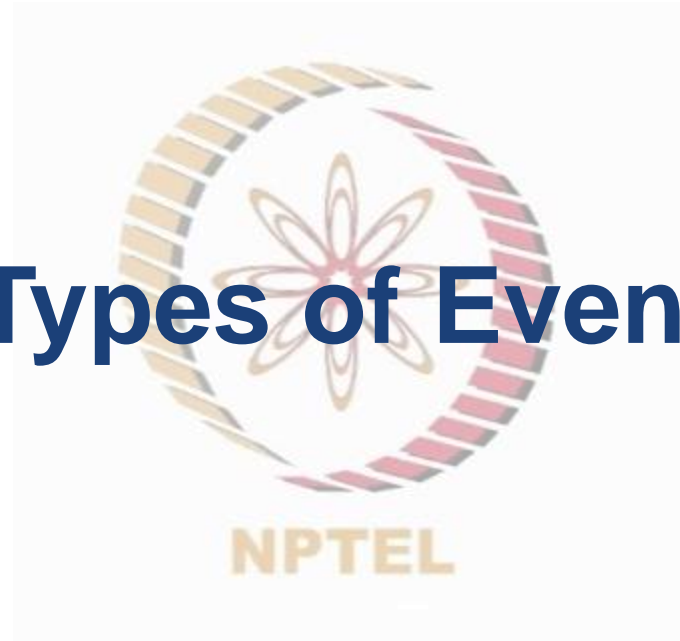
Combining two or more experiments

- For this question, you do not need the information about the next stage

- What is the probability of completing model development in exactly 3 months: $1/3$
- What is the probability that project will be completed in 8-months: $2/9$

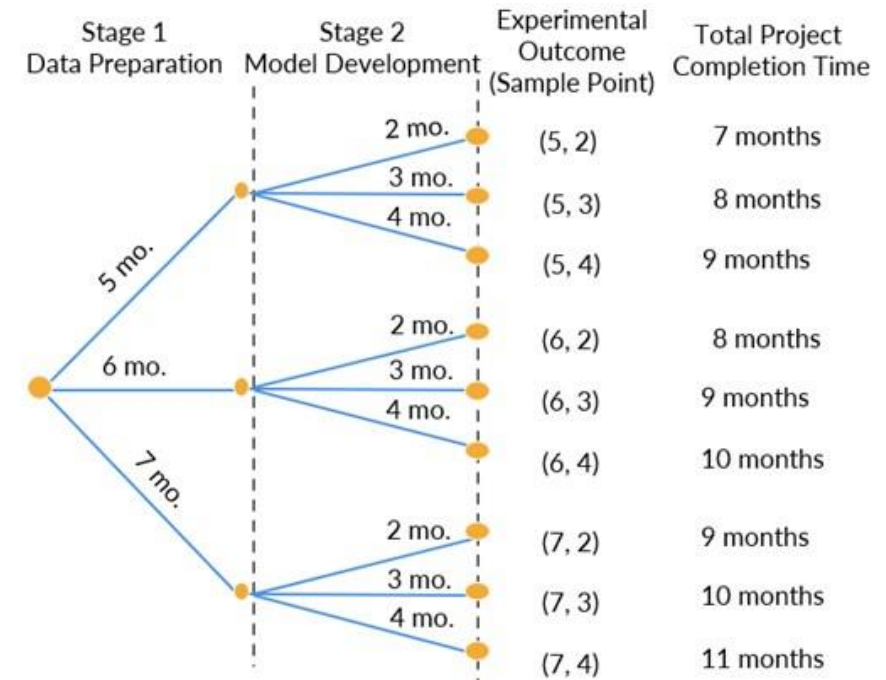
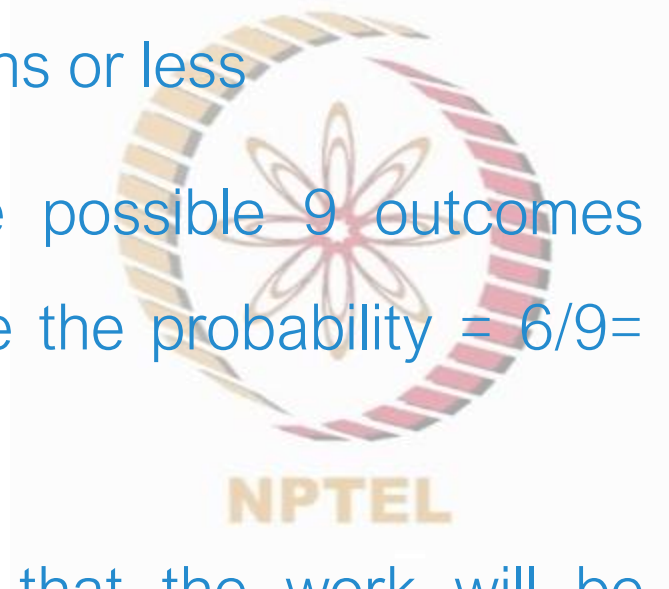


Types of Events



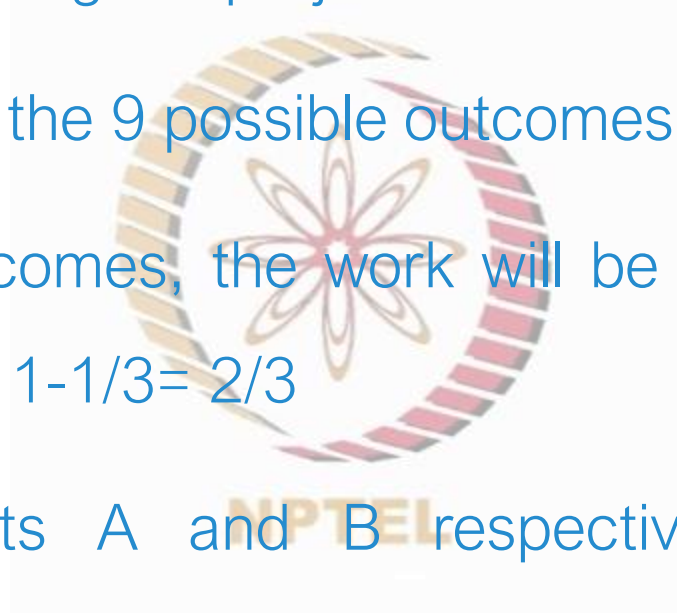
Types of Events

- If we are estimating the probability that project will be finished in 9-months or less
- There are 6 out of the possible 9 outcomes where it happens: hence the probability = $6/9 = 66.67\%$
- What is the probability that the work will be completed in more than 9-months; this question is also some how related to previous question



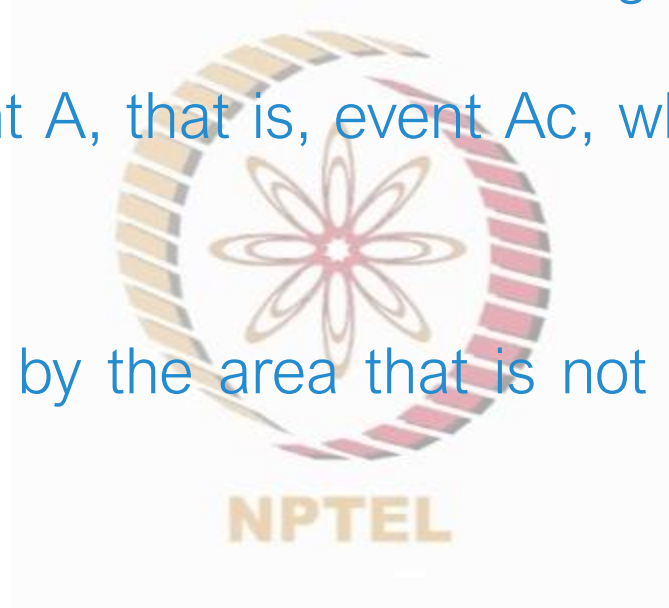
Types of Events

- The probability of completing the project in more than 9-months is $3/9=1/3$
- This is so because out of the 9 possible outcomes, 3 satisfy this condition
- In the remaining six outcomes, the work will be completed in 9-months or less; thus, the probability $1-1/3= 2/3$
- If we call these events A and B respectively, then A and B are complementary events
- That B can be referred to as 'Ac' or complementary to A



Types of Events

- This can very well be represented as a Venn diagram
- The complement of event A , that is, event A^c , which in our case is Event B
- Event B , is represented by the area that is not in A , which is the white portion within the rectangle
- Since the probability of complete sample space is 1: $P(A) + P(A^c) = 1$
- This takes us to the basic rule of probability, that is, the sum of the probabilities of all the events always add up to 1

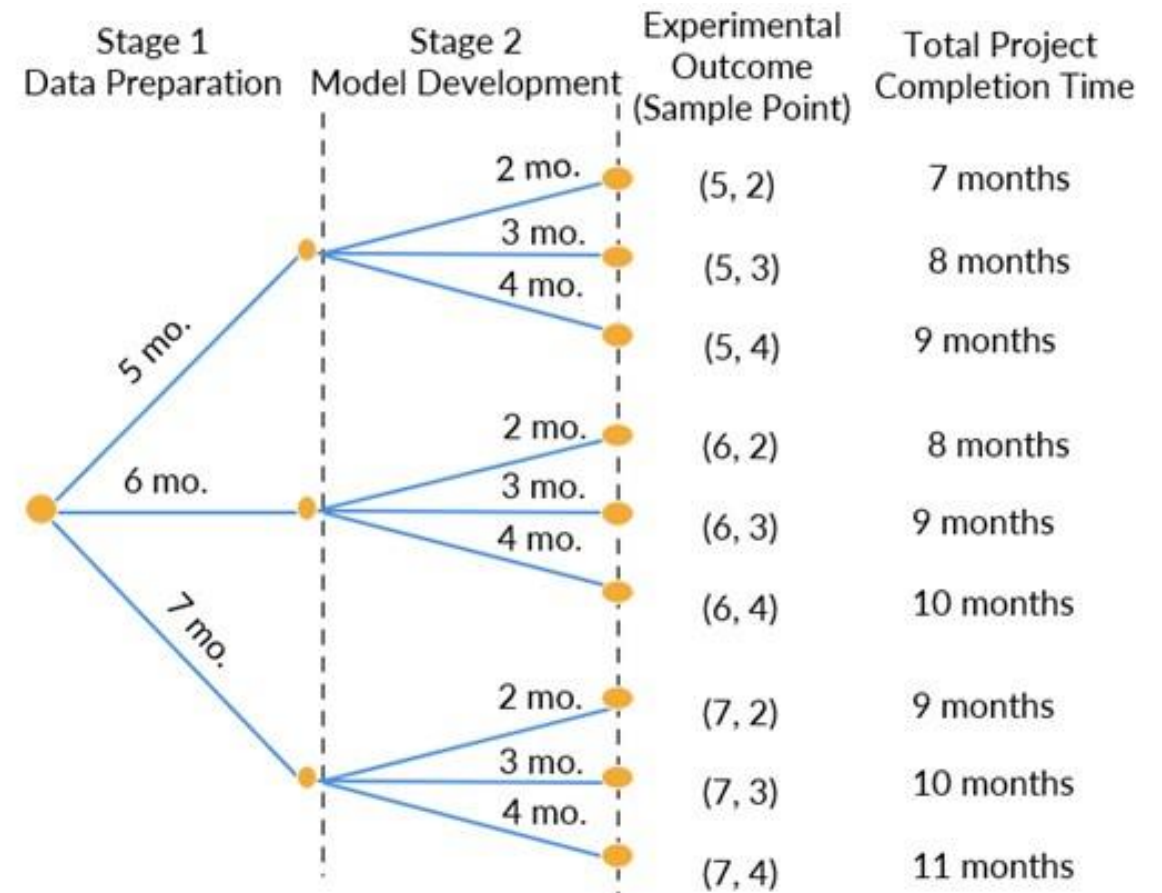


Intersection Events



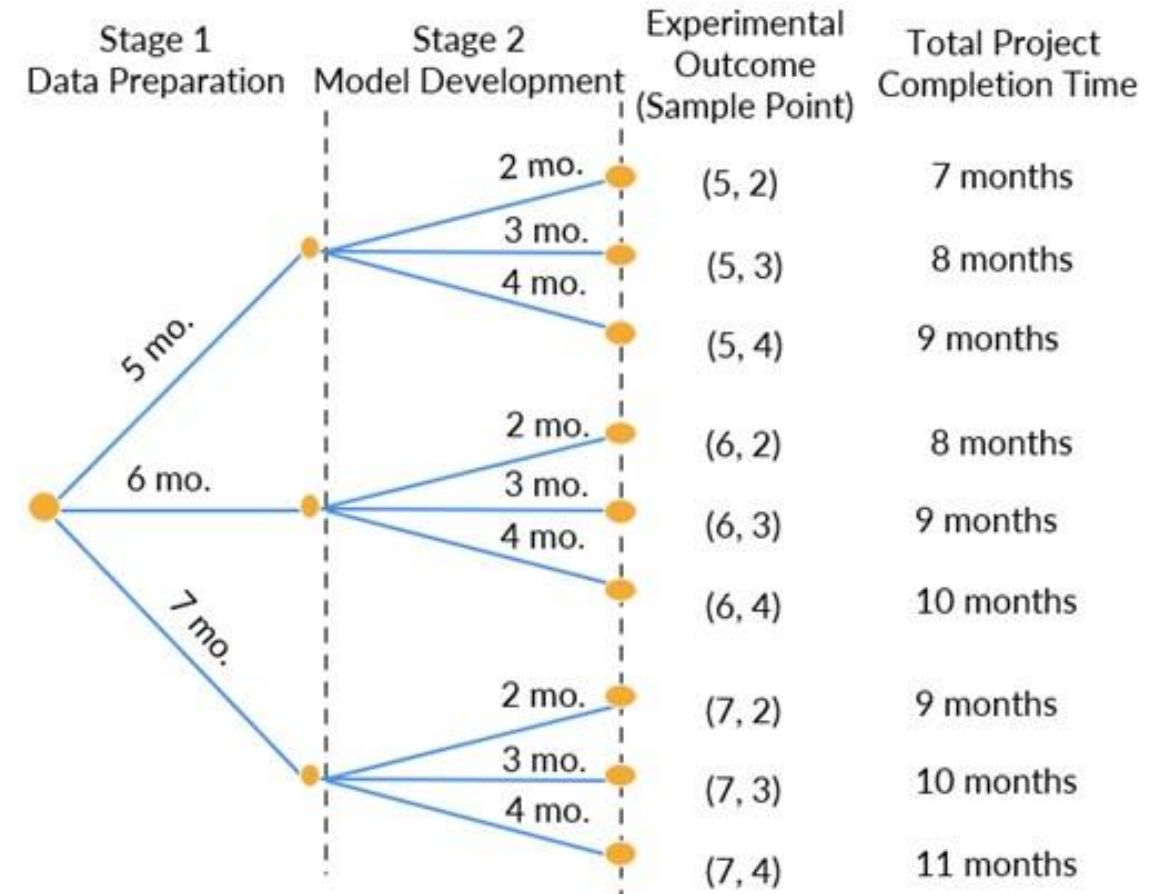
Intersection Events

- An event is a subset of all the possible outcomes of a random experiment
- Within an event, a simple event is with one possible outcome and a compound event is a combination of two or more simple events



Intersection Events

- Our machine learning projects, which is a multi-step experiment, a compound event like data preparation can be broken into 3 possible simple events
- This is how we compute the probabilities for compound events



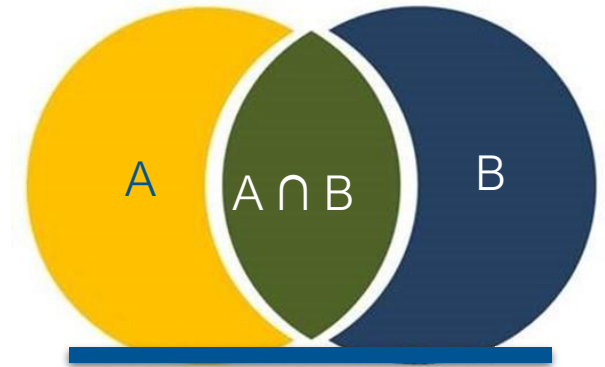
Intersection Events

- In the example of rolling a dice, the event, 'Rolling a one' is an example of a simple event
- The event, 'Rolling an odd number' will be an example of a compound event
- It can be broken down into three simple events, 'Rolling a one', 'Rolling a three' and 'Rolling a five'
- Consider the following event: Completing the Data preparation Stage in 5 months and the Model Development stage in 2



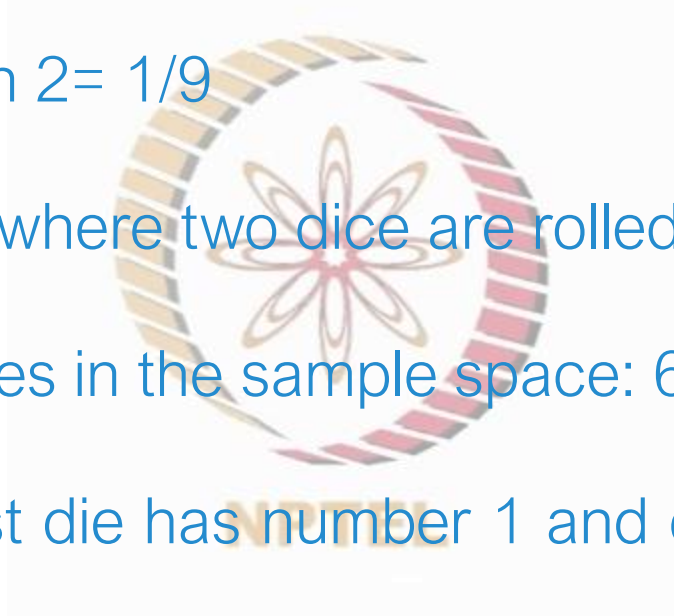
Intersection Events

- This is a composite event with intersection of two events
- Event A: completing the Data preparation Stage in 5 months
- Event B: completing the Model Development stage in 2 months
- For Event A, the sample outcomes are (5, 2), (5, 3) and (5, 4)
- For Event B, the sample outcomes are (5, 2), (6, 2) and (7, 2)
- The intersection or the common of these 2 events is (5, 2): $A \cap B$



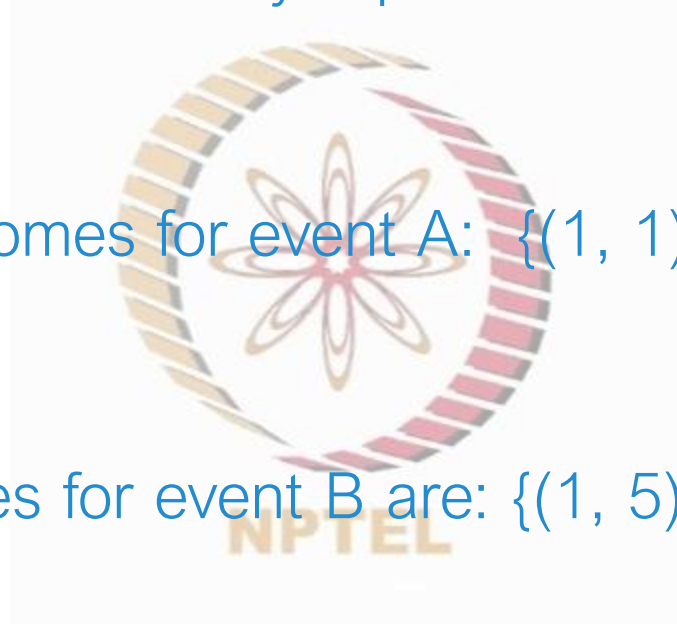
Intersection Events

- Hence, the probability of Completing the Data preparation Stage in 5 months and the Model Development stage in 2 = $1/9$
- Consider another example, where two dice are rolled simultaneously
- How many possible outcomes in the sample space: 6×6
- Consider event A where first die has number 1 and event B, where second die has the number 5



Intersection Events

- Each outcome has two values as they represent the individual outcomes of each of the two dice
- There are six possible outcomes for event A: $\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)\}$
- Again the possible outcomes for event B are: $\{(1, 5), (2, 5), (3, 5), (4, 5), (5, 5), (6, 5)\}$
- The intersection of events A and B ($A \cap B$) of these events is (1,5)

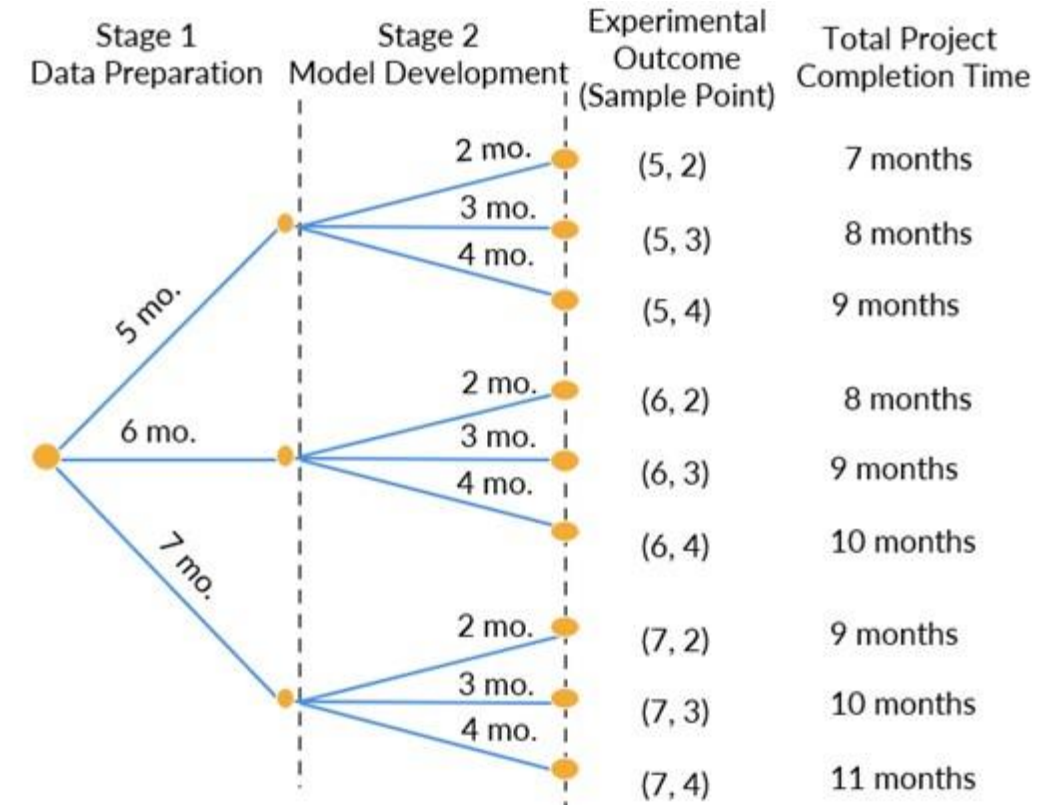


Union of Events



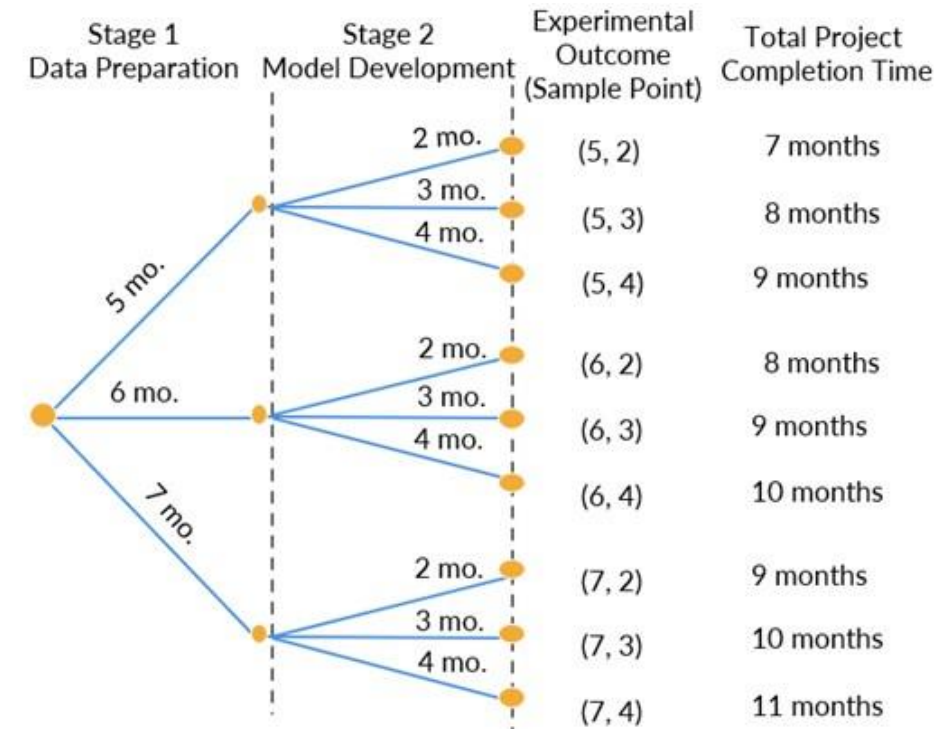
Union of Events

- For our Machine Learning project, what is the probability of completing the data preparation Stage in 5 months (Event A) or completing the Model Development stage in 2 months (Event B) ?
- For Event A, the sample outcomes are (5, 2), (5, 3) and (5, 4)
- For Event B, the sample outcomes are (5, 2), (6, 2) and (7, 2)



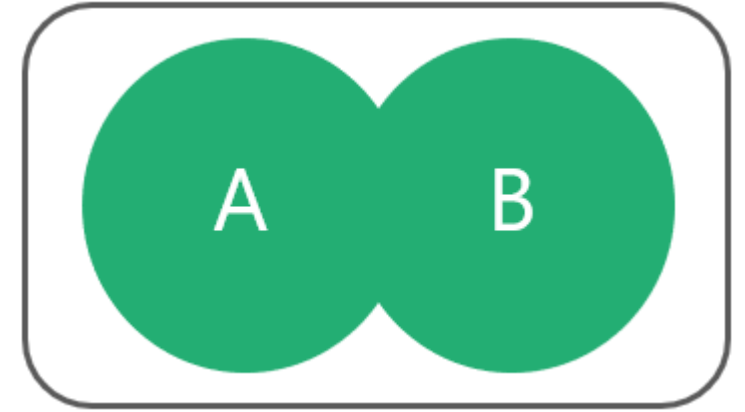
Union of Events

- The 'or' criteria represents the union of these events and we determine this by writing all the outcomes that belong to these 2 events which are $\{(5, 2), (5, 3), (5, 4), (6, 2) \text{ and } (7, 2)\}$
- Mathematically, the union of events A and B can be represented as $A \cup (\text{union}) B$ or 'A or B'



Union of Events

- The shaded green region between the events A and B represents union of A and B



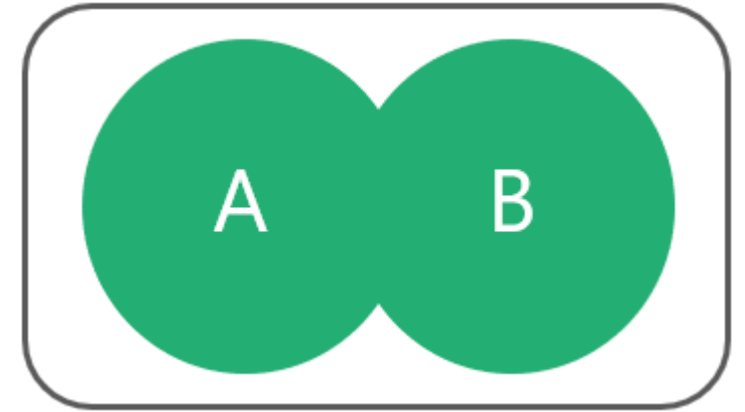
- Let us again consider the example of tossing two dice simultaneously



- Event A, where the first dice rolls 1 (6 possible outcomes): $\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)\}$
- Event B, where the second dice rolls 5 (6 possible outcomes): $\{(1, 5), (2, 5), (3, 5), (4, 5), (5, 5), (6, 5)\}$

Union of Events

- Event A, where the first dice rolls 1 (6 possible outcomes): $\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)\}$
- Event B, where the second dice rolls 5 (6 possible outcomes): $\{(1, 5), (2, 5), (3, 5), (4, 5), (5, 5), (6, 5)\}$



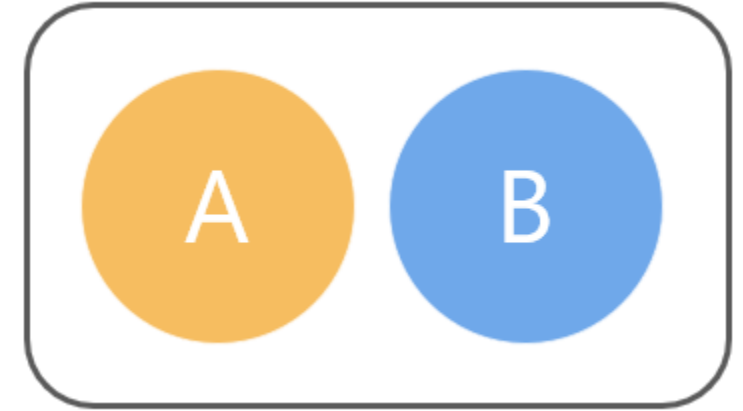
- All the possible $A \cup B$ outcomes are listed here: $\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 5), (3, 5), (4, 5), (5, 5), (6, 5)\}$

Mutually Exclusive Events



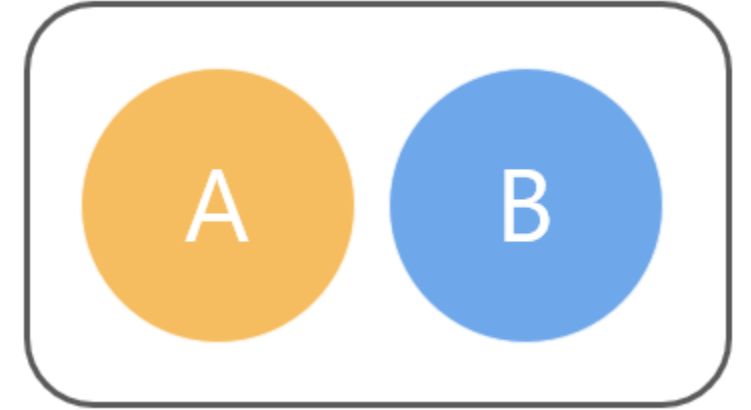
Mutually Exclusive Events

- We can say that two events are mutually exclusive when they do not occur at the same time; that is why they are also called as disjoint events
- For example, if a student has been awarded a grade of C in a subject in a given exam (Event A), she can not be awarded B in the same subject (Event B) in the same exam
- In the Venn diagram shown on the screen, you can see that events A and B do not have an overlapping region



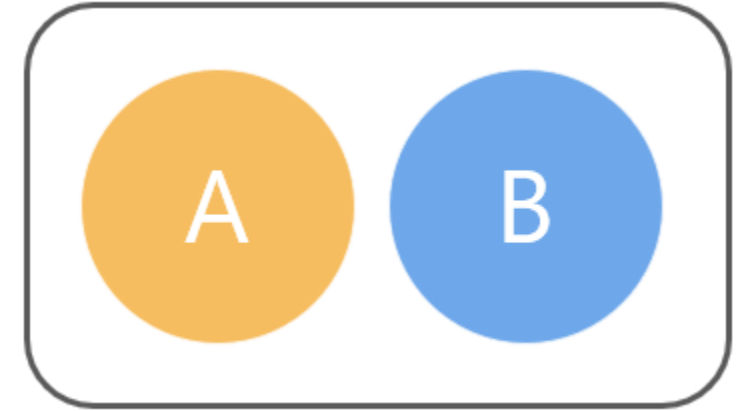
Mutually Exclusive Events

- This is a clear indication that the two events are mutually exclusive or disjoint
- Complement events are always mutually exclusive but not the other way round
- Let us go back to our rolling dice example again
- Event A is the roll where the first die has the number 1. Sample space: (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)
- Event B is Rolls where the first die has the number 3 and the second die has number 5. Sample space: {(3, 5)}.



Mutually Exclusive Events

- Event A is Rolls where the first die has the number 1.
Sample space: $(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)$
- Event B is Rolls where the first die has the number 3 and the second die has number 5. Sample space: $\{(3, 5)\}$.
- You can see that these 2 events have nothing in common
- There is no intersection between the two events; hence, these two events are mutually exclusive

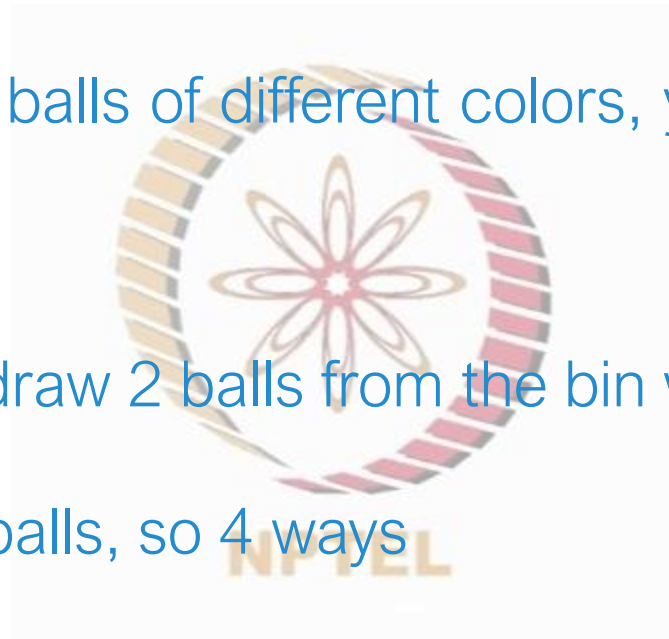


Probability for complex events



Probability for complex events

- Let us work through some classic problems related to complex events
- Consider a bin containing 4 balls of different colors, yellow, green, red and blue
- In how many ways you can draw 2 balls from the bin with replacement?
- In the first draw you have 4 balls, so 4 ways
- Similarly, you can draw 4 ways in the second draw. Hence, you can have $4 \times 4 = 16$ ways of drawing the 2 balls with replacement



Probability for complex events

- How many ways can you draw 2 balls with replacement such that both are yellow?
- In each draw you have only 1 yellow ball. So there is only 1 way in which both the draws are yellow balls
 - Number of ways to draw 2 balls with replacement is equal to 16
 - Number of ways to draw 2 balls with replacement such that both draws are yellow is equal to 1
- Number of ways to draw 2 balls with replacement such that both draws are yellow is equal to $1/16$



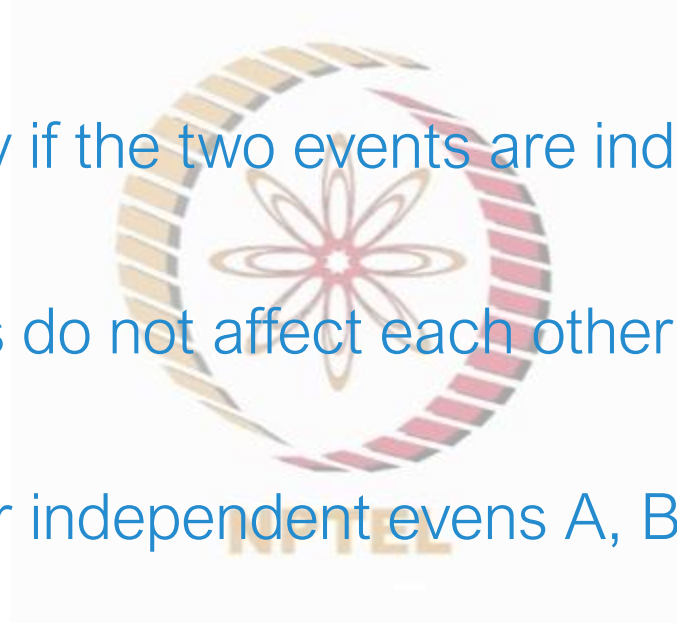
Probability for complex events

- There is another interesting way to solve this problem
- What is the probability of drawing a yellow ball? $P = \frac{1}{4}$
- Drawing 2 yellow balls with replacement is a compound event and is the intersection of the 2 simple events
 - Drawing yellow ball in 1st draw, Event A
 - Drawing yellow ball in 2nd draw, Event B
- And the probability of both events are $\frac{1}{4}$: thus, the overall event probability is $\frac{1}{4} * \frac{1}{4} = \frac{1}{16}$
- In other words $P(A \cap (\text{intersection}) B) = P(A) * P(B)$



Probability for complex events

- $P(A \cap (\text{intersection}) B) = P(A) * P(B)$ or multiplication rule
- Multiplication rule works only if the two events are independent
- That is, the two occurrences do not affect each other
- For example, if there are four independent events A, B, C, and D
- The probability that all these four events will happen together is $= P(A) * P(B) * P(C) * P(D)$



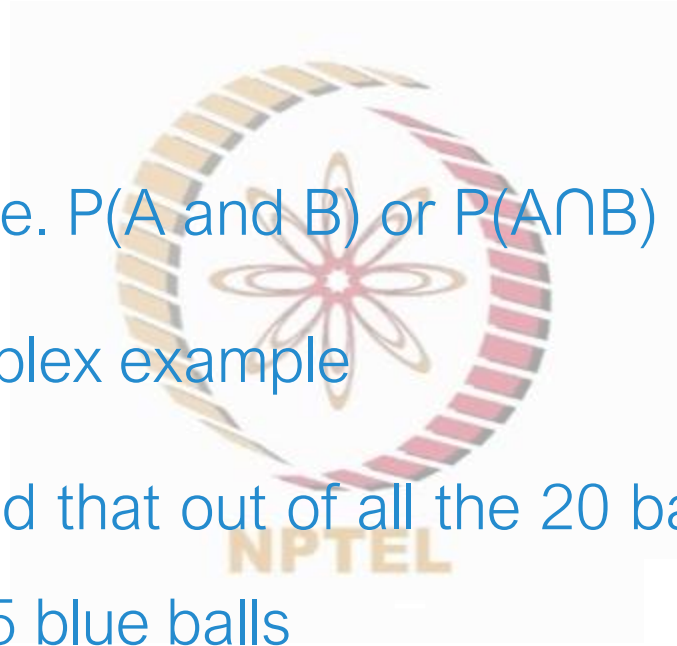
Probability for complex events

- The multiplication rule can also be applied for dependent events and the formula changes slightly
- This will be dealt along with conditional probabilities
- Independent events and mutually exclusive events are different
- Mutually exclusive events are events that cannot happen together
- Independent events are such that the occurrence of one event is not dependent on the other event



Probability for complex events

- What would happen to the probability of A and B happening together if the events are mutually exclusive?
- The probability of A and B, i.e. $P(A \text{ and } B)$ or $P(A \cap B)$ becomes zero
- Let us consider a more complex example
- The bin now has 20 balls and that out of all the 20 balls, there are 4 yellow balls, 5 green balls, 6 red balls and 5 blue balls
- What is the probability of drawing a yellow ball?
- You can draw 4 ways out of 20, so it's $\frac{4}{20} = 0.2$.



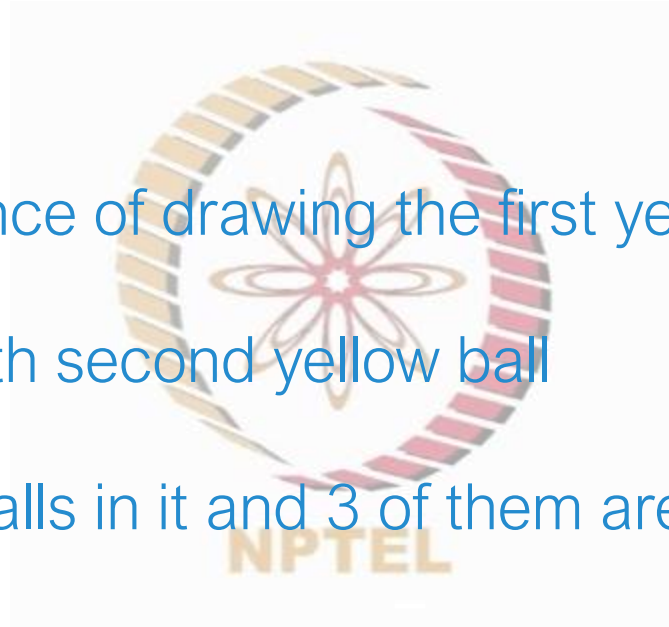
Probability for complex events

- What is the probability of drawing 2 yellow balls with replacement?
- This event is an intersection of 2 events
 - Event A: Drawing yellow ball in 1st draw, Event A
 - Event B: Drawing yellow ball in 2nd draw
- With multiplication rule: $P(A \cap B) = P(A) * P(B) = 4/20 * 4/20 = 1/25 = 0.04$
- What if the events are not independent
- What is the probability of drawing two yellow balls if we don't replace that ball after we have drawn it?



Probability for complex events

- What is the probability of drawing two yellow balls if we don't replace that ball after we have drawn it?
- We still have a one-fifth chance of drawing the first yellow ball
- The same is not the case with second yellow ball
- We have a bin that has 19 balls in it and 3 of them are yellow
- Event B probability of drawing the second yellow ball is $3/19$
- The overall probability is $= 1/5 * 3/19 = 0.032$

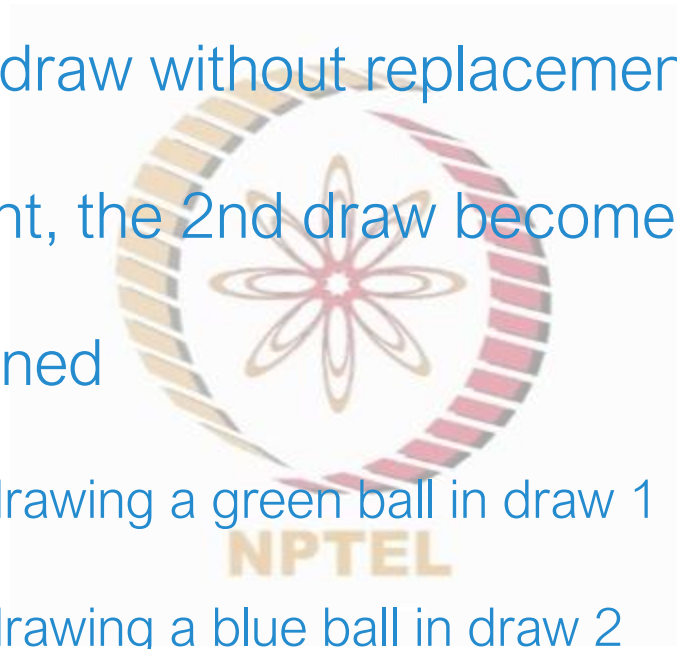


Conditional Probabilities



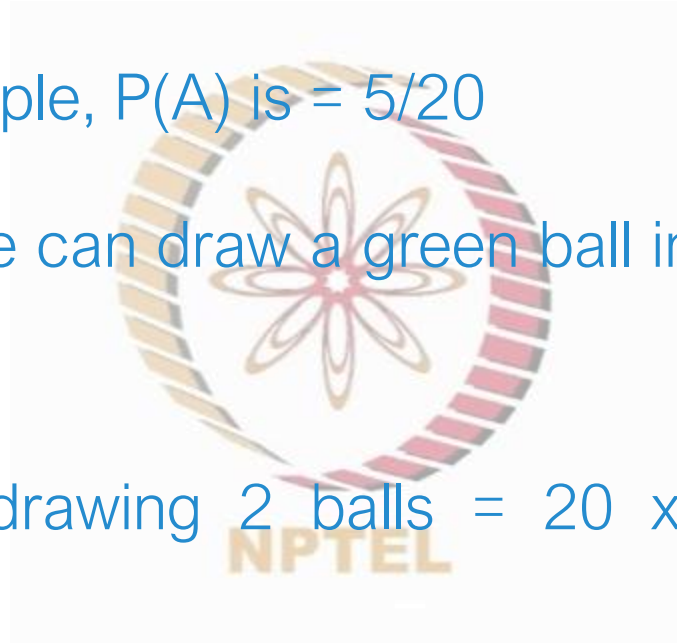
Conditional Probabilities

- Let's try to find the probability of drawing a blue ball in the 2nd draw given that you drew a green ball in the first draw without replacement
- Since there is no replacement, the 2nd draw becomes dependent on the 1st draw
- The following events are defined
 - Event A as the Probability of drawing a green ball in draw 1
 - Event B as the Probability of drawing a blue ball in draw 2
- Event B is conditional on event A due to the act of not drawing with replacement
- Mathematically: $P(B | A)$ or $P(B \text{ given } A)$



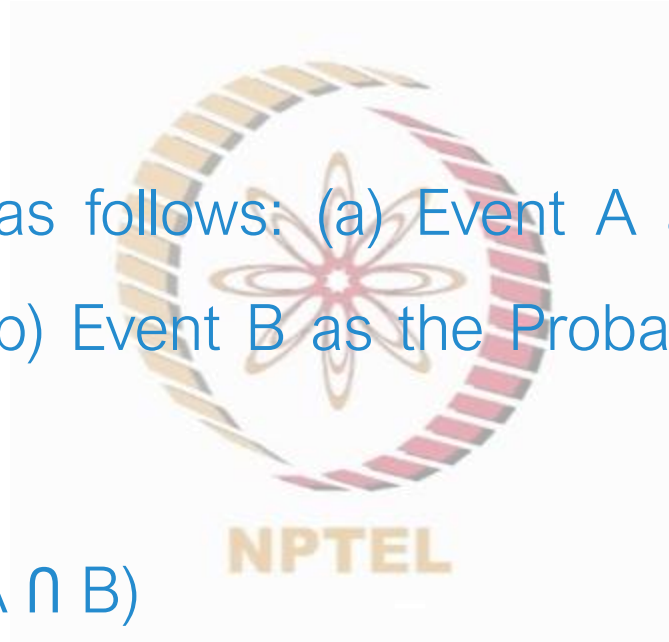
Conditional Probabilities

- The formula for $P(B|A) = P(A \cap B) / P(A)$
- Consider the following example, $P(A)$ is $= 5/20$
- The total number of ways we can draw a green ball in draw 1 and a blue ball in draw 2 is $5 \times 5 = 25$
- Total number of ways of drawing 2 balls $= 20 \times 19$, 19 because there is no replacement
- $P(A \cap B) = 25/380$, and $P(B|A) = P(A \cap B) / P(A) = (25/380)/(5/20) = 5/19$



Conditional Probabilities

- Let us look at another problem: What is the probability of drawing 2 yellow balls without replacement?
- We can define the events as follows: (a) Event A as the Probability of drawing a yellow ball in draw 1; and (b) Event B as the Probability of drawing a yellow ball in draw 2
- Drawing 2 yellow balls is $P(A \cap B)$
- Since we are doing this without replacement, we can use the conditional probability formula: $P(B \text{ given } A) = P(A \cap B) / P(A)$



Conditional Probabilities

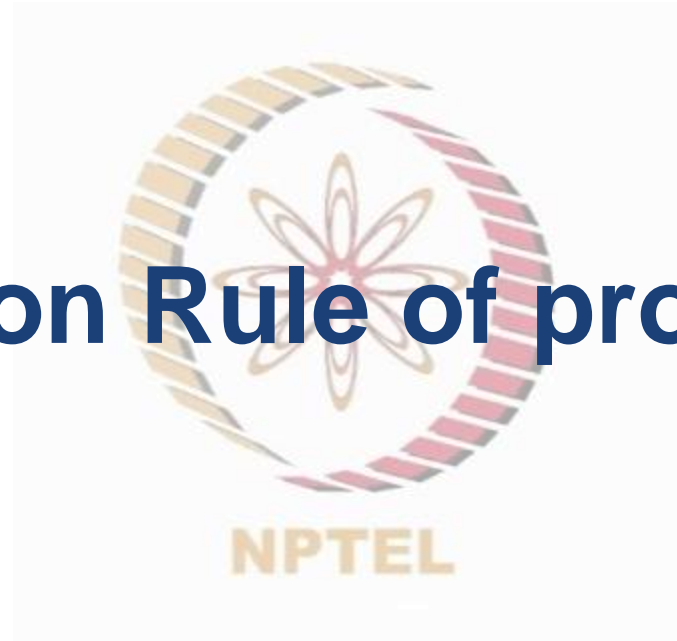
- We get $P(A \cap B) = P(B \text{ given } A) * P(A)$; where $P(A) = 4/20$
- Here $P(B/A)$: the probability of drawing a yellow ball in the 2nd draw given we have drawn a yellow ball in the first draw and not replaced it
- So, we have 19 balls left. Since, we have already drawn a yellow ball, 3 yellow balls are left: $P(B/A) = 3/19$
- So we get $P(A \cap (\text{intersection}) B) = 4/20 * 3/19 = 12/380 = 0.032$



Conditional Probabilities

- Let's end this discussion with 2 interesting notes
- If you look at the formula of $P(A \cap B)$ for dependent and independent events, the only difference is that $P(B)$ gets replaced with $P(B \text{ given } A)$
- Second, the condition for the conditional probability formula $P(B \text{ given } A) = P(A \cap B) / P(A)$ to work is that the denominator is non zero, that is $P(A)$ is not 0.

Addition Rule of probability



Addition Rules of probability

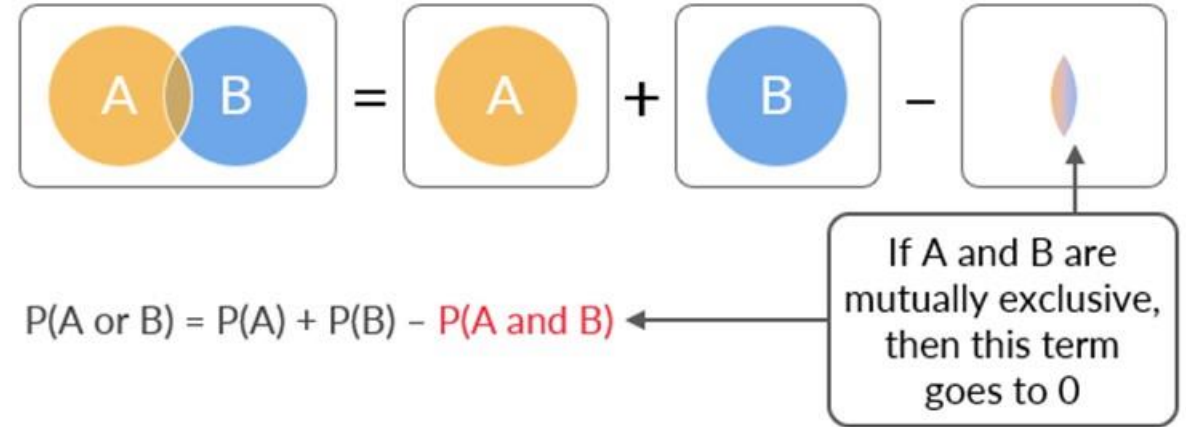
- Consider a simple Venn Diagram problem: Your city, which has 100 households, there are two newspapers. Times and Daily news. The circulation departments report that 25 of the city's households have a subscription to the Times and 35 subscribe to the Daily News. A survey reveals that 6 of all households subscribe to both newspapers. How many of the city's households subscribe to either newspaper?



Addition Rules of probability

- A is the number of households reading the Times and B is the number of households reading the Daily News

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

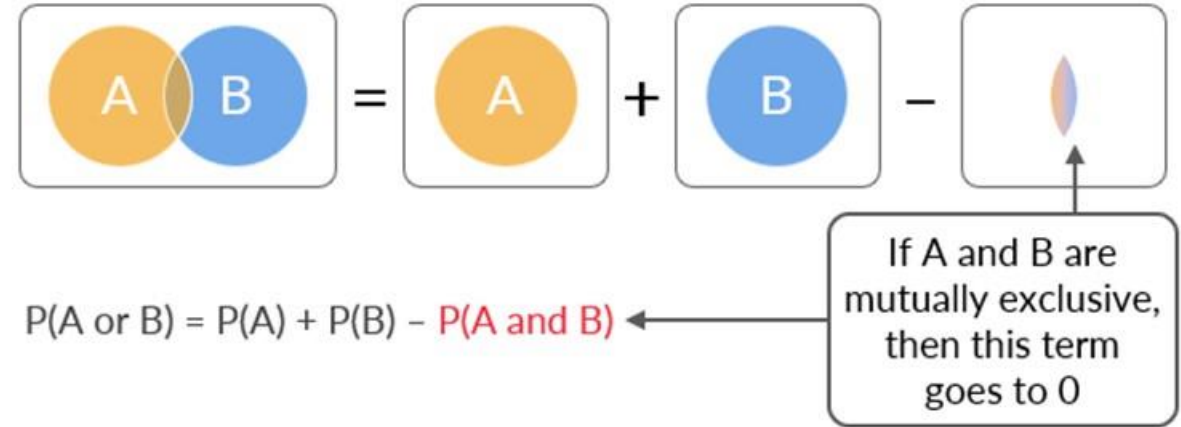


- The shaded region that overlaps A with B, that is $A \cap B$ indicates the households which have subscription to both
- The shaded region that overlaps A with B, that is $A \cap B$ indicates the households which have subscription to both
- $A = 25$, $B = 35$ and $A \cap B = 6$. Our objective is to find $A \cup B$

Addition Rules of probability

- Divide the three regions as small 'a' as those who read only the Times, small 'b' as those who read only The Daily News and 'c' as those who read both

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

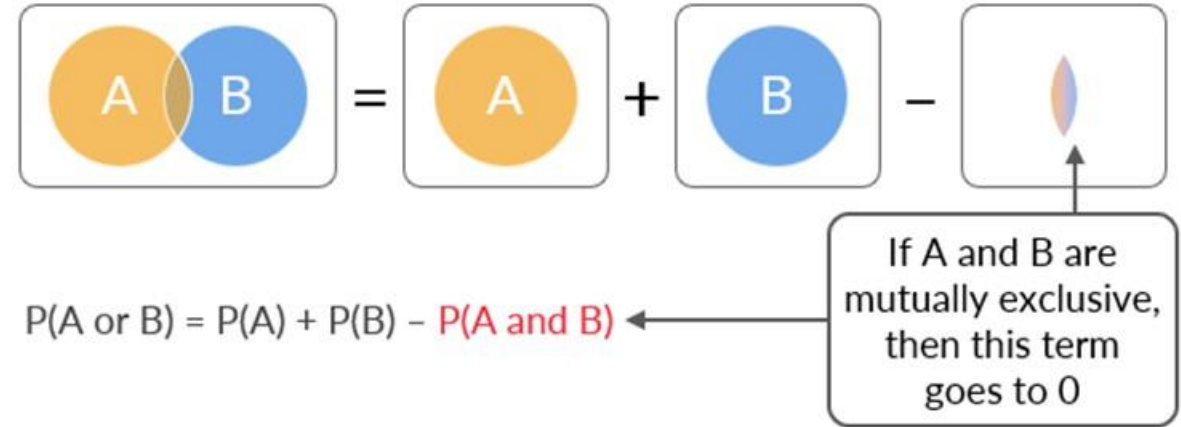


- $A = a + c = 25$, $B = b + c = 35$; $A \cup B = a + b + c$ and $A \cup B = A + B - c$
- Union of A and B ($A \cup B$) = $A + B - \text{intersection of A and B } (A \cap B) = 25 + 35 - 6 = 54$

Addition Rules of probability

- Divide by 100 to get the probabilities of these events (why?)
- $P(A \text{ union } B) = 0.54$; $P(A) = 0.25$, $P(B) = 0.35$; $P(A \text{ intersection } B) = 0.06$

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



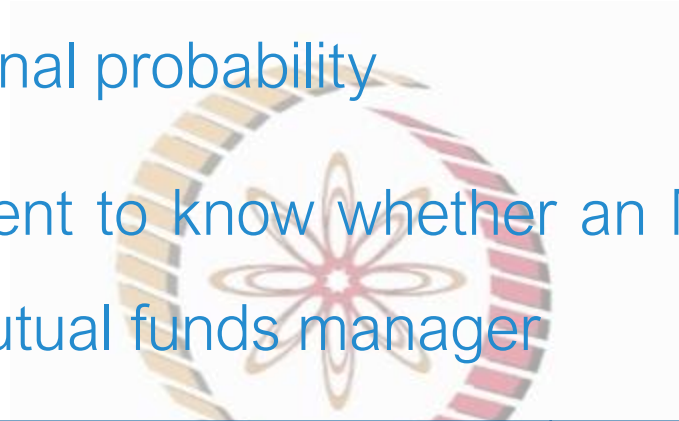
- In terms of probabilities: $P(A \text{ union } B) = P(A) + P(B) - P(A \text{ intersection } B)$
- This is known as the addition rule of probability
- If A and B are mutually exclusive, then $P(A \cap B) = 0$: $P(A \cup B) = P(A) + P(B)$.

Joint and Marginal probability



Joint and Marginal probability

- The three major types of probability include: (a) Joint probability, (b) Marginal probability, and (c) Conditional probability
- Let us consider an experiment to know whether an MBA degree could be a possible factor in the success of a mutual funds manager



	Fund outperforms the market	Fund does not outperform the market
Fund manager graduated from a top-30 MBA program	40	100
Fund manager did not graduate from a top-30 MBA program	16	244

Joint and Marginal probability

This can also be represented in another way

Fund manager graduated from a top-30 MBA program	Fund outperforms the market	40
Fund manager graduated from a top-30 MBA program	Fund does not outperform the market	100
Fund manager did not graduate from a top-30 MBA program	Fund outperforms the market	16
Fund manager did not graduate from a top-30 MBA program	Fund does not outperform the market	244

Joint and Marginal probability

Now before we calculate the different types of probabilities, we need to first convert this into a probability table

Fund manager graduated from a top-30 MBA program	Fund outperforms the market	0.10
Fund manager graduated from a top-30 MBA program	Fund does not outperform the market	0.34
Fund manager did not graduate from a top-30 MBA program	Fund outperforms the market	0.04
Fund manager did not graduate from a top-30 MBA program	Fund does not outperform the market	0.61

Joint and Marginal probability

We have also given notations to each of these events for simplicity

A1	B1	0.1
A1	B2	0.34
A2	B1	0.04
A2	B2	0.61



- The joint probability of event A1 and B1 is equal to $P(A1 \text{ intersection } B1)$ which is 0.11
- The joint probability of events A1 and B2 is equal to 0.34
- The joint probability of event A2 and B1 is equal to 0.04
- And the joint probability of events A2 and B2 is equal to 0.61

Joint and Marginal probability

Next let us understand the concept of marginal probability

A1	B1	0.1
A1	B2	0.34
A2	B1	0.04
A2	B2	0.61



- What is the probability that a mutual fund outperforms the market?
- The probability that the fund will outperform the market will be nothing but the sum of the probabilities in rows 1 and 3, which is equal to $0.1 + 0.04$ which is 0.14 or 14%
- This is marginal probability

Joint and Marginal probability

Marginal probability describes the probability of an event occurring, irrespective of the knowledge gained or the effect from previous or other external events

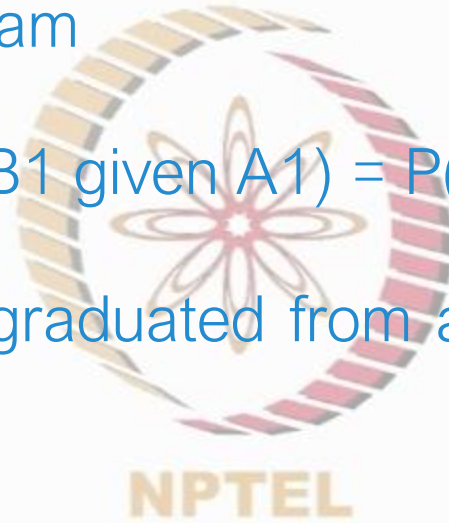
A1	B1	0.1
A1	B2	0.34
A2	B1	0.04
A2	B2	0.61



- Marginal probability of B1 is nothing but $P(B1)$ which is equal to 0.14

Joint and Marginal probability

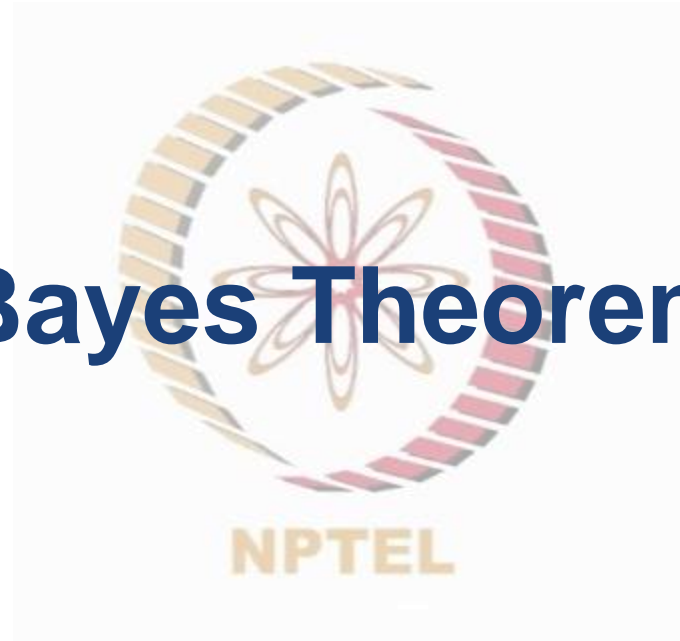
- Probability that a mutual fund will outperform the market given that the fund manager graduated from a top MBA program
- The following formula is used: $P(B1 \text{ given } A1) = P(A1 \cap B1) / P(A1)$
- The probability that a manager graduated from a top MBA program: $P(A1) = 0.10 + 0.34 = 0.44$
- Here, the joint probability $P(A1 \cap B1)$ from the table as 0.1
- Hence, $P(B1 \text{ given } A1) = 0.1/0.44 = 0.227$ (or 22.7%)



Joint and Marginal probability

- What happens to the conditional probability formula if the events are independent of each other?
- $P(A \text{ given } B) = P(A)$ since $P(A \cap B) = P(A) * P(B)$ for independent events
- Or alternatively, $P(B \text{ given } A) = P(B) * P(A)$
- If the event, 'Mutual Fund outperforms the market' and the event, 'Fund manager is from a top MBA school' are independent of each other or not?
- In the context of conditional probabilities, Bayes theorem assumes considerable significance

Bayes Theorem-I



Bayes Theorem

- Let us go back to our earlier experiment

Fund manager graduated from a top-30 MBA program	Fund outperforms the market	0.10
Fund manager graduated from a top-30 MBA program	Fund does not outperform the market	0.34
Fund manager did not graduate from a top-30 MBA program	Fund outperforms the market	0.04
Fund manager did not graduate from a top-30 MBA program	Fund does not outperform the market	0.61

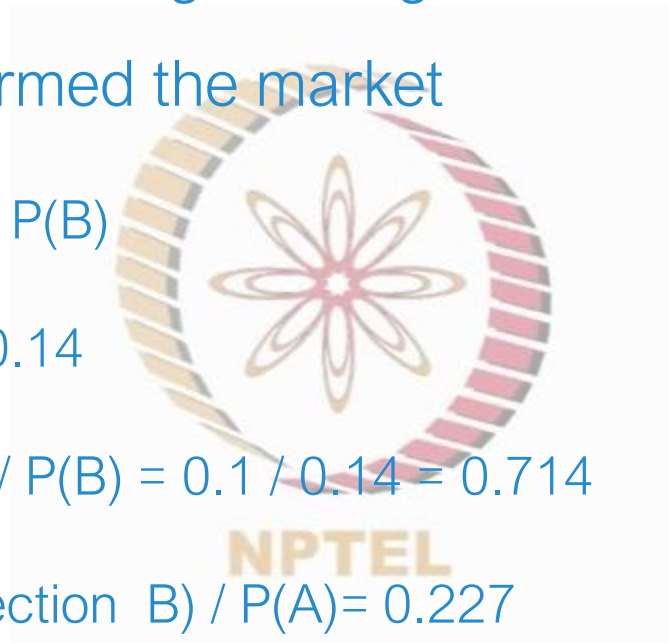
Bayes Theorem

- We calculated the probability that a mutual fund will outperform the market given the fund manager graduated from a top MBA program
 - Event A is for Fund manager graduated from a top-30 MBA program
 - Event B is for Fund outperforms the market
 - Also, $P(B \text{ given } A) = P(A \cap B) / P(A)$
- The joint probability of A and B: $P(A \text{ intersection } B)$ or $A \cap B = 0.10$
- $P(A) = 0.44$, $P(B \text{ given } A) = 0.1/0.44 = 0.227$



Bayes Theorem

- What is the probability that a manager has graduated from a top MBA program given that the mutual fund outperformed the market
- $P(A \text{ given } B) = P(A \text{ intersection } B) / P(B)$
- $P(A \cap B) = 0.1$; $P(B) = 0.10 + 0.04 = 0.14$
- $P(A \text{ given } B) = P(A \text{ intersection } B) / P(B) = 0.1 / 0.14 = 0.714$
- Similarly, $P(B \text{ given } A) = P(A \text{ intersection } B) / P(A) = 0.227$
- Also, $P(B \text{ given } A) P(A) = P(A \text{ intersection } B) = P(A \text{ given } B) P(B)$
- Rewriting differently, $P(B \text{ given } A) = P(A \text{ given } B) P(B) / P(A)$



Bayes Theorem-II



Bayes Theorem

- Bayes theorem comes into the picture when a direct calculation of a conditional probability is not possible due to lack of information
- Consider the following example, past research suggests that the probability of a middle-aged female developing breast cancer is 0.01 or 1% (Event B)
- A person can be tested positive for breast cancer but in actuality, one may not have it, and vice-versa
- Here, B is the event that a middle-aged female develops breast cancer; $P(B)=0.01$



Bayes Theorem

- So effectively, we are also asking how reliable is the testing methodology?
- This is a type of conditional probability because we are measuring the probability of breast cancer given the condition that the test came positive
- The event that a female tests positive as event A
- Thus, the probability that we are interested in measuring is basically, $P(B \text{ given } A)$
- In a sample of women who already had breast cancer, it was found that only 90% of these women tested positive. This means that the probability of a woman testing positive given she has breast cancer is 0.9: $P(A \text{ given } B) = 0.9$.

Bayes Theorem

- From the Bayes rule, $P(B \text{ given } A) = P(A \text{ given } B) P(B) / P(A)$
- We have $P(A \text{ given } B)$ and $P(B)$ with us; we need to calculate $P(A)$
- B' , which is the complement of B will be an event that a woman does not have breast cancer
- Hence this probability is nothing but $P(A \text{ given } B')$
- Also, $P(A) = P(A \cap B) + P(A \cap B')$



	B	B'
A	$P(A \cap B)$	$P(A \cap B')$
A'	$P(A' \cap B)$	$P(A' \cap B')$

Bayes Theorem

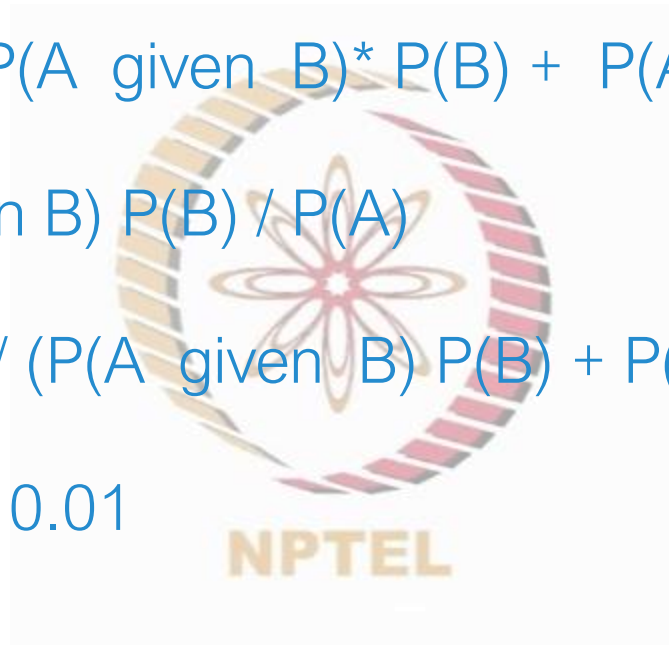
- From the Bayes rule, $P(B \text{ given } A) = P(A \text{ given } B) P(B) / P(A)$
- We have $P(A \text{ given } B)$ and $P(B)$ with us; we need to calculate $P(A)$
- B' , which is the complement of B will be an event that a woman does not have breast cancer
- Hence this probability is nothing but $P(A \text{ given } B')$
- Also, $P(A) = P(A \cap B) + P(A \cap B')$



	B	B'
A	$P(A \cap B)$	$P(A \cap B')$
A'	$P(A' \cap B)$	$P(A' \cap B')$

Bayes Theorem

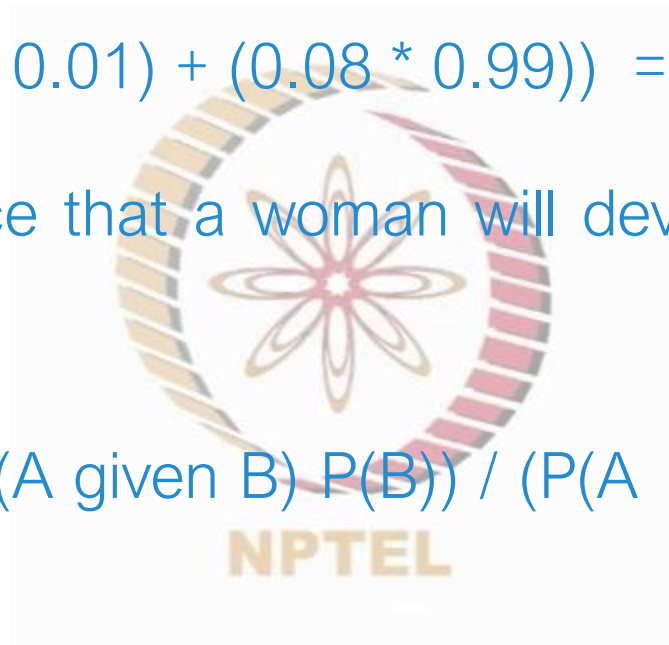
- $P(A \cap B) = P(A \text{ given } B) * P(B)$ or $P(A \cap B') = P(A \text{ given } B') * P(B')$
- $P(A) = P(A \cap B) + P(A \cap B') = P(A \text{ given } B) * P(B) + P(A \text{ given } B') * P(B')$
- Also, $P(B \text{ given } A) = P(A \text{ given } B) P(B) / P(A)$
- $P(B | A) = (P(A \text{ given } B) P(B)) / (P(A \text{ given } B) P(B) + P(A \text{ given } B') P(B'))$
- $P(A \text{ given } B) = 0.9$ and $P(B) = 0.01$
- $P(A \text{ given } B') = 0.08$
- $P(B') = 1 - P(B) = 0.99$



	B	B'
A	$P(A \cap B)$	$P(A \cap B')$
A'	$P(A' \cap B)$	$P(A' \cap B')$

Bayes Theorem

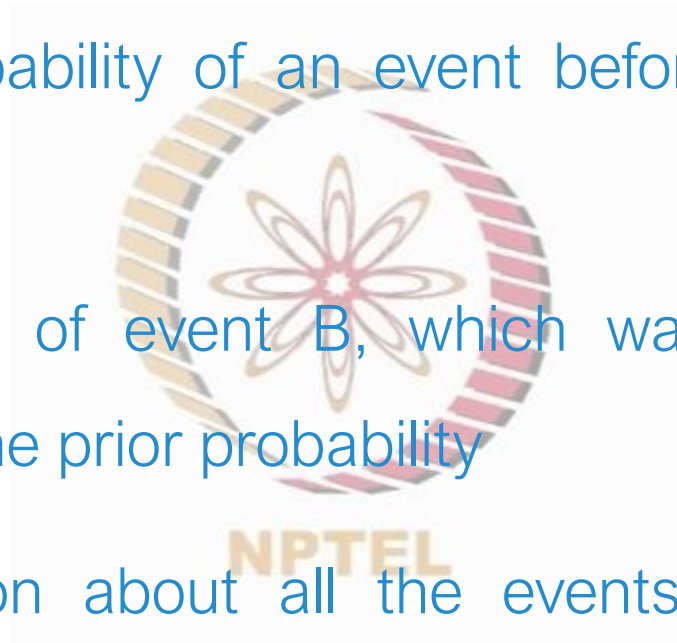
- $P(B | A) = (P(A \text{ given } B) P(B)) / (P(A \text{ given } B) P(B) + P(A \text{ given } B') P(B'))$
- $P(B | A) = (0.9 * 0.01) / ((0.9 * 0.01) + (0.08 * 0.99)) = 0.102 \text{ or } 10.2\%$
- There is only a 10.2% chance that a woman will develop breast cancer if she tests positive
- Bayes theorem : $P(B | A) = (P(A \text{ given } B) P(B)) / (P(A \text{ given } B) P(B) + P(A \text{ given } B') P(B'))$
- Let us understand the significance of this equation



	B	B'
A	$P(A \cap B)$	$P(A \cap B')$
A'	$P(A' \cap B)$	$P(A' \cap B')$

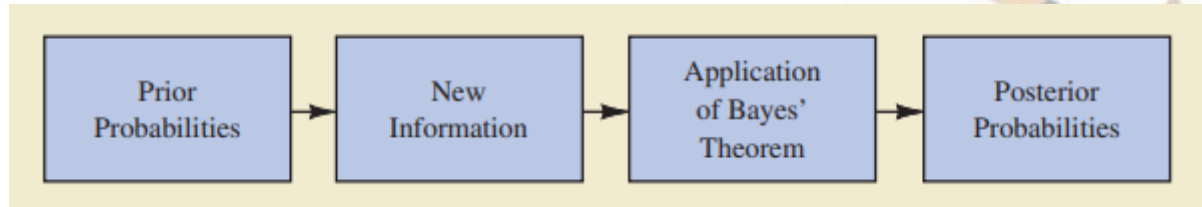
Bayes Theorem

- We start off with a set of initial or prior probabilities
- Prior probability is the probability of an event before we did our experiment and collected our new data
- In our case, the probability of event B, which was the probability of a female developing breast cancer is the prior probability
- We obtained new information about all the events for which we had the prior probabilities
- This information was given to us in the form of the conditional probabilities, $P(A \text{ given } B)$ and $P(A \text{ given } B')$



Bayes Theorem

- Our objective was basically to update these our prior knowledge about females developing breast cancer with this new information and calculate $P(B \text{ given } A)$



- This new probability that we calculated is referred to as posterior probabilities and Bayes' theorem is used for making these probability calculations
- In our case, the probability that a woman has or will develop breast cancer if she tests positive was the posterior probability
- $P(B | A) = \frac{(P(A \text{ given } B) P(B))}{(P(A \text{ given } B) P(B) + P(A \text{ given } B') P(B'))}$



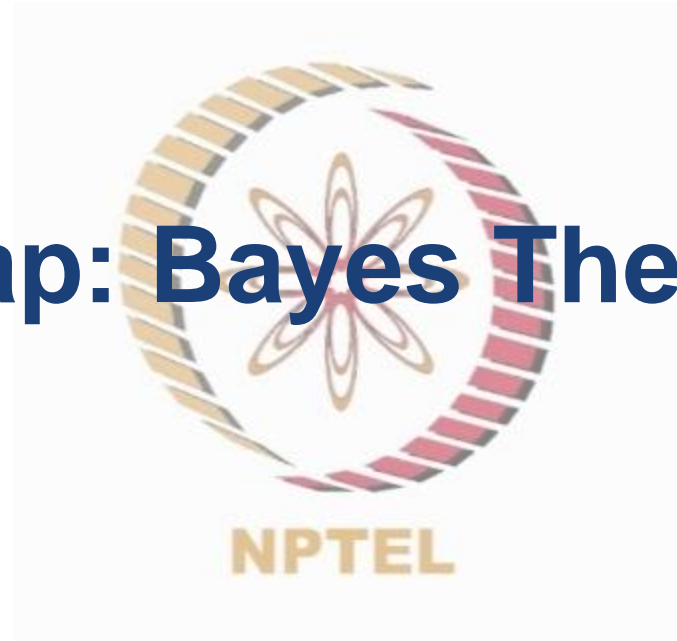
Probability Models

Prof. Abhinava Tripathi

NPTEL

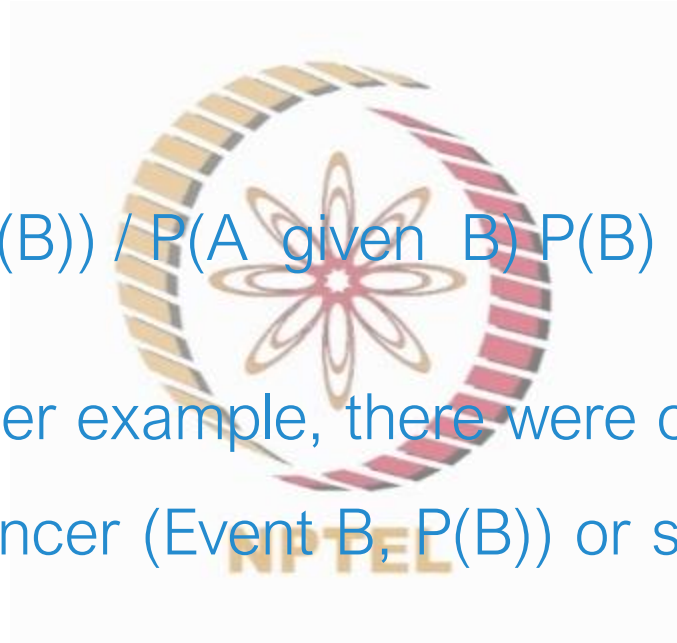


Recap: Bayes Theorem



Recap: Bayes Theorem

- Bayes theorem can be used to calculate posterior probabilities and the formula is



- $P(B | A) = \frac{P(A \text{ given } B) P(B)}{P(A \text{ given } B) P(B) + P(A \text{ given } B') P(B')}$
- In the earlier breast cancer example, there were only two possibilities, either the female has breast cancer (Event B, $P(B)$) or she doesn't have (Event B', $P(B')$)
- What if there are more than two possibilities, i.e., B1, B2, and B3

Recap: Bayes Theorem

- Consider Event B_i , where a patient is treated at hospital i (1, 2, 3). Thus, the probability that a patient was treated in hospital 1 is $P(B_1)$ which is 0.6. Similarly, the probability that a patient was treated in hospital 2 is $P(B_2)$ which is 0.3. And the probability that a patient was treated in hospital 3 is $P(B_3)$ which is 0.1. So these are our prior probabilities
- Here we have B_1 , B_2 and B_3 such that the probabilities of all these 3 events sum upto 1
- In this example, B is said to be polytomous in nature while in our earlier example, B is dichotomous in nature, i.e. we only had B and B' .

Recap: Bayes Theorem

- Now assume that hospitals have shared the probability of lawsuit being filed in the past. So we have the conditional probability of a malpractice suit being filed given the hospital is known to us



$P(\text{Malpractice suit given Hospital 1})$	0.002
$P(\text{Malpractice suit given Hospital 2})$	0.005
$P(\text{Malpractice suit given Hospital 3})$	0.007

NPTEL

- Let's say a malpractice suit is filed today. Then what is the probability that the suit was filed against hospital 1?
- So how are we going to solve this problem now?

Recap: Bayes Theorem

- Also let's call the event of filing a malpractice lawsuit as event A

$P(A \text{ given } B1)$	0.002
$P(A \text{ given } B2)$	0.005
$P(A \text{ given } B3)$	0.007

- Remember that we also have the prior probabilities that the patient was admitted to hospital 1, 2, 3. So we have: $P(B1) = 0.6$; $P(B2) = 0.3$; $P(B3) = 0.1$
- Given the information that a malpractice suit is being filed against one of the hospitals, what is the probability that hospital is hospital 1 or B1? In other words, we want to calculate $P(B1 \text{ given } A)$

Recap: Bayes Theorem

- Recall our earlier equation of Bayes theorem
- $P(B1 | A) = (P(A \text{ given } B1) P(B1)) / P(A)$
- We already know $P(A \text{ given } B1)$ and $P(B1)$. What we don't have with us is $P(A)$

	B	B'
A	$P(A \cap B)$	$P(A \cap B')$
A'	$P(A' \cap B)$	$P(A' \cap B')$

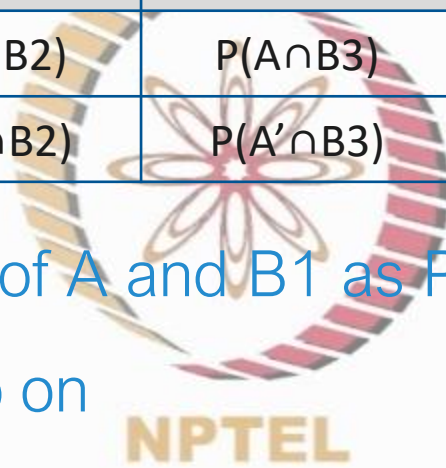
- We have $B1, B2$ and $B3$
- We have the joint probability of A and $B1$ as $P(A \cap B1)$, the joint probability of A and $B2$ as $P(A \cap B2)$ and so on

Recap: Bayes Theorem

- We can rewrite the joint probability as follows

	B1	B2	B3
A	$P(A \cap B1)$	$P(A \cap B2)$	$P(A \cap B3)$
A'	$P(A' \cap B1)$	$P(A' \cap B2)$	$P(A' \cap B3)$

- We have the joint probability of A and B1 as $P(A \cap B1)$, the joint probability of A and B2 as $P(A \cap B2)$ and so on
- Hence, the marginal probability of A comes out as $P(A) = P(A \cap B1) + P(A \cap B2) + P(A \cap B3)$



Recap: Bayes Theorem

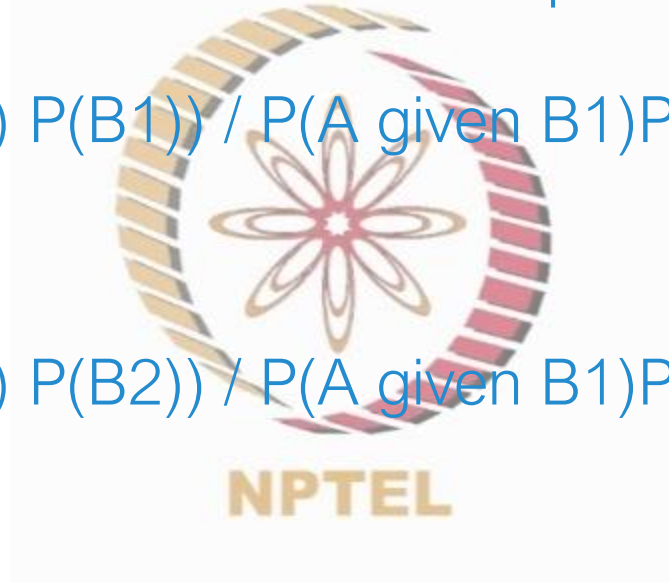
- We also know how to write joint probabilities in terms of their conditional probabilities
- $P(A \cap B_1) = P(A \text{ given } B_1)P(B_1)$; $P(A \cap B_2) = P(A \text{ given } B_2)P(B_2)$; $P(A \cap B_3) = P(A \text{ given } B_3)P(B_3)$
- Substituting these values back into $P(A)$, we get
- $P(A) = P(A \text{ given } B_1)P(B_1) + P(A \text{ given } B_2)P(B_2) + P(A \text{ given } B_3)P(B_3)$
- Now: $P(B_1 | A) = (P(A \text{ given } B_1) P(B_1)) / P(A)$
- Also, $P(A) = P(A \text{ given } B_1)P(B_1) + P(A \text{ given } B_2)P(B_2) + P(A \text{ given } B_3)P(B_3)$



Recap: Bayes Theorem

Thus we obtained the expression for conditional probabilities as shown here

- $P(B1 | A) = \frac{(P(A \text{ given } B1) P(B1))}{P(A \text{ given } B1)P(B1) + P(A \text{ given } B2)P(B2) + P(A \text{ given } B3)P(B3)}$
- $P(B2 | A) = \frac{(P(A \text{ given } B2) P(B2))}{P(A \text{ given } B1)P(B1) + P(A \text{ given } B2)P(B2) + P(A \text{ given } B3)P(B3)}$
- $P(B3 | A) = \frac{(P(A \text{ given } B3) P(B3))}{(P(A \text{ given } B1)P(B1) + P(A \text{ given } B2)P(B2) + P(A \text{ given } B3)P(B3))}$



Recap: Bayes Theorem

Let us do the numbers now

- We have Probability of A given B1 as 0.002, Probability of A given B2 as 0.005 and Probability of A given B3 as 0.007. We also have probability of B1, B2 and B3
- $$P(B1 | A) = \frac{0.002 \cdot 0.6}{0.002 \cdot 0.6 + 0.005 \cdot 0.3 + 0.007 \cdot 0.1} = \frac{0.0012}{0.0012 + 0.0015 + 0.0007} = 0.3529$$
- Thus, if a malpractice suit takes place, then there is a 35.29 percent chance that this suit took place against hospital 1
- Now we can calculate $P(B2 | A)$ and $P(B3 | A)$ as well

Recap: Bayes Theorem

In the industry, scenarios with multiple possible events are more common

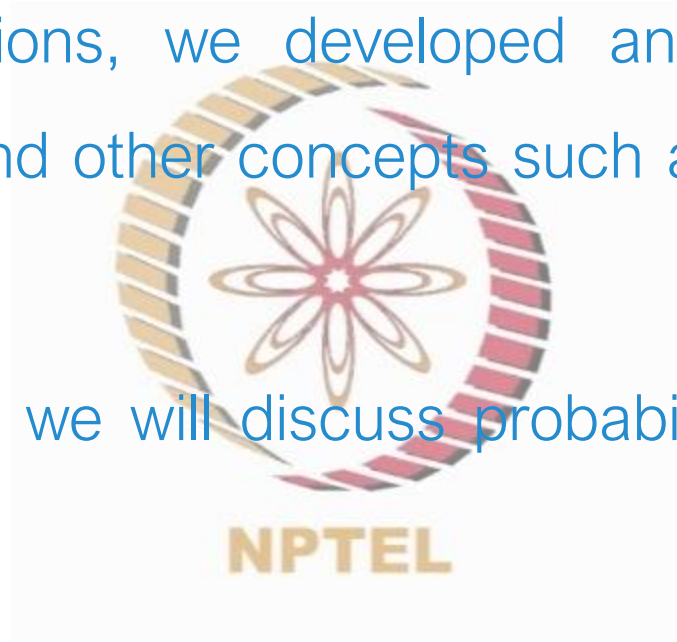
- Bayes theorem can be extended to such complex scenarios
- Consider, n mutual exclusive events, $B_1, B_2, B_3, \dots, B_n$
- For any event 'i', $P(B_i \text{ given } A) = P(A \text{ given } B_i) P(B_i) / P(A)$
- We can generalize the expression for $P(A)$: $(P(A \text{ given } B_1) P(B_1) + P(A \text{ given } B_2) P(B_2) + \dots + P(A \text{ given } B_n) P(B_n))$, thus we obtain
- $P(B_i \text{ given } A) = P(A \text{ given } B_i) P(B_i) / (P(A \text{ given } B_1) P(B_1) + P(A \text{ given } B_2) P(B_2) + \dots + P(A \text{ given } B_n) P(B_n))$

Introduction to Random Variables and Probability Distributions



Introduction

- In the previous discussions, we developed an understanding of basic concepts of probability and other concepts such as Bayes' theorem and its applications
- In the next set of topics, we will discuss probability distributions and their properties
- First we will learn about the concept of Random Variables and Probability Distributions

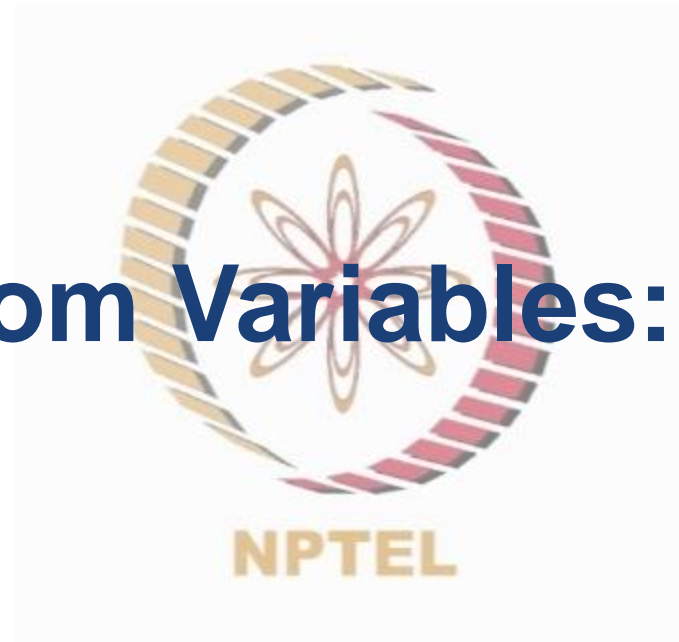


Introduction

- The next concept that we will see is that of the Expected value and Variance
- We will also discuss a discrete distribution, that is, the binomial distribution



Random Variables: I



Random Variables

- “The house always wins”: One or two people may end up winning large sums of money, but the machines are designed such that the rest of the people collectively lose more
- Consider a bag filled with 3 balls, 2 red and 1 blue; Each participant had to take out a ball, note its color, and then put it back in
- A participant who got a red ball all 4 times would receive 150 rupees; the participants who got any other result would have to pay 10 rupees



Random Variables

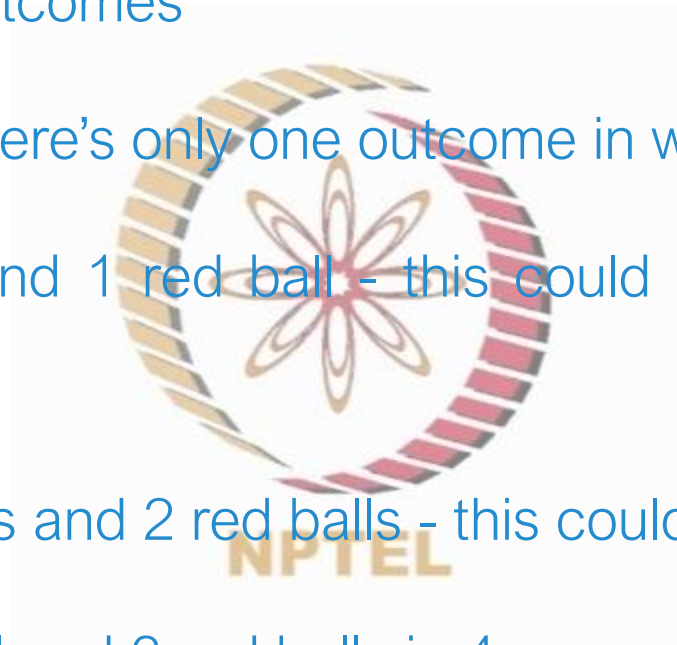
- We'll approach the problem in three steps
 - First we will see what all kinds of combinations can come up
 - Then we will see the probabilities of these combinations
 - We will use these probabilities to estimate the profit loss of a player playing this game once
- The one outcome all the players would want is to have all four ball being red
- But there are other possible out comes, what are they?



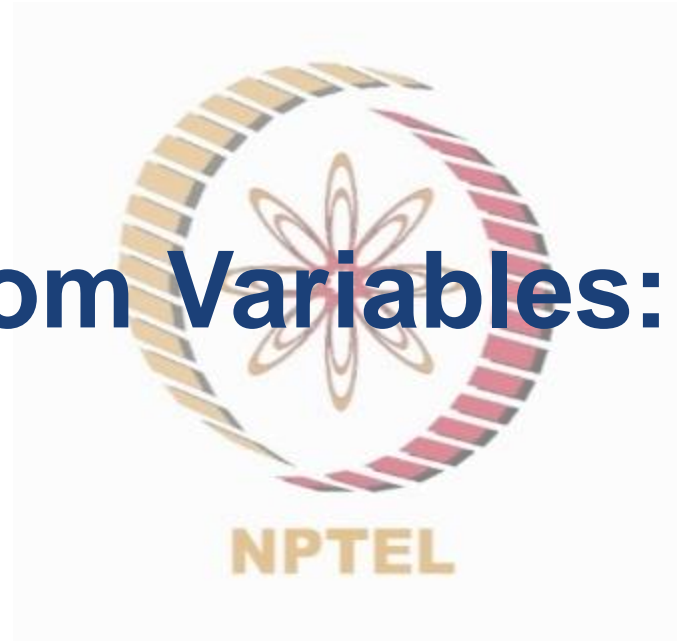
Random Variables

Let us look at all the possible outcomes

- We could get 4 blue balls - there's only one outcome in which this happens.
- We could get 3 blue balls and 1 red ball - this could happen in 4 ways - RBBB, BRBB, BBRB and BBBR.
- We could also get 2 blue balls and 2 red balls - this could happen in 6 ways.
- Also, we could get 1 blue ball and 3 red balls in 4 ways
- And there's only 1 way in which we could get 4 red balls
- In total, there are 16 possible outcomes

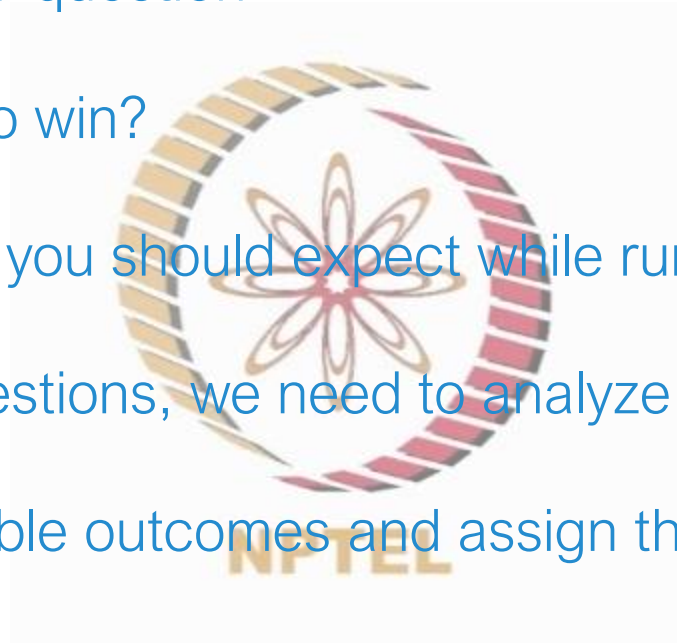


Random Variables: II



Random Variables

- Let us go back to the original question
 - How likely is the house to win?
 - What is the profit or loss you should expect while running such a business?
- In order to answer these questions, we need to analyze some probabilities
- We can quantify these possible outcomes and assign them to a variable
- In the previous example, what are the possible outcomes with 3 red balls: $X=3$
- In statistical terms, this 'X' converts possible outcomes to a number and is a random variable



Random Variables

- Such random variables can be defined in multiple ways
- For example, the number of blue balls that we drew from the bag or the number of red balls minus the number of blue balls that we have drawn from the bag
- The right way to choose the variable depends upon the nature of information you are interested in
- In this case we are interested in selection of red balls (X)
- If we want to know whether we will win or lose, we need to find different values of X and their probabilities

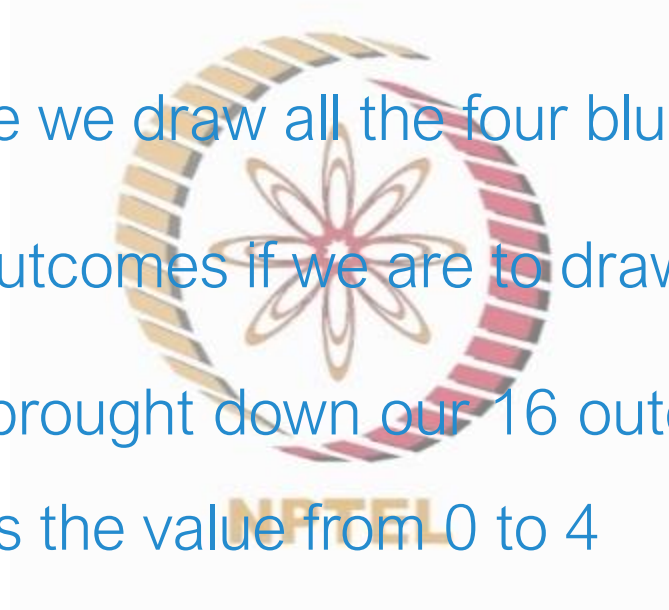


Probability Distributions



Probability Distributions

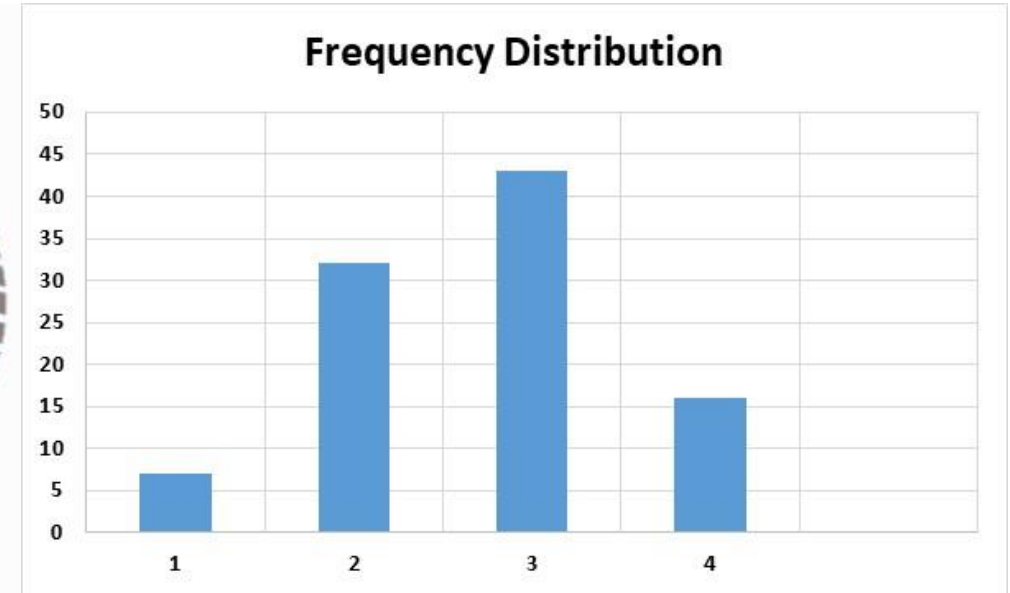
- We defined 'X' as the random event and number of red balls drawn
- So, for an outcome where we draw all the four blue balls only, $X=0$
- There are four possible outcomes if we are to draw one red ball then $X=4$
- In this manner, we have brought down our 16 outcomes into 5 groups where the random variable takes the value from 0 to 4
- To find out whether we loose money or win, we need to find the likelihood of each of these values



Probability Distributions

- The following results are obtained

X	Individual runs	Probabilities
0	2	0.02
1	7	0.07
2	32	0.32
3	43	0.43
4	16	0.16
Total	100	1

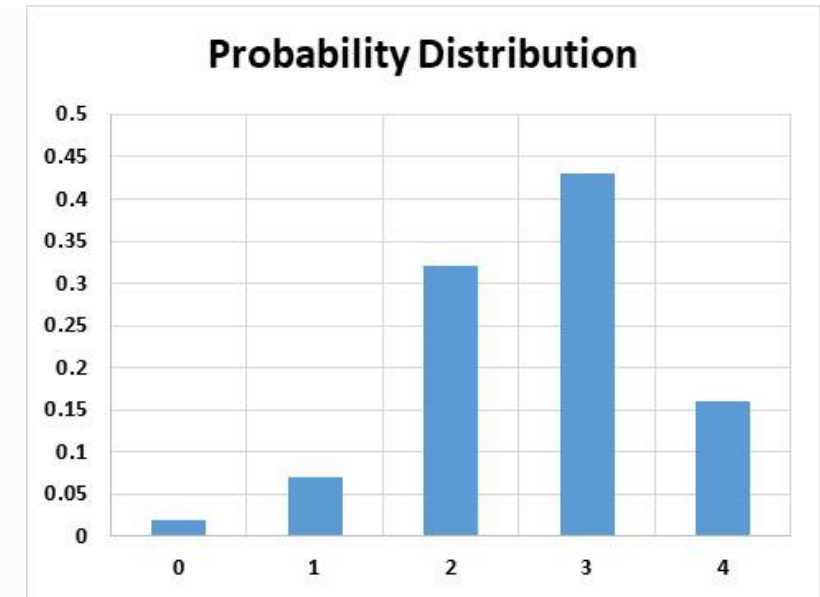


- The probability of any outcome is the number of favorable outcomes divided by the total number of outcomes
- For example, the probability of drawing two red balls is $32/100=0.32\%$

Probability Distributions

- These table and chart are also called probability distribution table and chart

X	Individual runs	Probabilities
0	2	0.02
1	7	0.07
2	32	0.32
3	43	0.43
4	16	0.16
Total	100	1

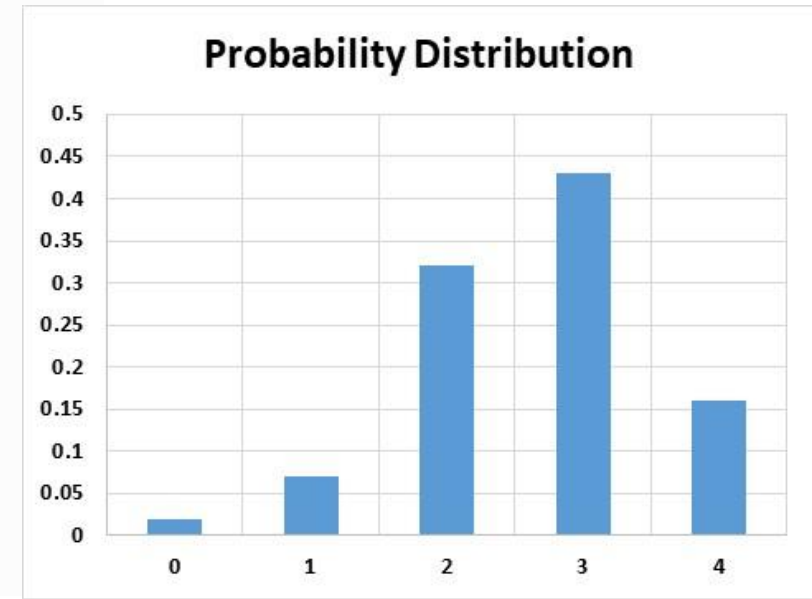


- On the X axis here, we have our 5 random variable values and on the y axis, we have our probability values

Probability Distributions

- Each value represents the probability of getting a certain number of red balls

X	Individual runs	Probabilities
0	2	0.02
1	7	0.07
2	32	0.32
3	43	0.43
4	16	0.16
Total	100	1



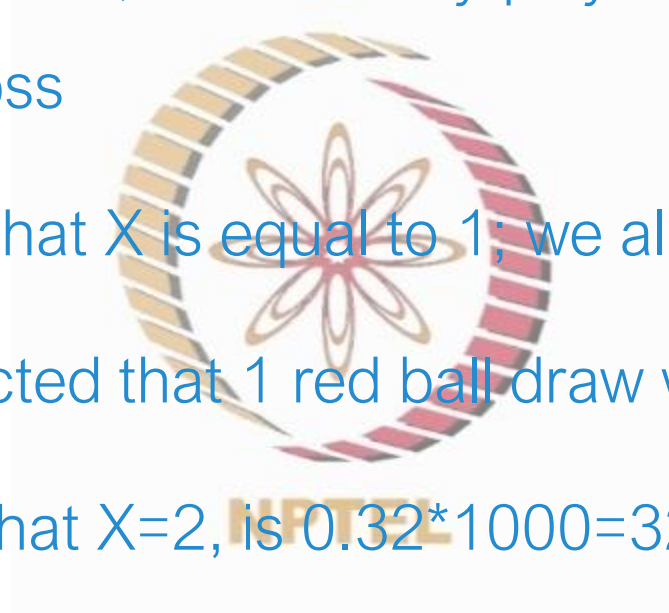
- Using the probability distribution, we can answer the question would we in the long run make money or lose money'

Expected Value



Expected Value

- Remember the red ball game; if somebody played it 1000 times, what will be their on average win or loss
- Consider the probability that X is equal to 1; we already know that it is 0.07
- In 1000 draws, it is expected that 1 red ball draw will occur 70 times
- Similarly, the probability that $X=2$, is $0.32 \times 1000 = 320$
- In this fashion, 160 people will draw 4 balls

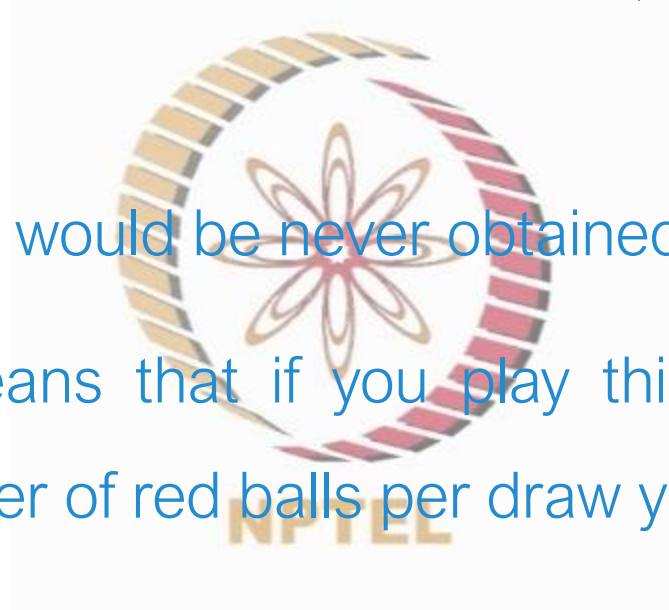


Expected Value

- The total number of red balls we will get after 1000 attempts at the game will be $0*20 + 1*70 + 2*320 + 3*430 + 4*160=2640$
- Alternatively, we can say that we obtain 2.64 red balls per experiment
- The expected value of an event is defined as: $X_1*P(X=x_1) + x_2*P(X=x_2) + x_3*P(X=x_3) + \dots$ so on, till $\dots x_n*P(X=x_n)$
- Expected value of the event X (X=0, 1, 2, 3, 4): $EV = 0*P(X=0) + 1*P(X=1) + 2*P(X=2) + 3*P(X=3) + 4*P(X=4)$: $0*(0.02) + 1*(0.07) + 2*(0.32) + 3*(0.43) + 4*(0.16) = 2.64$

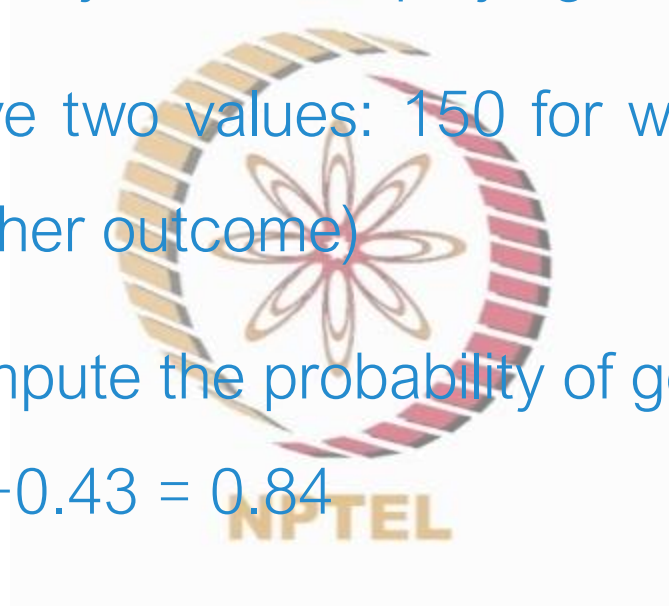
Expected Value

- The expected value for our random variable X , that is, the number of red balls, is 2.64
- Interestingly, this number would be never obtained in any of the draw
- This expected value means that if you play this game infinite number of times, the average number of red balls per draw you expect is 2.64
- In our case, a more efficient random variable would be the expected amount won or lost in a game



Expected Value

- So we can define: X = money won after playing the game once
- For this variable, we have two values: 150 for winning (getting 4-red balls) and -10 for losing (any other outcome)
- For example, we can compute the probability of getting 0, 1, 2, 3 red balls as follows: $0.02+0.07+0.32+0.43 = 0.84$
- Also, the probability of getting 4 red balls is 0.16
- Now we know that for winning probability is 0.16 and for losing the probability is 0.84



Expected Value

- We can compute the expected value as: $(150 \cdot 0.16) + (-10 \cdot 0.84)$, which is equal to +15.6 dollars
- That is, a player would expect to win \$15.6 dollars by playing, and therefore, the game is not profitable for the gambling house
- If the house makes money, it needs to ensure that the expected value won by the player is negative
- What to do: decrease the prize money, increase the penalty, decrease the players chances of winning, etc.



Expected Value: Example

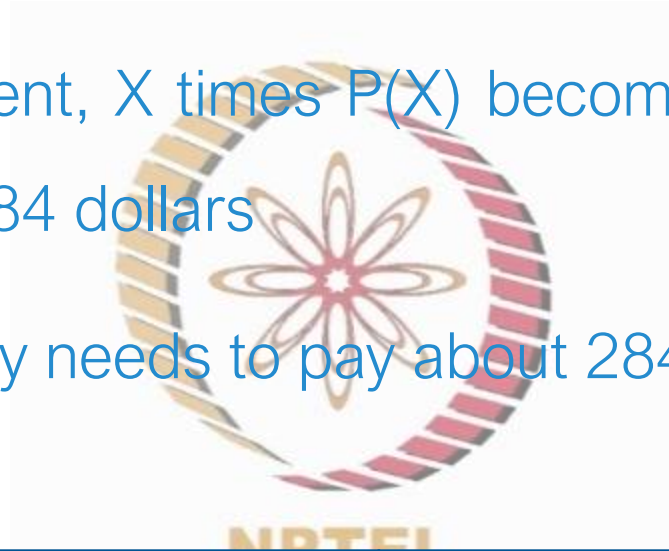
- An insurance company estimates the probability that an accident will occur within the next year is 0.00071. Basis this information, what premium should the insurance company charge to break even on a \$400,000 1-year term policy?



Event	x	$P(x)$	$xP(x)$
Live	0	0.99929	0.00
Die	400,000	0.00071	284.00
Total		1.00000	284.00

Expected Value: Example

- In case of no accident, X times $P(X)$ becomes 0 times 0.99929 which is 0
- And in case of an accident, X times $P(X)$ becomes 0.00071 times 400,000 which is approximately 284 dollars
- On average, the company needs to pay about 284 dollars to settle a policy



Event	x	$P(x)$	$x \cdot P(x)$
Live	0	0.99929	0.00
Die	400,000	0.00071	284.00
Total		1.00000	284.00



Variance

Variance

- Similar to the mean or expected value of a measure, we also examine the spread or variability of the data
- This can also be computed using the probability distribution of the random variable
- $\text{Var}(x) = (x - \mu)^2 \text{ times } P(x) \text{ summed over all the values of } X$
- Or alternatively $\sum_{i=1}^n (x_i - \mu)^2 * P(X = x_i)$, here μ (or sometimes \bar{x}) is the expected value or mean of the random variable (X)



Variance

- Recall our probability table

X	Probabilities	$(x - \mu)^2 * P(x)$
0	0.02	$(0 - 2.64)^2 * 0.02$
1	0.07	$(1 - 2.64)^2 * 0.07$
2	0.32	$(2 - 2.64)^2 * 0.32$
3	0.43	$(3 - 2.64)^2 * 0.43$
4	0.16	$(4 - 2.64)^2 * 0.16$
Total	1	0.8104

Binomial Distribution: I



Binomial Distribution

- Let us examine a very commonly occurring probability distribution: Binomial distribution
- Consider a simple set-up with 2 red and 1 blue balls, what is the probability of drawing 1, 2, and 3 red balls
- The probability of getting a red ball in one trial is $\frac{2}{3}$ and the probability of getting a blue ball is $\frac{1}{3}$
- If you are drawing 4 balls, what is the probability of drawing 4 balls from the bag



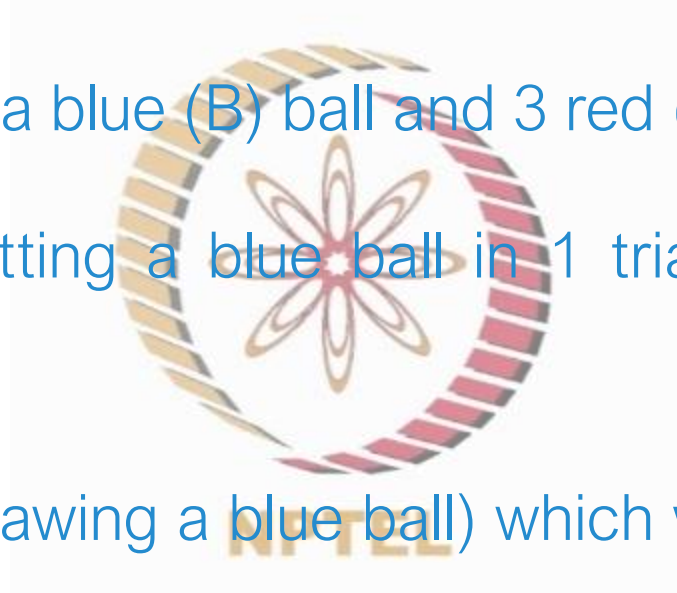
Binomial Distribution

- As per the multiplication rule, the probability of events 1 (E1) and 2 (E2) happening is $P(E1)*P(E2)$
- Here, we have four events, getting 4 red balls, the probability of the event is: $P(\text{Event 1})*P(\text{Event 2})*P(\text{Event 3})*P(\text{Event 4})$
- the probability of getting 4 red balls after 4 trials = Probability of getting a red ball in the first trial * Probability of getting a red ball in the 2nd trial * Probability of getting a red ball in the 3rd trial * Probability of getting a red ball in the 4th trial = $\frac{2}{3}*\frac{2}{3}*\frac{2}{3}*\frac{2}{3}$
 $= 0.197$



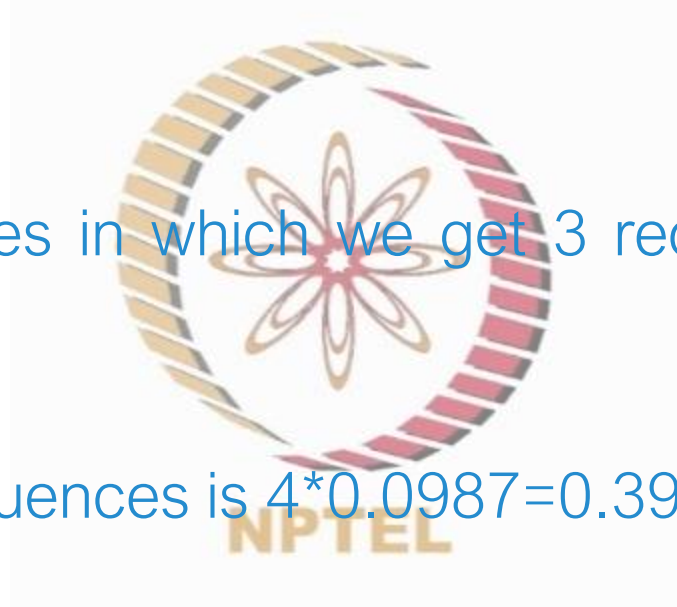
Binomial Distribution

- Remember that we are replacing the red ball again after drawing it
- Consider a case of drawing a blue (B) ball and 3 red (R) balls 'BRRR'
- $\frac{1}{3}$ is the probability of getting a blue ball in 1 trial and $\frac{2}{3}$ is the probability of getting a red ball in 1 trial
- With multiplication rule: $P(\text{drawing a blue ball})$ which would be $\frac{1}{3}$ times $P(\text{drawing a red ball})$ which would be $\frac{2}{3}$ times $P(\text{drawing a red ball})$ which would again be $\frac{2}{3}$ times $P(\text{drawing a red ball})$: $(\frac{1}{3}) * (\frac{2}{3}) * (\frac{2}{3}) * (\frac{2}{3}) = 0.0987$



Binomial Distribution

- There are various possible combinations in which we can get 3 red balls and 1 blue ball
- There are 4 such sequences in which we get 3 red balls and 1 blue ball- RBRR, RRBR and RRRB
- The probability of these sequences is $4 \times 0.0987 = 0.3948$
- This is rule of addition, that is, for independent events E1 and E2, the probability of E1 or E2 is $P(E1) + P(E2)$

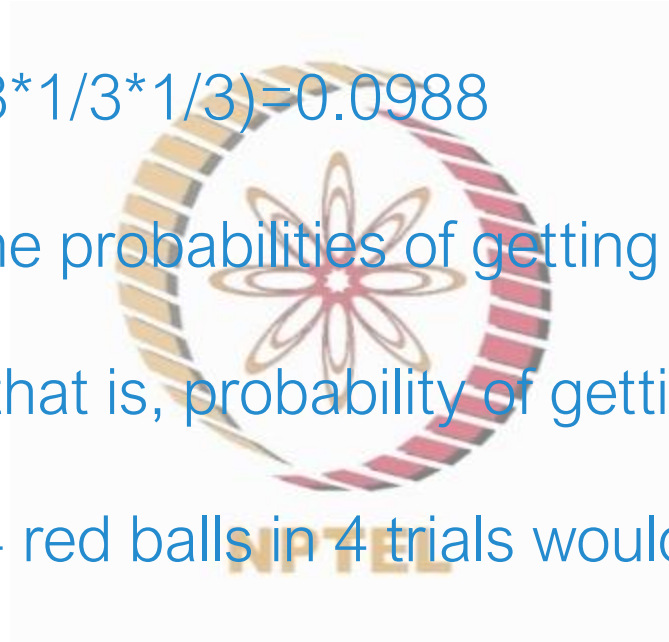


Binomial Distribution: II



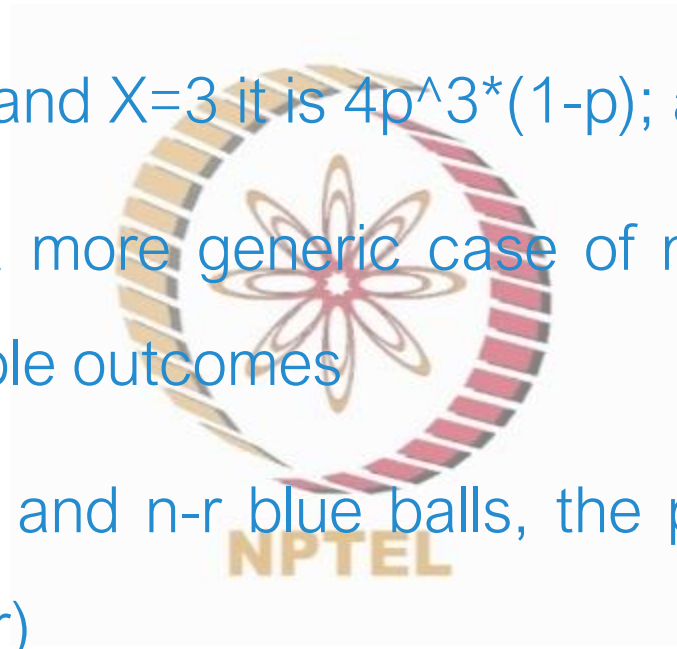
Binomial Distribution

- Consider an event $X=1$, one red ball and three blue balls
- The probability is $4 \cdot (2/3 \cdot 1/3 \cdot 1/3 \cdot 1/3) = 0.0988$
- Similarly, we can compute the probabilities of getting 0, 2 or 4 red balls
- Let us generalize this case, that is, probability of getting a red ball is p
- Thus, probability of getting 4 red balls in 4 trials would be $= p^4$
- Probability of getting 3 red balls $= 4 \cdot p^3 \cdot (1-p)$
- The probability that 0 red balls are drawn in 4 trials is $(1-p)^4$



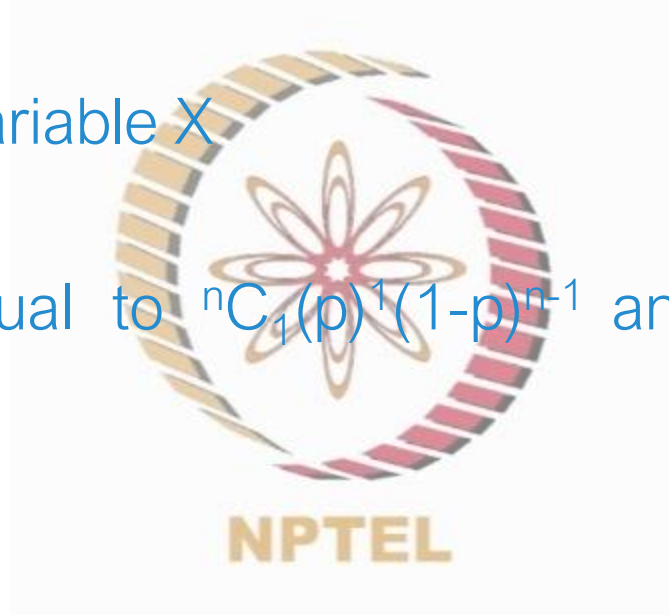
Binomial Distribution

- Probability of taking 1 red and 3 blue balls is $4 \cdot p \cdot (1-p)^3$
- For $X=2$, it is $6 p^2 (1-p)^2$; and $X=3$ it is $4 p^3 (1-p)$; and $X=4$, it is p^4
- Now let us extend this to a more generic case of making n draws with a success probability of p and r favorable outcomes
- For $P(X=r)$, e.g., r red balls and $n-r$ blue balls, the probability of getting one such combination is $p^r (1-p)^{(n-r)}$
- Also, there are nCr combinations of getting r red balls out of total n balls
- The resulting probability is $nCr \cdot p^r (1-p)^{(n-r)}$



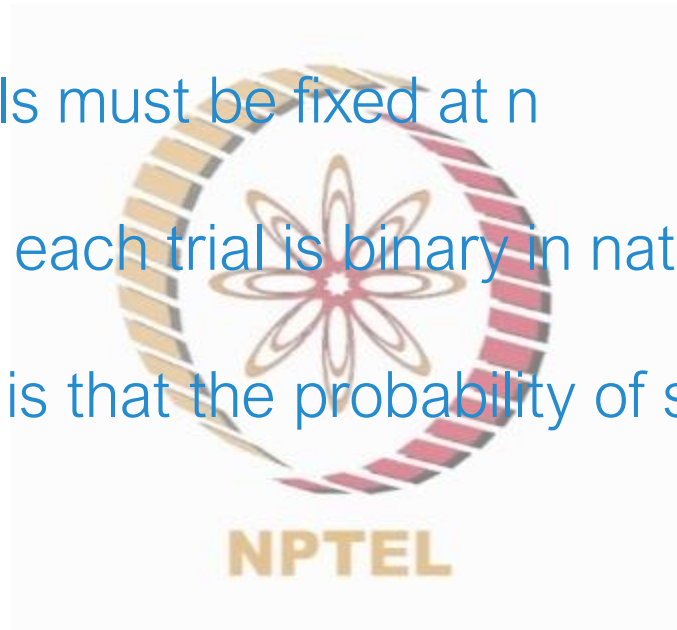
Binomial Distribution

- Using this formula for different values of $r=0,1,2,3\dots n$, we can find the probability distribution of the random variable X
- E.g., $P(X=1)$ would be equal to ${}^nC_1(p)^1(1-p)^{n-1}$ and $P(X=n)$ would be equal to ${}^nC_n(p)^n(1-p)^{0}$
- This probability distribution is called the binomial probability distribution
- Under what conditions we can use the binomial distribution?

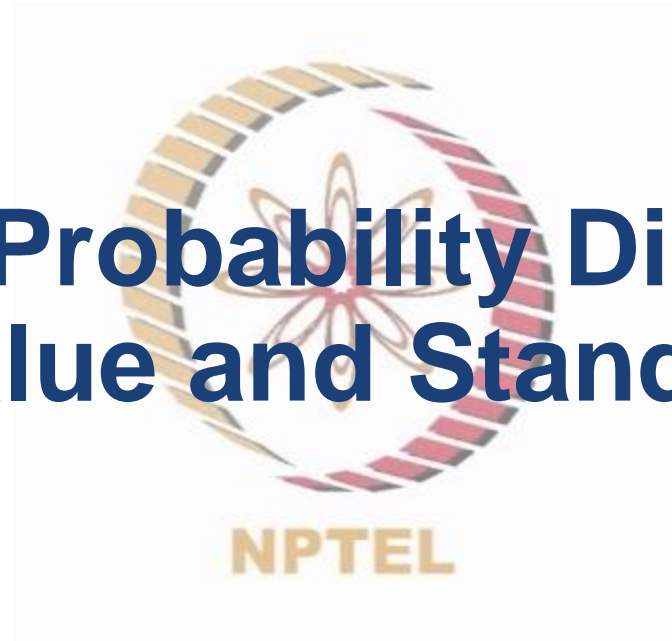


Binomial Distribution

- Under what conditions we can use the binomial distribution?
- First, the total number of trials must be fixed at n
- The second condition is that each trial is binary in nature
- The third and final condition is that the probability of success is the same in all trials, denoted by p
- When all of these conditions are satisfied, then the random variable will follow a binomial distribution and the probability for $X = r$, that is, getting r successes in n trials, can be calculated as ${}^nC_r(p)^r(1-p)^{n-r}$, as given by the binomial distribution

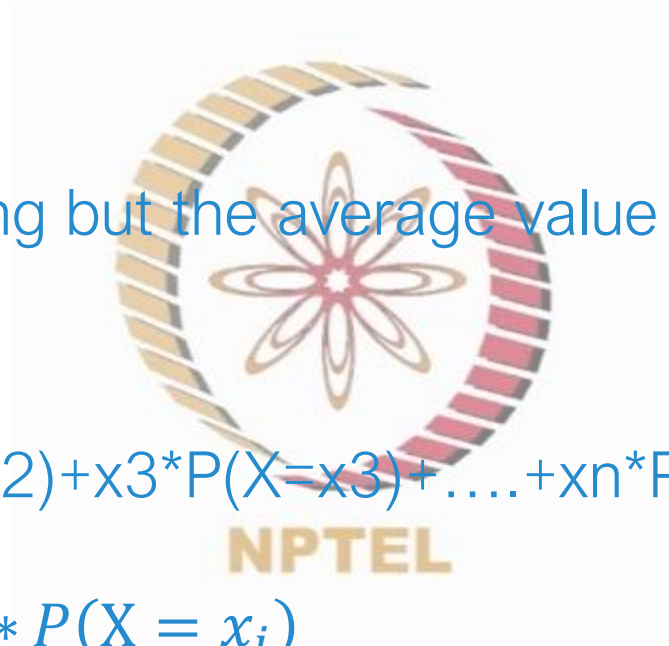


Binomial Probability Distribution – Expected Value and Standard Deviation



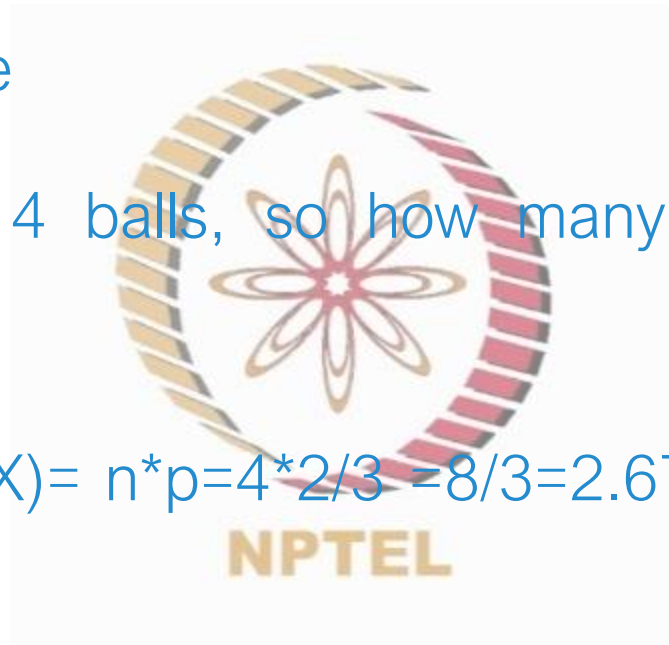
Expected Value and Standard Deviation

- In our earlier experiment, where in we got 2.64 red balls per game as the expected value
- The expected value is nothing but the average value that we would 'expect' to get for a random variable
- $E(X) = x_1 * P(X=x_1) + x_2 * P(X=x_2) + x_3 * P(X=x_3) + \dots + x_n * P(X=x_n);$
- And, $Var(x) = \sum_{i=1}^n (x_i - \mu)^2 * P(X = x_i)$
- For binomial distributions: $E(x) = n * p$ and $Var(x) = np(1-p)$

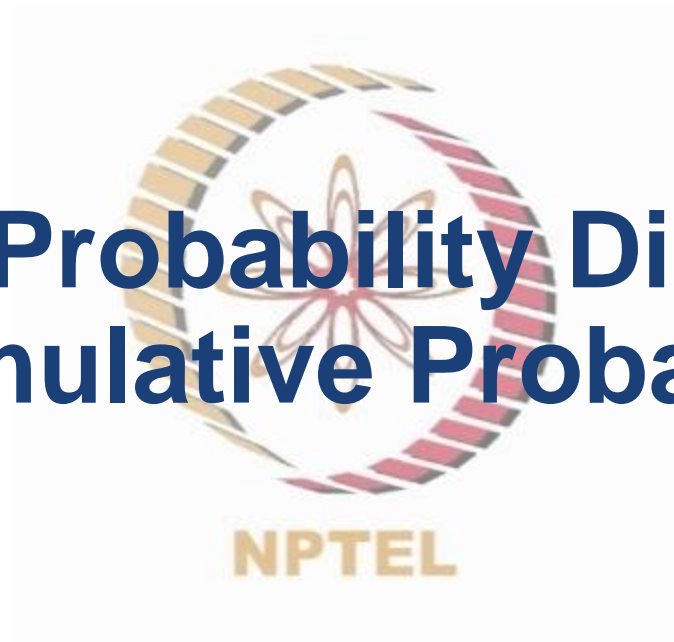


Expected Value and Standard Deviation

- Let us revisit our example to calculate the number red balls to calculate the expected value and variance
- The participant can draw 4 balls, so how many red balls on average, the participant can draw
- Here $n=4$ and $p=2/3$, so $E(X) = n \cdot p = 4 \cdot 2/3 = 8/3 = 2.67$, i.e., the expected value of getting a red ball in a game
- Similarly, $\text{Var}(x) = n \cdot p \cdot (1-p) = 4 \cdot 2/3 \cdot 1/3 = 8/9 = 0.88$

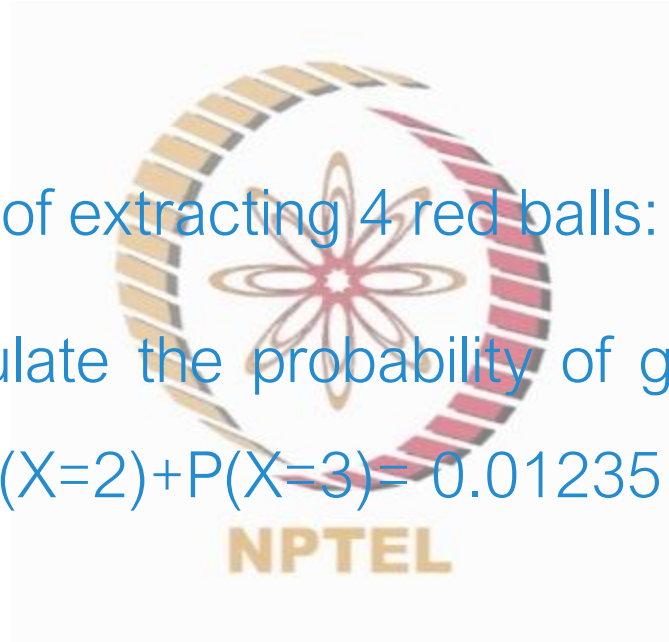


Binomial Probability Distribution – Cumulative Probability



Cumulative Probability

- Using the binomial distribution, we were able to calculate the probability of getting an exact value
- For example, the probability of extracting 4 red balls: ${}^4C_4(2/3)^4(1/3)^0 = 0.19753$
- What if we wanted to calculate the probability of getting less than equal to 3 red balls, i.e., $P(X=0)+P(X=1)+P(X=2)+P(X=3) = 0.01235 + 0.09877 + 0.2963 + 0.39506 = 0.80247$
- That is an 80.2% chance that any randomly selected participant will have selected maximum 3 red balls while drawing 4 balls



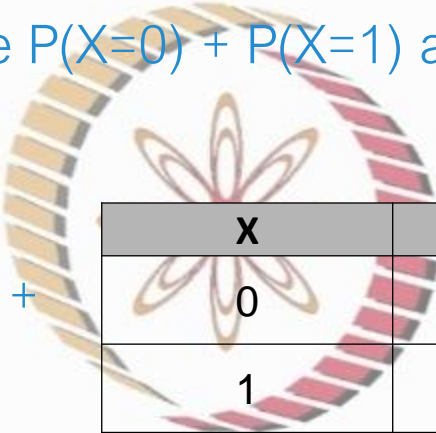
Cumulative Probability

- Any probability where we have to determine the likelihood of X being less than a certain number is called a cumulative probability
- For example, 0.802 is the cumulative probability for $X \leq 3$. And instead of saying $P(X \leq 3)$ is 0.802, we can use $F(X = 3)$ is 0.802, where F represents the cumulative probability
- For $X=0$, $F(X=0)$ is also same as $P(X \leq 0)$, since $X \leq 0$ takes only one value, $X=0$, then this is also same as $P(X=0)$

X	$P(X=x)$	$F(X=x)$
0	$= {}^4C_0 * \left(\frac{2}{3}\right)^0 * \left(\frac{1}{3}\right)^4 = 0.01235$	0.01
1	$= {}^4C_1 * \left(\frac{2}{3}\right)^1 * \left(\frac{1}{3}\right)^3 = 0.09877$	0.11
2	$= {}^4C_2 * \left(\frac{2}{3}\right)^2 * \left(\frac{1}{3}\right)^2 = 0.2963$	0.41
3	$= {}^4C_3 * \left(\frac{2}{3}\right)^3 * \left(\frac{1}{3}\right)^1 = 0.3951$	0.80
4	$= {}^4C_4 * \left(\frac{2}{3}\right)^4 * \left(\frac{1}{3}\right)^0 = 0.1975$	1.00

Cumulative Probability

- Next, let's calculate $F(X = 1)$. Here, $F(X = 1)$ means $P(X \leq 1)$, which will be $P(X = 0) + P(X = 1)$, which will come to 0.11111
- Similarly, we have $F(X=1)$, which will be $P(X=0) + P(X=1)$ and $F(X=2)$, which will be $P(X=0) + P(X=1) + P(X=2)$
- Thus, we can write $F(X=2)$ as $F(X=1) + P(X=2)$
- Hence, $F(X=2)$ will be $0.111 + 0.296 = 0.41$
- And $F(X=4)$ will be $F(X=3) + P(X = 4)$ which is $0.80247 + 0.19753 = 1.00$



X	P(X=x)	F(X=x)
0	$=4C_0 * \left(\frac{2}{3}\right)^0 * \left(\frac{1}{3}\right)^4 = 0.01235$	0.01
1	$=4C_1 * \left(\frac{2}{3}\right)^1 * \left(\frac{1}{3}\right)^3 = 0.09877$	0.11
2	$=4C_2 * \left(\frac{2}{3}\right)^2 * \left(\frac{1}{3}\right)^2 = 0.2963$	0.41
3	$=4C_3 * \left(\frac{2}{3}\right)^3 * \left(\frac{1}{3}\right)^1 = 0.3951$	0.80
4	$=4C_4 * \left(\frac{2}{3}\right)^4 * \left(\frac{1}{3}\right)^0 = 0.1975$	1.00

Continuous Probability Distributions: Continuous random variables



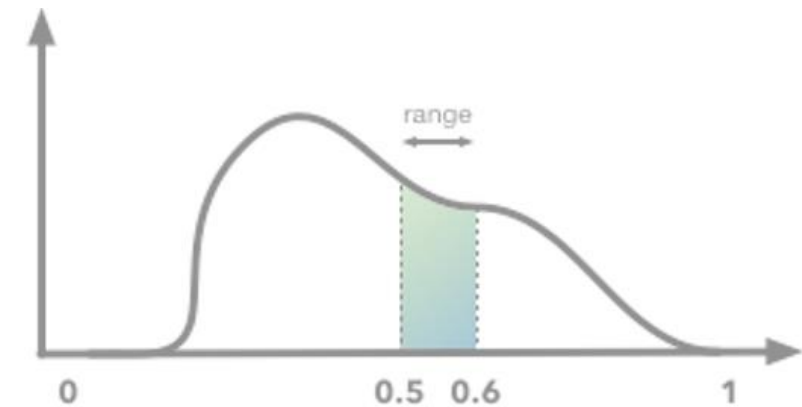
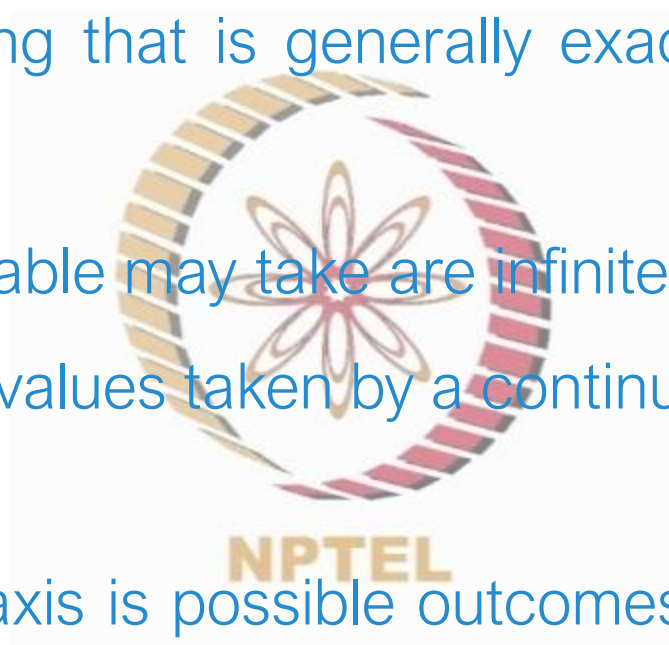
Continuous random variables

- Now we will discuss continuous probability distributions
- The random variable used to define the continuous probability distribution is called the continuous random variable
- If you are a pizza delivery manager, you are concerned with the average time it takes for a pizza delivery to reach a customer. Let us call this delivery time as random variable 'X'
- This variable can take on various values like 5-min, 30-min, 15.13-min
- Here defining the variable exact time, is defined as continuous random variable, i.e., it can be defined even up to last milliseconds



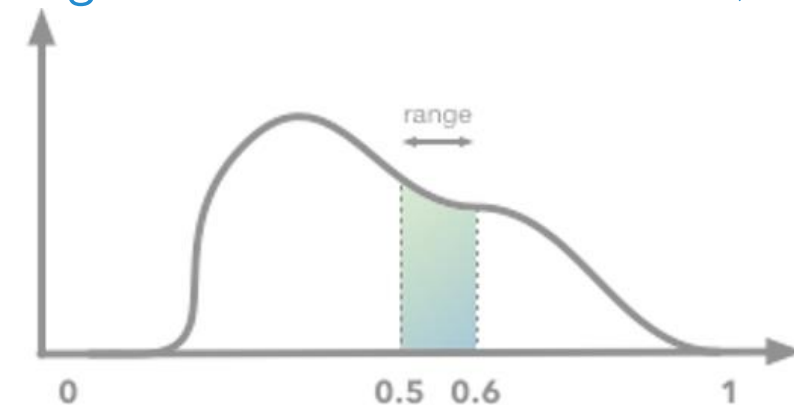
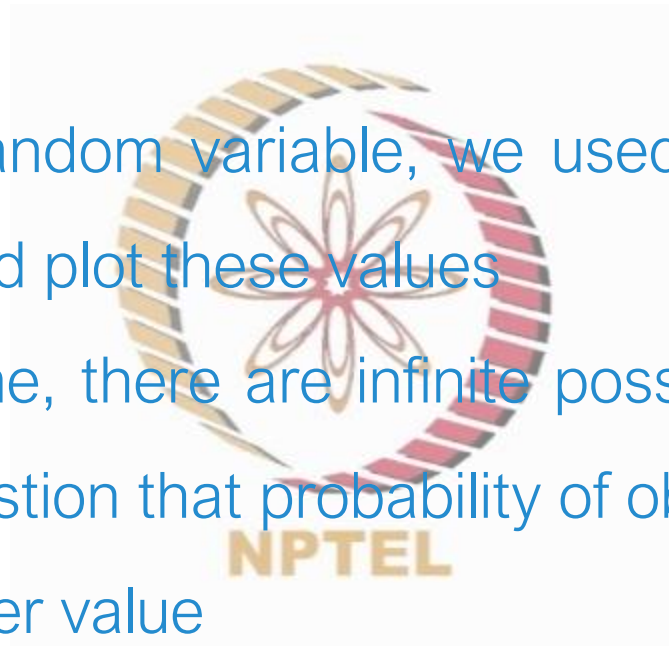
Continuous random variables

- The 'amount of water present in a bottle' or 'the exact stock price at the end of a trading session', or anything that is generally exact, this is always going to be a continuous random variable
- The values that random variable may take are infinite. For example, even between 20 to 20.1 there can be infinite values taken by a continuous random variable
- Let us draw a plot where x-axis is possible outcomes of random variable
- In case of continuous random variables, it is difficult to define/obtain the probability of a specific value of X



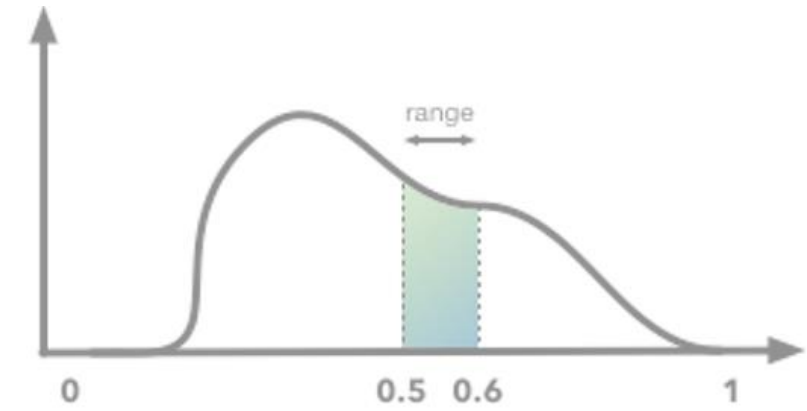
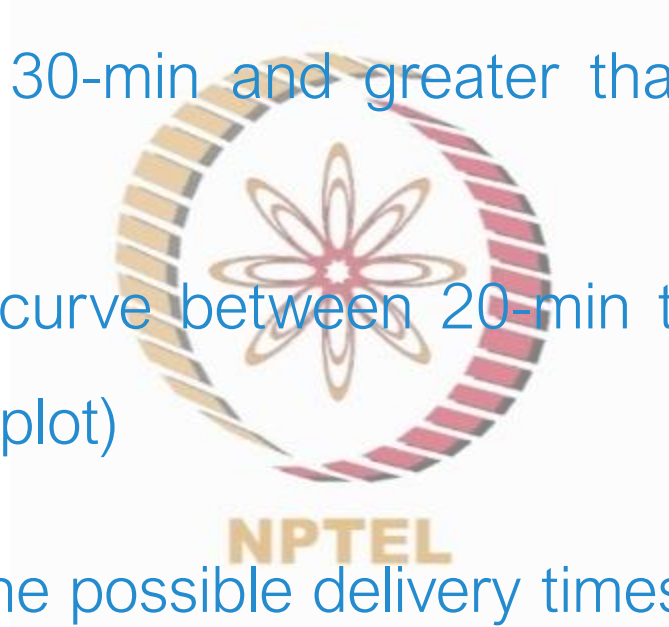
Continuous random variables

- Thus, a continuous random variable is represented by a continuous line known as probability density function
- In the case of a discrete random variable, we used to calculate the probability by saying $P(X=0)$ or $P(X=1)$, and plot these values
- In case of pizza delivery time, there are infinite possible outcomes for a continuous random variable; so the question that probability of observing 15-min is close to zero, and same is true for any other value
- So instead of X being any specific value, we measure the specific probability of X lying in a certain interval



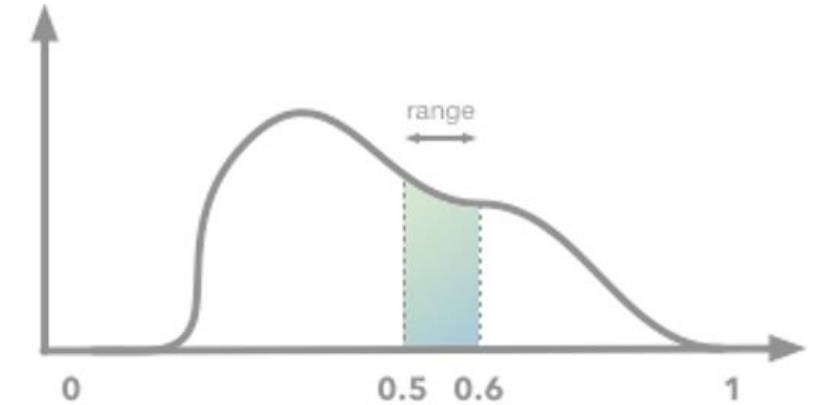
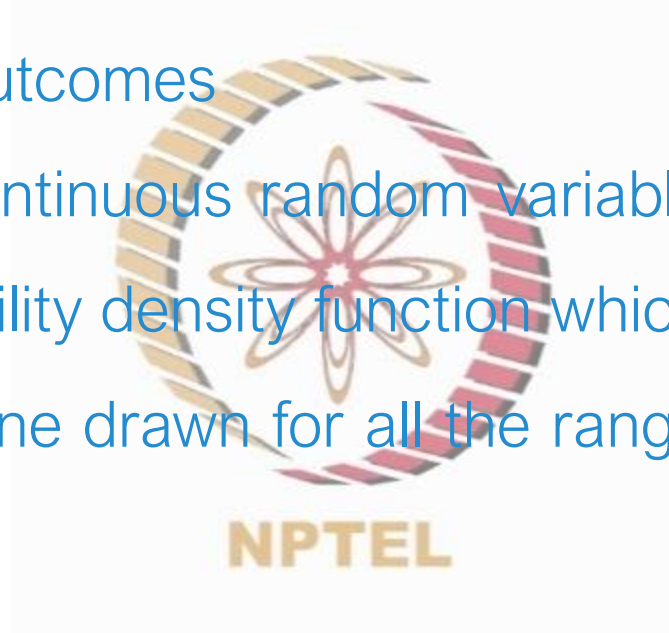
Continuous random variables

- For example, we will define the cumulative probability as X is less than equal to 30-min and greater than equal to 20-min
- This is the area under the curve between 20-min to 30-min (colored area in the plot)
- Also, if we assume that all the possible delivery times range in between 0 to 1 hour, the X will range from 0 to 1-hour, and the area of the curve in this range will be 1



Continuous random variables

- First, a continuous random variable is such that it can have an infinite number of outcomes
- Second, we represent a continuous random variable using what we call a probability density function which is essentially a continuous line drawn for all the range of values that X can take



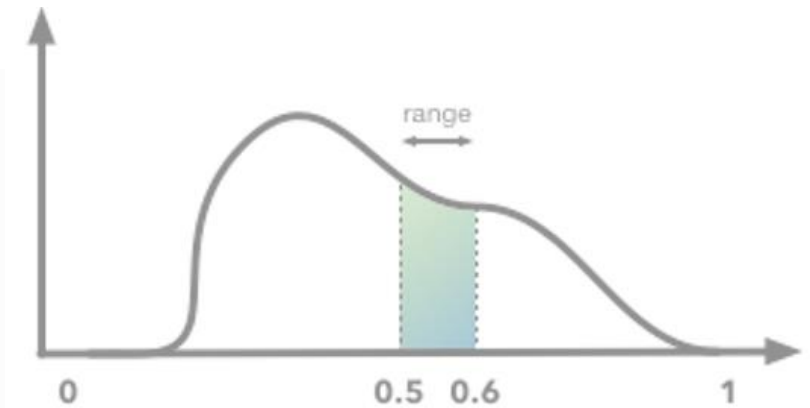
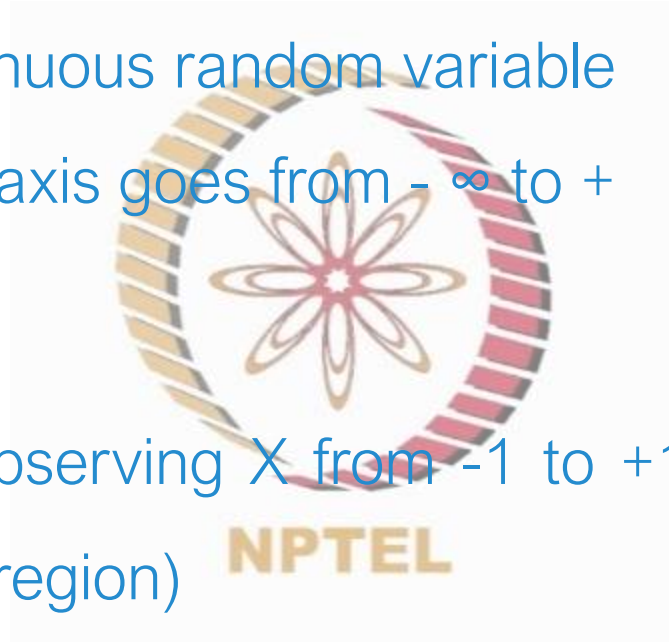
- Thirdly, the area under the curve, represents the probability that the random variable lies in that interval
- Finally, the total area under the curve will always be equal to 1

Continuous Probability Distributions: Cumulative probability for continuous Random Variables (RV)



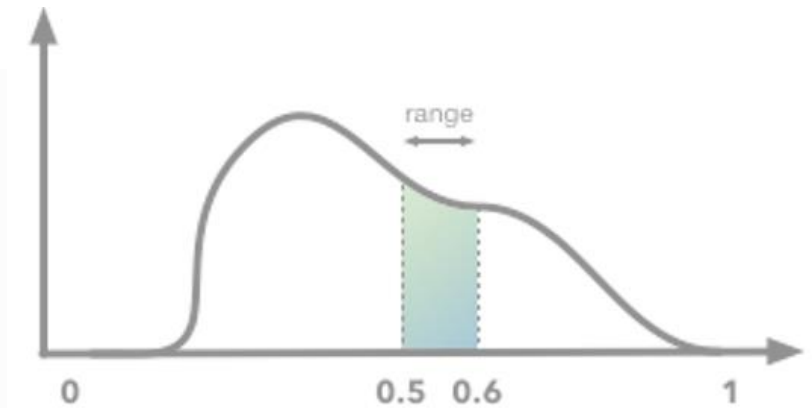
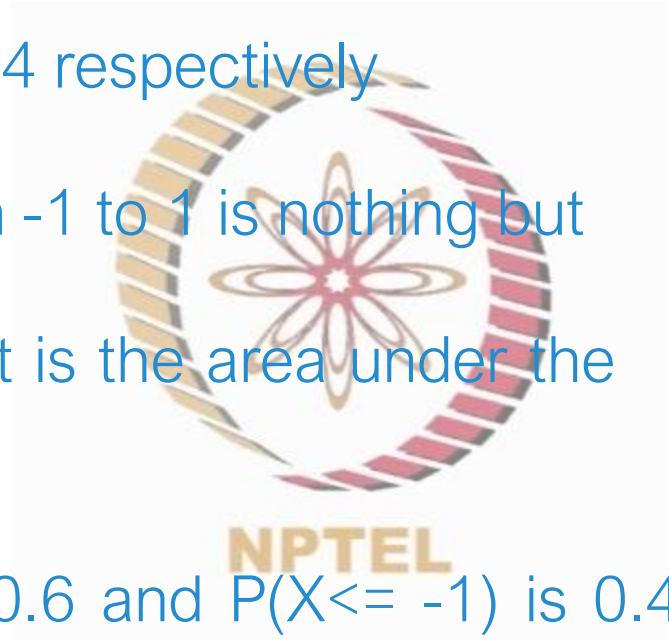
Cumulative probability for continuous RV

- Consider the following example. The probability density function of our continuous random variable is plotted on x-y axis. The x-axis goes from $-\infty$ to $+\infty$.
- What is the probability of observing X from -1 to $+1$, that is area under the curve from $X=-1$ to $X=+1$ (colored region)
- When we talk about cumulative probabilities, we are always dealing with ranges

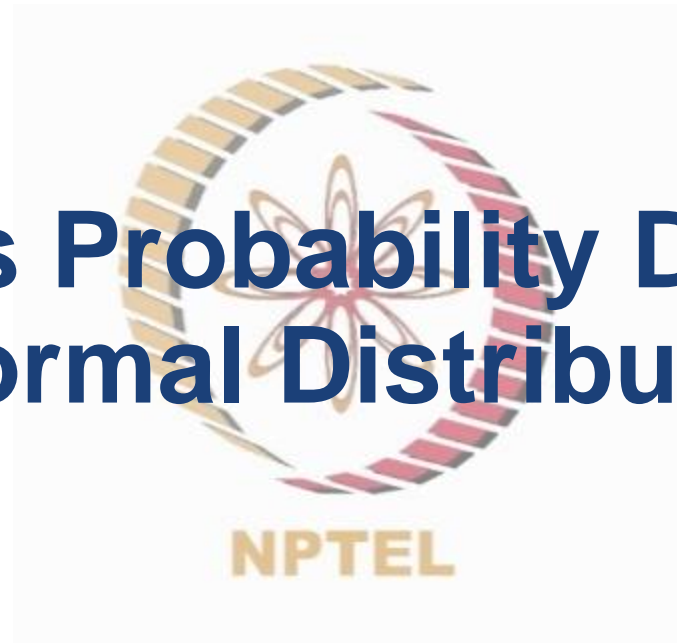


Cumulative probability for continuous RV

- If you are given that the cumulative probability for $X=1$ and $X=-1$, as 0.6 and 0.4 respectively
- The probability of X between -1 to 1 is nothing but
- $P(X \leq 1) - P(X \leq -1)$, that is the area under the curve from -1 to 1
- In this case, $P(X \leq 1)$ is 0.6 and $P(X \leq -1)$ is 0.4, so the probability of X lying between 1 and -1 is $(0.6-0.4)$ which is 0.2

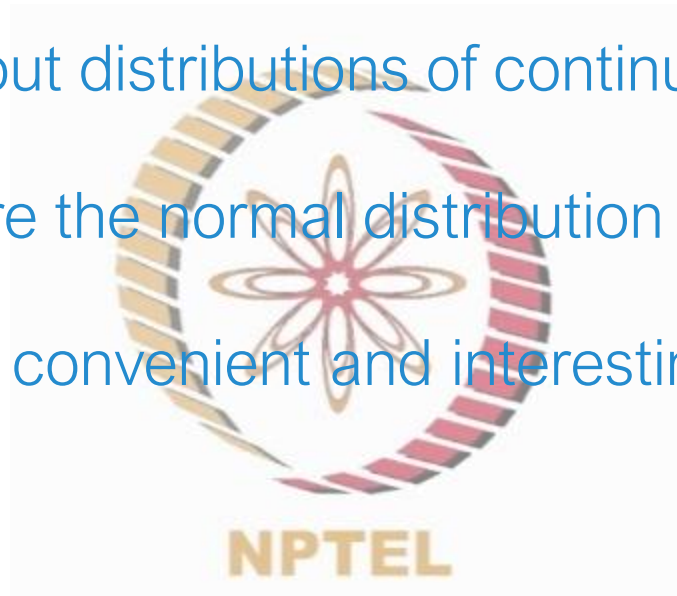


Continuous Probability Distributions: Normal Distribution



Normal Distribution

- The normal distribution is perhaps the most widely used and most important distribution when talking about distributions of continuous random variables
- There are many places where the normal distribution appears naturally
- Normal distribution has very convenient and interesting properties that makes it easy to work with them
- Also, normal distribution is an integral part of central limit theorem



Normal Distribution

- Let us go back to our earlier example of the commute time for delivering pizza to the customer's houses
- Remember the 30-minute discount scheme to deliver the pizza. If the time is more than 30-minutes, then pizza is free or available with large discounts
- As a manager at one of the pizza outlets and given this 30-minute guarantee, you want to ensure that most of the pizzas are delivered well before these 30 minutes
- Let us consider only those commutes from the pizza outlet to the customer's location and not include the deliveries where the delivery boy has to visit multiple locations as he is delivering multiple orders

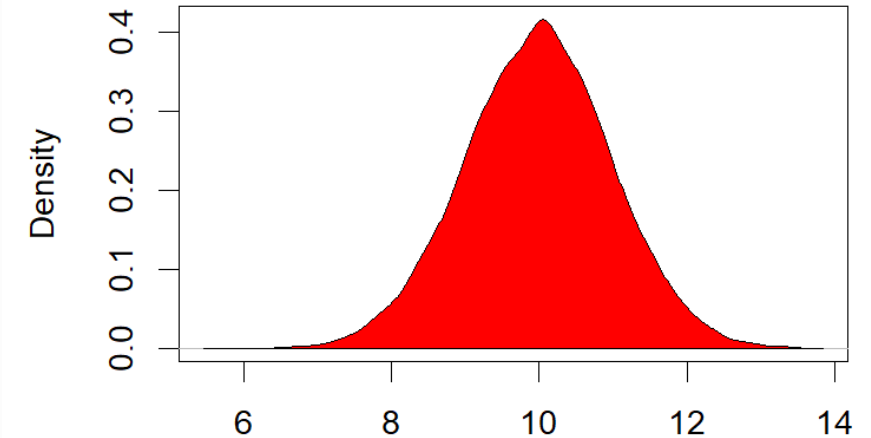


Normal Distribution

- Let us assume that it takes a maximum of 10 minutes to make the pizza. That leaves us with only 20 minutes to deliver the pizza from the outlet to the customer's location
- The commute time may vary from location to location
- The probability density function for such a scenario may appear like a normal distribution : Bell shaped curve



Normal Probability Density Function

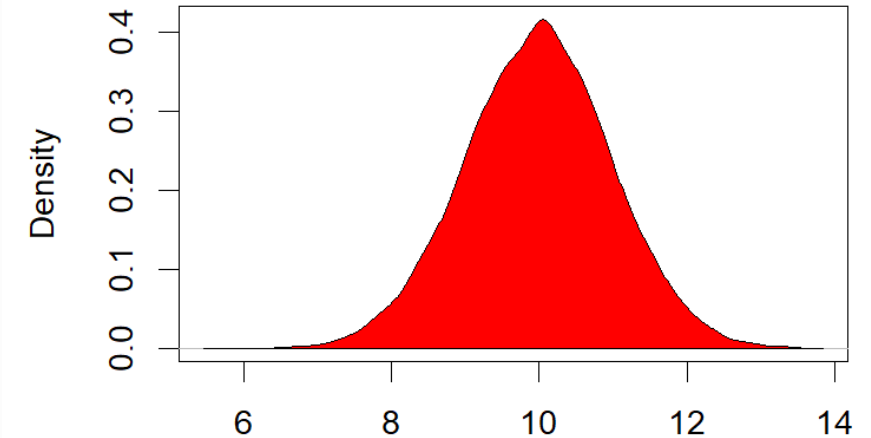


Normal Distribution

- This chart is known as a normal distribution
- You can see that the probability density is highest at 10 minutes
- We can also see that the probability density starts decreasing as we move towards the right or left
- Just by looking at this distribution, we can make several observations about the characteristics of this distribution

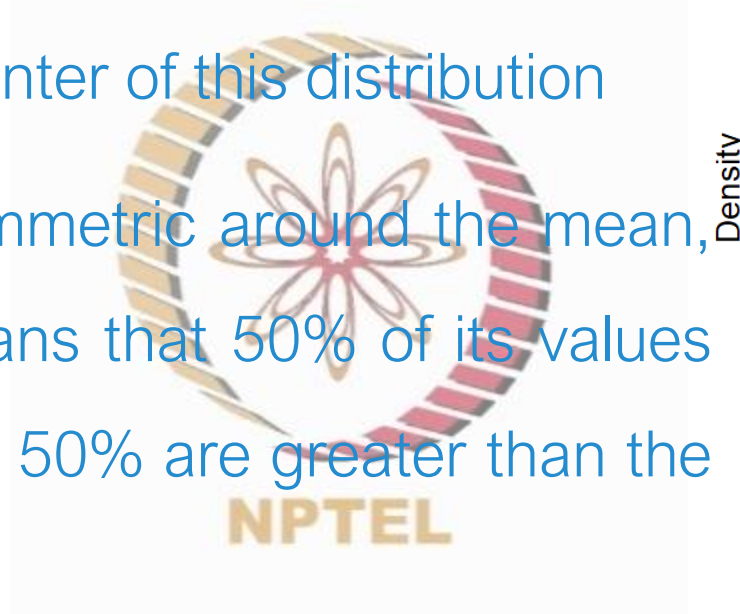


Normal Probability Density Function

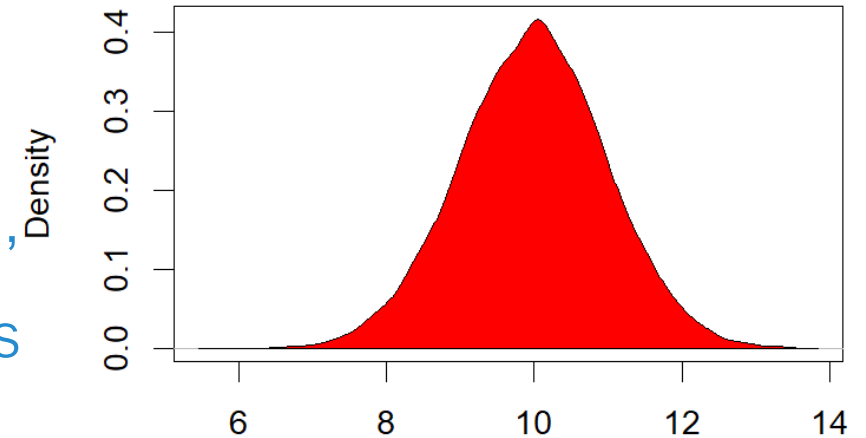


Normal Distribution

- The mean of this distribution, which in our case is 10 minutes, lies in the exact center of this distribution
- Since the distribution is symmetric around the mean, which is 10 minutes, it means that 50% of its values are less than the mean and 50% are greater than the mean
- we can see that the probability density is the highest at the mean, and decreases exponentially as we move further away from the mean

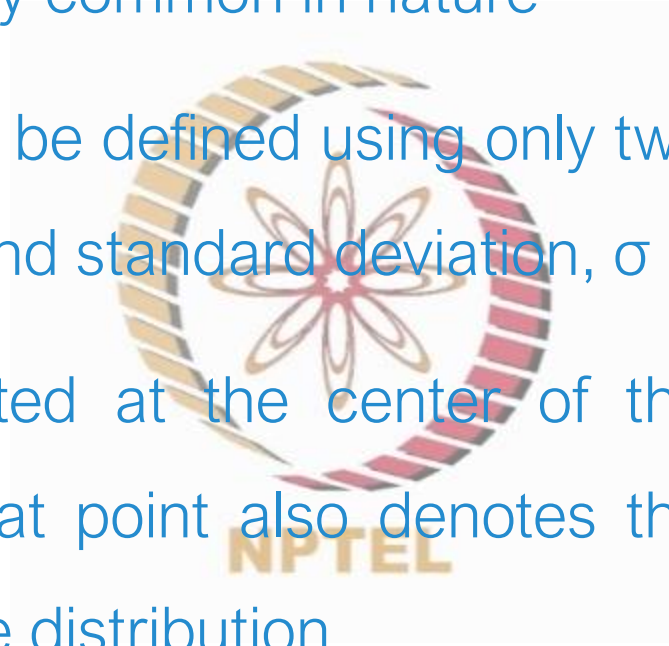


Normal Probability Density Function

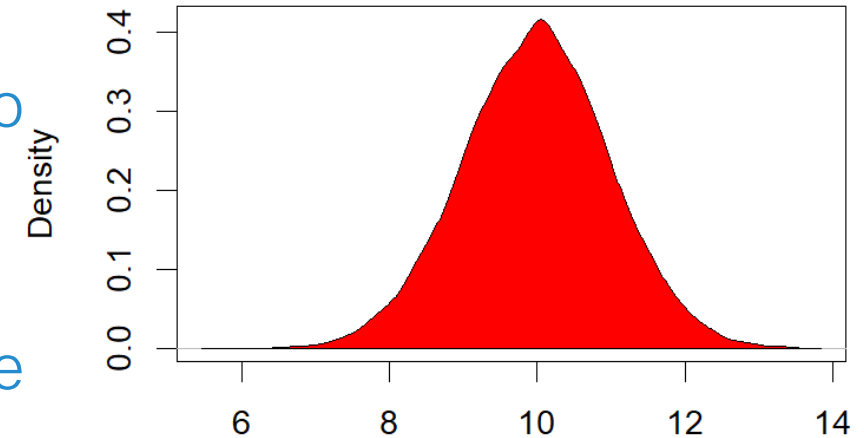


Normal Distribution

- Normal distributions are very common in nature
- Any normal distribution can be defined using only two parameters - the mean, μ and standard deviation, σ
- The mean ($\mu=10$) is located at the center of the normal distribution, and that point also denotes the median and the mode of the distribution
- The standard deviation (σ) determines how flat and wide the normal distribution curve is
- An increased σ indicates a more dispersed and flat distribution



Normal Probability Density Function

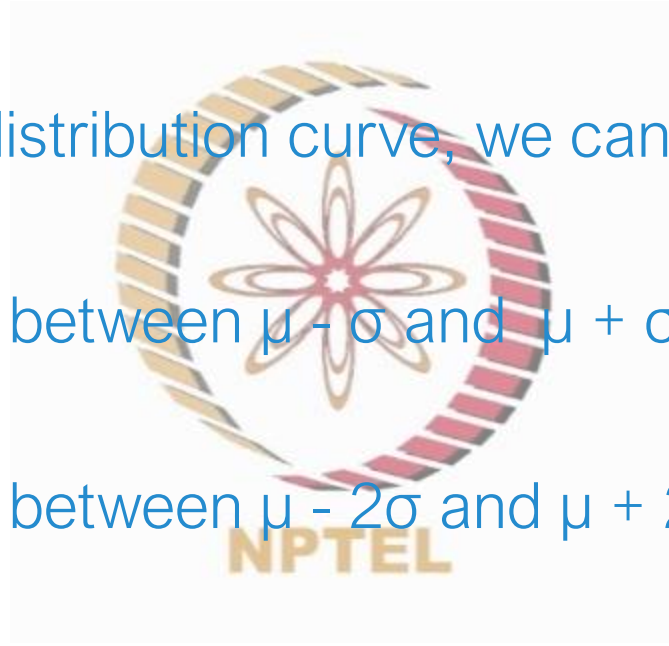


Probabilities for a Normal Distribution



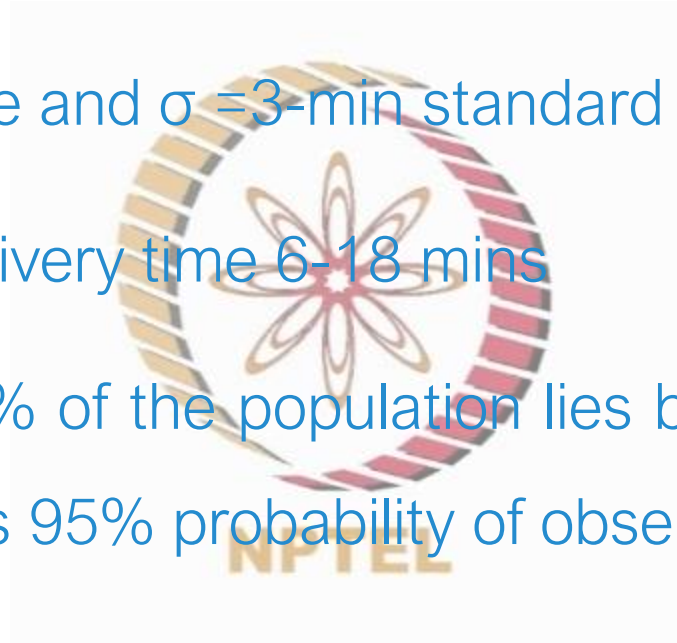
Probabilities for a Normal Distribution

- Assume a normal distribution with a mean of ' μ ' and standard deviation given by ' σ '
- Using properties of normal distribution curve, we can calculate probability values
 - The probability of X lying between $\mu - \sigma$ and $\mu + \sigma$ is around 68% or 0.68
 - The probability of X lying between $\mu - 2\sigma$ and $\mu + 2\sigma$ is around 95% or 0.95
 - And the probability of X lying between $\mu - 3\sigma$ and $\mu + 3\sigma$ is around 99.7% or 0.997



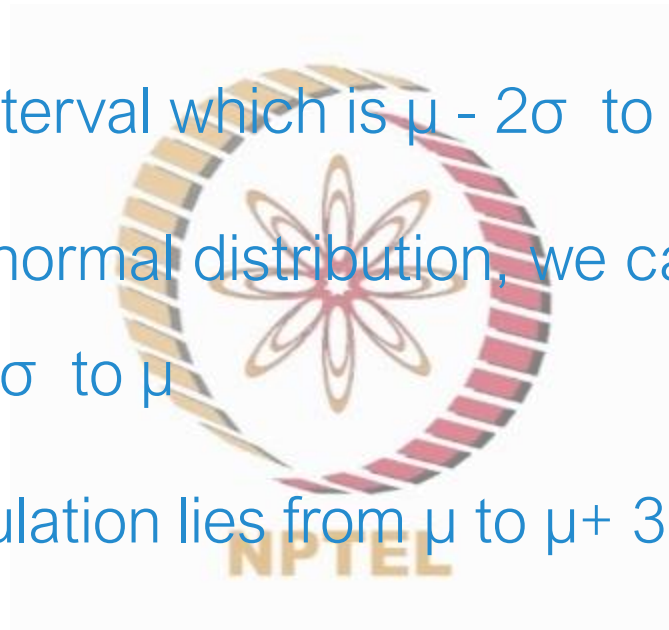
Probabilities for a Normal Distribution

- Let us revisit the Pizza delivery time example, assume a normal distribution with $\mu = 12$ -min average deliver time and $\sigma = 3$ -min standard deviation
- What is the probability of delivery time 6-18 mins
- Using our empirical rule 95% of the population lies between $\mu - 2\sigma$ and $\mu + 2\sigma$, i.e., 6-18 minutes; in other words 95% probability of observing a delivery time between 6-18 mins
- Similarly, 99.7% probability of observing a delivery time between 3-21 mins ($\mu - 3\sigma$ and $\mu + 3\sigma$)



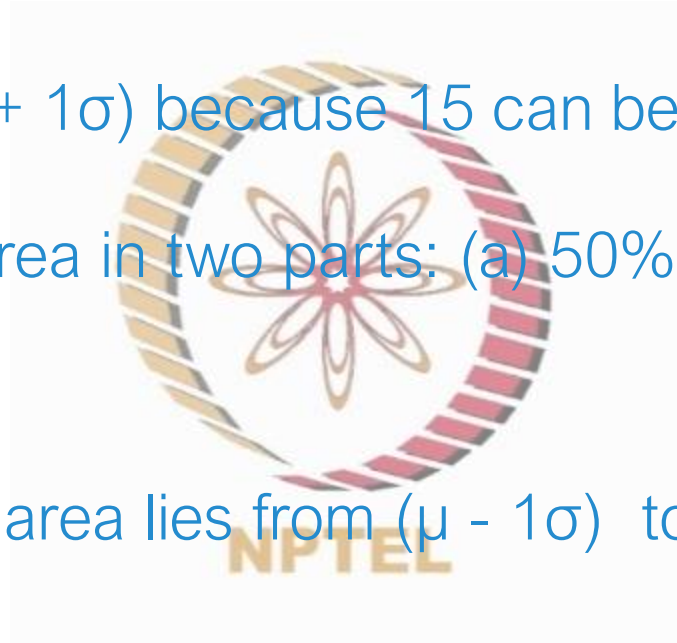
Probabilities for a Normal Distribution

- What is the probability of observing a delivery time between 6 and 21 minutes
- Now we are looking at the interval which is $\mu - 2\sigma$ to $\mu + 3\sigma$
 - Due to symmetry of the normal distribution, we can say that $95\%/2=47.5\%$ of the population lies from $\mu - 2\sigma$ to μ
 - And $99.7\%/2$ of the population lies from μ to $\mu + 3\sigma$
 - Adding the two, we get 97.35% of the population lie between 6-21 mins , i.e., $\mu - 2\sigma$ to $\mu + 3\sigma$



Probabilities for a Normal Distribution

- What is the probability of observing a delivery in less than 15-mins, i.e., $P(X \leq 15)$ or
- Probability of X less than $(\mu + 1\sigma)$ because 15 can be written as $(\mu + 1\sigma)$
- We can divide the desired area in two parts: (a) 50% of area up to $\mu=12$, and (b) 12-15 mins (μ to $\mu + 1\sigma$)
- As noted earlier 68% of the area lies from $(\mu - 1\sigma)$ to $(\mu + 1\sigma)$ or 34% area lies from μ to $\mu + 1\sigma$
- Therefore, $50\% + 34\% = 84\%$ of pizza deliveries are taking place in less than 15-minutes

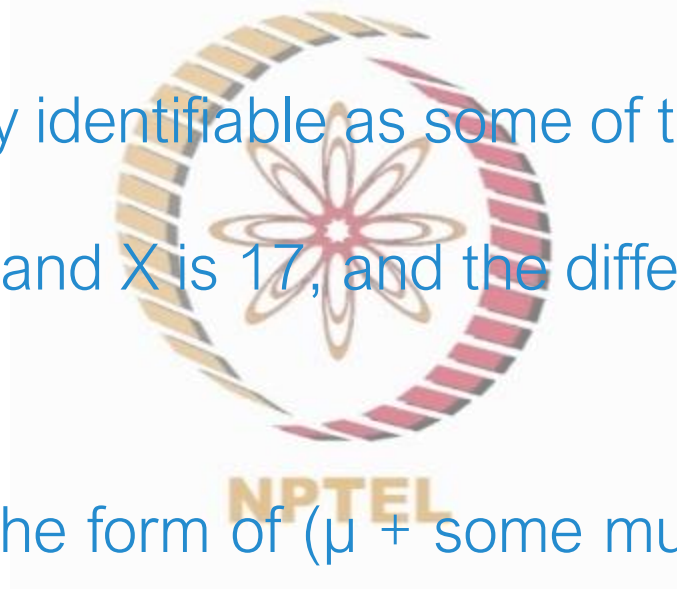


Standard Normal Distribution: Part I



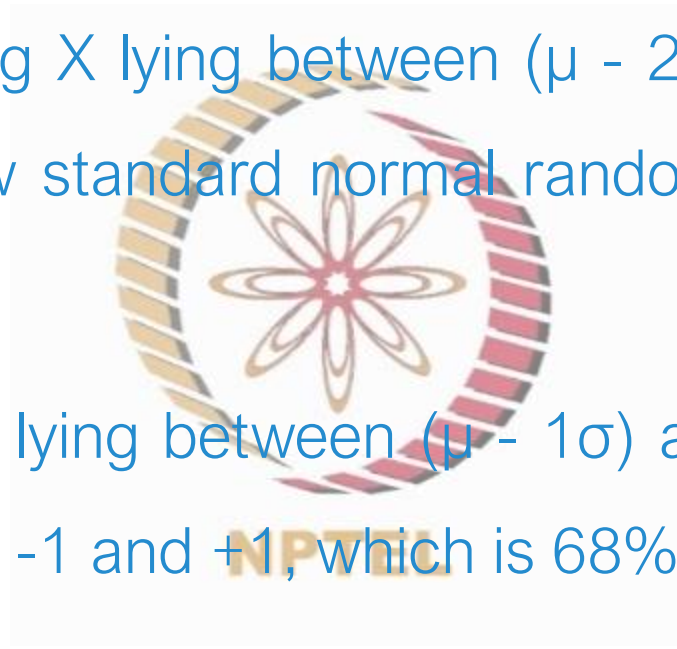
Standard Normal Distribution

- What if you want to find, 'what is the percentage of deliveries where the commute time is between 6 and 17 minutes or between 6 and 16.95 minutes
- These values are not as easily identifiable as some of the earlier examples
- Here we can see that μ is 12 and X is 17, and the difference between these two values should be 5
- if we have to represent X in the form of $(\mu + \text{some multiple times } \sigma)$, then we can find this multiplication factor by using $(X - \mu)$ divided by σ
- In our case, $(X - \mu) / \sigma$ comes out to be $(17 - 12) / 3$, which equates to 5 by 3, which is around 1.67



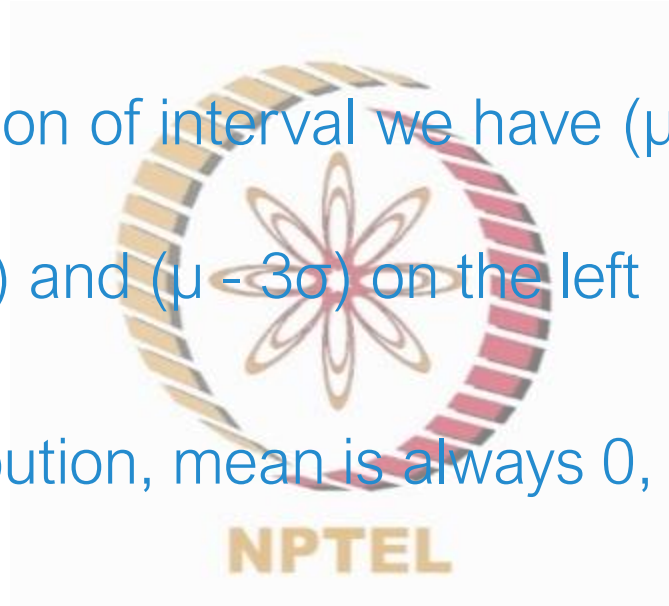
Standard Normal Distribution

- The value $(X - \mu) / \sigma$ is denoted by z and z is called the standard normal variable
- Now, the probability of finding X lying between $(\mu - 2\sigma)$ and $(\mu + 2\sigma)$ is same as the probability of finding the new standard normal random variable z between -2 to $+2$, which is 95%
- Similarly, the probability of X lying between $(\mu - 1\sigma)$ and $(\mu + 1\sigma)$ is the same as the probability of Z lying between -1 and $+1$, which is 68%
- For a random variable X , we can find the probability of X within a certain range, in the form of a standard normal variable $z = (X - \mu) / \sigma$
- Z follows standard normal distribution



Standard Normal Distribution

- What is the difference between normal distribution and standard normal distribution
- In a normal distribution notation of interval we have $(\mu + 1\sigma)$, $(\mu + 2\sigma)$ and $(\mu + 3\sigma)$ on the right and $(\mu - 1\sigma)$, $(\mu - 2\sigma)$ and $(\mu - 3\sigma)$ on the left
- In the standard normal distribution, mean is always 0, and standard deviation is 1
- $(\mu + 1\sigma)$ will mark to 1, $(\mu + 2\sigma)$ will mark to 2 and, similarly, on the left hand side as well, $(\mu - 1\sigma)$ will mark to -1 and so on



Example

- How do we proceed to solve the problem of calculating the probability that an employee will take less than 17 minutes to commute to the office?
- First, let us convert 17 into the Z value; Remember that the μ and σ for this distribution was 12 and 3 minutes
- Thus, calculating $(X - \mu) / \sigma$, we get $(17 - 12) / 3$, which equates to 5 by 3, which is 1.67
- $P(X \leq 17)$ is the same as the probability, $P(Z \leq 1.67)$; we are essentially calculating the cumulative probability for $Z = 1.67$
- This value can be found from excel [=1-NORM.DIST(-1.67, 0, 1, TRUE)] or Z-table

