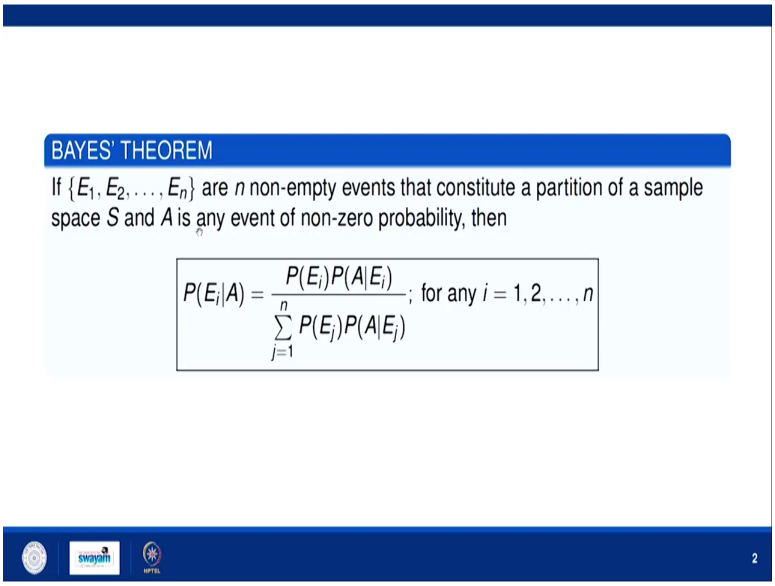
Essential Mathematics for Machine Learning
Prof. S. K. Gupta
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture – 32
Bayes' Theorem and Random Variables

Hello friends. So, welcome to lecture series on Essential Mathematics for Machine Learning. In the last lecture we have seen the basic concepts of probability. We have seen that if the classical definition of probability is basically number of favourable outcomes upon total number of outcomes and how we can define conditional probability, how we can see I mean multiplication rule, all those things were defined in the last lecture. Now, let us come to Bayes' Theorem, in this Bayes' Theorem and Random Variables.

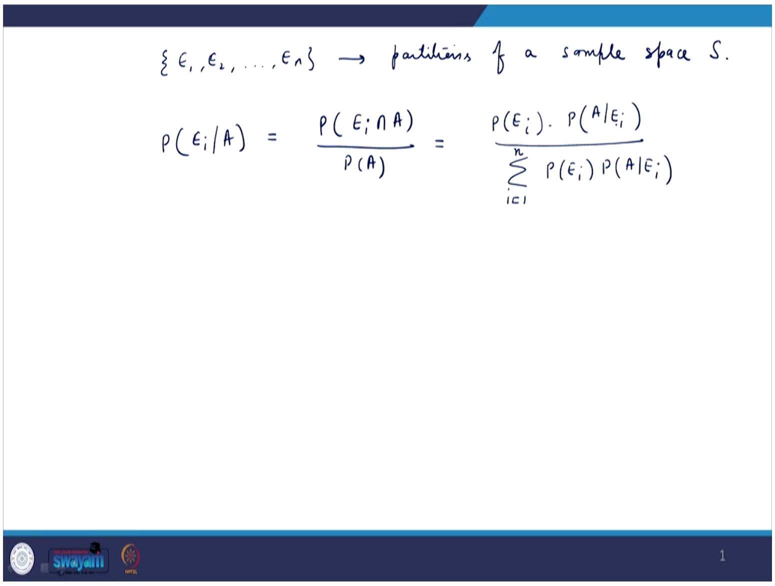
(Refer Slide Time: 00:59)



So, what is Bayes' Theorem basically, let E 1 E 2 and E n be n non-empty events that constitute a partition of a sample space S. So, we have already discussed the partition is basically what if you are saying that E 1 E 2 and up to E n are the partition; that means, the union of E 1 E 2 up to E n is S that is sample space and the intersection of any two different event is phi so; that means, a partition.

So, suppose these are the partition and A is any event of the non of non-zero probability then probability of any E i i may be 1, 2, 3 up to n given A is given by probability of E i only for that i into probability of A given E i upon sum of all this.

(Refer Slide Time: 01:58)

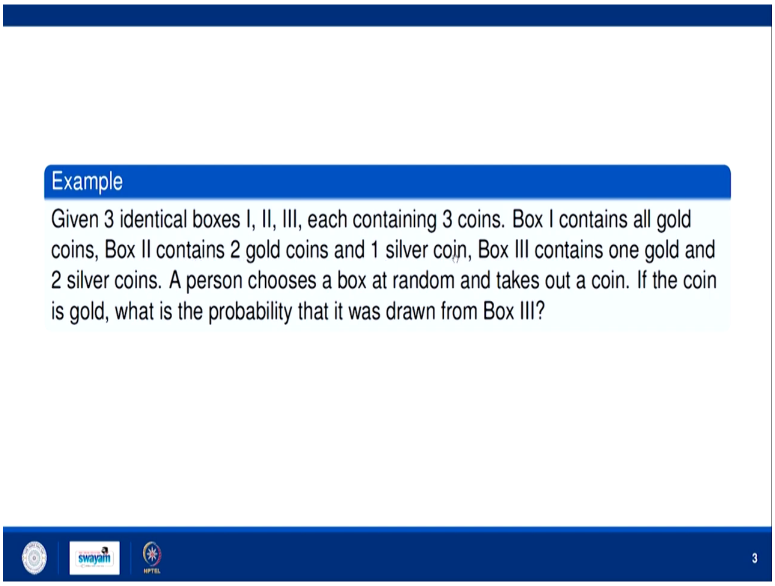


So, how you obtain this, given to assist E 1 E 2 up to E 1 these are partition of partitions of a sample space S. Now, we want to calculate probability of as we as see here probability of E i given A. So, as per a definition of conditional probability this is nothing, but probability of E i intersection with A upon probability of A ok.

Now what is probability of E i intersection of with A by the multiplication rule? This is nothing, but probability of E i probability of E i multiplied by probability of A given E i and this P A from the total probability will be given by summation i from 1 to n P of E i probability of E i multiplied by A given E i.

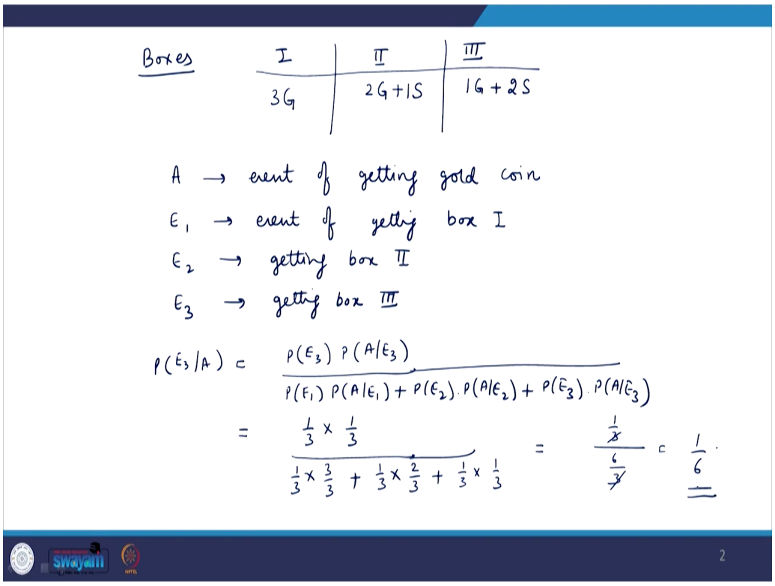
So, this we have already discussed in last class that how we can find out P A I mean from the total probability. So, that is given by this expression and that is simply by the multiplication rule we can say that it is this expression. So, in this way this is the required expression here which is basically the Bayes theorem.

(Refer Slide Time: 03:46)



Now, let us discuss one example based on this. So, suppose you are having 3 identical boxes 1, 2 and 3 each containing 3 coins. So, suppose you are having 3 boxes.

(Refer Slide Time: 04:01)



Box I, box II, box III each is having 3 coins box I contains all gold coins so; that means, 3 coins; that means, 3 gold all the 3 gold coins are here, box II contain 2 gold coins and 1 silver coin; that means, box II contain 2 gold and 1 silver ok. In box III it is 1 gold and 2 silver coins. So, in box III it is 1 gold and 2 silver coins. So, that is given to you. We are having 3 boxes and these are the different coins box I 3 gold, box II 2 gold and 1 silver, box III 1 gold and 2 silver.

Now, a person chooses a box at random and takes out a coin, if the coin is gold what is the probability that it is drawn from the box III. So, let us suppose A is the event of getting gold coin gold coin, suppose E 1 is the event of getting box I getting box I you can write, E 2 is getting box II, and E 3 is getting box III.

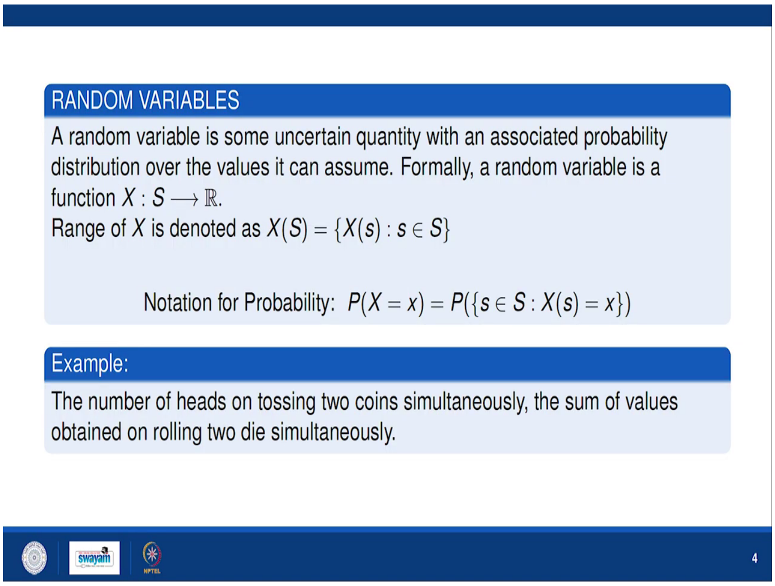
Now, you want to; you want to find out what is the probability that it is drawn from the box III; that means, you want to calculate E 3 given A and that will be nothing, but P of E 3. So, probability of E 3 multiply A given E i E given E 3 that is from the box III divided by a probability of E 1 probability of A given E 1 plus probability of E 2 multiplied by the probability of A given E 2 plus probability of E 3 into property of A given E 3.

So, now what is probability of E 3, E 3 means probability of getting box 3. So, there are 3 box you are having identical boxes I, II and III. So, all are equal probabilities. So, probability of choosing one box any one box is 1 by 3 definitely. So, it is 1 by 3 into probability of getting a gold coin when it is taken from the box III.

So, probability of getting a gold coin from the box III is basically 1 upon 3 because number of gold coins are 1 and total coins are 3. So, it is basically 1 upon 3 divided by. So, here again P of E 1 is 1 upon 3, now probability of A that is getting gold coin from box I, from box I it is all the coins are gold. So, it is 3 upon 3 plus now probability of E 2 is again 1 by 3 into probability of A when it is getting from the box II.

So, from box II it is II gold coins total coins are 3. So, it is 2 upon 3 plus probability of E 3 is 1 by 3 probability of A; that means, getting gold coin from probability of from box III that we have already computed it is 1 by 3. So, now, if you solve it so that will be nothing, but 1 upon 3 upon that is 3 5 6 that is 6 upon 3. So, that is 5 plus 1 6. So, it is what? It is if you cancel 1 by 3 from numerator and denominator. So, it is 1 by 3 upon 3 plus 2 5 plus 1 6 6 upon 3. So, that is basically 3 3 cancels out. So, it is 1 by 6. So, that is the required probability.

(Refer Slide Time: 08:24)

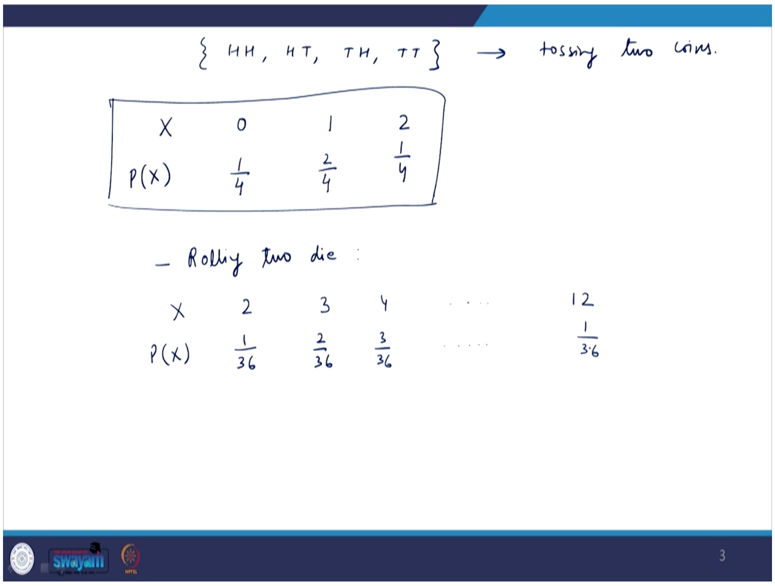


Now random variables. So, a random variable is some uncertain quantity with an associated probability distribution over the values it can assume. Formally, a random variable is a function X from S to R, basically we defined random variable as a function from S to R.

Now, the range of X is denoted by X S which is all X S such that S belongs to X that is basically range of X random variable X. Now if we in the in terms of probability of X when small x is given to you equal to small x is nothing, but probability of all those X belongs to s where X s is equal to small x ok.

So, that is how you can define out random variables in terms of probability now let us discuss this with an example. Suppose you are finding number of heads in tossing two coins ok, you are tossing a coin two times. So, you are finding number of heads. So, now, if you are tossing coins two coins.

(Refer Slide Time: 09:48)



So, sample space will be what either head, head tail, tail head or tail these are the only four combinations only four possibilities tossing two coins. Now, if you are saying X is a random variable which is the number of heads.

So, number of heads may be 0, number of heads may be it 1, number of heads may be 2, it cannot be 3 because we are tossing only two coins. Now if you are finding P of X X may be 0, X may be 1, X may be 2. So, probability if you are; if you are taking 0 heads so, how many cases for 0 heads. So, there is only one case or both are tail.

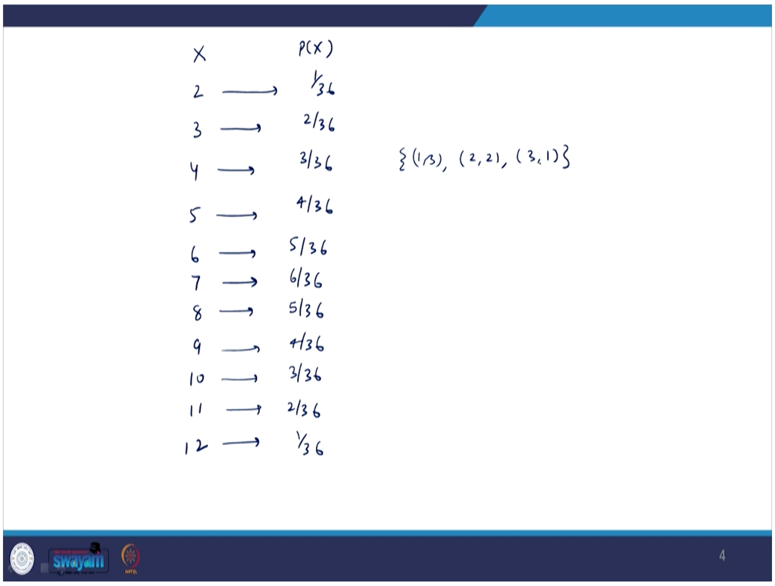
So, one case out of how many? Out of 4 so that will be 1 by 4 favourable upon total favourable is one case total are 4, 1 head for 1 head there are two cases. So, that is 2 upon 4 or 1 by 2 for 2 heads. So, it is one only one case. So, it is 1 by 4. So, that will be basically P X for this problem.

Now for a second example suppose the sum of values obtained. So, suppose you are rolling two dies simultaneously rolling two dies. So, now, the sum is either starts the minimum sum is 2 when it is 1 and 1 and the maximum sum is 12. So, X is a random variable which can take values from 2 to 12, 2, 3, 4 up to 12.

Now, if we want to compute P of X. So, for 2 it is only one case one when both the die will turn out to be 1 and total probability is I mean total cases are 36 6 on one die, 6 on other die, 6 into 6 is 36. So, probability will be I mean 1 upon 36, now for 3 it is 1 2 or 2 1 so; that means, 2 upon 36.

Similarly, 3 upon 36 and the last case for 12 it will be for 12 it will be 1 upon 36. So, it will first increase and then decrease see if you want to compute the complete table of this.

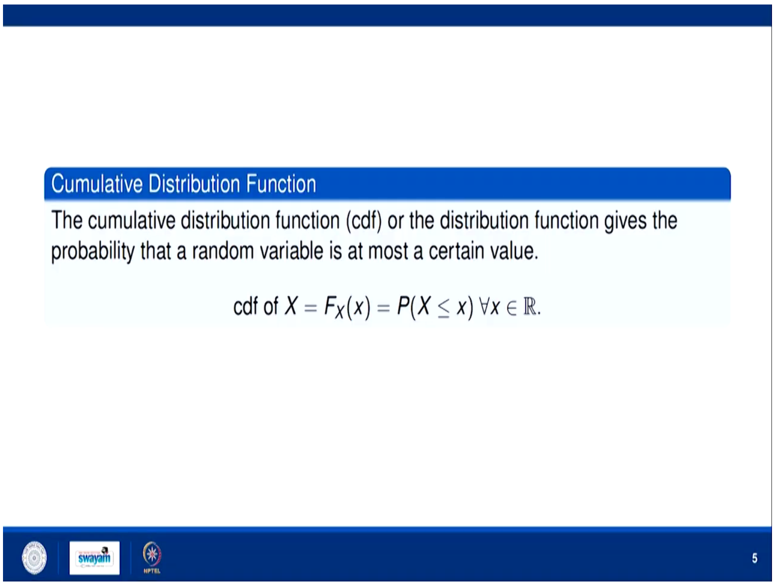
(Refer Slide Time: 12:31)



So, it will be like this you will take X as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and probability of X will be for 2 it is 1 by 36, for 3 it is 2 by 36, for 4 it is 3 by 36, for 5 it is 4 by 36, for 6 it is 6 by 36 and then start decreasing ok.

For 7 it is 6 by 36, for 8 it is now 5 by 36. So, now, it is 1 by 36 here, it is 2 by 36 here, it is 3 by 36 here, it is 4 by 36 here that the cases you can easily obtain, see if you want to; if you want to draw the cases for this 4. So, for 4 it is 1 3, 2 2 then it is 3 1. So, you can similarly draw all the possible cases. So, that will be the basically you can get the probability for different X for this case.

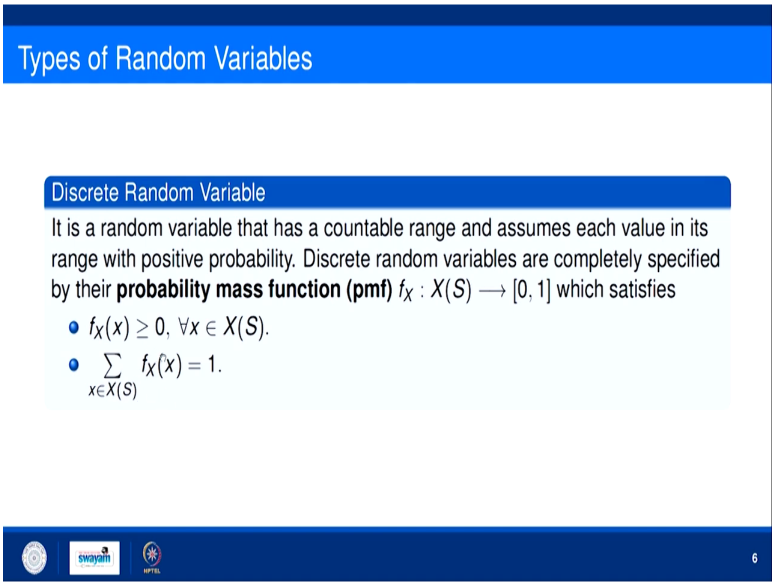
(Refer Slide Time: 13:43)



Now what is cumulative distribution function cdf, the cumulative distribution function cdf or the distribution function gives the probability that a random variable is at most a certain value, see we are having a small x.

So, the probability that the random variable is always less than equal to small x is simply the is simply a cdf of that X for every X in R. So, that is how we can define cdf. Now we will discuss some examples to for the better understanding of this, first we will cover all the definitions. So, now, what is discrete random variable.

(Refer Slide Time: 14:22)

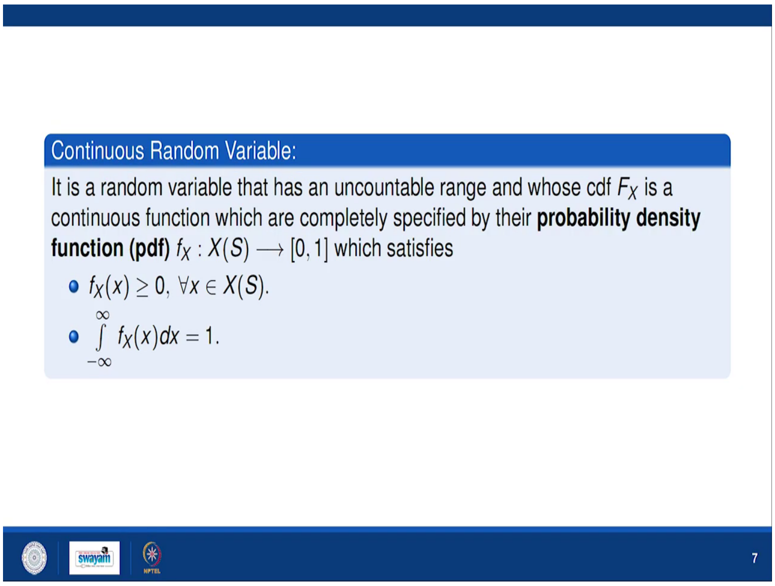


So, it is a random variable that has countable range and assumes each value in its range with positive probability. So, it is discrete; that means, has a countable range and it assumes in its range with positive probability as mentioned. Discrete random variables are completely specified by their probability mass function and what is a probability mass function, it is a function from X S to 0 1, which satisfy the following two properties.

The two properties are, number 1 the probability mass function must be greater than equal to 0 for every X and number 2 the sum of all the probability mass functions I mean the sum of all this functions must be 1. So, these are the two basic properties and then we can say that it will be a probability mass function for discrete random variables.

Now, how can you defined it for continuous case, I mean continuous random variables.

(Refer Slide Time: 15:34)

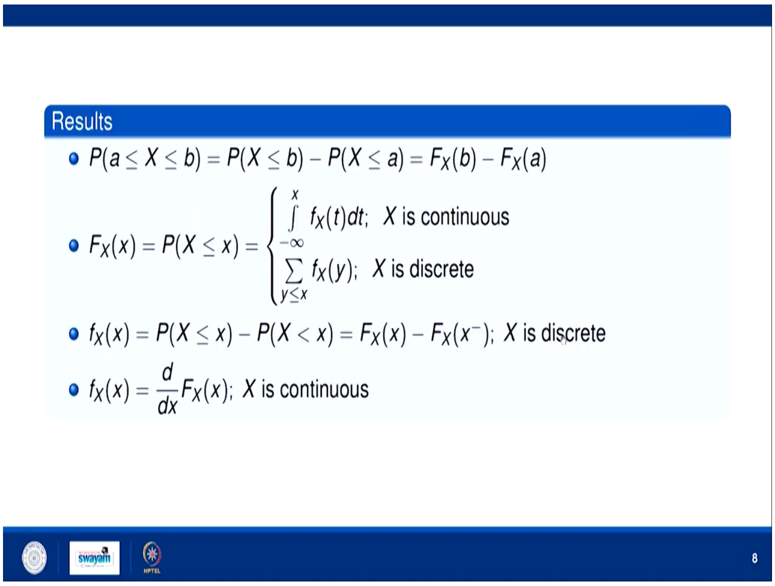


So, for continuous random variable we can define it like this a continuous random variable is a random variable that has an uncountable range and whose cdf that is F x capital F x is a continuous function which are completely specified by their probability density function pdf here we call as pdf for discrete case it is probability mass function, for continuous case it is probability density function and which again satisfy two properties.

The two properties are, number 1 it must be non negative for every X and number 2 because it is continuous. So, here comes out to be integration. So, the integral from minus infinity to plus infinity should be 1. So, in other words we can say this as area if you take the probability density function if you draw the area the entire area under the curve must be 1.

So, the main difference for probability density function and probability mass function is; mass function is for discrete case it is for continuous case, the second one is course to the first property is same that f x must be greater than or equal to 0 for every x here in probability mass function you are having sum of all the; sum of all the functions and here we are having integral from minus infinity to plus infinity that is the only difference.

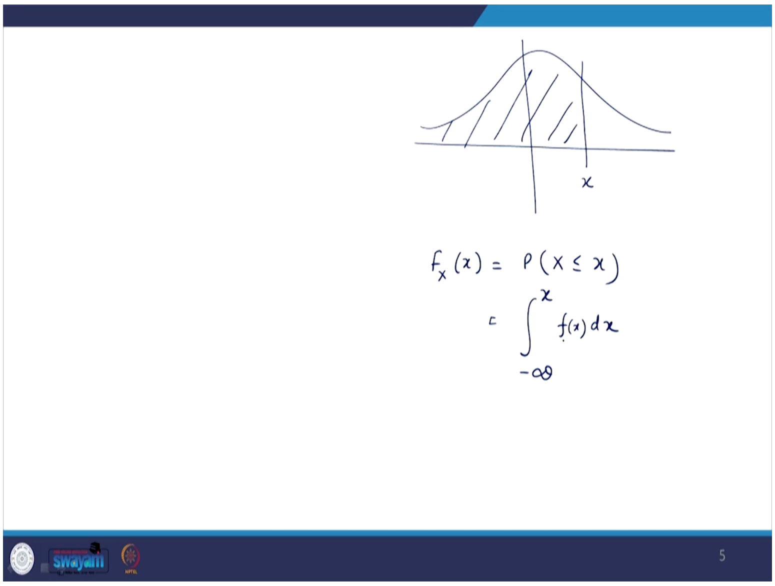
(Refer Slide Time: 17:02)



Now, we are having certain properties. The first property is if you want to find out between a and b; that means, you take from b up to b and subtract up to a that will give from a to b and that is nothing, but cdf at X equal to b and cdf at X equal to a, this capital F is cumulative distribution function ok. So, it is cdf at X equal to b and cdf at X equal to a this is the first property.

The next property is suppose you want to compute cdf at X equal to x at capital X equal to small x. So, that is nothing, but for the continuous case it will be nothing, but integral from minus infinity to x because you want to find out up to x up to small x. So, we want to find out up to a small x; that means, you have to integrate from minus infinity up to x. So, that area from minus infinity to up to x that will give cdf when capital X is small x that you can understand from here also.

(Refer Slide Time: 18:12)



Suppose you want to compute. So, suppose it is some continuous case and for this x for this small x you want to compute cdf. So, F x F at x equal to x; that means, probability of X less than equal to x means here this area. So, you want to compute this area.

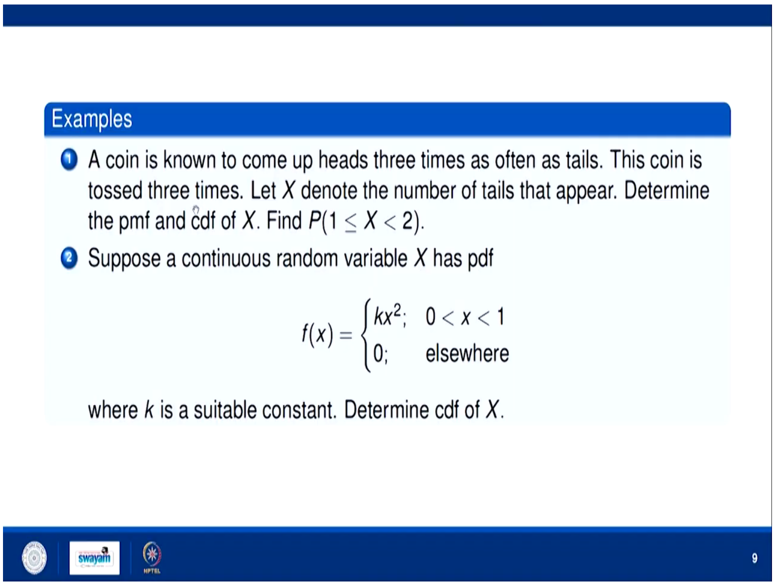
So, this area means what this area is basically integral minus infinity to 2 x f x dx f x you can write x here or you can write f x x here ok. So, that is f x t. So, that is not a that is notational problem only. So, basically that function will come here and that is minus infinity to x.

Now, if X is discrete so; that means, you have take the sum, sum up to; sum up to x sum up to x you have to find. So, that will give basically cdf. So, these are the basically relation between cumulative distribution function or probability mass function or probability density function ok.

Now if you want to compute this probability density or mass function for any small x then that is that will be given by you first find probability when X is less than equal to x that is up to x; up to x and when you subtract with X still less than x so that will give probability when X equal to x.

So, in notationally we can also write it like this when X is a discrete and for continuous case it is simply derivative because this is simply this is simply integral. So, basically derivative will come in the case of continuous if distribution is a continuous distribution.

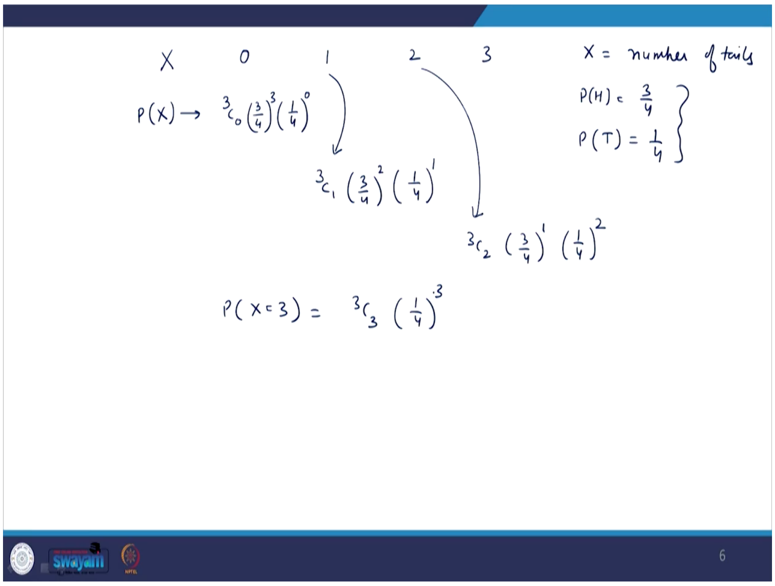
(Refer Slide Time: 20:09)



So, now let us come to few examples, now first problem is a coin is known to come up heads three times as often as tails. This coin is tossed 3 times, let X denote the number of tails that appear.

Determine the probability mass function and cumulative distribution function of f of x and also find probability when X is greater than equal to 1 and less than 2. So, let us discuss this example. So, what is what X denote X is a number of tails that appear three times we are throwing ok.

(Refer Slide Time: 20:44)



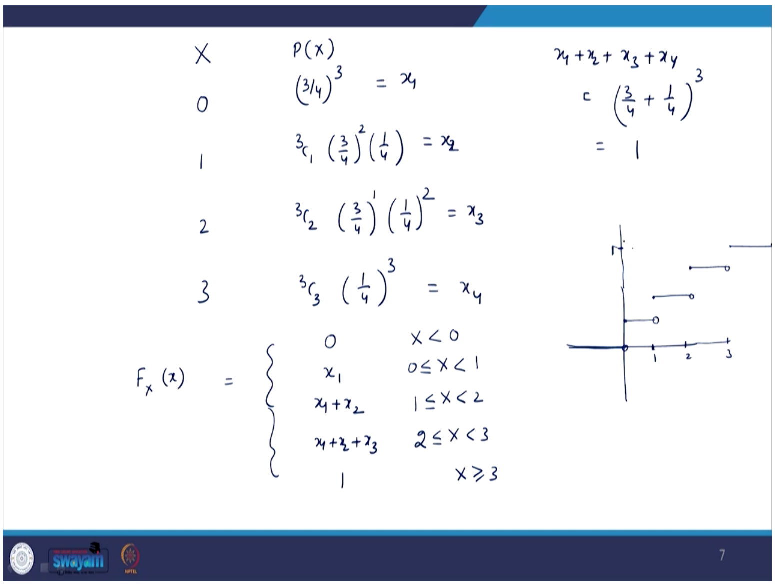
So, X may be either 0 1 2 or 3 these are the X is number of tails number of tails ok. Now, what will be probability of X? First it is not the equal probability of head and tail, here the probability of head come out to be head is 3 times as often as tails. So, basically probability of getting head is 3 by 4 and a probability of getting tail is 1 by 4 ok.

Now, if there is if we are finding 0 tail. So, out of 3 coins there is no tail and probability of getting head is 3 by 4 and probability of getting tail is 1 by 4. So, that is very clear because you see that it is 0, you can see here because there is no tail; that means, all 3 are head so; that means, 3 by 4 into 3 by 4 into 3 by 4.

Now, if you are taking so it this is basically this is P X for this now for 1 when X equal to 1, when X equal to 1 out of 3 cases any one case it may be take out to be 1, there are we are tossing 3 coins we want 1 head it may be in the first place, second place or third place. So, that is 3 C 1 total cases and 3 by 4 getting head is simply getting head is simply 2 and 1 by 4 is 1 so, 3 by 4 into 3 by 4.

Now for 2 so, for 2 it will be 3 C 2 3 by 4 raise to power 1 and 1 by 4 whole square. So, that will come 2 tails and for X equal to 3 for P X equal to 3 that will be sorry that will be 3 C 3 and 1 by 4 raise to power 3 ok. So, that will be basically distribution function of this I mean that will be simply what we call that will be simply pmf ok.

(Refer Slide Time: 23:19)



So, what we have; what we have seen when X equal to 0 1 2 or 3 let us write it neatly. So, P X will be here it is; here it is 3 by 4 whole raise to power 3, here it is 3 C 1 3 by 4 square into 1 by 4, here it is 3 C 2 3 by 4 3 by 4 it is 1 1 by 4 whole square, it is 3 C 3 1 by 4 whole raise to power 3.

Now, suppose this value is x 1 suppose this value is x 1, x 2, suppose this value is x 3 suppose value is x 4. So, first of all probabilities are all probabilities are greater than equal to 0 of course, and if you sum up all probabilities. So, this will come out to be 1. So, you can check. Basically if you take x 1 plus x 2 plus x 3 plus x 4 so, that will be nothing, but 3 by 4 plus 1 by 4 whole raise to power; whole raise to power 3 and that is 1 this is simply by binomial theorem.

So, hence it is basically a probability mass function now if you want to compute cdf of this. So, what will be cdf, cumulative distribution function? So, we want to calculate probability. So, when X is suppose you want to compute when X is less than 0. So, when X is less than 0. So, we want to compute cdf, it is cdf basically. So, cdf we denote by F x x ok.

So, if you if X is less than 0. So, it is 0 because X can take the value 0 1 2 or 3 if suppose X is negative. So, there is no probability that is 0. Now if X is less than 1 or greater than equal to 0 ok, if X is between 0 and 1, then in that case suppose in that case it simply take the value when X equal to 0 and that is simply x 1. So, that will be simply x 1 because if you take any other value between 0 and 1 there is no other no probability except 0.

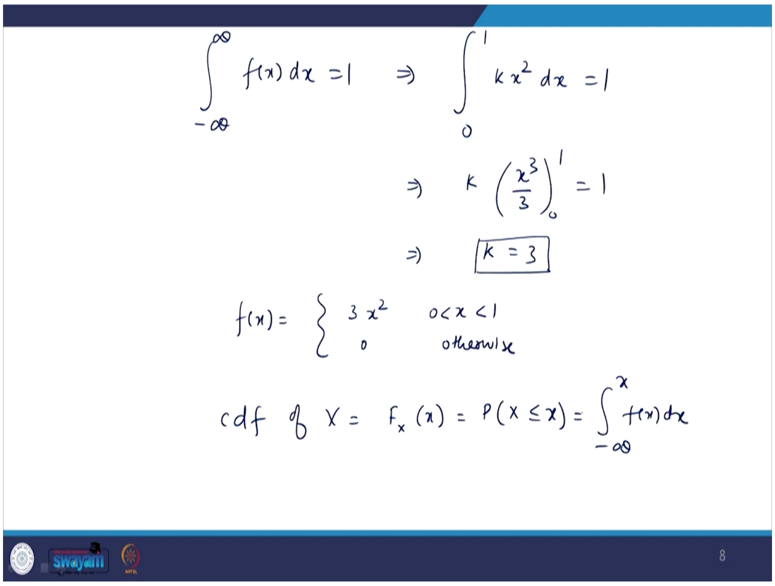
Now, when X is between 2 and 1 it will be simply x 1 plus x 2. Now if you take X less than 3 greater than equal to 2. So, that will be nothing, but x 1 plus x 2 plus x 3 and when x is greater than equal to 3 that is 1. So, this will be the cdf of this X. So, basically if you draw it so, it will be some steps step type function when x is 0 when x is less than 0 it will be 0 excluding this.

When x is between 0 and 1 so, it is some x 1 which is somewhat here suppose this is inclusive when X is between 1 and 2 1 and 2 it will be something like 1 and 2. So, it is hollow here comes here and when between 2 and 3 2 and 3 it will go here 2 and 3 it is hollow here hollow here and after 3 it will be 1. So, this is basically 1 ok. So, this is this type of function.

Now, next we want to compute X probability when X is between 1 and 2. So, that is given here when 1 and 2 is simply x 1 plus x 2. So, you simply add these two terms and you will get the answer of the probability when X is between 1 and 2. So, this is for the discrete case.

Now, let us discuss 1 for continuous case, now suppose a continuous random variable X has a pdf given by this expression. So, first you find first you want to find out for which k it will be a pdf. So, for pdf first of all for all x this must be greater than equal to 0. So, it is equal to 0 and it must be greater than equal to 0; that means, k must be greater than equal to 0 that is certain. Next is total area must be 1, now total area means now here elsewhere it is 0; that means, integral from 0 to 1 must be 1 so, that will give the value of k.

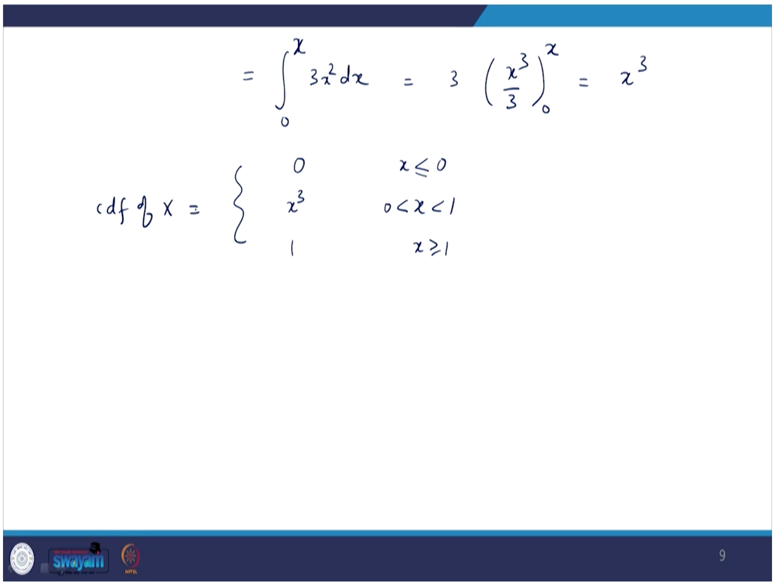
(Refer Slide Time: 28:02)



So, integral basic definition is integral this should be 1, now this implies integral from 0 to 1 k x square dx should be 1 and this implies k into x cube upon 3 from 0 to 1 should be 1. So, it implies k is equal to 3 so; that means, k the value of k is 3.

Now, if you want to find out the cdf of this x. So, what is f x now, f x now is 3 it is 3 x square when x is between 0 and 1 and 0 otherwise. So, how can you find cdf of x. So, cdf of x will be simply F x x. So, that will be probability when x is less than equal to x and that is given by minus infinity to x fx dx. So, here what is F x, F x is now 3 x square and that is defined between 0 and 1. So, between 0 and 1 so, we can say that from minus infinity to 0, it will be 0, because otherwise it is 0.

(Refer Slide Time: 29:29)



So, it will be defined only for 0 to x. So, we can say that it will be further equal to 0 to 1 it is 3 x square dx and that will be 3 x cube by 3 0 to 1. So, that will be 0 to x say that will be 0 to x sorry because it is it is 0 to x.

So, 0 to x it will be; it will be x cube. So, what will be the cdf now? So, cdf will be basically when x is less than 0 it will be 0 less than equal to 0 it will be 0 and between 0 and 1 it will be x cube and when x is greater than equal to 1 it is 1. So, this is how we can define the cdf of x ok.

So, we have seen so, in this lecture we have seen that what is Bayes' theorem and how we can find out what is probability mass function, density function and how we can find cumulative distribution function, if pdf is known, if probability mass function is known or vice or vice versa. So, in the next lecture we will see how we can find expectation variance, covariance etcetera.

Thank you.