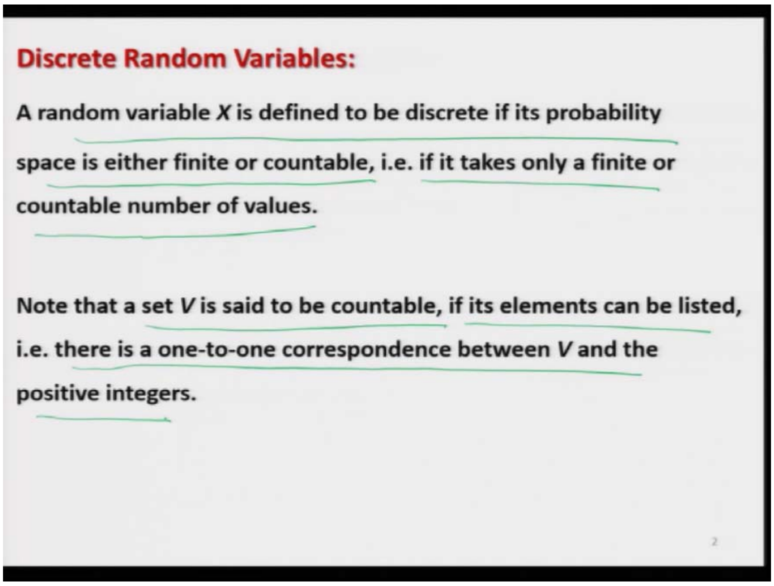
Essentials of Data Science with R Software - 1
Professor Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Lecture No. 24
Discrete Random Variables, Probability Mass Function and Cumulative Distribution Function

Hello, friends. Welcome to the course Essentials of Data Science with R Software 1, in which we are trying to understand the concepts of probability theory and statistical inference. So, you can recall that in the last lecture, we considered the continuous random variable and we defined the probability density function, probability distribution function or cumulative distribution function. Now, the same thing I am going to do for a discrete random variable in this lecture.

I will try to define the probability mass function, which has a similar role what the probability density function had in the case of a continuous random variable. And then I will try to show you that how the distribution function under a discrete random variables is different from the distribution function under the continuous random variable. So, let us try to begin this lecture and and I will try to take couple of example to explain you in a much simpler way. Let us begin.

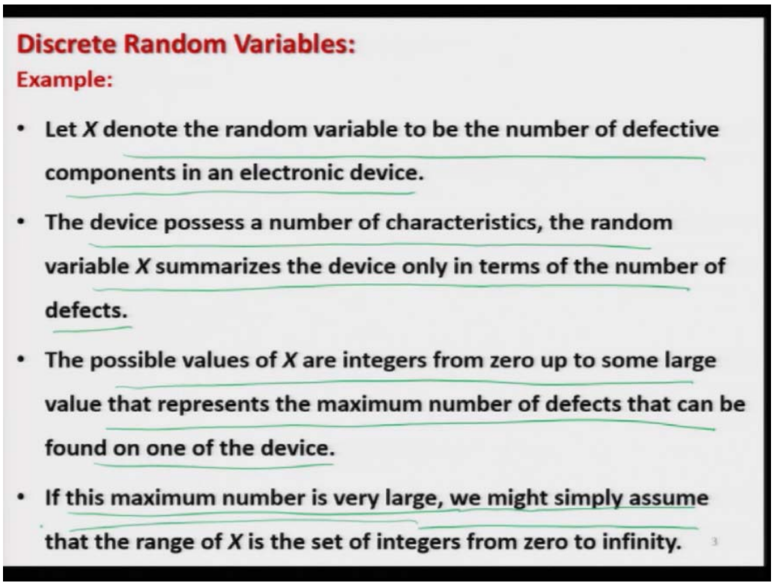
(Refer Slide Time: 01:23)



So, now, first I will try to take the same definition that I have given you earlier, about the random variable to be called as a discrete random variable. So, a random variable capital X is defined to be discrete, if it is probability space is either finite or countable. That is, if it takes only a finite or countable number of values. Now, what is the meaning of this countable?

So, a set capital V is said to be countable, if its elements can be listed, that is, there is a one-to-one correspondence between V and the positive integers. That is what I was trying to tell you by giving you different types of example, that when we are trying to roll dice, we are trying to take the values of random variable to be 1, 2, 3, 4, 5, 6 or when we are trying to flip a coin, we are trying to give the random variable the value 0, 1 and so on.

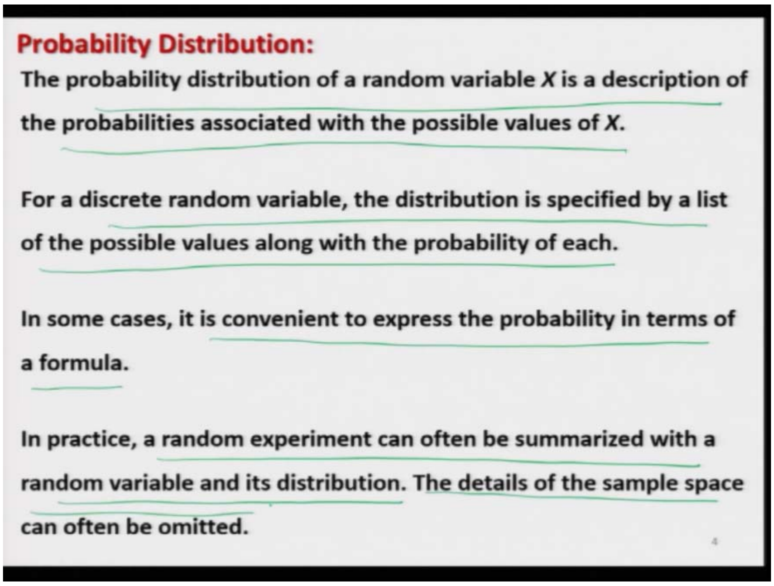
(Refer Slide Time: 02:12)



So, there can be various examples of this type. So, suppose X indicate the random variable to be the number of defective components in an electronic device. You know that this number can be either there are two defects or three defects that cannot be 2.3 defects. The device possesses a number of characteristics, the random variable X summarises the devices only in terms of the number of defects. There, there are two defect, there are three defect, there are four defects and so on.

So, the possible values of X are integer from zero up to some large value that represent the maximum number of defects that can be found on one of the device. And suppose, if this maximum number is very large, then you may simply assume that the range of X is the set of integers from 0 to infinity. That is all, as simple as that.

(Refer Slide Time: 03:05)

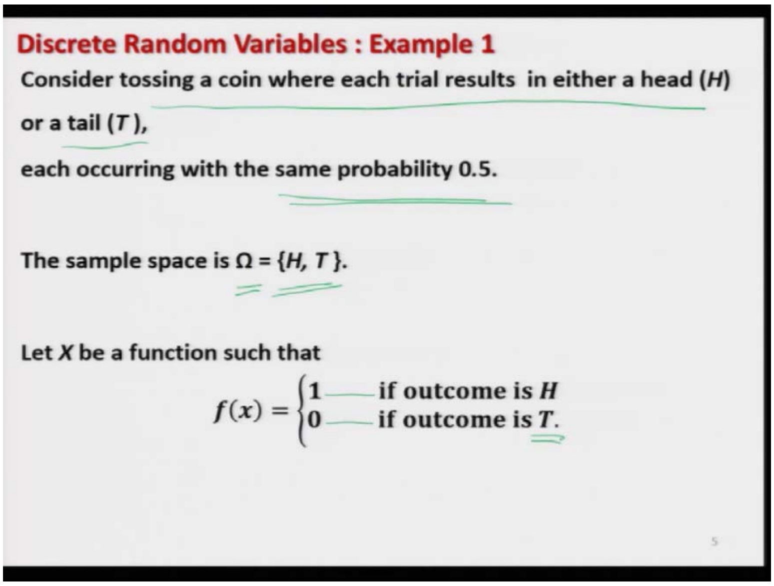


Now, when this random variable is going to take such a values, then what we now have to do? We have to describe the probability. So, we had discussed towards the end of the last lecture that this is called as probability distribution, because it is trying to distribute the probabilities over different values of all the ranges of the random variable. So, the probability distribution of a random variable X is a description of the probabilities associated with the possible values or capital X.

So, for a discrete random variable, the distribution is specified by a list of the possible values along with the probability of each. I agree that we always consider the example where we are trying to give you the formula of the probability or sometime, we also discuss that a given set of data you can count the total number of favourable cases, to compute the probability.

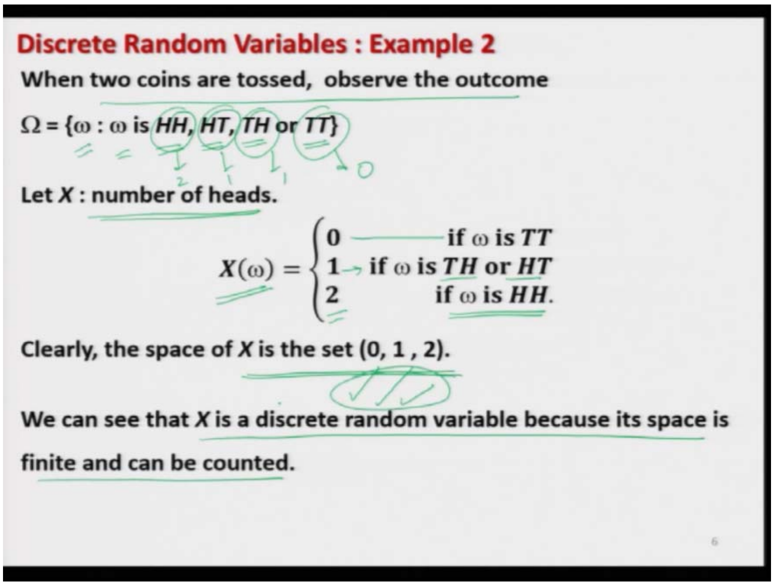
But in most of the cases, it is convenient to express the probability in terms of a formula. That we will try to see very soon, that we will try to give you the formula so called quote unquote "formula". That if you want to determine the probability of a particular type of event, then you can simply substitute the values and can find out the value of the probability. So, in practice, a random experiment can often be summarised with a random variable and its distribution. The details of the sample space can often be omitted in practice.

(Refer Slide Time: 04:36)



So, for example, let me try to take an example to illustrate what I am trying to say. Consider an experiment where we toss a coin, whether trials are resulting either in head or tail. And we know now that the probability of getting head or tail is always 0.5. So, now the sample space of this experiment is Ω which is head or tail, H or T. Now, we are trying to define here function fx such that fx takes value 1 if the outcome is H, that is head and it takes value 0, in case if the outcome is T, that is tail.

(Refer Slide Time: 05:12)

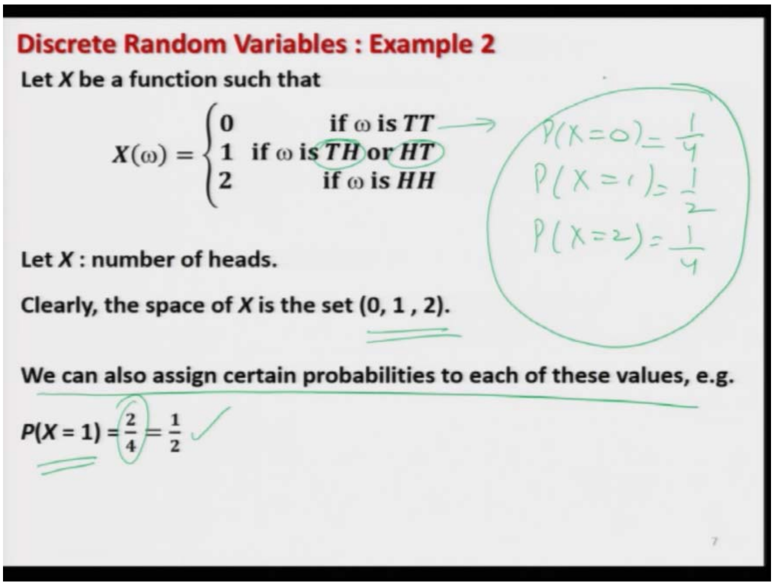


Now, similarly, I can take one more example, that suppose two coins are tossed and we try to observe the outcome. The outcome will be something like two heads; first head, second tail; first tail, second head; and both are tail. These are the 4 options which are available. So, this sample space Ω will be consisting of the points Ω, such that ω is HH, HT, TH or TT. Now, the same thing can be associated with a random variable. Now, suppose I define the random variable as the number of heads. So, the number of heads in the first value HH, is 2; in the second value HT, is 1; in the third value TH, is 1; and in the fourth value, TT is 0.

So, now I can define here a function like X (ω), which takes value 0, 1 and 2. It takes value 0, if ω is TT means both are tails, that means there is no head. And it takes value 1, if the ω is either TH or HT, that means at least one of the values is the head. And it takes value 2, if ω is HH, that means both the tosses results in heads. So, now, what is the space of X. The space of X now becomes simply here three values 0, 1 and 2.

And by looking at that any of these values 0, 1 or 2, means we can define that or we can understand that what has really occurred. So, in this case, we can see that X is a discrete random variable, because its space is finite and can be counted 1, 2 and 3. There are three values in this space 0, 1, 2. That is all. This is what we are trying to mean in a very simple lucid language.

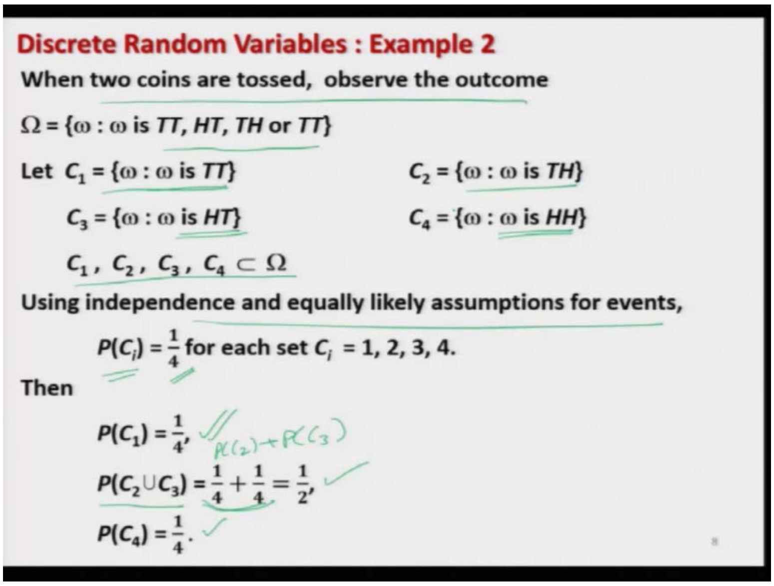
(Refer Slide Time: 06:50)



Now, in the same case, we can also now assign certain probabilities to each of these values. For example, you can see here that in this case, the outcome is going to be one of the 4 outcomes. So, here I can define here probability of X equal to 0, probability of X equal to 1, probability of X equal to 2. So, what it is going to be? Suppose if I want to define here the probability of X equal to 1. So, there are two events out of four that is TH and HT, that can occur. So, this will become here 2/4, which is equal to here 1/2. So, I can write down here 1/2.

And similarly, for for X equal to 0, the probability will be 1/4 and probability for X equal to 2 is 1/4. So, now, I have obtained here all these probabilities. Now, my objective is this I want to make you understand that why this is called as probability distribution. So, now you can see here that this event has taken the three possible values of probabilities, depending on the values of random variable.

(Refer Slide Time: 07:56)

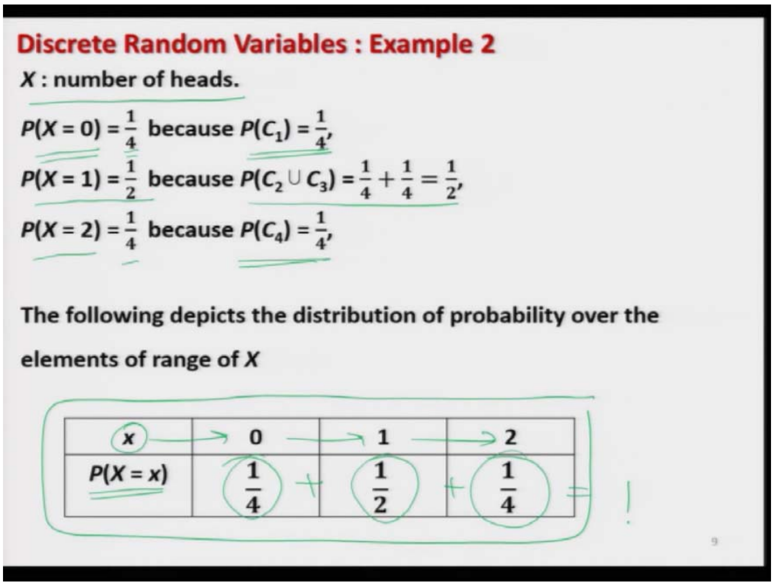


Now I tried to make it more formal. Well, I am doing it so that you can easily understand it. And yeah, means at this moment, I am trying to explain you in more detail, after some time, once you become conversant with these things, I will be quick and brief. So, suppose, now, we consider the same experiment that we have tossed two coins and these outcomes are obtained TT, HT, TH or TT. Now, I try to define here 4 possible events, C_1, C_2, C_3, C_4. C_1 is the simple event that ω is TT, C_2 is ω is TH, C_3 is ω HT and C_4 is ω is HH. All these C_1, C_2, C_3, C_4 they belong to ω. No issues.

Now, using the independence and equally likely assumption of this events, we can see that the probability of C_i, that probability of C_1, probability of C_2, probability of C_3, and probability of C_4, they are 1/4 for each of the event. So, now you can compute here, what is the probability of C_1, this is here, 1/4. Now, what is about here probability of $C_2 \cup C_3$, because they are disjoint events, so this will be $P(C_2) + P(C_3)$.

Now you can see what you had learned earlier that is now being used. So, this will become her 1/4 + 1/2, this 1/2. And P(C_4) here is 1/4. And you can see here that in this earlier slide, this is exactly what you had obtained.

(Refer Slide Time: 09:26)

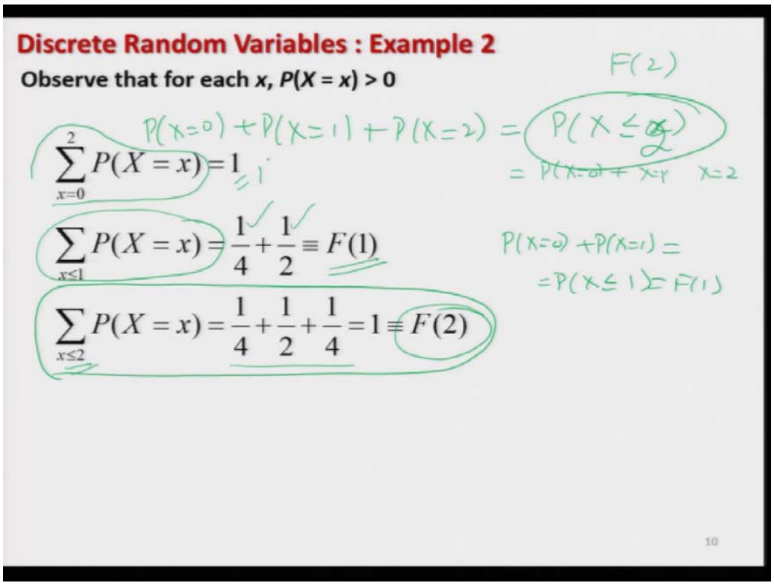


And now from here, I try to rewrite the same thing in a different way. That I have here a random variable X, which is the number of heads in the trial of two coins. And now, probability that X is equal to 0 takes value 1/4, because P(C_1) is 1/4, probability that X takes value 1/2, because P(C2 union C3 is 1/2, probability that X takes value 2 is 1/4, because P(C4 is 1/4.

So, now I can describe all these probabilities and their values of random variables in the form of a table. And I can write down here, a small x takes value 0, 1 or 2. And when it takes value 0, then P(X equal to small x is 1/4, 1/2 and 1/4 respectively. So, you can see here, do not you think that this small table is going to give you the distribution of the probabilities with respect to the values of the random variable that it can take. So here, the random variable can take only 3 possible values 0, 1, and 2.

So, and this is the entire distribution of the total probability, because the total probability is always going to be 1. And this total probability has been distributed into three segments, 1/4, 1/2 and 1/4. And if you try to add them together, this will always be equal to 1. So, this table is indicating the distribution of the probabilities with respect to the respective values of random variables. So, that is why this is called as probability distribution.

(Refer Slide Time: 11:01)

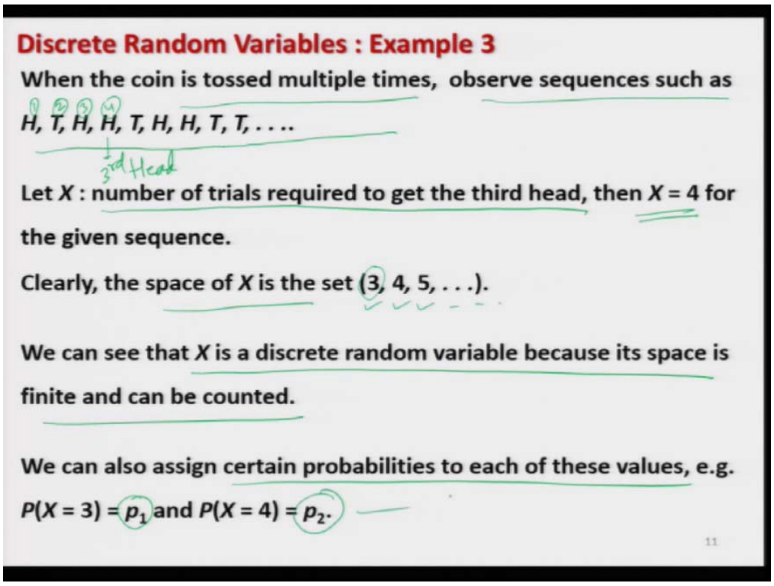


So, I hope I have made it quite clear. And now in case if you want to see that, how are you going to find out the cumulative distribution function. So, in this case, if you try to see, there are 3 possible values, which the random variable is trying to take 0, 1 and 2. So now in case if you try to take the probability that X probability that X takes value 0, plus probability that X takes value 1 plus probability that X takes value 2. So, do not you think that this is going to be the value of probability that X less than or equal to x, and where your this x is equal to here 2.

And since this is a discrete, so, these probabilities are going to be P(X) equal to 0 plus for X equal to 1 plus X equal to 2. So, this is here your what? This is nothing but your here F2. So, you can see here, if you try to write down here, like this, $\sum_{x\le 2} P(X=x)$. This is something like here 1 and this is the value of cumulative distribution function at X equal to 2. And similarly, if you try to write down here the value of a probability that X equal to 0 plus probability of X equal to 1, which is nothing, but the probability that X less than or equal to 1 and which is nothing, but your here F1.

So, you can see here that this, the same probability X equal to small x and the summation is going over x is less than equal to 1. This will be here 1/4 + 1/2, which is equal to here, the F(1). And obviously, in case if you try to sum all the probabilities, probability X equal to 0 plus probability at X equal to 1 and probability at X equal to 2 that is always going to be the 1, because P(ω is always equal to 1 and this is indicating the Ω.

(Refer Slide Time: 12:59)

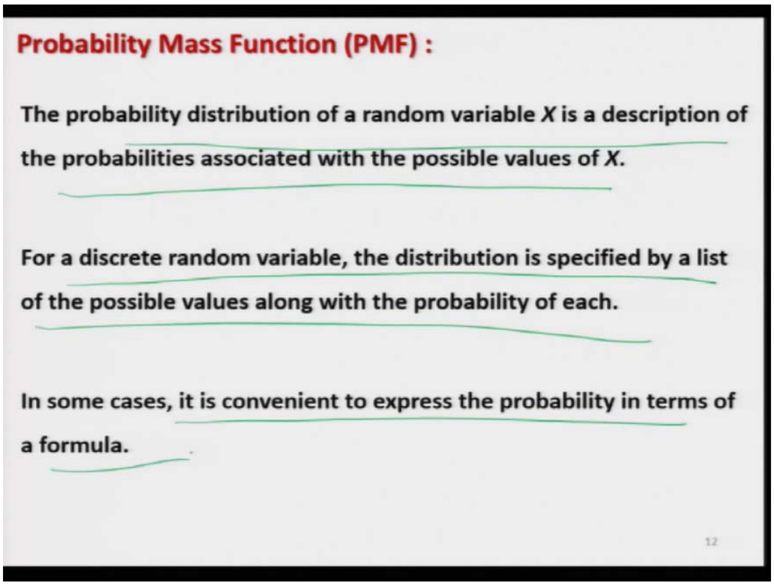


And now, I try to take one more example, suppose the coin is tossed multiple times, and we observe a sequence of heads and tails like this H, T, H, H, T, H, H, H, T, T … and so on. Now, suppose if I say, X is the number of trials required to get the third head, so, in this case, what will happen here? This is the head number 1, now there is here in the second trial, there is no head. Now, we have a third trial, where we have a head; and we have a fourth trial where we have the third head. So, now, X is going to take the value 4, for this given sequence.

So. now, in case if you try to look at this sequence, means the, you can create different types of sequences or there can be different types of sequence. But definitely, if you are trying to get the third head, the minimum number of trials that you need is three. So, the sample space of X will become here 3, 4, 5, and so on. You cannot say here that this three can be 0, 1, 2. Because, if you are not tossing the coin for at least for 3 times, you will not get 3 heads. So, in this case, you can see that X is a discrete random variable, because it is space is finite and can be counted.

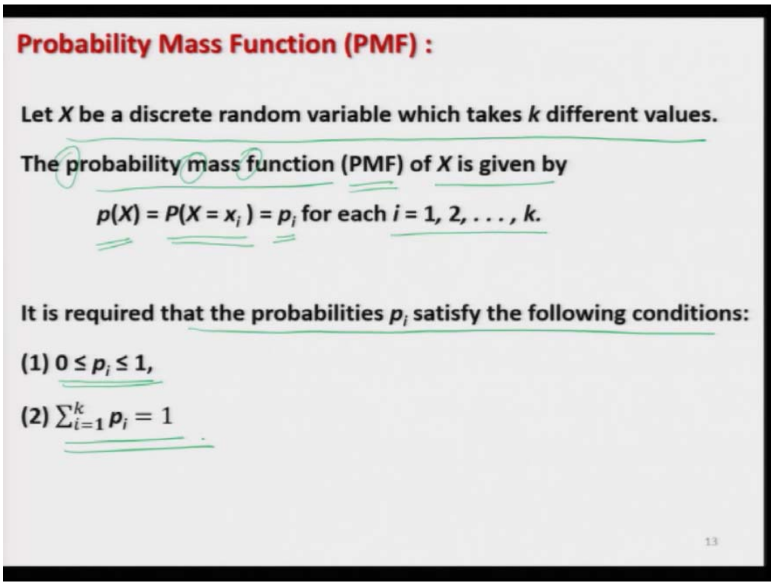
And now in these cases, we can also assign the probabilities to each of this event. So, we can always assign certain probabilities to each of these values, probability that for example, X equal to 3 can be small p_1, and probability that X equal to 4 can be small p_2 and so on. So, p_1 and p_2 are going to be some numbers, that have to be determined.

(Refer Slide Time: 14:52)



So, now based on these examples, we can now have a definition of probability mass function. The probability distribution of a random variable X is the description of the probabilities associated with the possible values of X. That is what we have understood. So, now for a discrete random variable the distribution is specified by a list of the possible values, along with the probability of each value.

(Refer Slide Time: 15:25)

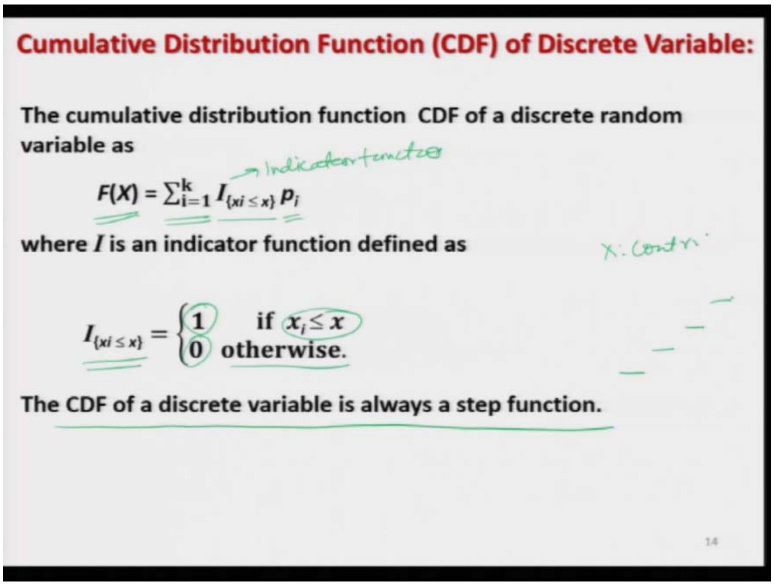


So, in some cases, it is convenient to express the probability in terms of formula, and based on that now, we are trying to give a formula for the probability, in the case of discrete random variable, which is going to be called as probability mass function. And before I go further, let me try to recall that in the last lecture, we had done the probability density function. So, the probability density function is defined for a continuous random variable. And probability mass function, that is defined for a discrete random variable.

So, now, I can formally give the definition of probability mass function. Let capital X be a discrete random variable which takes k different values, then the probability mass function which is also indicated as a PMF, P is coming from probability, M is coming from mass, and F is coming from function. The PMF of X is given by $p(X) = P(X = x_i) = p_i$ for each $i = 1, 2, \ldots, k$

And for this function PX to become the probability mass function, it is required that the probabilities pi satisfy the following condition. That each value of the p_i will be lying between 0 and 1. 0 and 1 are inclusive. So, p_i is greater than or equal to 0 and less than equal to 1. And sum of all the p_i's over i go from 1 to k. This will be equal to 1. That means p_1 plus p_2 plus p_k that is always going to be equal to 1 because that is your probability of Ω which is equal to 1. So, so, this is the definition of probability mass function.

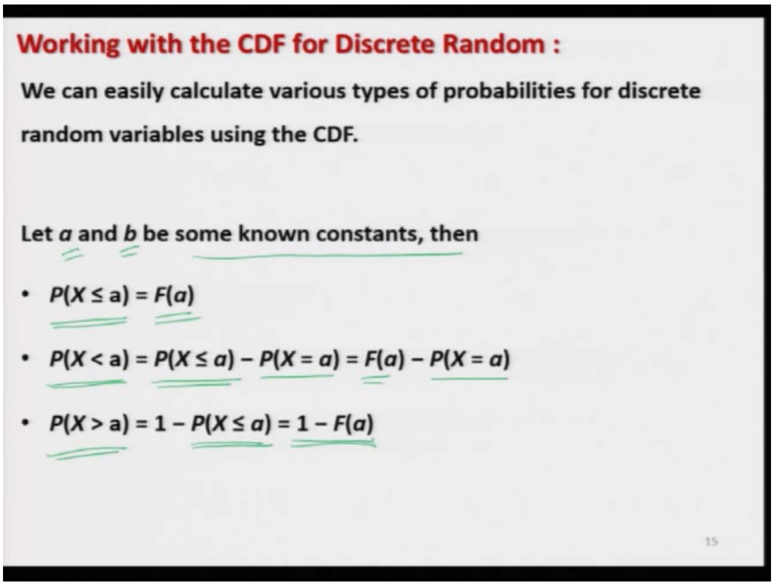
(Refer Slide Time: 17:05)



Now, similarly, based on this definition, we can define the cumulative distribution function for a discrete random variable. So, suppose X is a discrete random variable, then this CDF can be defined here is as follows $F(X) = \sum_{i=1}^{k} I_{\{xi \le x\}} p_i$. That the indicator function takes $I_{\{xi \le x\}} = \begin{cases} 1 & \text{if } x_i \le x \\ 0 & \text{otherwise.} \end{cases}$

One thing you have to keep in mind, means earlier if you remember, in the case, when X was a continuous random variable, I had very easily described the continuous curve for indicating the PDF. But in case of discrete the CDF of a discrete random variable is always a step function. Step function means, you know the steps like this, which we would want to move on a stair, and the stair has the steps.

(Refer Slide Time: 18:16)

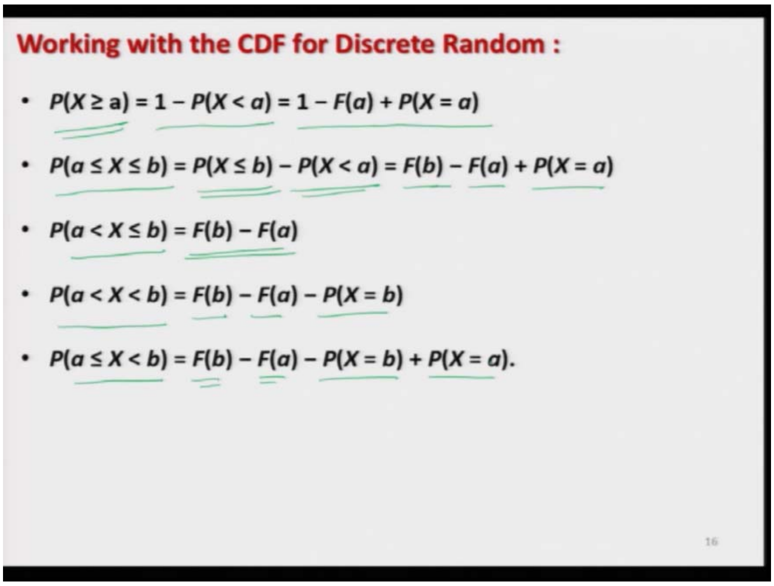


And using this CDF, we can also compute different types of probabilities just like as we have done in the case of continuous random variable. For example, if I say, let small a and small b be some constant, which are known to us. Then probability that X is less than equal to small a is your the value of CDF at X equal to a.

And similarly, if you want to find out the $P(X \leq a) = F(a)$, then this is $P(X < a) = P(X \leq a) - P(X = a) = F(a) - P(X = a)$. And similarly, if you want to find out $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$

So, this is how we can compute the different probabilities, if a CDF is given to us.

(Refer Slide Time: 19:15)



Similarly, in case if you want to find out here $P(X \geq a) = 1 - P(X < a) = 1 - F(a) + P(X = a)$. Similarly, $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a) + P(X = a)$

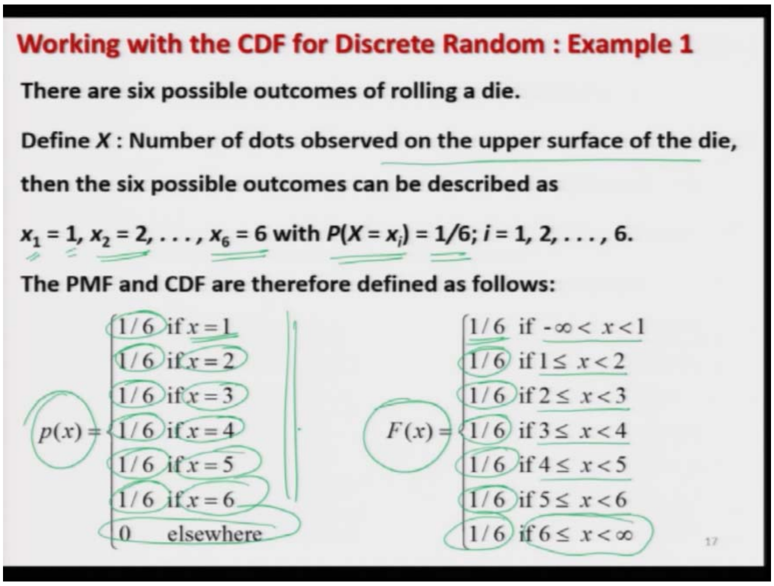
$$P(a < X \leq b) = F(b) - F(a)$$

$$P(a < X < b) = F(b) - F(a) - P(X = b)$$

$$P(a \leq X < b) = F(b) - F(a) - P(X = b) + P(X = a).$$

So, using these expressions, you can very easily compute the different types of probabilities using the CDF. So, in case if CDF is given to you, you can now compute the probabilities either in the case of discrete random variable as well as continuous random variable.

(Refer Slide Time: 20:33)



So, now, let me try to take take an example and try to show you here how it will look like. So, for example, suppose we roll a dice and so there are 6 possible outcomes as the number of heads on the upper surface of the dice. So, there can be 6 possible outcomes which I am going to describe by the value $x_1, x_2, x_3, x_4, x_5, x_6$. So, x_1 is the value 1, x_2 is the value 2 and similarly, here x_6 is the value 6. And we know that probability that X equal to x_i is simply 1/6, that when probability of observing 1, 2, 3, 4, 5 or 6. This is 1/6.

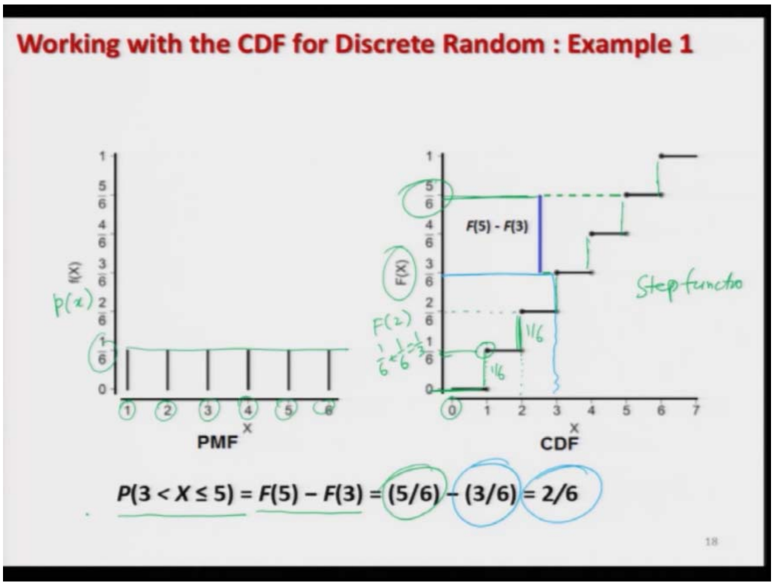
Now, the probability mass function this is very easy to find that, when x takes value 1, the probability is 1/6, when x takes value 2, the probability is 1/6, when x takes value 3, the probability is 1/6; when x takes value 4, the probability is 1/6; when x takes value 5, the probability is 1/6; when x takes value 6, the probability is 1/6; and 0 at all other places. If you say that what will happen if x takes the value 7, 8, 9, 10 etc., etc., I will say, the dice cannot have these values. So, the value of p(x) is going to be simply 0.

Now, in case if you want to find out the cumulative distribution function, so, remember what is the meaning of cumulative? As the meaning suggests that you have to accumulate the probabilities up to that point. So, in case if I say you your value x lies between minus infinity and 1. So, this value is going to be here only 1/6.

And similarly, if your x is going to be greater than or equal to 1 and less than 2 that way it can take only here one possible value because x cannot take the value like 1.2. Because you can see here from this probability mass function that it can take only integer value. So, this probability is going to be here 1/6 and similarly, x is greater than or equal to 2 less than 3, this is again going to be 1/6 because this is simply the probability of x equal to 2. And similarly, here the probability of x equal to 3 can be obtained as that x lies between say 3 inclusive and less than 4, this will be 1/6.

Similarly, x greater than or equal to 4 less than 5, this is 1/6. x greater than or equal to 5 less than 6, this is 1/6. And similarly, if x is greater than or equal to 6 then less than infinity, this is again 1/6, because there is only one value in these intervals. So, now, you try to have a look that how this px and this capital Fx will look like.

(Refer Slide Time: 23:22)



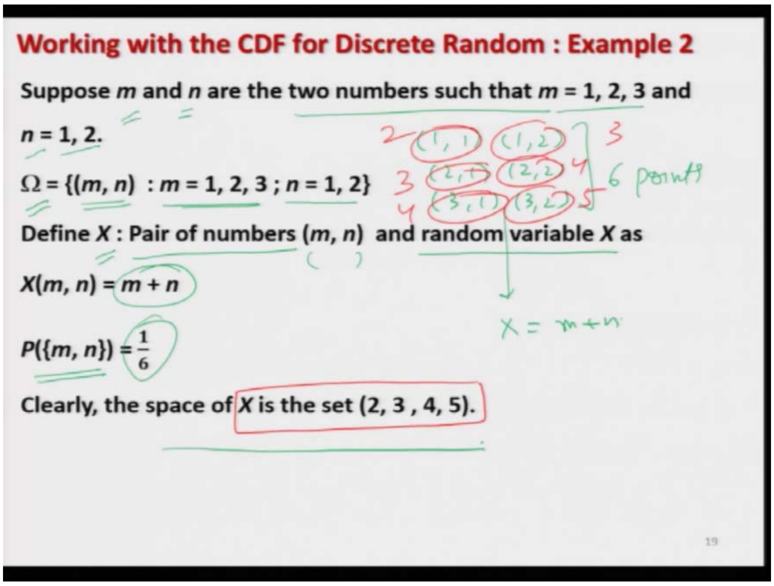
So, we tried to plot these functions with respect to the value of X. So, you can see here when this X is equal to 1, 2, 3, 4, 5, 6 it is taking the value here 1/6, which is here like this. So, this is how the graph between say here p(x) and X will look like. Now, I tried to look at the function this here Fx and I tried to plot it. So, you can see here this value is increasing by 1/6, 1/6 and so on. So, in case if you try to plot here the CDF, it will also take the value between 0, 1, 2, 3, 4, 5, 6 and so on. So, you can see here at 0, the value here is 0, and this will remain 0 up to here 1, at 1 it will take here a jump and it will come to 1/6.

And now, it will continue up to here point 2 and if you come to here point 2, once again it will take here a jump off 1/6. So, now this value is your actually here F(2) that will be 1/6 plus 1/6, which is equal to here 1/3. And similarly, if you try to come here, say 3 you can see here at 4, at 5, at 6 and so on. So, it will try to take a jump at an interval of 1/6. So, the, this CDF F(x) will look like a step function. Do not you think that it looks like as a stairs like this one. Means is if you want to climb on it, you can climb it.

And from here, in case if you want to find out suppose the probability that X lies between 3 and 5, that is X is greater than 3 and and less than or equal to 5. This can be obtained as F5 minus F(3). So, now the value of F(5) and F(3) they can be obtained directly from this curve. So, value of F(5) here you can see here, this is somewhere here. So, this is 5/6. So, I am writing here 5/6. And the value of here F(3), this is here somewhere here, I will use a different colour of pen, you can see it this is here you can see here.

So, this is here 3/6, so now, if you try to solve it, this will come out to be here 2/6. So, this is the graphical representation of the PMF CDF and as well as how do we try to find out the probabilities. Similar to the case of PDF, where I had shown you that the probabilities from the PDF are the area under the curve between the limits of the integral.

(Refer Slide Time: 26:02)

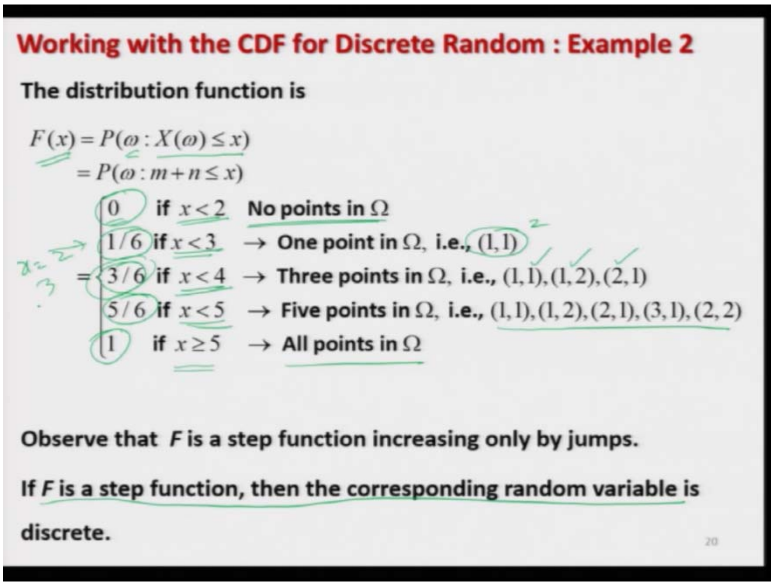


So, now, let me take here one more example before I conclude the lecture, to explain you these things in more lucid way. Suppose small m and small n are the two numbers such that m takes value 1, 2, 3 and small n takes value 1 and 2. And we define here a random variable here X, which is the pair of numbers something like mn. And the random variable X has a form that it is the sum of both the numbers m and n.

So, now, first I have to determine what is the sample space? So, the sample space Ω will consist of all pairs of numbers of m and n, where small m goes from 1, 2, 3; n goes from 1 to 2. So, this is something like 1, 1; 1, 2 and then here 2, 1; 2, 2 and then here 3, 1 and like here 3, 2. So, you can see here all together there are 6 points. And assuming that all are equally likely, means I can write down here the probability of observing any pair of the point is 1/6.

So, now, in case if you try to look at this set, and try to find out the value of your X as m plus n, you can see here that it is trying to take the value here 2, here 3, here 3, here 4, here 4, and here 5. So, that means the space of X will have the values only 2, 3, 4, 5. Or we can say that the space of X is the set of the point 2, 3, 4 and 5.

(Refer Slide Time: 27:51)

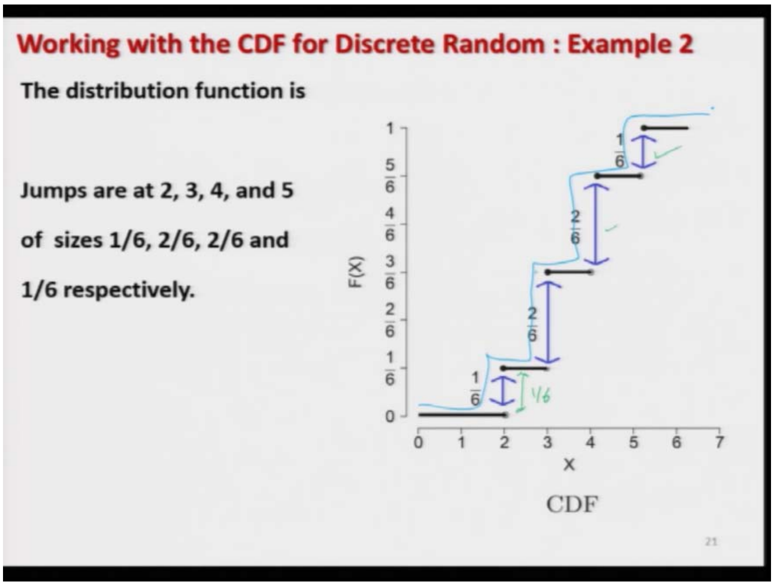


So, now, we try to find out the cumulative distribution function of x. So, we know by definition that a capital Fx is given by the probability of that ω for which X(ω) is less than or equal to x. So, now, we can write it that you can see here that it is taking here the four values means something like 2, 3, 4 and 5. So, when x is equal to hear 0, that means x is less than 2, there are no points in Ω for this value here is 0. And if I try to take here x less than 3, so that means there are there is only one point 1 and 1, which is giving me the value here 2. So, this probability is simply here 1/6.

And if I try to take here x less than 4 that means all the values of x for which the value here is 0, 1, 2, 3. So, there are only 3 possible values 1, 1; 1, 2; and 2, 1 that means there are 3 points, it will contribute to this event, so the probability will become here 3/6. And similarly, for here x say less than 5, there are 5 points in ω which are satisfying the condition. So, this probability becomes here 5/6.

And in case if you try to take x greater than equal to 5, that means all the points are in ω and this probability becomes here 1. So, now in this case also you can see here that this F(x) is a step function and which is increasing only by jumps. And this exactly does similar thing which I explained you here in this curve. Although here in terms of probability, but now I am taking her in terms of this jumps. So, one thing now I can conclude that if F is a step function, then the corresponding random variable is discrete.

(Refer Slide Time: 29:54)



And it will look like actually here like this. If you try to plot it, it will be, if you try to look here at say, say X equal to 2, this is 1/6; and for 3, it is a 3/6; and so on. If you try to plot it, it will look like this. So, the you can see here the jumps here at 1/6 and here it is 2/6, 2/6, and 1/6. And if you want to compute any probability you can compute it and if you try to create here a staircase you can very easily obtain like a like this here. So, if you want to climb on this one you have steps here, that is why this is called here as a step function.

So, now, we come to an end to this lecture, and you can see that here I have considered the concepts related to the discrete random variable, and that was practically a copy of the last lecture, where I had explained you the same concept in the case of a continuous random variable. So, there is no difference actually in the conceptual part the only thing is this, how are you going to handle it in terms of discrete and continuous random variable.

So, once again, I will request you try to take some example, try to look into books, like try to look into assignments, and try to practice it. The more you practice it, the more you will understand. Do not get confused that when I am asking you to solve very simple problem, how are you going to learn the data science? I promise you, you will learn the data science because once you are trying to practice, you have to take very simple and a small example, so, that you can understand what is really happening with the conceptual point of view.

Once you are clear about the concepts, then then whatever is happening inside the computer with those large number you should be confident, you will be confident actually. So, try to practice it and I will see you in the next lecture. Till then, good bye.