Data Analysis and Decision Making - I
Prof. Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 07
Baye's Theorem and Distributions

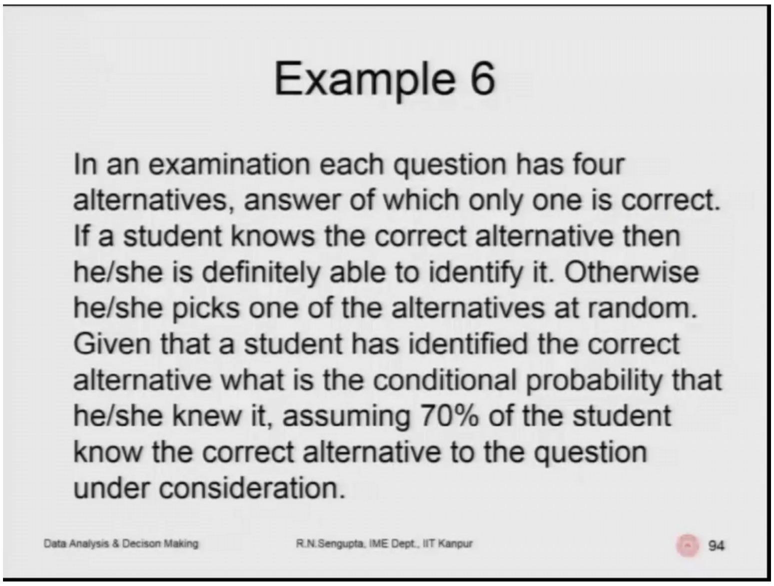
Welcome back my dear friends; a very good morning, good afternoon, good evening to all of you. And this is the DADM 1 lecture, and this is the 7th lecture. As you know this is the course of 30 hours which will be spread over 12 weeks each week we have 5 lectures, each lecture being of 30 minutes. So, if you remember in the last lecture which is the sixth-one we were discussing the concept of Bayes theorem. And I did mention that how the concept of Bayes theorem could be understood.

So, Bayes theorem is basically coming from the actual background mean conditional distribution or conditional probability whatever you want to say. So, with condition and probability we went with the concept of Bayes theorem. Now, rather than basically going through the equation, I basically do a neat diagram, explain that how if the overall universal set is broken down by n number of mutually exclusive and exhaustive events being given from by B 1 to B n, and then we can actually have an event a which is basically a conglomeration of some intersection of B's such that we can find out the probability of B j; j being from 1 to n.

So, all those n's which were basically mutually and exhaust mutually exclusive and exhaustive. We can find out the probability of B j's provided a has also already occurred. And it can be extended into the second stage, third stage by the word of second stage third stage you means and they are different type of combinations from one stage to the other; that means, given B you want to find A, given A you want to find out C, given C you want to find out D and so on and so forth.

Now, let us consider a very simple example Bayes theorem, and then I basically I will try to build up a more such conceptual problems which can be solved using the Bayes theorem. So, the problem is like this.

(Refer Slide Time: 02:13)

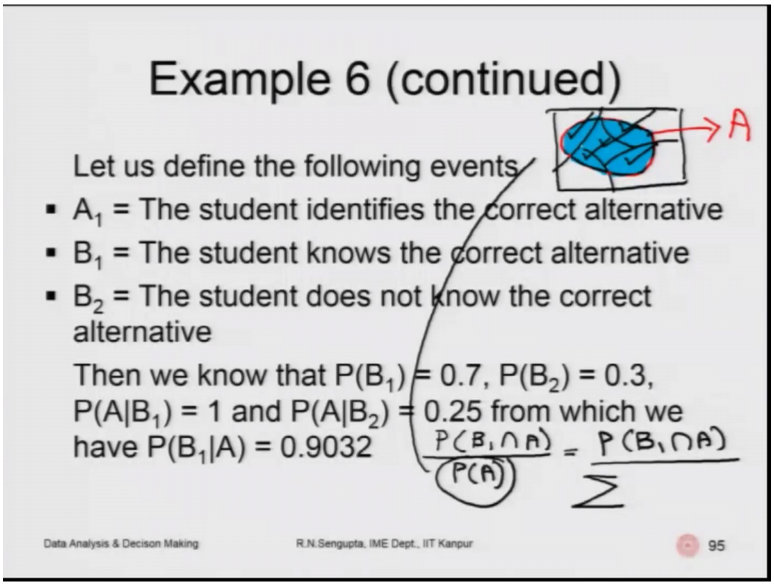


In an examination each question has four alternatives. So, it can be more than four also less than four also, but for our problem you have four alternatives of which only one is correct. Now, the problem again can be extended with more than one as correct multiple choice questions, but let us keep it simple. So, if a student knows the correct alternative then he or she is definitely able to identify it. Otherwise he or she picks one of the alternatives at random. So, we will consider some probability distribution of probability values of picking the random answer.

And there would be a simple concept a very simple practical concept for that also. Now given that a student has identified the correct alternative. We on and what is the conditional probability that we have we are required to find out. So, what is the conditional probability that he or she knew it, knew the answer assuming 70 percent of the student know the correct alternative to the question under consideration.

So now with this what are the assumptions we will consider and how we will proceed. So, let me come to the problem directly and it is a very simple one. So, let us define the following concepts. So, what are those are the assumptions? Let us consider A suffix 1 or A 1 be the event that the student identifies the correct alternative.

(Refer Slide Time: 03:32)



B 1 and B 2 which are technically the mutually exclusive and exhaustive set, try to remember the diagram which had drawn on the fag end of the 6th class. So, B 1 and B 2 are the mutually exclusive and exhaustive. So, B 1 denotes the student knows the correct alternative. B 2 denotes the student does not know the alternative correct alternative.

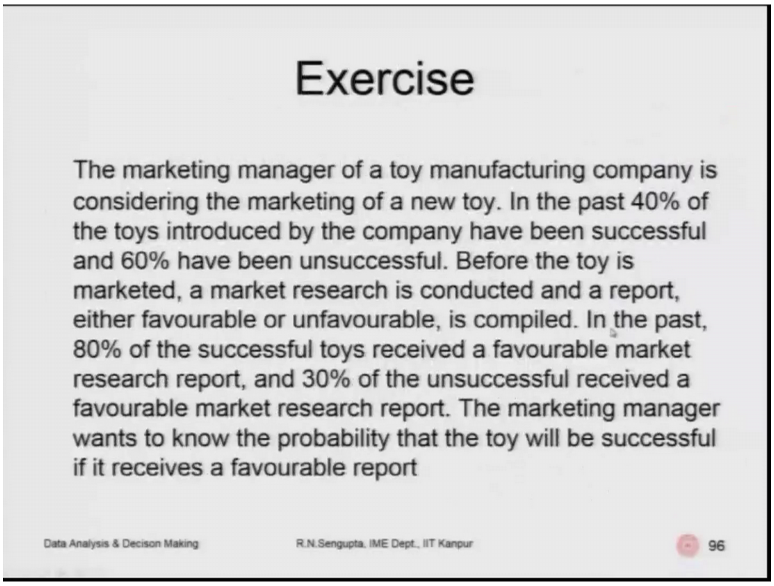
So, what are the corresponding probabilities let the answer it very simply, then we know if the student knows the correct alternative. It is given by 70 percent as and the question so, it is 0.7. And the complement set of that is the student does not know the alternative is 0.3 which is 30 percent. Now what are the scenarios? Given the student knows the correct alternative, and if A is the event then the student A 1 is the event, the student identifies the correct alternative; obviously, if we knows it we will answer it correctly hence the probability is 1. So, probability of P A or P A 1 given B 1 the answer is 1.

Now, let us come to the second information. Now if this student does not know the alternative; that means, he is are not sure. So, if he has not sure in front of him he has four alternatives. So, the property of unsealing anyone one of them is basically one fourth. Hence the probability that of A; that means, answering and the correct alternative by the student provided he or she was not aware of the alternatives is given by one-fourth which is 0.25. So, if this is the case now, we will try to answer the question what is required. The answer required is the probability that the students knows the correct alternative provided he has and identify the correct alternative we will just formulate it as a simple base theorem given the problem and solve it.

So, the actual so, this is what we want to find so, this A if you remember and go into this diagram. So, these are the B 1 B 2 B 3 and I want to basically have A. So, this is A, I want to find out the probability of A so; obviously, and use the highlighter use the yellow color. Let me use the green color it would be or the blue it would be easy. So, I want to find out this probability.

Then this probability, I am going to add them up. So, what I am marking I want to find out this probability, I want to find out this probability, I want to find out this probability. So, if you remember for A let me use the black. So, this one this probability of A, if you basically go through it is basically the sum of this, this, this, this and this. So, this can be found out put it into the value and the answer comes out to be 0.9032. I am not solving them. But I am trying to give you a pictorial diagram and you can understand in much better.

(Refer Slide Time: 07:28)



An exercise which I would love the students to do, remember one thing. As we have started now there would be many exercises which come up as we are discussing, these exercises are not part and parcel of the assignment, not and part to and parcel of the end semester also.

But please solve them try to understand, because as we are going through the concepts you are understand that how they would make some sense based on the fact that we had covered one concept. We will solve one small problem and then we immediately come up with the small like exercises and that you can understand it very clearly.

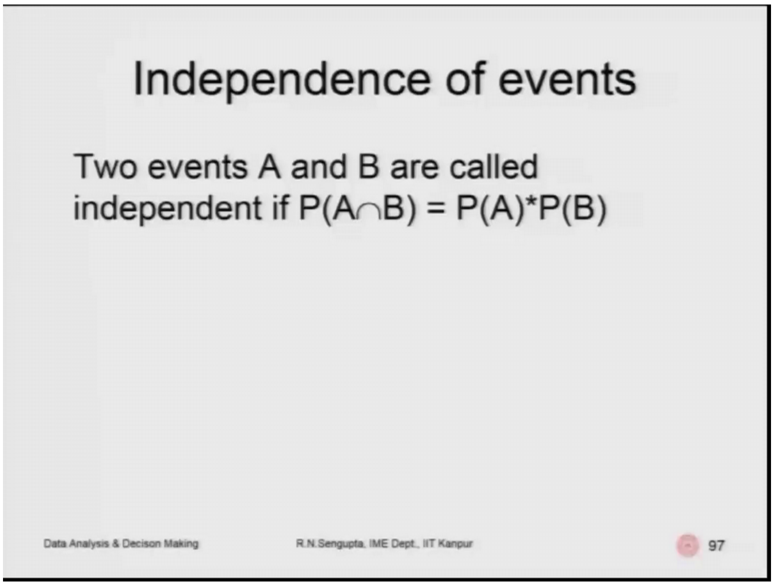
The problems I am telling you before and are very simple, easy understand the concept and solve them. So, the exercise or the example is like this. The marketing manager of a toy manufacturing company is considering the marketing of a new toy. In the past 40 percent of the toys introduced by the company have been successful and 60 have been unsuccessful.

Before the toy is marketed a market research is conducted and a report either favorable on unfavorable is submitted is compiled. In the past 80 percent of the successful toys received a favorable market research report. And 30 percent of the unsuccessful if we see would a favorable market research report. So, it may be possible that you do a marketing survey. So, it may it is not 100 percent that you are able to predict everything. So, even if the marketing survey the data may give you that the market is actually bad, but still it is given a good set of information to you or it may be possible the market is good.

But still the marketing survey is give you giving your bad survey report. So, any combinations are possible. And obviously, other two combinations are the marketing survey does really, analyze the market, and it is good and it gives you a good a report. And it unless the market the market is bad and it he gives you a bad report. So, those are actually what it is and you are getting the information. But the other two instances where it is good, but you get a bad report or it is bad you get a good report are the problematic area. But and obviously, we will consider such problems later in the hypothesis testing.

So, the marketing manager wants to know the probability that the toy will be successful if it receives a favorable report. So, basically you would analyze the from market which one should be B 1, which one should be 2 which should be B 3 such that B 1 B 2 B 3 or whatever how however, different values of n you have you basically combine them and try to find out that they are mutually exclusive and exhaustive and solve the problems accordingly. Now, what we mean by independence of events?

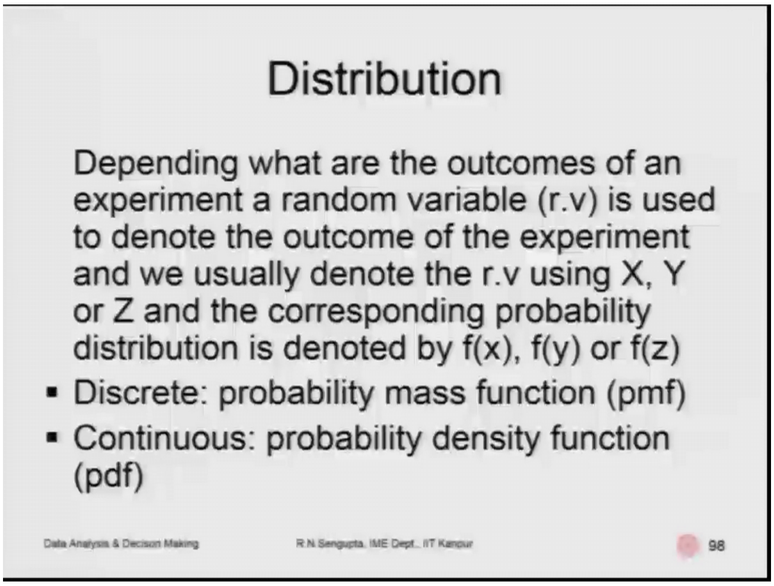
(Refer Slide Time: 10:15)



Now, here be careful what I mentioned and note down it very carefully. So, we will say 2 events are independent if the pro the probability of the intersection of A and B, A and B are in events which are and part and parcel of the universal set. So, we will see A and B are mutually independent of each other taking two at a time if the probability of a intersection B is given by the multiplicative value of probability a into probability B.

Now if I am saying that if there are 3 events if they are independent, how would we analyze? It will be given the prop that 3 events A 1, A 2, A 3 are independent of each other. If probability of A 1 intersection A 2 intersection A 3 is given by probability of A 1 multiplied by probability A 2 multiplied by the probability of A 3. So now, further on now what we will now consider is basically a little bit different we will consider the different above distributions.

(Refer Slide Time: 11:16)



And we will come to the uses of distribution in different general examples. Now, depending on what is the outcome of an experiment, you will basically have a random variable to denote that what is the actual thing we are going to study, what we are going to learn? And you will consider that is to be random, because the fact that we do not know what should the outcome be of the experiment which you are trying to study. So, this random variable which is r.v you would be used to denote the outcome of an experiment.

And we will try to basically denote the random variable by either the capital symbol of the variable X or capital Y or capital Z. Now the reason why we use capital one is basically denotes the random variable which is unknown before we do the experiment or as we are doing the experiment. The moment we complete the experiment the value which is in front of us is known. It is known as the Rayleigh's value, and the random variable may take those Rayleigh's value depending on what is the actual probability function based on which you are trying to map, the random variable and the different type of Rayleigh's value.

So, this basically a mapping where you take a random variable x, put it into the function that function what do you want to find out with the probability distribution function, and the output which will get would base be based on the fact that what capital X takes that is small Rayleigh's value is small x. Based on when you put the small Rayleigh's value small x into the function you get a actual function and value which is known as the probability distribution function.

So, let us consider a very simple example, consider a coin. And it is in unbiased one. So, there is a head, there is a tail. Consider that rather than head and tail the coins in one face you have red color another case based face you are basically yellow color not. So now, before you toss or as the coin is tossed in the air, we are interested to find out that what face will come out.

So, we will denote the random variable capital X by the face which will mean that we want to study the face which will be coming out on the top. And we know that, for that face which is a random variable, which we do not know as the experiment is being done. But what we actually know is that once the face actually comes out they are basically 2 Rayleigh's values one is the red. One is the yellow one.

Now, you may say that why you are basically trying to depict them as colors, but then how does the value of numbers come in to as the probability sense. Now consider the example that those red and yellow are removed. But you put that in the case of a coin where the faces are now black and white. Or consider you consider the coin is in such a way that the faces in one face there is one dot and another face there is another there is 2 dots.

So, your actual aim is to find out the outcome, but what you are trying to do is that that outcome is mapped onto a function which is basically a mathematical functional formulation, which gives us that what is the outcome and what is the corresponding probability based on which we are trying to study that outcome.

Now, consider their second example, again the probability example not of the coin, but of a dice. Consider the faces are marked 1 to 6. So, obviously, this why I am saying 1 to 6 or why I told one dot for the coin or 2 dots for the coin is basically to give you an example such that the underlying factor of the experiment which you are trying to do need not have any quantitative values assigned to them, but the actual output based on which we are trying to understand the concept of probability are all quantitative in nature.

Because they are some functional form such that once the capital X whether qualitative or quantitative whatever it is takes is Rayleigh's value, once you put the Rayleigh's value into the functional form you get an output which is in number terms that is the probability. Now consider the dice example, if you roll the dice the probability of getting any face that is 1 to 6 is given by one-sixth. Now replace that experiment with the same site of sort of dice, say exact same density, same weight, same shape, whatever it is same material.

But now the faces are marked by the letters ABCD so, A for 1, B for 2 and so on and so forth, now you again do the experiment. In the initial case, the experiment outcomes were dots 1 to 6 any one of them. Now, but now in this second experiment that the they are not the dot, but they are the faces are the alpha v alphabetic values ABCD whatever.

So, when you are doing the experiment in both the cases the capital X the random variable where, in one case the random variable denotes the outcome which is unknown. In the second case also the outcome is unknown, but when actually x takes place and you know that in the first case small x which is the Rayleigh's value gives us the point which is either 1 or 2 or 3 or 4 or 5 or 6.

In the second case when the Rayleigh's values comes they are the value either A or B or C or D or E or F but, the corresponding functional mapping when you put these values 1 to 6 or A to F into the functional function f of x. The outcome is the probability which is 1 by 6 in both the cases provided they are unbiased point. Now we will consider that the outcomes which you are doing for the experiment can be either continuous or they can be they discrete.

So now you will be asking what do you mean by continuous. And discrete is the x values which are the outcomes can be either be the discrete or continuous based on the fact hence we will have the probability mass function for the case when it is discrete, and probability distribution function for the case when they are continuous. So, these would basically given by the functional form of small f of x in both the cases, and we will analyze them in details later on.

So, let me continue reading it. So, we will usually denote the random variable using x y or z, and the corresponding probability we will given as I just mentioned that by small f of x or small f of y or small f of z.

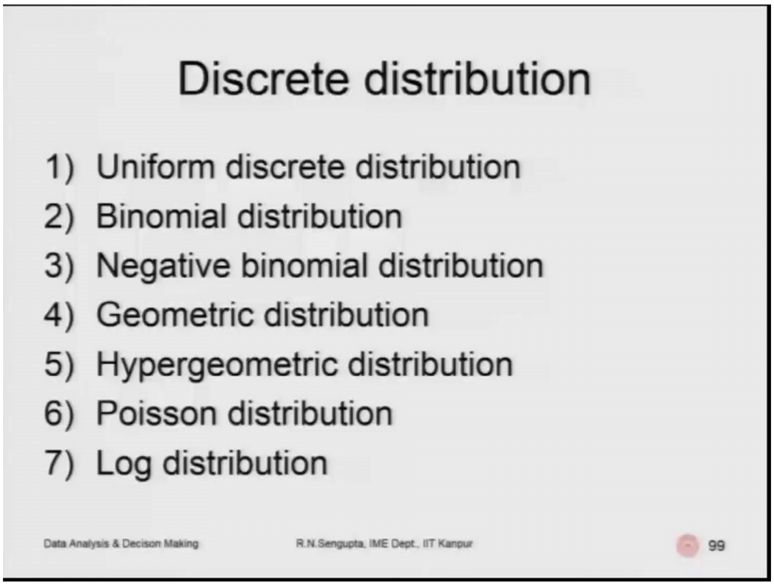
Now, this f of small x, f of small y, f of small z, they are not small f of capital X, or not small f of capital Y because, those capital values denotes the random variable. Once you plug in the values which is the Rayleigh's one that will give you what is actually the pdf or the pmf or the distribution we are trying to consider.

Now, if it is a discrete distribution, we will have basically the probability mass function; that means, the probabilities are concentrated on masses. And if it is continuous it is known as the probability density function; where the random variables if you are trying to understand or trying to study are continuous in nature. Now let us come back to the example of a probability mass function. This is the example which we concern these 2 examples of a tossing a coin or rolling a dying.

Now, when you basically me roll the dye, the probabilities are concentrated at the values of the Rayleigh's values which are either any one of them either one. So, hence the corresponding probability which is being assigned, or the masses which are being assigned to the value one, and the corresponding probability is 1 by 6. If you come to the example of say for example, when you are trying to analyze the probability for the Rayleigh's value as 2 again it is 1 by 6. Where it is 3, it is again 1 by 6, 4 it is again 1 by 6, 5 again 1 by 6 6 again 1 by 6 which means, that the probabilities are concentrated at the values 1 to 6, but any probability values at the point other than 1 to 6 are 0.

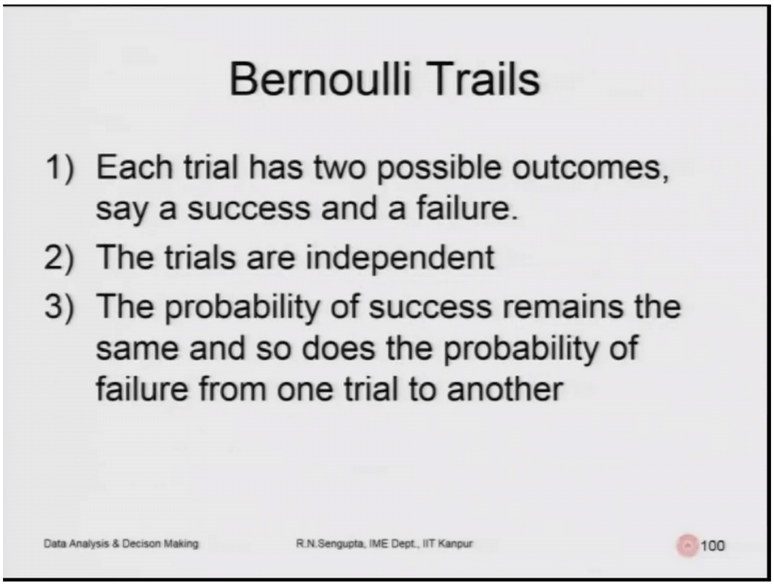
Now, you may be asking that what does it mean. It does not show such then when you are trying to a plot the value of the cdf which is the cumulative distribution function; I will come to that later on. But in the case when you are trying to analyze the property, density function we do not consider the densities are concentrated on only points, and unlike basically analyze that in later on in more details. So now, we will consider the discrete distributions.

(Refer Slide Time: 20:16)



And I will and whatever I have talked about showing that once life for about 10 minutes I will go into the details about the discussion later on also. So, the discrete distribution which will consider for our example in this DADM 1 course would be the uniform discrete distribution the binomial distribution, the negative binomial distribution, the geometric distribution the hyper geometric distribution, the Poisson distribution, the log distributions and there are other distribution also which we will not consider in details. These would be more than sufficient to basically at least cover the probability mass function.

(Refer Slide Time: 20:53)



Now before we discuss the probability and discrete probability distribution or the pmf probability mass functions. We will like to basically go into the concept very simple concept of Bernoulli trails. Now, in the Bernoulli trail, we have basically two outcomes. It can be more than two outcomes also not for the Bernoulli trails, but for the experiment, but we are not going to go into the details of that. So, there are un experiment or trial which you do or an x, or some random experiment it you do, and there are two possible outcomes. Say, there is a success and a failure so that the very good example would be a head and a tail.

But if you are considering rolling the die, it is not, if and only if we are able to divide the rolling of the die into real now by into even numbers and odd numbers; obviously, there are two categories, then it can be said depending on how we have been able to formulate the experiment as Bernoulli trail. But let us not go into such details, but let us come to the simple example. So, each trial or the Bernoulli trail I have put has two possible outcomes which is a success or a failure, or a head or a tail, or a yes or a no whatever it is.

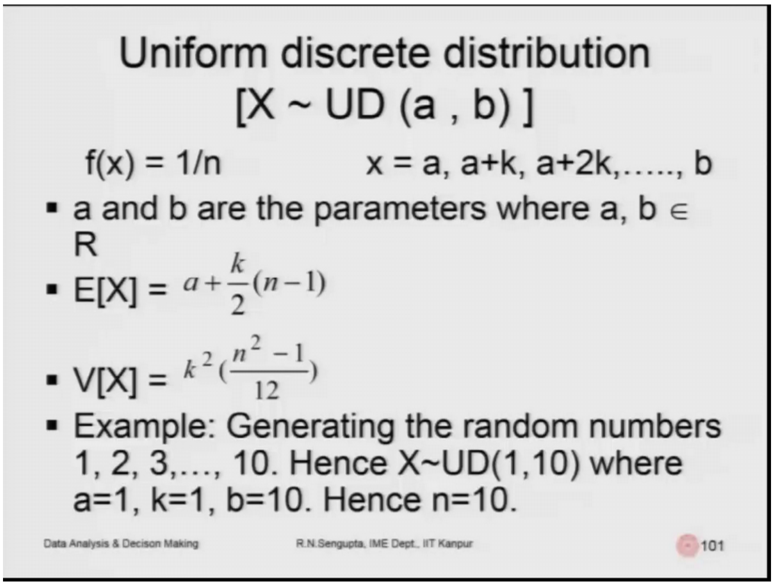
This trials are independent; that means, when we do the experiment the outcome of the head or outcome the yes, or outcome of the red, be or outcome of white whatever it is does not affect the complimentary event of that that is for the yes it is a no, for a head is a tail for a white is a black whatever it does not affect that and vice versa. Also, that means, the trials are independent. Also which basically means that if I do the experiment repeatedly; like, I roll the die and I want to find out the probability of getting a even number or odd number or the experiment which I do is basically I toss the coin I want to find out the head or the tail.

If I keep doing this tosses these are each trials, these trials are independent of each other; that means, one trial does not affect the next or the next trial does not affect the form what. So, they are independent of each other. Now the point third point I have already mentioned it means the probability of a success or a failure remains the same and it is not does not change from experiment to experiment.

Now, there are other important points which I also mentioned. Consider that you have a coin and you keep tossing it. So, the trials do not affect each other they are independent. That was point number 2 as mentioned in the slide. Point number 3 is that the outcome of ahead or outcome of a tail, whatever the outcomes are we have denoted in toss in the coin are independent of each other do not affect; that means, it is not that if the outcome of the head and the tail for an unbiased coin is half and half the probability does not change from trial to trial. It remains half and head does not affect the tail in the next outcome or head does not affect the head from the last outcome and or tails, tail does not affect the tail or tail does not affect the head and it continues in that way.

In case if you have a trial whether 3 outcomes the same concept can be utilized where the trials will be independent of each other there will be 3 tails whatever the probabilities are they can be P 1, P 2, P 3. And the probabilities of getting the first one which is P 1 that would not affect P 2 or P 3 and vice versa in the sense P 2 would not affect P 1 and P 3. Similarly, P 3 would not affect P 1 and P 2 and it will continue in this way. So, let us consider the first discrete distribution which is the uniform discrete case

(Refer Slide Time: 24:37)



So, x is the random variable, and we denote as I am just hovering my pen here. It means UD which is the uniform distribution. And this a and b are the parameters based on which we are trying to basically find out the uniform discrete distribution case.

So, U is uniform D is discrete; a and b are the parameters. Now the f of x which is the pmf is denoted by 1 by n. N is basically the number of such trials which we have. And the Rayleigh's values of x which is the random variables are in the in the least case, you can have basically the minimum value which is a and the corresponding masses are concentrated at points which are a plus k; that means, the next jump is basically from a to a plus k the next jump further on would be a plus k plus k which is a plus 2 k and it will go on till the last value which is b.

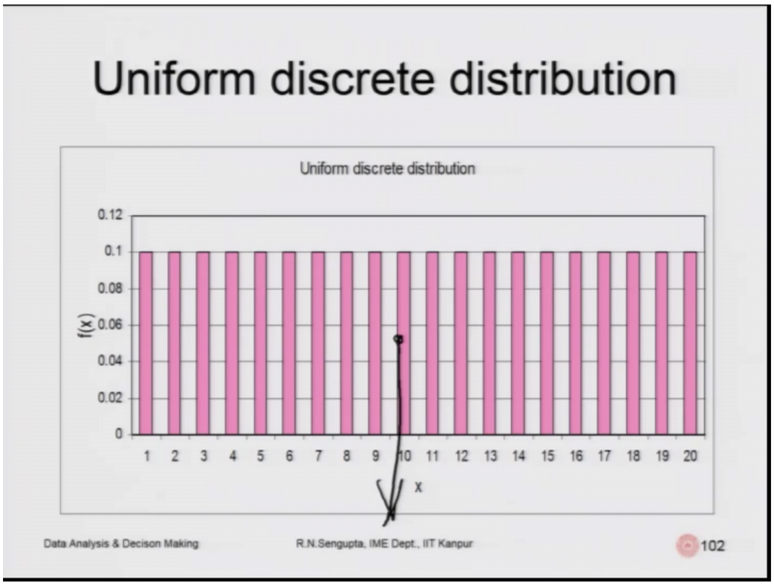
So, that means, the masses are concentrated at the Rayleigh's values which are a plus k a plus 2 k so on. And so, till b and they are such n number of such masses where the probabilities can be found out. Here a and b are the parameters which are on the real line. The expected value I will come to the expected value and the variances concept later on when we solve the problems.

So, they and they will be utilized later all also. The expected value is given by a plus k by 2 into n minus 1 and the variances given by k square into n square minus 1 divided by 12. So, what can be the example? Example can be generating the random numbers 1, 2, 3, 4 till 10. Here the random distribution of the probability mass function will be given by x is in uniform distributed between 1 and 10.

So, a is 1, b is 10; where the values of k provided they are only discrete number like 1, 2, 3, 4 in that case a is 1, then a plus k is 2. So, hence k would be 1, and you can find out that what are the different numbers of ns which you have based on which you can formulate the ppm ppm f which is the probability mass function we can find out the expected value you can find out the variance also.

Now, remember one thing for the uniform distribution case. If you are able to draw it how it will look like I will come to that later on and as also signify the main concept of the expected value. There is a reason for that and I will come to the variances expected value for the other distributions later on also. So, this is a uniform discrete distribution I have do it in very simple x n.

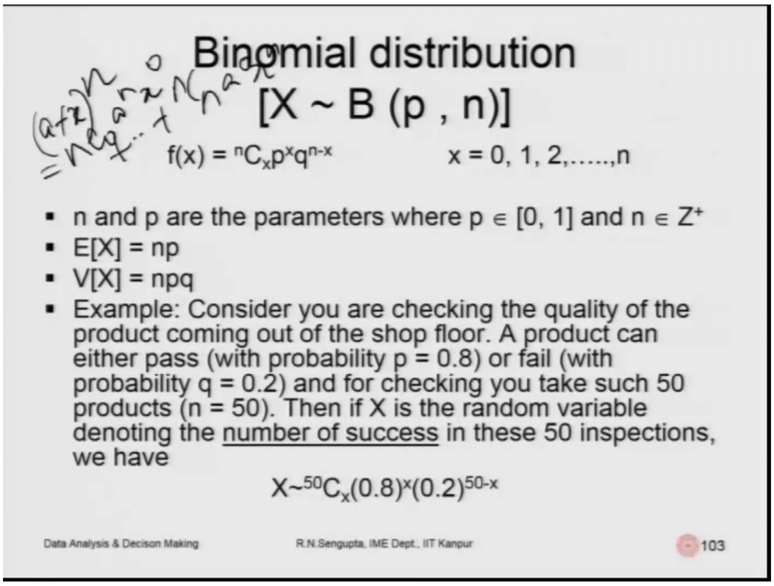
(Refer Slide Time: 27:19)



So, the values are of the x's are from 1 to 20. The corresponding probabilities are given by 0.1. So, these are uniform discrete so the heights are all equal. So, f of x is plotted along the y axis, and x values are plotted along the x axis. And if you find out the expected value would be the center of gravity of this figure. So, obviously, the expected value would be denoted by somewhere in between. So, is this I am just using a very simple concept. Center gravity which we know in physics are very simple plus 10 or 12 concept of simple sciences.

So, so center gravity is basically where you can basically balance it and the weight would basically be kept balance. Like, if you have a scale of one feet which is 12 inch or 30 centimeters, if you balance it at the point which is 15 centimeters it will be keep balance because the weights on the left hand side and the right hand side are equal. So, it is the same way what we denote. So, we will consider now the binomial distribution.

(Refer Slide Time: 28:28)



So, binomial distribution is given by if you remember the binomial expansion which we have done in class. So, it is given by basically a plus x to the power n. So, you want to basically expand it. So, the first term would basically be n C 0 a to the power n x to the power 0. And the last term would basically be n Cn a to the power 0 x to the power n. So, basically you find out all the powers of x and a as they keep changing such that x increases from 0 to n, and simultaneously a basically changes from n to 0, such that the power values when they added up for a and x are always n.

Now, here n np are the parameters, and the random variable is x, B is the binomial distribution the parameters are p and n. Now p is the probability of the of the good thing or the bad thing whatever it is; however, you denote or yes or no. So, if it is p is for yes, then probability of no would be 1 minus p. If p is no, then the probability of s would be 1 minus p, whatever way you denote. Now the f of x is given by n Cx B to the power x q to the power n minus x. And the x values are from 0 to n.

So, x being 0 means the corresponding probability of getting the good item is 0, there is no good items and prob and the values of the all the bad items are all the n. And if basically x is n it means, all the good items are there while the number of bad items are 0. So, n and p are the parameters. So, p is between 0 and 1, n is integers positive, expected value is given by n into p variance is given into n p q.

So, consider an example. You are checking the quality of the product coming out from the shop floor, a product can either pass with the probability a 0.8 and fail with probability of 0.2. It can also be the other warehouse; like we want to find out the probability of or the bad item. So, obviously, in that case p is 0.2 and q is 0.8. Now you are checking 50 such items which n is 50. If x is the random variable which denotes the number of success, then for the inspection of such 50 items it will be 50 Cx which is n is 50 p to the power x which is 0.8 to the power x, and q to the power n minus x which is 50 minus x would be given by 0.2 to the power 50 minus x.

Now, remember if you want to find out the bad items again the formula remains the same initial part 50 Cx would remain the same. It will be 0.2 to the power x and 0.8 to the power 50 minus x, but in that case x is a number of bad items while in this case which is shown on the slide is the number of good items. So, with this I will end this 7th lecture and continue the discussion further on with examples from the binomial distribution and so on and so forth. Have a nice day.

Thank you very much.