Essentials of Data Science with R Software- 1
Professor Shalabh
Department of Mathematics & Statistics
Indian Institute of Technology Kanpur
Lecture No. 32
Bernoulli and Binomial Distributions

Hello friends welcome to the course Essentials of Data Science with R software 1 in which we are trying to understand the basic concepts of probability theory and statistical inference. So, now, you can recall that in the last lecture, we had initiated a discussion on different types of probability mass functions and we had considered the discrete uniform distribution. So, now, in this lecture, we are going to consider some more probability mass functions and basically we are going to consider the but Bernoulli distribution and binomial distribution.

So, the first question comes under what type of conditions these two probability functions are going to generate the probability or I can say under what type of conditions you can use these probability mass functions to compute the probabilities and different types of properties of the data. So, now you can recall that in the last lecture, we started with degenerate distribution.

So, in the degenerate distribution, there is only one point at which the probability is concentrated. Now, I will try to extend it and I will try to consider that suppose the probability is concentrated at two different points. Now, as soon as I make this statistical statement, you get badly confused that what is the meaning of this and this is going to happen in the entire course.

But my suggestion to you is that try to think carefully what I am trying to say, I am simply trying to say that the probability is concentrated at two different points, only two points. Now, can you think that is there any practical condition a practical situation where such a phenomena can happen? We have had numerous examples not one not two.

For example, if I say whenever somebody is appearing in an examination, there are only two possibilities success or failure and definitely success and failures have some associated probability when can always say that, there are 70 percent chances that the student will pass the examination and there are 30 percent chances obviously 30 percent will come as 100 minus 70 percent. So, 30 percent chances are there that the student may not succeed in the examination.

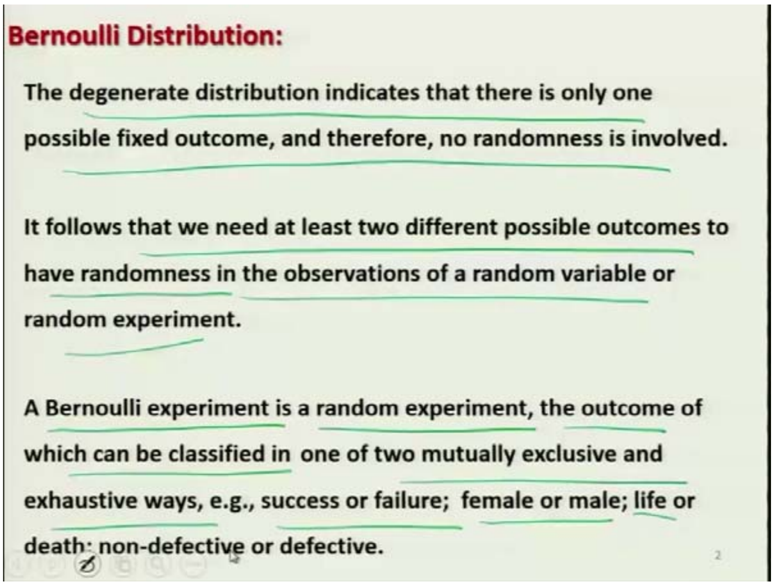
Suppose, if you want to know whether it will rain today or not, you can say, there are 80 percent chance that it may rain today. So, obviously there are 20 percent chance that it may not rain today. So, in all these cases, if you try to see the probabilities are concentrated only at two different points. So, you can see it is pretty simple, the only thing is this you have to develop a statistical thinking, how to think statistically.

So, now in this case, I will say simply that we can use the Bernoulli distribution and similarly, in case if you try to repeat this experiment for certain number of times, then this Bernoulli distribution will help us in getting a binomial distribution that will also be used under some special type of conditions. Well, what are those conditions? That is what we have to learn here.

So, one thing I would say that whenever I am trying to initiate a discussion on any probability mass function or say any probability density function, try to divide your thought process into two parts. One is theoretical and second is application oriented that where it can happen, although I am trying my best to take couple of examples in every probability function, so that I can convince you that these things are going to really happen in practice.

And I will also try to show you that the same probabilities that you are computing theoretically using the probability mass functions how they can be completed in the R software directly. Well, those things will start now. So, let us begin our lecture and first we try to understand the Bernoulli distribution.

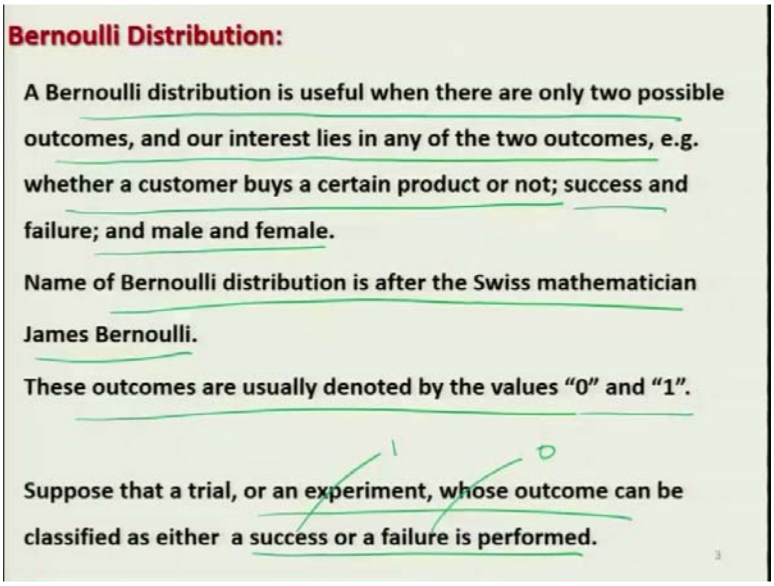
(Refer Slide Time: 05:16)



So, now as I said the degenerate distribution indicates that there is only one possible fixed outcome and therefore, there is no randomness involved in it and it follows that we need at least two different possible outcomes to have randomness, at least you should have two different outcomes in which the probabilities can be divided. Otherwise, there is no variation in all those events like sun will rise in the east direction, it will always happen there is no randomness, so we need at least two values, so that we can incorporate the uncertainty in the observations.

Now, after this thing we have to understand what type of mathematical functions can be used that can illustrate this type of phenomena. So, now I try to introduce here Bernoulli experiment, this Bernoulli experiment is a random experiment in which the outcome of the experiment can be classified in one of the two mutually exclusive and exhaustive ways like as success or failure male or female life or death defective, non-defective etc.

(Refer Slide Time: 06:50)

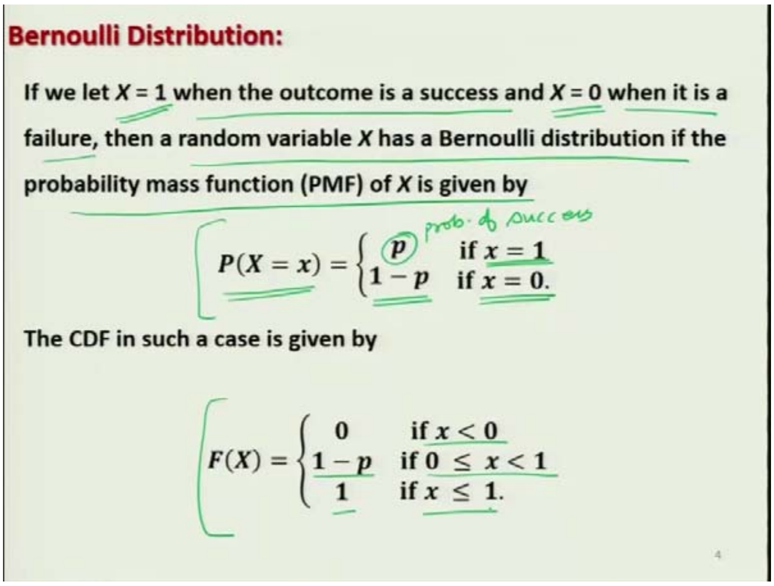


So, this but Bernoulli experiments will give us an idea about the Bernoulli distribution. So, a Bernoulli distribution is useful when there are only two possible outcomes and our interest lies in any of the two outcomes, for example there is a success or the failure, whether a customer buys a certain product or not, there is success or failure or the newly born baby is male or female.

The name of this Bernoulli distribution is after the Swiss mathematician James Bernoulli, just for your information and in general these outcomes are usually indicated by the values 0 and 1. Well, we already have seen that whatever are the outcome that we are getting, they have to be converted into some numerical value and that is what we had discussed when we initiated the discussion on the random variable.

So, suppose that there is a trial, there is an experiment whose outcome can be classified as either as a success or a failure. So, for example now, success can be given the value 1 failure can be given the value 0 or even vice versa also.

(Refer Slide Time: 08:11)

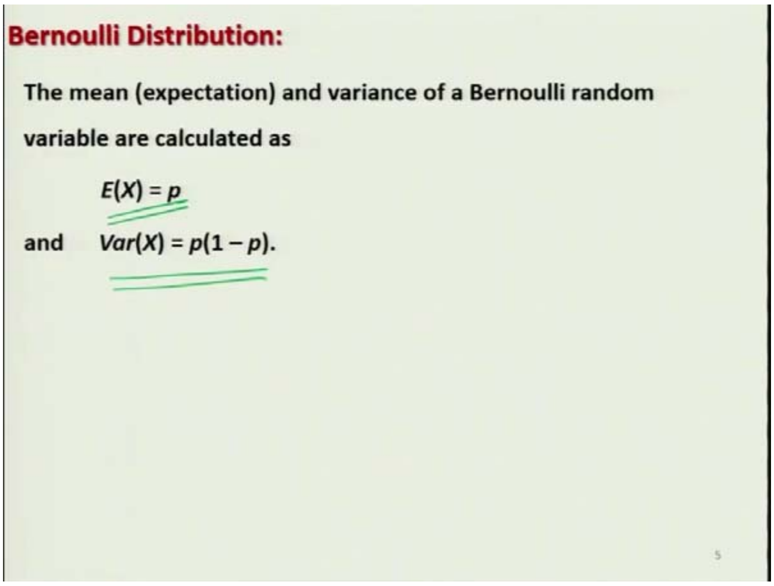


So, now in case if we assume that X equal to 1, there is a random variable X, which takes value when 1 when the outcome is up to success and the random variable X takes value 0, when the outcome is a failure. One thing I would say once again that it is not necessary always to take X equal to 1 as success and 0 as failure. You can also take X equal to 0 as success and X equal to 1 as failure, it depends on you how do you try to define the this random variable and the associated values, but means psychologically it is more easy to understand that X equal to 1 is indicating the success.

So, in this case the random variable X has a Bernoulli distribution if the probability mass function of X is given by this mathematical function and if you try to say this is a very simple thing, you are trying to say that probability that X equal to X when X equal to 1, that is success, it takes the value P which is the probability of success. So, p is the probability of success and in case if x equal to 0, then we know that the probability of success plus probability of failure is equal to 1, because both these events are disjoint.

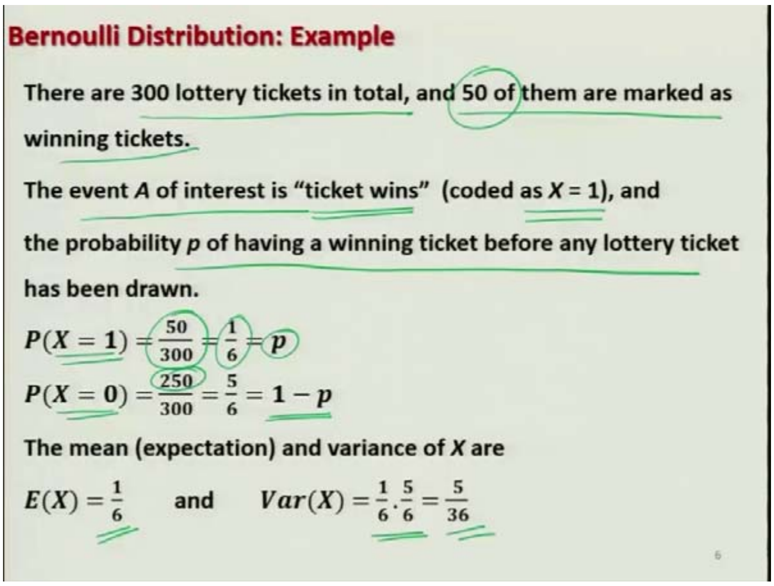
So, this probability that x equal to 0 will simply be 1 minus p and although I am not giving you here the proof but the cumulative distribution function in such a case is given by this function here capital F(X) which is equal to 0, if x is less than 0, it takes value 1 minus p if x the line between 0 and 1 and it takes value 1 if x is equal less than or equal to 1.

(Refer Slide Time: 10:07)



And in case if you try to find out the mean and variance of a Bernoulli random variable, then expected value of X will come out to be p and variance of X will come out to p into 1 minus p, there is no issue. Well, I mean, so you can ask me at this stage that this p and that is the probability of success, how it is going to be known to us. So, in practice, it is unknown to us and then we have different ways to estimate it on the basis of a random sample that we are going to discuss in the forthcoming lectures.

(Refer Slide Time: 10:42)



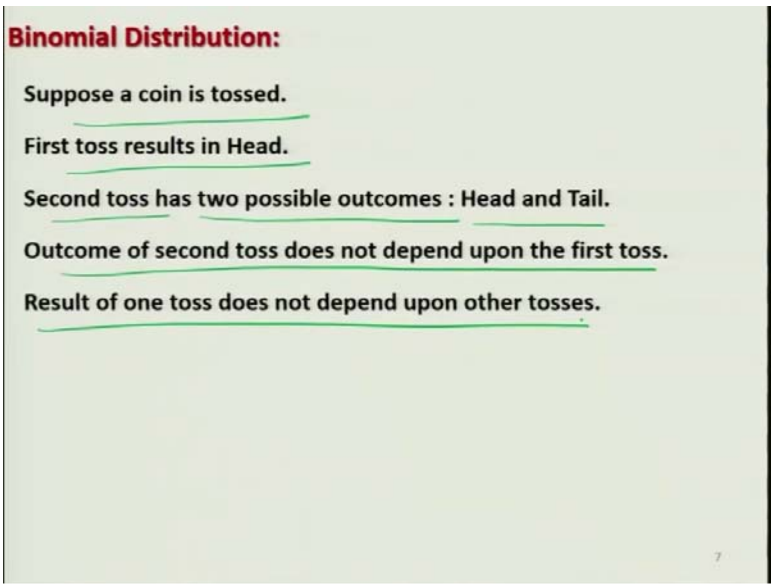
Now, look at we try to take a very simple example, to give you some idea that under what type of condition these types of properties can be used. So, suppose there are 300 lottery tickets and if 50 of them are marked as winning ticket, that means if anybody gets one ticket out of those 50 tickets, it will be a win.

So, suppose we define an event A as ticket wins and this is coded as X equal to 1 that means X equal to 1 is going to indicate that the ticket is coming from these 50 winning tickets and the probability is p of having a winning ticket before any lottery ticket has been drawn means somebody has to buy the ticket before the draw.

So, in this case if somebody wants to compute the probability of winning the lottery or probability of losing the lottery, that can be obtained as probability of X equal to 1 which is the probability of success, probability of getting a winning ticket that will be 50 upon 300, that will be 1 upon 6 and this 1 upon 6 is actually here p and the probability of failure that mean the probability of getting a ticket which does not win that will be probability X equal to 0 that will be 300 minus 50, which is 250 divided by 300 and this will be 5 upon 6, which is actually the value of 1 minus p.

So, the mean and variance in such a case will be given by expected value of X is equal to p which is equal to here 1 upon 6 and variance of X is going to be $p(1 - p)$ which is 1 upon 6 into 5 by 6 which is 5 by 36. So, this is how you can characterize the entire phenomenon once you have got the probability of success and the average value variability, possibly you will get a fair idea about that phenomenon.

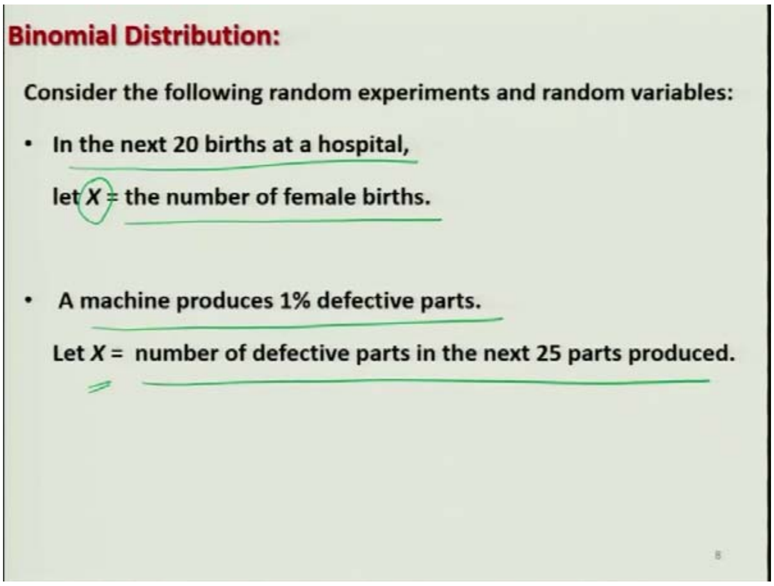
(Refer Slide Time: 12:51)



Now, how these things can be done in R software and, I mean these probabilities, you do not have to compute manually, but how to get them in R software case, this I will try to explain you after I consider the binomial distribution, because this Bernoulli and binomial they are very closely related. So, now we consider the binomial distribution and suppose a coin is tossed and suppose the first toss results in head. Now, what will happen to the second toss? The coin toss has two possible outcomes, head and tail.

Now what is going to be the outcome that you do not know unless and until the experiment is completed and in this case, if you try to see one very important characteristic that the outcome of the second toss does not depend upon the outcome of the first toss and the same story is going to continue if you try to continue to toss the coin, the outcomes of the given toss do not depend on the outcomes of the earlier tosses. So, in general, I can say that the result of one toss does not depend upon other tosses.

(Refer Slide Time: 14:20)

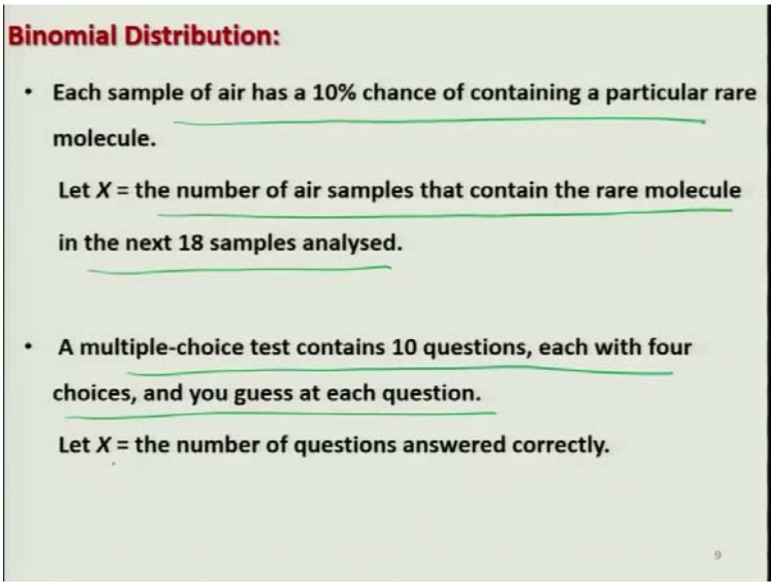


So, suppose, now we try to consider an experiment and we try to define the random variables, which are going to ultimately give you an idea about the binomial distribution or what type of probabilities can be considered under the binomial distribution. So, consider the following random experimental and the random variables. Say, for example, in the next 20 births at a hospital, we want to know what is the number of female birth that how many girls are born and let this number be actually here X.

So, if you try to see whenever a baby is born, there are two possibilities either the baby is a boy or a girl. So, that means don't you think that it is looking like that that every outcome is following a Bernoulli distribution, it has only two possible outcomes and suppose if I take 1 more example that a machine produces 1 percent defective parts. So, suppose the capital X is the random variable which is indicating that number of defective parts in the next 25 parts produced.

So, every time in both the cases, the outcome is going to be governed by the Bernoulli distribution or the Bernoulli trials. So, all the outcomes are the Bernoulli trials and if you try to see we are interested here in a certain number of the occurrence or non-occurrence of the events.

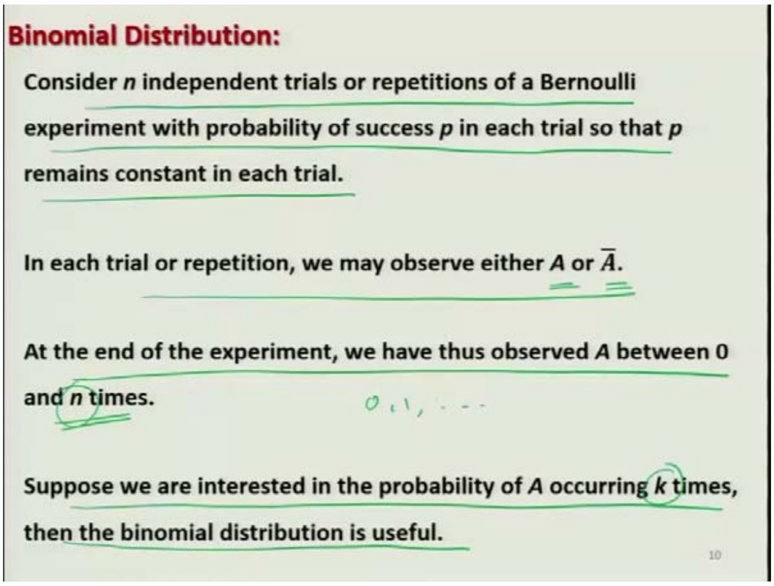
(Refer Slide Time: 16:00)



And in case if I try to convince you and I try to take two more examples, suppose any sample of here has a 10 percent chance of containing a particular type of rare molecule. So, now let X be the number air sample that contained a rare molecule in the next 18 analysed samples. So, once again means, you can see the molecule is absent or present that is sort of a Bernoulli trial and then you are interested in the number, number of times the event is occurring.

And similarly, in a multiple choice tests, suppose it 10 questions and each question has 4 alternatives, 4 choices and you have to guess each question you do not know the answer, often you have the question, then in every question, there are two possibilities, whether your answer is correct or incorrect. So, in case if you consider here a random variable, there are number of questions that are answered correctly, then you are trying to find out the probability of certain event that has occurred.

(Refer Slide Time: 17:17)



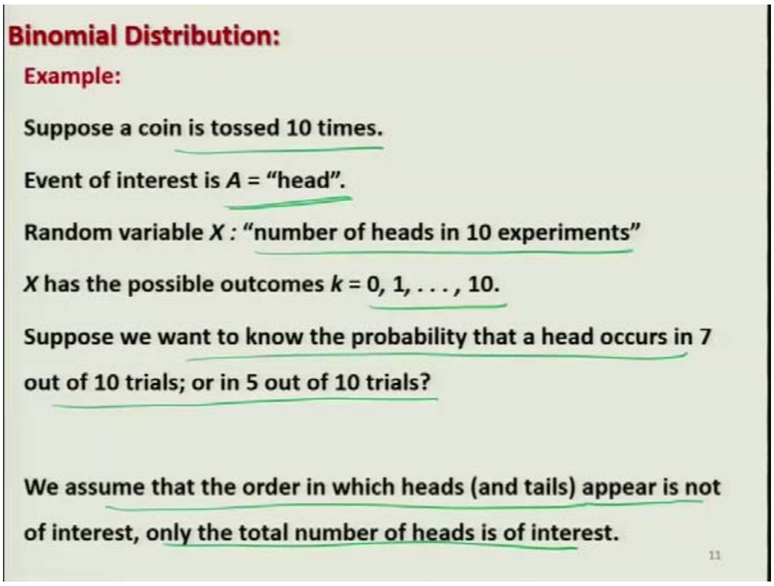
So, in all these cases, you can see that you are trying to repeat the Bernoulli trials. So, now we try to extend it and we try to make it more general and we try to consider a small n number of independent trials or repetitions of a Bernoulli experiment with probability of success p in each trail.

So, obviously this probability p remains constant in each of the trial. So, now this depends on the objective of the experiment, that in each trial or repetition, we can observe either the event A or the event A compliment, for example either we are interested in the outcome of the newly born babies are male or if compliment is female, in case if you are interested in the number of successes, then its compliment is going to be number of failures.

So, at the end of the experiment, we have this observed A between 0 and small n times, because every event is resulting into 0 and 1 and you are trying to repeat the experiment small n number of times, so there will be something like 0111 and so on. But the maximum number of times one can occur in all the trials will be that every trial is having only 1. So, this number is going to be a small n and suppose, we are interested in the probability of A occurring k number of times, then in under such type of cases, the binomial distribution is used.

For example, if you try to take here means any example, in this example also there are 10 possible question. So, the success and failure can be say indicated by 0 and say 1 at the most what will happen that no solution is correct. So, X is going to take the value 0 and if all the solutions are correct, then X is going to take the value say 10. So, that is what I was trying to explain you here.

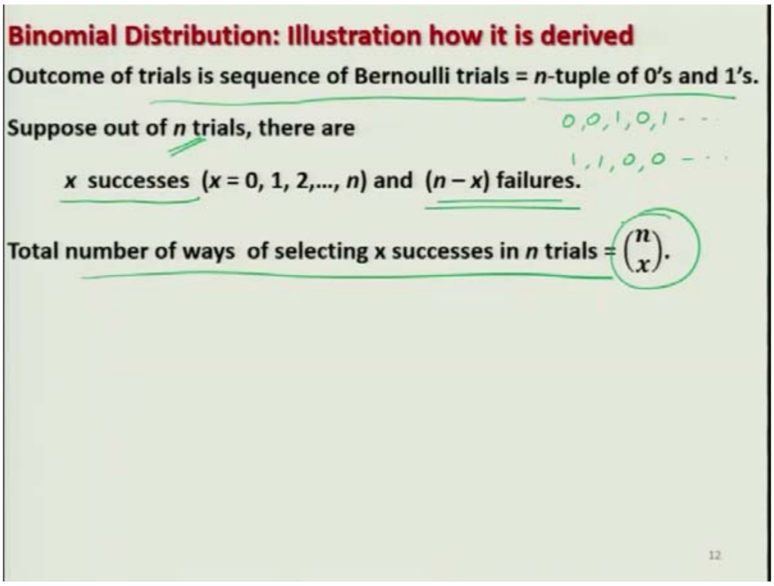
(Refer Slide Time: 19:47)



So, now suppose a coin is tossed suppose 10 times and the event of interest of interest here is the number of heads. So, the random variable in this case is going to be number of heads in the 10 experiments and X is going to take the possible values as say k equal to 0, 1, 2 up to here 10 and now suppose we want to know the probability that head occurs in 7 out of 10 trials or say in 5 out of 10 trials.

So, under this type of condition, we can use the binomial distribution to compute the probabilities and we assume that the order in which the heads or the tails appear is not of interest, only the total number of heads is of interest. So, we are not bothered that what is the individual outcome, whether success or failure, but we are interested only in the number.

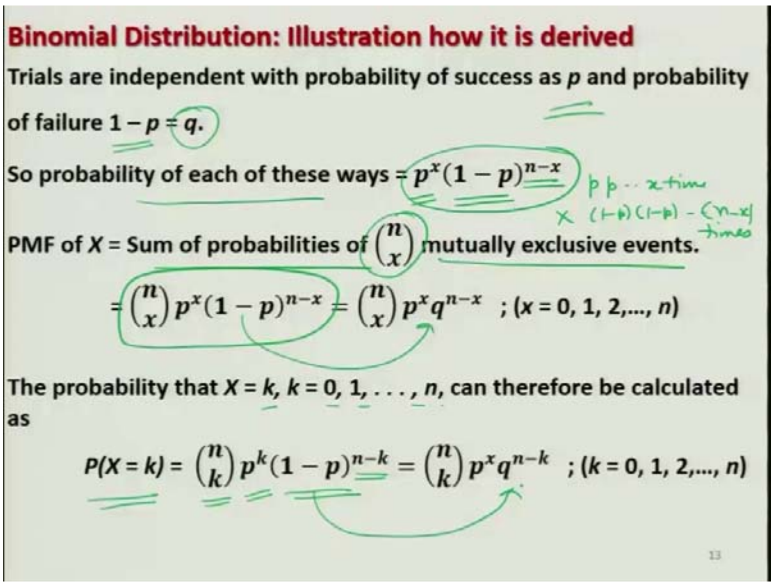
(Refer Slide Time: 20:42)



So, you can see here the outcome of these types of trial is actually a sequence of Bernoulli trials and the outcome is going to look like as n-tuple of 0s and 1 like 0, 0, 1, 0, 1 up to here n times, then the next outcome can be say 1, 1, 0, 0 and so on.

So, out of these smaller n number of trials, means if you see that there are small x successes, that means x is going to take the value say 0, 1 here up to n and depending on the success, the remaining say n minus x events will indicate the complement or the failure. So, obviously now, you can see the total number of ways of selecting, x successes in n trials is going to be n choose x obviously, that you already have done.

(Refer Slide Time: 21:43)



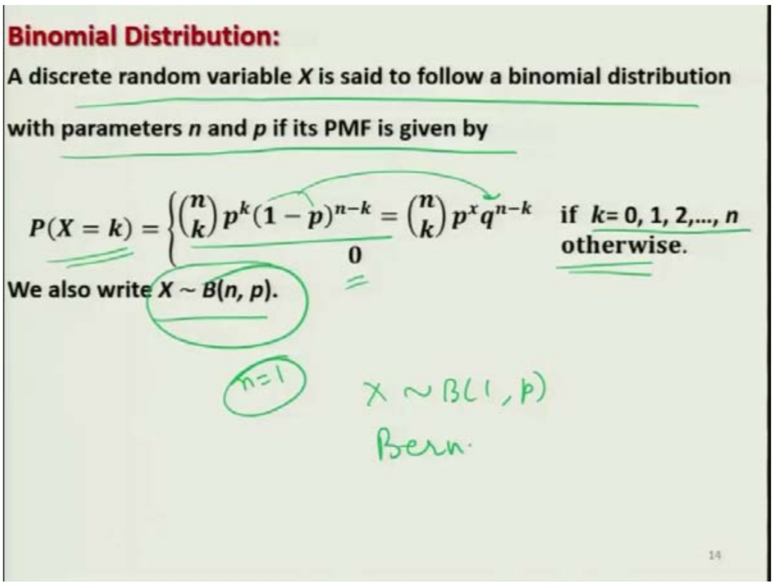
Now, the trials are independent and the probabilities of success are very small p and the probability of failure here is 1 minus p that we are trying to indicate by here small q. So, because all these events are independent, so you know that if there are two events A and B, which are independent, then probability of A intersection B is probability of A into probability of B.

So, the probability of each of these ways will simply be here raise to power of here x into 1 minus p raise to power of here and minus x. So, this is indicating here there are x times successes, so it will go p, p, p, p, p x times and multiplied by say n minus x failures. So, 1 minus p, 1 minus p, up to here n minus x times.

So, now in case if I tried to find out the total probability that will give us the probability mass function of x that is going to be some probabilities of $\binom{n}{x}$ mutually exclusive events and if you try to sum this p raise to power of x into 1 minus p raise to power of x into 1 minus p raise to power n minus, $\binom{n}{x}$ time you get here a function like n choose p raise to the power here x 1 minus p raise to power of n minus x and this 1 minus p can be replaced by here q.

So, now I can say the probability that the random variable x takes the value of k, where k is equal to 0, 1 up to here and can be can therefore be calculated as probability that x equal to k is equal to $\binom{n}{k}p^k(1-p)^{n-k}$ or that can also be written by replacing the 1 minus p by q.

(Refer Slide Time: 23:38)



So, if you try to see what I have done, I have just considered a particular type of probability and I have illustrated here how you can compute the probability and ultimately you have now converted you can see here in the last part, you have converted into a very general statement how to compute the probability and now I can just say that this is going to generate the probabilities of these types of events in which you are interested in finding out the probability of success or failure out of the number of trials, where the trials are going to be been Bernoulli trials.

So in this case, because this was, because at this moment we are beginning so I have given you here the complete idea that this probability mass function or equivalently, the probability density function, they are not coming from the sky, we are simply trying to compute the probability of certain type of events and finally we are trying to write them in a nice mathematical form, which can be used directly without any problem.

Because I personally feel that sometimes I see that students are confused that how the probability functions are coming into picture and they say that, there are so many probability function and how are they going to know that where to use how to choose, when to use etc. So, that is what I am trying to cover in the beginning and as soon as we feel that we all are now mature enough, I will try to get used to the discussion.

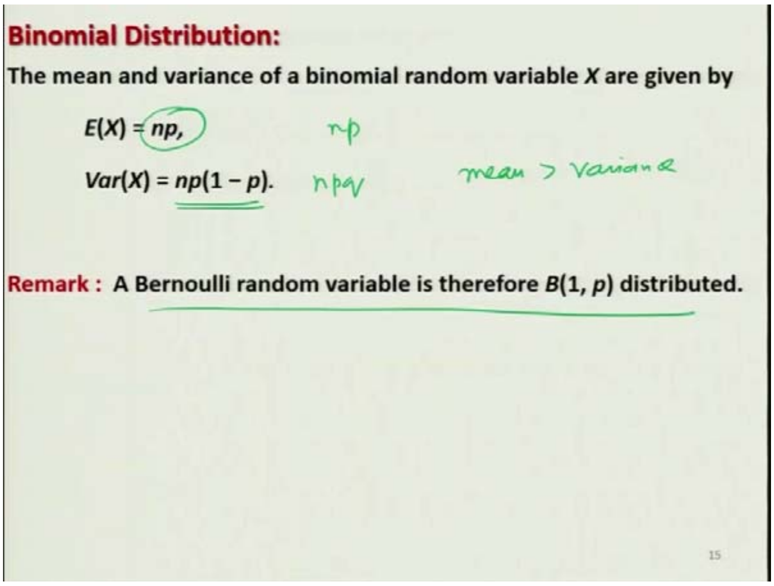
After that as soon as I say that, let there be a random variable which is following this probability mass function, you can very well assume that this is coming from some through proper methodology proper concepts, although I may not be able to give you the complete proof or derivation of all the probability mass function or the probability density function.

So, that is why in this case I have given you detail information, but it may not be possible for me every time to give you such detailed information, but if you want you can simply pick up a book and then try to understand them. So, now I can define the binomial probability mass function.

So, our discrete random variable X is set to follow a binomial distribution with parameters small n and small p if its probability mass function is given by $P(X=k)=\begin{cases}\binom{n}{k}p^k(1-p)^{n-k}=\binom{n}{k}p^x q^{n-k} & \text{if } k=0,1,2,\ldots,n\\ 0 & \text{otherwise}\end{cases}$, like and if you want to use here 1 minus p here as q, you can simply replace it and we indicate symbolically that X follows the binomial distribution with the parameters n and p by writing this x equivalence and inside the parenthesis is small n, small p.

So, that is a standard language, a standard symbolic language in which we try to represent this thing and if you try to take here n equal to here 1, what do you get? If I write down here x binomial is 1 p, do not you think that you are going to get here the Bernoulli distribution? Think about it. Answer is yes.

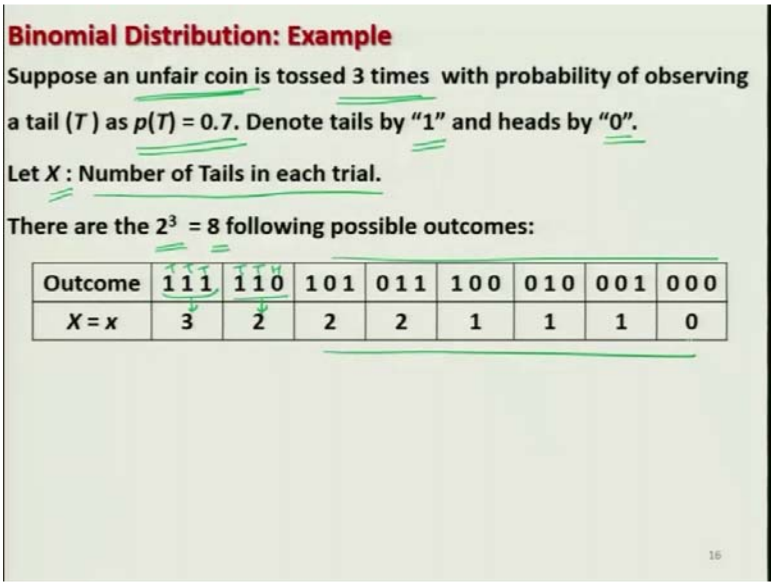
(Refer Slide Time: 27:19)



So, in the case of binomial distribution the mean of X can be found to be as np and the variance of X can be found as np(1 – p) or we write down here say npq. So, the mean here is np and variance here is npq and you can see here that p and q they are always going to live between 0 and 1. So, you can see here in this case, mean is greater than variance.

So, and if you want to have the results for the Bernoulli distribution, you can simply substitute here n equal to 1 and a Bernoulli random variable is therefore, binomial 1, p distributed and that is the thing that we are going to use in the software also.

(Refer Slide Time: 28:11)

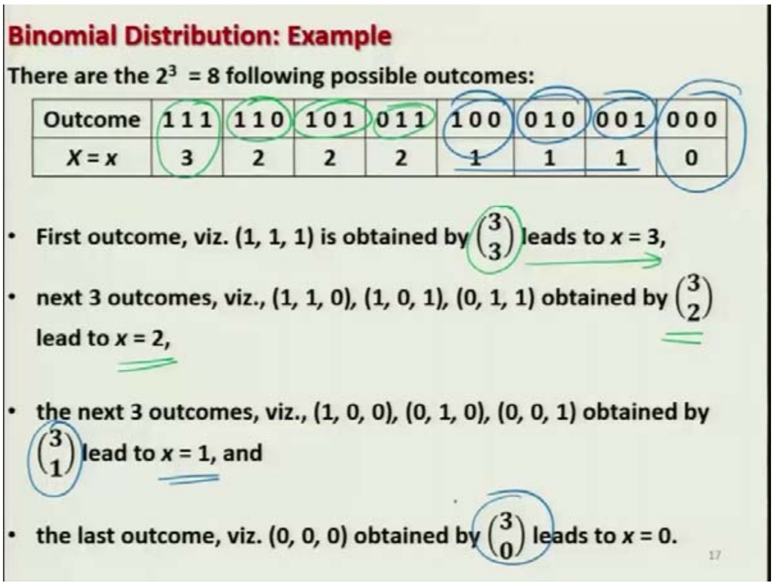


So, can we try to take here an example and try to show you and the same thing I will try to do in the R software also. So, suppose there is a coin which is unfair in the probability of head and tail are not exactly the same. But suppose the probability of tail is 0.7 and suppose the coin toss 3 times.

So, now the outcomes of that tail are going to be indicated by 1 and the outcome of heads are going to be indicated by 0. So, now the unfair coin is tossed 3 times. So, we try to define here the random variable as the number of tails in each trial. So, following are the outcomes why because the coin is tossed 3 times, so every time you can have different types of outcomes.

So, here I am trying to illustrate all possible outcomes which are actually 2 cube that is 8. So, you can get here tail, tail, tail that you are trying to indicate by 1, 1, 0 you are getting tail, tail and head that is being indicated by 1, 1, 0 and similarly, I try to write down here are all the possible outcomes and then I tried to define here the value of capital X. So, in this case the number of details are 3. So, this really comes out to be here 3. In the second case there are 2 tails. So, there is a capital X takes the value say 2 and so on. So, these are the values of the number of tails that can be obtained in every trial.

(Refer Slide Time: 29:53)

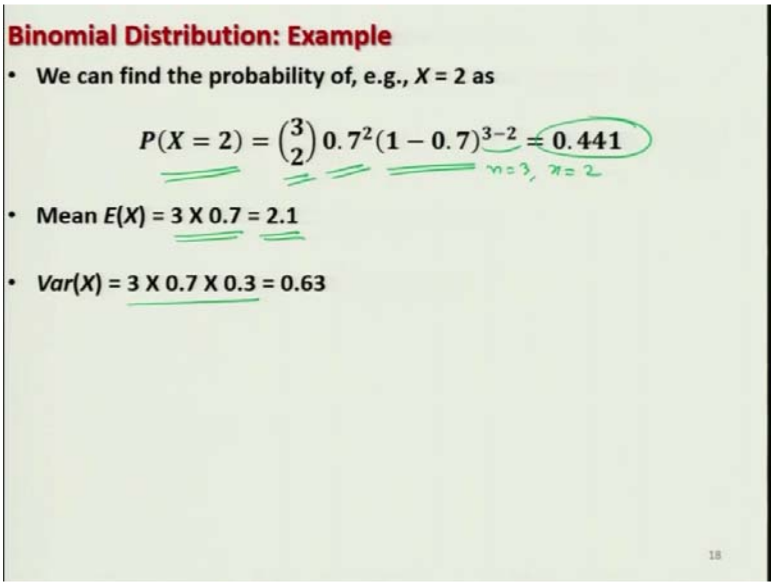


And now, let me try to give you some details of this computation. For example, if you can see here in the first outcome, there are 3 possible tails, there are 3 tails. So, this number 1, 1, 1 is obtained by the competition $\binom{3}{3}$ that is leading to x equal to 3 you are not observing it, but you are trying to say that in the 3 trials, there are 3 tails.

Now, similarly if you try to look at here, here and here, in this cases what is happening that the total number of cases in which you can obtain 2 tails in a sequence of 3 trials that can be obtained by say $\binom{3}{2}$ and that is a leading to x equal to 2 and similarly, in the these 3 trials see here 1, 0, 0, 0, 1, 0, 0, 0, 1 the values here are 1.

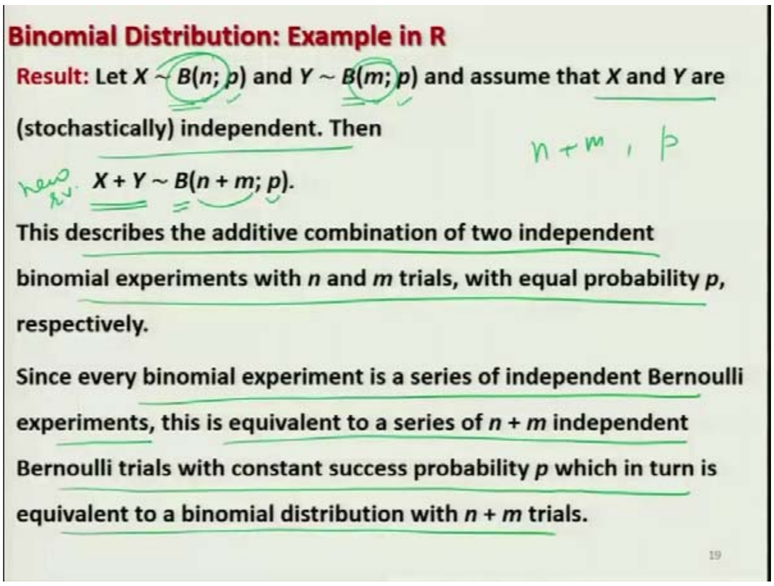
So, in these 3 cases, the total amount of tails in the 3 possible trial they are obtained by $\binom{3}{1}$ and that is leading to x equal to 1 and similarly, in the last outcome say 0, 0, 0 means you are trying to obtain the 0 number of tails in the 3 tails and that leads to x equal to 0.

(Refer Slide Time: 31:27)



So, now, in this case if you want to compute the probability for example, X equal to 2 that can be simply obtained by $\binom{3}{2}$ into the probability of getting a tail which is 0.7 into 1 minus 0.7 and minus x. So, n here is 3 and x here is 2 and this can be computed as 0.441 and the mean is going to be 3 into p that is 3 into 0.7 2.1 and the variance is going to be 3 into 0.7 into 0.3 which is 0.63. Well, I have taken this example and I have computed it here mathematically to convince you how these values are obtained and I will try to do the same thing in the R console also.

(Refer Slide Time: 32:19)



But, before I leave, let me try to give some important results. So, one important result in the binomial distribution is that, if there are two stochastically independent variable X and Y and suppose, both are follow following a binomial distribution. So, X is following a binomial distribution with parameters is small n and small p and Y is following a small m and small p. The probability in both the random variable remains the same as p, but only the numbers m and n are different.

In this case, if you want to find out the distribution of X plus Y, that is going to be a new random variable, then the X plus Y will have a binomial distribution, whose parameters will be n plus m and the probability will remain the same as p. So, this is actually the additive property and this describes the additive combination of two independent binomial experiment with an n, m number of trials with equal probability say p and this can be extended also.

And if you try to see this is not really difficult to understand although there are rigorous proof, but here you can also see that these results can be viewed as follows, for example every binomial experiment is a series of independent Bernoulli experiments and this is equivalent to a series of n plus m independent Bernoulli trials with constant success probability p, which in turn is equivalent to a binomial distribution with n plus m trials.

So, in this case you have a small n of Bernoulli trials that is giving you in binomial, in the case of Y there is small n number of Bernoulli trials that are giving you binomial and if you can consider that instead of n or m you have n plus m number of Bernoulli trials with the probability of occurrence as p that remain constant in all the trials then you will have a binomial m plus n and trials n to a probability p.

So, now we come to an end to this lecture and you can see here I have tried my best to give you the detailed development of the Bernoulli and binomial distribution and I also tried to give you some examples, so that I can show you that under what type of condition they can be used and there are some statistical properties also there is a long list, but I have just given you here one is statistical property and actually this is useful in data sciences.

So, in fact, in every probability function, there will be lots of statistical properties of those distribution, but I will try to give you here only those property which I personally feel that they are going to be useful when you are trying to deal with data science and you are trying to handle the problems in the data science and they are useful from the computational point of view for example, in this case you can see that if you have got two experiments which had been conducted at two different places, but you see that, both are going to follow the binomial distribution and the only thing is, they say there are number of trials are differing, but the probability of success is the same at both places. For example, if you try to see, the probability of getting a head in say New Delhi and probability of getting a head in say, Mumbai they are the same. So, somebody has made to say this 20 observation in Delhi and 40 observation in Mumbai you can club it together and the new random variable will follow the binomial distribution with say the n is equal to 20 plus 40, that is 60 and with the same probability of success p.

So, this is what I want and this is what exactly is in my mind while I am trying to prepare the slides and giving you the lecture, after some time it may change, I do not know, but this is what I have done. So, now, this is your turn, I will say simply try to pick up a book and try to read the chapter on Bernoulli and binomial distributions, you will see there are many more things and for me it is not possible to cover each and everything, but the more you learn, the better you are.

So, you try to take some exercises from the assignment from the book and try to solve them. The main thing is this you have to understand why are you trying to employ here the binomial distribution or Bernoulli, why not Poisson, geometric etc. that we are going to discuss in the later lectures and that will help you using these tools in their real applications.

By looking at the nature of the data, you can have a fair idea that what type of probability mass function of probability function has to be used, so that you can compute the probabilities correctly. Once you can compute the probabilities correctly, you have solved the problem. So, you try to practice it and I will see you in the next lecture. Till then, goodbye.