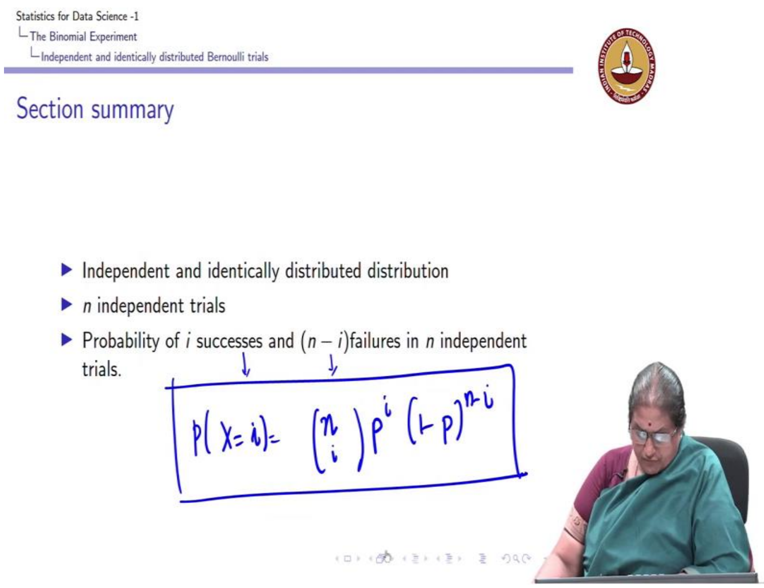
INDIAN INSTITUTE OF TECHNOLOGY MADRAS • सिद्धिर्भवति कर्मजा •

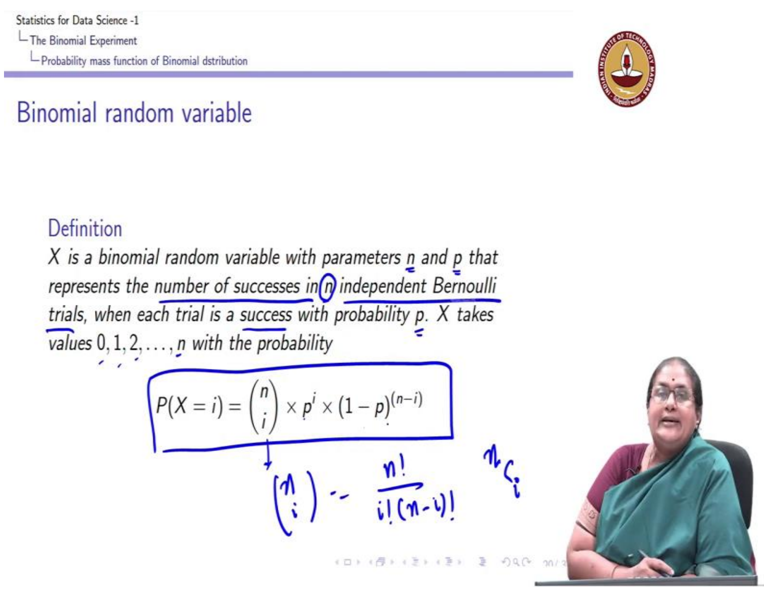
IIT Madras

ONLINE DEGREE

Statistics for Data Science - 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 10.3
Binomial Distribution- Distribution of Binomial Random Variable

(Refer Slide Time: 00:14)

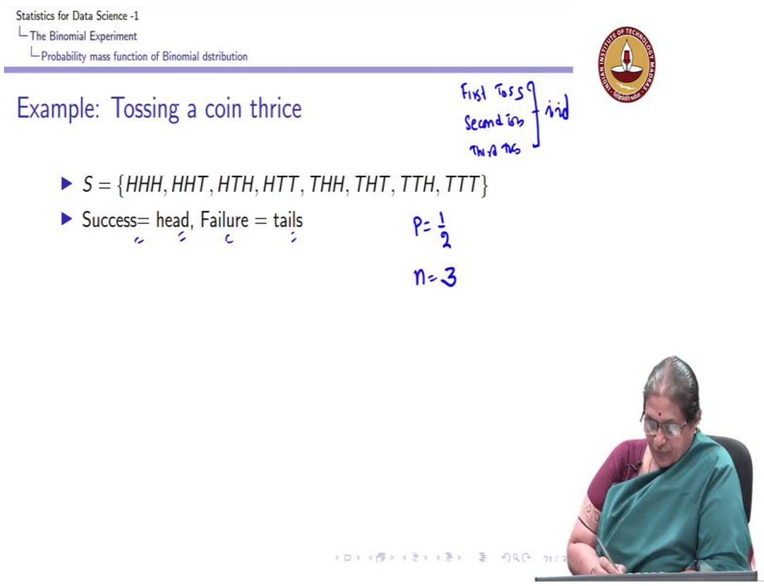




Now, the next thing is, what we focused on is to get hold of the probability distribution of n identically distributed Bernoulli random variables and we defined X as the sum of these n identically or sum of outcomes counted the number of successes and these n independent trials. So, now we are in a position to formally define what is a binomial random variable.

I formally define X as a binomial random variable, what are the parameters of the binomial random variable n and p it represents the number of successes in n independent Bernoulli trials each trial is a success with probability p, X takes the value 0, 1, 2 up to n with this probability probability X equal to i is n choose i this is the binomial coefficient, we know n choose i is n $\frac{n!}{i!(n-i)!}$, some books referred to it as ${}^{n}C_{i}$, $p^{i} \times (1-P)^{(n-i)}$.

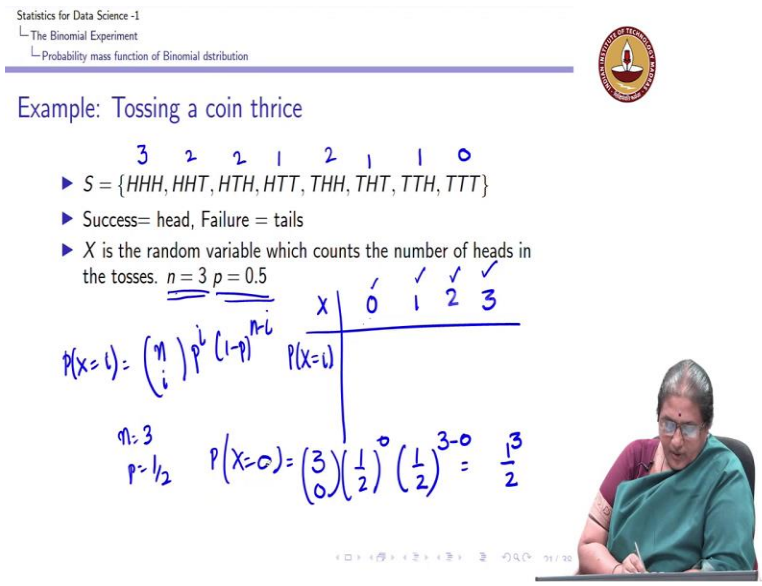
(Refer Slide Time: 1:37)



Now, let us go back to our example of tossing a coin 3 times. Again, remember in our discussion a few minutes back we said if I am tossing a coin the first toss, the second toss and the third toss they are independent and identically distributed that the outcomes are independent they are also identically distributed. So, when I am tossing a coin a Bernoulli trial I have also said a coin toss is a Bernoulli trial with success being the appearance of a head and failure being the appearance of a tail. I am talking about the fair coin here.

So, the probability of success is equal to the probability of failure which I can take S equal to half. Now, if I am tossing a coin thrice I am repeating this Bernoulli experiment 3 times. So, my n can be 3.

(Refer Slide Time: 2:44)

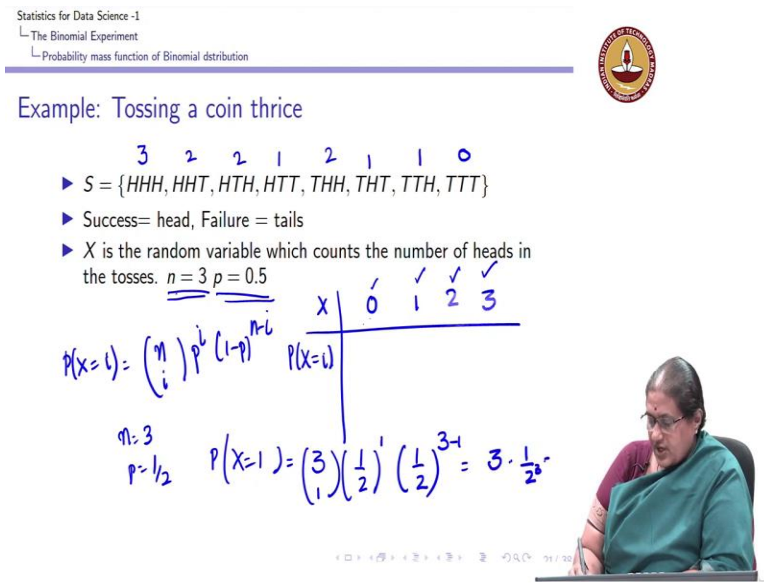


So tossing a coin thrice can be viewed as a binomial experiment with parameter n equal to 3 and P equal to half. Now, what are the values as random variable can take, we see again the random variable X takes values 0, 1, 2 up to n, here my n is 3 so it takes the value 0, 1, 2, 3. Now, again, what is X doing, X is counting the number of successes. So, the number of successes here is counting the number of heads. Again, when I count the number of heads I see there are 3 heads here, 2 heads here, 2 heads here, 1 head here, 2 head here, 1 head here, 1 head here and no head here.

So, I can again see that my random variable which is counting the number of successes which is equivalent to counting the number of heads takes the value 0, 1, 2, 3 and since n equal to 3, it is a binomial random variable which takes values 0, 1, 2 and 3. Let us apply the formula to find out what are the probabilities recall $P(X = i) = {}^{n}C_{i}\, p^i (1 - p)^{(n-i)}$, here my n is 3 and p is half.

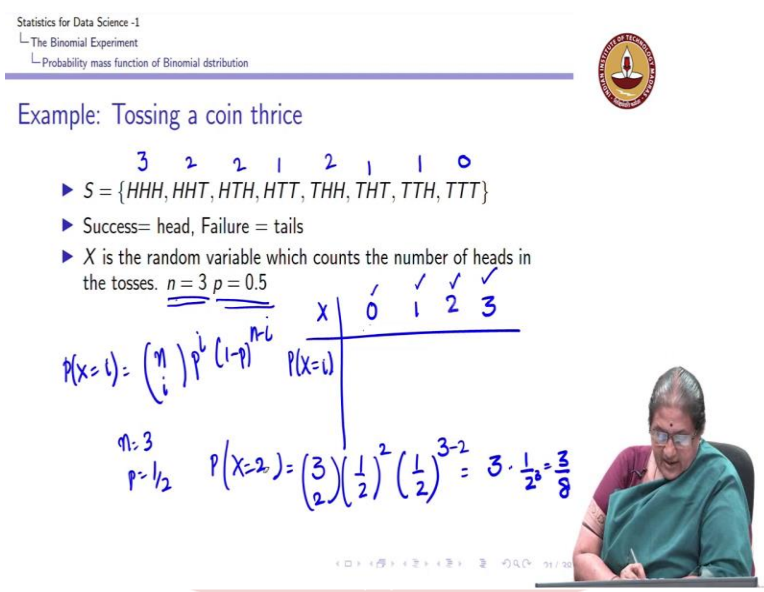
So, what is probability X equal to 0, n is 3, 3 choose 0, p is half half to the power of 0, 1 minus p is again half, 3 minus 0, which is 1 by 2 to the power of 3, which is going to be 1 by 8.

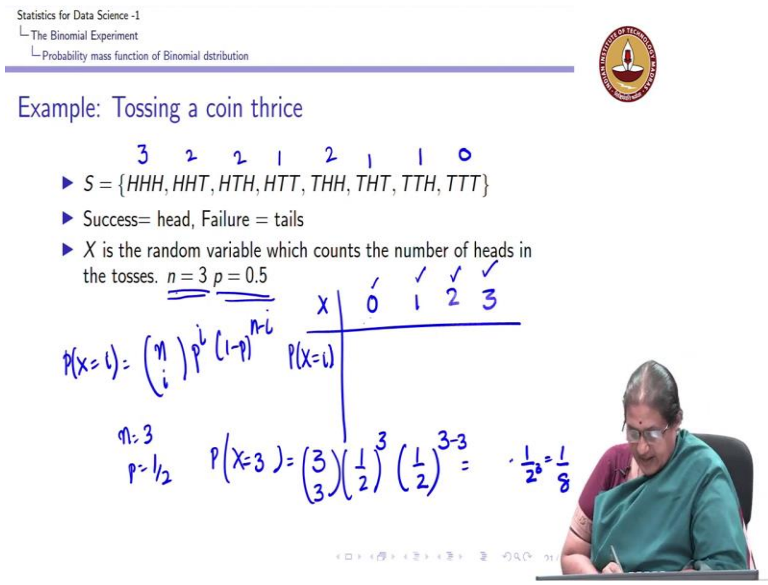
(Refer Slide Time: 4:38)



$P(X = 1)$ is going to be $^3C_1 \left(\frac{1}{2}\right)^1 \times \left(\frac{1}{2}\right)^{(3-1)}$, which is going to give me $3 \times \left(\frac{1}{2}\right)^3$, which is 3 by 8.

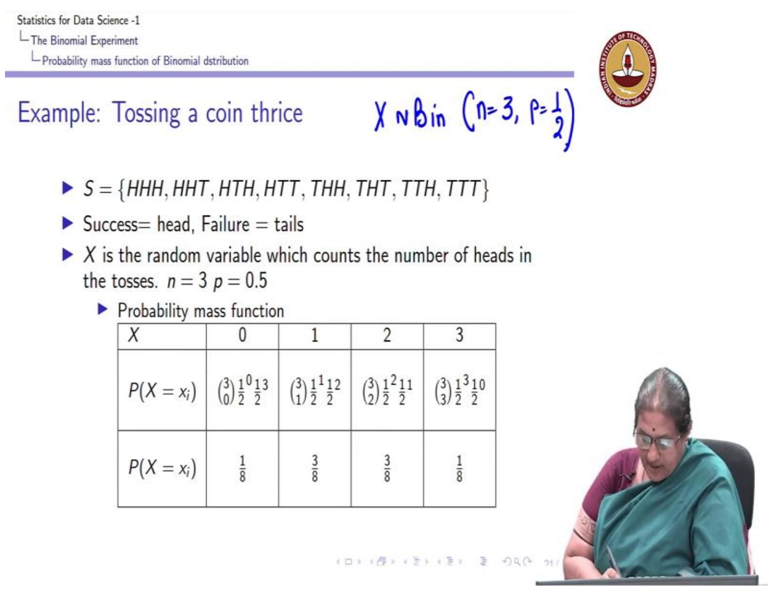
(Refer Slide Time: 4:55)





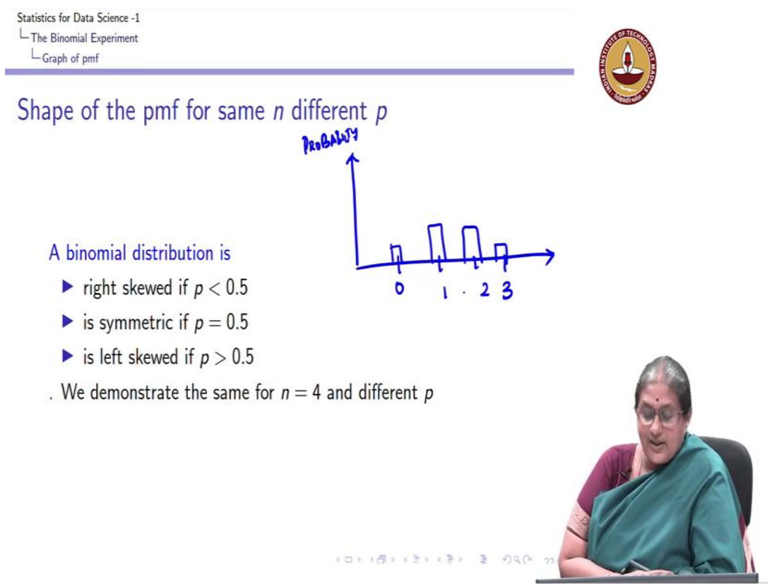
$P(X = 2)$ would be ${}^3C_1 \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^{(3-2)}$. So, I am going to again get a 3 into 1 by 2 to the power of 3, which is again 3 by 8 and $P(X = 3)$ is going to be 3 choose 3, 1 by 2 to the power of 3, this minus 3 which is going to give me 1 by 2 to the power of 3 which is again going to be a 1 by 8.

(Refer Slide Time: 5:27)



And hence I can verify that this is precisely what was the probability mass function of tossing a coin thrice. So, I can view the experiment of tossing the coin thrice and noting the number of heads in my tosses as a binomial experiment with parameter $n = 3$, and $p = \frac{1}{2}$. So, what we have discussed earlier as tossing a coin thrice is in effect a binomial experiment with parameters $n = 3$, and $p = \frac{1}{2}$.

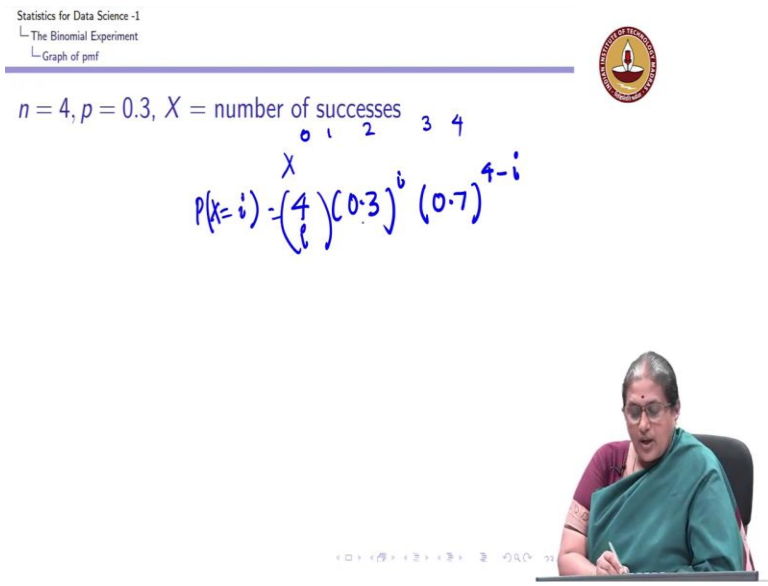
(Refer Slide Time: 06:07)

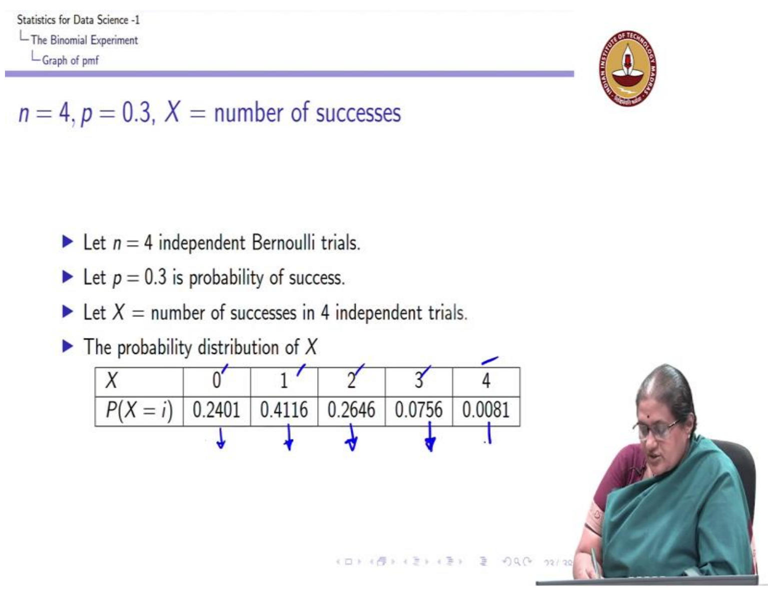


Now, let us go back and look at what would be the shape of this probability mass function. What do we understand by the shape of a probability mass function, recall a probability mass function takes on the X values, the values of X and on the Y values it gives me the probability with which it takes these values. For example, if X takes the value 0, 1, 2 and 3 with probability 1 by 8, 1 by 8, 3 by 8, and 3 by 8, this is the probability mass function of the tosses of the coin 3 times.

So, now let us look at a binomial distribution and see the effect of this binomial distribution on my probability mass functions of the parameters.

(Refer Slide Time: 07:03)

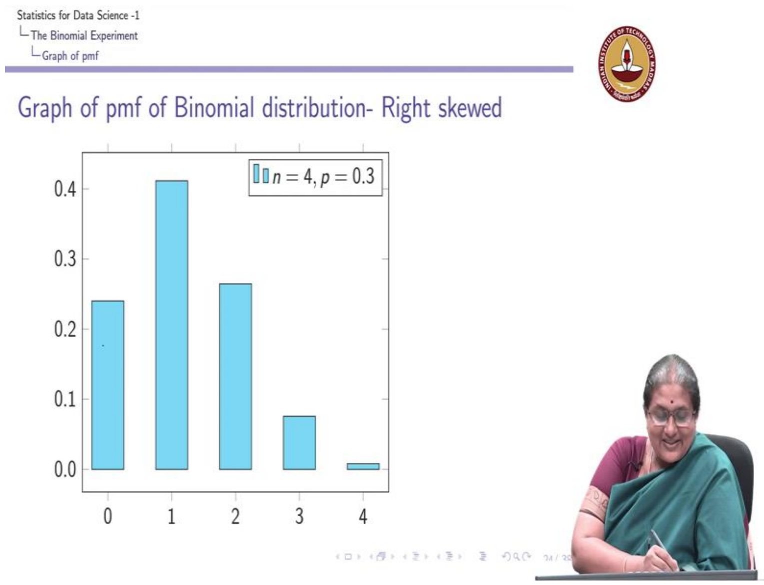




So let us start with $n = 3$, and $p = 0.3$. In other words I am having again 4 Bernoulli trials. So, my X can take the value 0, 1, 2, 3, 4. Again, the probability of X taking the value i is going to be 4 choose i, p equals 0.3 to the power of i and 0.7 to the power of 4 minus I ^4C_i. I am just applying the formula $P(X = i)$ is $^4C_i 0.3^i \times 0.7^{(4-i)}$.

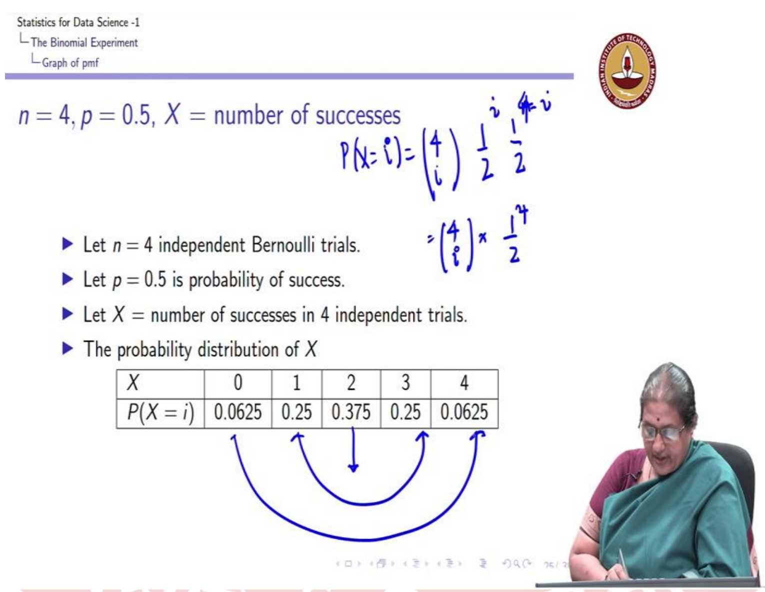
So, if we are going to apply this we can see that the probability or the value X takes 0, 1, 2, 3, 4 with the following probabilities, what we notice in this probabilities is $P(X = 0)$ is 0.24, $P(X = 1)$ is 0.4, $P(X = 2)$ is 0.26 this is 0.07 and point 0.088.

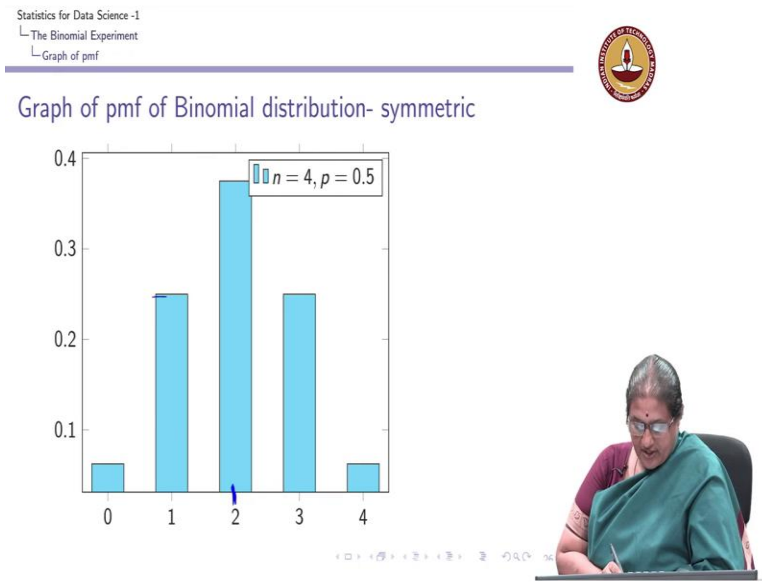
(Refer Slide Time: 8:20)



If I plot it as a graph, you can notice the skewness in the distribution and this graph is called a right skewed graph.

(Refer Slide Time: 08:35)

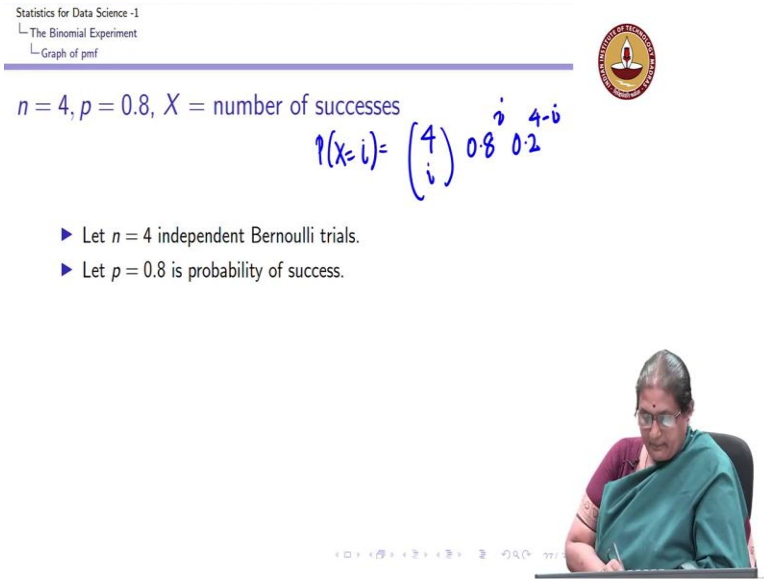


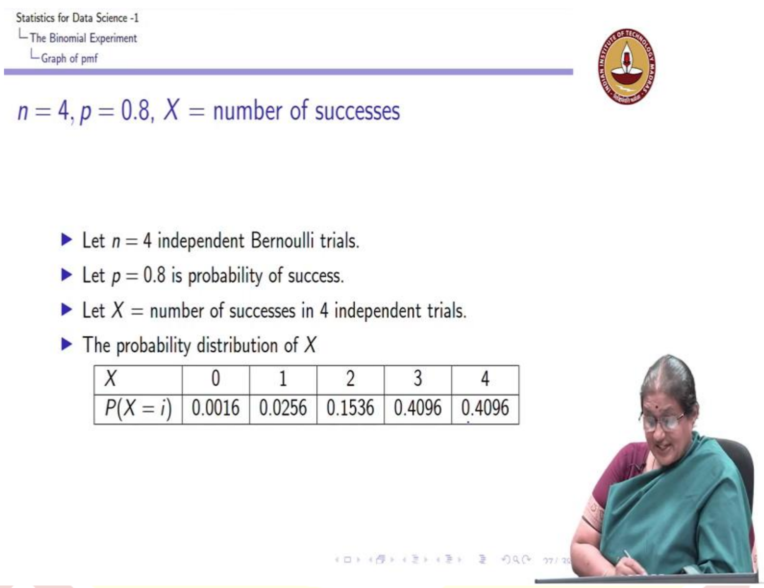


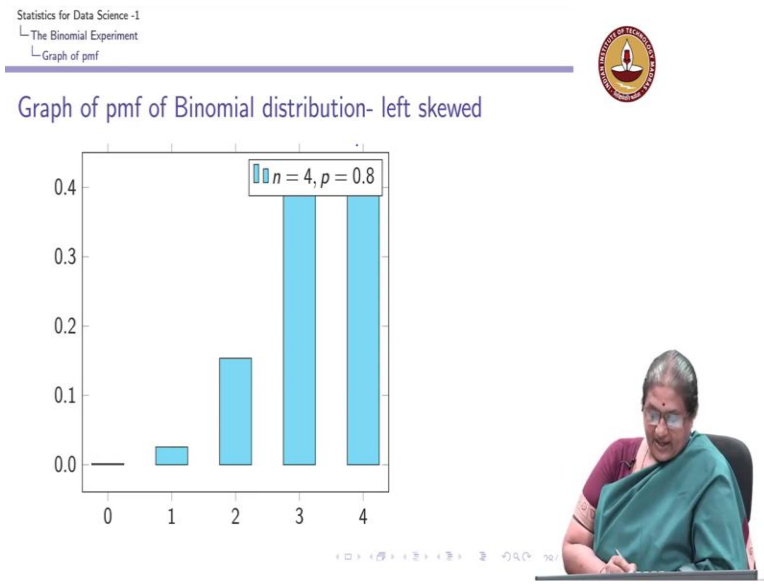
Now, for the same value of n, which is equal to 4, if I take p equal to half recall, p equal to half says that equally likely chance of getting a success and a failure, that is what my p equal to half represents. So, if I have a p equal to half, I notice that distribution again my probability of X taking a value i is going to be $^4C_i 0.5^i \times 0.5^{(4-i)}$ which is going to be $^4C_i 0.5^4$. Because n equal to 4 so it is going to be 4 choose i into 1 by 2 to the power of i.

What you notice here is the highest value occurs where X equal to 2, X equal to 1 and 3 take the same probability X equal to 0 and 4 take the same probability. So, when I plot the probability mass function, you can see that there is a symmetricity around this point X equal to 2 you can see that symmetric distribution which is demonstrated here.

(Refer Slide Time: 10:01)

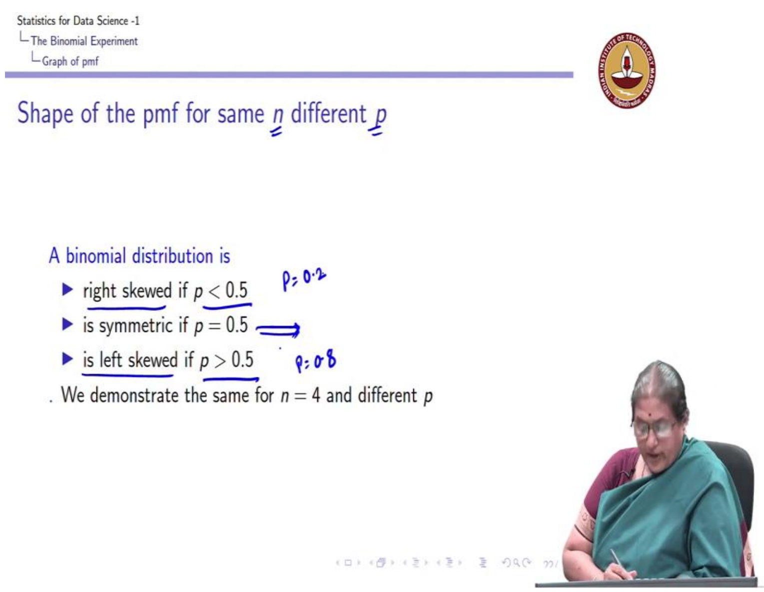






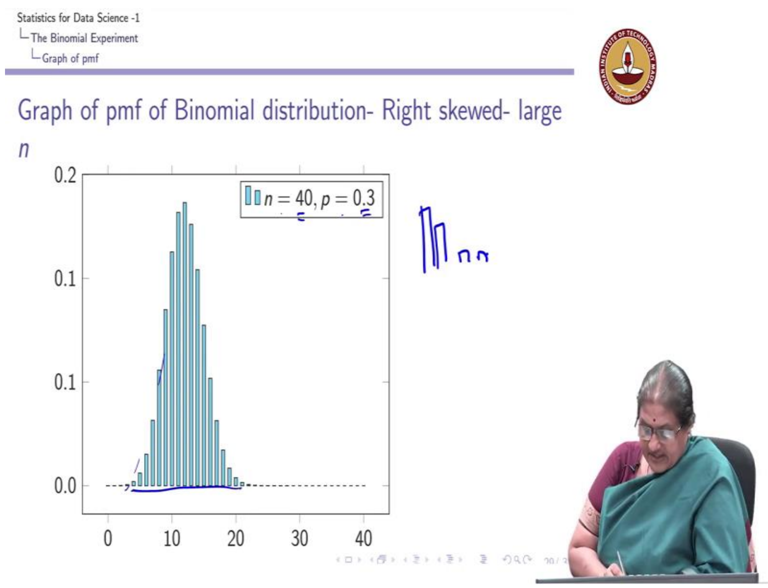
Now, again and I take n equal to 4 and p equal to 0.8, I notice that again probability X equal to i is going to be 4 choose i, 0.8 to the power of i and 0.2 to the power of 4 minus i, this is going to be my probability mass function, the probability distribution is given in this way and when I plot the graph I see there is a left skewedness. Again, I see a skewed distribution.

(Refer Slide Time: 10:51)



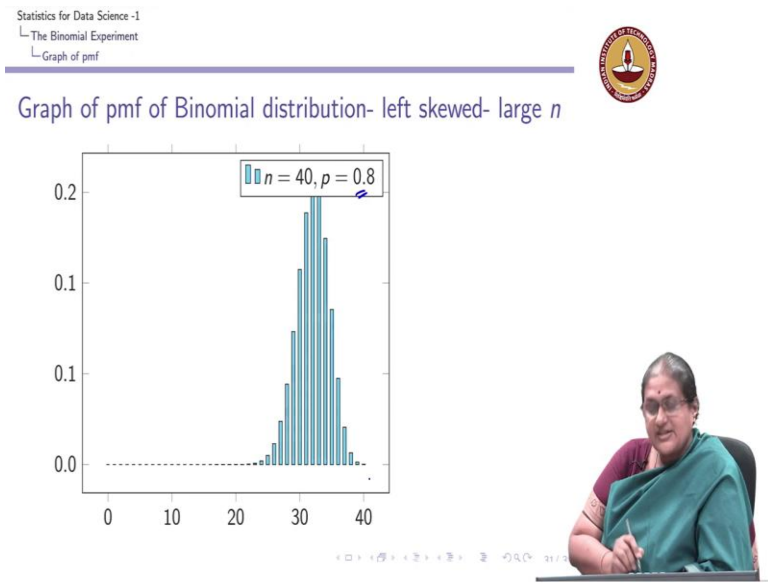
In summary, I can summarise this to say that for the same value of n, here, I have chosen a small value of n, n was equal to 4 and they vary the probability parameter p I call it a right skewed, if $p < 0.5$, left skewed is $p > 0.5$, we observed for $p = 0.3$ and this was $p = 0.8$, 0.2 and 0.8 and we observed symmetricity $p = 0.5$, we demonstrated the case for p equal to 0.3, 0.5 and 0.8 were for a value of n equal to 4. Now, let us look at what happens to the shape of the distribution. When I increase my n earlier my n was equal to 4. Now, I am going to take n equal to 40. So, I am looking at a large value of n.

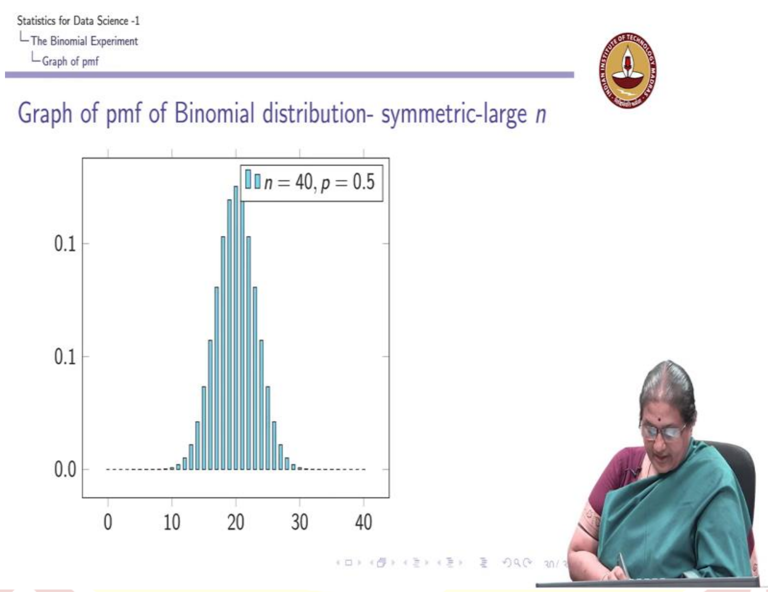
(Refer Slide Time: 11:45)



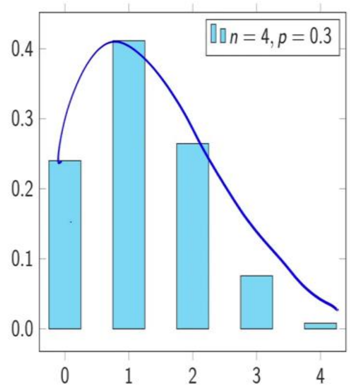
So, what happens when I am taking a large value of n. Remember, when I took n equal to 40 and p equal to 3 I had 0.3 I had a right skewed distribution. But now, you can see that this distribution, what is happening I want you all to notice what is happening for a large value of n, you notice that this when n equal to 40 and p equal to 3, I noticed some sort of a symmetricity it is shifted to the left but there is a sort of symmetricity.

(Refer Slide Time: 12:23)

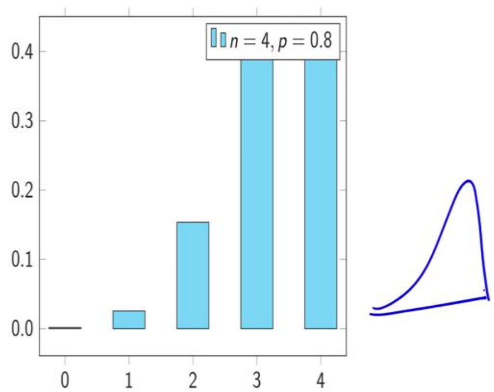


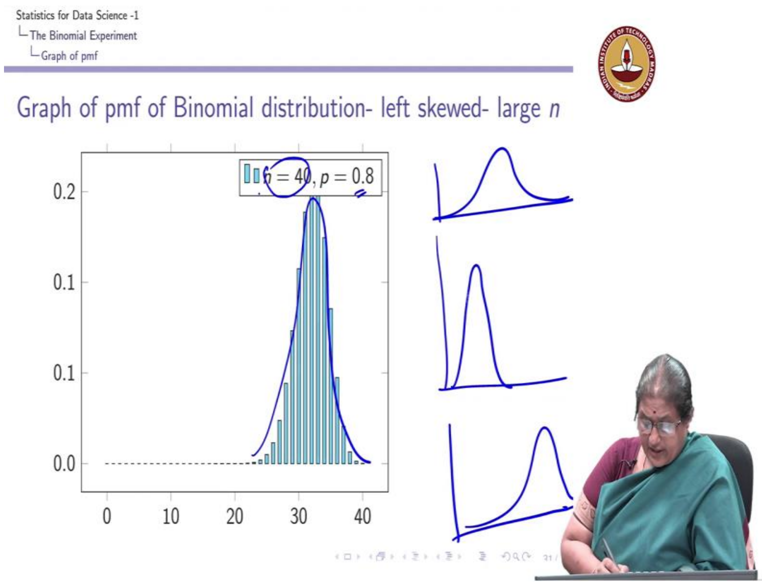


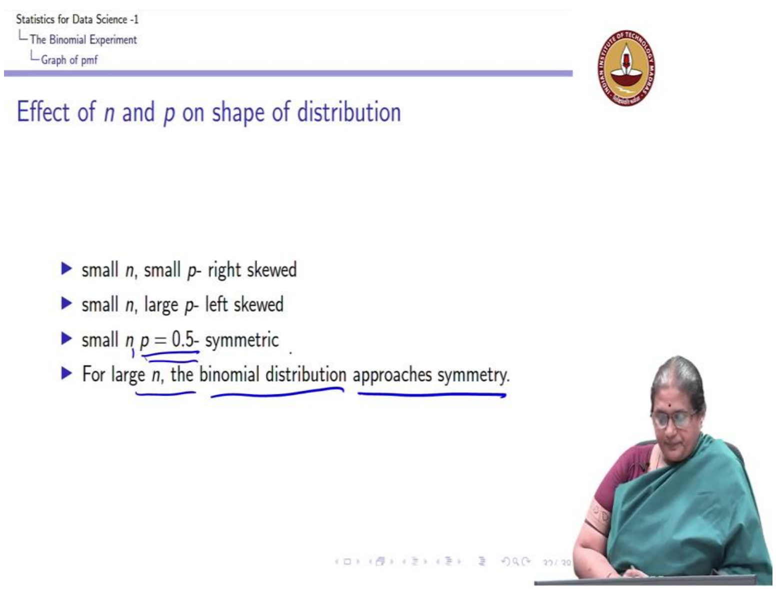
Graph of pmf of Binomial distribution- Right skewed



Graph of pmf of Binomial distribution- left skewed





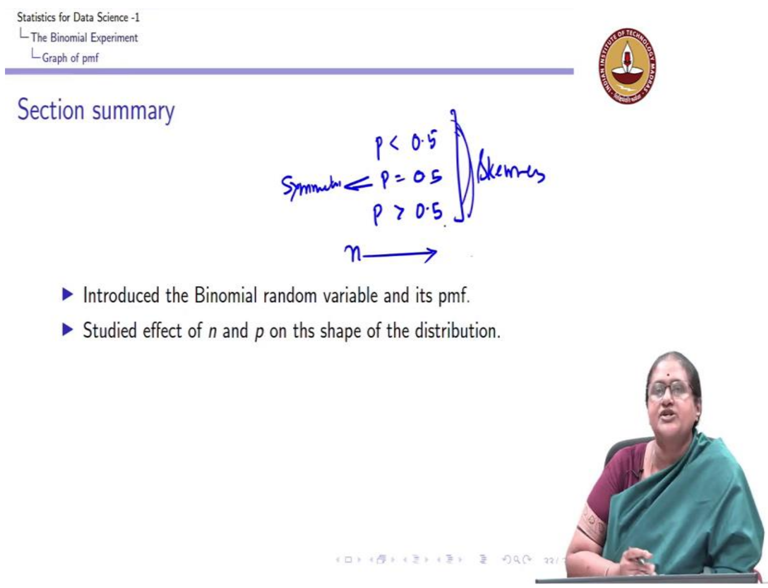


Similarly, when I go and look for a large n with P equals 0.8, I noticed that symmetricity again of course for large n and if p equals 0.5 I have again a symmetric distribution I repeat when I am taking for the same p but n equal to 40 earlier I noticed that right skewed was very predominant. So, when n was point n equal to 4 and p equal to 0.5, we noticed that the skewness was of this kind.

Now, when I have n equal to 40 and p equal to 0.3, I notice what is happening is I noticed some sort of a symmetric it is tending to a symmetric distribution. Similarly, when n was equal to 4 and p was equal to 0.8, I noticed a skewness of this kind and now I notice that that is also becoming a symmetric distribution. So, for large, so I can say that for large n I noticed the skewed distribution is tending to be symmetric and for same p, I am having a symmetric distribution, this is something which I want you to observe this would be formalised in your advanced courses, but this is a very important observation.

For same value of p if I vary n, I notice that for large n the binomial distribution approaches symmetric behaviour. However, for small n and p equal to 0.5 itself I noticed a symmetric behaviour. So, for large n I am going to continue to notice this symmetric behaviour.

(Refer Slide Time: 14:21)



So, what we can conclude is we introduced as the binomial random variable trail as the sum of independent and identical Bernoulli random variables, we obtained what was the probability mass function, we saw that this probability mass function the graph of the probability mass function, when for I first saw that I kept n same and a varied p to demonstrate for small values of n if p is less than 0.5 and if p is greater than 0.5 you get skewness and if p equal to 0.5 we get symmetry and we saw that as n increases or as n becomes large, the shape of all the distributions start exhibiting symmetry.