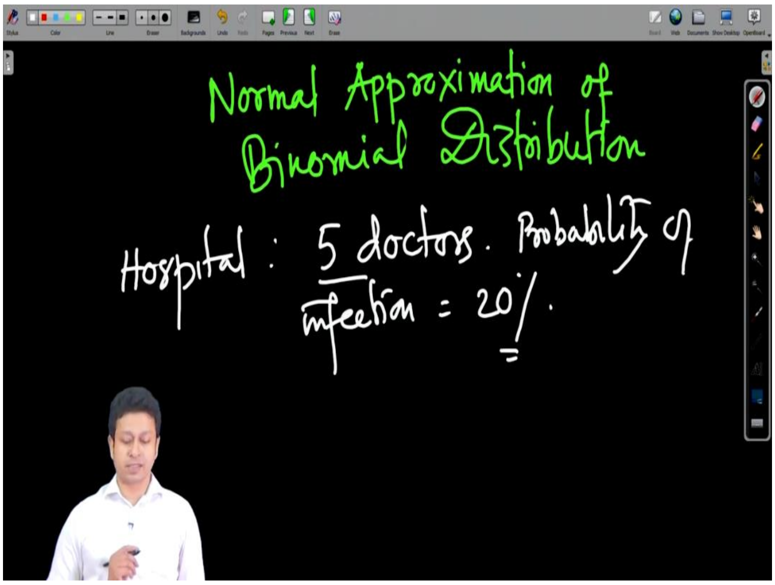
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Module - 2
Lecture - 22
Normal Approximation of Bernoulli Distribution

Hello and welcome back to the lecture on Applied Econometrics. So, we are dealing with a very interesting topic. We are talking about probability distribution and we have seen Bernoulli trial and Bernoulli distribution and how we can use those concepts for constituting a binomial distribution. Now, when we dealt with binomial distribution, one problem that we faced is that, when we have, when we deal with a large number of trials and we have to get something like, we have to get the probability for in each and individual events, it is going to be really a difficult task for us to actually calculate that.

And we said that there is a possibility, there is a way to actually approximate a binomial distribution to a normal distribution. So, now, before we actually do that, before we actually get into the theories, let us try to intuitively understand it, let us try to visualise it, how it is happening, and how we can actually see that in real life. So, I will do that, as usual with an example, because that will actually give us a sense of what is happening here. So, again, like the; we take an example.

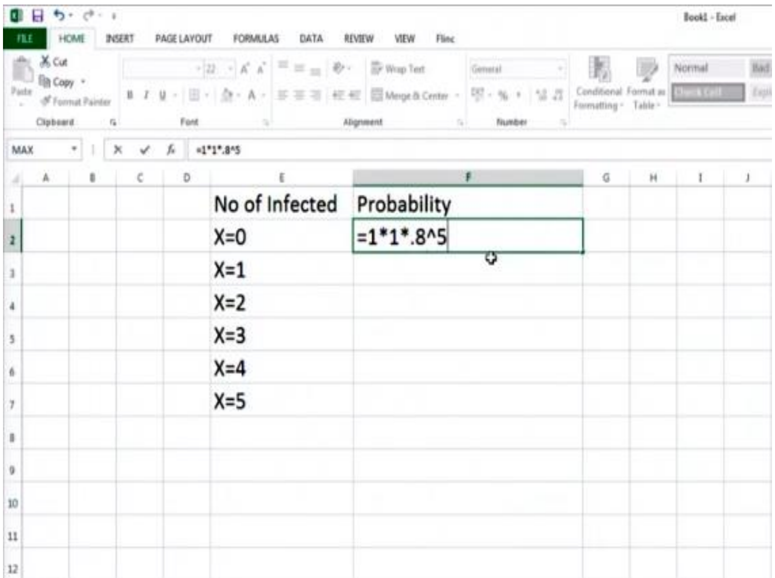
(Refer Slide Time: 01:31)



Let us say, this is a hospital. And hospital has, let us say 5 doctors. Now, it is the time of COVID, so, the hospital is really afraid that if the doctors are infected, if all the doctors are infected or if most of the doctors are infected, how it will run its business or how it will treat people during this tough time. Now, there are 5 doctors, and the probability of getting infected, probability of infection, let us say 20%.

So, now, hospital has, let us say call one of you as a statistician to actually help them to constitute a probability table of doctors getting infected. So, how many doctors will be infected and what is the probability of each of these individual events? So, what do you do? You basically, let us say we use a simple excel sheet here, and I have already written it down here.

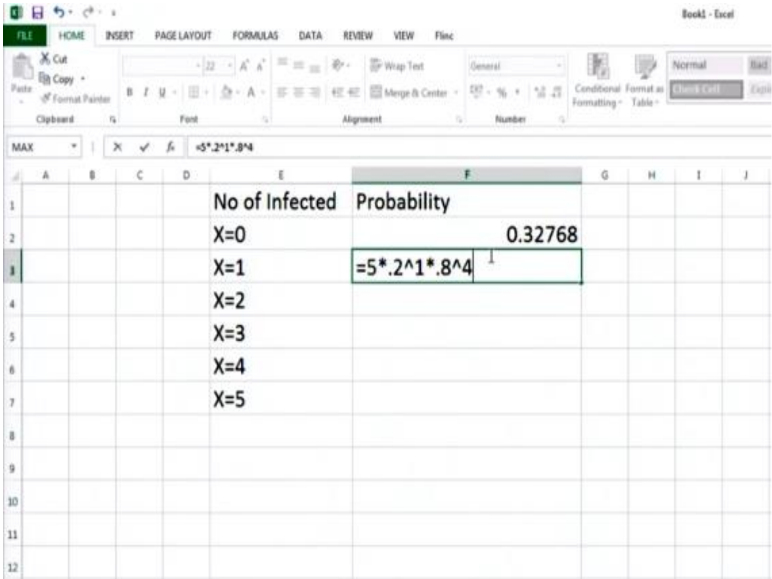
(Refer Slide Time: 02:41)



So, number of infected and their corresponding probability. So, let us say my random variable X, it can be 0, no infection, no doctor infected. X could be 1, X could be equal to 2, X could be equal to 3, X could be equal to 4 or X could be equal to 5; all 5 doctors could be infected. And you need to; so, these are your X's, and you need to create your probability, you need to get your probability. So, what do you use?

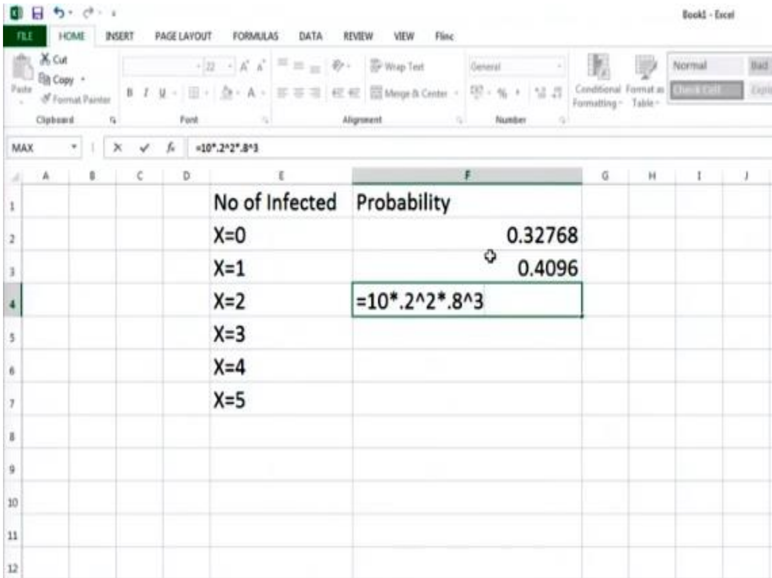
You already have learned binomial distribution, you know how it is to be used and you use the binomial distribution to write down the probabilities. Now, so, the first term is basically, n is 5 here, x is 0. So, 5 c 0, the first term 5 c 0 is equal to 1 into your 0 success, or whichever we say; so basically 0.2 power 0 which is equal 1. And then I have 0.8 power 5. So, that is the probability of 0 person getting infected. So, which is 0.32. Now, what about 1 person getting infected?

(Refer Slide Time: 04:01)



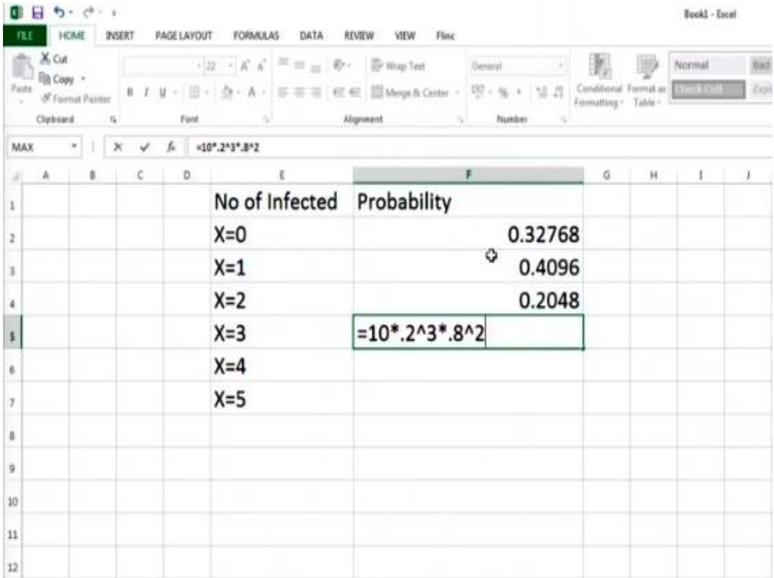
Probability of 1 person getting infected is 5 c 1 which is 5 into 0.2 is the probability of getting infected to power 1 into 0.8 to the power 4. So, the 4 persons are not infected and 1 person is infected. So, it is this. This is the value we are going to get. What about 2 persons getting infected?

(Refer Slide Time: 04:31)



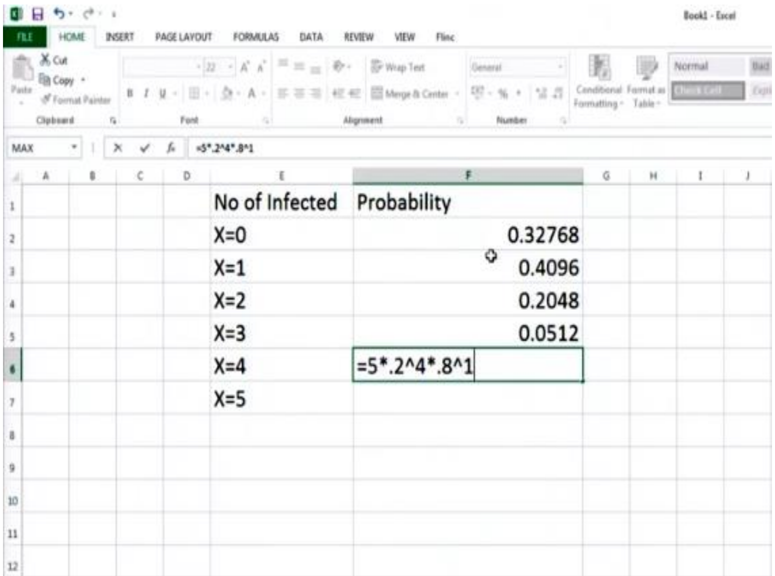
So, now it is 5 c 2 which is 10 into 0.2 power 2, I have 2 persons infected and I have 0.8 to the power 3, because 3 persons are not infected. So, the probability is going to be 0.2, 0.20 something. Now, what about 3 persons getting infected?

(Refer Slide Time: 04:57)



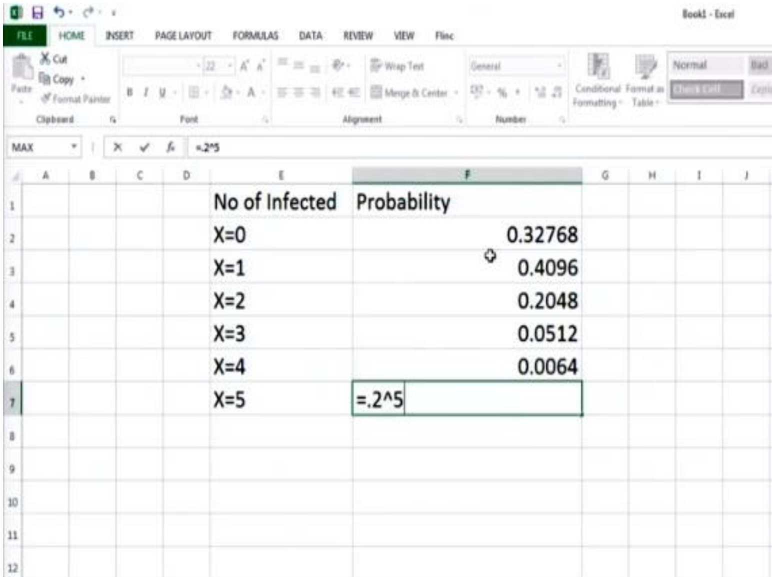
So, it is going to be 5 c 3. 5 c 3 is again 10. 5 c 3 into your 0.2 power 3 into 0.8 power 2. So, that is going to give you something like a 0.05. What about 4 persons getting infected?

(Refer Slide Time: 05:19)



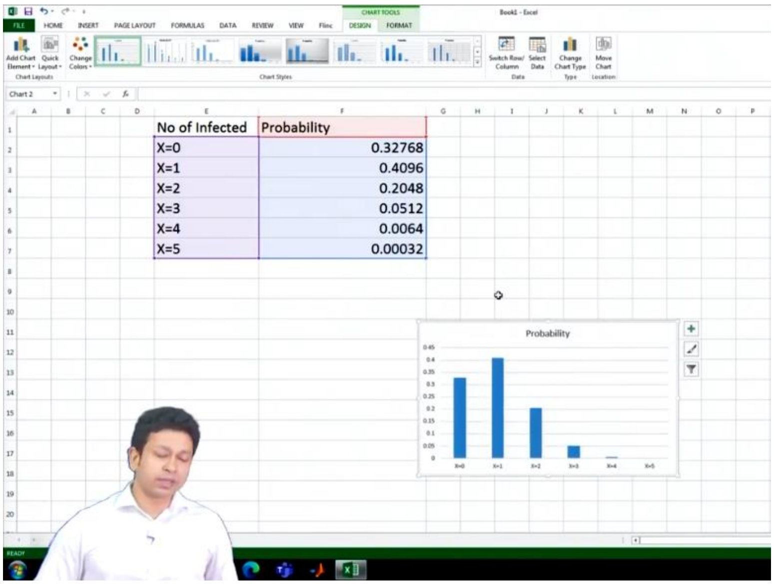
So, 5 c 4 is equal to 5 into 0.2 power 4 into 0.8 power 1. So, that is going to give you a value of 0.006. Now, what about 5 persons, all 5 persons getting infected?

(Refer Slide Time: 05:47)



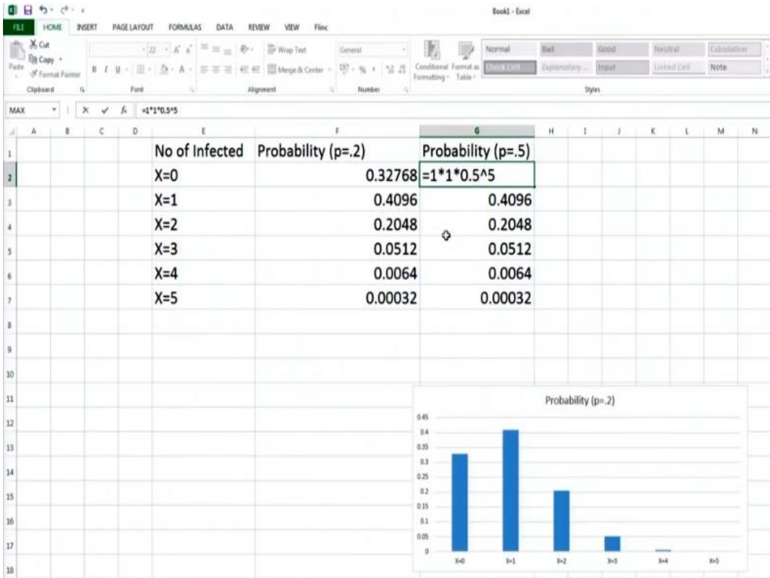
So, if all 5 persons are to be infected, so, we have to have 5 c 5 which is 1 and 0.2 power 5. And the last term is going to be 0.8 power 0, which is 1. So, basically what you will have is 0.2 power 5. And that is going to give you this value. So, if you add them up, of course, this total will be 1. Now, the reason I actually took the pain of actually deriving each of these probability values is to give you a sense how the distribution looks like.

(Refer Slide Time: 06:22)



If I really plot it, it is like somewhat skewed; you can think of a normal distribution somewhat is skewed. Now, so, this is what we are going to see if I have the uh these values here. Now, we will do some more intuitive understanding. We will build some more intuitive understanding here.

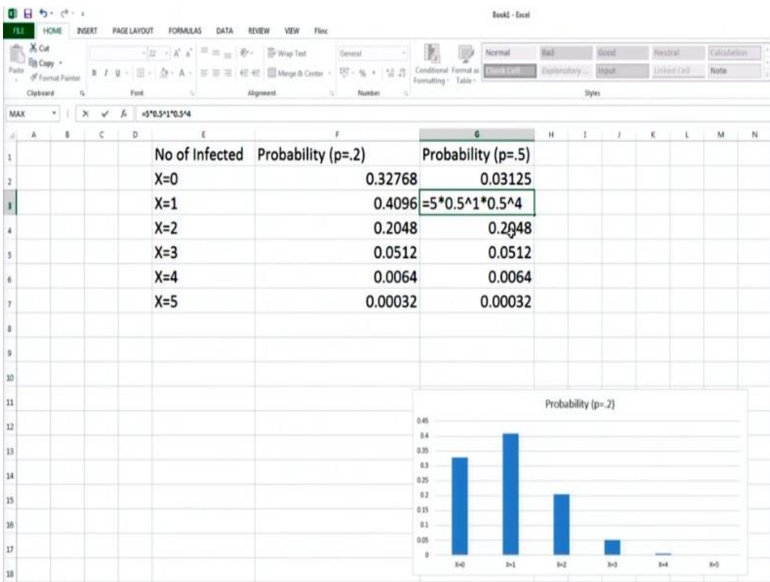
(Refer Slide Time: 06:41)



So, one thing we have done is that we have actually got a probability of successes to be 0.2. Now, I want to actually change that. I want to make it, let us say 0.5, it is like symmetric. So, success and failure is going to be symmetric. So, here, p = 0.2. And I am going to do another probability calculation. Probability, where p = 0.5. And here, what I will see is; let us again do the exercise.

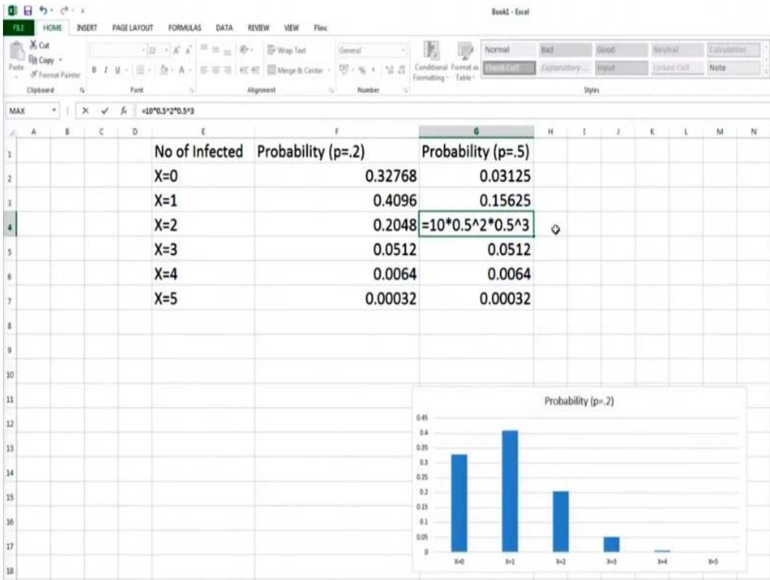
And there is a reason we want to do the exercise, and we will just see that. So, again, if X = 0, so, we will simply use the formula. So, it is 0.8. So, let me actually copy the whole thing here. So, here I have, instead of 0.8, I will simply replace it with 0.5. So, the success rate is 0.5.

(Refer Slide Time: 07:50)



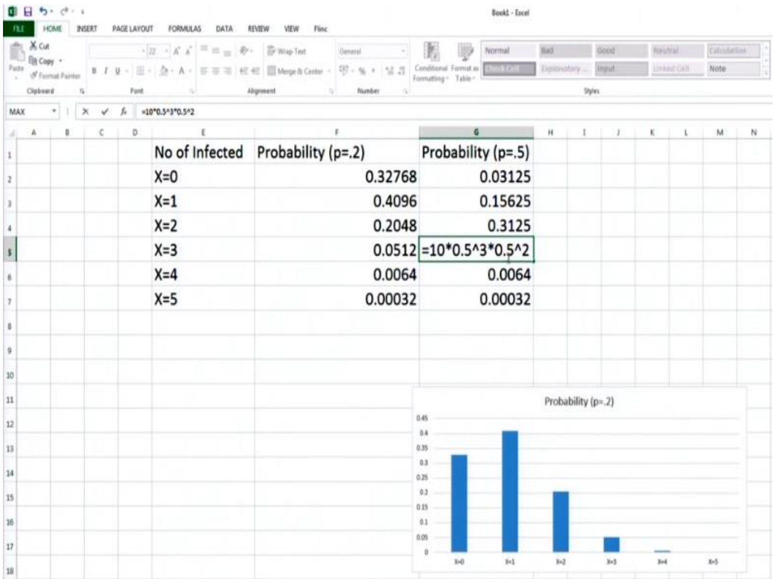
Instead of 0.8, here it is going to be 0.5, and this is also 0.5. So, you get a value here. I am doing it mechanically.

(Refer Slide Time: 08:00)



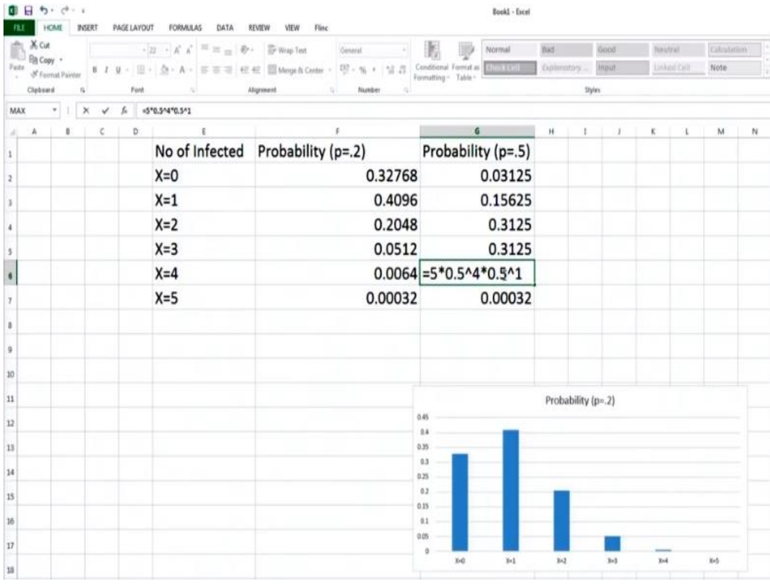
So, 0.5. All I am doing is, I am changing the probabilities. The n remains same.

(Refer Slide Time: 08:07)



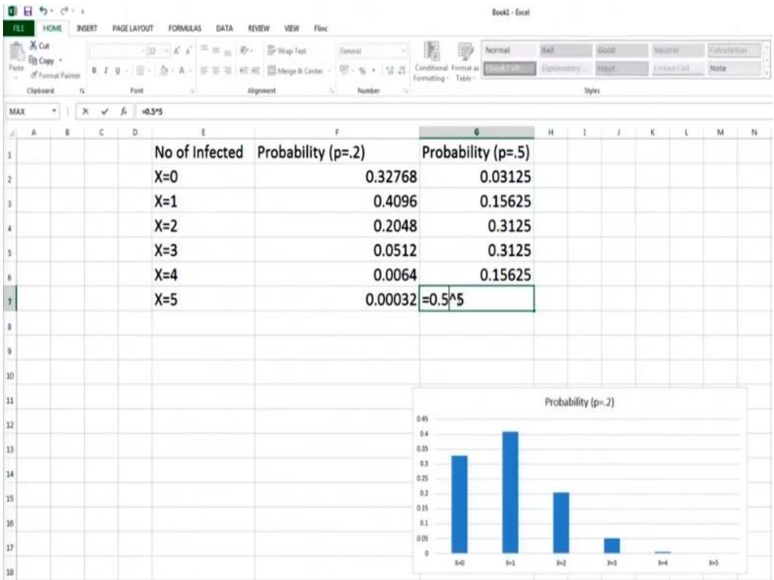
0.5, 0.5.

(Refer Slide Time: 08:14)



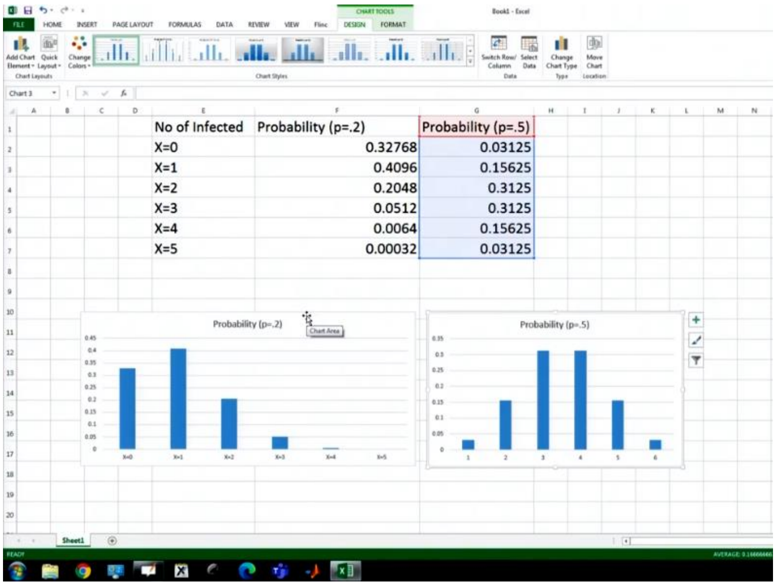
And here again, I change to 0.5 and 0.5. And we see this.

(Refer Slide Time: 08:21)



And here, of course, I change it to 0.5. So, I see this.

(Refer Slide Time: 08:26)



Why did I do that? Now, let me actually plot it, and let me see, let us see what we get if we plot it. So, I am going to plot this one. Insert; can I not add one? Yeah, here I get it. So, what is happening here, you see. You see the, this one is skewed, but this one is more symmetric. This particular, when I had probability equal to 0.5, I get a more symmetric distribution vis-a-vis what I got when I got probability p = 0.2.

So, the point I am trying to make here is that, if I have the probability of success and probability of failure almost equal, I am going to get a symmetric distribution. And the second point I am going to make is, which I am not going to show you is that, if I keep on increasing the number of trials; here I have only like 6 trials 0 to 5; but if I keep on increasing the trials, what I am going to see is, I will be seeing more sort of this curve to be more coming close and getting densed, and we can see a more symmetric distribution, like a bell shaped curve we will see.

And that bell shaped curve is nothing but our normal approximation of a binomial trial. And that is really helpful, because we cannot really afford to calculate all the probabilities for all the different events. So, for example, if instead of 5 doctors, I had 60 doctors let us say. And out of the 60 doctors, I had to calculate that at least 40 doctors will be healthy, they are not infected.

So, I have to basically calculate 20 possible events where the doctors could be infected, and I have to subtract that from 1. So, that is really difficult task. But instead, I can use a normal distribution and where I have fortunately for us all the z's. I can convert that into a standard normal distribution; I can get the z values. And we will talk about all these details in details in the next lecture.

So, there, we can simply use the z values and we have already a z table for us. And we can use the z table to actually see the probability corresponding to each of these z value. So, that is where we will see the results are coming similar or same, basically for binomial as well as for normal. So, that is the idea of normal approximation of the binomial distribution. And with this, we end our discussion on binomial distribution.

We will again revisit when we introduce the normal distribution. But before that, in the next lecture, we are going to talk about some fundamental theories, theorems that are really crucial for building many models that we are going to build going forward. And that is the building block. So, we are going to talk about these theorems in the next lecture. With this, we end our lecture on binomial distribution. Thank you.