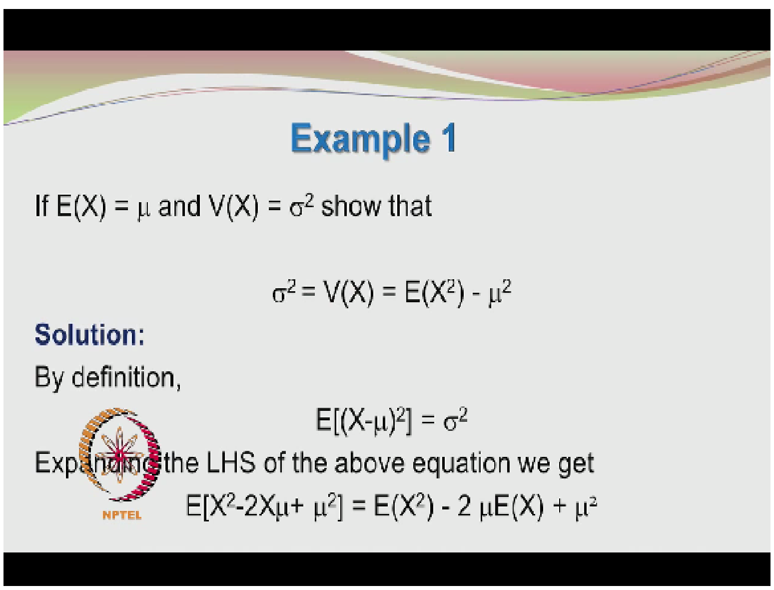
Statistics for Experimentalists
Prof. Kannan. A
Department of Chemical Engineering
Indian Institute of Technology - Madras

Lecture – 04
Example Set - I

Hello again. In today's lecture, we will be looking at the few typical problems involving discrete probability distributions and cumulative probability functions. It is suggested that you read the problem statement and try to solve the problem on your own and if your answers are correct, well and good, you have understood the course material.

Otherwise, please look at the solution which will also be provided in this lecture and compare it with your answers. You can find out where you made the mistake, okay. It is not a one-way street. There maybe even a mistake in my calculation but I have checked it a couple of times. So I am pretty sure that these problems are error-free, okay.

(Refer Slide Time: 01:29)

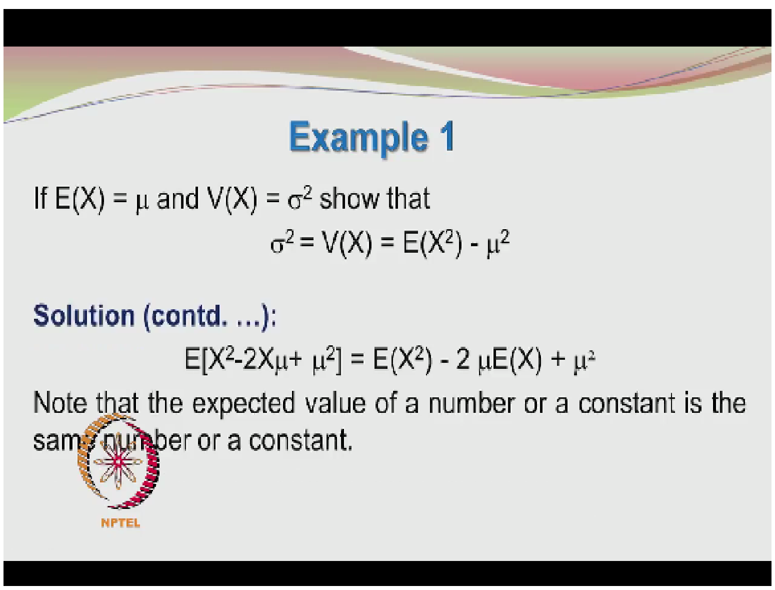


So the first example is a simple derivation, okay. We know that the expected value of X=mu, the mean value, okay and the variance of the random variable X=sigma squared, right. So do not put sigma here. Sigma means standard deviation and sigma squared means variance square root of variance is the standard deviation. So you have to show that the variance=the expected value of X squared-mu squared, okay.

The value of E of X squared need not be = E of X whole square, okay. You should not confuse this with E of X whole squared, okay. E of X squared is different from E of X whole squared. Now we know by definition the expected value of X-mu whole squared = sigma squared. What we will do is we will take the left-hand side of this equation and expand it, A-B whole squared is A squared-2AB+B squared.

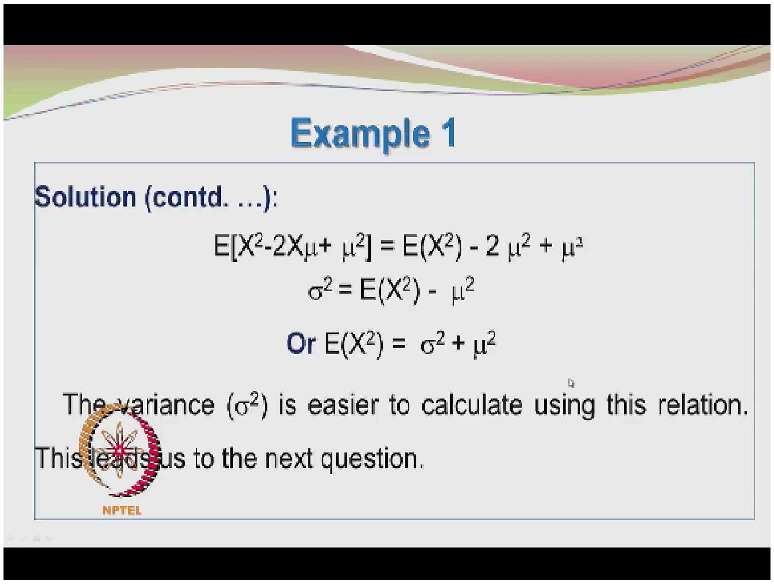
So you have expected value of X squared-2X mu+mu squared and then we take the expected value term by term. So we get E of X squared-2mu E of X+mu squared. You may be asking, wait a second, what is this E of mu squared? How did you write mu squared? The expected value of a constant is a constant, okay. So we can write expected value of mu square as mu squared directly and that also enabled us to take out the 2mu when we put the expected value or applied the expected value on 2X mu. So the 2mu came out and we simply had E of X, okay.

(Refer Slide Time: 04:04)



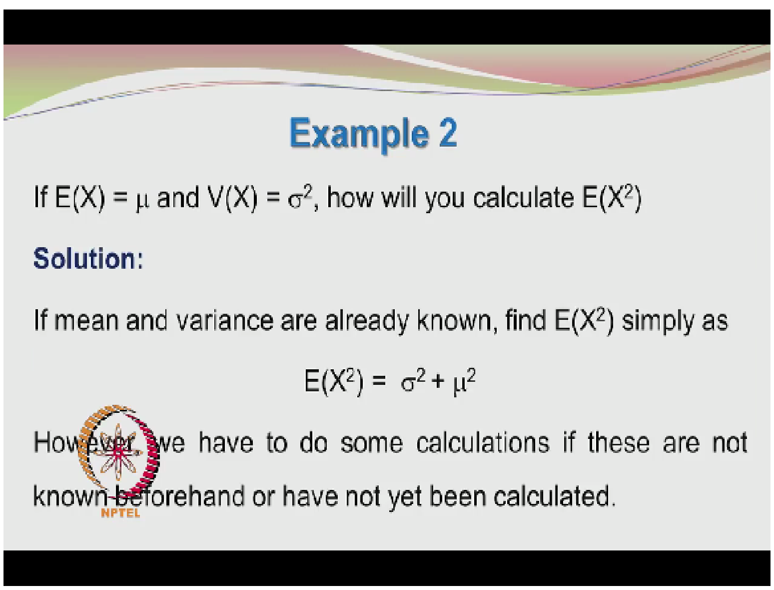
This is the question which is being repeated here for your convenience. So we end up with E of X squared-2mu E of X-mu squared. Repeating what I said a bit earlier the expected value of a number or a constant is the same number or a constant, okay. Now if you look at this expected value of EX=mu. So we can put mu here and we will get E of X squared-2mu squared+mu squared.

(Refer Slide Time: 04:46)



That is what I have shown in the slide. You can see that it is E of X squared-2mu squared+mu squared which leads to E of X squared-mu squared. So the expected value of X squared = sigma squared + mu squared, okay.

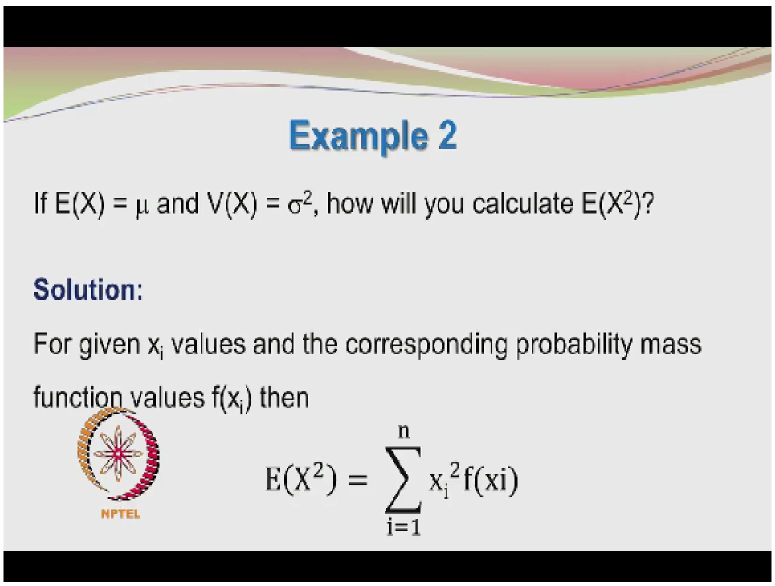
(Refer Slide Time: 05:20)



Let us see the next question. If E of X = mu which we know very well and V of X = sigma squared, how will you find E of X squared, okay? A solution is, if you are already having the mean and variance with you, you can use the result from the previous example and take it as E of X squared=sigma squared+mu squared, okay. Since sigma squared and mu square are already available, you can just add them up and then say that the expected value of X squared is the sum of these 2, okay; however, that is one way of doing it, okay.

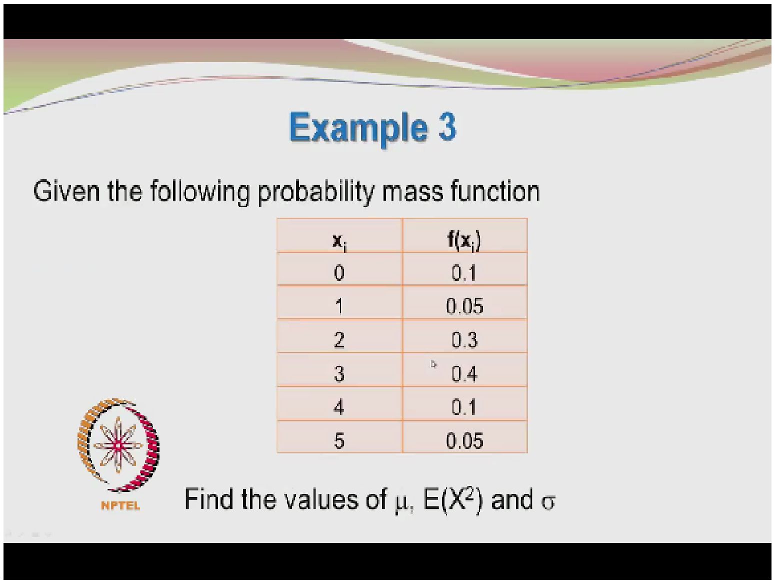
Anyway if those values, these 2 values are not available to you, then you have to do some calculations and find mu squared and sigma squared, okay. I am going to show you how to find E of X squared independently. So rather than doing somewhat or relatively tedious calculations for sigma squared, what you can do is you can first find out E of X squared, you already know or you can calculate mu squared rather easily and then you can subtract mu squared from E of X squared to get the variance sigma squared, okay. So let us see the calculations.

(Refer Slide Time: 06:56)



An independent way to find out E of X squared is to use the definition of E of X squared which is nothing but sigma the summation going from I=1 to n xi squared f of xi, okay.

(Refer Slide Time: 07:38)



x_i	$f(x_i)$
0	0.1
1	0.05
2	0.3
3	0.4
4	0.1
5	0.05

Now let us go to the third example where we will numerically calculate the mean variance and the expected value of X square, okay. So you are given the probability mass function, okay. The random variable X, capital X can take these 6 values, 0 1 2 3 4 and 5, okay. These are discrete values and please note that these are the only permitted values of the random variable. You cannot have a value like 0.5 or 1.5, okay and this is f of xi.

F of xi, if you recall, is the probability of the random variable taking the value 0, okay. For this particular case, f of xi is 0.1, okay and in general, f of xi is the probability that the random variable takes the value xi. Xi can be, in this case, any number involving 0 1 2 3 4 and 5, okay. So among all these cases, you can very easily verify that the probability values are not the same, okay. 0 is having 0.1, 1 is having 0.05, 2 is having 0.3 and so on, okay.

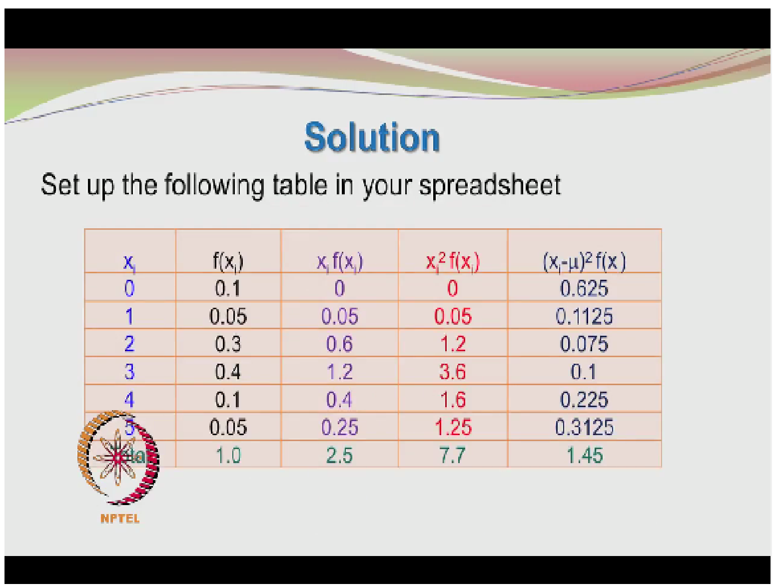
So whenever you are given problem like this, the first thing you should do is to see whether the problem is correct, okay. That is very important, okay. We can still carry out the routine mathematical exercises and give the answer but there may be a printing mistake or there may be a genuine mistake and the problem itself may be erroneous, okay. So we have to check the problem and see whether the statement is correct and the data given are correct.

The first check you have to make here is to see whether the probabilities all add up to 1. So I have given the table again and these are the probability values or f of xi values and when I am adding these values, I get 1. The number are not too difficult, so I can even add it up manually, 0.1+0.05 is 0.15; 0.15+.3 is 0.45; 0.45+0.4 is 0.85; 0.85+0.1 is 0.95; and 0.95+0.05 adds up to 1, good. So there is nothing wrong in the given probability distribution. So the equation is correct.

Now we can go ahead and find the other statistical parameters like the mean variance, standard deviation and the expected value of X squared. You might recollect that I told you that there are 2 different ways to find E of X squared. First way is to do in the following manner. You can calculate mu squared. You can also calculate sigma squared. Add the 2 and get E of X squared. That is one way of doing it. Another way of doing it is to multiply f of xi with xi squared, okay and then add, the total sum will give expected value of E of X squared, okay.

So let us do it in both the ways, right.

(Refer Slide Time: 12:03)



So this is your table. Now you can setup the problem and its data in a suitable spreadsheet. The data is given here. The first column contains the xi values; the second column is the column of probabilities. You can see that the sum of the probabilities = 1 and then I find xif of xi, okay. The purpose of doing that is to find the mean. Now what you have to do is find out xif of xi, right.

Please note that sigma xif of xi will not be = 1, okay. Only sigma of f of xi will be = 1. Sigma of xif of xi will be = the mean, okay, the average value of the given discrete probability distribution.

So you can see that I multiply 0*0.1, I get 0; 1*0.05, 0.05; 2*0.3, 0.6; 3*0.4, 1.2; 4*0.1, 0.4; 5*0.05, 0.25. Whereas sigma xif of xi, you can directly do the summing from the spreadsheet; otherwise, you can do it manually for the simple case. 0.65, 1.85, 2.25+0.25 is 2.5.

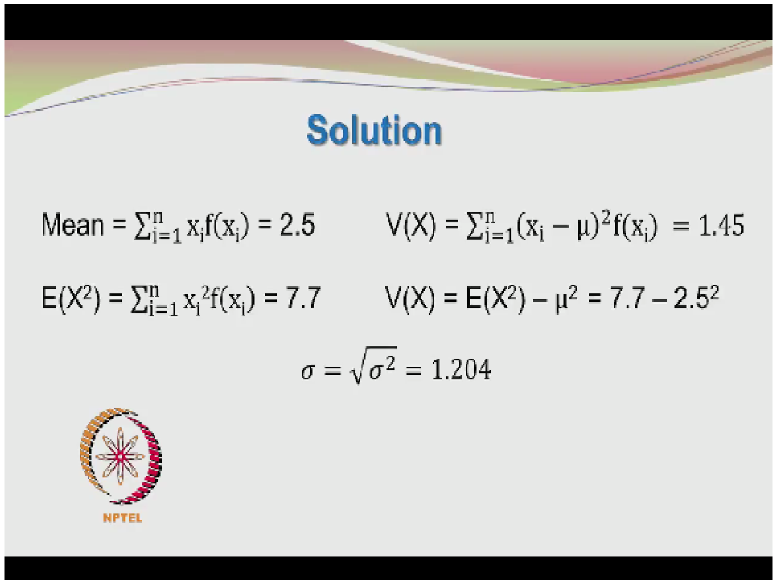
So the average or mean value is 2.5. Now you have to calculate the expected value of x squared which means we have to find E of x squared. So what you do here is, you take the squire of xi, xi squared*f of xi. So 0 squared*0.1is 0; 1*0.05 is 0.05; 4*0.3 is 1.2; 9*0.4 is 3.6; 16*0.1 is 1.6; 25*0.05 is 1.25. When you add this up, it comes to 7.7 and this is the expected value of x squared. Now to calculate the variance, you how to find x of i-mu whole squared*f of xi. What you are doing is, find the deviation of the x of I value from the mean value, okay, square it, okay.

So you have to find 0-2.5 that means it is -2.5. When you square it, it will become 6.25. When you multiply it by 0.1, it becomes 0.625. Then 1-2.5 is -1.5, so -1.5*-1.5 is 2.25, that you multiply by 0.05, you get 0.1125 and then you do 4-2.5 that is 1.5, okay and the square of that is 2.25 and then you do with multiplied by 0.3. Similarly, you can do it for all the other values. I will just pass a moment to do some calculations here, 2-2.5. Actually I made a mistake, okay.

It is not 4-2.5, that is a mistake. It is 2-2.5 which is -0.5. So -0.5 squared will be 0.25 and then 0.25*0.3is 0.075, okay. So these are some typical mistakes you can make when you are doing hand calculations. These mistakes are unlikely when you do the calculations directly in the spreadsheet, cutting and pasting the formula in the cells. Similarly, you can do 3-2.5 is 0.5; 0.25*0.4 is 0.1 and similarly you can get the other values.

So the total sum of the squared deviations multiplied by the respective probability distribution values is 1.45. This is nothing but the variance, the sum is the variance, okay.

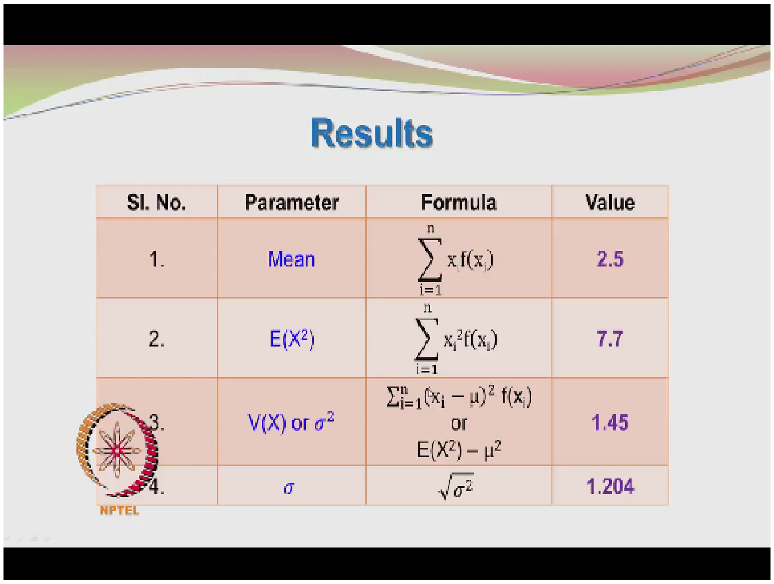
(Refer Slide Time: 17:39)



So summarising the values we have got, we found the mean to be 2.5, the variance we found from the previous example is 1.45, the expected value of X squared from the previous example was 7.7 and we can also crosscheck whether this variance of X, okay, this variance of X = 1.45. This is from the direct method. Again you can see that there is a mistake here which I will correct now, right. So earlier this f of xi was missing.

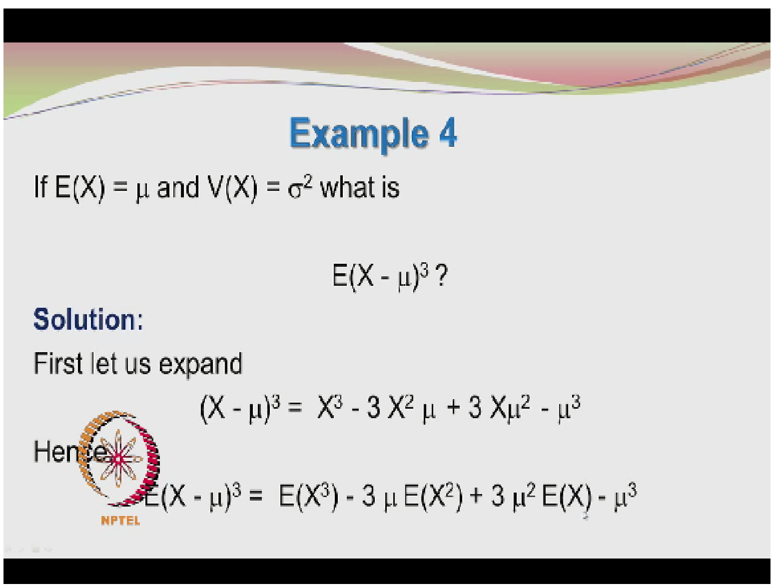
Now I have put variance of X = sigma i=1xi-mu whole squared*f of xi that is coming to 1.45. When you crosscheck it, you can see that variance of x is again coming to 1.45, when we subtract mu squared from E of X squared, we get E of X squared as 7.7-mu squared, 6.25. So 7.7-6.25 is 1.45. The standard deviation will be the square root of the variance and square root of 1.45 is 1.204.

(Refer Slide Time: 19:34)



The results are nicely summarised in the form of a table here. The parameter is mean E of X squared. Variance of X or sigma squared and sigma. The formula is also given here. Again you have to make a correction here. So you can see that it sigma i=1-n xi-mu whole squared*f of xi and for that we got a value of 1.45, okay. So these are the results for the different parameters and they are summarised in this table.

(Refer Slide Time: 20:33)



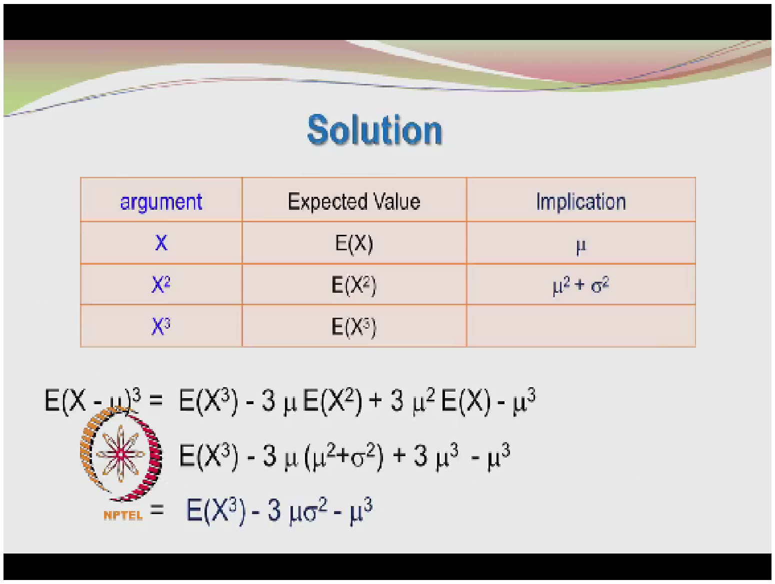
Now we got one interesting problem. You are having E of X=mu and variance of X=sigma squared. So you have to find out the expected value of E of X-mu cubed, right. So I hope all of you remember what is A-B whole cube, okay. A-B whole cube is A cube -3A squared B+3AB squared - B cube. So using that, instead of A and B, we put X and mu. X-mu whole cube is X cube-3X squared mu+3X mu squared-mu cube.

So we have to again apply the expectation on this, x-mu whole cube, we get E of X cube-3mu E of X squared+3mu squared E of X-mu cube. Please recollect that when you apply the expectation on a constant or a number, then you get the same constant or the number, okay. Only for a variable, a random variable, you have the expectation of that X having a non-constant value, you have to do some mathematical calculations with that, okay.

Once you are given the random sample and you are given the data, then you can find the appropriate expectation. So as of now, we are not given any data. We are only given X-mu whole cube and we have to find the expectation, okay. So to summarise, we get E of X cube-3muE of X squared+3mu squared E of X-mu cube. We do not want to leave the result here because there is some possibility of further simplification.

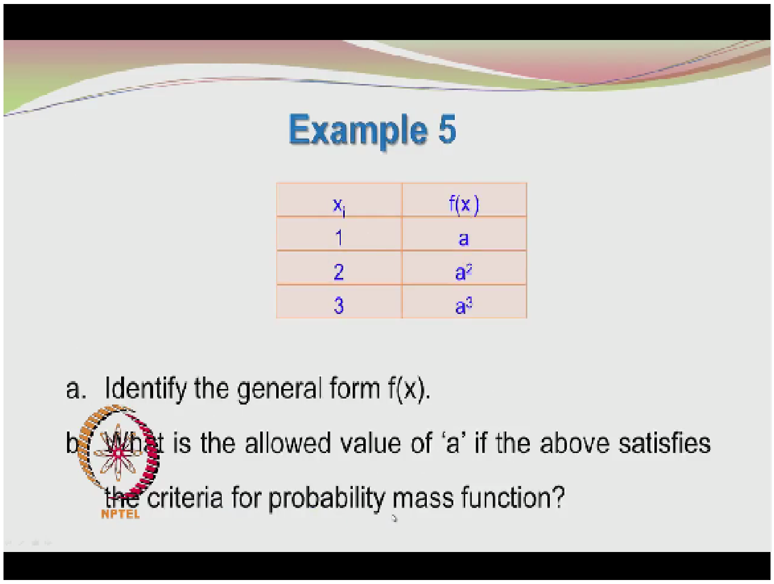
So you know that the expected value of X squared=mu squared+sigma squared. We saw that in one of the previous examples.

(Refer Slide Time: 23:15)



As what is summarised in this table here, So YOU have expected value of X=mu, expected value of X square=mu squared+sigma squared. So in this E of X squared, you substitute mu square+sigma square-3mu*mu square will become -3mu cube that will cancel out with this 3mu cube and you will be having E of X cube-3mu sigma squared-mu cube, right.

(Refer Slide Time: 23:50)

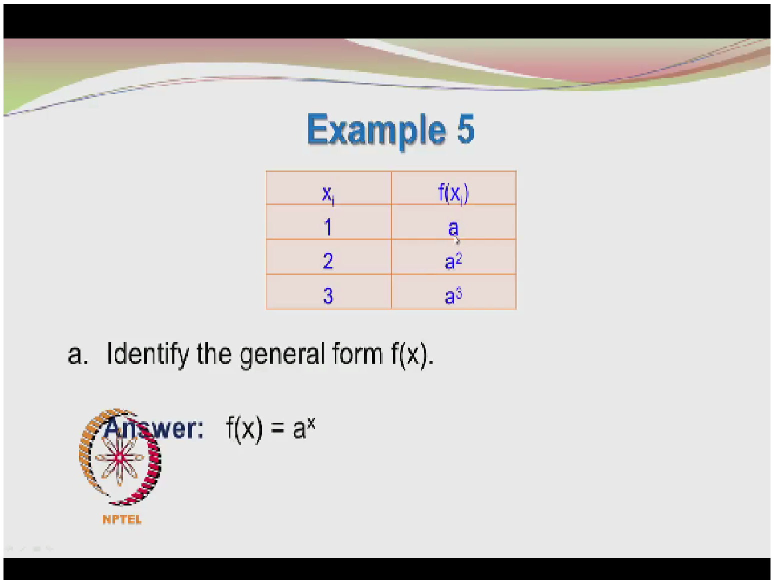


Now we come to a quite an interesting problem. I setup the problem myself. You can see whether you can do it on your own. You may want to pause and then do the problem yourself. Even the previous problem I cratered my own data and the rest was routine you have to just do the numerical computation. The one involving E of X-mu whole cube, is a standard result, okay. For example, Ogunnaike's random phenomena book has a dozen example, sorry, an exercise problem, okay.

So now if you look at this problem, the first subdivision states identify the general form f of x, okay. You want to give a mathematical expression which will describe this given data, okay, when x=1, the probability distribution takes a value a. If it is 2, it takes a value a to the power of 2. If it is 3, it takes the value a to the power of 3. So the general form of x is easy to identify. It is simply a to the power of x where xi values can be 1, 2 and 3, okay.

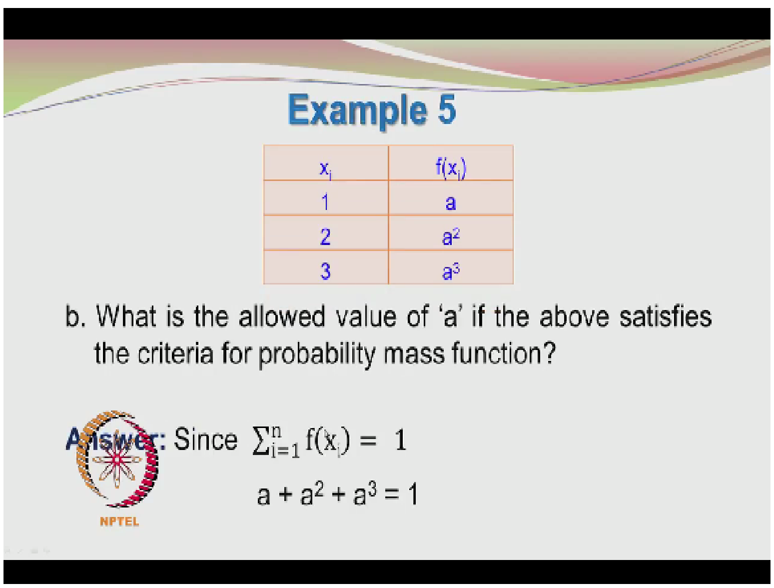
The second question or the second subdivision states that what is the allowed value of a if the above satisfies the criteria for probability mass function, okay. So we have 2 determine a value for a, okay, using the conditions that need to be satisfied for a probability distribution function, okay.

(Refer Slide Time: 25:58)



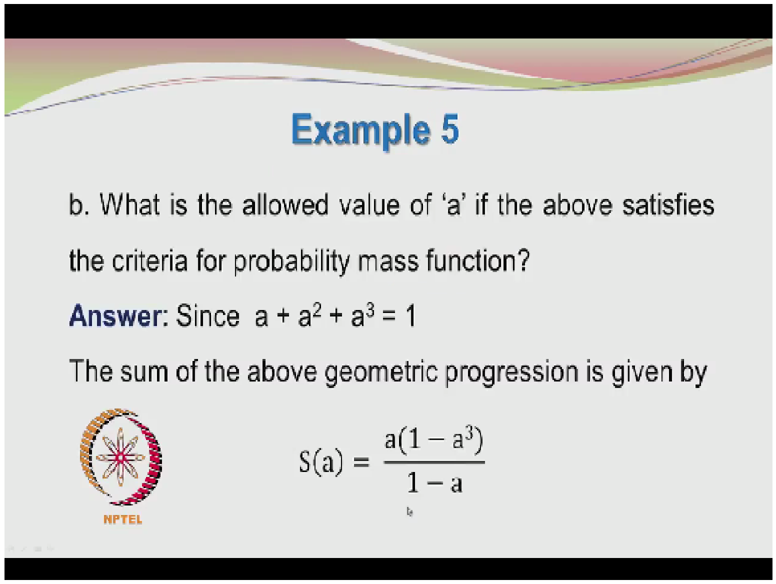
Identify the general form of f of x as I said a short while earlier f of x=a power x, okay, a, a square, a cube. So this expand this matching with the x value, so it becomes a power x, okay.

(Refer Slide Time: 26:18)



So the important condition that needs to be satisfied for this function to be a probability distribution function is the sum of the probabilities should be = 1. There is another criterion that these f of xi values should be positive, okay. So we will see if both these conditions are being met. So what we have to do is sum the f of x values overall the possible values of x. So a+a squared+a cube that should be = 1. So what is a value of a which will satisfy this equation.

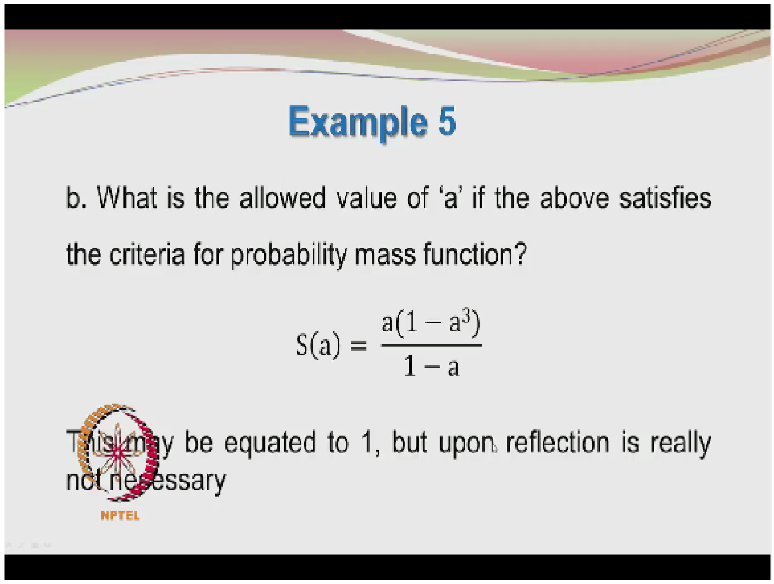
(Refer Slide Time: 27:07)



It is geometric series. So you may want to take the summation formula and do it, okay and you will get a*1-a whole cube/1-a. It is a more compact expression, okay and the you have to find the value of a such that this equation will become = 1. You cannot put a=1 because then this summation expression will blow up, okay. You may think of expanding 1-a cube, okay. You know 1-a cube can be expanded into 1-a*1+a squared+a or to put it in other way, 1-a whole cube can be expanded into 1-a*1+a+a squared, right.

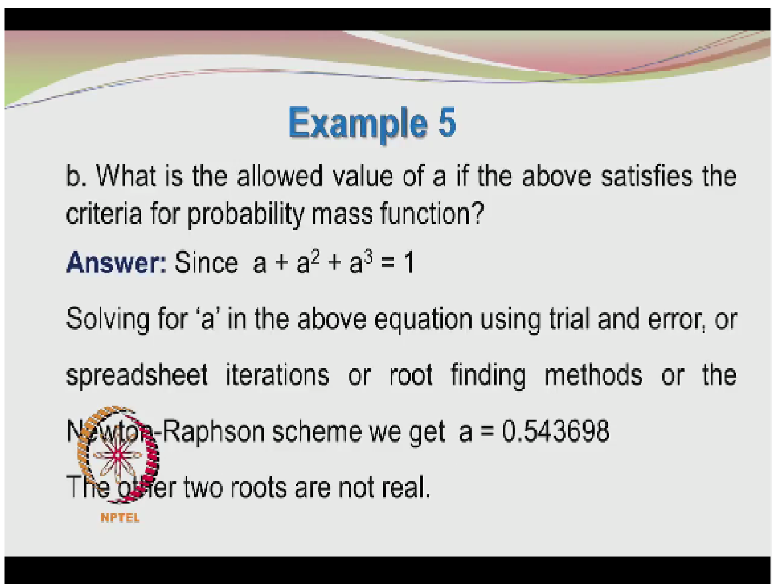
So the 1-a will cancel out and you will have again a+a squared+a cube. So we are back where we started. In fact, there is no need for us to take the summation of a using the geometric progression, okay. We could have directly taken this as the equation and try to solve for a such that this equation satisfies the constraint that the 3 terms should add up to 1. So we do not really need to do the summation and take the summation.

(Refer Slide Time: 28:43)



So it is not really necessary.

(Refer Slide Time: 28:48)

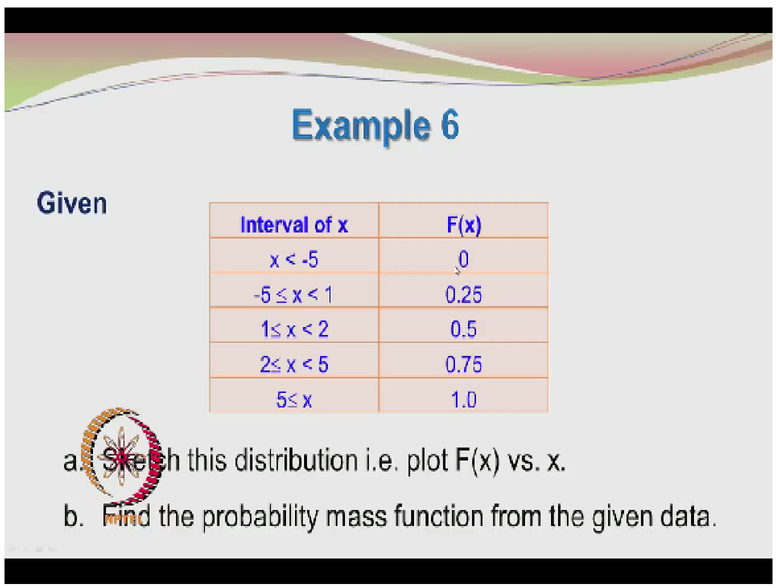


Now there are various ways to skin the cat. The first way is to simply use trial and error. You do not know where to start, okay. Obviously the value of a is likely to be between 0 to 1. If you put a=0, you get the sum to be zero and if you put the value of a=1, the sum becomes 3. So you are ranging between 0 and 3. You want to find the value of a such that the sum = 1. So you can do trial and error or you can use Newton-Raphson method or you can use appropriate algorithm in Matlab, okay and eventually find the root as reported in one of the softwares as 0.543698.

We do not really need all these decimal values. You can report the answer as 0.5437, okay. So after finding this root, we need to do a check. What you need to do is take the obtained value and plug it back into the equation you are trying to solve and see whether the equation is satisfied with this identified value. So what you may want to do is take 0.5437, then square it and cube and add all these 3 numbers and see whether the answer comes to be 1.

In our case, this answer would be around 1.000 and some decimal number which is pretty close to 1. So we have identified the value of a to be 0.5437 and then what we can do is, instead of giving the value as a a square and a cube, we can directly enter the value of a and then give the actual numbers.

(Refer Slide Time: 31:01)



Let us do the next example. Here we are going to do an example involving the cumulative distribution function, okay. So the problem statement is like this. You are having different intervals of x okay and then the cumulative probability distribution function value is also reported for each of these intervals, okay. Now the first interval is x<-5 and the f of x value which is the cumulative distribution function value that is = 0.

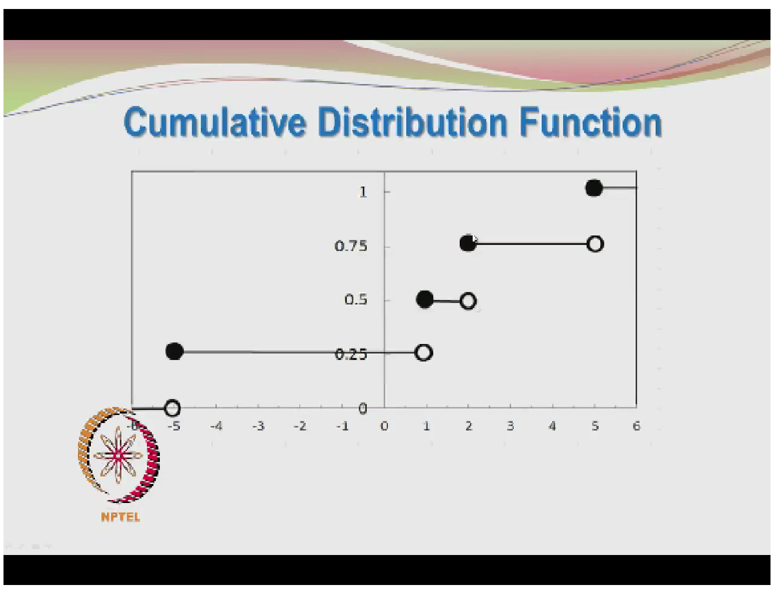
Then you have the interval -5 <= x<1, you have the cumulative distribution function value to be 0.25 and then you have 1 <= x<2, the cumulative distribution function value is 0.5. Then you have 2 <= x<5. The cumulative distribution function value is 0.75 and then 5 <= x 1.0. So one thing you may notice about this interval is it is involving no inequality sign. It is only x < -5.

Then the next interval is having -5 <= x<1.

So you are having <= here. Then you have 1 <= x<2. 2 <= x<5. 5 <= x, okay. So in the last 4 intervals, you are having the first as <= sign and then the other part is not having the equality contribution, it is just directly <, <, < for the 3 intervals, okay. Just an observation. So now we want to sketch this distribution plot of f of x versus x, okay and then find the probability mass function from the given data.

We are given the cumulative distribution function. So we need to extract the probability distribution function from the given cumulative distribution function, okay. As I told earlier, we need to check whether the problem statement is correct, okay. We are not making any mistakes such as putting X here, okay. So you cannot put X here because X is the abstract random variable and once the data is available, it takes a value x, okay. That is something which we have to watch out for. It is also cap, F of small x here.

(Refer Slide Time: 34:17)



This is a very interesting diagram. You can see some lines with some headers and the footers or you may want to call it as a tail and the head. The head sometimes a hollow circle or it is a full circle, okay. That is because the cumulative distribution function takes a value of 0 until x < -5, okay. At x value of -5, it takes the value 0.25 and continues to take that value until it reaches 1 and once it has reached a value of 1, it jumps to 0.5, okay.

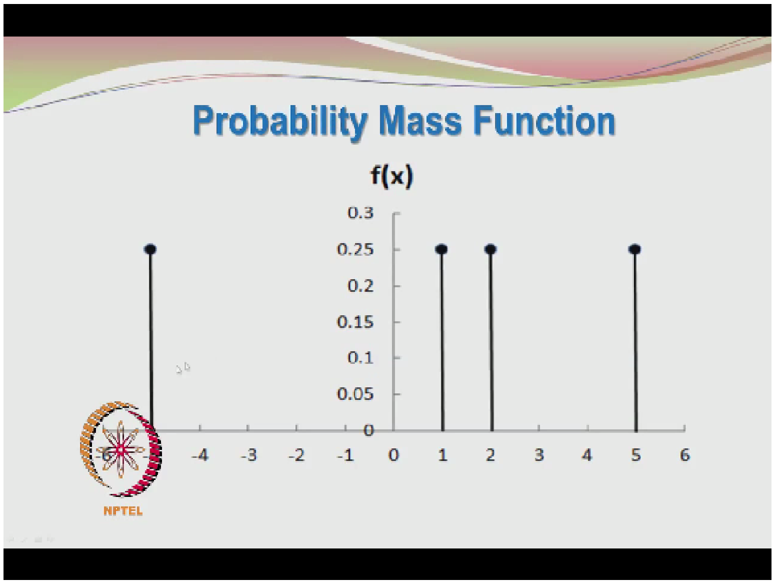
So this is the cumulative distribution function having a value of 0 and at -5, there is a discontinuity. The function is not continuous. If the function was continuous, the value of F of x, whether you approach a value either from the left or from the right, should be the same but there is a discontinuity here, the function value when you are approaching -5 from the left is 0. If you are approaching -5 from the right, it is 0.25.

So this is inevitable discontinuity at -5 and all the other points where the interval switches, okay. So at -5 and up to 1, the cumulative distribution function takes a value of 0.25 and then it jumps to 0.5 until you reach the value of 2, okay. So it takes a value of 0.5 until you reach the value of 2 and once again there is a discontinuity at the value of 2 and the functional value jumps from 0.5 to 0.75. So up to -5, it is 0. At -5, it jumps to 0.25. It goes up to 1 and then takes another jump.

So there is a discontinuity here, takes a value of 0.5, goes to 2. There is a discontinuity here. It takes a jump, it goes to 0.75 and then it goes all the way up to 5 where it takes another jump, where it reaches a value of 1 at which point it saturates, okay. Cumulative distribution function is the sum of the probabilities and once you have reached the end of the permitted values for the random variable x, then you should have reached 1, okay.

So there is no further change or jumps or discontinuities beyond this point, okay. Now the next step is to retrieve the probability distribution function, okay and how to do that.

(Refer Slide Time: 37:50)

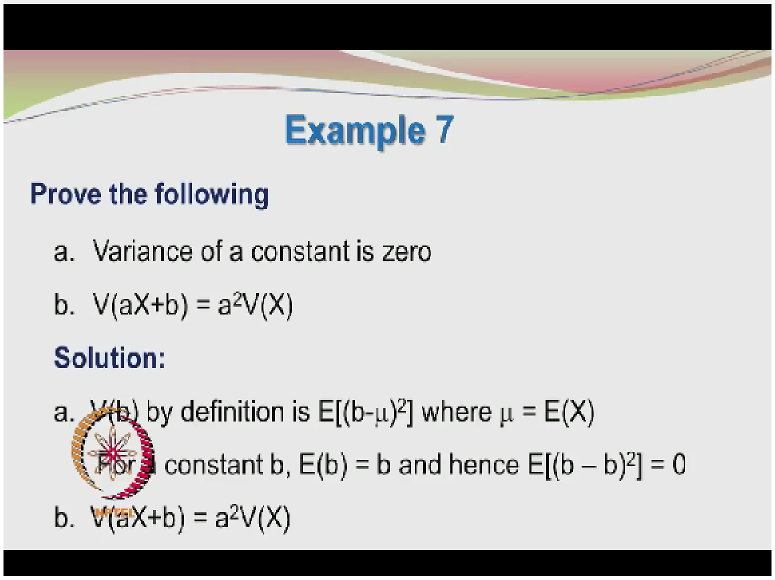


So you can see that the probability mass function or the probability distribution function is defined at -5 1 2 and 5, okay. So the permitted values of the random variable x are -5 1 2 and 5. At each of these values, the probability is 0.25, okay. You can see that whenever we reach the assigned value for the random variable, the jump occurred by 0.25 units. So from 0 you went to 0.25, that means the probability value corresponding to -5 is 0.25.

Since you are adding the probabilities, when you went to 1, the probability value is again 0.25 at 1 because you are going from 0.25 to 0.5. The difference is 0.5-0.25 which is again 0.25. So at 1, the probability is again 0.25, okay and the total sum is 0.25+0.25 which is 0.5 and that is what you are shown here and when you go to the value of 2, there is a jump to 0.75. So from 0.5, you have gone to 0.75, an increment of 0.25 and that is the probability value at x=2.

Similarly, at the x value of 5, the probability is again 0.25, okay. In this particular example, it is clear that all the values of x have equal probabilities, okay. You have 1 2 3 4, okay. We are having 4 values of x -5 1 2 and 5, okay. 1 2 3 4, each of them is having a probability value of 0.25. So when you add 0.25 4 times, you get 1. So the sum of f of x=1 as it should be.

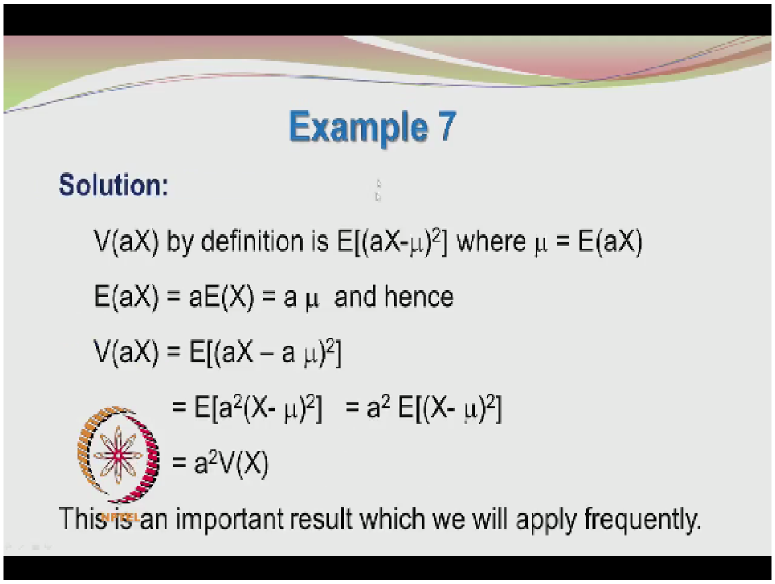
(Refer Slide Time: 40:31)



Now let us go to the seventh example, okay. You have to prove that the variance of a constant is 0 and the variance of a linear combination of random variable, in this case, it is a simple linear combination. aX+b, okay. For the particular case of variance of aX+b, you have to simply show that it is equal to a squared variance of X. This is an important result; we will be frequently encountering that in our course. First let us go to part a, variance of a constant is 0, okay.

So the variance of b by definition is expected value of b-mu whole squared, okay, where mu=expected value of X. For a constant b, expected value of b=b, okay. So you have expected value of b-b whole squared=0. In this case, the X is taking a value of b, okay. So expected value of b-b whole square=0. Let us go to the next part, variance of aX+b=a squared variance of X.

(Refer Slide Time: 42:11)



So we use the definition for variance again, expected value of, instead of X, we have aX. So we just put aX here, aX-mu whole squared, where mu=expected value of a of X, okay. So expected value of a of X is nothing but a*mu. So we have variance of a of X=expected value of a of X-a*mu whole squared, okay. So we can take within the bracket, a is common. So you can take a outside.

So it becomes a squared when you square it, a squared*X-mu whole squared. You can take a squared outside the bracket and so you have a squared*expected value of X-mu whole squared which is equal to a squared*variance of X, okay. This is an important result which we will apply frequently. So this concludes our session on example problems involving discrete probability distributions.

We will not be extensively using these discrete probability distribution functions, okay because in our real-life problems, we assume the random variable to be continuous, okay and when you have continuous distributions, instead of calling it as the probability distribution function, we call it as the probability density function, okay and in the discrete probability distribution cases, we were using the sigma notation or the summation symbol, okay.

When you go in for continuous probability distributions or the density functions, you will be using the integral sign, okay. There is a lot of use in the discrete notation also. Sometimes when you are not able to do the integration analytically, you may want to do the calculations for finding the integral using numerical schemes and these numerical schemes involve the summation of different quantities.

So even though we are not going to use the discrete probability distributions that much, it forms the necessary background for the next step that is the continuous probability distributions, okay. There also we have the expected value of the X, X squared, variance of X, variance of aX and so on, okay. These are just illustrative problems. You can find a lot of problems, okay, in the reference book I mentioned by Montgomery and Runger.

These problems also have answers, okay. The odd-numbered problem answers are given if I remember correctly. So you can work out some of these problems and gain practice, okay. The advice is do the problems even if it appears to be simple, preferably by hand okay and then you can also verify your hand calculation with the spreadsheet calculations, okay. Oftentimes when you are doing statistical analysis of data, there may be a silly mistake creeping somewhere, okay.

And sometimes when the mistake is made at the very beginning of your data processing, it becomes difficult to detect at a later stage and sometimes your results may not be as you expect, just because you have calculated the mean wrongly or the variance wrongly, okay and the current trend is to go in for softwares, okay. Many softwares are available, I mentioned Minitab, okay.

But these kind of hand calculations are essential so that you appreciate the significance of various parameters, okay and you really have a feel for what they really mean, okay. So to summarise, do as very problems as your time permits, make sure that whatever you have done, you have done correctly and crosscheck with the answers. We will go to the next lecture session from now.