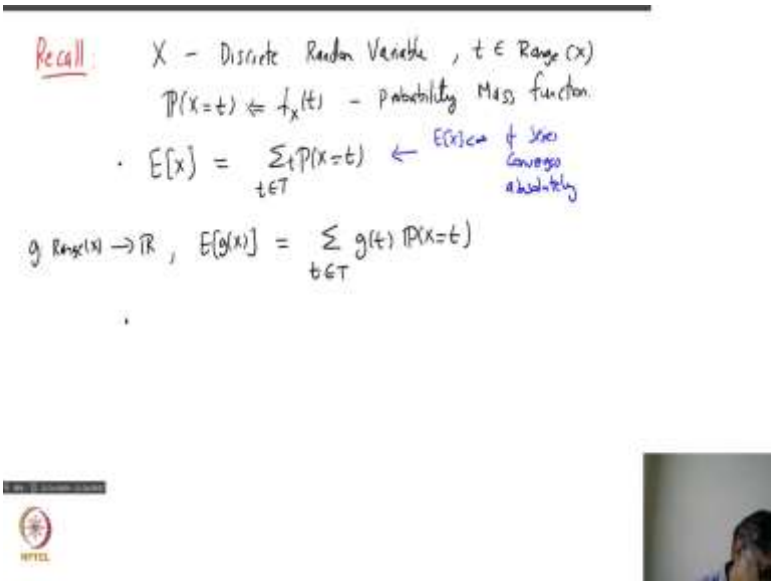
Introduction to Probability – With Examples Using R
Professor. Siva Atherya
Theoretical Statistics & Mathematics Division
Indian Statistical Institute, Bangalore
Lecture No. 24
Variance of Discrete Random Variable

So, we were discussing variance last time. So, we just finished the discussion.

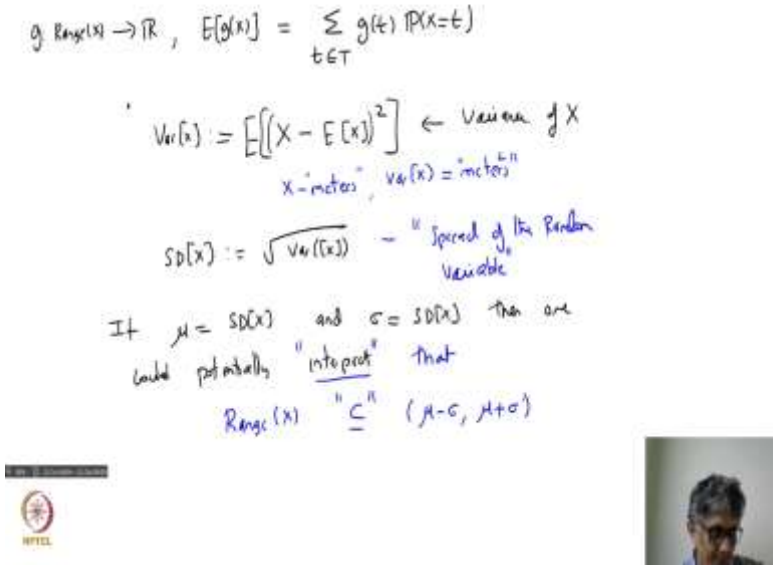
(Refer Slide Time: 00:23)



So, we had just start this idea recall. So, until now we have always been discussing X as a discrete random variable. And so the range of X is countable. And then we know that the probability X equal to t we call that as f X of t. We define in the reverse way we define this as the probability mass function. And then we had this the mean of X the actual value of X for the sum over all t and t so t is the range of X sum over t and t are the chance that t times the chance that X equal to t was the mean of X. And this we interpreted as whenever the series converges absolutely.

So, E of X is finite if series converges absolutely. And we set infinity if the series diverges and so on and so forth. If it does not converge at all to infinity of infinity we said that effect does not exist. Then we also said the following that we had this idea that E of X was one second so if here if you had a function of g from the range space of X at the real line. The real line then we have that E of g of X you got to go and find the distribution of g of X you could just use a change through a variable idea and get that g of t times the chance that X equals one thing. And then the other thing we discussed that once we understand this we thought of E of X as sort of the place where the random variable is centered.

(Refer Slide Time: 02:58)



So, we had two ideas about how to understand how the random variable is far from its actual mean. How close it is to its series. So, towards this I said that what we are doing is standardized into squares. So, you know the exact distances of the random variable from its mean there is no cancellation on (())(03:11).

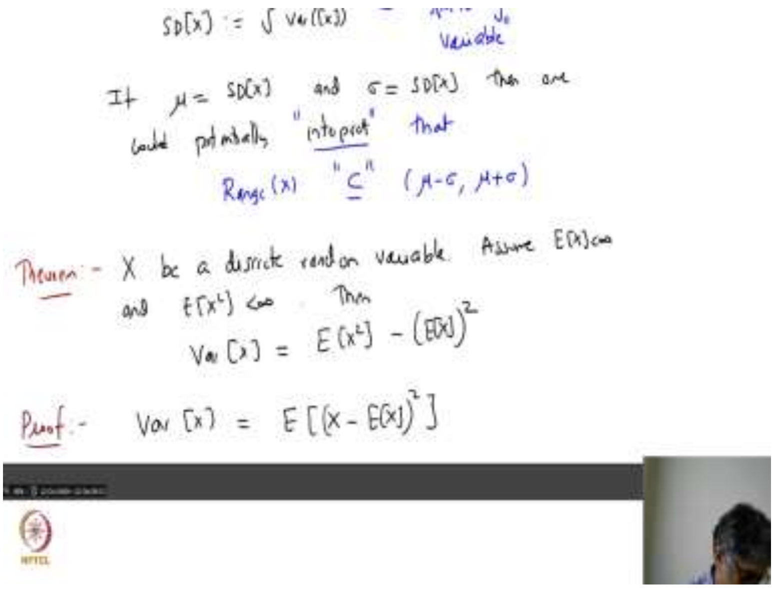
And then you take the average value of that that is what we call that the variance of this. So, this is called the variance of X and we notice one thing is that in this idea was that if X was in meters then the variance of X was not in meter square. Square and in that sense, it does not really determine the true spread of the random variable.

Of how far the random variable is from this mean on average so for that we had this notation we said we take standard deviation of X. Which is just the square root of the variance of X. And this we will know this we will use to understand what we call as the spread of random variable. And the idea being that that if if mu was equal mu was SD of X as a notation and let us say sigma was the standard deviation of X and as location as a symbol.

Then then one could one could potentially interpret let me do something with caution it is not quite a proper that the range of X is again it is a wrong idea if its contains in the interval mu minus sigma mu plus sigma. That was understanding of what a standard deviation actually means. It signifies the spread and you can use it to understand how far random variables must be used.

This is where we were last time. So, I wanted to sort of go a little bit further and just do a couple of quick properties of random variables or variances and then and then actually move on to ideas of how to understand many more things of a random variable. How do you understand how far it is mean? One way is the variance are the other ways can you compute the probability of the chance that X from its mean is away in a certain certain number.

(Refer Slide Time: 06:16)

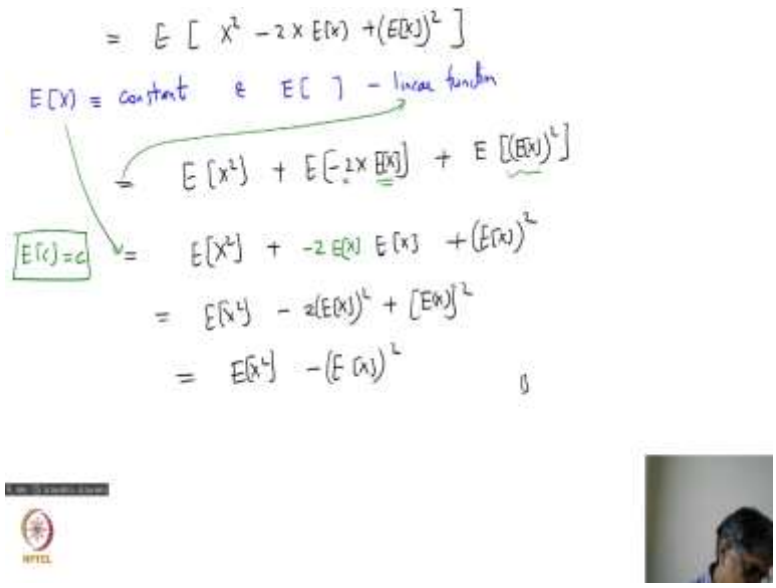


So, I will try and discuss all that so the first result I would like to sort of illustrate the following. So, like we discussed last time to compute the the the any functional you would just put g of t times probability X equal to complete E of g of X. And so for the variance you would put the function X minus the mean whole square. We did a computation last time. But today I want to sort of give you another interpretation for the variance.

Another simple calculation that is quite useful. So, X be a discrete random variable let us I assume E of X is finite and E of X square is finite. Let us they both are finite then then to compute the variance of X and you do not have to do your X minus here X whole square. It is just the same as E of X square minus E of X the whole square. So, this is a simple fact what I can easily show.

So, I will try and prove it today so let us try and do the provenance let us try to prove. Here is the proof of this. So, let us say here so what is the variance the variance of X is E of X minus E of X the whole square. This is what the variance of X was.

(Refer Slide Time: 07:56)



That is the same as E of how do you expand you would expand E of X minus E of X whole square that is X square plus minus 2 times X E of X plus E of X the whole square. That is what the that is what the the variance. So, if you expand this this order the expression right here you would just get X square 2 X E actually. But now you just think a little bit it is just a it is a simple computation.

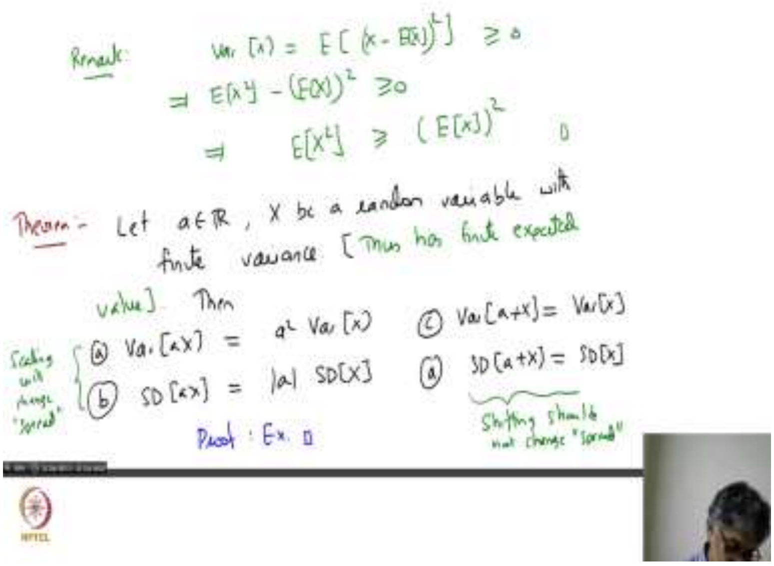
This here you notice that E of X is a constant, E of X is a constant it is not random. So, and and we have shown that E is linear. So, we have shown that we have shown earlier that expression is a linear function. So, now from this about the it is quite easy to see that this guy is going to be E of X square then plus E of first unified linear you have X minus 2 E X then plus E of E of X the whole square. That is what the linearity will give you.

So, that is what linearity gives you so let us let is get this. The linearity then then you use the fact (())(09:23). Then you define that E and X is a constant. So, first is E of X square that remains the same nothing happens. Here if you notice minus 2 is a constant E of X is a constant. So, you can remove here it is a constant and let us say minus 2 the constant. So, we know that E of a constant is is the constant itself.

This we know this fact you can just pull it out and say this is minus 2 E of X and what is left is E of just X as a Laplace so then plus here E of X is a constant this will be the constant so it just comes out. So, it is just going to be E of X. So, now all I will get is I get E X square minus here I get minus 2 times E of X the whole square. And here I add E of X the whole square.

So, that completes it. So, that gives you E of X square plus E of X E of X square minus E of X the whole square so that is the simple complication there. So, one important aspect of this is the following that since this variance is holding in variance so one one sort of a simple remark is the following that.

(Refer Slide Time: 11:01)



So, so variance of X is E of X minus E of X whole square. So, it is always non-negative because it is a square it is a non-negative sum. So, it is always non-negative that means we have shown that E of X square minus E of X the whole square is always non-negative. And from this about definition so we have an interesting idea that we something maybe I need to see.

If you take square of the expectation of the square that is always going to be larger than expected expectation and then square. This is sort of a interesting inequality it holds in much more generality but it is something that one could observe. Let me just quickly use one more thing before I recap the variance. So, one last fact which I will not prove about variance is the following (())(12:07).

This is the standard results. I will not prove it in class but I will understand. So, we all know that expectation is linear but variance is not linear variance square. So, so let us say let, let a be in R let X be a random variable with finite variance. And by the previous understanding thus has something you have to think about but it should be obvious if it has finite variance thus has finite exponent value.

Then a; the variance of a times X raised to a times X is a square times variance of X. So, the variance scales in squares (())(13:35). So, then then expectation was linear but the variance scales. Then a little careful this is also kind of obvious to you that if I take standard deviation of a X from the above it is square root of the variance and square the variance means square root of a square variance but for the square root out you have to be careful you get the mod a outside. So, you put mod a outside mod a times the variance of X. Sorry.

Mod a times the standard deviation. So, I think one can easily show and then variance does not change. Example, if I take a if I take a random variable and I shift it by a constant so then intuitively the the spread should not change. Because, E of X is linear so E of a plus X is a times a plus E of X. So, this is just it will just be the same as the variance of X. It will not change.

Similarly, again the substantial thing for standard deviation the spread should not change if I just move the random variable by approximating. So, there is some simple ideas that I will not show but I will leave the proof of exercise. But, the formal thing is that shifting should not change spread it should not change spread should not change spread that is what this signifies.

Here what signifies is that scaling will change spread. So, if you multiply by a constant that is like scaling random variable it will change spread. It will change spread in a square fashion for the radians you will change in a linear fashion further. So, let me just recap this discussion in double splits view go back and just check it and see if it comes very quickly.

So, just to recap variance and standard deviation and to understand how far random variable spreads out. And if you have a discrete random variable then the variance has an interesting expectation expression it is E of X square minus E of X the whole square. That is a simple computation and that is useful to do because all you have to do is compute the second movement which is called E of X squared.

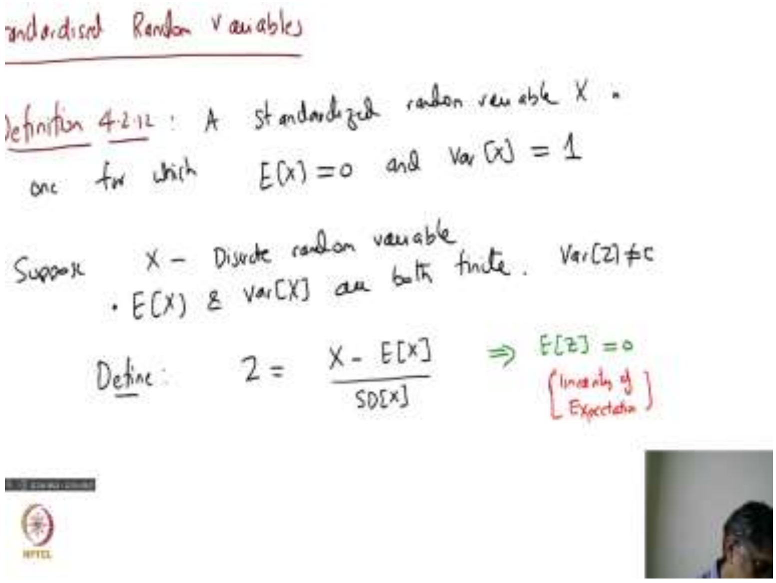
And E of X the whole square and once you do this you have a simple idea that variance of X is always non-negative. So, E of X square is always bigger than E of X the whole square so this part of a very general inequalities we will discuss that later in the course perhaps but it works from larger class of functions for the first square function it definitely works. As if we square random variable take its mean you will always get a larger expectation then if you take the expectation and square it.

Then there is this idea of how about how variance and standard deviation identify themselves with scaling and shifting. So, if you shift a random variable the spread should not change the length of the spread should be almost the same. So, that is exactly what happens variance of a plus X is variance of X. Standard deviation of a plus X is standard deviation of X. And if you scale the variance will change.

Because, the length of the interval changes but for the variance it will go by square for standard deviation will just change by the absolute value of the scale change. We just get back to next part. So, now we understand these ideas so variance is one sort of concept spread of random variables. And it tells you how far the random variables mean and so on an so forth.

So, now I want to something little different I want to I want to first think of random variables in a standardized form. So, what does that mean so let us just this is a very important concept in statistics but in probability also it is quite a it is a very understandable very useful notion if I am wrong here.

(Refer Slide Time: 19:06)



So, this is called standardized random variables. So, here is one thing so here what happens is that so first thing we we think of standardized means that the kind of the spread is 1 and the mean is 0. That is the first definition let me write that down. The first definition let me call this section itself as 4.2.3. So, its definition is 4.2.12. So, we standardized it.

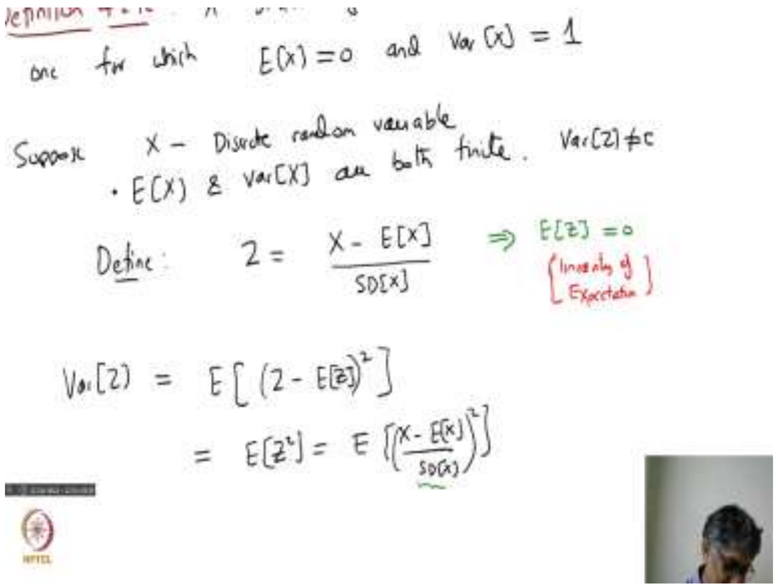
A standardized random variable so X is one for which E of X is equal to 0. And the spread so I can use variance of standardising. Because, I want to say the variance of X is 1. So, standard random variable is one in which the variance is is 1 and the mean is 0. So, why do I why do I say this is standardized. I will come to it in a second. Now, this is important you can you can always scale random variable to make it standardized.

So, how do I do that so using the previous theorem that I said. So, so what you do is suppose Z is another a discrete random variable. And let us say E of Z and variance of Z are both finite. And let us say variance of Z is not 0. So, here what happens is that now maybe I will I will use notation X.

So, let X be random variable let us say E of X is some number then (()) (21:42). Then these are both numbers now. So, what I do is I define Z to be X minus E of X. So, one thing you can just think a little bit if I if I just subtract the mean from E of X the mean of Z should be 0. Because, equation is linear and just pull the expression across you will get linear. So, this will give you that the mean is linear the mean is 0. So, mean of Z is 0.

But, now what I do is I let me write this in green observation so mean that is that is because expression is meaning. But, now what I do is variance of Z is still not 1 so but I can scale it. So, what I do is scale it by the standard deviation. Now, scaling also will we know that E of a X is a times E of X so this pulls out E of z is 0. If I scale by something drops to 0. So, this remains is equal to c. So, this is linearity of expectation gives you this. Now, let us check the variance. So, what is the variance of that?

(Refer Slide Time: 23:21)



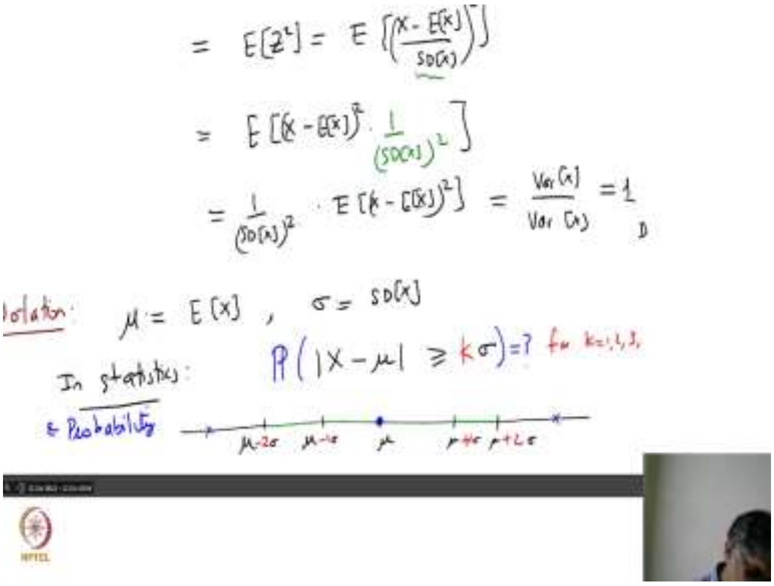
The variance of Z we we know in in many ways is just is just you can do it in in any way you want let us do it the same way as before. Is variance over its variance of Z is just again let us say expected value of Z minus E of Z the whole square that is what the variance of X. But, we know E of Z is 0. So, this is just E of Z square. But, what is that square that is just E of X minus E of X by SD of X the whole square. But, SD of X is a constant.

So, that will just come out as a square. So, this is a constant this is a constant. So, what happens it will just come out of the square so maybe you can write it properly. So, that is just E of at the top I have X minus E of X whole square. And I can think of this as 1 divided by SD of X the whole square. That is just a constant for explanation the constant just comes out so I get 1 over SD of X the whole square and then I have E of X minus E of X the whole square.

But, the top part is just variance of X and SD square is again variance of X. So, I just get 1. So, this one way of standardizing a random variable. You can you can show that variance of Z is 1. But, so this is kind of a very useful fact and probability that if the variance of our mean of finite then you can standardize to variance 1 by shifting it by the mean and dividing it by standardizing.

It is a very useful fact. So, once we are in standardized units it helps because then you can always think of it as a you could think of if you say mean is mu mean is 12 sigma is 5 then you know exactly how the spread so and so forth. But, if you standardize it then you know exactly between 0 and 1 you know. So, let me try and understand how to use the standardized units. So, from now on so here is notation from now on notation.

(Refer Slide Time: 26:29)



Here is mu we will always think mu as E of X whenever E of X is defined mu will be notation. And sigma will denote standard deviation so whenever I say sigma I mean standard deviation whenever I have E of X I mean. So, now if I have mu and X max so in statistics the one is these two numbers in a very very very clear and specific way like I said if sigma if we use some number sigma is a number then mu minus sigma and mu plus sigma is kind of the golden idea of how the random variable behaves.

So, in statistics one is always interested in understanding how far is X from the mean. So, if you look at X minus mu that is a distance from mean. You look at its absolute value that is the distance from the mean. How many multiples of it is it from standard deviation? So, what is interested in understanding how, how far it is from sigma sometimes that is is it larger than sigma is a smaller sigma that is how many points the random variables are outside this interval mu minus sigma w plus sigma or and sometimes they understand it in terms of multiple signals.

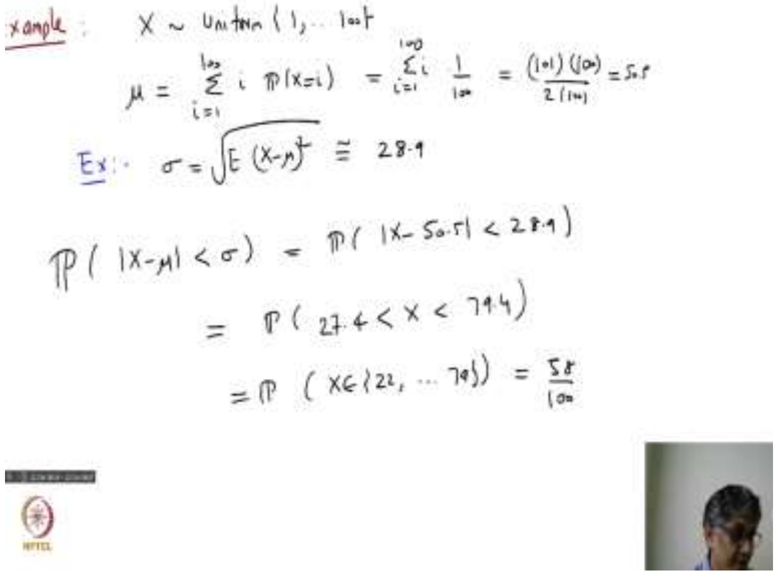
So, let us say k for some k being equal to k equal to 1 2 3 and so on. So, that is the idea is that you have the number mu number line here the number mu and then you have these points here this is mu minus sigma mu plus sigma. This number is again you go a little bit further down it will be minus 2 sigma plus 2 sigma. So, you are interested in understanding how far the random variable is from the distance mu. So, are there points outside mu minus sigma and mu plus sigma?

So, in statistics often what will happen is that one will be trying to understand this is the mean we understand that are there points here. Are the random variables taking positive values very high probability outside this range. Somehow, would be for the spread be so large that sigma and mu do not really capture the support. This is one interesting idea. So, one always interested is the fact is what is the chance that X is outside this what is this chance. This is a very important question in statistics and in any application.

Even probability if you want to understand if a random variable is close to is mean or not one tries to understand. So, how far a random variable is from spread? And one way of understanding it is is with regard to variance because we know that X minus mu by sigma is kind of standardized. So, so you want to understand if the random variable is far from its mean in terms of multiple design.

So, this is something that one can sort of try and see so let us see let us try and see how let us try and see let us try and give an example to understand this question. So, here is an example.

(Refer Slide Time: 30:09)

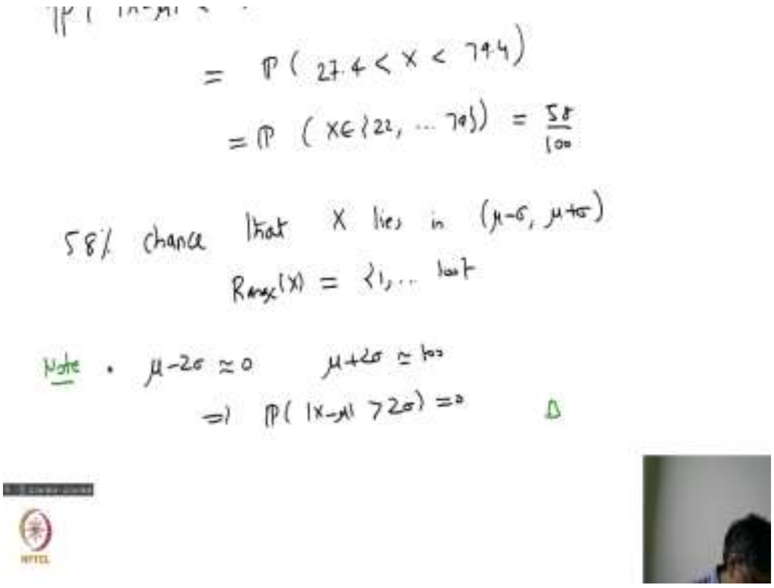


Suppose, X is uniform let us say 1 up to 100 this example. So, then let us say I want to understand let us let is do me what does mean mean mu is again you can just the computation is just the sum over i equal to 1 to 100 i times the chance that X equal to i. And that is going to be i summation over i equal to 1 to100 and chance is 1 over 100. And that is just 101 into 100 divided by 2 times 100 and that gives you some 50.5.

So, I will leave it as an exercise for you to check that that sigma which is a value expression of X minus mu the whole square square root that is going to be the same as 20 approximately 28.9. That is something you can check. So, now if I if I look at this probability that X minus mu it is looking the the let us say X minus mu is within a sigma. I want to know the chance that the random variables range is like mu minus sigma to mu plus sigma.

So, that is the same as saying the chance of X minus mu is less than sorry X minus mu is our 50.5 is less than sigma just 28.9. So, if I do the computation that is the same as X is between here I get on that side I get let us say 79 plus 5 is 14 79.4 on this side. On this side I will get that 14 minus 9 again 4 and this is like around 22. So, it is like a 9 minus 8 is 27(())(32:28). Something like that. So, as saying the chance that X is between 22 to 79. And now this we know is 58 by 100. That means the 58 chance that X lies between the intervals minus 1.

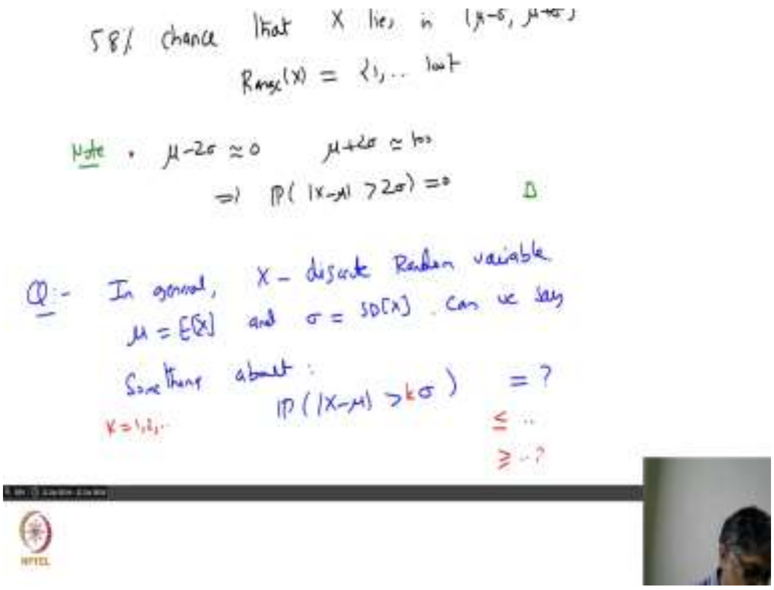
(Refer Slide Time: 32:53)



58 percent chance that X lies in mu minus sigma to mu plus sigma. And but then we know we know the range of X is is actually from 1 to 100. So, that means we clearly know that if you take mu to be use mu minus 2 sigma you know mu minus 2 sigma in this case mu minus 2 sigma is negative. Because, sigma is 28 you know sigma is like close to 0 it is close to 0. And mu plus 2 sigma is close to 100.

So, that means this will imply that the chance that X minus mu is bigger than 2 sigma is our front is actually 0. So, these are two observations one can make if you lose computation for uniform it is within sigma is like 58 with about two sigma the probability is different. So, now the next step I will try and do is can I can I generalize this. So, can I understand if I give you without doing computations and do the following can we generalize this?

(Refer Slide Time: 34:26)



The question is in general X discrete random variable mu is expectation and sigma sigma standard deviation of X. I give you this. Can we say something about all I give you this distribution of X mu and sigma of X without computation can I say something about X minus mu is bigger than sigma.

Or bigger than sigma is what. And then in general let us say k is k is 1, 2 and 3. If I put a k here what will I can I say can I can I give any quality can I say it is smaller than something can I say it is larger than something what can I say. These are very powerful tools both in statistics and probability we will try and understand that in next class.