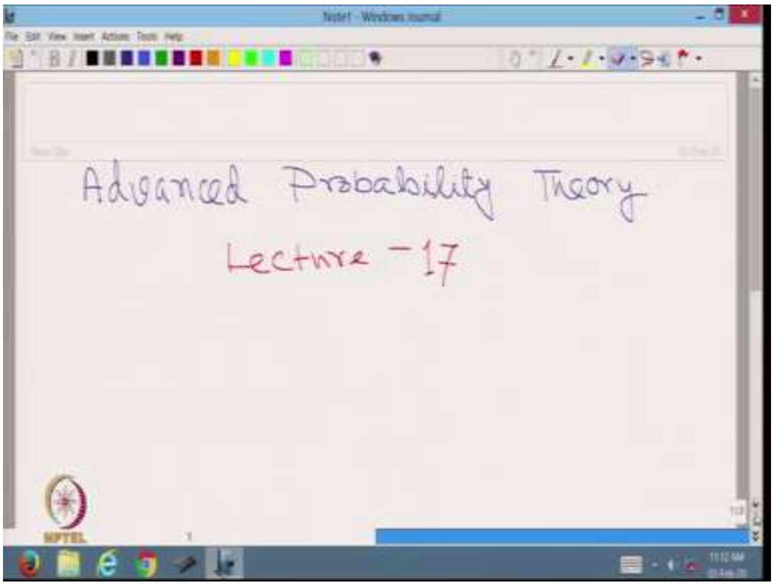
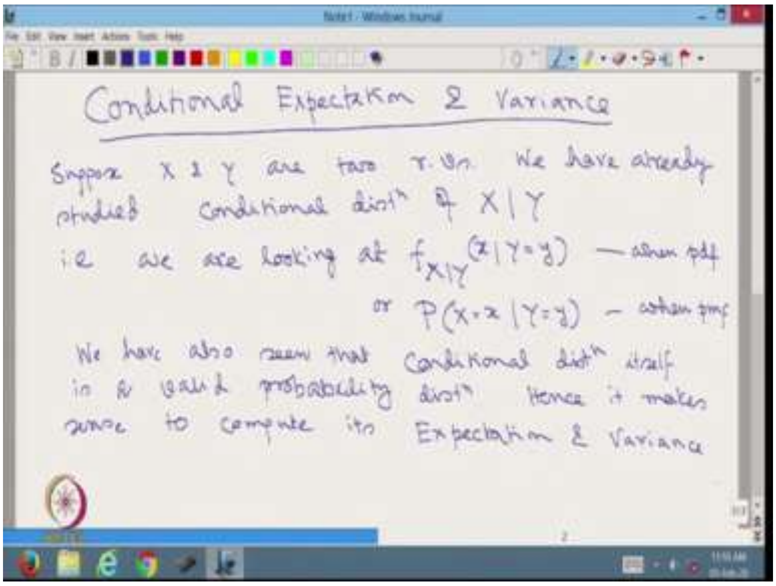
Advanced Probability Theory
Professor Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology Delhi
Lecture 17

(Refer Slide Time: 00:27)



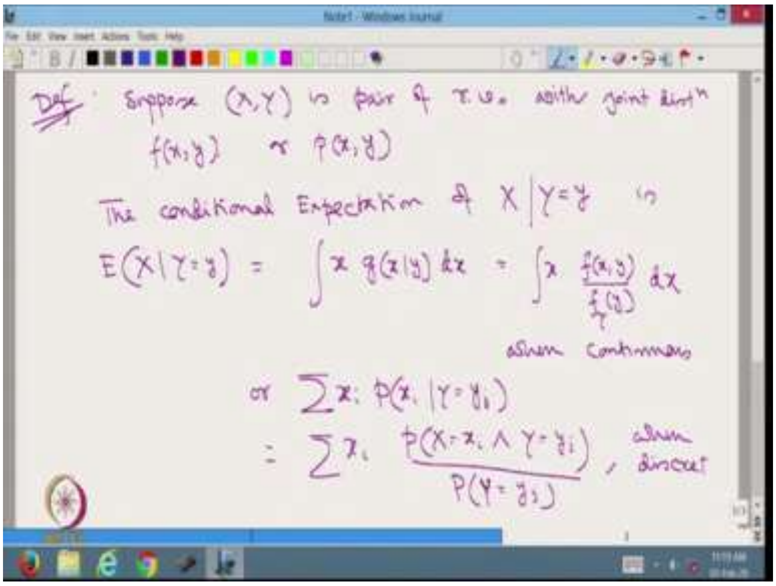
Welcome students to the MOOC series of lecture on Advanced Probability Theory. This is lecture number 17.

(Refer Slide Time: 00:32)



As I said in the last class that today we shall start with Conditional Expectation and Variance. Suppose x and y are two random variables, we have already studied conditional distribution of X given Y, that is, we are looking at f of x given that Y is equal to y, when it is a pdf or probability X is equal to x given that Y is equal to y when pmf. We have also seen that conditional distribution itself is a valid distribution. Hence it makes sense to compute its expectation and variance.

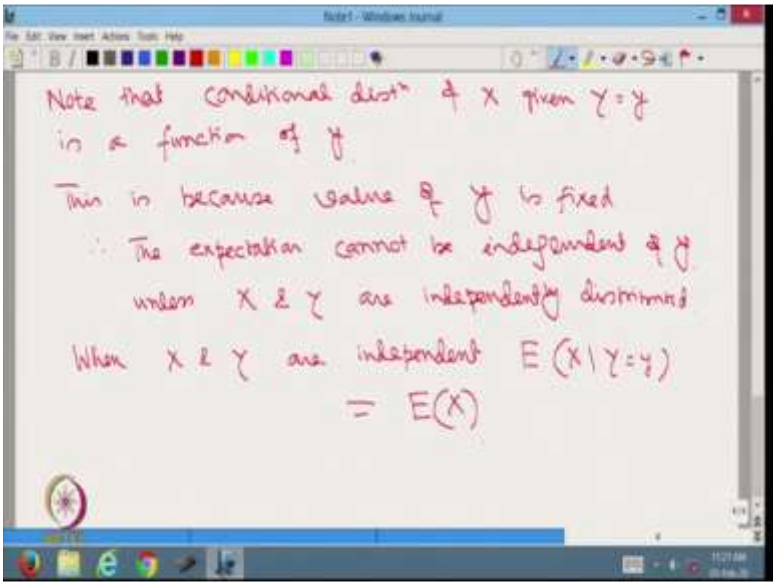
(Refer Slide Time: 03:09)



So definition, suppose x, y is a pair of random variables with joint distribution fxy or pxy when discrete, therefore the conditional expectation of X given Y is equal to y is the expected value of X given Y is equal to y is equal to integration of x g of X given Y dX is equal to x into f x y upon f y of y dx when continuous.

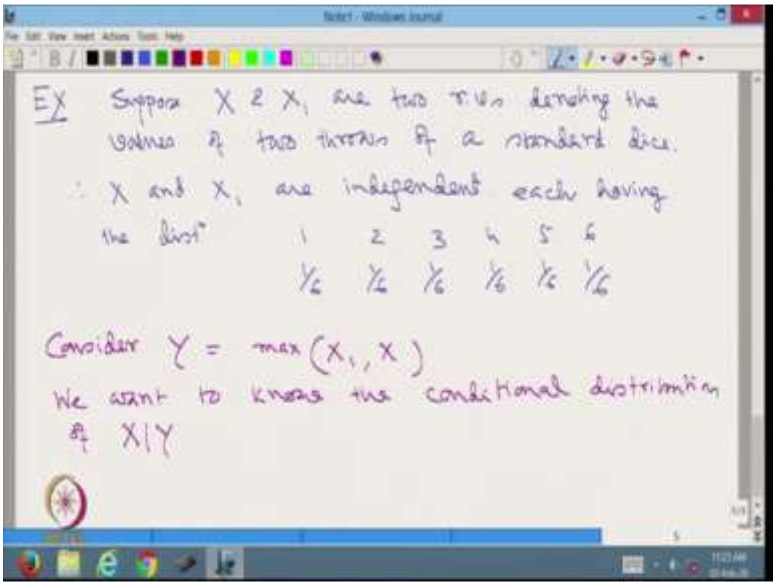
Or sigma xi into probability of xi given y is equal to y j is equal to sigma xi probability x is equal to xi and Y is equal to y j divided by probability of Y is equal to yj when discrete.

(Refer Slide Time: 05:18)



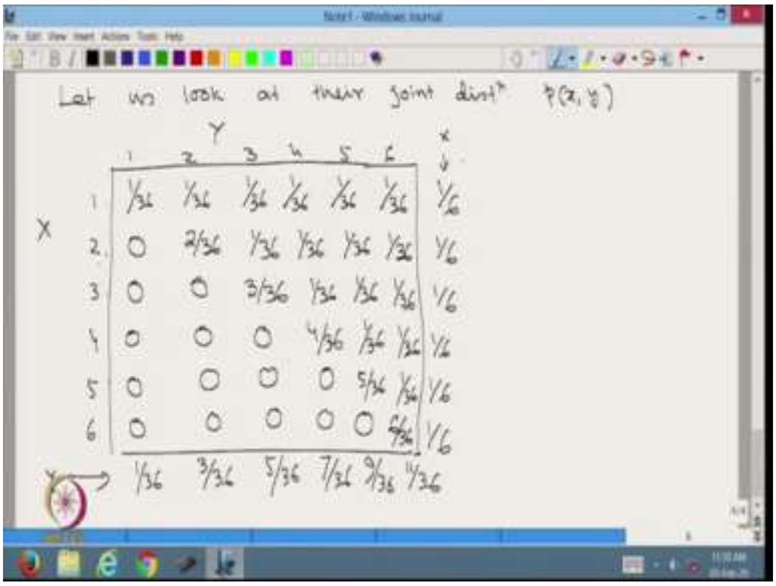
Note that conditional distribution of X given Y is equal to y is a function of y. This is because value of y is fixed. Therefore, the expectation cannot be independent of y unless x and y are independently distributed. When x and y are independent, expected value of X given Y is equal to y is same as expected value of x that should be very clear. Let us start with an example.

(Refer Slide Time: 07:06)



Suppose x and x1 are two random variables denoting the values of 2 throws of a standard dice. Therefore, x and x1 are independent, each having the distribution 1, 2, 3, 4, 5, 6 with probabilities 1 by 6, 1 by 6, for all the 6 values. Consider y is equal to maximum of x1 and x. We want to know the conditional distribution of X given Y.

(Refer Slide Time: 09:00)



So, let us look at their joint distribution p x, y. As we can understand, y takes values 1, 2, 3, 4, 5 and 6 and x also takes values 1, 2, 3, 4, 5 and 6. Note that when x is equal to 1, x1 can be 1, 2, 3, 4, 5 and 6, any one of them and value of y will depend upon the probability of x1, taking that particular value.

So, it is very easy to see that these probabilities are coming out to be 1 by 36, 1 by 36, 1 by 36, 1 by 36, 1 by 36 and 1 by 36, this is because, suppose, what is the probability that x is equal to 1 and y is equal to 4? That means, probability of x taking the value 1 and x1 taking the value 4 therefore, because they are independent, so, 1 by 6 into 1 by 6, that is 1 by 36. Therefore, marginal distribution of x is 1 by 6 for x is equal to 1. Now, suppose x is equal to 2, therefore y cannot be less than equal to 2.

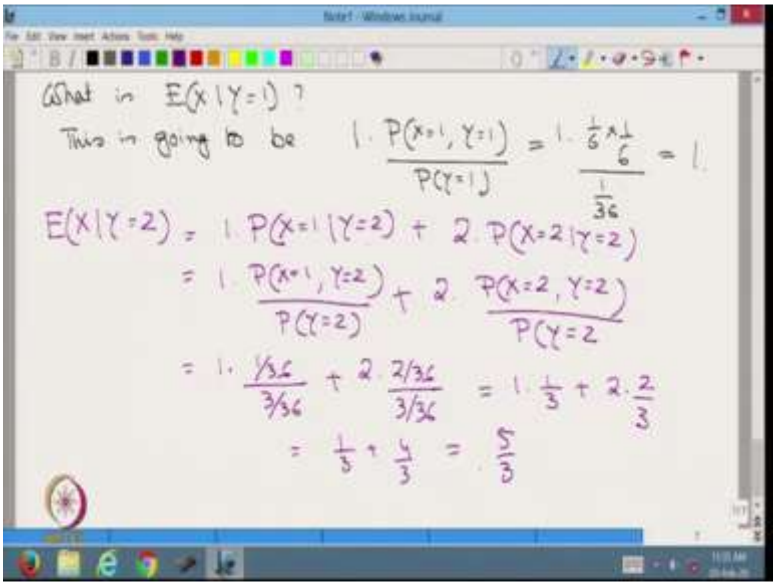
Therefore, probability of y is equal to 1 is 0. Probability, y is equal to 2 can happen if x is 1, or x is 2. Therefore this is going to be 2 by 36, y is going to be 3. If x is 2, and y is 3, therefore 1 by 36, y is going to be 4, when x is equal to 2, if x1 is equal to 4, therefore, 1 by 36, 1 by 36, 1 by 36. When x is equal to 3, y cannot be 1, y cannot be 2.

Now consider x is equal to 3 and y is equal to 3. This can happen in 3 different ways when x is equal to 3 and x1 is equal to 1, x is equal to 3, x1 is equal to 2, and x is equal to 3 and x3 is equal to 3. In all these cases, the maximum will be 3. Therefore, this is 3 by 36, y can be 4 if x is equal to 3 and x1 is 4, therefore this is 1 by 36, 1 by 36, similarly 1 by 36.

Convince yourself that these probabilities will come out to be like this, x is 5 and y is, anything less than or equal to 4 is not possible, but y can be 5 when x1 is any of 1, 2, 3, 4, 5, therefore 5 by 36, 1 by 36 and x is 6, then for all possible values of x1, we get y is equal to 36. Therefore, we can see that x and y are dependent on each other, and the marginal distributions for x are to be 1 by 6, 1 by 6 for all the 6 values.

However, the marginal distribution of y is changing. In this case, it is 1 by 36, in this case, it is 3 by 36, in this case, it is 5 by 36, in this case it is 7 by 36, 9 by 36 and 11 by 36. So this is marginal distribution of y. This is marginal distribution of x.

(Refer Slide Time: 14:20)

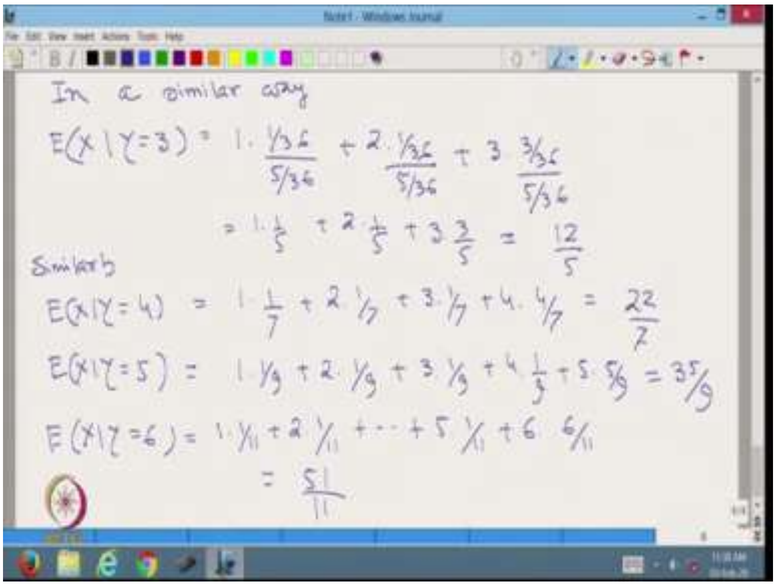


What is the expectation of X given Y is equal to 1? Quite naturally, this is going to be 1 in 2, probability x is equal to 1, y is equal to 1 divided by probability y is equal to 1, is equal to 1 times 1 by 6 into 1 by 6 upon 1 by 36 is equal to 1.

Expected value of X given Y is equal to 2 is equal to, so we are looking at probability x is equal to 1, and y is equal to 2 is equal to when x is equal to 1, probability of x is equal to 1 given y is equal to 2 plus 2 times probability x is equal to 2, given y is equal to 2 is equal to 1 times probability x is equal to 1, and y is equal to 2 divided by probability y is equal to 2 plus 2 times probability x is equal to 2, and y is equal to 2 divided by probability y is equal to 2.

Is equal to 1 times 1 by 36 divided by 3 by 36 plus 2 times 2 by 36 upon 3 by 36 is equal to 1 times 1 by 3 plus 2 times 2 by 3 is equal to 1 by 3 plus 4 by 3, is equal to 5 by 3.

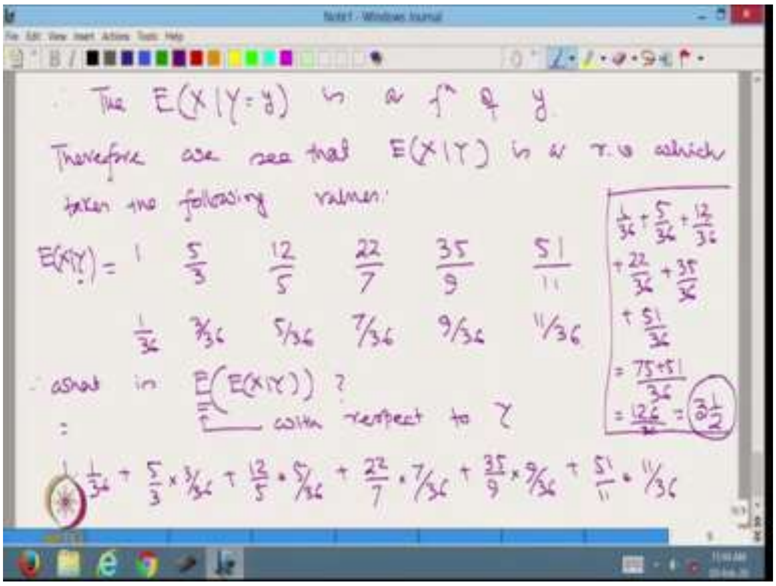
(Refer Slide Time: 16:59)



In a similar way, expected value of X given Y is equal to 3 is equal to, 1 into, 1 by 36 upon 5 by 36 plus 2 into 1 by 36 upon 5 by 36 plus 3 into 3 by 36 upon 5 by 36 is equal to 1 into 1 by 5 plus 2 into 1 by 5 plus 3 into 3 by 5 is equal to 12 by 5. Similarly, expected value of X given Y is equal to 4 is going to be 1 into 1 by 7 plus 2 into 1 by 7 plus 3 into 1 by 7 plus 4 into 4 by 7 is equal to 22 by 7.

The expected value of X given Y is equal to 5 is equal to 1 into 1 by 9 plus 2 into 1 by 9 plus 3 into 1 by 9 plus 4 into 1 by 9 plus 5 into 5 by 9 is equal to 35 by 9 and finally expected value of X given Y is equal to 6 is equal to 1 into 1 upon 11 plus 2 into 1 upon 11 plus up to 5 into 1 upon 11 plus 6 into 6 by 11 is equal to 51 by 11.

(Refer Slide Time: 19:33)

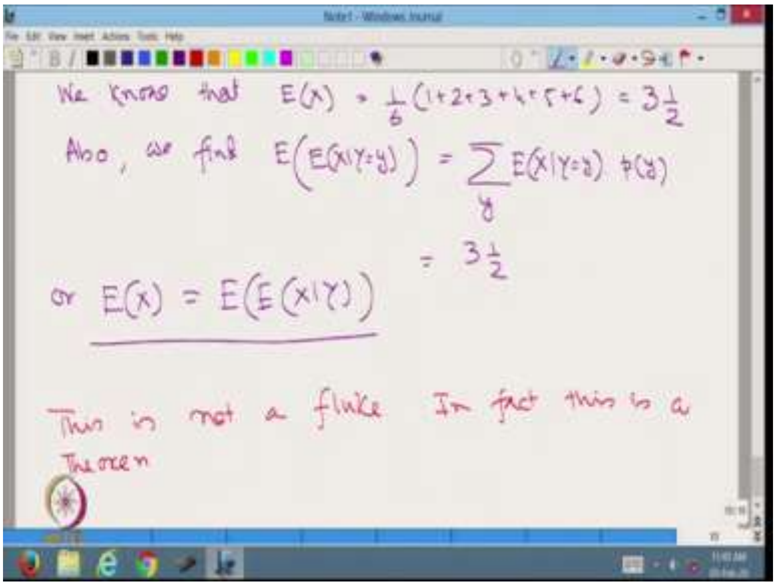


So, we have seen that the expected value of X given Y is equal to y is a function of y. Therefore, we see that expected value of X given Y is a random variable which takes the following values 1, 5 by 3, 12 by 5, 22 by 7, 35 by 9 and 51 by 11. And these are the values and the probabilities for them is given by what is the probability that y is taking that particular value.

Therefore, these probabilities are 1 by 36, 3 by 36, 5 by 36, 7 by 36, 9 by 36 and 11 by 36. Therefore, what is expected value of expected value of X given Y? That is a very interesting question. This is equal to, note that this expectation is with respect to y.

Therefore, this we can write it as 1 into 1 by 36 plus 5 by 3 into 3 by 36 plus 12 by 5 into 5 by 36 plus 22 by 7 into 7 by 36 plus 35 by 9 into 9 by 36 plus 51 by 11 into 11 upon 36 which is coming out to be 1 by 36 plus 5 by 36 plus 12 by 36 plus 22 by 36 plus 35 by 36 plus 51 by 36 is equal to 1 plus 5, 6, 6 plus 12, 18, 18 plus 12, 40. 40 plus 35 is equal to 75 is equal to 75 plus 51 upon 36 is equal to 126 upon 36 is equal to 3 and a half. So that is a very interesting thing.

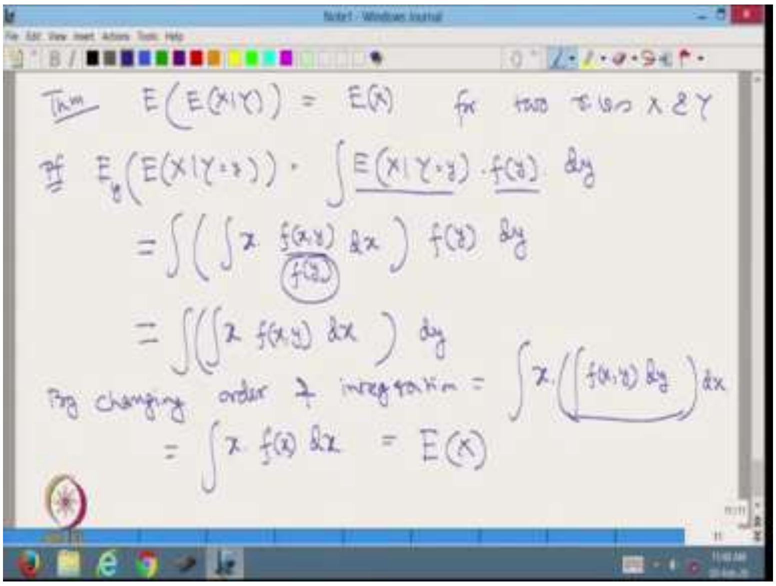
(Refer Slide Time: 23:16)



We know that expected value of x is equal to 1 by 6 into 1 plus 2 plus 3 plus 4 plus 5 plus 6 is equal to 3 and half. Also, we find the expected value of X given Y is equal to y, which is sigma over y, expected value of X given Y is equal to y into probability of y is equal to 3 and a half or the expected value of x is equal to the expected value of X given Y.

This is an interesting observation from this example. But this is not a fluke. In fact, this is a theorem. What is that?

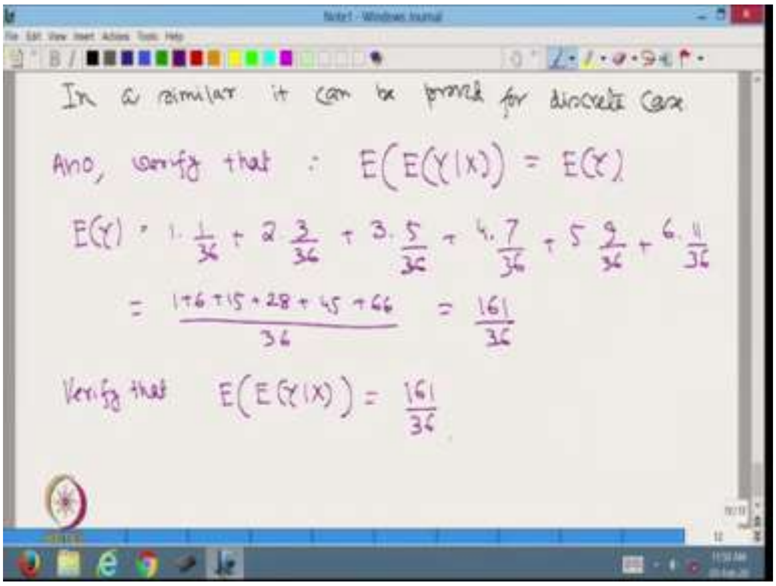
(Refer Slide Time 24:45)



The theorem is that expected value of X given Y is equal to expected value of x for 2 random variables, x and y. Proof, expected value with respect to y of the expected value of X given Y is equal to y is equal to integration of expected value of X given Y is equal to y multiplied by fy into dy, this is the value of the random variable, this is the probability is equal to integration of x f x, y upon fy dx multiplied by fy dy is equal to, since this is independent of x, we can take out of the integration.

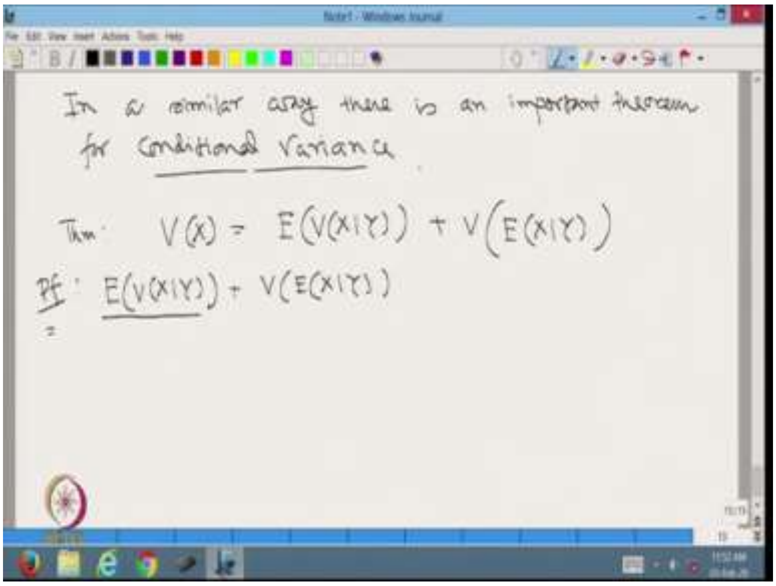
So, we can write it as integration of x, f x, y dx into dy. By changing order of integration is equal to integration of x integration of f x, y dy dx is equal to, now this integration gives us marginal density of x. Therefore, this is going to be x into the fx dx is equal to expected value of x.

(Refer Slide Time 27:07)



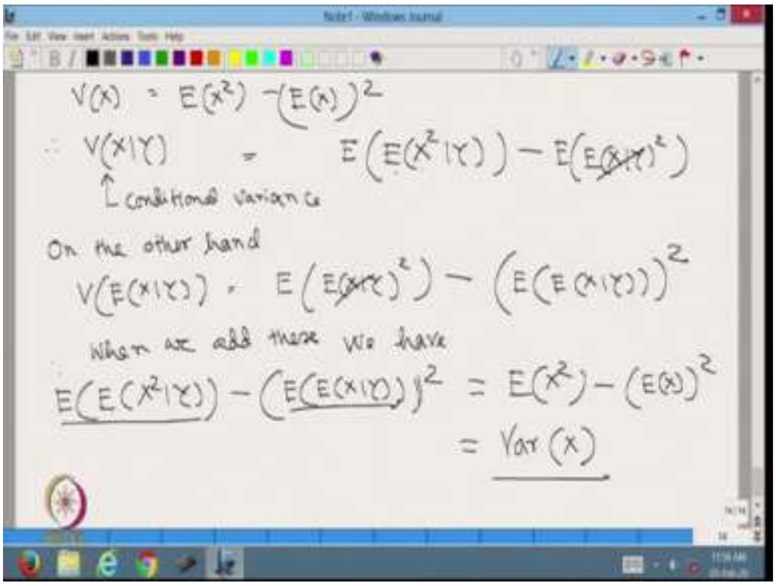
In a similar way, it can be proved for discrete case. I leave it as an exercise. Also verify that expected value of Y given X is equal to expected value of Y. Expected value of Y is we know that 1 into 1 upon 36 plus 2 into 3 upon 36 plus 3 into 5 upon 36 plus 4 into 7 upon 36 plus 5 into 9 upon 36 plus 6 into 11 upon 36 is equal to 1 plus 6 plus 15 plus 28 plus 45 plus 66 upon 36 is equal to 161 upon 36. Verify that expected value of Y given X is equal to 161 upon 36.

(Refer Slide Time: 29:03)



Now, in a similar way, there is an important theorem for conditional variance. I want you to remember the result. The result is that variance of x is equal to expected value of variance of X given Y plus variance of expected value of X given Y. Proof, expected value variance of X given Y plus variance of expected value of X given Y is equal to, so, let us look at the first part.

(Refer Slide Time: 30:18)

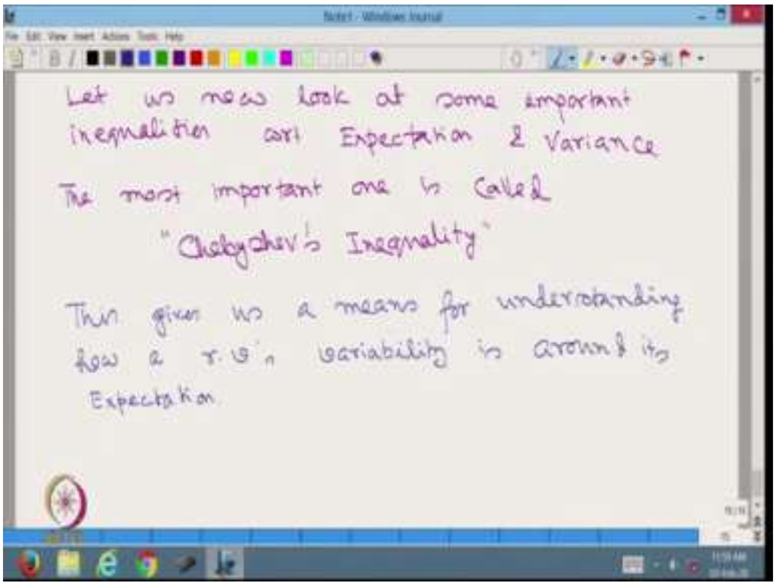


We know that variance of x is equal to expected value of x square minus expected value of x whole square. Therefore, variance of X given Y, that is the conditional variance is equal to expected value of, expected value of X square given Y conditional expectation of X square minus expectation of, expectation of X given Y whole square.

On the other hand, variance of expected value of X given Y is equal to expected value of, expected value of X given Y whole square minus expected value of expected value of X given Y whole square. Therefore, when we sum them, we have, this cancels with this. Result is expected value of expected value of X square given Y minus expected value of , expected value of X given Y whole square.

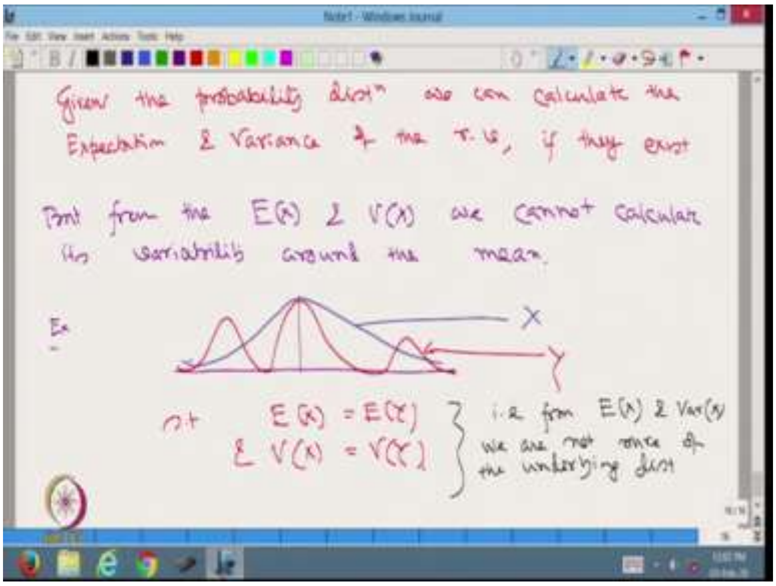
Now, this by our earlier result is the expected value of x square. As we have seen that expected value of a random variable conditioned in y its that expectation is the expectation of the original random variable minus this is expectation of X. Therefore, expectation of x whole square which is equal to variance of x. Thus we prove another important result with respect to conditional distribution. So, these are some important results to remember.

(Refer Slide Time: 33:10)



Let us now look at some important inequalities with respect to the expectation and variance. The most important is called Chebyshev's Inequality. This gives us a means for understanding how a random variables variability is around its expectation. Let me illustrate.

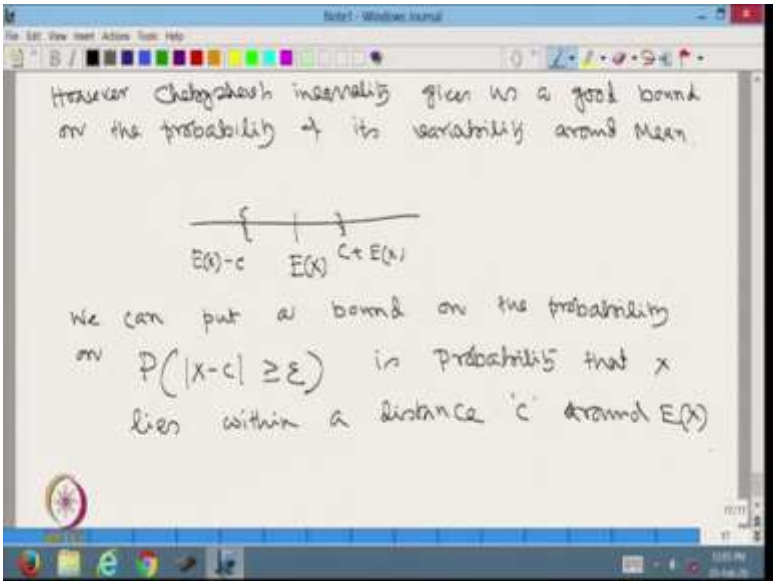
(Refer Slide Time: 34:51)



Given the probability distribution, we can calculate the expectation and variance of the random variable if they exist. But from the expectation of x and variance of x, we cannot calculate its variability around the mean. For example, suppose a random variable has this expectation but suppose its distribution is something like this.

Suppose this is for x, there can be another random variable, say y, which has the following distribution such that the expectation of x is equal to expectation of y and variance of x is equal to variance of y that is from E X and variance of x, we are not sure of the underlying distribution.

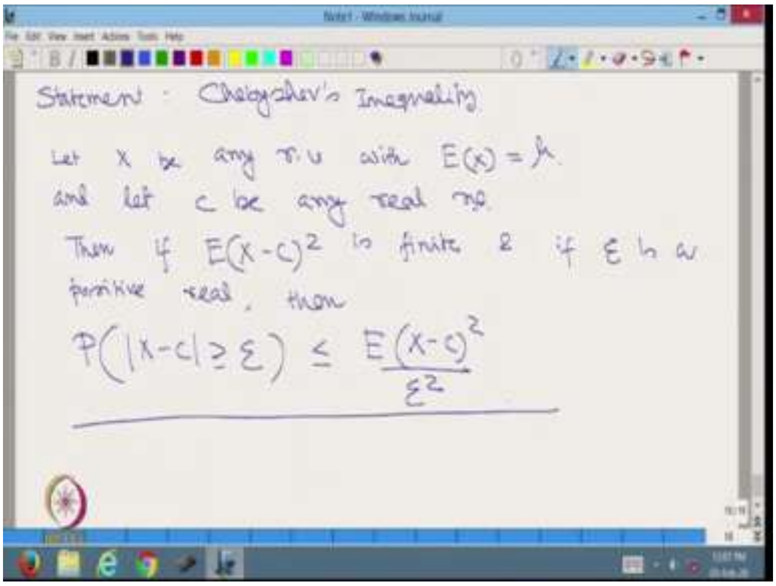
(Refer Slide Time: 37:36)



However Chebyshev's inequality gives us a good bound on the probability of its variability around mean that is expectation. Say, for example, suppose this is the expected value of x, we consider an interval around its expectation.

We can put a bound on the probability on the probability, on probability modulus of x minus c greater than or equal to epsilon, that is probability that x lies within a distance C of or around the expected value of x.

(Refer Slide Time: 39:52)

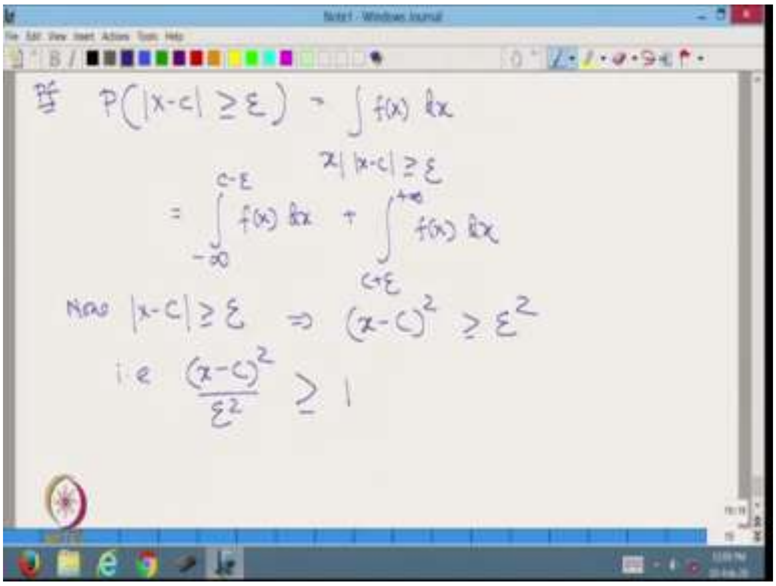


So here is the theorem of Chebyshev's inequality. Let x be any random variable with the expectation of x is equal to mu and let C be any real number, then if expectation of X minus

C whole square is finite and if epsilon is a positive real, then probability modulus of x minus c greater than equal to epsilon is less than equal to the expected value of x minus c whole square upon epsilon square. So, that is the result.

So, this gives an upper bound on how much is the probability that x lies at a distance more than epsilon around C.

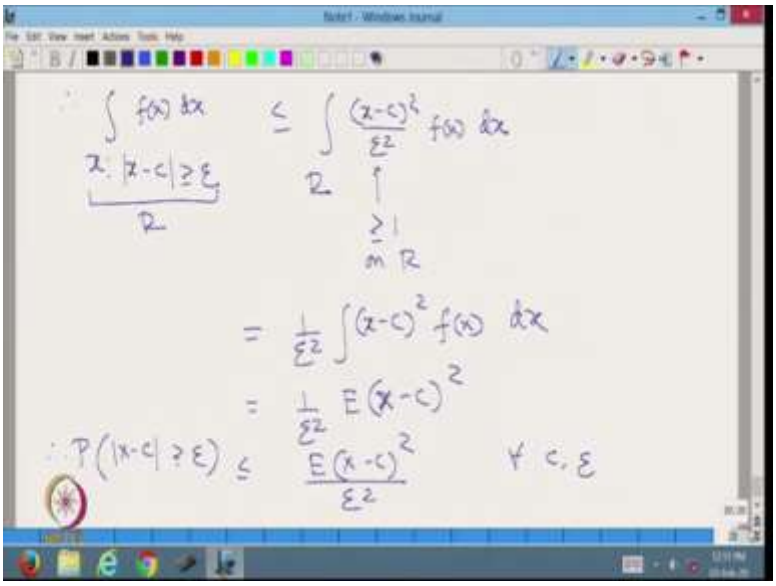
(Refer Slide Time: 41:36)



Proof, probability modulus of x minus c greater than equal to epsilon is equal to integration of fx dx on the space of x such that modulus of x minus c is greater than equal to epsilon which is is equal to minus infinity to C minus epsilon fx dx plus c plus epsilon to plus infinity of fx dx.

Now, modulus of x minus c greater than equal to epsilon implies x minus c whole square is greater than equal to epsilon square that is x minus c whole square upon epsilon square is greater than or equal to 1.

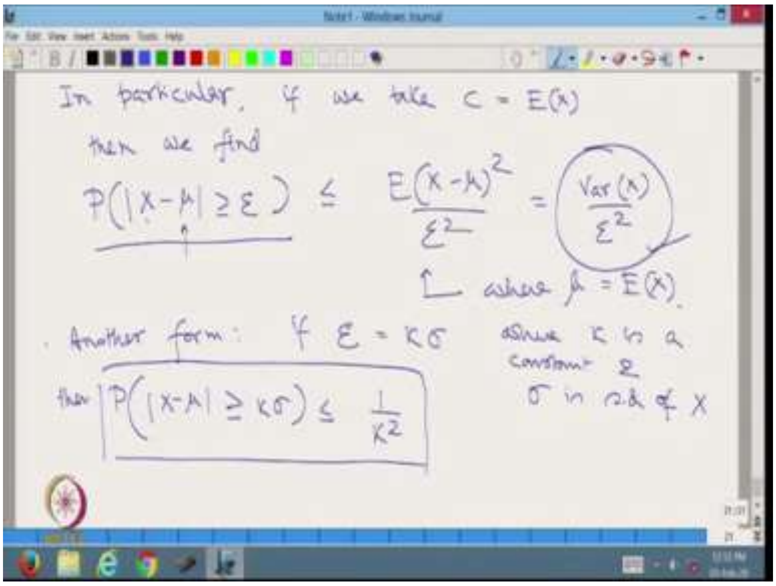
(Refer Slide Time: 42:54)



Therefore, integration of x such that modulus of x minus c greater than or equal to epsilon, fx dx, let us call it R, less than equal to integration over R x minus c whole square upon epsilon square fx dx, because this quantity is greater than or equal to 1 upon R, which is equal to 1 upon epsilon square integration of x minus c whole square fx dx, which is equal to 1 upon epsilon square expectation of x minus c whole square.

Therefore, probability modulus of x minus c greater than equal to epsilon is less than equal to the expected value of x minus c whole square upon epsilon square for all C, epsilon.

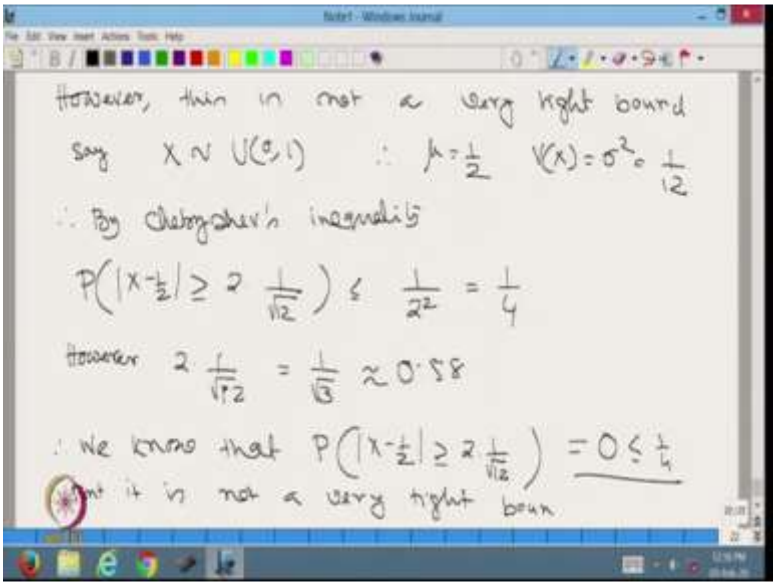
(Refer Slide Time: 44:19)



In particular, if we take C is equal to the expected value of x, then we find probability modulus of x minus mu greater than equal to epsilon is less than equal to the expected value of x minus mu whole square upon epsilon square is equal to variance of x upon epsilon square where mu is equal to expected value of X.

Thus, we can put a bound that the random variable x is at a distance greater than epsilon from its expectation mu that is bounded by this probability variance of x upon epsilon square. Another form, if epsilon is equal to k sigma, where k is a constant and sigma is standard deviation of x, then probability modulus of x minus mu greater than equal to k sigma is less than equal to 1 upon K square by putting epsilon is equal to k sigma. This is another form of using Chebyshev’s inequality.

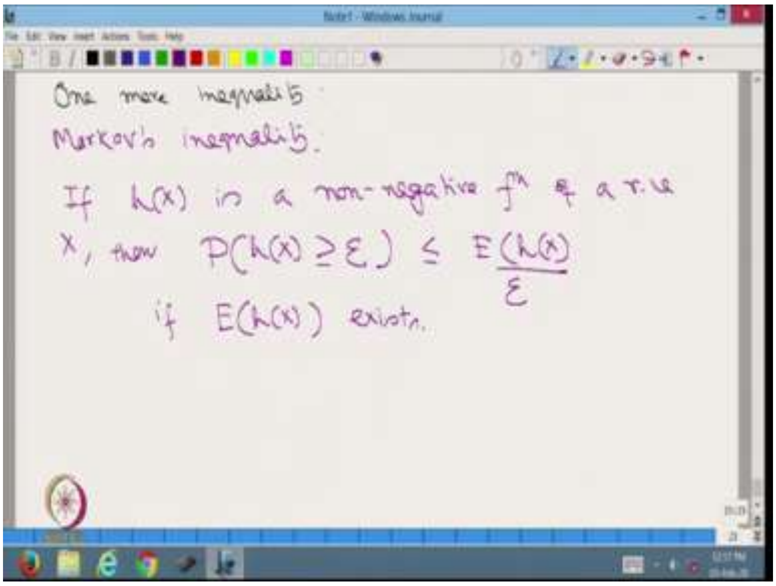
(Refer Slide Time: 46:26)



However, this is not a very tight bound. Say x is uniform 0, 1, therefore, mu is equal to half, variance of x is equal to sigma square is equal to 1 upon 12. Therefore, by Chebyshev's inequality, probability modulus of x minus half greater than equal to 2 times 1 upon root 12 is less than equal to 1 upon 2 square, which is equal to 1 by 4.

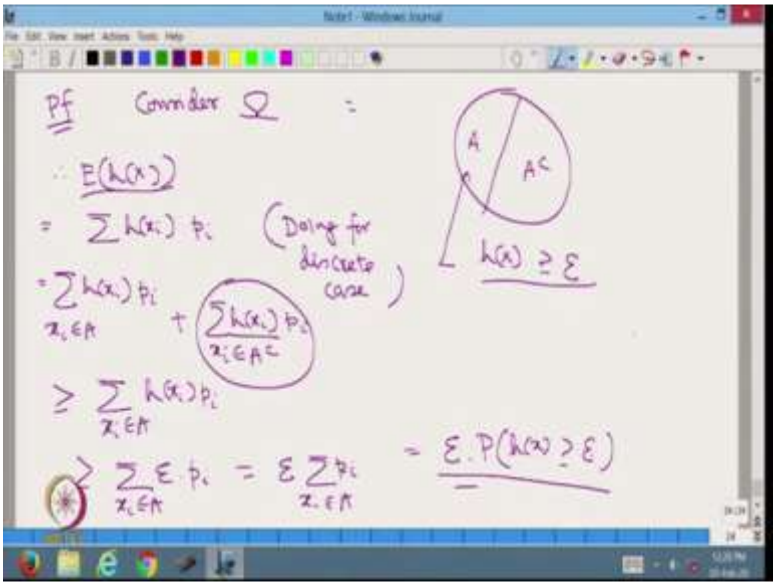
However, 2 into 1 upon root 12 is equal to 1 upon root 3 is equal to roughly 0.58. Therefore, we know that probability modulus of x minus half greater than equal to 2 upon root 12 is equal to 0. Of course this is less than or equal to 1 by 4, but it is not a very tight bound.

(Refer Slide Time: 48:27)



One more inequality this is called Markov's inequality, which states that if hx is a non-negative function of a random variable x, then probability h of x greater than equal to epsilon is less than equal to the expected value of h of x by epsilon if expected value of hx exists.

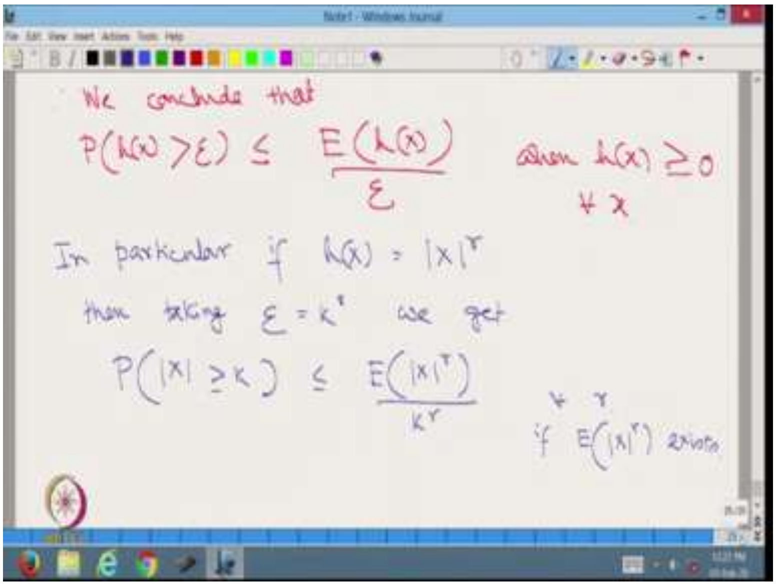
(Refer Slide Time 49:31)



Proof, consider the omega is equal to like this. Suppose we divided it by 2 parts, A and A complement, A is such that h of x is greater than equal to epsilon. Therefore, expected value of h of x is equal to sigma h of xi multiplied by pi doing for discrete case. You can do similarly for non-discrete case, is equal to sigma over h of xi into pi xi belonging to A plus sigma over h of xi into pi xi belonging to A complement, which is greater than equal to sigma over xi belonging to A, h of xi into pi.

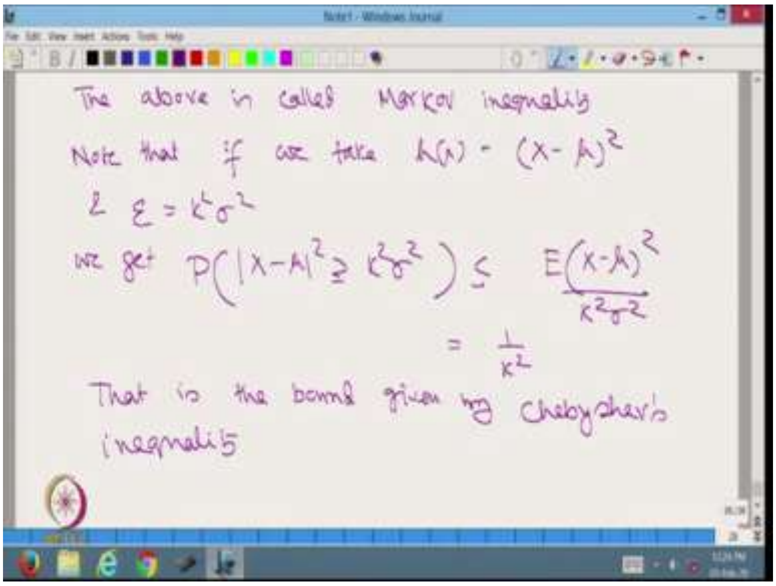
This is because h is a non-negative function and I am deleting that this part is positive which is greater than equal to sigma over Xi belonging to A epsilon times pi. This is because on A, hx is greater than equal to epsilon, which is equal to epsilon times sigma over pi, xi belonging to a is equal to epsilon times probability of hx greater than equal to epsilon. Therefore, the expected value of hx is greater than equal to epsilon times this.

(Refer Slide Time: 51:38)



Therefore, we conclude that probability h of x greater than epsilon is less than equal to the expected value of h of x upon epsilon when h of x is greater than equal to 0 for all x. In particular, if h of x is equal to modulus of x to the power r, then taking epsilon is equal to k to the power r, we get probability modulus of x greater than equal to k is less than equal to the expected value of modulus of x to the power r divided by k to the power r for all r, if expectation of modulus of x to the power r exists.

(Refer Slide Time: 53:01)



The above is called Markov inequality. Note that if we take h of x is equal to x minus mu whole square and epsilon is equal to k square sigma square, we get probability modulus of x minus mu whole square greater than equal to k square sigma square is less than equal to the expected value of x minus mu whole square upon k square sigma square which is equal to 1 upon K square.

That is the bound given by Chebyshev's inequality. Thus, Chebyshev's inequality can be derived from Markov inequality, but not vice versa. Okay friends, I stop here today. In the next class, I shall discuss by variate normal distribution, which is very important from probability point of view. Thank you so much.