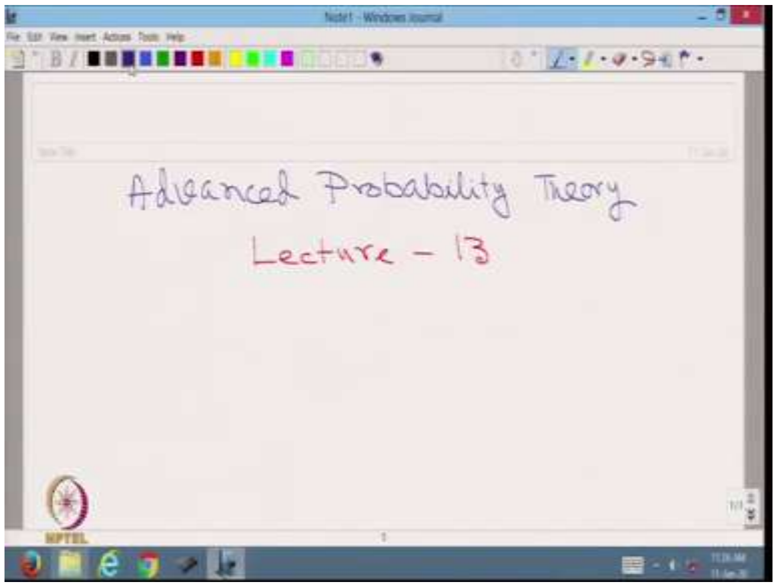
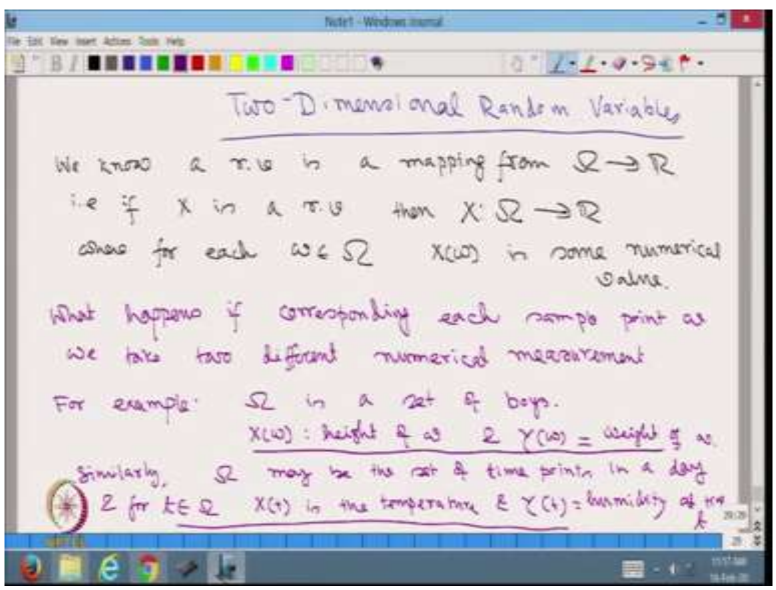
Advanced Probability Theory
Professor Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi
Lecture 13

(Refer Slide Time: 00:26)

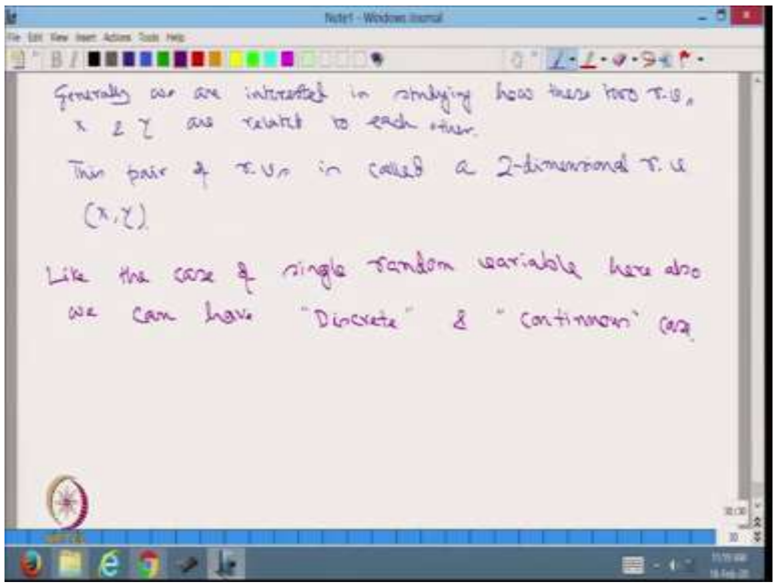




Welcome students to the MOOCs course on Advanced Probability Theory. This is lecture number 13. So as I said, at the end of last class, we shall study today, 2 dimensional random variables. We know a random variable is a mapping from omega to real line that is if x is a random variable then x is from omega to R where for each omega belonging to capital omega, x omega is some numerical value.

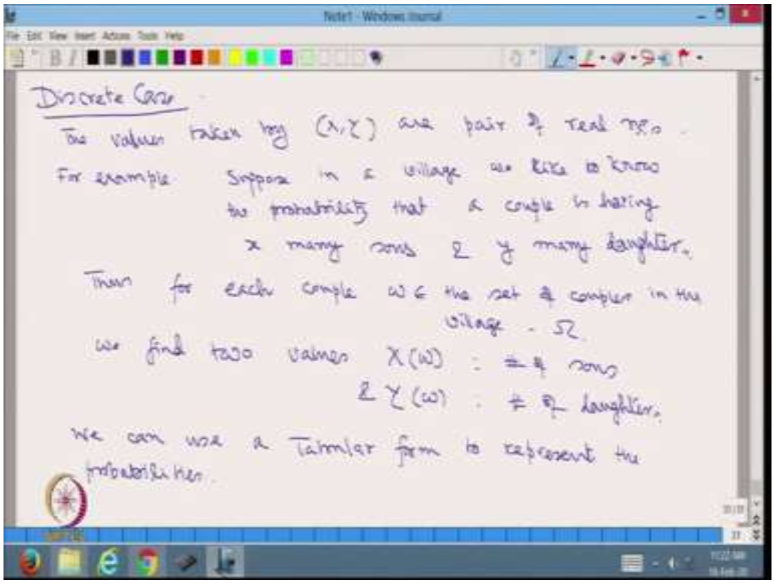
Well, what happens if corresponding to each sample point omega we take 2 different numerical measurement. For example, omega is a set of boys and x omega is height of omega and y omega is the weight of omega. Similarly, omega maybe the set of time points in a day and for t belonging to omega xd is the temperature and yt is equal to humidity at time t y. Thus we can see that we are taking 2 different measurements for the same sample point.

(Refer Slide Time: 03:51)



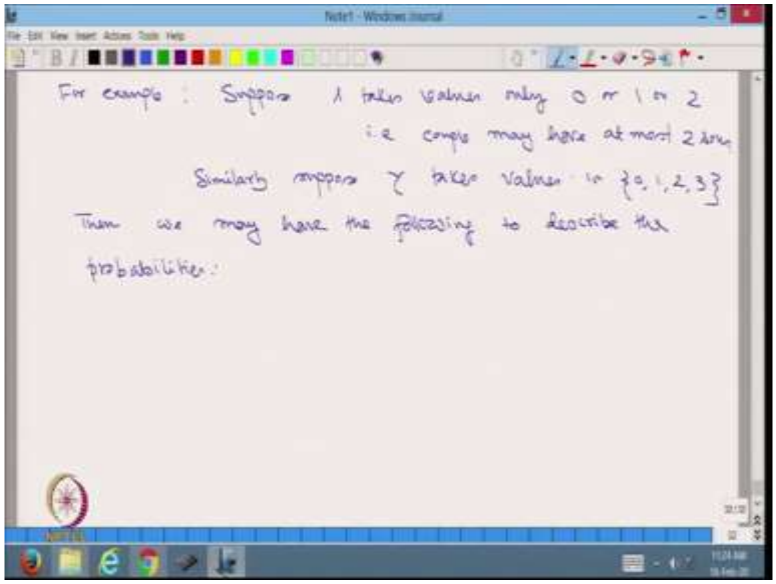
Generally we are interested in studying how these 2 random variables x and y are related to each other. This pair of random variables is called a 2 dimensional random variable x comma y, like the case of single random variable here also we can have discrete and continuous case.

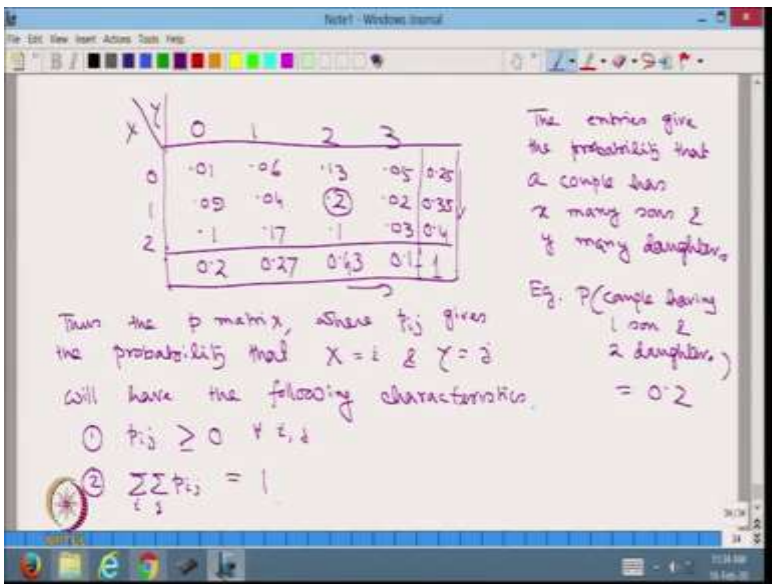
(Refer Slide Time: 05:26)



So, discrete case the values taken by x y are pair of real numbers. For example, suppose in a village we like to know the probability that a couple is having x many sons and y many daughters thus for each couple omega belonging to the set of couples in the village which let us call it omega we find 2 values x omega number of sons and y omega number of daughters. We can use a tabular form to represent the probabilities.

(Refer Slide Time: 08:04)



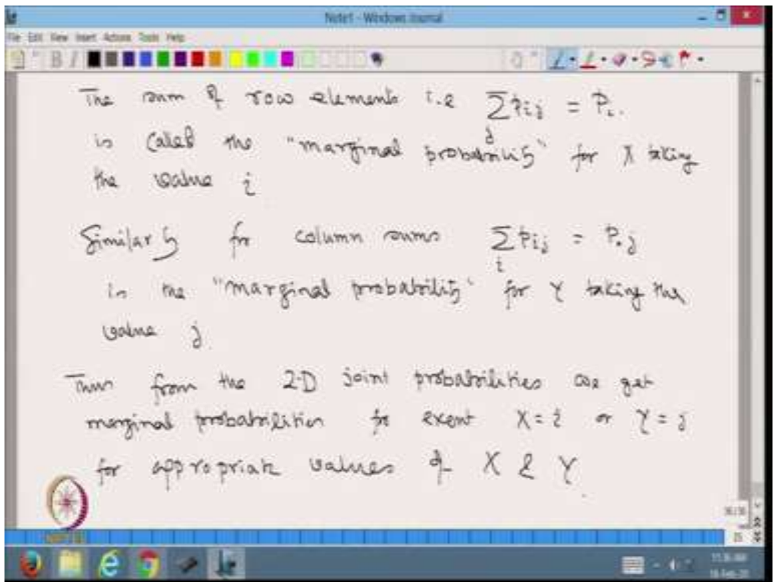


For example, suppose the x takes values only 0 or 1 or 2 that is couple may have at most 2 sons similarly suppose y takes values in 0 or 1 or 2 or 3 then we may have the following table to describe the probabilities. So, considered this table this gives the values of y 0, 1, 2, 3 and this gives the values for x 0, 1, 2. Suppose these values are point 0.01 0.06, 0.13 and 0.05, 0.09, 0.04, 0.2 and 0.02, 0.1, 0.17, 0.1, 0.03.

What does it mean? The entries give the probability that a couple has x many sons and y many daughters. For example, this entry is 0.2 therefore probability couple having 1 son and 2 daughters is equal to 0.2. Now let us add the rows, so, this gives me 0.25 this gives 0.35 and this gives 0.4 and if we add the columns, this gives us 0.2, this gives 0.27, this gives 0.43 and this gives this 0.1

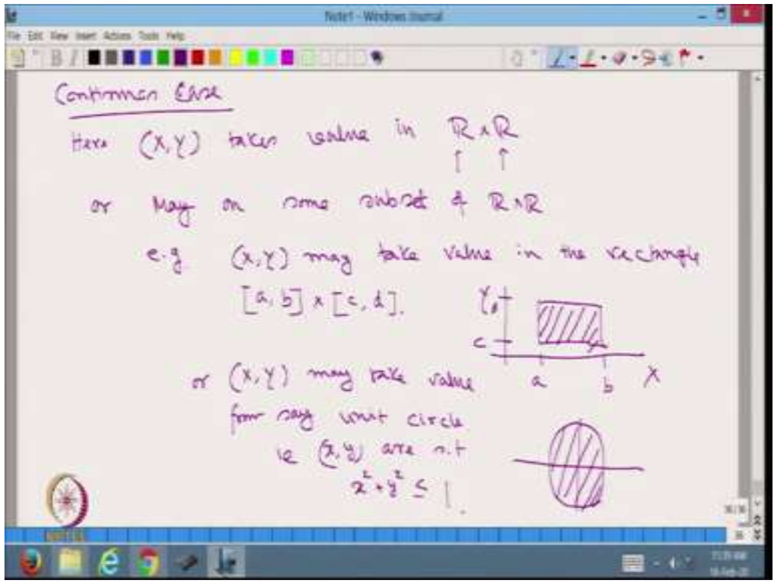
Now, if we add these values and if we add these columns, both will give us what, thus the P matrix where Pij gives the probabilities that x is equal to i and y is equal to j will have the following characteristics. One, Pij greater than equal to 0 for all ij and 2 sigma over i and sigma over j, Pij is equal to 1.

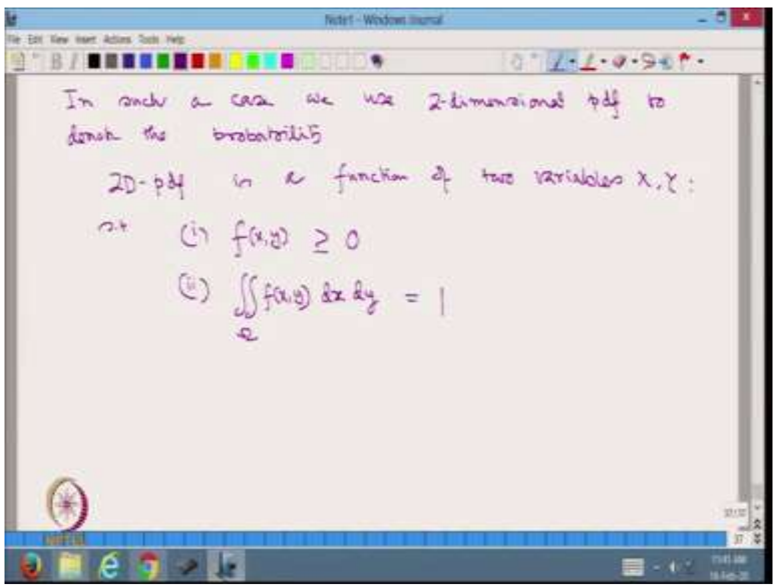
(Refer Slide Time: 13:21)



The sum of row elements that is sigma Pij summing over j is equal to Pi dot is called the marginal probability for x taking the value i. Similarly, for column sums sigma Pij over i is equal to call p dot j is the marginal probability that for y taking the value j. Thus from the 2D joint probabilities we get marginal probabilities for events x is equal to i or y is equal to j for appropriate values of x and y.

(Refer Slide Time: 15:51)



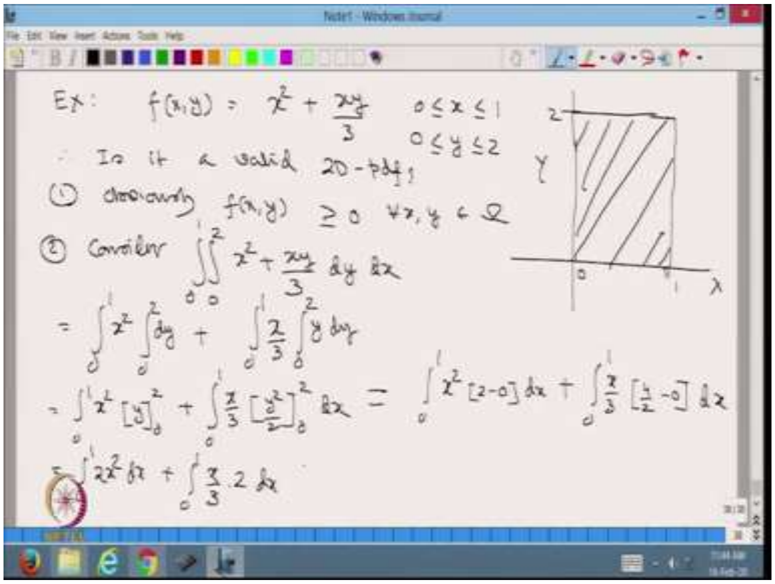


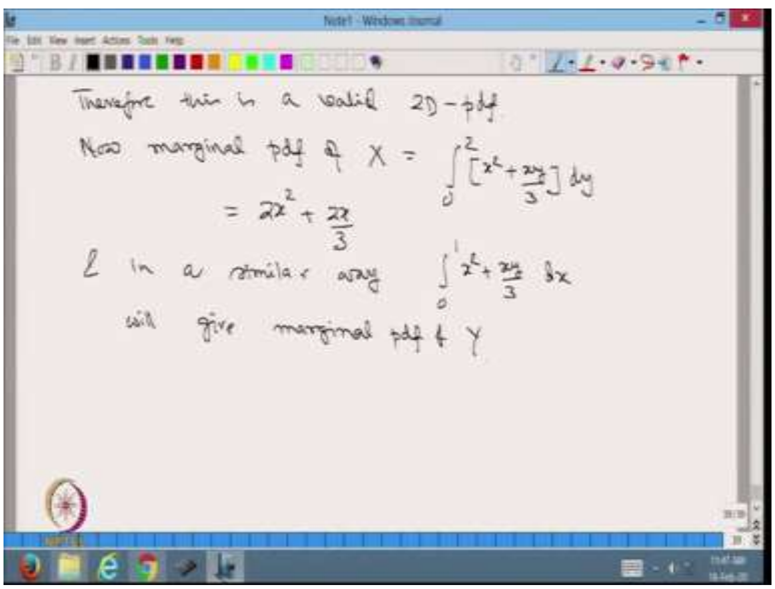
Continuous case, here x, y takes value in R cross R that means, x can take value any real numbers or y can take value in any real number or maybe on some subset of R cross R for example, x, y may take value in the rectangle a comma b plus c comma d.

That is, if this is the axis we are looking at a rectangle of the form and x, y values will be from this rectangle or x, y may take value from say unit circle that is, if this is the 2D plane and this is the unit circle, then x y can take value any point from inside the circle that is x, y are such that x square plus y square is less than equal to 1. In such a case, we use 2 dimensional PDF to denote the probability density function to denote the probability.

So, 2D PDF is a function of 2 variables x and y, such that 1 f x, y is greater than equal to 0 and 2 integration if x y on the space, let us call it omega is equal to 1.

(Refer Slide Time: 19:23)



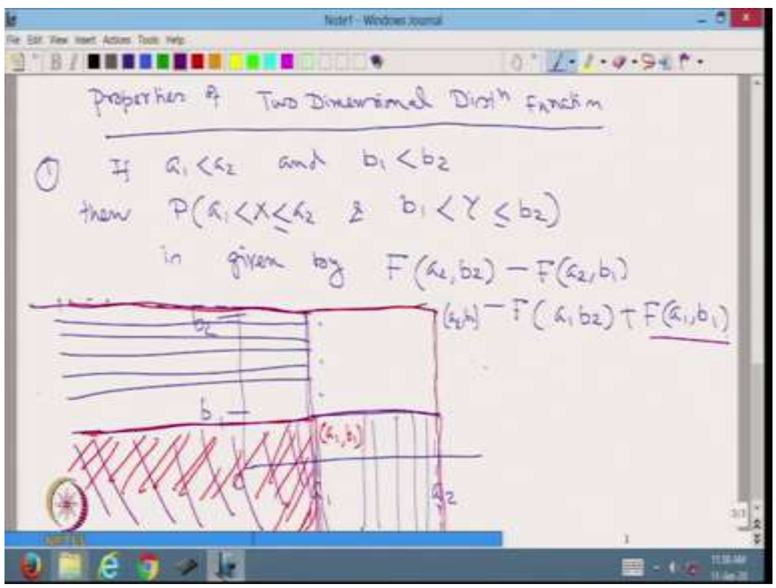


We illustrate with an example, if x, y is equal to x square plus x y upon 3, when 0 less than equal to x less than equal to 1, and 0 less than equal to y less than equal to 2, that is, if this is the 2D plane, and we have 0 and 1 on x axis here and 0 and 2 on the y axis here.

Therefore, the omega is the points in this rectangle. Therefore is it a valid 2 dimensional PDF? Obviously, f x, y greater than equal to 0 for all x y belonging to this omega now considered integration x is equal to 0 to 1, y is equal to 0 to 2 x square plus x y by 3 dy dx is equal to integration 0 to 1 x square integration 0 to 2 dy plus integration 0 to 1, x by 3, integration 0 to 2 y, dy is equal to integration 0 to 1 x square into y from 0 to 2 plus integration 0 to 1 of x by 3 of y square by 2 from 2, 0, dx is equal to integration 0 to 1, x square into 2 minus 0 dx plus integration 0 to 1 x by 3 into 4 by 2 minus 0 dx is equal to integration 0 to 1, 2x square dx plus integration 0 to 1 x by 3 into 2 dx is equal to 2 into x cube by 3, 0 to 1 plus 2 by 3 into x square by 2, 0 to 1 is equal to 2 by 3 plus 1 by 3 is equal to 1 therefore this is a valid 2D PDF.

Now, our marginal PDF of x is equal to integration 0 to 2 x square plus x y by 3 dy which will come out to be 2 x square plus 2x by 3 and in a similar way, integration 0 to 1 x square plus x y by 3 dx will give marginal PDF of y. With this background, we shall now study certain properties of joint distribution.

(Refer Slide Time: 24:14)



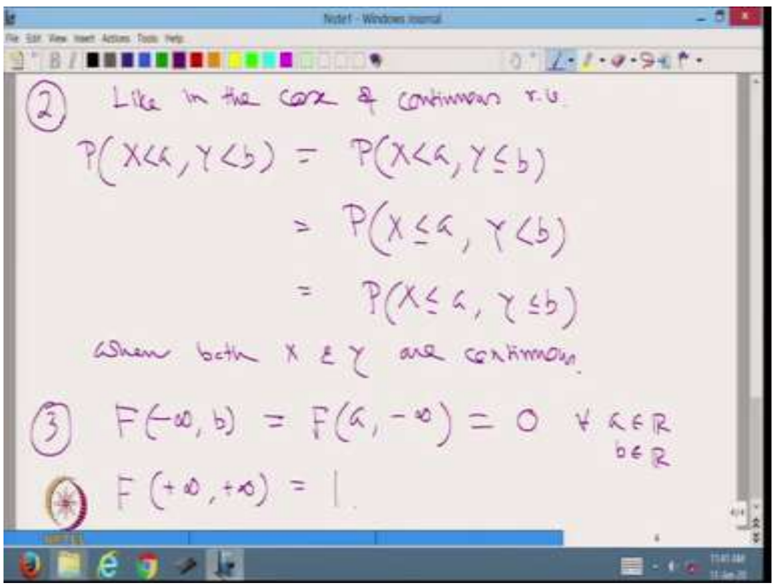
Some properties of 2 dimensional distribution function. One, if a1 less than a2 and b1 less than b2, then probability a1 less than x, less than a2 and b1 less than y less than equal to b2 is given by f at a2 b2 minus f at a2 b1 minus f at a1 b2 plus f at a1 b1, I am not going to prove it, but I am trying to give an explanation why it happens.

So, consider these are 2 plane and suppose, these are the points a1 a2 on the x axis and the these are the points b1 and b2 on the y axis, then effectively the event a1 less than x less than a2 and b1 less than y less than equal to b2 is going to be the probability x and y together lines in this rectangle.

On the other hand if this is a2 b2 then the inter area below this is going to give the probability that x less than equal to a2 b2. On the other hand this region is going to give us the probability that x less than equal to a1 and y is less than equal to b1. Therefore, we need to subtract from this rectangle, the rectangle as I am marking it here with red minus this rectangle, which now I am marking with blue line and also this rectangle which I am marking with blue line.

Therefore, we cannot define this and this rectangle explicitly. So, we try to subtract this rectangle and this rectangle but in doing that, we have subtracted this part twice. Therefore, to compensate for that we had f a1 b1 thus we get the formula as I have given earlier.

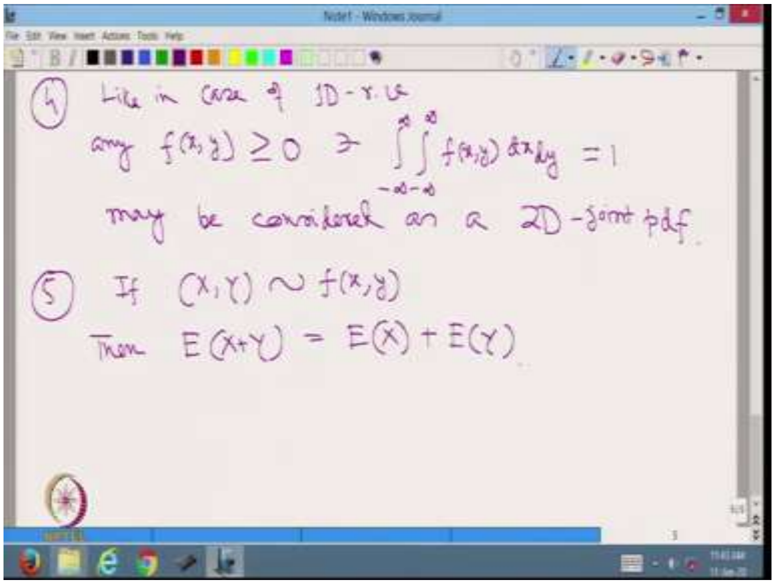
(Refer Slide Time: 28:43)



Property 2, like in the case of continuous random variable probability x less than a y less than b is equal to probability x less than a, y less than equal to b is equal to probability x less than equal to a, y less than b, which is equal to probability x less than equal to a, y less than equal to b when both x and y are continuous.

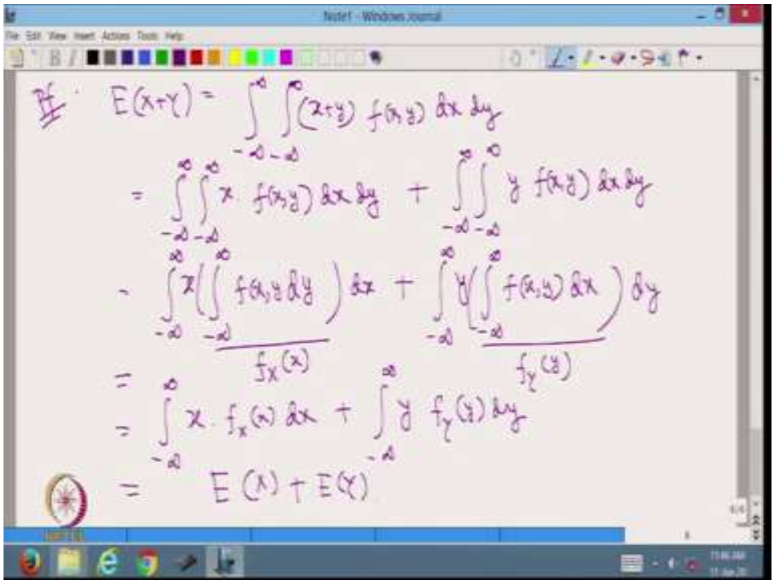
Three, f of minus infinity comma b is equal to f of a comma minus infinity is equal to 0 for all a belonging to R and be belonging to R. However, if plus infinity, plus infinity is equal to 1, these are very obvious properties and I hope you can understand them and their significance very easily.

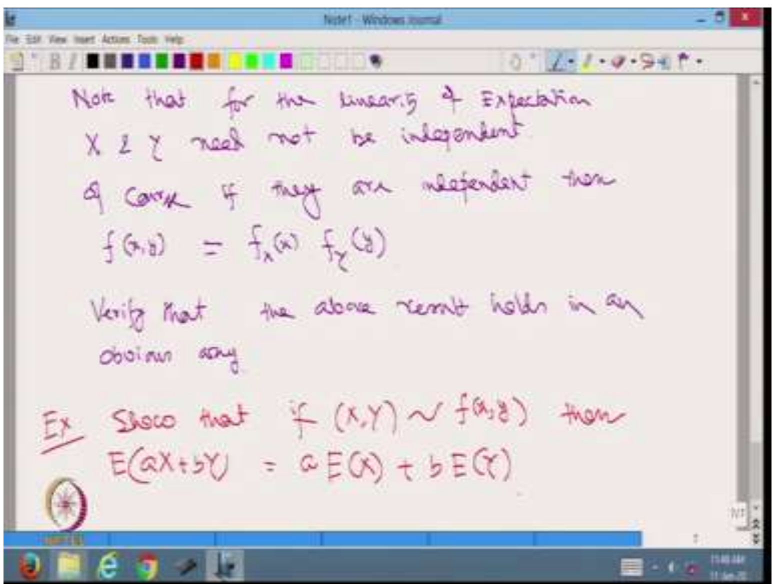
(Refer Slide Time: 30:27)



Like in case of 1 dimensional random variable in any f x, y greater than equal to 0 such that integration minus infinity to infinity f x, y dx dy equal to 1 maybe be considered as a 2D joint PDF. Five, if x y is jointly distributed as a f x, y then expected value of x plus y is equal to expected value of x plus expected value of y.

(Refer Slide Time: 31:42)



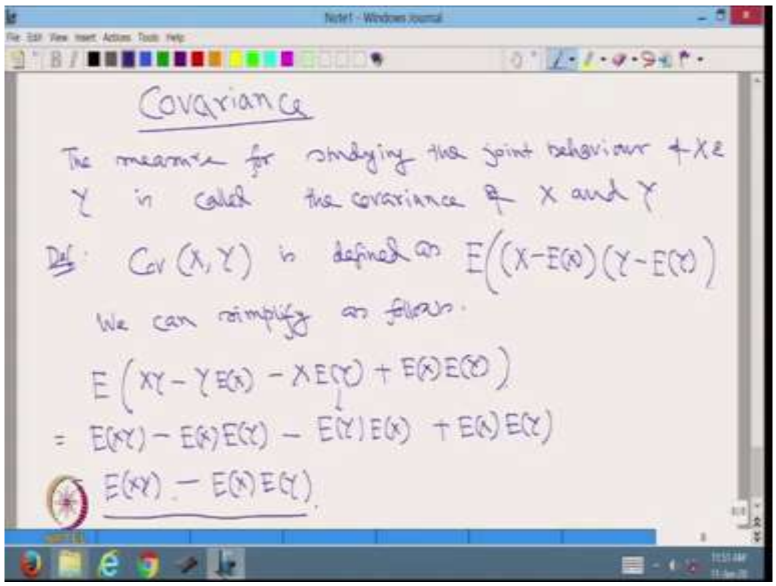


Proof, expected value of x plus y is equal to minus infinity to infinity, x plus y, f x, y dx dy is equal to minus infinity to infinity x times f x, y dx dy plus minus infinity to infinity y times f x, y dx dy is equal to let us first integrate with respect to y. Therefore, we can take out the x part of it minus infinity to infinity f x, y dy and then we integrate with respect to x plus, in a very similar way we take out y and then we integrate it with respect to y.

Now this internal part for a fixed x is going to give us the marginal density of x and this is going to give us the marginal density of y. Therefore this can be written as minus infinity to infinity x times a f x of x dx plus minus infinity to infinity y times f y, y dy which are nothing but expected by x plus the expected value of y.

Note that for the linearity of expectation x and y need not be independent. Ofcourse, if they are independent, then if x y we will be writing as f x x into a f y y and I want you to verify that the above result holds in an obvious way. I will give you an exercise show that if x y is jointly distributed with a f x y then expected value of ax plus by where a and b are constants is equal to a times expectation of x plus b times expectation of y.

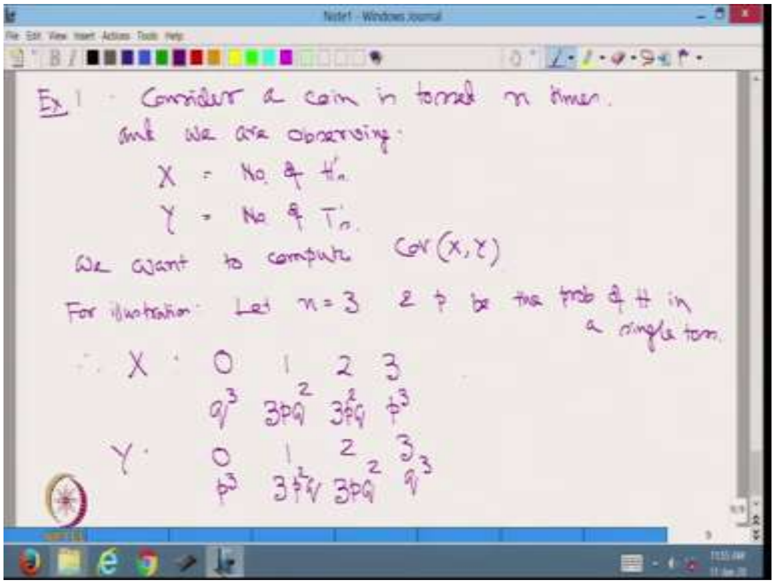
(Refer Slide Time: 36:05)

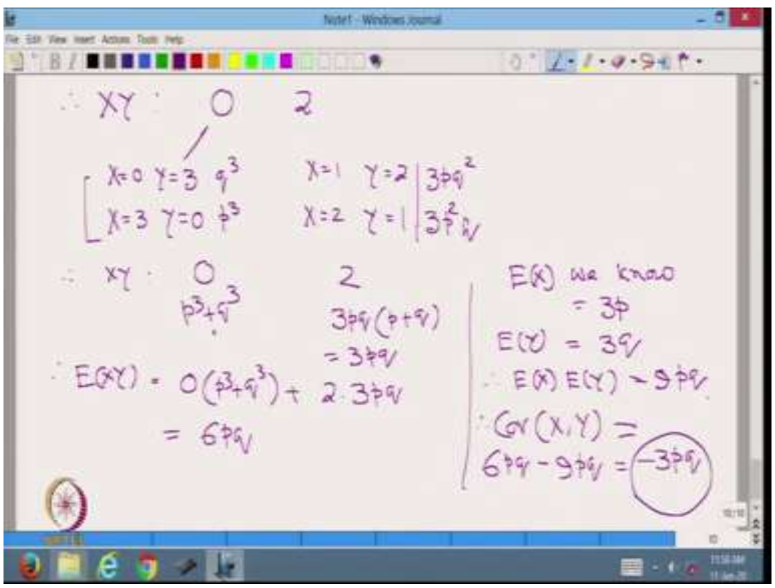


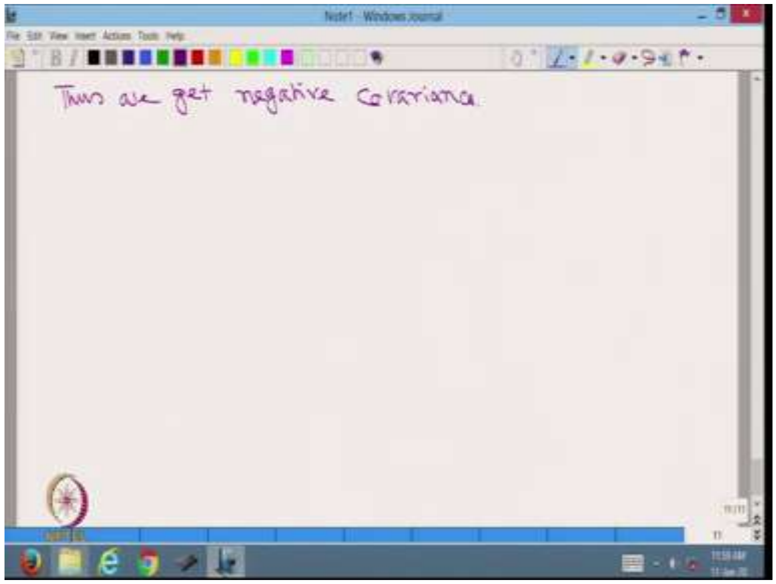
Concept of covariance as I said in the last class that we are often interested to see how x and y are behaving together and the measure for studying the joint behavior of x and y is called the covariance of x and y definition covariance between x and y is defined as the expected value of x minus expected value of x into y minus expected value of y and we can simplify it as follows.

This is equal to expected value of x y minus y times expected value of x minus x times the expected value of y plus the expected value of x, expected value of y is equal to you expected value of x y minus expected value of x, which is a constant. Therefore, we take out and then we take expectation of y minus expectation of y which is a constant we take out and multiplied by expectation of x plus expectation of x into expectation of y which is equal to expectation of x times y minus expectation of x into the expectation of y.

(Refer Slide Time: 38:47)







So, let me illustrate this example one, consider a coin is tossed n times and we are observing x is equal to number of heads and y is equal to number of tails. We want to compute covariance between x and y. So, how we will do that? For illustration let n equal to 3 and P be the probability of head in a single toss. Therefore, x is a binomial 3 comma p. Therefore, it takes values 0, 1, 2, 3 and their probabilities are q cube, because all 3 tosses ended up in tail 3 p q square 3 p square, q and p cube.

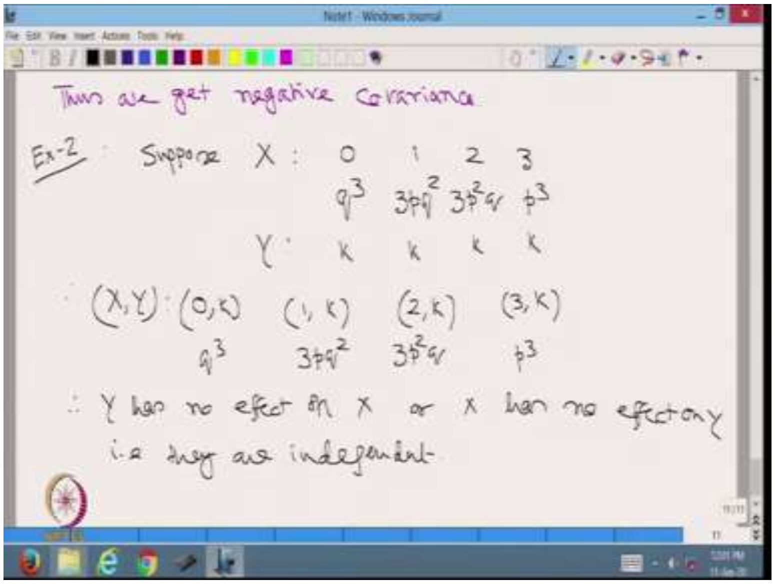
Therefore, y is that number of tails, y also can be 1, 2 and 3 and probability y is equal to 0 means all 3 are heads probability y is equal to 1 means 3 p square q, 3 p, q square, and this is going to be q cube. Therefore x, y can take the values 0 or 2, 0 comes when x is equal to 0 and y is equal to 3 this probability is q cube and it can come when x is equal to 3 and y is equal to 0 that probability is p cube.

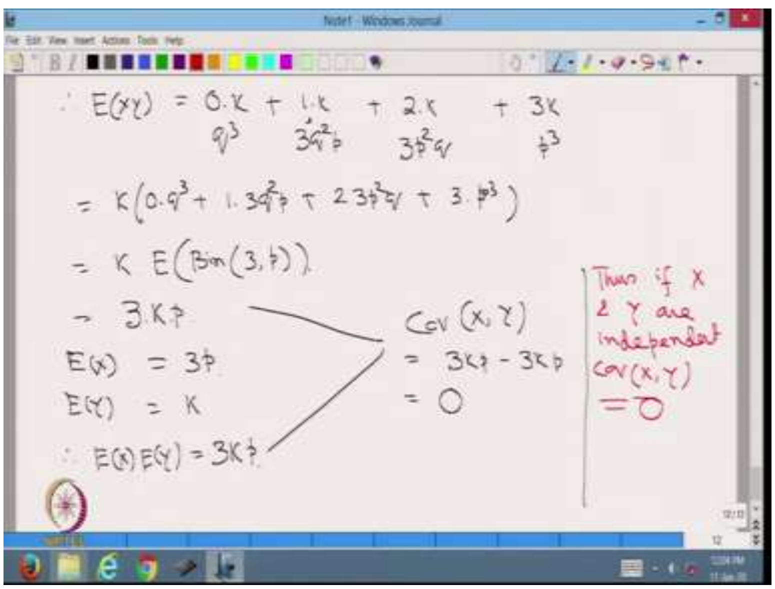
On the other hand, the x y can be 2 if x is equal to 1 y is equal to 2 and that probability is 3 p q square x is equal to 2 and therefore, y is equal to 1 and that probability is 3 p square q. Therefore, x y has two values 0 and 2 with probabilities p cube plus q cube and here it is 3 pq into p plus q is equal to 3 p q.

Therefore, expected value of x y is equal to 0 times p cube plus q cube plus 2 times 3 p q is equal to 6 p q, but we know that expected value of x we know is equal to NP, which in this case is 3p and expected value of y. If you are smart enough, you can easily guess it is going to be 3q.

Therefore, the expected value of x into the expected value of y is equal to 9 p cube. Therefore, covariance between x y is equal to 6 pq minus 9 pq is equal to minus 3 pq which is negative.

(Refer Slide Time: 44:14)





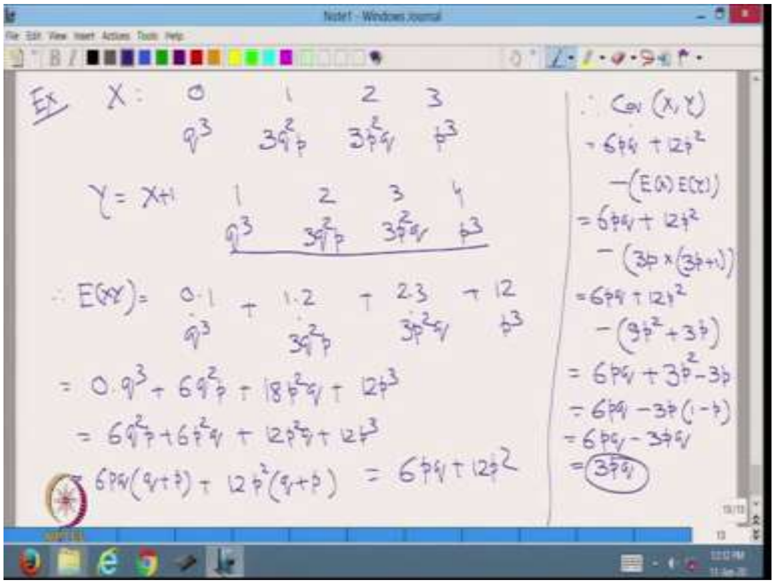
Example two, suppose x takes values 0, 1, 2, 3 and their probabilities like before p cube 3p q square, 3p square q and p cube. Let us assume corresponding values of y is that some constant k in all the 4 cases that means, x y takes values 0 comma k with probability q cube 1 comma k with probability 3 pq square, 2 comma k with vulnerability 3 p square q and 3 comma k with probability p cube.

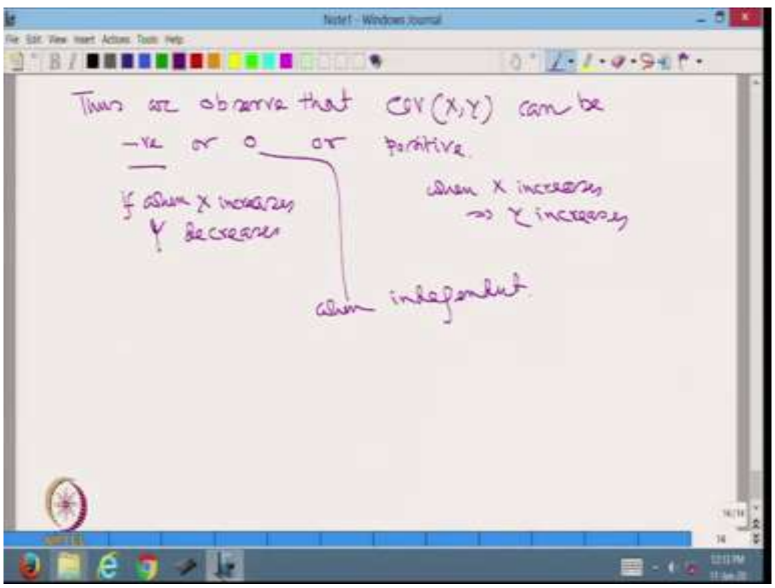
What does it mean? That means that y has no effect on x or x has no effect on y that is they are independent therefore, expected value of x y is equal to 0 times k with probability q cube plus 1 times k with probability 3q square I am multiplying by this corresponding probabilities 2 times k multiplied by 3 p square q plus 3 k multiplied by p cube is equal to k times 0 into q cube plus 1 into 3 q square p plus 2 into 3 p square q plus 3 into p cube.

Is equal to expected value of a binomial random variable with parameters 3 comma p is equal to 3 k p, now expected value of x as we know is equal to 3 p expected value of y, because it takes k, the same value in all the cases therefore, the expected value of y is k therefore, expected value of x into expected value of y is also 3 k p.

So, together from these 2, we can see that covariance of x y is equal to 3 k p minus 3 k p is equal to 0 thus if x and y are independent covariance of x y is equal to 0. This is not a proof, this is an example to give you the insight between covariance.

(Refer Slide Time: 48:28)





Now, I take a very similar example, suppose x takes the value 0, 1, 2, 3 the same case I am dealing with, with the probabilities q cube 3 q square p 3p square q, p cube and let y is equal to x plus 1. Therefore, y takes the values 1, 2, 3, 4 with the probabilities q cube, 3q square p, 3p square q, p cube.

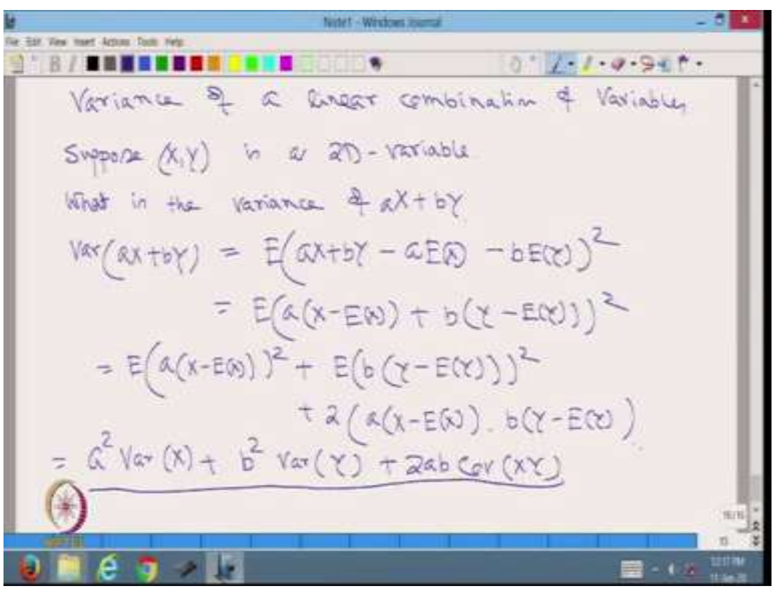
Therefore, expected value of x y is equal to 0 into 1 with the probability q cube plus, we will be multiplying 1 into 2 multiplied by the probability 3q square p plus 2 into 3 multiplied by 3 p square q plus 12 multiplied by p q is equal to 0 times q cube plus 6 into u square p plus 18 into p square q plus 12 into p cube is equal to 6 q square p plus 6 p square q plus 12 p square q plus 12 p cube is equal to 6 p q multiplied by q plus p plus 12 p square into q plus p, which is equal to 6 pq plus 12 b p square.

Therefore, covariance of x y is equal to 6 pq plus 12 p square minus expected value of x into expected value of y is equal to 6 p q plus 12 p square minus expected value of x is going to be 3 p multiplied by expected value of y since y is equal to x plus 1, therefore, expected value of y is going to be expected value of x plus 1 that this is going to be 3 p plus 1 is equal to 6 pq plus 12 p square minus 9 p square plus 3 p is equal to 6 p q plus 3 p square minus 3 p is equal to 6 p q minus 3 p into 1 minus p is equal to 6 p q minus 3 p q is equal to 3 p, q.

Since P and q both are positive, we see that in this case, dark covariance between x and y is coming out to be positive. Thus we observed that covariance between x y can be negative or 0 or positive. Can you guess when it is going to be negative? It is going to be negative if when x increases but y decreases it is positive when x increases implies y increases and it is going to be 0 when independent.

Thus just from the looking at the sign of the covariance, we can understand the underlying relationship between x and y, at least at a very crude level.

(Refer Slide Time: 53:51)

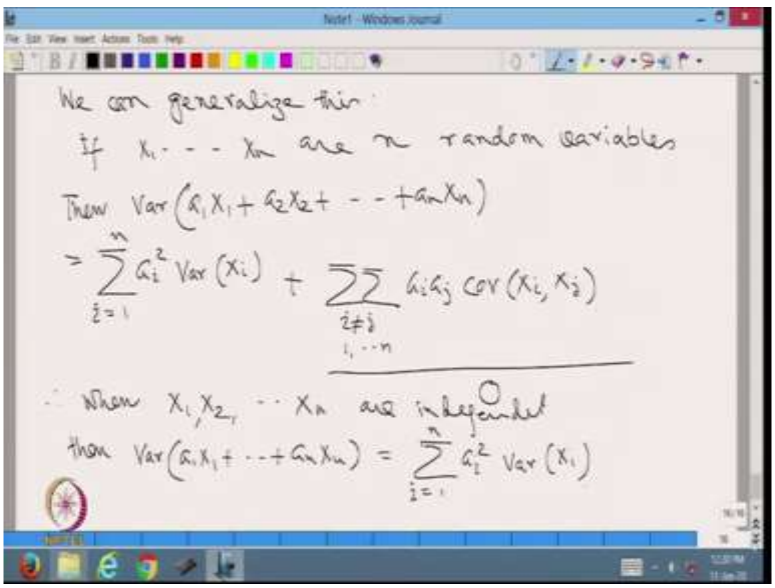


What is the variance of linear combination of variable. Suppose x and y is a 2D variable, what is the variance of a x plus b y, we know variance of a x plus b y is equal to the expected value of a x plus b y minus a times expected value of x minus b times expected value of y whole square.

Which is equal to expected value of a times x minus expected value of x plus b times y minus expected value of y whole square is equal to expected value of a times x minus expected value of x whole square plus expected value of b times y minus expected value of y whole square plus 2 times a into x minus expected value of x into b into y minus the expected value of y which is very easy to understand is going to be a square times variance of x plus b square times variance of y plus 2 a b times covariance of x y.

Which is coming from here thus we get a formula for the variance of the linear combination of 2 variables, which is coming out to be a squared times variance of x, b squared times variance of y plus 2 a b times covariance of x y.

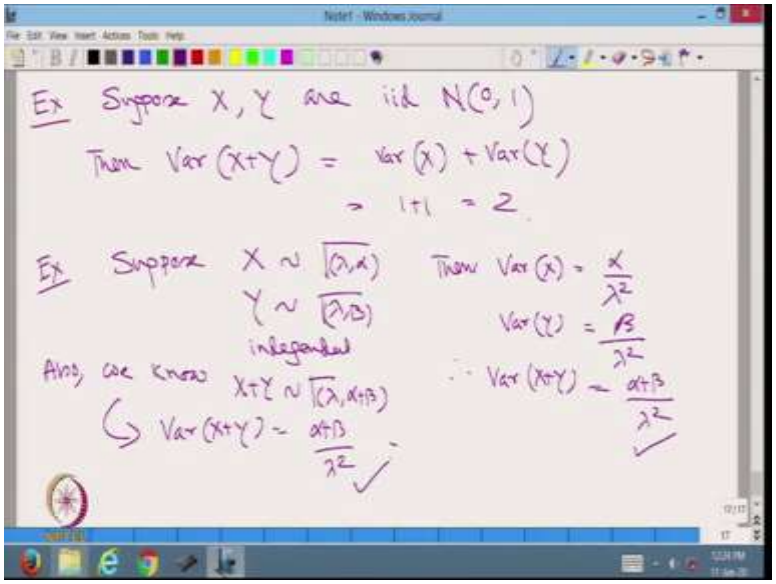
(Refer Slide Time: 58:50)



We can generalize this if x1, x2, xn are in random variables need not be independent, they are just in random variables then variance of a1 x1 plus a2 x2 plus an xn is equal to I am writing you the formula, you please verify in a very similar way it is coming out to be sigma ai square variance of xi, i is equal to 1 to n plus sigma-sigma is not equal to j from 1 to n that means we are looking at n into n minus 1 many different combinations i and j cannot be equal and what we will get is a i, a j covariance between xi xj.

Now, if all the xi's are independent of each other, then this term becomes 0. Therefore, when x1, x2, xn are independent then variance of a1 x1 plus an xn is equal to sigma i is equal to 1 to n ai square variance of xi.

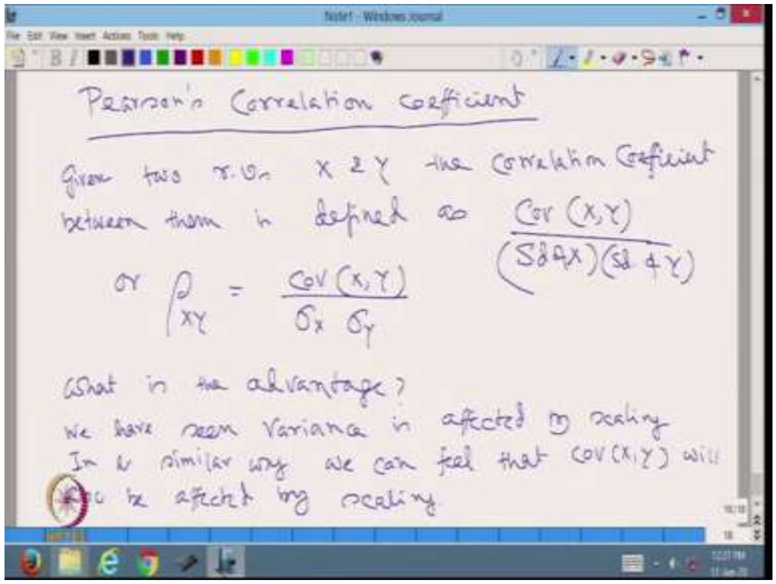
(Refer Slide Time: 59:08)



Example suppose x and y are independent normal 0 1 then variance of x plus y is equal to variance of x plus variance of y is equal to 1 plus 1 is equal to 2. Another example, suppose x is gamma lambda alpha y is gamma lambda beta and they are independent then variance of x is equal to alpha upon lambda square variance of y is equal to beta upon lambda square.

Therefore, from the above formula we get variance of x plus y is equal to alpha plus beta upon lambda square. On the other hand, we know x plus y will be distributed as gamma with lambda alpha plus beta. Therefore, from here we can conclude variance of x plus y is also going to be alpha plus beta upon lambda square. So, you can see that these 2 results are coming out to be equal.

(Refer Slide Time: 61:15)

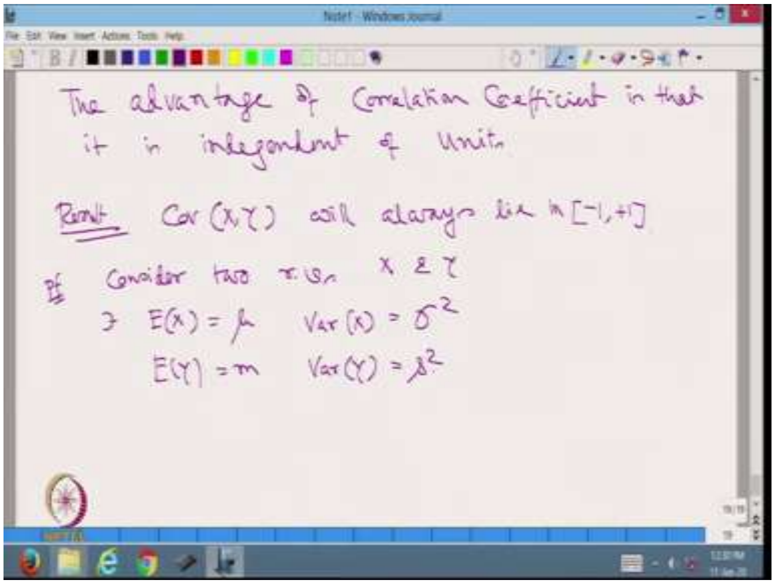


Let us now study another important concept, which is called the Correlation Coefficient due to Karl Pearson. So, given 2 random variables x and y, the correlation coefficient between them is defined us covariance of x y upon standard deviation of x into standard deviation of y or by notation rule x y is equal to covariance of x comma y upon sigma x sigma y, where sigma x is the standard deviation of x, sigma y is the standard deviation of y.

What is the advantage? We have seen variance is affected by scaling. In a similar way, we can feel that covariance of x y will also be affected by scaling that makes comparison or association of 2 random variables little bit difficult. With respect to some unit you may feel they have a huge variance with respect to some other skill, they may have a very small variance. For example, if we measure the length in terms of centimeters as opposed to in terms of meters, we will find that the variance in the first case, the numerical value is going to be very high.

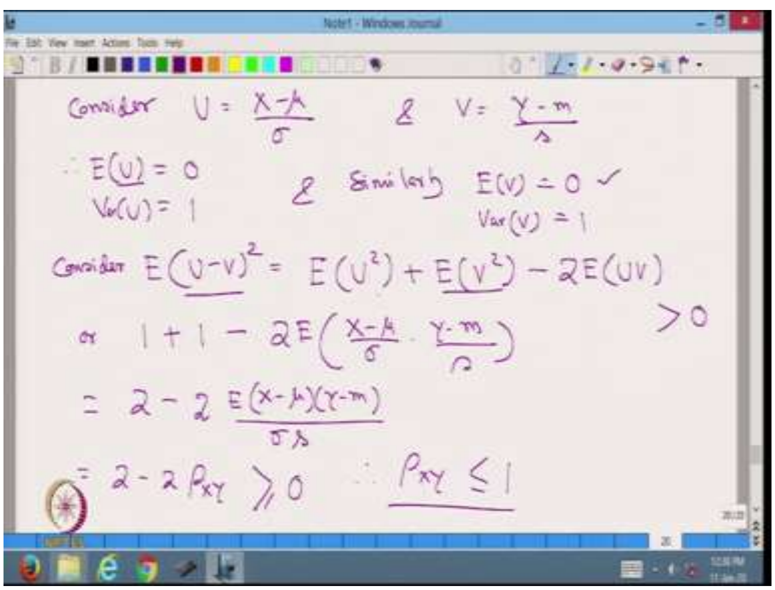
Whereas, when it is measured in meters, then the variance value is going to be much smaller. Same is with respect to covariance.

(Refer Slide Time: 64:34)



The advantage of correlation coefficient is that it is independent of units. So, this gives us a better objective view of the association of x and y. Result, covariance of x y will always lie in minus 1, to plus 1. Proof, consider 2 variables x and y such that expected value of x is equal to Mu variance of x is equal to sigma square and expected value of y is equal to m and variance of y is equal to s square.

(Refer Slide Time: 66:08)

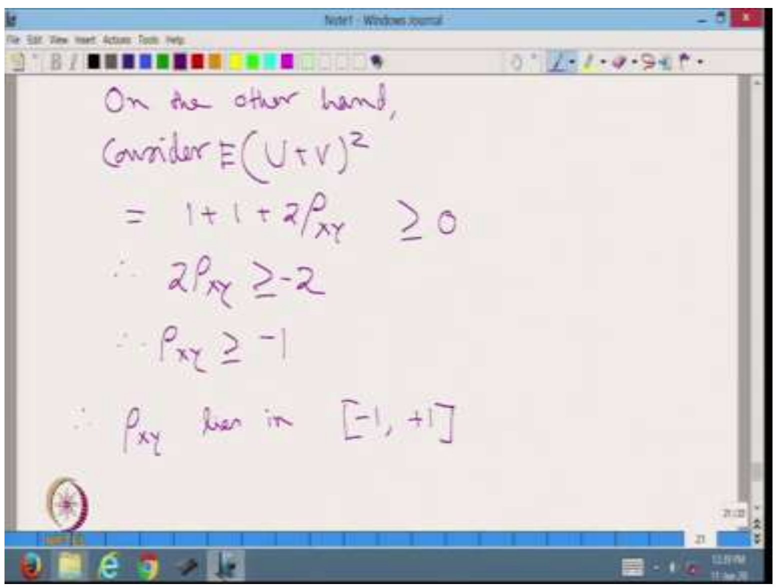


Consider u is equal to x minus mu upon sigma and v is equal to y minus m upon s. Therefore, expected value of u, this is going to be 0, variance of u is going to be 1 by sigma square into variance of x, therefore, it is going to be 1 and similarly, expected value of v is equal to 0 and variance of v is equal to 1.

Consider expected value of u minus v whole square. This is equal to expected value of u square plus the expected value of v square minus 2 times expected value of uv and since it is a positive quantity, this is going to be greater than 0, or since the expected value of u square is equal to variance of u, because if the expectation is 0, this quantity is going to be 1.

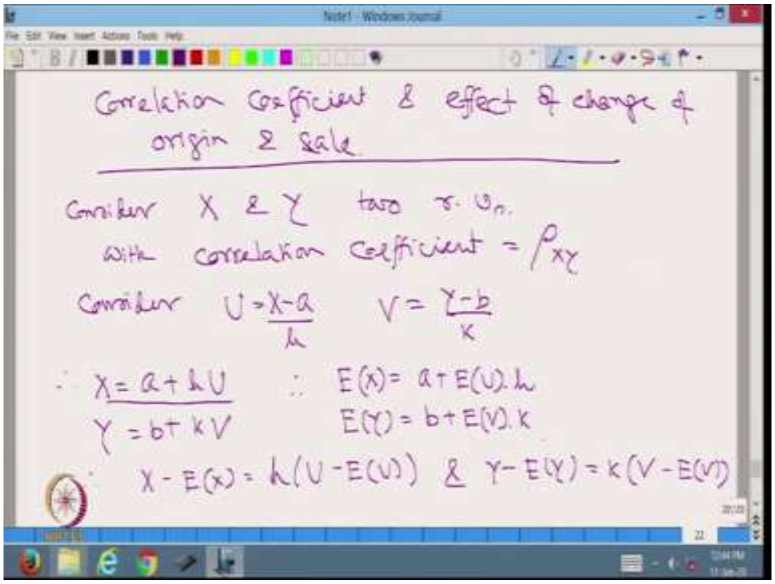
In a similar way, expected value of v square is equal to because expected value of v is equal to 0 is equal to 1 minus 2 times the expectation of x minus mu by sigma into y minus m upon s. Therefore, this quantity is equal to 2 minus 2 times expected value of x minus mu into y minus m upon sigma s is equal to 2 minus 2 times rho x y, which is greater than equal to 0, they can be, it can be 0 if they are independent. Therefore, rho x y is less than equal to 1.

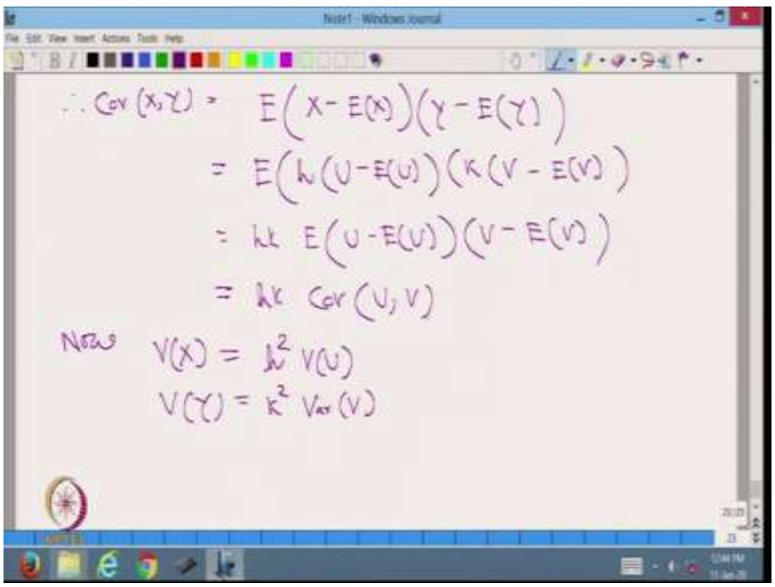
(Refer Slide Time: 68:50)

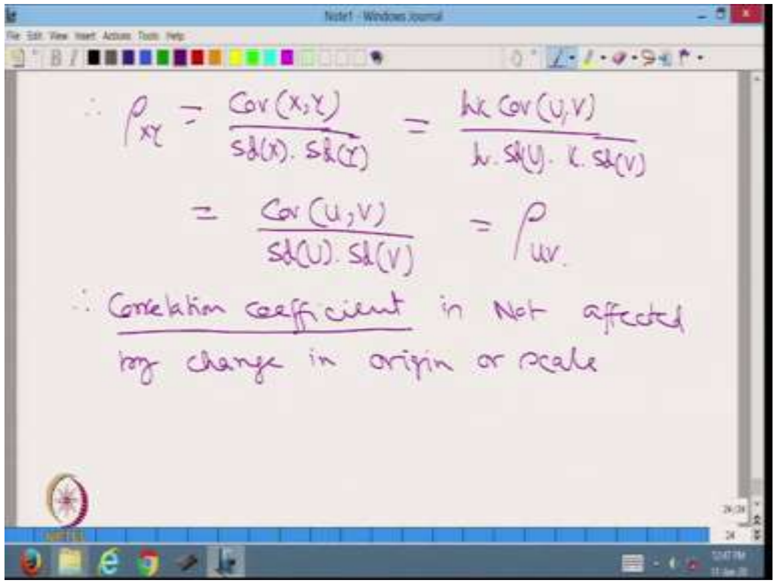


On the other hand consider u plus v whole square and take its expectation in a very similar way it is going to be 1 plus 1 plus 2 times rho x y, which is greater than equal to 0. Therefore, 2 times rho x y is greater than equal to minus 2. Therefore, rho x y is greater than equal to minus 1. So, rho x y, which is the correlation coefficient between the x and y lies in minus 1 to plus 1.

(Refer Slide Time: 69:49)







Correlation coefficient and effect of change of origin and scale, consider x and y, two random variables with correlation coefficient is equal to rho x y consider u is equal to x minus a upon h and v is equal to y minus b upon k. Therefore, x is equal to a plus h u and y is equal to b plus k times v therefore expected value of x is equal to a plus expected value of u multiplied by h and expected value of y is equal to b plus expected value of v multiplied by k.

Therefore, x minus expected value of x is equal to h into u minus expected value of u and y minus expected value of y is equal to k into v minus expected value of v. Therefore, covariance between x comma y is equal to the expected value of x minus expected value of x into y minus the expected value of y is equal to expected value of h into u minus expected value of u into k times v minus expected value of v is equal to h k if we take out then it is going to be expected value of u minus expected value u into v minus expected value of v is equal to h k into covariance between u comma v.

Now variance of x is equal to h square times variance of u, because x is a plus h u and we know that variance is affected by the square of the coefficient, variance of y is equal to k square into variance of v. Therefore, rho x y, which is the correlation coefficient between x and y is equal to covariance of x comma y divided by standard deviation of the x into standard deviation of y is equal to h k times covariance of u v upon h time standard deviation of u into k times standard deviation of v is equal to covariance of u v divided by standard deviation of u into standard deviation of v is equal to correlation between u and v therefore correlation coefficient is not affected by change in origin or scale.

So that is the major advantage of correlation coefficient because it gives a stable measurement of the association between the 2 random variables, okay, friends, I stop here today from the next class I shall discuss the concept of generating functions in particular with a lot of emphasis on moment generating function. Till then thank you so much. Thank you.