Stochastic Structural Dynamics

Prof. Dr. C. S. Manohar

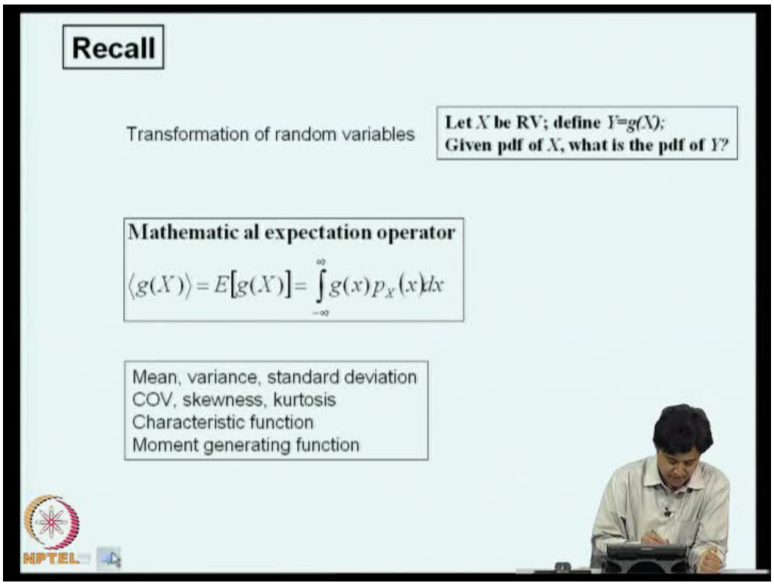
Department of Civil Engineering

Indian Institute of Science, Bangalore

Lecture No. # 04

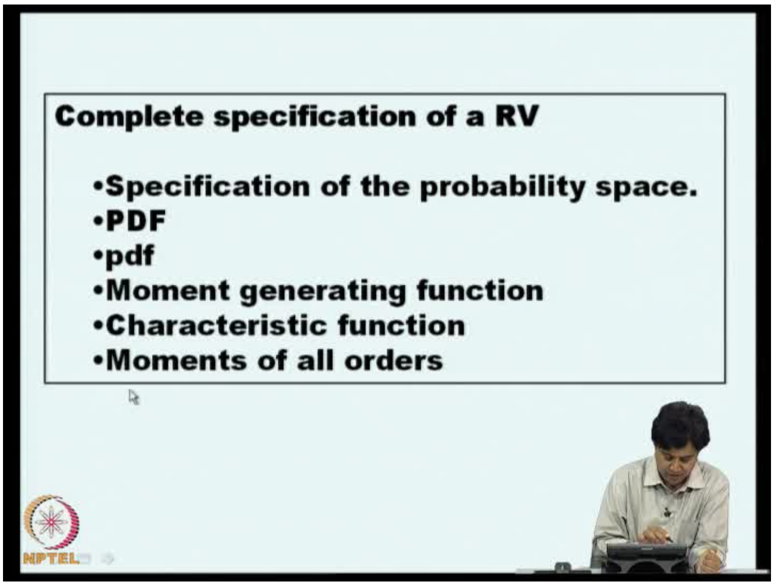
Multi-Dimensional Random Variables – 1

(Refer Slide Time: 00:21)

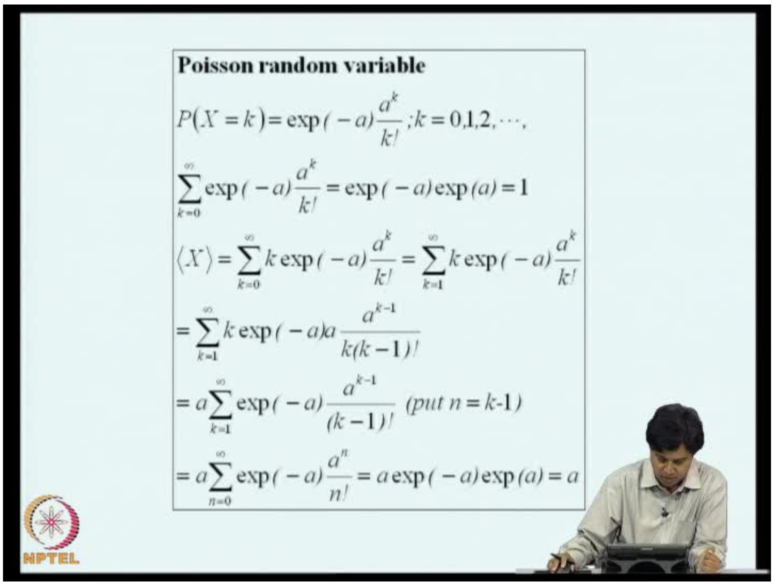


In this lecture, we will continue with our discussion on description of random variables; before that, we will quickly recall what is that we are trying to do in the last lecture. So, we consider the problem of transformation of random variables. So, the basic problem was if X is a random variable, whose probability density function is given and we introduce the transformation Y is equal to g of X, where g is a given function, the question is - what is the probability density function of Y? So, this problem we have tackled in the last class. We also introduced the notion of mathematical expectation operator, where this is read as expectation of g of X, also written as E of g of X, is actually the integral g of x p X of x dx.

(Refer Slide Time: 01:37)



(Refer Slide Time: 02:18)



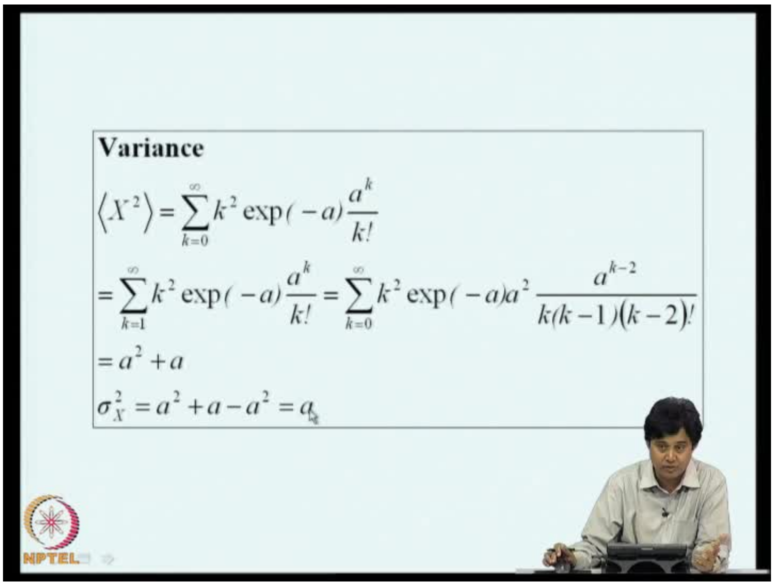
Based on this notion, we introduce the description of random variable in terms of moments, namely: mean, variance, standard deviation, coefficient of variations, skewness, kurtosis and we also introduced what are known as characteristic function and moment generating function.

So, at this stage, a complete specification of a random variable, based on what we have learnt so far, is through specification of the probability space or specification of the probability distribution function or through probability density function or through moment generating function or characteristic function or moments of all orders, when they exist. So, when we start modeling to describe any random variable, we can use any one of these descriptions, depending on the convenience in a given context. So, we will continue with this, and now, consider deriving the moments of a Poisson random variable.

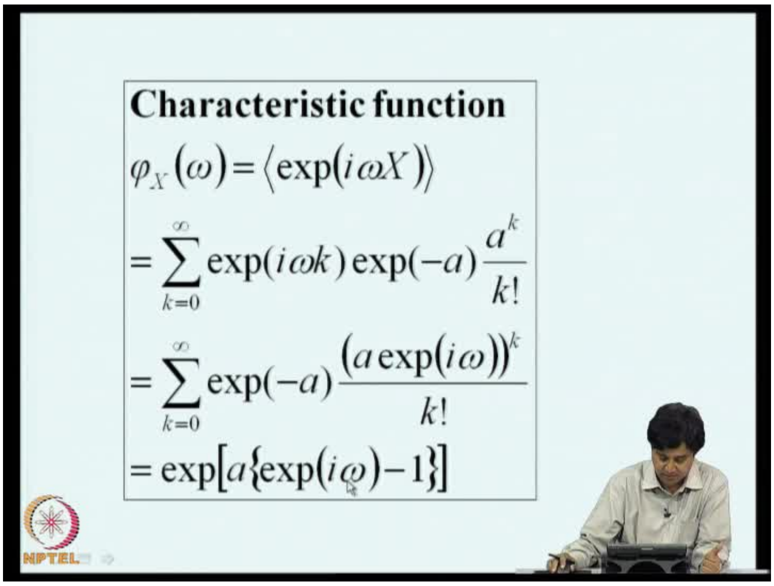
A Poisson random variable is a discrete random variable with countably infinite sample space, and it is probability mass function, is given through this expression, where a is a parameter. Now, we can first verify that, this is a valid probability mass function, by considering the question, whether probability of sample space is 1 or not, that would mean, if you sum all the probabilities, this should go to 1 and indeed, it happens to go to 1, as you see here.

Now, the expected value of X is given by this expression k into probability of X equal to k, which is exponential minus a a to the power of k by k factorial and since at k equal to 0, the term would varnish, we can start the summation from k equal to 1, and with little bit of rewriting, we can introduce a substitution of variables n is equal to k minus 1, and we can show that, the expected value of X is equal to the parameter a; that means, the parameter a that appears in the description of Poisson random variable, has the interpretation of being its mean value also.

(Refer Slide Time: 03:46)

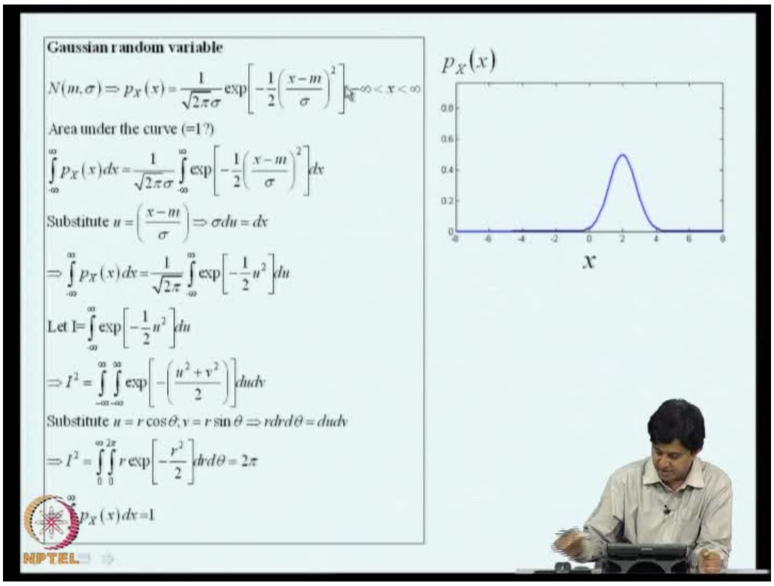


(Refer Slide Time: 04:40)



To compute the variance, we can start by computing the mean square value, which is now the summation k square exponential minus a a to the power of k by k factorial. And again by arranging the terms in a convenient manner, we can show that, this mean square value is given by a square plus a, the variance is the mean square value minus square of the mean and it turns out that this is also a; there is no contradiction here, because the basic random variable is a integer, which is used to count occurrences; so, there is no question of inconsistency of units, because mean is also a and variance is also a, the question may occurred to you, whether the units are consistent. The characteristic function is defined as the expected value of exponential i omega capital X, where omega is the real constant.

(Refer Slide Time: 05:43)



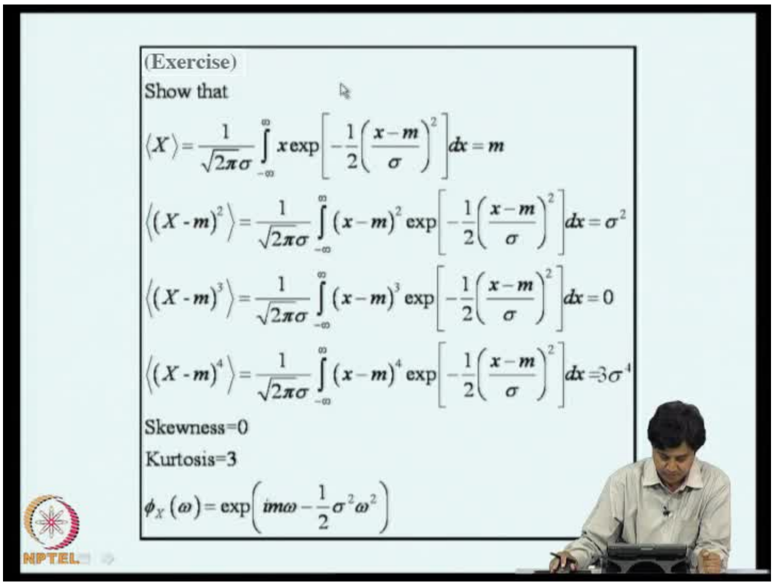
Now, this again is summation k equal to 0 exp of i omega k into the probability mass function, and by arranging the terms, by taking exp i omega along with this a raise to k, we can write this as this entire thing to the power of k by k factorial, and from this, it follows that, this is the characteristic function of a Poisson random variable. From this, we can again compute the moments, and this matter of differentiation with respect to omega, and putting omega equal to 0, and you can verify, whether the mean and variance etcetera, that we just calculated are consistent, with this expression for characteristic function.

Let us, now, consider another random variable, the Gaussian random variable; so, a Gaussian random variable, we write it as N of m sigma, where m, N, sigma are parameters of the probability density function. So, the probability density function itself is given by this expression, where X takes value from minus infinity to plus infinity and parameter m appears here, and parameter sigma appears here, and here.

First, we will verify if we plot this, a typical plot will be the familiar bell, like curve shown here. The first question we can ask is, whether the area under this function is indeed equal to 1, because probability of sample space has to be 1. Now, this is matter of integration, but there are few simplifications possible, that would be useful later; so, we can go through that. We are interested in finding area, under the probability density function, that is minus infinity to infinity exponential of this term in the bracket.

Now, we make this substitution u is equal to x minus m by sigma, and modify this integral, we get, this to be area under the curve is same as the this value of this integral. Now, if we call this integral as I, and if we square it, we get this, it becomes a double integral now. And now, if you make the substitutions u is equal to r cos theta, and v equal to r sin theta d u d v will be r d r d theta, so you substitute that; we can show that the value of I square will be 2 pi. And therefore, the value of this integral which is the area under probability density function is 1 by square root of 2 pi into square root of 2 pi which is 1; so, this is the valid probability density function.

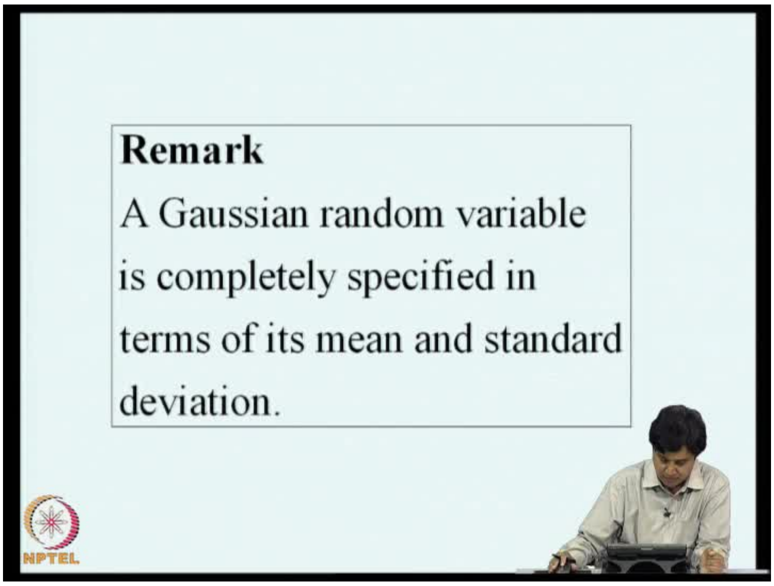
(Refer Slide Time: 07:42)



How about the mean, is question of now integration, I will state the results, I leave the actual calculations, as derivation as an exercise. So, we can show that, that expected value of X turns out to be the parameter m; similarly, the variance turns out to be the parameter sigma square. So, in a Gaussian density function, there are two parameters m and sigma and now they have the meaning, that m is the mean and sigma is the standard deviation.

We can continue to evaluate the third order moment and we can show that, this is equal to 0, this is not surprising, because this is an odd function, this is an even function and you can show that, this indeed equal to 0. The fourth moment can be shown to be equal to 3 into sigma to the power of 4. We have introduced the notion of skewness, which is the third moment divided by standard deviation whole cube and that for a Gaussian random variable is 0, and similarly, the kurtosis for a Gaussian random variable, which is ratio of fourth moment and fourth power of standard deviation, which is 3.

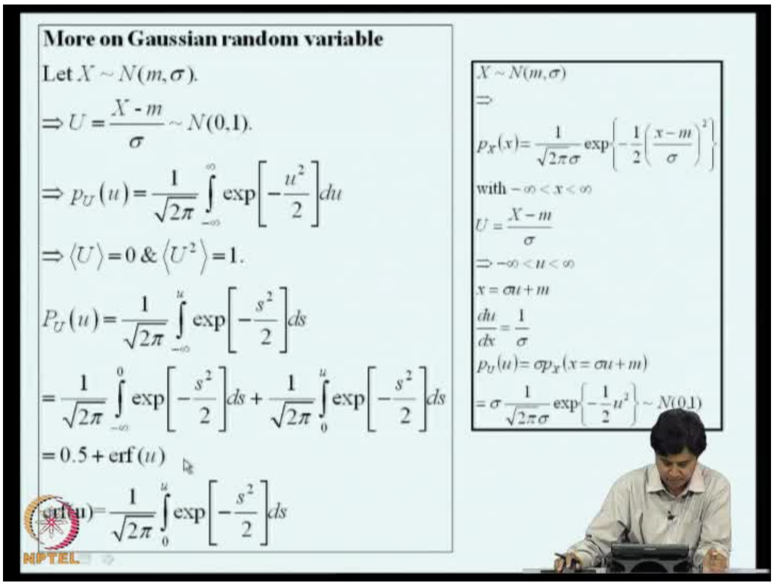
(Refer Slide Time: 09:36)



So, in modeling of a skewness and kurtosis are used to check non-Gaussian, it departure from value of skewness value of 0 and kurtosis value of 3 indicates, that you are dealing with a non-Gaussian random variable. We can also evaluate the characteristic function which is shown here, and from this, of course, we can again derive all the moments using the property of the characteristic function.

We can now ask the question, what constitutes a complete description of a Gaussian random variable? A Gaussian random variable has two parameters, namely - mean and standard deviation, that is m and sigma, which just now we have verified, that they have the meaning of being mean and standard deviation. So, a Gaussian random variable therefore is completely specified in terms of its mean and standard deviation. So, often when you use a Gaussian random variable, we talk only about mean and standard deviation, given that there enough to characterize an underline Gaussian density function. If a random variable is not Gaussian, then mean and standard deviation are not adequate to completely specify the random variable.

(Refer Slide Time: 10:27)



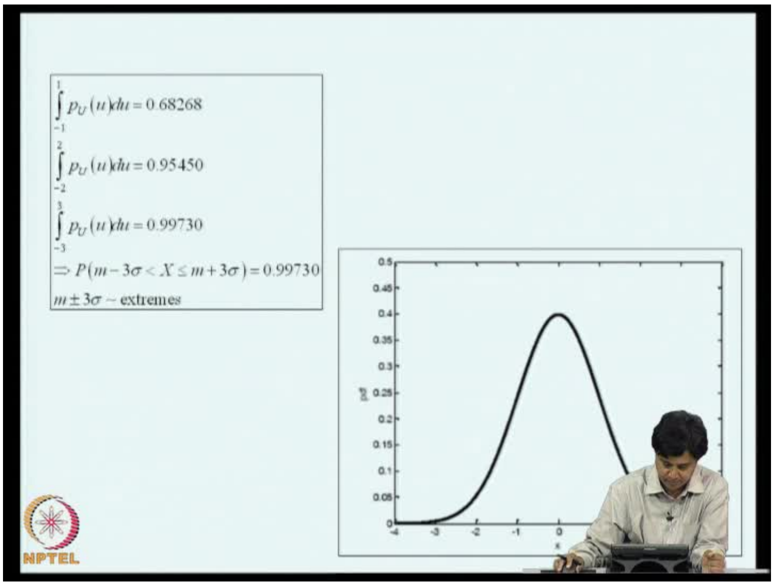
We will discuss a bit more on Gaussian random variable, let X be a normal random variable, with mean m, parameter m and sigma; now, I introduce a new random variable X minus m by sigma.

We will show now, that this has this remains Gaussian with mean 0 and standard deviation 1; now, that proof of that, is in this box. So, when we say X is a Gaussian, what is meant is the probability density function, is given by this, and if we make the transformation u is equal to X minus m by sigma, it follows that u takes values in minus infinity to plus infinity, because X takes values in minus infinity to plus infinity, and using the rules of transformation of random variables, we can show that the probability density function of u is given by this function, which is nothing but the probability density function of a Gaussian random variable, with mean 0 and standard deviation unity.

Another important property of u can be noted, namely that u is non dimensional, that is, because we have now got a meaning for sigma, that it is a standard deviation and m is the mean; so, both mean and standard deviation have the units of X; therefore, this ratio of X minus m by sigma will be non-dimensional. So, this is an important simplification which we often exploit in modeling. Now, this is the probability density function of this u, which is also known as standard normal random variable, by that what is meant, is the mean is 0 and standard deviation is 1; so, mean is 0 and mean square value is 1 here, since mean is 0, that mean square value itself is the variance which is unity.

Now, the probability distribution function associated with this density function is given by this integral, which is minus infinity to u into this probability density function. This can be split as minus infinity to 0 plus 0 to u of this quantity; since, probability density function is symmetric about u equal to 0 and area under the probability density function is 1, it follows that the value of this integral is half and this integral is known as error function this erf of u, should be write as error function evaluated at u; this u appears as an upper limit in this integral, that is this, in this a tabulated function and in many software, it is a function like any other trigonometric or exponential functions.

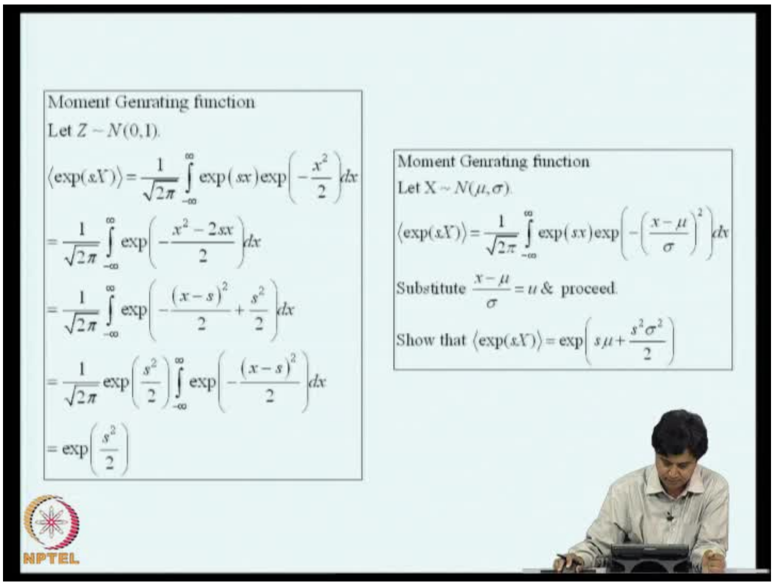
(Refer Slide Time: 13:26)



Now, how does a standard normal random variable look like? This is how it appears, this is a probability density function, and this is X, and it takes at 0, which is the mean value. And if you now evaluate the area under the curve from minus 1 to plus 1, that means, remember, that 1 is a standard deviation here; that means, the probability content in minus sigma to plus sigma, that is 1, standard deviation turns out to be 0.68268. If you compute the probability between two sigma levels, minus 2 sigma to plus 2 sigma, sigma is 1, it turns out to be 0.95450. If you compute between minus 3 sigma to plus 3 sigma, this area is particulars to 1.99730.

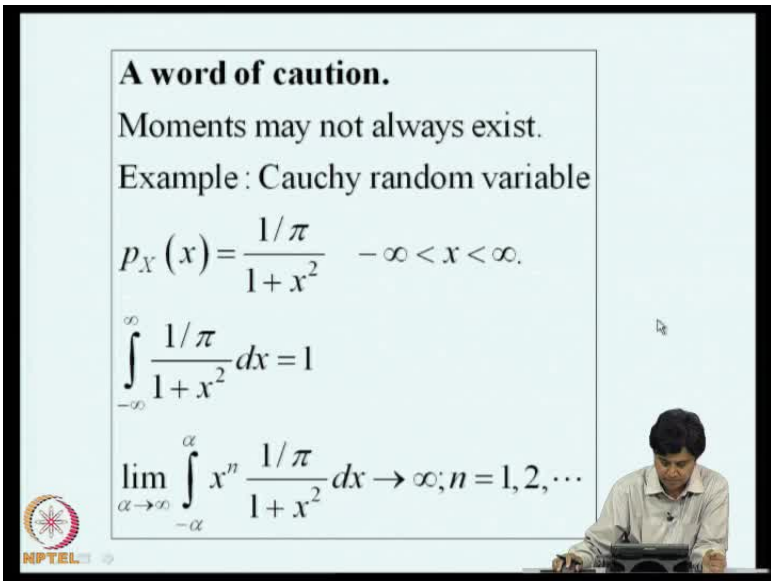
So, the probability that X takes values between mean minus 3 sigma and mean plus 3 sigma, here I am derived the probability with respect to u, but in terms of X, we can show that for a Gaussian random variable, the probability that X takes values between m minus 3 sigma and m plus 3 sigma is 0.99730; so often, mean plus minus 3 sigma are taken to be the extremes of a Gaussian random variable; mean plus 3 sigma is roughly the highest value and mean minus 3 sigma is roughly the lowest value. Remember, a Gaussian random variable takes values from minus infinity to plus infinity.

(Refer Slide Time: 15:07)



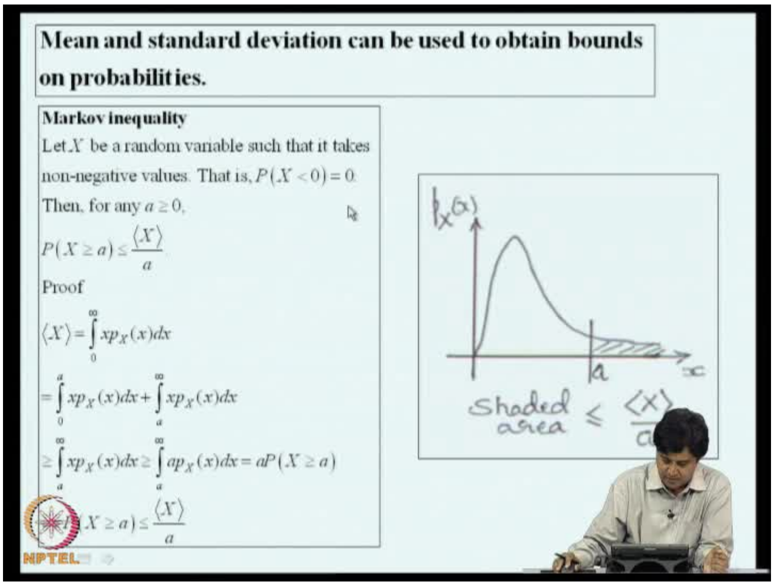
How about the moment generating function for a Gaussian random variable, this you recall is like a Laplace transform of the probability density function, which is given by expectation of exp s X, and if we evaluate this integral, we can show that, this is given by exp of s square by 2, this is for a standard normal random variable. And for a normal random variable with parameter mu and sigma, one can show that, this is given by exponential of s mu plus s square sigma square by 2.

(Refer Slide Time: 16:08)



So, to completely specify the Gaussian random variable, we can as well specify the moment generating function, which again contains only two parameter mu and sigma. While talking about moments a word of caution is necessary, one should not assume that in a modeling exercise that moments always exist. There are random variables for which moments do not exist, for example, a Cauchy random variable is given by 1 by pi 1 plus x square, where x takes value from minus infinity to plus infinity, it can easily be verified that area under this curve is unity; therefore, this is the valid probability density function, but if you try evaluating the expectation of x to the power of n, it turns out that; suppose, n equal to 1, this integral actually diverges, so the moments do not exist. So, if we are actually dealing with a Cauchy random variable and you try modeling its mean and standard deviation things like that, it will not work, there will be problem in your modeling.

(Refer Slide Time: 17:09)

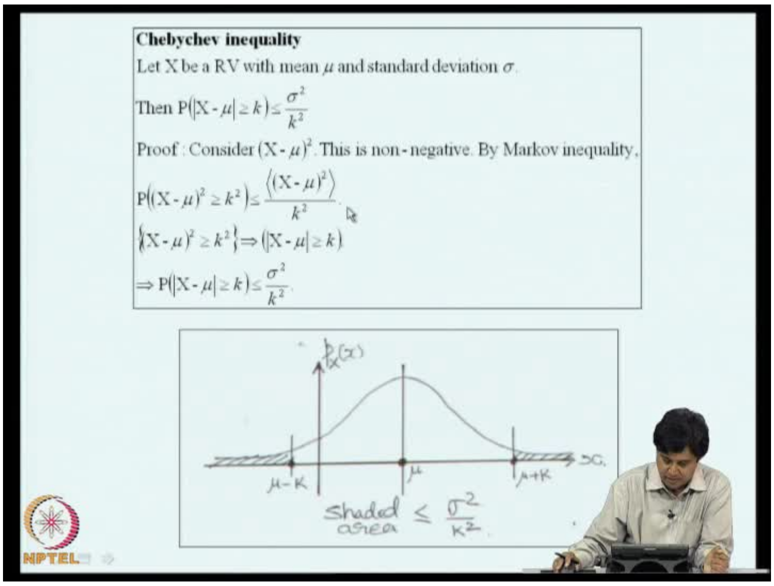


Now, mean and standard deviations are important descriptors of random variable, from another point of view namely, that they can be used to obtain bounds on probabilities, remember that, mean and standard deviations are simply first and second moments associated with the first and second moments, they, themselves do not directly enable you to make any statements on probabilities. There is one inequality known as Markova inequality that can be described as follows.

Let us consider a random variable X, such that it takes non-negative values; that mean, probability of X less than equal to 0 is 0. Then, for any a greater than 0, probability of X less than or equal to, greater than or equal to, a is less than or equal to this ratio; this is the statement of Markova inequality. This can easily be verified by looking at the expression for the mean, the expected value of X is 0 to infinity x p X of X dx, this can be written as 0 to a plus a to infinity x p X of x dx.

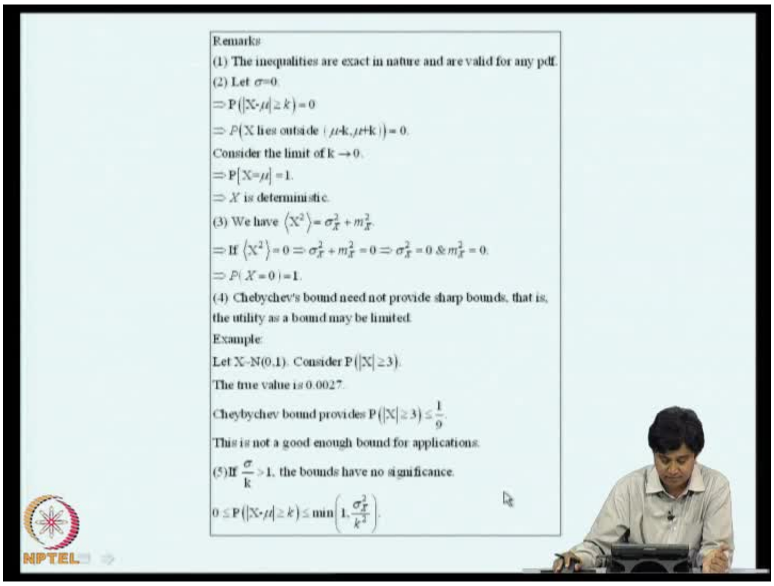
Now, from this it follows, that this expected value is greater than or equal to this value, because we are removing a non-zero term from this sum; therefore, this become greater than or equal to this, and this itself or a is somewhere here, and I am trying to find area a to infinity of x p X of dx, now this inequality is true, because I can replace x by a and I can get this and therefore this in fact is equal to a into probability of X greater than or equal to a. So, from this, it follows that probability of X greater than or equal to a less than this ratio, and this is valid for any random variable X, which satisfy this requirement, that probability of X less than or equal to 0, is 0.

(Refer Slide Time: 19:25)



Based on this, we can also derive what is known as Chebychev inequality, to see that, we will consider a random variable X, with mean mu and standard deviation sigma. According to Chebychev inequality, the probability, that modulus of X minus mu is greater than equal to k, is less than this ratio; so, here you can see there are two parameter mu and sigma, that are entering the inequality. To prove this, we consider the quantity X minus mu whole square; this again a random variable, which is strict probability of this being negative is 0; therefore, I can use Markova inequality and make this statement because this statement of Markova inequality. Now, the probability of X minus mu whole square greater than or equal to k square itself, can be written in this form; this event can be written like this, and therefore it follows, that the statement of Chebychev inequality is proved.

(Refer Slide Time: 20:50)



So, what this inequality says, this, the shaded area which is actually probability that X takes values, this interval union, this interval is less than this quantity; this is exact and it does not depend on distribution of X.

So, we will quickly make few remarks, the Markov and Chebychev inequalities are exact in nature and they are valid for any probability distribution function. Now, let us consider a situation where standard deviation is 0; so, probability of X minus mu greater than or equal to k is 0, because inequality is less than or equal to sigma square by k square. Now, sigma is 0, therefore the equality holds, because this probability cannot be negative, so equality holds; so, that means, probability of X lies outside mu minus k to mu plus k is 0.

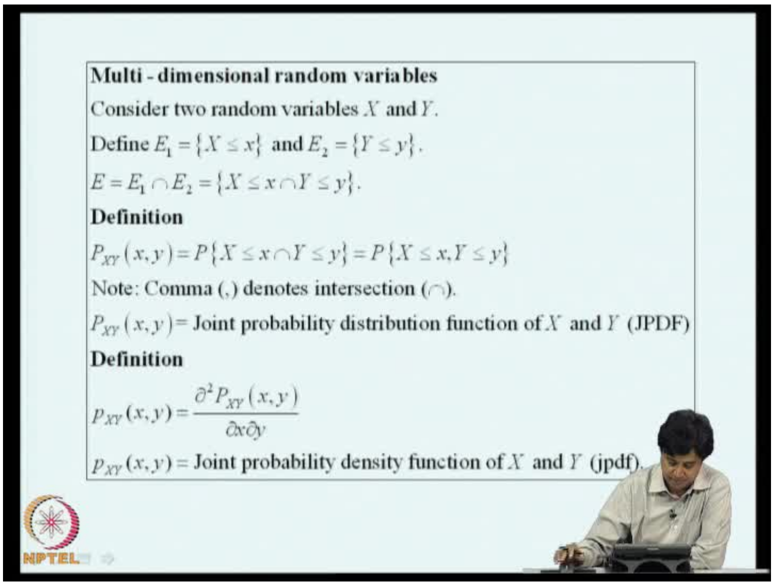
Now, if we consider limit of k going to 0, it means probability of X equal to mu is 1. So, if standard deviation is 0, we can treat X as deterministic. So, this is an important property of standard deviation, because it is not immediately apparent, if standard deviation is 0, it is quite possible that higher moments may not be 0, I mean, you can (( )), that situation at least in mind, but from, it, this follows that X is actually deterministic standard deviation is 0.

Now, similarly, we have X mean square value of sigma X square plus m X square; so, if mean square value is 0, it again follows probability of X equal to 0 is 1; this another property of mean square value, again a consequence of this inequalities.

A word of caution again Chebychev's bound need not provide sharp bounds, that is, the utility of Chebychev bound, as a bound may be limited. To show that, let us consider a normal random variable, with mean 0 and standard deviation 1 and let us consider the probability of modulus of X greater than or equal to 3, the true value of this is, you can easily compute, it is 0.0027.

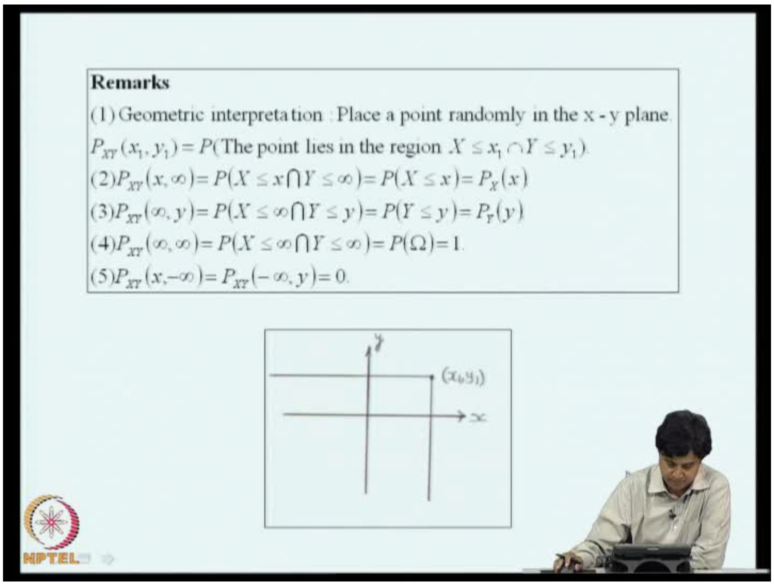
Now, according to Chebychev bound, this is actually 1 by 9 and 1 by 9 is way of from 0.0027; so, 1 by 9 as a bounds for this number, this not a very good bound. Again, if sigma by k is greater than 1, the bounds are no significance, because probability it cannot be great than 1 itself, is a bound imposed by axiomatic definition; so, this carries those special meaning, in such situation, to make that precise we can say that, probability of this event is greater than or equal to 0 and it is less than equal to minimum of this quantity; here, I have combined the axioms of probability with the Chebychev inequality.

(Refer Slide Time: 23:40)



So for, we have been talking about one-dimensional random variables, we can talk about more than one random variable as well; let us consider, two random variables X and Y, let us define the E 1, E 1 is X less than equal to x and E 2 is Y less than equal to y. And consider the event E, which is intersection of E 1 and E 2, that is X less than or equal to x and Y less than or equal to y. By definition, the probability of this event probability of X less than or equal to x intersection Y less than equal to y is known as the joint probability distribution function of the random variables X and Y. And in our further discussion, we replace this intersection by a comma, comma is should be interpreted as an intersection. Now, we denote the joint probability distribution function of X and Y in upper case JPDF; this is a notation, that we will follow. Associated with this joint probability distribution function, we can now define a joint probability density function of X and Y as the partial derivative of P x y x, y dou x dou y; this is denoted as lower case j pdf.

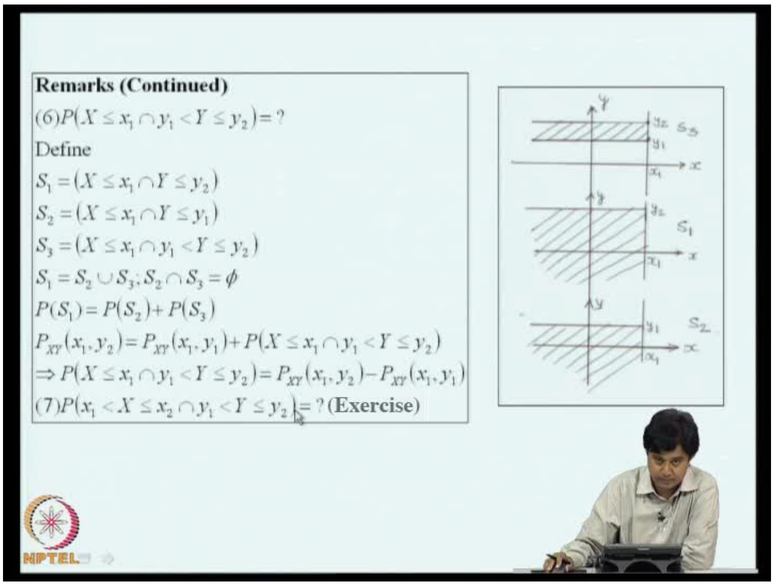
(Refer Slide Time: 25:24)



Now, to interpret the meaning of joint probability distribution function, we can provide a geometric interpretation. So, you consider a plane and consider the Cartesian coordinates system x and y; let us think of a random experiment, where we place a point randomly in this plane. So, we define the probability, that the point lies in the region X less than equal to x 1 intersection Y less than equal to y 1 as P x y (x 1, y 1); that means, point lies in this region below this two lines. From this, now we can see P x y of x, infinity is nothing but probability of X less than or equal to x intersection Y less than equal to infinity. And this is simply probability of X less than equal to x, because when you place a point randomly, it has to lie in the region y less than equal to infinity; so, this is P x of x. This provides you with the relationship between joint probability distribution function and so-called marginal probability distribution function associated with the joint PDF.

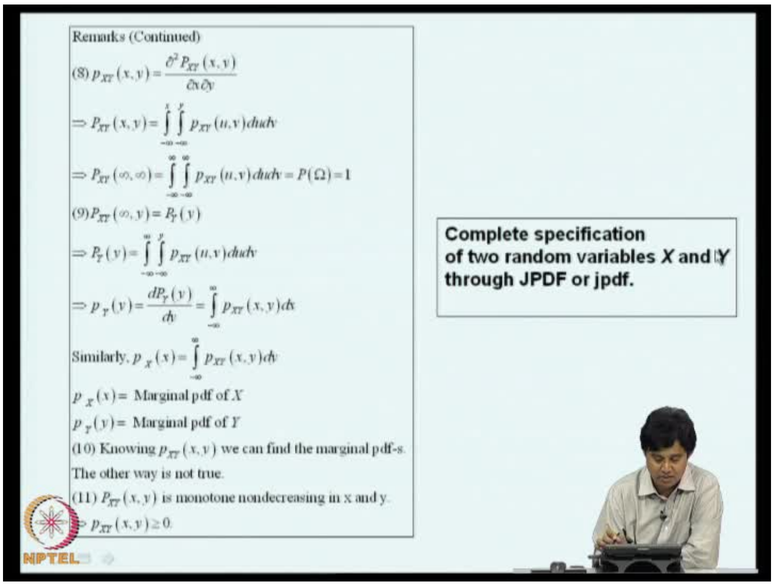
Now, if we now consider P x y of infinity , Y, this is probability of the event X less than or equal to infinity intersection Y less than equal to y, which is nothing but probability of Y less than equal to y this is P Y of y. If you now consider the probability distribution evaluated at X going to infinity and Y going to infinity, we are looking at essentially the sample space. If you place randomly a point in a plane, it will indeed, it will always lie in this region; therefore, that probability is 1. It also follows P x y of x, minus infinity, P x y of minus infinity, y, both are 0, you cannot place a point, you know, below minus infinity.

(Refer Slide Time: 27:37)



We can continue with this geometric interpretation and consider what is the probability of X less than or equal to x 1, that means, region to the left of this line and Y lying between y 1 and y 2, that means, basically this shaded region. To evaluate this, we define three sets S 1 is X less than or equal to x 1 intersection Y less than equal to y 2, that means, this is S 1. S 2 is X less than or equal to x 1 intersection Y less than or equal to y 1, which is shown here. Now, we can verify that S 1 is S 2 union S 3, that means, you take union of S 2 and S 3 you will get S 1 and also intersection of S 2 and S 3 is a null set; therefore, these two sets are mutually exclusive. Therefore, probability of S 2 union S 3 is sum of probability of S 2 and probability of S 3. If we now write this probability in terms of probability distribution functions, we get probability of S 1 is x 1, y 2 here probability of S 2 is x 1, y 1 here and the probability that we are looking for basically here. So, rearrangement of this term shows that the probability, that we are looking for is indeed P x y x 1, y 2 minus P x y x 1, y. So, using this logic, we can try finding the probability that X lies between x 1 and x 2 and Y lies between y 1 and y 2. The exercise is here to find out this probability, in terms of the joint probability distribution function of X and Y.

(Refer Slide Time: 29:29)



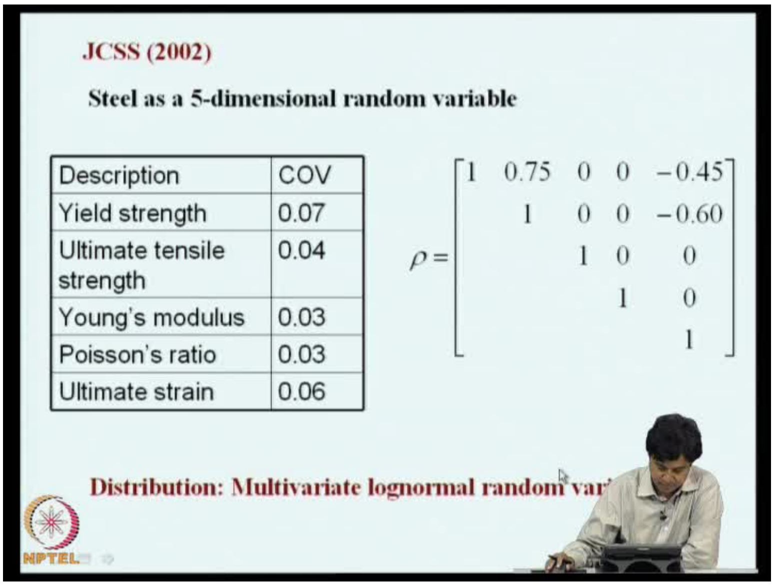
Now, let us look at the joint probability density function; so, this is given as this derivative, and from this, it follows the distribution function is given by this integral, where the probability density function. Now, we have already shown P x y of infinity, infinity is nothing but probability of a sample space which is 1, therefore it follows that this integral minus infinity to plus infinity minus infinity to plus infinity of the joint density function is 1; that means, the volume under the joint density function is 1.

Now, we also have now P x y of infinity, Y is P y of y, so P y of y is given by this, and now if you differentiate this with respect to y, by definition that is the probability density function of y, and we can show that, this is actually the area of the joint probability density function, the integral of this over the variable x. So, similar argument will show that the probability density function of x is given by this integral. We call this P x of x and P y of y as a marginal probability density functions, associated with the joint density function x and y.

It is very clear that, if you know the probability density function, joint probability density function of x and y, you can easily find the marginal density functions of x and y, but if you know the marginal density functions, you cannot construct the joint density function. We can also show that the probability distribution function is monotone non-decreasing in x and y using the arguments, that we use to show the property of a one-dimensional probability distribution function, and from this, it follows the joint density function is strictly, is actually non-negative.

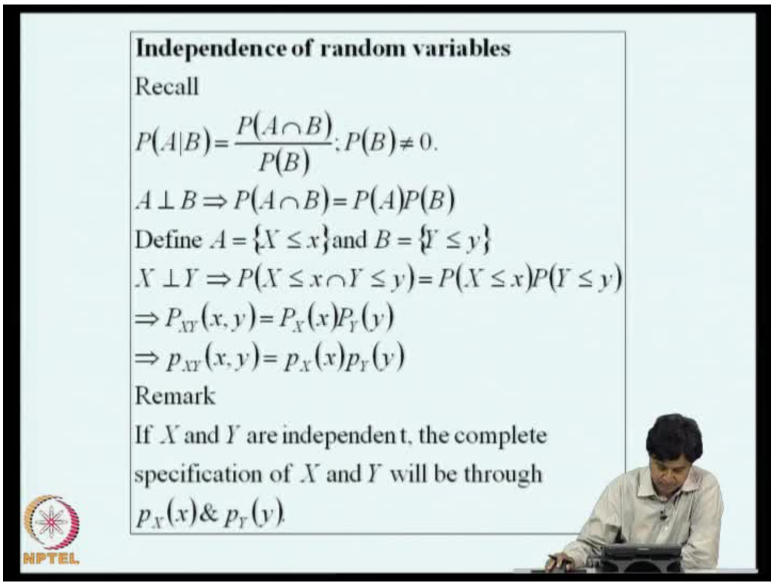
So, at this stage, you can now ask the question what constitutes the complete specification of two random variables X and Y. The answer at this stage, is that, you should know, their joint probability distribution function or the joint probability density function, it is not adequate; if you know the marginal probability density function or distribution of X and Y, you should have the two-dimensional distribution or density function.

(Refer Slide Time: 32:06)



Just say an example, in characterizing steel in structural engineering applications, there is a joint committee on structural safety which specifies, that the steel as a material can be modeled as a five-dimensional random variable; the five characteristic of steel are yield strength, ultimate tensile strength, Young's modulus, and Poisson's ratio, ultimate strain. It is recommended here, that they can be modeled as multivariate lognormal random variables and they are suggested parameter of this, this multi-dimensional random variable.

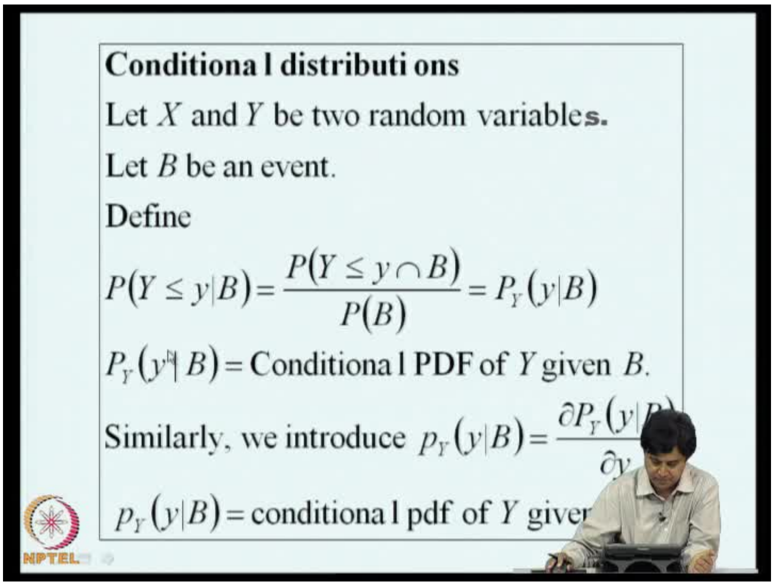
(Refer Slide Time: 32:50)



We would now consider the notion of independence of random variables; we have defined the notion of independence for events; so, if we recall if A and B are events probability of A condition on B, we have defined as probability of a intersection B divided by probability of B, with probability of B being not 0. So, we say that A and the events A and B are independent, if probability of A intersection B is probability of A into probability of B. Now, what we will do is, we consider two random variables X and Y, and define the event A as the event X less than or equal to x, and B as the event Y less than or equal to y.

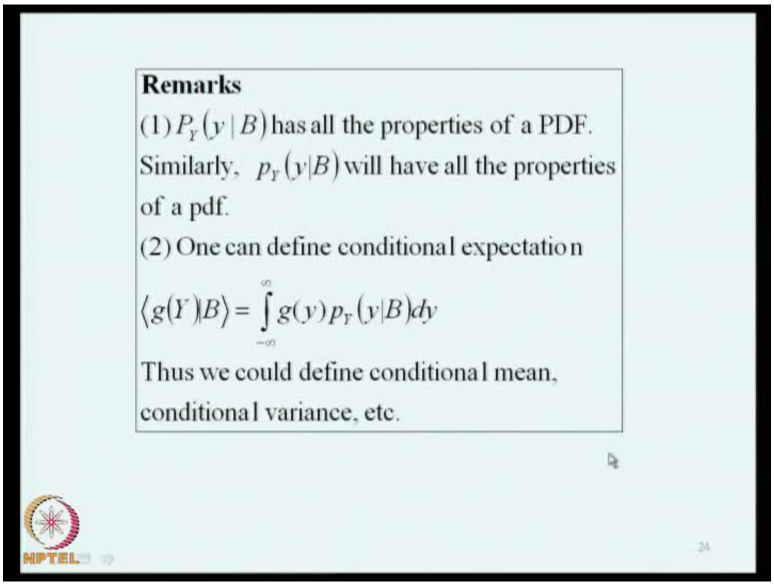
We say that the random variables X and Y are independent, denoted as here, if this probability of X less than or equal to x intersection Y less than or equal to y is given by the product of probability of X less than or equal to x probability of Y less than or equal to y for every X and Y or in other words, the joint probability distribution function of x and y, if they can be expressed as product of marginal probability density distribution functions, then we say X and Y are independent. This, this can be stated in terms of density functions also, if X and Y are independent, then the joint density function is obtained as product of marginal density function. Now, if X and Y are independent, the complete specification of X and Y will be through their marginal density function, because if you know P X of x and P Y of y, you can always construct the joint probability density function.

(Refer Slide Time: 34:48)



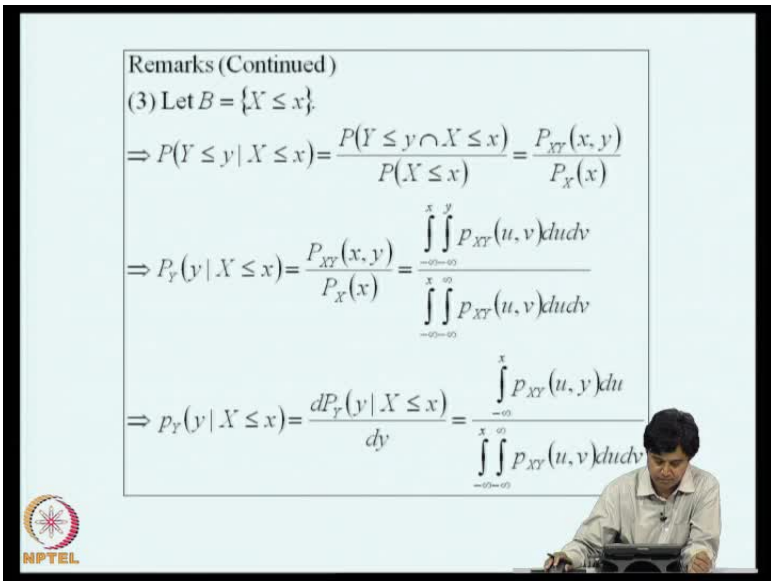
Now, let us continue with the notion of conditional probabilities and let us consider two random variables X and Y. Let B be an event. And now we will consider the probability that Y less than equal to y condition on B ,is according to all definition is this. And this probability is defined as conditional probability distribution function of Y given the event B. And associated with this, I can derive a conditional probability density function of Y given B, as the derivative of this distribution with respect to y.

(Refer Slide Time: 35:33)



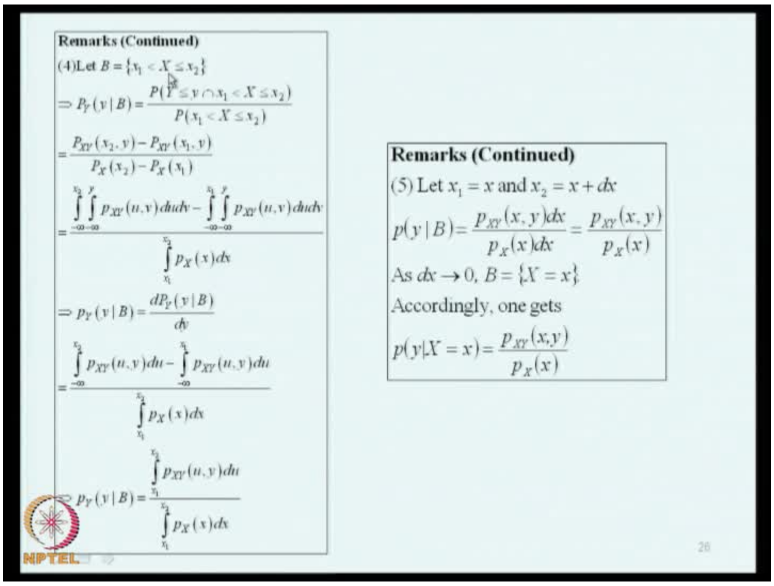
Now, this conditional probability distribution functions and for conditional probability density functions, will have all the properties of a distribution function or a density function. And we could, in fact define a conditional expectation, that is, expected value of g of Y conditioned on B is given by minus infinity to infinity g of Y into the conditional probability density function of Y condition on B, this integral. So, thus we can talk about conditional mean, conditional variance, conditional characteristic function, so on and so forth.

(Refer Slide Time: 36:15)



Now, let us give some specific interpretation to the event B. To start with, let us consider B is X less than or equal to x, the event X less than or equal to x. Now, if we run through this calculation, you can show that the conditional probability distribution of Y condition on B is, now the ratio of the joint distribution of x and y and the marginal distribution of x. This in terms of density functions, can be written as shown here, and if I now differentiate this with respect to y, I will get the conditional probability density function of Y conditioned on the event X less than or equal to x, and that is given by this ratio. We have to differentiate this quantity with respect to Y, which appears as a limit in this integral.

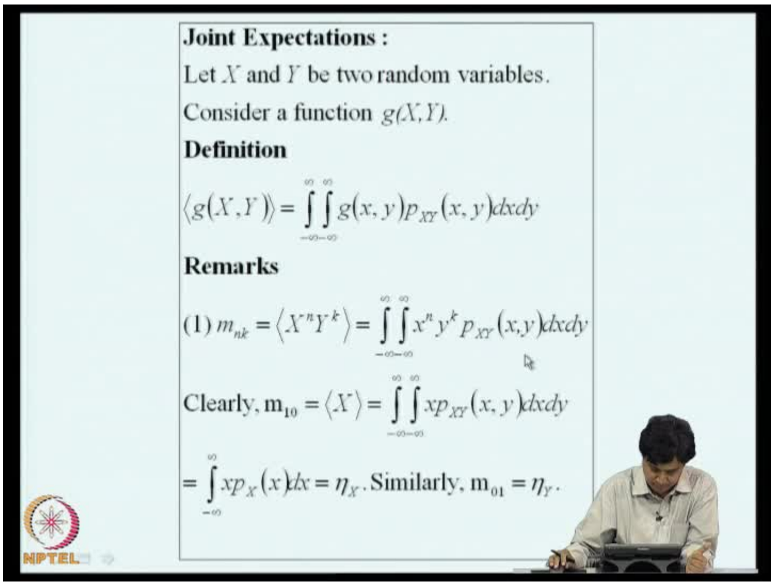
(Refer Slide Time: 37:05)



Now, let us interpret B as the event X greater than x 1 less than or equal to x 2. Now, the probability distribution of Y conditioned on B, is given by this, and we can write this, in terms of the joint probability distribution functions, and the, and the marginal probability distribution function, and now in terms of density functions. And if we differentiate this with respect to Y, I will get the conditional probability density function, which is again given by this ratio and the simplification shows that, this is this.

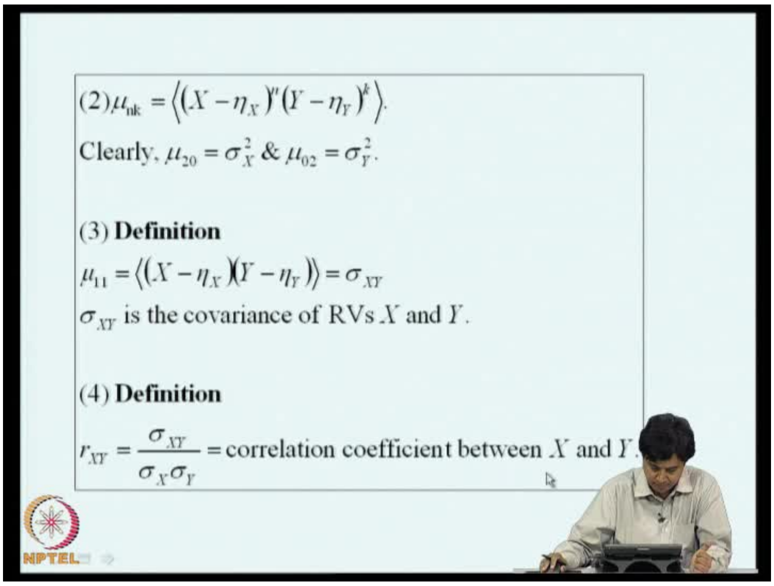
Now, what we will do now is, in the definition of B, I will take x 1 as x and x 2 as x plus dx. Now, I consider the conditional probability density function of y given B, and based on this definition, it turns out that, this integral can be written as now P X Y x y into dx, and this P x of x into dx, and this becomes this ratio, and this ratio independent of dx; therefore, if I take dx going to 0, the event B itself will be X equal to x. Here, if I take dx going to 0, this event is nothing but X equal to small x. So, we get an important result, namely the conditional probability density function of y condition on the event X equal to x, is given by this ratio. This is very important result in modeling; we will have opportunity to use this later.

(Refer Slide Time: 38:54)



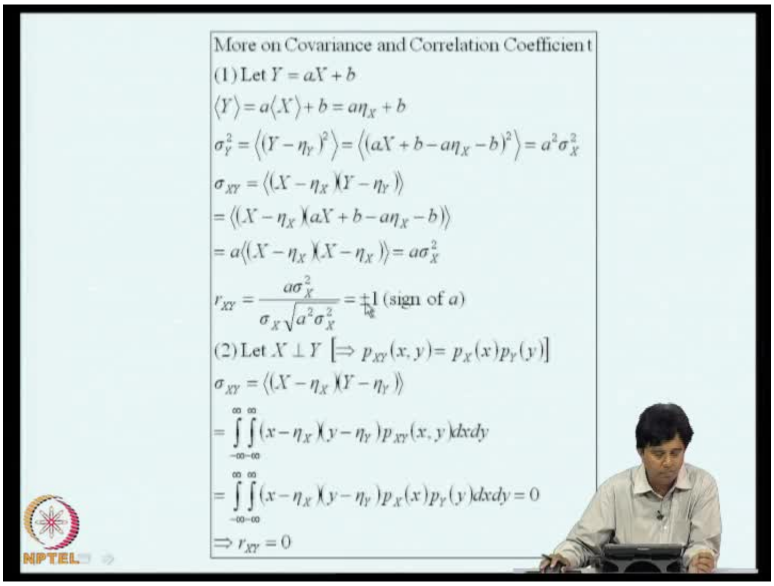
We have talked about the notion of expected value of a random variable, that notion now can be extended to joint expectation of two random variables. Suppose, again let us consider X and Y to be the two random variables and now let us consider a function g of X, Y. By definition, the expected value of g of X, Y is given by this integral; this is minus infinity to plus infinity double integral g of X Y into the joint probability density function. Now, by given specific interpretations to this function g, we can derive certain useful results, for example, if I now take expected value of X to the power of n into Y to the power of k, call it as m n k, this become this integral; and in this, for example, if I take k equal to 0 and n equal to 1, this is nothing but expected value of X is now given in terms of joint density function X into P x y x, y dx dy, but we already know that if we integrate joint density function with respect to y, you get the marginal density function of x; therefore, this double integral is nothing but x into P x of x dx, which is what we defined as expected value of x earlier. So, this definition of expectation is consistent with what we have done earlier for a single random variable x. So, this in, this we denote as eta of X, that is mean of X; similarly, the quantity m 01 will be the mean of Y.

(Refer Slide Time: 40:38)



We can now define mu n k as expected value of X minus eta X to the power of n Y minus eta Y to the power of k. And from this, you can verify that mu 20 is expected value of X minus n eta X to power of 2 into 1 is nothing but variance of X; similarly mu 0 2 is variance of Y. We introduce a definition, now we call the quantity mu 11, which is expected value of X minus eta X into Y minus eta Y, has the covariance of random variables X and Y. The covariance will have units of product of units of X and multiplied by product units of Y. To make it non-dimensional, we divide this quantity by the standard deviation of X and standard deviation of Y and this quantity is known as correlation coefficient between X and Y.

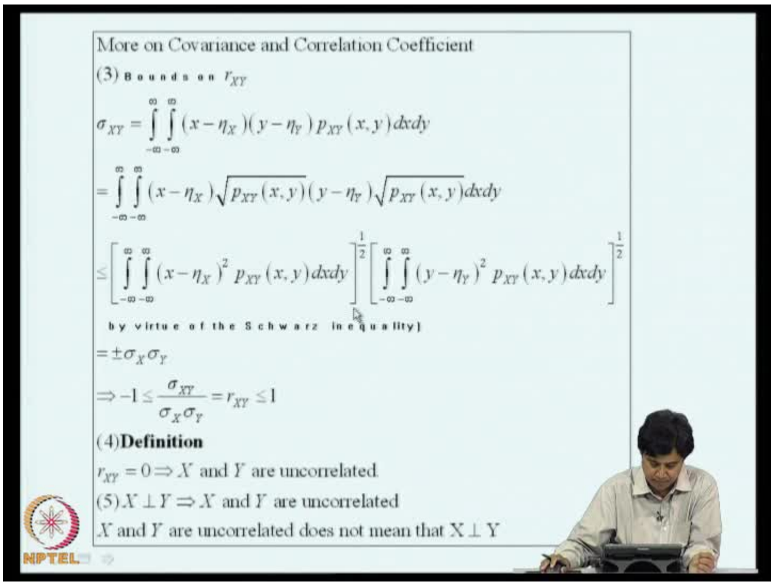
(Refer Slide Time: 41:48)



What does this covariance and correlation coefficients mean? To see that, let us consider two random variables X and Y, which are related through this relation Y equal to a X plus b. Now expected value of Y is a eta X plus b variance of Y is Y minus eta Y whole square, where Y a X plus b I, substitute this, I get a square sigma X square covariance will be expected value of X minus eta X Y minus eta, Y for Y minus eta Y, i am writing a X plus b minus a eta X plus b, that is this, and this turns out to be a into sigma X square. If you now substitute in this into the definition of correlation coefficient, we see that the correlation coefficient is either plus 1 or minus 1, depending on sign of a, that would mean, if there are two random variables X and Y, which are linearly related the correlation coefficient will take value either minus 1 or plus 1.

Now, what happens if X and Y are independent, here the joint density function is nothing but product of a marginal density functions. So, if you now consider the covariance, it turns out that, this will be 0, because this X minus eta X Y minus eta Y; p X Y x, y dy, this p X Y x, y is actually the product; therefore this double integrals splits into single integrals and each of this integral is 0. Therefore, the correlation coefficient in this case turns out to be 0, that is if X and Y are independent, the correlation coefficient is 0; if X and Y are linearly related the correlation coefficient is either plus 1 or minus 1.

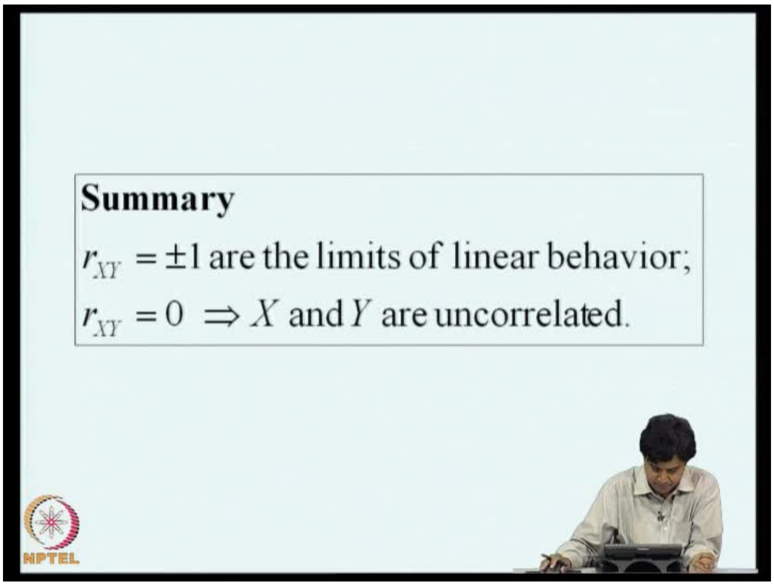
(Refer Slide Time: 43:43)



Now, indeed we can show that, plus 1 r X Y actually bounded between minus 1 to plus 1. To show that, we start with the definition of covariance and rewrite, this function in a slightly different manner, we write this as x minus eta X into square root of this density function and y of eta Y into square root of density function; so, this is exactly equal to this. Now, we apply on this integral, this equality and it turns out that the sigma X Y will be less than or equal, to these individual integrals, product of these two integrals. These are nothing but the standard deviations, this is standard deviation of X and this is standard deviation of Y.

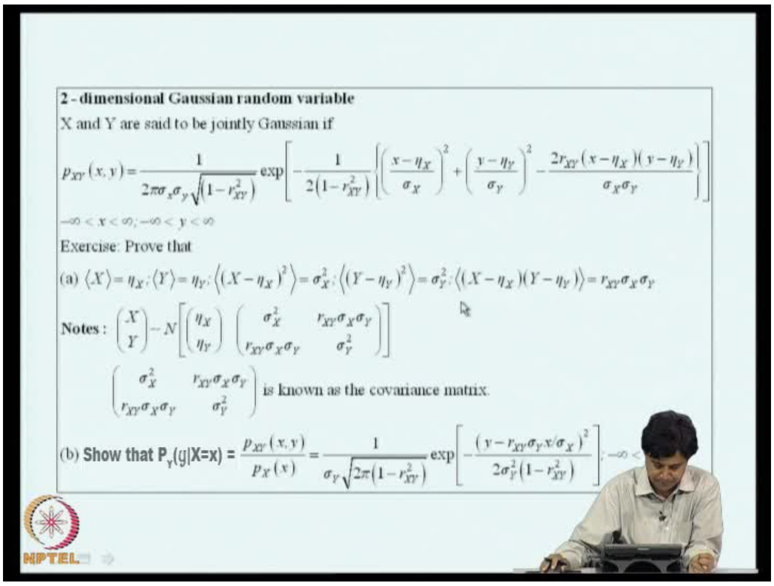
So, from this, it follows that correlation coefficient lies between minus 1 to plus 1. So, we can now introduce a definition, if correlation coefficient is 0, we say that X and Y are uncorrelated. Actually, if X and Y are independent, we already shown that, they are uncorrelated, but you must bear in mind that if X and Y are uncorrelated, it does not mean that X and Y are independent, is a much stronger property correlation, is the property defined on moments independence, is the property defined on probability density function; therefore, it is a stronger property.

(Refer Slide Time: 45:13)



So, in summary r X Y equal to plus minus 1 are the limits of linear behavior and if r XY equal to 0 X and Y are uncorrelated. So, this is a very useful modeling tool, if we are dealing with data and modeling of uncertainties.

(Refer Slide Time: 45:32)

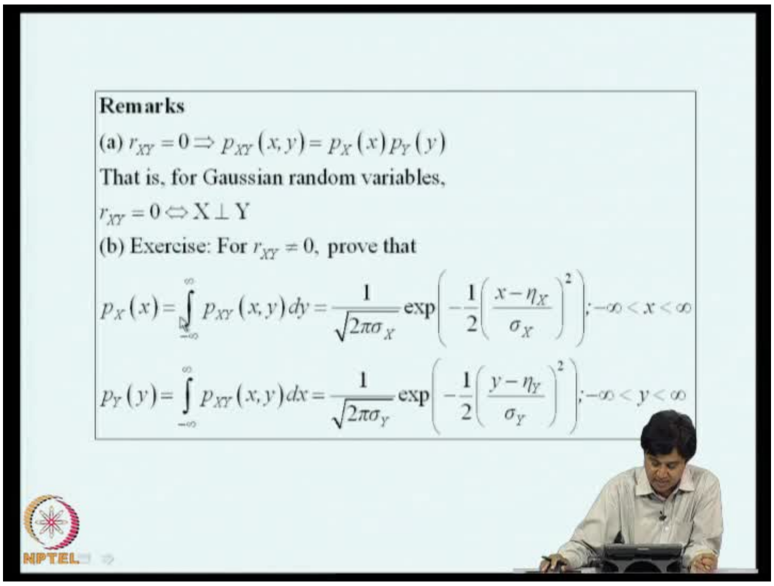


To elaborate the meaning of a multi-dimensional random variables, let us consider a two-dimensional Gaussian random variable, the definition is as follows X and Y are said to be jointly Gaussian, if the joint probability density function of X and Y is given by this lengthy expression, where you will see, there are, now five parameters sigma X, sigma Y, r X Y, eta X and eta Y.

So, a joint density function of two normal random variables requires five parameters. Now, these parameters themselves have a direct meaning, you can show that, this parameter eta X is nothing but mean of X, the parameter eta Y is nothing but mean of Y, and this parameter sigma X is the standard deviation of X sigma, Y is the standard deviation of Y, and this r X Y is the correlation coefficient between X and Y.

So, we write that a two-dimensional normal random variable is represented as five parameters eta X, eta Y, and this quantity written as a matrix, variance here, and covariance here, and variance here, this quantity is known as the covariance matrix. So, a two-dimensional Gaussian random variable is completely specified, in terms of its mean vector and the covariance matrix.

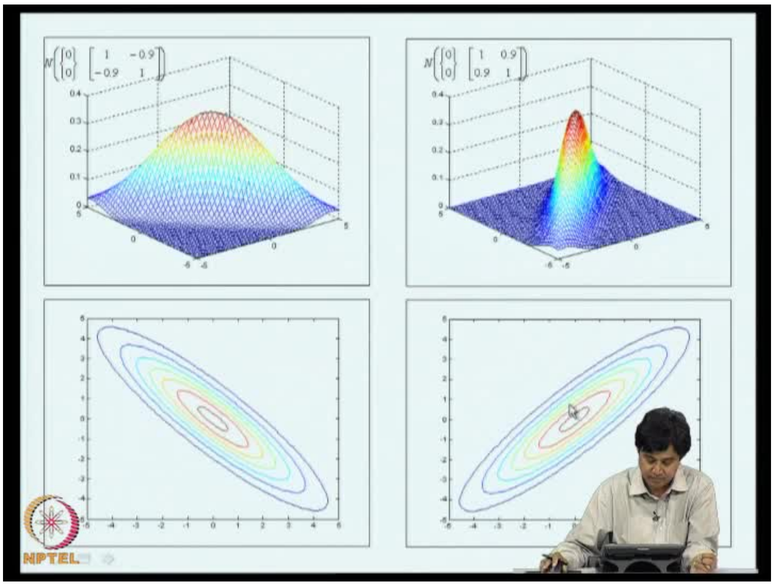
(Refer Slide Time: 47:36)



It can be shown that, if you consider the conditional probability density function of Y condition on X equal to x, given by this is actually another Gaussian probability density function; this I leave it as an exercise. Now, if in the two-dimensional Gaussian probability density function, we put correlation coefficient as 0, we get that the joint probability density function is indeed, the product of the marginal density function; that is, for Gaussian random variables uncorrelatedness and independence are synonyms, that means, if X and Y are uncorrelated, it also means X and Y are independent;

independence of X and Y automatically implies uncorrelatedness. Now, if correlation coefficient is not 0, I leave it as an exercise, that if we evaluate the marginal probability density functions, you will recover the one-dimensional normal density functions, that we have described earlier, which is as it should be; so, this exercise of carrying out this integral is left as an exercise.

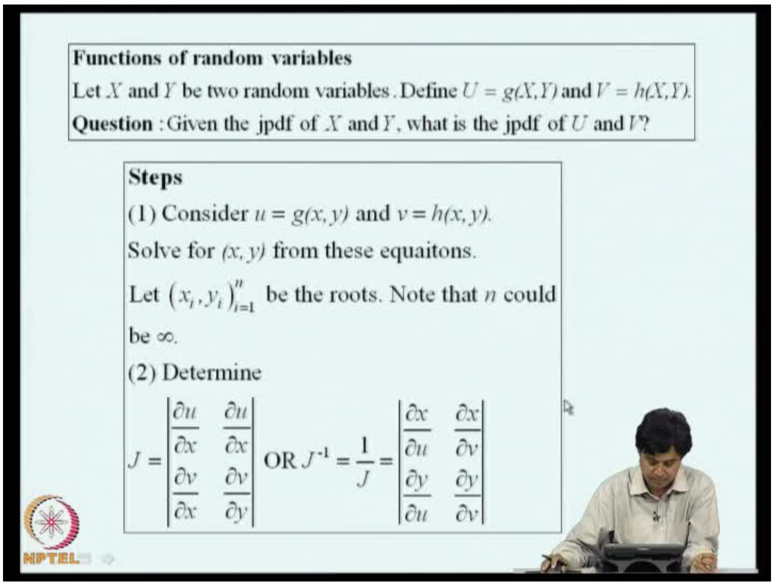
(Refer Slide Time: 48:32)



How does this two-dimensional probability density function look like? So, if we have a normal two-dimensional, normal random variable with mean 0, and unit standard deviation, and covariance is 0 that is uncorrelated. It will be a hill like structure, symmetric about the vertical axis at X and Y equal to 0 and the contours look like this; they are circle, if these axes are sampling, they are circles.

Now, if we now introduce a correlation coefficient of 0.5, the contours now get distorted, they are symmetric about now in inclined line. And if you put minus0.9, there inclined in the other way; the contours are elliptical here as you see. So, similarly with plus 9, it is in this shape.

(Refer Slide Time: 49:31)



Now, when we discussed single random variables, we talked about functions of random variables. The problem of functions of random variables or transformation of random variables can be now considered for two-dimensional random variables also; this statement of the problem is as follows. Let us consider two random variables X and Y and we have the joint probability density function of X and Y given by P x y of X,Y, and we introduce two functions u is g of X, Y and V is h of X, Y; these functions g and h are given.

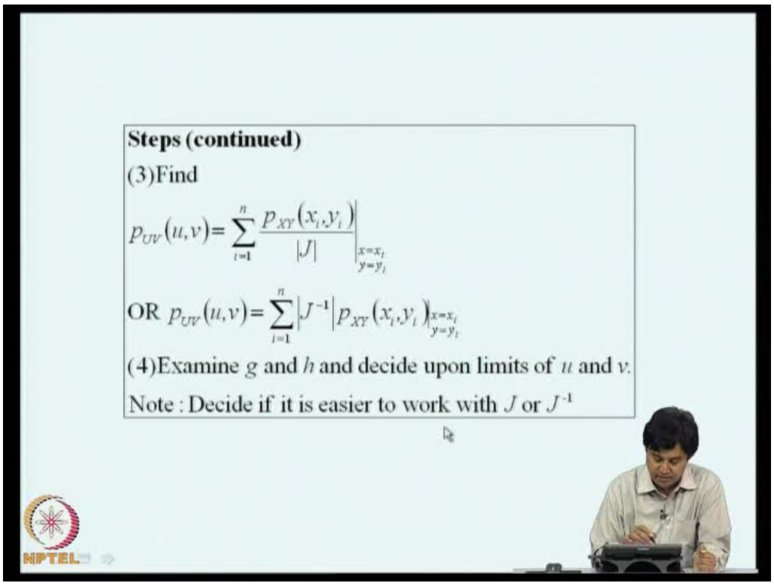
So, what are given, the joint probability density function of X and Y is given the definition of g and h are given, based on this the question is to determine the joint probability density function of U and V. So, this is the statement of problem of transformation of two-dimensional random variables.

The steps to solve this problem are quite similar to what we did for the case of scalar random variable; so, I will outline the steps. The first step would be to consider the functions u equal to g x y and v equal to h x y, and find the roots of these equations, that by solving these equations for x and y. And if we say that x I, y i equal to 1 to n are the roots, n can be infinity, these are the roots; in the first step, we find the roots.

In the second step, we evaluate the Jacobean, this is a determinant dou u by dou x dou u by dou y dou v by dou x dou v by dou y are the reciprocal of that, given by this;

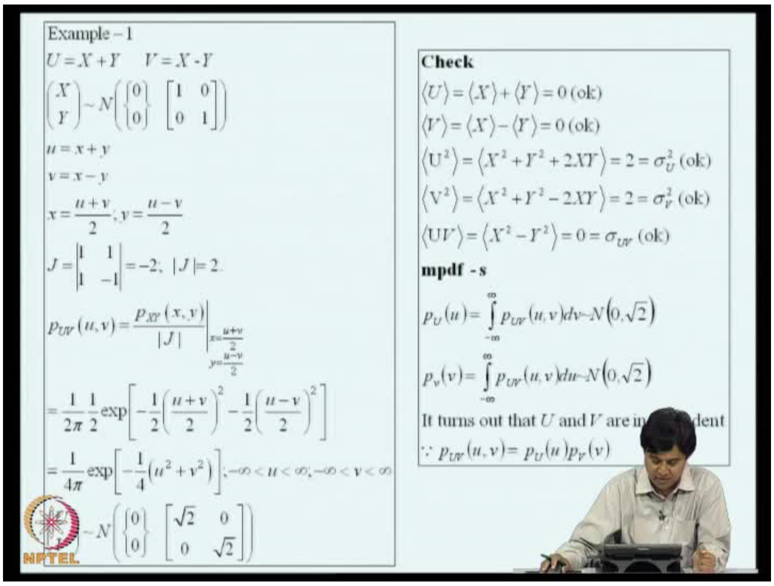
depending on the context, it may be easier to find this or easier to find this, so you have to see which is easy in a given problem.

(Refer Slide Time: 51:37)



Then we evaluate the joint density function of u and v using this formula; this is P x y of x i, y I, and these roots have to expressed in terms of u and v, and based on that, we get the requisite joint density function. After we have arrived at this step, we have to now fix the limits for u and v, for that we have to examine the nature of g and h, and based on that inspection we have to assign the limits of u and v.

(Refer Slide Time: 52:23)



So, this can be illustrated using a few simple examples; so, I will start with one example. Consider X and Y to be normal random variables with 0 mean vector, and identity as it is covariance matrix, and we introduce the transformation U equal to X plus Y and V equal to X minus Y. So, a first step is u is x plus y and v is x minus y, and I solve for a x and y in terms of u and v, J is dou u by dou X is 1 dou u by dou X is 1; similarly, dou v by dou X is 1, dou v by dou X is minus 1, so from this, I get this as minus 2 and this modulus is 2.

So, therefore P U V of u, v is p X Y of x, y divided by this J, evaluated at x equal to u plus v by 2 y equal to u minus v by 2; so, based on the known description of the joint probability density function of X and Y, we write this and make this substitutions. And after rearranging we get this expression; based on the inspection of these results if X and Y are Gaussian, they take values from minus infinity to plus infinity; so, it automatically follows U and V also take values from minus infinity to plus infinity; so, we fix the limits as well.

Now, comparing this with the joint density function of a Gaussian, appear of Gaussian random variables, it can be deduced that U and V are also normal with 0 mean vector and covariance matrix given by this. This again is an illustration of a basic result, namely that Gaussian random variables are closed under linear transformation. So, actually what we are doing is, we are applying a linear transformation on X and Y which are Gaussian and we are getting joint density function of U and V to be Gaussian; that means, the Gaussian nature is retained in linear transformation.

We can check whether this is true by using the definition of expectations, if I now find expected value of U it is expected value of X plus expected value of Y, which is 0; similarly, expected value of V is again 0, so these 0, 0 are ok. Now, mean square value of U is X plus Y whole square which is this, and this is 2, because expected value of X square is 1, this is Y square is 1, and this is independent, because of diagonal terms are 0; so, this is 2, this is 2, since mean is 0, mean square value themselves are the variances; therefore, this covariance matrix is also ok.

We could also now find the marginal density function by integrating this joint density function over v, if you do you get u and if you do our u, you get v and here we get 0 mean, and square root of 2 as standard deviation, and in turns out that, of that, in this case, U and V are independent, because if you take P U of u into P V of v, you actually get the joint density function P U V of u, v. So X and Y are Gaussian, U and V are also Gaussian; X and Y are independent, U and V are also independent, in this case. So, with this, we will conclude this lecture.