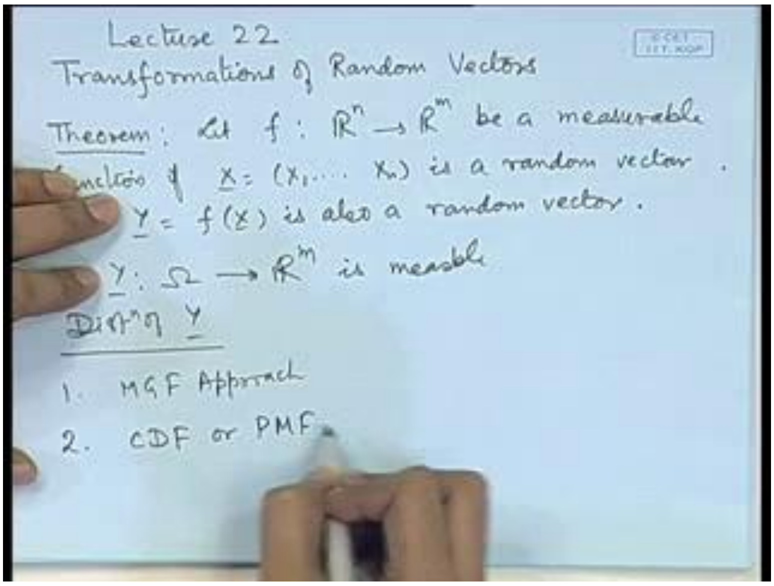
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 43
Transformation of Random Variables

We have seen the distributions of several random variables, many times we are not interested in the original random variable itself, but certain function of it. For example, sums of random variables are say different are any linear function of those random variables.

(Refer Slide Time: 00:41)

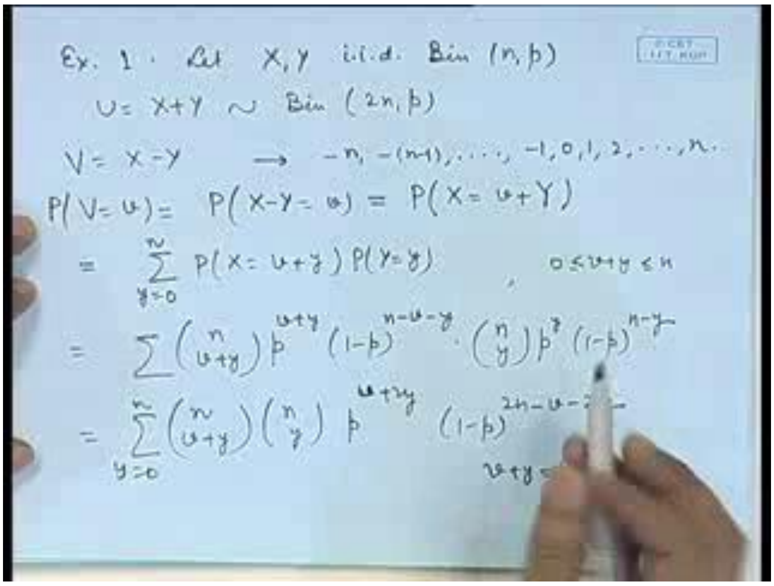


So, in general if I have a measurable function of random vector X 1, X 2, X n then it will also be a random variable. So, we stated in the form of following theorem: let f from say R n to R m be a measurable function, so if X is equal to say X 1, X 2, X n is a random vector, then let us call it say Y; Y is equal to f X is also a random vector. This is so, because random variable X is a measurable function from omega n to R n and a measurable function of a measurable function is measurable function. So, Y becomes a measurable function from basically omega into R n. So, this is measurable and so this is a random vector.

So, now the methods of determining the distribution of Y. So 1 is the MGF approach: we have already seen application of this approach in determining distributions of sums of certain random variables. So, if we are having certain independent random variables and we want the distribution of the sum then it is the distribution, it is the product of the individual MGF and in many cases where the product of the MGF can be determined and explicit form as an identifiable MGF, then the distribution of sum can be determined. It can also be used for distribution of difference etcetera where the forms are very defined.

In the case of discrete distributions or in certain other cases where the CDF can be directly used, then we can use directly the CDF or the probability mass function.

(Refer Slide Time: 03:18)

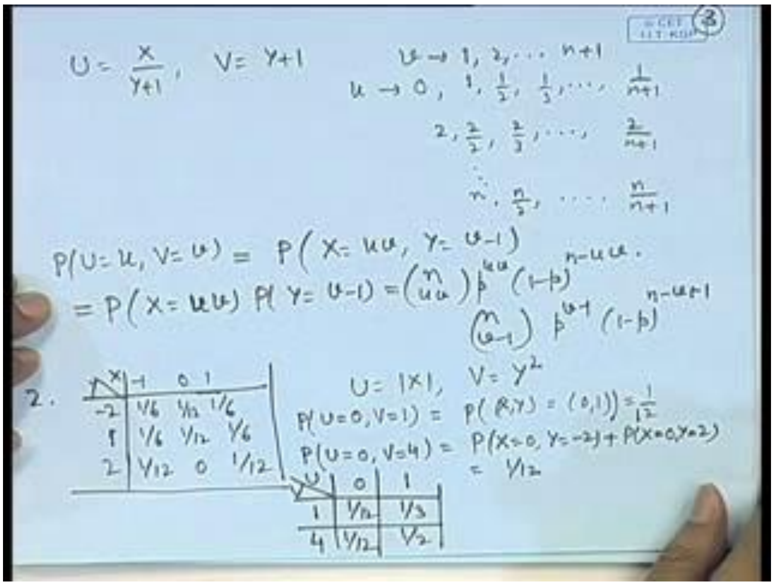


Let me give an example of this suppose X and Y are independent and identically distributed binominal n, p variables. Suppose we want the distribution of U that is X plus Y then from MGF approach we are able to determinatives binominal 2 n, p. Now suppose we want the distribution of say V that is X minus Y, then let us look at the set of values of V this will vary from minus n, minus n minus 1, minus 1, 0, 1, 2 up to n; because each of X and Y can take values 0, 1, 2 n. So, probability of V is equal to says small v that is probability of X minus Y is equal to v; this we can write as X minus is equal to v plus Y. Now Y can take values using a binomial distribution n p. So, we can use the theorem of total probability here and write it as probability X is equal to say v plus y into probability of Y is equal to y; this is because of independence I can split for y is equal to 0 to n.

Now, this is subject to the condition that v plus Y is also lying between 0 to n. So, this is equal to n c v plus y, P to the power v plus y, 1 minus p to the power n minus v minus y, n c y, P to the power y into 1 minus p to the power n minus y. So, this is equal to sigma n c v plus y; n c y, p to the power v plus 2 y and 1 minus p to the power 2 n minus v minus 2 y, where y is equal to 0 to n, subject to the condition that v plus y is also taking value 0 1 2 n, because v plus y denotes the value of the random variable X here. So, this shows that in the case of discrete random variables, directly the probability mass function can be use to determine the distribution of a function.

Let us take another case here.

(Refer Slide Time: 06:02)



Suppose I define say U is equal to X by Y plus 1, and V is equal to say Y plus 1. I want the joint distribution of U and V here, where X and Y follow independent binomials. So, here you look at the set up values, we will follow since Y is binomial n p, Y takes value 0 1 2 n. So, we will take values 1, 2 up to n plus 1, where as a values of u will be now here X can take value 0, X can take value 1 in that case Y plus 1 can take values all these. So, 1, 1 by 2, 1 by 3 and so on 1 by n plus 1; X can take value say 2.

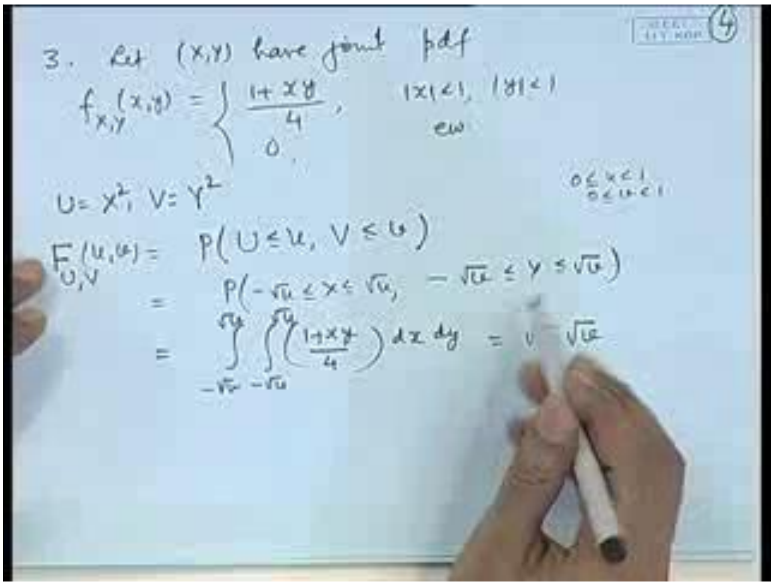
So, these values can be 2, 2 by 2, 2 by 3 and so on up to 2 by n plus 1 and so on n, n by 2 and so on n by n plus 1. So, these are the possible values taken by U. So, we look at probability of say U is equal to small u, V is equal to small v where small u and small v take these values, then this can be expressed as probability X is equal to u v, and Y is equal to v minus 1. So, X and Y are independently distributed so this becomes product of that is equal to n c u v, p to the power u v, 1 minus p to the power n minus u v, then n c v minus 1 p to the power v minus 1, 1 minus p to the power n minus v plus 1. So, this is the joint distribution of u and v, where u and v take these values.

Let us take another example here say X and Y have the joint mass function, the probabilities are 1 by 6, 1 by 12, 1 by 6, 1 by 6, 1 by 12,1 by 6, 1 by 12, 0 and 1 by 12. So, X takes values minus 1 0 and 1, and Y takes values minus 2 1 and 2. Suppose I define U is equal to modules of X and V as Y square, then the possible values of U are 0 and 1 and possible values of V are 1 and 4. So, the joint distribution that is probability say U is equal to 0, V is equal to 1, that is simply probability of X, Y equal to 0, 1 that is 1 by 12. If we look at what is the probability of U is equal to 0, V is equal to 4. It is a sum of X is equal to 0; Y is equal to minus 2, plus probability X equal to 0, Y is equal to plus 2. So, if we add these probabilities we get 1 by 12.

In a similar way we can obtain probability of U is equal to 1, V is equal to 1, U is equal to 1, V is equal to 4 and the joint distribution turns out to be we can explicit as U V you can take values 0 and 1, we can take value 1 and 4 . So, the distribution is the 0, 1 is 1 by 12; 0, 4 is 1 by 12, 1 1 is 1 by 3 and this is half and from here we can derive the marginal distribution of U and V.

So, in the case of discrete distributions etcetera it is possible to derive the distribution of the function of random variables by directly considering the probability mass function; sometimes it is easy to use the direct cumulative distribution function also I can give 1 example here.

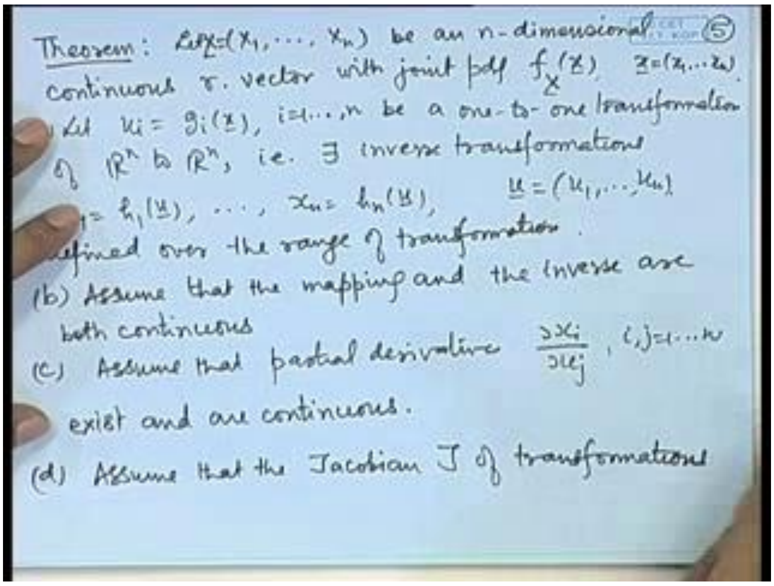
(Refer Slide Time: 10:42)



Let us consider say let X and Y have joint probability density function say f X ,Y given by 1 plus x, y by 4; where modules x is less than 1 and modules y is less than 1; 0 elsewhere. So, we want say the distribution of U is equal to X square and V is equal to Y square, let us consider say CDF of U and V that is probability of U less than or equal to small u, V less than or equal to small v. Now notice here that both x and y lay between minus 1 to 1. So, here the valid region for U and V will be between 0 and 1. So, we consider that, 0 less than u less than 1 and 0 less than v less than 1. So, for this case this is nothing, but probability of X lying between minus root u 2, plus root u and Y lying between minus root v to plus root v. So, this is nothing, but the integration of the joint density over this region. So, that is integral 1 plus x, y by 4, d x d y over minus root u to plus root u minus root v to plus root v and we can evaluate it to be root u root v. So, the joint CDF can be obtained, from here we can determine the density of u and v.

In general cases when we have continuous random variable and we make a transformation of that, it may not be so easy to look at the joint CDF etcetera. In that case like in the case of univariate random variables, we have an approach the so called Jacobian approach for determining the distributions of random variables. So, we stated it in the form of the following theorem.

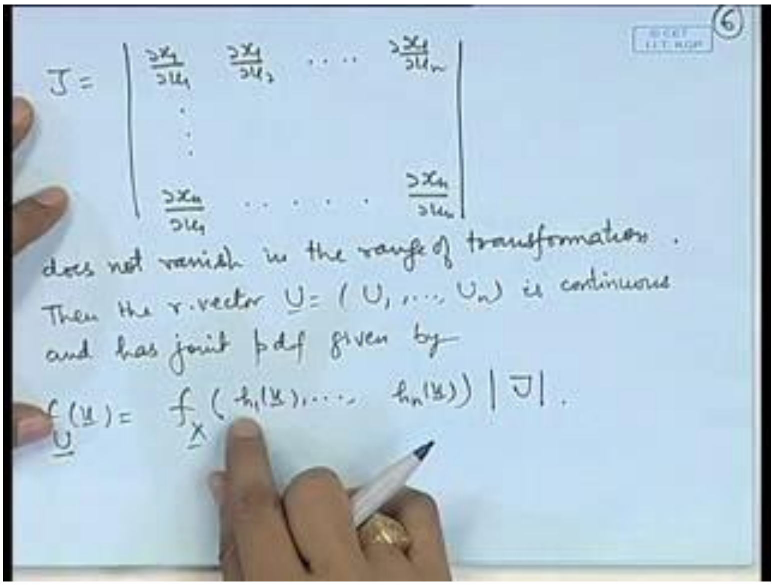
(Refer Slide Time: 13:17)



Let X 1, X 2, X n be an n dimensional continuous random vector with joint probability density function say f X x. So, here X is denoting the vector X 1, X 2, X n small x is denoted the vector x 1, x 2, x n; let u i is equal to g i of x, i is equal to 1 to n be a one-to-one transformation of R n to R n; that is if I am taking one-to-one then there exist inverse transformations it is call it is a x 1 is equal to say h 1 of u and so on, x n is equal to h n of u; where u is u 1, u 2, u n define over the range of transformation.

Let us assume that the mapping and the inverse are both continuous. Further assume that the partial derivatives del x i over del u j for i j is equal to 1 to n, that is all partial derivatives del x i by del u 1, del x 1 by del u 2, del x n by del u 3 and so on all the partial derivatives exist and are continuous, then we define assume that the Jacobian J of transformations.

(Refer Slide Time: 16:26)

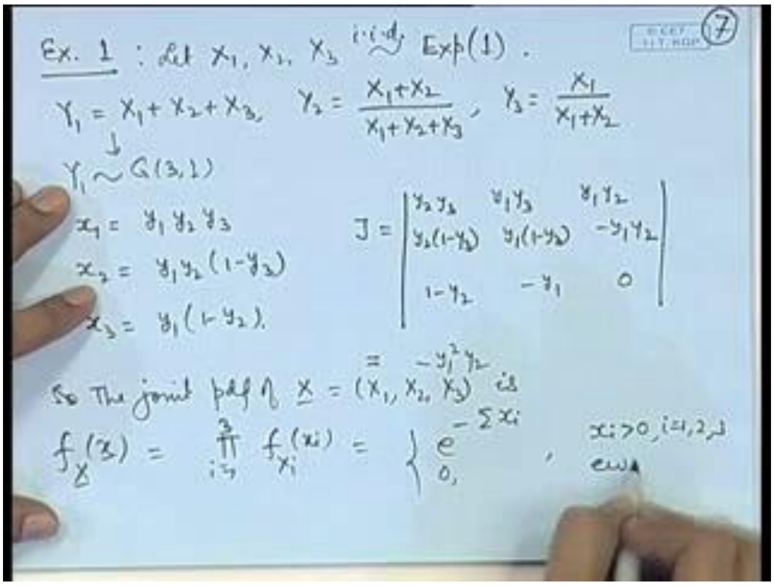


Which is defined by J is equal to del x 1 by del u 1; del x 1 by del u 2 and so on, del x 1 over del u n and so on, del x n over del u 1 and so on, del x n over del u n. Assume that this Jacobian does not vanish in the range of transformation, then the random vector U is equal to U 1, U 2 U n is continuous and has joint pdf given by. So, you write it has f U is equal to f X now in place of X 1, X 2, X n replace it by h 1 u, h 2 u, h n u multiplied by the absolute value of the Jacobian over the range of the transformation. If you see it carefully it is a state forward generalisation of the result for one dimensional case.

In the one dimensional case we had consider a one-to-one transformation and we had looked at the d x by d y term. So, the density of the transform variable was obtained as the density evaluated at x equal to g inverse y, multiplied by the absolute value of d x by d y term. So, when we have a n dimensional random vector and n dimensional transformation, so if it is a one-to-one case, we look at exactly the inverse function and calculate the determinant of the partial derivatives called Jacobian, substitute the values of X 1, X 2, X n in terms of U i's and multiply by the Jacobian term absolute value of the Jacobian, that yields the joint density function of the transform random vector.

So, let us look at a few applications here, let X 1, X 2, X 3 follow exponential with lambda is equal to 1.

(Refer Slide Time: 19:07)

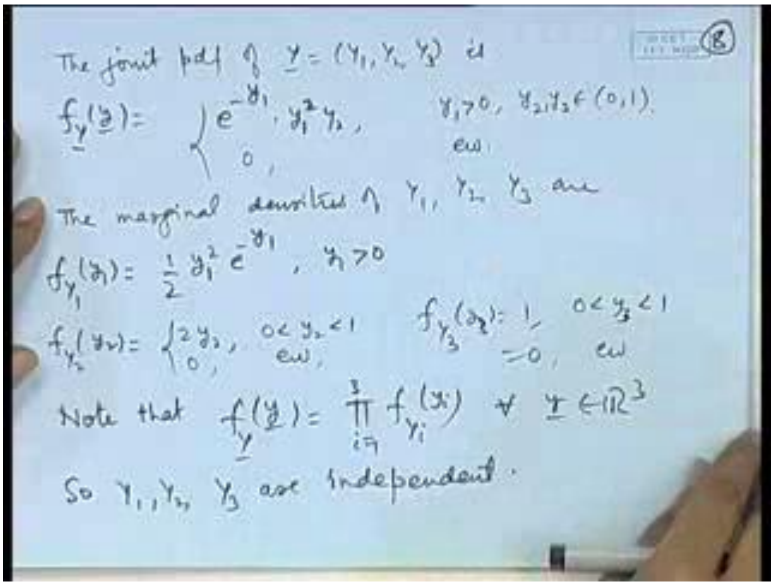


Suppose they are independent and identically distributed random variables, let me define Y 1 is equal to say X 1, plus X 2, plus X 3; Y 2 is equal to say X 1, plus X 2 divided by X 1, plus X 2, plus X 3; and Y 3 is equal to say X 1 by X 1 plus X 2. We are interested in the joint and marginal distributions of Y 1, Y 2 and Y 3 of course, here if we are interested only in a distribution of Y 1, then that is directly obtain because of sums of independent exponential is a gamma. So, Y 1 will follow a gamma distribution with parameter 3 and 1. So, that is directly known; however, that does not yield the distribution of Y 2 or Y 3.

So, we observe here that it is a one-to-one transformation and inverse functions can be written as x 1 is equal to y 1, y 2, y 3; x 2 can be written as then y 1, y 2 into 1 minus y 3; and x 3 can be written as y 1 into 1 minus y 2. So, we can determine the Jacobian of the transformation dou x 1 by dou y 1 is y 2, y 3; dou x 1 by dou y 2 is y 1 y 3 and so on y 2 into 1 minus y 3, y 1 into minus y 3, minus y 1 y 2, 1 minus y 2, minus y 1 and 0. So, if you evaluate this it turns out to be minus y 1 square y 2. So the firstly, we write down the joint density function of X 1, X 2, X 3. So the joint pdf of X 1, X 2, X 3; so since X 1, X 2, X 3 are independently distributed, the joint density is nothing, but the product of the individual density functions of X 1, X 2, X 3 that is product of f X i that is equal to e to the power minus sigma x i each x i is positive.

Therefore, the joint density of Y 1, Y 2, Y 3 can be obtain from here by substituting the inverse function of X 1, X 2, X 3 and the corresponding range and multiply by the Jacobian.

(Refer Slide Time: 22:32)

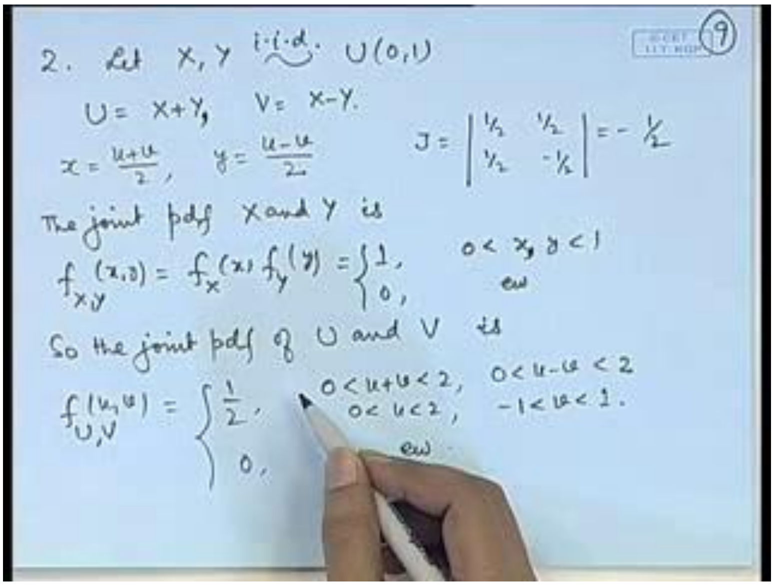


So, the joint pdf of Y is equal to Y 1, Y 2, Y 3 is f, e to the power minus y 1 into y 1 square y 2. The range of the variables we can observe here that each of the X I is a positive random variable, so each of Y I is also a positive random variable; further if x 2 is positive then y 3 will be less than 1 and similarly y 2 will also be less than 1. So, the ranges are y 1 greater than 0, y 2 and y 3 they belong to the interval 0 to 1. So, we have been able to determine the joint distribution of y1, y 2, y 3; in order to get the marginal distributions we notice here that if we integrate with respect to y 3 from 0 to 1, we get the same term and therefore, if we integrate this will respect to y 1 y 2, it should give the density of y 3 as 1 on the interval 0 to 1.

So, the marginal distributions the marginal densities of Y 1, Y 2 and Y 3 are obtained as f Y 1 as half Y 1 square, e to the power minus Y 1, which is nothing, but it gamma distribution with parameters 3 and 1, f Y 2 is 2 Y 2, for Y 2 between 0 and 1 and f Y 3 is equal to 1 between 0 and 1. So, this is a uniform distribution, one interesting feature we can notice here that if I look at the product of the marginals it is equal to the joint, note that f of Y is equal to the product of. So, Y 1, Y 2, Y 3 are independent.

So, here we are able to obtain the distribution of a three dimensional function of a of three random variables here, the important thing to notice here is that apart from substitution in the density function and multiplying by the Jacobian, we are also judiciously determine the ranges of the variable; like one may simply say that Y 1 is positive, Y 2 is positive, Y 3 is positive without noticing that Y 2 and Y 3 are less than 1 also, in that case if we will evaluate the integrals of this density it will not give us one. So, that will be not determining the density correctly.

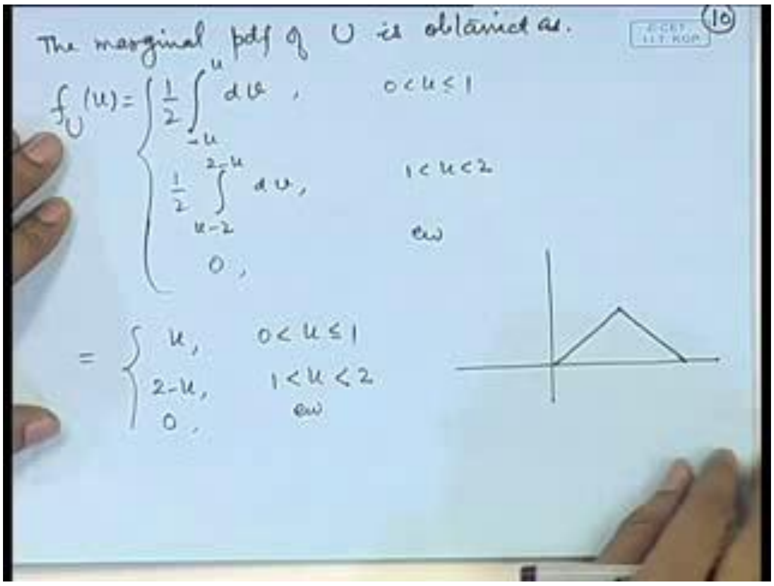
(Refer Slide Time: 26:11)



Let us take uniform distributions let X and Y be independent and identically distributed uniform random variables.

Let us define say U is equal to X plus Y, and V is equal to say X minus Y. Now clearly this is a one-to-one transformation x is equal to u plus v by 2, and y is equal to u minus v by 2. So, if we look at the Jacobian term dou x y by dou u is half, half, half and minus half, which is equal to minus half. So, the joint pdf of say X and Y that is f X Y, it is the product of the individual distributions of X and Y both are uniform 0 1. So, it is simply 1, for 0 less than x y less than 1 and 0 elsewhere. So, the joint pdf of U and V is it will become half, for 0 less than u plus v less than 2, 0 less than u minus v less than 2. Now the ranges of u and v we can notice further here that since X and Y are between 0 to 1, U will be between 0 and 2 and V will be between minus 1 and 1 this gives the joint density function of U and V.

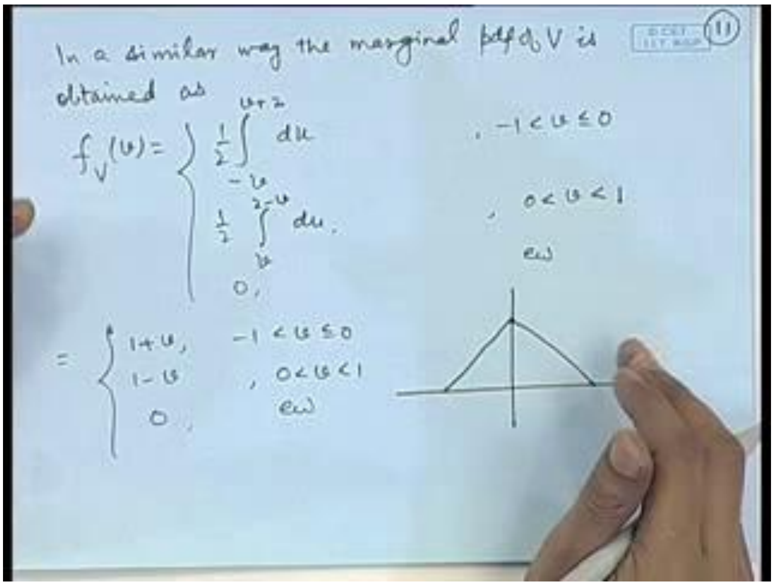
(Refer Slide Time: 28:32)



Suppose we are interested in the marginal distributions of U and V. So, in order to get the marginal distribution of U we need to integrate this will respect to the variable V. So, the marginal density of U is obtained as. So, f U integral of this joint density that is half d v; now notice here the range of v, v is absolute range is from minus 1 to 1, but here v lies between minus u to 2 minus u and v is less than u and v is also greater than u minus 2. So, if we determine the region, it is from minus u to u if u is between 0 to 1, it will be half, u minus 2 to 2 minus u, d v, if v is between 1 and 2 and 0 elsewhere.

So, after simplification this turns out to be u for 0 less than u less than are equal to 1, it is 2 minus u for 1 less than u less than 2 and 0 elsewhere. Notice here this is a triangular distribution 0 to 1 and 1 to 2. So, the distribution of the sums of 2 independent uniform random variables is actually a triangular distribution, in a similar way we can obtain the marginal density of v also if we integrate to respect to u.

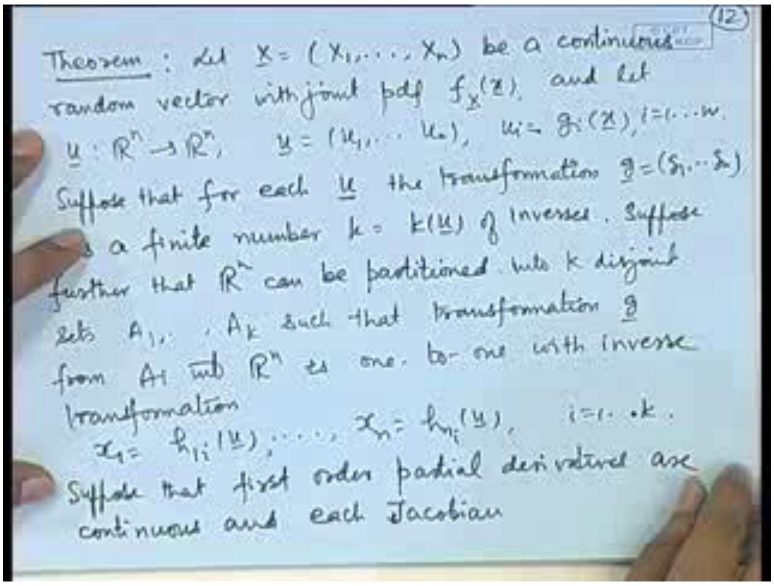
(Refer Slide Time: 30:37)



In a similar way the marginal pdf of V is obtained as f V v, it is integral of half with respect to u from minus v to v plus 2 for minus 1 less than v less than are equal to 0, it is half from v to 2 minus v, d u for 0 less than v less than 1 and 0 elsewhere. So, after simplifications this turns out to be 1 plus v, for minus 1 less than v less than are equal to 0 and 1 minus v for 0 less than v, less than 1 0 elsewhere. This is again a triangular distribution on the interval minus 1 to 1. So, minus 1 to 0 the density is 1 plus v and between 0 to 1 the density is 1 minus v. So, we notice here that the sums and differences of independent uniform random variables are again are triangular distributions and obviously, they are not independently distributed here, because the joint distribution of u and v is not equal to the product of marginal distributions of u and v here.

Now, in many cases the function from R n to R n need not be one-to-one for example, we have consider the discrete case where u was modular x and v was y square. So, it is not a one-to-one transformation, the other it is a 4 to 1 transformation over the range of them variables. So, in that case we have a result similar to the case of univariate, in the case of univariate when we had a many one transformations we split the domain into disjoint regions, such that from each region to the range we have a one-to-one transformation, we consider the inverse transformation using that we calculate the density function in each region of the domain, disjoint regions and we add all of this that gives the joint distribution. So, generalization of this result is available for the n dimensional case also and we stated in the form of the following theorem.

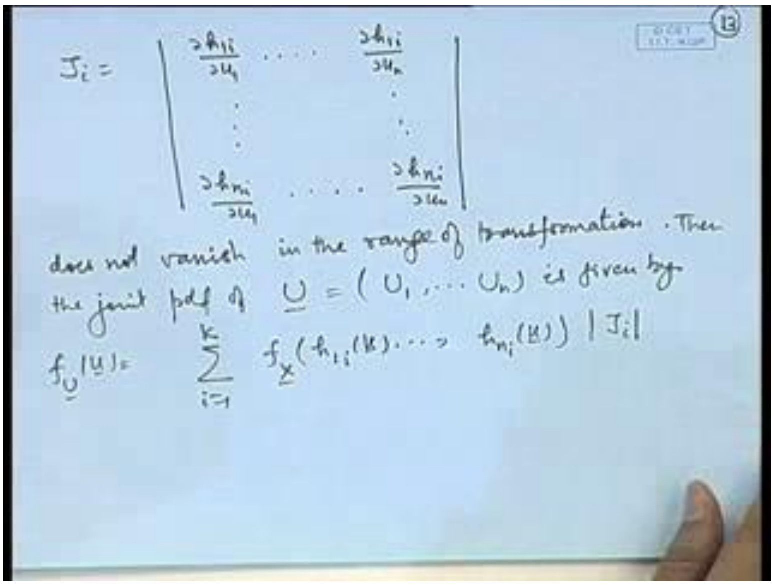
(Refer Slide Time: 33:37)



Let X is equal to X 1, X 2, X n be a continuous random vector with joint pdf f of X and let u be a mapping from R n into R n, where u is equal to u 1, u 2, u n u i is equal to sum g i of X for i is equal to 1 to n. So, we are not assuming that it is a 1 1 on to function, suppose that for each u the transformation g that is g 1, g 2, g n has a finite number say k of inverses. Suppose further that R n can be partitioned into K disjoint sets say A 1, A 2, A k such that transformation g from A i into R n is one-to-one with inverse transformation say x 1 is equal to h 1 i u and so on, x n is equal to h n i, u for i is equal to 1 to k. As in the case of previous theorem we have to assume that the mapping and in the inverses are continuous and this first partial derivatives are continuous.

Suppose that first order partial derivatives are continuous and each Jacobian that is J i.

(Refer Slide Time: 37:02)



That is del h 1 i by del u 1 and so on, del h 1 i over del u n and so on, del h n i over del u 1 and so on, del h n i over del u n does not vanish in the range of transformations then the joint pdf of U is equal to U 1, U 2 U n is given by f U u is equal to sigma f of X h 1 i u and so on h n i u multiplied by absolute value of the Jacobian i is equal to 1 to K.

So, note here if we consider this term it is the density determined by the one-to-one transformations from A i into R n. So, we calculate this density for each region A 1, A 2, A k and add this give a joint distributions U 1, U 2, U n we consider an example for example, we can considered distribution of orderly statistics.

Thank you.