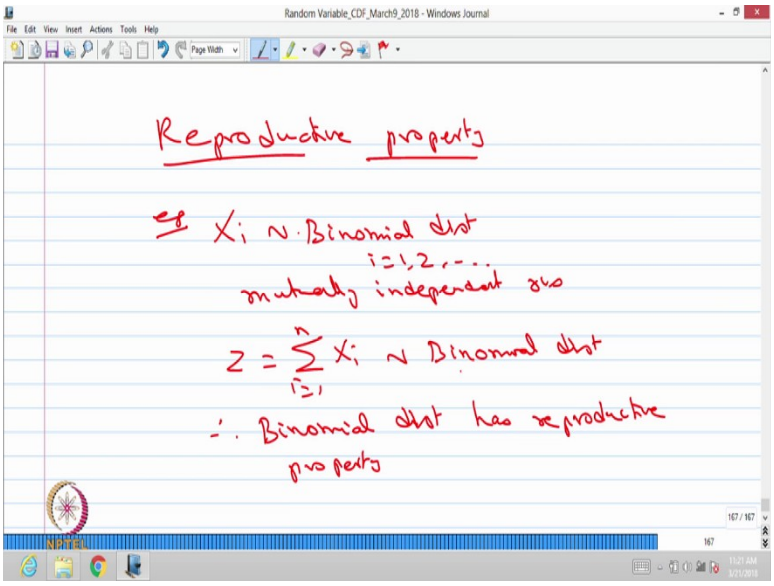
(Refer Slide Time: 04:03)



What the reproductive property says that if you have sequence of random variables and if you make a sum of those few of the variables out of it and all are having some distribution and after making the summation you are getting the same distribution of same as X_i's, or the original sequence of random variable then we conclude this random variable has; this particular random variables has a reproductive property.

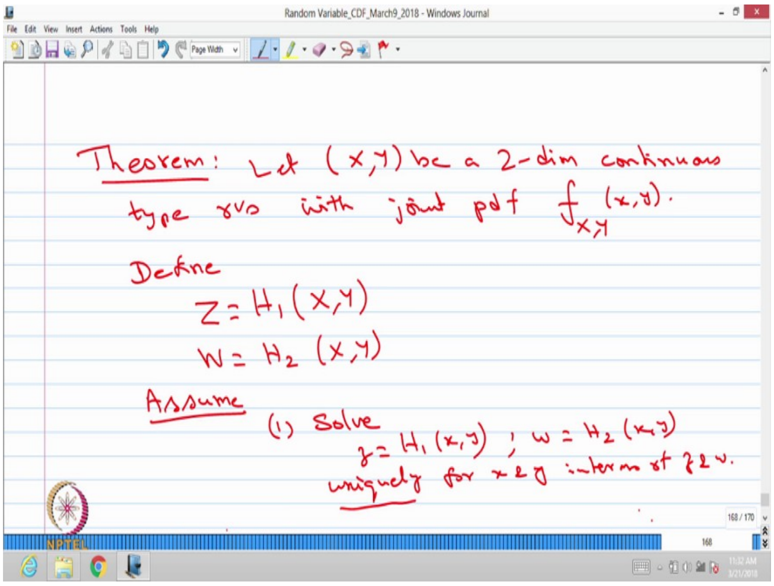
That means for example, each X_i's are a binomial distributed and I have many random variables. All are mutually independent. I make the assumption all the random variables are mutually independent. Then if I make a random variable as the sum of few random variables out of this collection if that is also follows a binomial distribution. So, we can conclude binomial distribution has reproductive property.

Similarly, one can say the Poisson distribution also has a reproductive property whereas the Bernoulli distribution does not have a reproductive property. Because if you have a Bernoulli distributed random variable all are mutually independent, if you make n such random variable as a summation then that is going to be a binomial distribution, no more Bernoulli distribution. Therefore, Bernoulli distribution does not have a reproductive property.

Similarly, one can go for some common continuous type random variables. If you have a normal distributions all are mutually independent, if you make a summation then that is also going to satisfies; the reproductive property. That means, summation is also going to be a normal distribution. So, like that we can make a list of standard or common distributions satisfying the reproductive property and not satisfying the reproductive property.

Now, we will move into distributions of functions of several variable when each random variable is of the continuous type. So, for that I am going to give one important result as a theorem. After I introduce a theorem then I will go for giving some examples we are not going to prove the theorem. So, I am going to give the important result or the theorem as some sort of results, then I am going to give some examples.

(Refer Slide Time: 07:36)



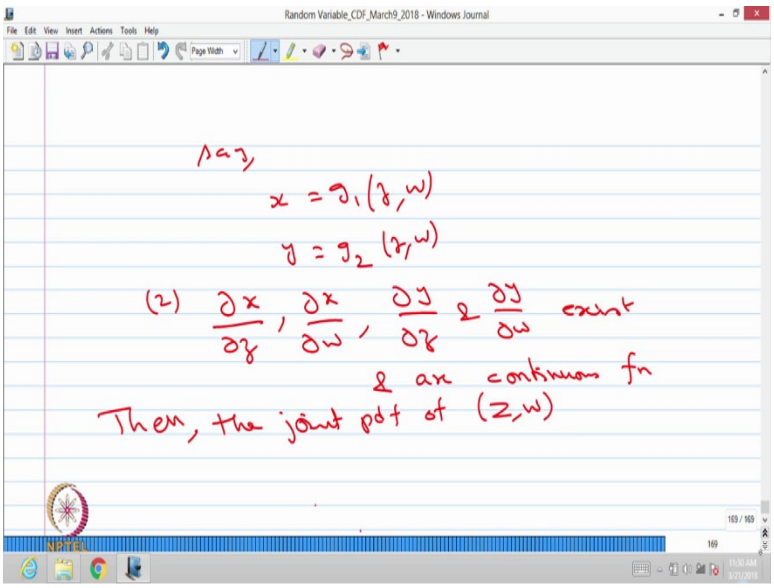
So, let me make it as the theorem we are not going to give the proof of this theorem. Let me start with this theorem for only two dimension random variable then the same concept can be extended for n dimensional. So, let me start with the two dimensional. That is let X, Y be a two dimensional continuous type random variables with the joint probability density function that is $f_{X,Y}(x,y)$.

I am going to define new set of random variables that is the first random variables Z = $H_1(X,Y)$. The another random variable W ¿H_2¿X, Y). We can assume that both $H_1 \wedge H_2$ are a Boral measurable functions, so that Z and W are going to be random variables.

I am going to make a few assumptions so that I can able to get the joint probability density function of Z and W directly with the help of the joint probability density function of X, Y. I am going to make a few assumptions if those assumptions are satisfied, then that makes Z is a continuous type random variable as well as W is a continuous type random variables not only that I can find the joint probability density function of Z and W with the help of the joint probability density function of X, Y, so that is what I am going to give it as theorem.

In the first assumption I can solve Z as a function of X, Y and the W as a function of X, Y. This equation can be solved uniquely for x and y in terms of z and w. I can solve the same thing I am going to I am replacing capital Z by z capital X and Y by x and y. Therefore, whatever I made the transformation of the random variable Z = $H_1(X,Y)$, W ¿ H_2¿X, Y). I am going to solve those with smaller letters because I am consistently using the capital letter for the random variable. So, I am solving this equation uniquely for x and y in terms of z and w.

(Refer Slide Time: 11:36)

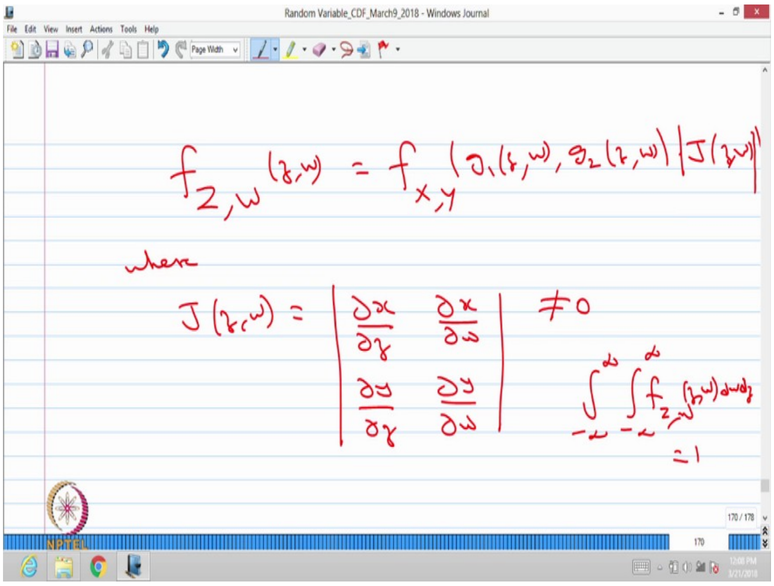


So, whatever I am getting the solution that I am going to write it as say x is going to be; say the answer which I am going to get x in terms of z and w that I am going to write it as the sum function of z, w g_1. Similarly, I am going to write y as some function of z comma w. So, this is after solving those two equations.

The second assumption the x in terms of z and w, and y in terms of z and w, I can go for finding out the partial derivative with respect to z, w for x and y. I can find the partial derivative of x with respect to z and w.

Similarly, the partial derivative of y with respect to z and w. Here I am making the assumptions partial derivative exist, not only exist all this partial derivative are continuous functions. Not only the partial derivative exist, it as to be continuous functions also this is the second assumption.

(Refer Slide Time: 13:35)



With this assumption I am going for concluding joint probability density function of Z, W can be written as the probability density function of Z, W as a function of z and w in terms of the joint probability density function of X, Y by replacing x by. If you see we made the, we got by after solving x ¿$g_1(z,w)$, y you are getting g_2. Therefore, in the joint probability density function of X, Y. I am going to replace small x by $g_1(z,w)$.

Similarly, y I am going to replace by $g_2(z,w)$, multiplied by the absolute of the determinant that is called Jacobian as a function of z, w where I can define the Jacobian as a function of z, w that is nothing but the determinant of the partial derivative which we have got it, $\frac{\partial x}{\partial z}$, $\frac{\partial x}{\partial w}$, $\frac{\partial y}{\partial z}$, $\frac{\partial y}{\partial w}$.

This determinant is the Jacobian whereas in the probability density function of Z and W you substituted the absolute of this Jacobian. This is going to be the joint probability density function of Z, W; that means, this theorem says whenever you have a continuous type random variable and you know the joint probability density function of the continuous type random variables.

As long as these two assumptions are satisfied. The word uniquely is very important if that is not satisfied then we have another remark over it. So, here if the assumption 1 as well as the assumption 2 are satisfied, then we can directly conclude the Z, W is going to be continuous type variables and one can get the joint probability density function of Z comma W. By substituting an x by g_1 and y by g_2 in the joint probability density function of X, Y with the product of absolute of Jacobian.

The product absolute of Jacobian that is nothing, but the normalizing constant; that means,

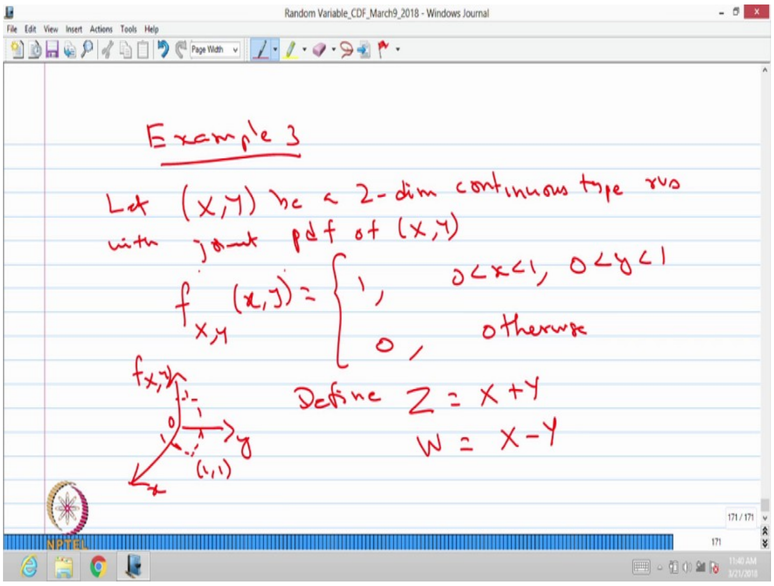
$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Z,W}(z,w)\,dzdw$ has to be 1. So, this is going to be 1 whenever you multiply the absolute of Jacobian. Therefore, the absolute of Jacobian is nothing, but the normalizing constant.

There is another remark, some books use instead of product of Jacobian they use divided by absolute of Jacobian. In that case they make the Jacobian in the determinant form not the $\frac{\partial x}{\partial z}$ and $\frac{\partial x}{\partial w}$ they make the $\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}$. Find the determinant of Jacobian of that inverse then substitute in the formula with the divider in the denominator.

Both the results are one and the same because the result is the Jacobian matrix, Jacobian this determinant are the inverse one if you make a product that is going to be 1. Because you have n dimension random variable again you are transforming another n dimension random variable by satisfying the two conditions, that is you are solving those equation uniquely and the partial directive exists and continuous that makes whether you use the Jacobian or the inverse.

Therefore, the formula changes either in the multiplication in the numerator, or multiplication in the denominator form. Because the Jacobian and the inverse Jacobian that determinant value product is always going to be 1. Now, let us go for one easy example to explain how this theorem works.

(Refer Slide Time: 19:20)



Let X, Y be two dimensional continuous type random variables with joint probability density function of X, Y that is given as f(x, y), that takes a value 1, when x lies between 0 to 1 and y lies between 0 to 1, otherwise it is 0.

So, this is the joint probability density function of X, Y you can verify you can verify by just you know this is X, this is Y this is joint probability density function it takes a value 1 between the interval 0 to 1, and y is also 0 to 1. So, in the x y plane the region is a square with the vertex 0 comma 0, 1 comma 0, 0 comma 1 and 1 comma 1.

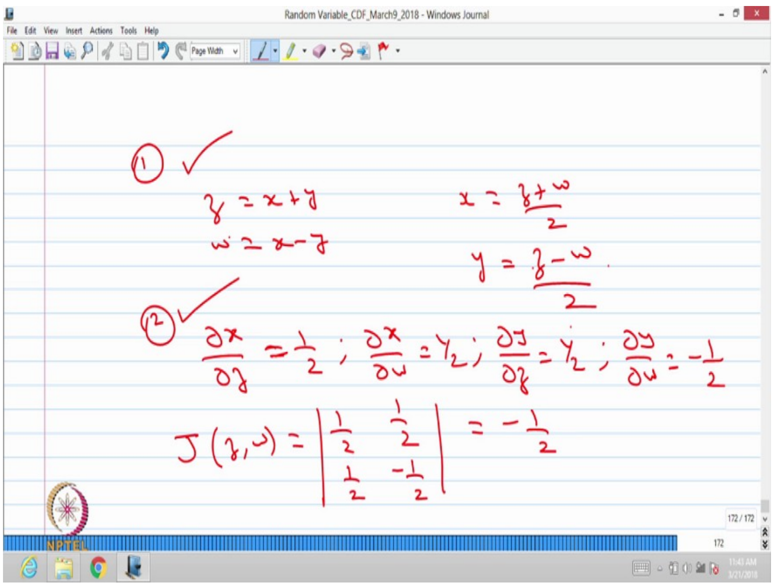
And at the height 1, the surface is at the height 1 over the square in the x y plane. And if you find the volume below the surface that plane 1, above the square shape that is going to be 1, it is a cube. Volume of the cube; therefore it is easy to verify this is a joint probability density function of two dimensional continuous type random variable. The question is we are going to create another two dimensional random variable and then we are finding the distribution of a the new set of random variables that is also two dimensional.

So, I am going to define the new set of random variables you can use the same notation Z= X+ Y that is the function $H_1(X,Y)$. The second function that is capital $H_2(X,Y)$ that is X-Y, both X and Y are continuous random variable the way we have defined Z = X+Y and W = X-Y.

You can immediately say both are going to be again continuous type random variables, therefore either you can find the cdf of Z, W. If the question is find the distribution, if you know that both the random variables are of the continuous type you can find the joint probability density function. So, here we are going for finding the joint probability density function of two dimensional continuous type random variables Z and W.

We can a make sure whether the assumption of the previous theorem satisfied. If it is satisfied, then you can use the theorem and get the result. If it is not satisfied, then you cannot find the joint probability density function using that theorem to apply the theorem you have to make sure that the assumption satisfied. Now, we will go for whether the first assumption is satisfied or not.

(Refer Slide Time: 23:37)

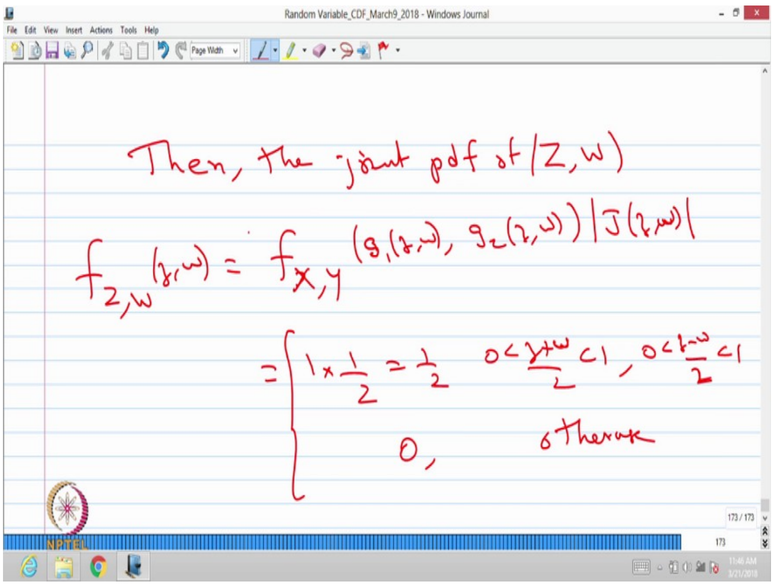


So, you try to find out that is a said Z= X+Y and W = X-Y. You solve for these two equations for x and y in terms of z and w; that means, you can get x ¿$\frac{z+w}{2}$. If you add this two equations y will be cancelled, so 2 x =z+ w. Therefore, x ¿$\frac{z+w}{2}$, y is going to be you subtract, x will be cancelled. So, you will get the 2 y. Therefore $\frac{z-w}{2}$, that is going to be y. So, you are able to solve this equation uniquely and you can get the answer x and y in terms of z and w. So, the first condition is satisfied.

We will go for second condition. Find out the partial derivative whether it exits or not. The partial derivative with respect to z of the function x that is one by two exist which is continuous constant here that is ok. Similarly, you find out the partial derivative with respect to w, partial derivative of y with respect to z, partial derivative of y with respect to w, all exist and are continuous functions also in particular. Here it is constant therefore; the second condition is also satisfied.

Now, we can go for writing the joint probability density function of Z and W with the help of joint probability density function of X and Y. We will find out the determinant of a Jacobian that is Jacobian as a function of z, w that is determinant of; I will substitute all the partial derivatives in the correct order. Whether you write like this or in the transpose ways it does not matter because at the end you are finding the determinant. That is – 1/4-1/4, therefore, it is -1/2 this is a Jacobian.

(Refer Slide Time: 26:30)

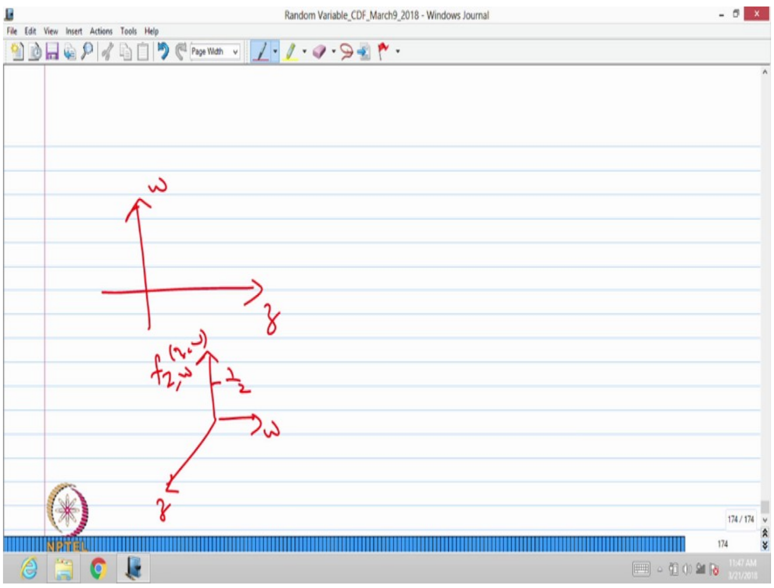


Now we can go for writing since the two assumptions are satisfied you can give the joint probability density function of Z comma W as first write down the $g_1(z,w), g_2(z,w)$ mu ltiplied by the absolute of Jacobian. The Jacobian has to be a non zero, it is also very important condition. Because if it is 0, then the probability density function will become zero. So, as long as the Jacobian is going to be a non-zero quantity, we can go for it.

Now, you substitute in this problem the joint probability density function is function of x and y is 1 between these intervals, otherwise it is 0. So, you can replace x by $\frac{z+w}{2}$ y by $\frac{z-w}{2}$ within that range of z and w lies between 0 to 1, the value is going to be 1. So, this is going to be; since it is a constant you cannot substitute the x by $g_1 \wedge g_2$. Therefore, this is going to be again 1, and the Jacobian quantity is - 1/2 and we have to substitute which is absolute quantity. Therefore, it is ½ multiplication provided this joint probability density function provided x lies between 0 to 1. So, here it is $0 < \frac{z+w}{2} < 1$.

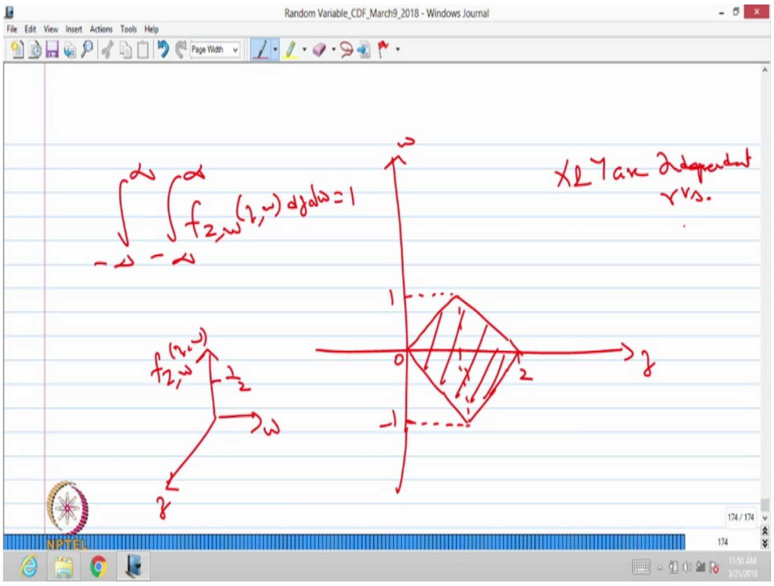
Similarly, y is lies between 0 to 1 that is $0 < \frac{z-w}{2} < 1$. So, as long as z and w satisfies these two conditions in which the joint probability density function is 1/2, otherwise it is 0. So, the joint probability density function is 1/2 when z and w satisfies; $0 < \frac{z+w}{2} < 1$, $0 < \frac{z-w}{2} < 1$. That means, now you can think of a how the joint probability density function of Z and W look like.

(Refer Slide Time: 29:33)



Before that we can go for what is the region of Z and W in which the joint probability density function is greater than 0 that is 1/2. First, we will identify basically what we want is z, w the joint probability density function of Z and W. For that first we are making a what is the region in which the joint probability density function is going to be the value is 1/2. So, the region is if you simplify these two in equalities you can identify the region of z and w, z and w 0, 1, 2 and -1.

(Refer Slide Time: 30:18)



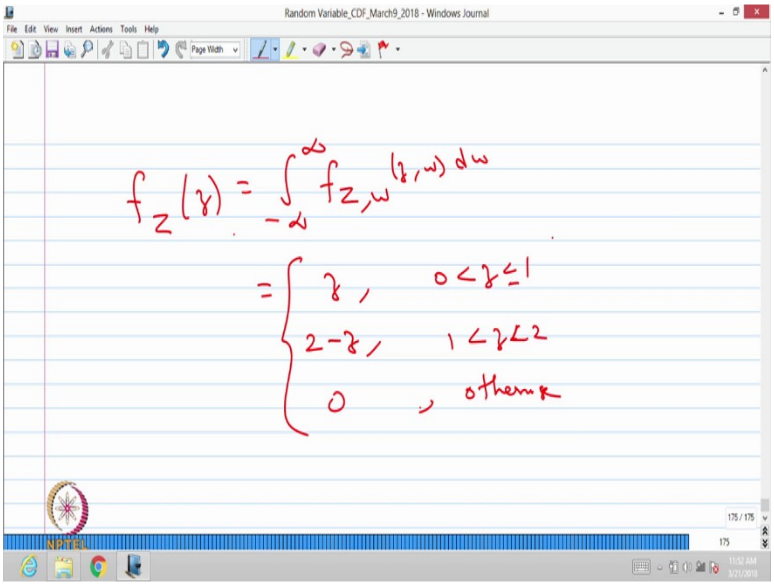
So, if you simplify those two inequalities; you can identify the region is going to be; I am not drawing the diagram in correct scaling way. Just for the illustration purpose. So, this is the shaded region, is the region of z and w; that means, z and w plane this is the region in which the joint probability density function is 1/2, otherwise it is 0.

Now we can verify the $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Z,W}(z,w)\,dz dw$ is going to be 1 because the x y plane is this diagram above that it is 1/2. So, the volume below that surface is 1/2 constant, over the region in which this diagram shaded region is there the volume is going to be 1.

So, this type of graphical representation is possible only for two dimensional random variable, not for any n dimensional random variable 3, 4 and so on it is very difficult to visualize. So, this is easy to visualize one more observation in this problem given X and Y you can see it. The joint probability density function is one if you find out the probability density function of X that is going to be one between the interval 0 to 1 for x.

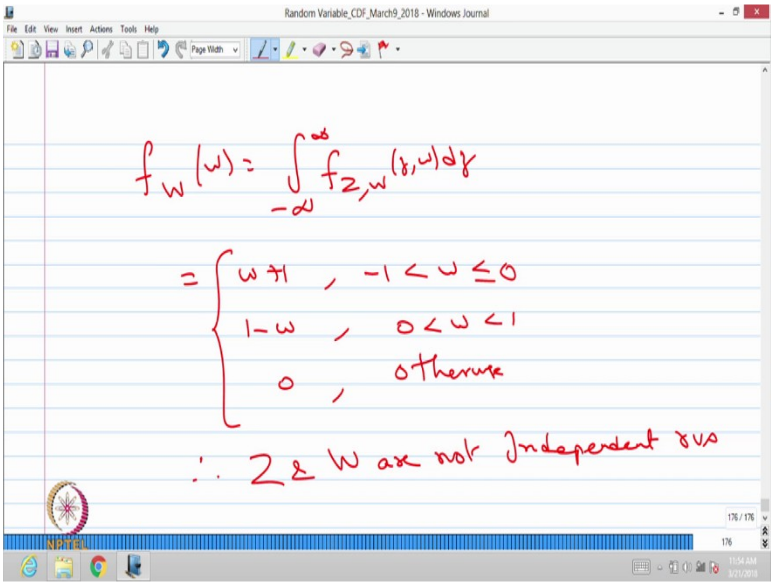
Similarly, if you find the marginal distribution of Y that is probability density function of Y that is also one between the interval 0 to 1 of y, otherwise 0. If you multiply the probability density function of X and Y that is same as joint, so we can conclude X and Y are independent random variables, where as the joint probability density function of Z and W is 1/2 between this interval.

(Refer Slide Time: 33:24)



If you find out the marginal distribution of Z if you do the simple exercise finding the probability density function of Z by $\int_{-\infty}^{\infty} f_{Z,W}(z,w)\,dw$. One can get I am not doing the derivation by substituting the joint probability density function substitute the correct interval then integrate one can get the answer that is a z when z lies between 0 to 1 that is 2 - z when z is lies between 1 to 2, otherwise it is 0. So, I can make less than or equal to here. So, this is going to be probability density function of Z between the interval 0 to 1 that is z, and 1 to 2 it is 2 - z, otherwise it is 0.

(Refer Slide Time: 34:41)



Similarly, one can compute the probability density function of W from $\int_{-\infty}^{\infty} f_{Z,W}(z,w)\,dz$. If you do that you will get the probability density function of W that is w + 1, when w in the range from -1 to 0 and it is 1 - w between 0 to 1, otherwise it is going to be 0.

The interval of z and w that you can get it you can feel it from the diagram itself, the range of z is 0 to 2 whereas, the range of w is -1 to 1. So, therefore, we are getting the probability density function like this for Z and probability density function of W in this form. The way the probability density function of Z and W is like this if you make a multiplication you would not get the value 1/2 that is joint probability density function of Z and W.

Therefore, you can immediately conclude Z and W are not independent random variables X and Y are independent random variable the way we defined Z = X+Y, W = X-Y they are not independent random variables. So, with this example we are explaining how the theorem works.

But sometimes the assumption first assumption that is solve uniquely it may not satisfy; that means, you may have a more than one set of values instead of z and w in terms of x and y uniquely in that case for every set of pairs you have to identify what is the density function with the corresponding Jacobian. And you have to keep adding how many pairs of solution you are going to get those many summations you have to make it to get the joint probability density function of Z and W.

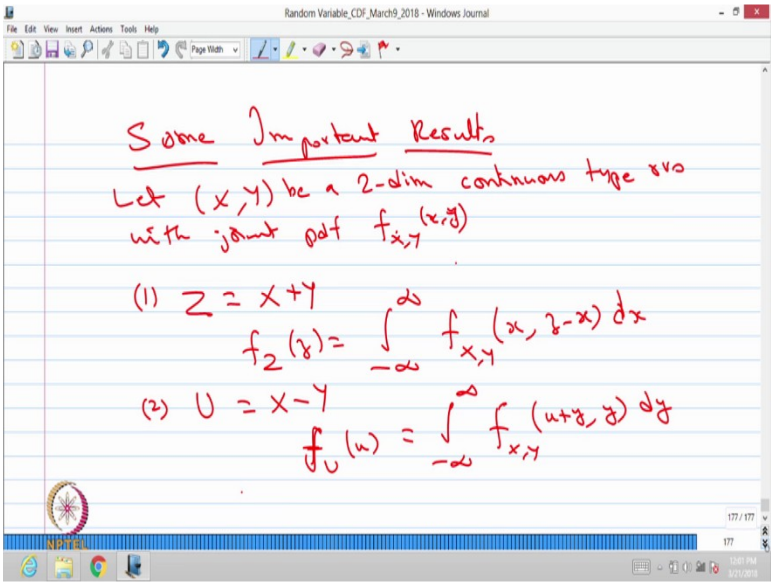
It is similar to what we have done it with the function of a random variable for a continuous type whenever it is not satisfying, whenever the function is monotonically increasing, or decreasing or decreasing or increasing form the same technique he has to be applied for the multidimensional random variable.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 31

Now, we are going to give few important results. Some important results on distributions of functions of several variables.

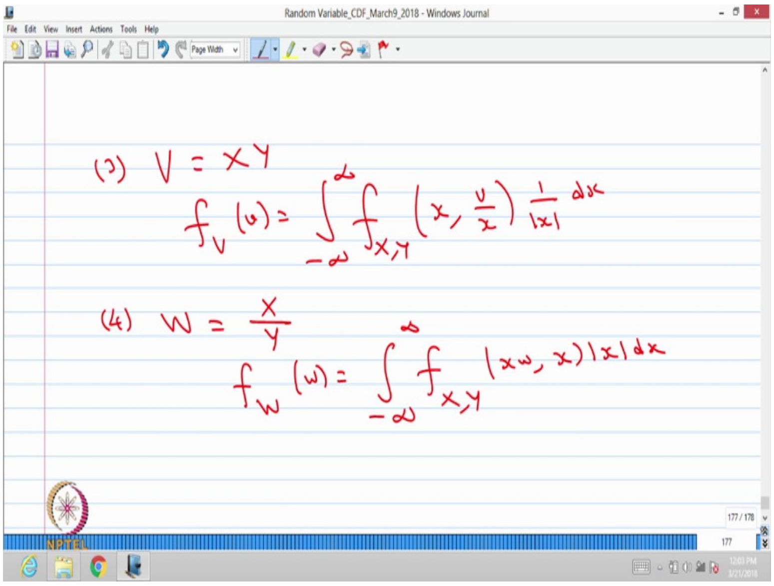
(Refer Slide Time: 00:06)



The first one, let me start with a continuous type random variable let (X, Y) be two dimension continuous type random variables with joint probability density function is given or is known. The first result, suppose I make a random variables Z = X+Y we can find the distribution of Z directly that is the probability density function of Z is going to be $\int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\,dx$ by replacing x by x whereas, y by z - x, so this is going to be the probability density function of Z when Z = X+Y and X and Y are two dimensional continuous type random variable with a joint probability density function f.

The second result suppose I have another random variable U = X - Y. Then the probability density function of U as a function of u that is going to be $\int_{-\infty}^{\infty} f_{X,Y}(u+y, y)\,dy$ by replacing x by u + y and keep the other variable y. This is for the difference of two random variables by knowing the joint probability density function of X and Y you can directly get the probability density function of U.

(Refer Slide Time: 03:01)



Similarly, the third result that is the V is a new random variable that is product of two random variables. You can get the probability density function of V by $\int_{-\infty}^{\infty} f_{X,Y}\left(x, \frac{v}{x}\right)\frac{1}{|x|}dx$, by replacing y by v divided by x and multiplied by 1 divided by absolute of x, this will give the probability density function of V.

Similarly, in the last result the random variable W $= \frac{X}{Y}$ then you can get the probability density function of W by the similar manner, $\int_{-\infty}^{\infty} f_{X,Y}(xw, x)|x|dx$. You see in all these four examples we started with the two dimensional continuous type random variable and we define one random variable Z or U or V or W.

You can get the probability density function of Z or U or V or W by using the previous theorem which I have explained you. The only difference is to apply the theorem you need n dimensional random variable of transformation into another dimension random variable. I have explained the theorem with only two dimensional random variable, the same concept can be extended to n dimension random variable.

So, here we started with the two dimensional random variable, we are finding the probability density function of a random variable Z. You can apply the previous theorem by introducing one dummy random variable dummy means you can create any random variable along with Z = X+Y such a way that a both the assumptions are satisfied and the Jacobian is a non zero. Then find the joint probability density function of Z with that dummy random variable.

After that from the joint probability density function you can always able to get the probability density function of any one random variable using that concept you can get the probability density function of Z. That means, we are applying the theorem by suitable inclusion of one dummy random variable getting the joint distribution then getting the original distribution from the joint distribution. So, that is what this result says the probability density function of Z is $\int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\,dx$ by replacing x by x and replacing y by z - x.

One more remark over this results I have not made the assumptions of this random variables are independent or not. If these random variables are independent random variables then you can replace the right hand side with joint probability density function by product of probability density functions.

Even though I have given results for in general if the random variables are mutually independent random variables you can replace the joint probability density function by product of probability density functions. So, these four results you can remember and you can use it whenever you want to go for finding the sum, or difference, or multiplication, or division of two random variables.

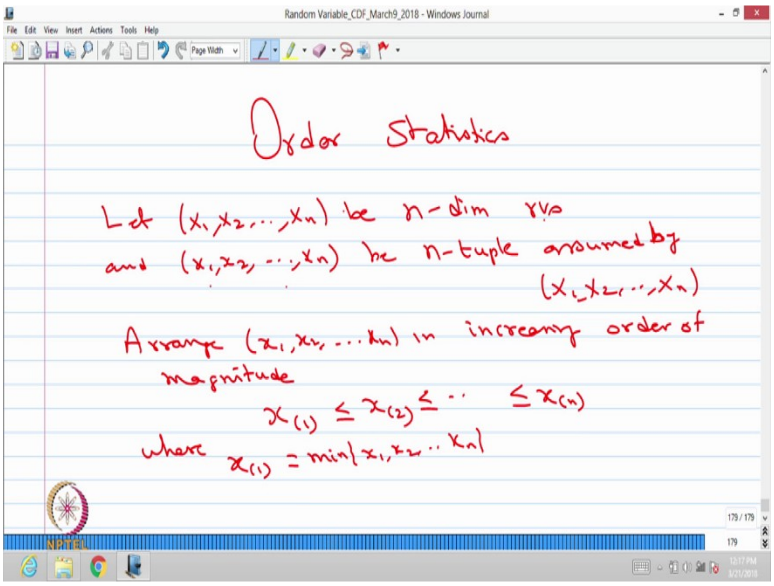
Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 32

In the last class, we have discussed distributions of functions of several variables. In that we have started with the discrete type random variables. If you have n dimensional discrete type random variables and one can find the distributions of the new set of n dimensional random variable of the discrete type.

Then later, we have discussed when you have a continuous type of random variable of n dimensional and you have another set of new continuous type n dimensional random variable using nice theorem, one can get the joint probability density function of new set of n dimensional random variable. By applying the theorem, one can get the joint distribution, then you can get the margin.

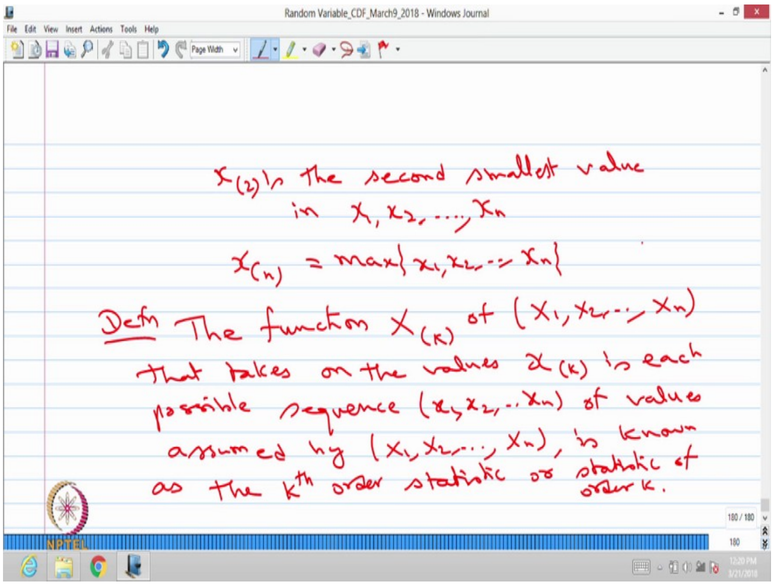
(Refer Slide Time: 01:13)



Now, in this lecture, we are going to discuss one particular type of functions of several random variables that is called Order Statistics. Why the name is called the order statistics? Why the order statistics has to be in this lecture as the functions of several random variables? You will understand. So, for that I am going to give first, what is the meaning of order statistics from the scratch.

Let $(X_1, X_2, \ldots, X_n)$ be n dimensional random variables and $(x¿¿1, x_2, \ldots, x_n)¿$ be n tuple assumed by the random variables $(X_1, X_2, \ldots, X_n)$. That means, the possible variables of a the n dimensional random variables are small $(x¿¿1, x_2, \ldots, x_n)¿$. Arrange these possible values in increasing order of magnitude. So, that I can make out that is, make it as a $(x¿¿(1), x_{(2)}, \ldots, x_{(n)})¿$ where the $x_{(1)}$is nothing, but a minimum$\{x_1, x_2, \ldots, x_n\}$.

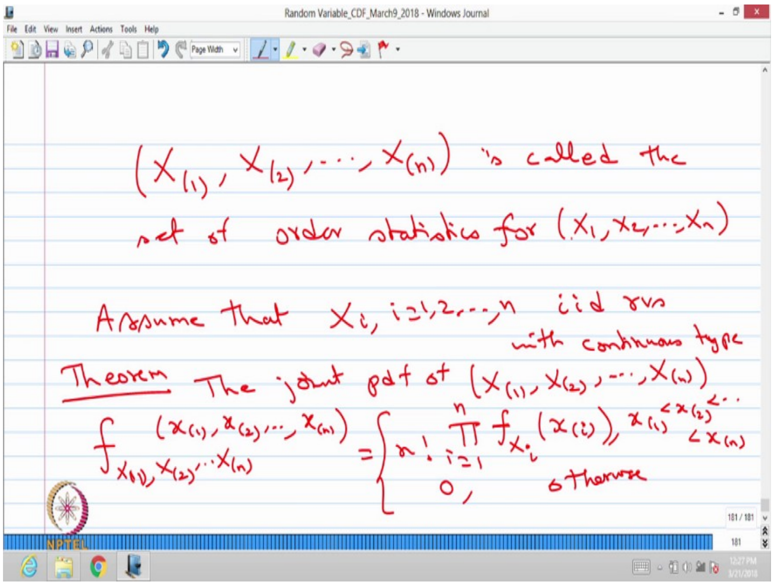
(Refer Slide Time: 03:44)



Similarly, $x_{(2)}$ is the second smallest value in $\{x_1, x_2, \ldots, x_n\}$. Similarly, $x_{(n)}$ that is nothing, but a maximum$\{x_1, x_2, \ldots, x_n\}$. That's a difference between x_i's and $x_{(i)}$'s fine.

Now I am going to define, the order statistics. That is the function $X_{(k)} of (X_1, X_2, \ldots, X_n)$, that takes on the values $x_{(k)}$ in each possible sequence $(x¿¿1, x_2, \ldots, x_n)¿$ of values assumed by the n dimensional random variables $(X_1, X_2, \ldots, X_n)$. That function, the function $X_{(k)}$; that $X_{(k)}$that is known as the kth order statistics or statistics of kth order; statistics of order k.

(Refer Slide Time: 06:54)



Now, the $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ that is called the set of order statistics for the random vector $(X_1, X_2, \ldots, X_n)$. So, from the given n dimension random vectors, we are getting another n dimension random vectors $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$. In which each $X_{(k)}$'s is the function of X_i's in which the $X_{(1)}$is a minimum$\{X_1, X_2, \ldots, X_n\}$, $X_{(n)}$the maximum$\{X_1, X_2, \ldots, X_n\}$. This set of n random variable is called the order statistics.

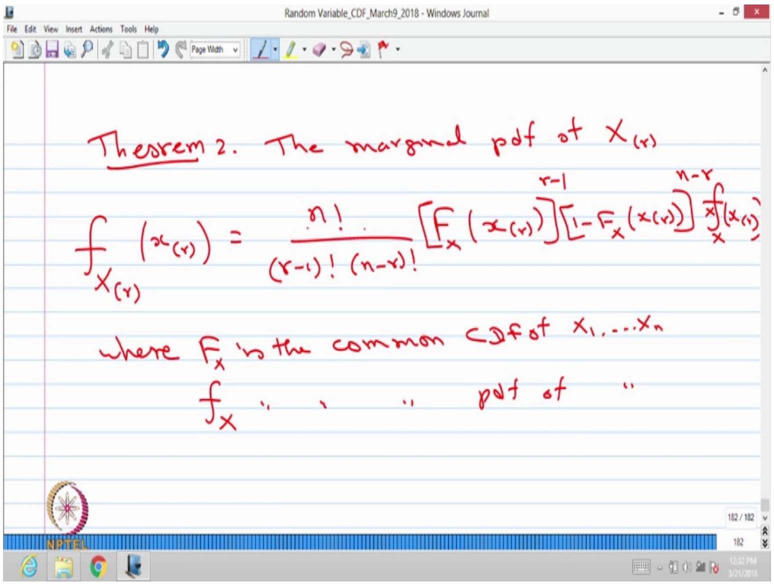
We will go for some important results on order statistics. For that I am going to make the assumption. Assume that the random variable X_i's, i is equal to 1 to n are iid random variables. We have already given the definition of iid; that means, each random variable having the same distribution and all the random variables are mutually independent. All the random variables are the mutually independent and having identical distribution. Therefore, they are called iid random variables with continuous type; that means, all the random variables are of the continuous type as well as they are iid random variables.

Going to give the result as the following theorem. One can find the joint probability density function of the order statistics as the joint probability density function with possible values, in terms of the probability density function of a X_i's that is $n!\prod_{i=1}^{n} f_{X_i}(x_{(i)})$. The product of this probability density function multiplied by the n! that is going to be the joint probability density function within the interval when $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$; otherwise it is going to be 0.

So, this result also comes from the theorem. One can find the joint probability density function of order statistics, in terms of the joint probability density function of $X_1, X_2, \ldots, X_n$, but since these random variables are i i d random variables. Therefore, in the right hand side, instead of joint probability density function of X_i's; we have a probability density function of X_i itself, the $n! \prod_{i=1}^{n} f_{X_i}(x_{(i)})$ by substituting the value x_i by $x_{(i)}$ between the interval $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$. It is a non zero joint probability density function, otherwise it is 0.

One can verify, the multidimensional integration with respect to $dx_{(1)}, dx_{(2)}, \ldots, dx_{(n)}$ that is going to be 1. So, here we are finding the joint probability density function of order statistics.

(Refer Slide Time: 12:42)

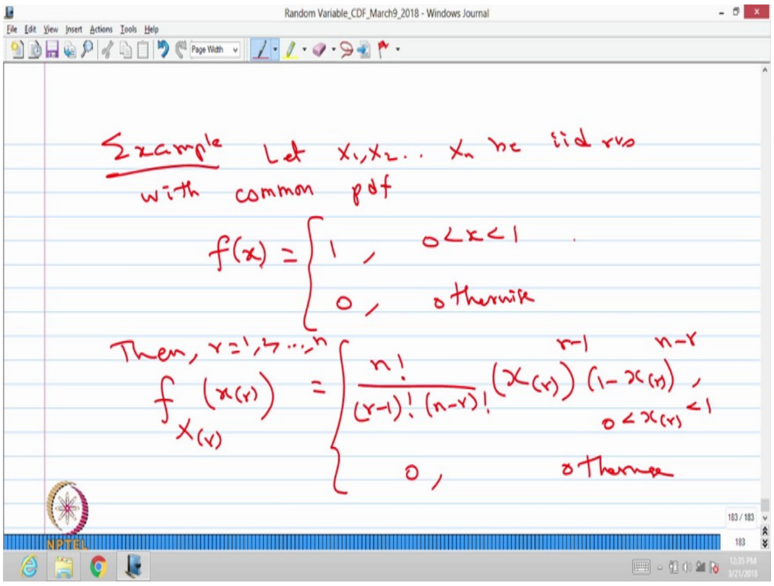


The next result, one can get the marginal distribution of order statistics that has second theorem. The theorem 2, the marginal probability density function of any rth order statistics is given by $f_{X_{(r)}}(x_{(r)})$ that is going to be, since you know the joint probability density function of order statistics; one can easily, not one can easily, one can get the marginal probability density function of any one rth order statistics from the n dimensional joint probability density function of order statistics. So, this is going to be $\frac{n!}{(r-1)!(n-r)!}$ multiplied by the CDF of since all the random variables are iid random variables, you do not need to mention or you can just mention x or any x_i evaluated at $x_{(r)}$, ¿¿. All are identical random variable..

This is going to be the probability density function of rth order statistics by using the CDF of any one random variable and the probability density function of any one random variable. So, instead of writing X_i, I am writing X where F is the common; since they are identical, it is a common CDF of the random variables $X_1, X_2, \dots, X_n$.

Similarly, the f_X is the common probability density function of the random variable $X_1, X_2, \dots, X_n$. So, one can get the probability density function of order statistics with the help of the probability density function and the CDF of the random variables; common CDF and the common PDF of the random variable $X_1, X_2, \dots, X_n$.

(Refer Slide Time: 16:24)



As an example, we are going to make, let $X_1, X_2, \dots, X_n$ be iid random variables with common with common probability density function; that means, I am going for all the random variables of the continuous type. So, the common probability density function that is 1, when x is lies between 0 to 1; otherwise it is 0. All are iid random variable as well as all are continuous type random variable with the common probability density function f(x) is this.

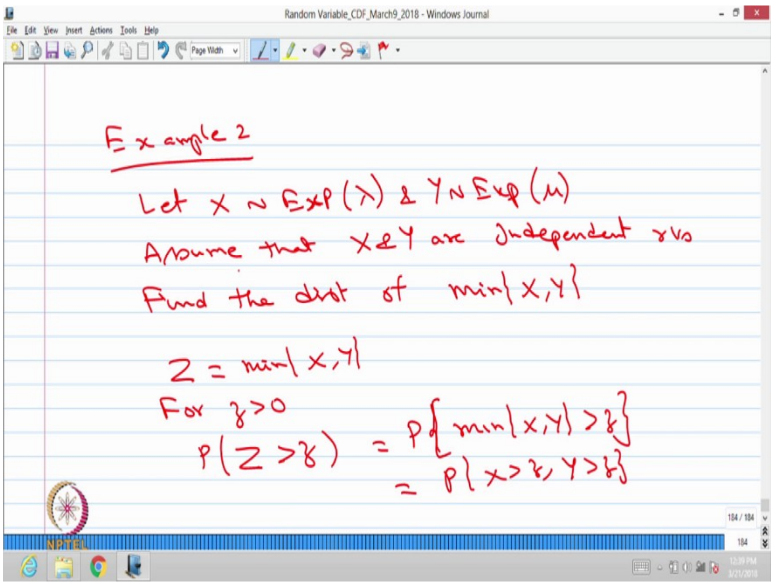
By using the previous result, one can get the probability density function of any rth order statistics, i.e., the probability density function of any rth order statistics; that is going to be

$\frac{n!}{(r-1)!(n-r)!}$ and then you have to substitute. See the previous result, you have to substitute the CDF as well as the probability density function.

And if you see this example; you can make out, the distribution of the X_i's is a uniform distribution between the interval 0 to 1. If the random variable is uniform distributed between the interval 0 to 1, you can get the CDF easily. So, you substitute $\left(x_{(r)}\right)^{r-1}\left(1-x_{(r)}\right)^{n-r}$ and multiplied by the probability density function that is 1.

So, this is valid whenever the rth order statistics is going to be 0 to 1 and this is true for r is equal to 1, 2 and so on till n; 0 otherwise. So, this result that the probability density function is non-zero, when $x_{(r)}$is lies between 0 to 1, otherwise it is 0 and this way you can find out the probability density function of r is equal to 1 to n.

(Refer Slide Time: 19:45)



We will go for second example. The second example is let X follows exponential distribution with the parameter λ and Y follows exponential distribution with the parameter μ. Assume that both the random variables are independent. We have two exponential distributed random variables with the parameters λ and μ respectively. Our interest is to find the distribution of, find the distribution of minimum{X, Y}.

Since it is only 2 random variables, you will have a minimum of 2 random variables as well as we will have a maximum of 2 random variables. Both together that is a set of order statistics. In that, I am interested to find out the distribution of minimum of 2 random variables. I can use the previous theorem, I can get the joint distribution; then I can get the marginal distribution by applying the theorem, but I am not going to apply the theorem.

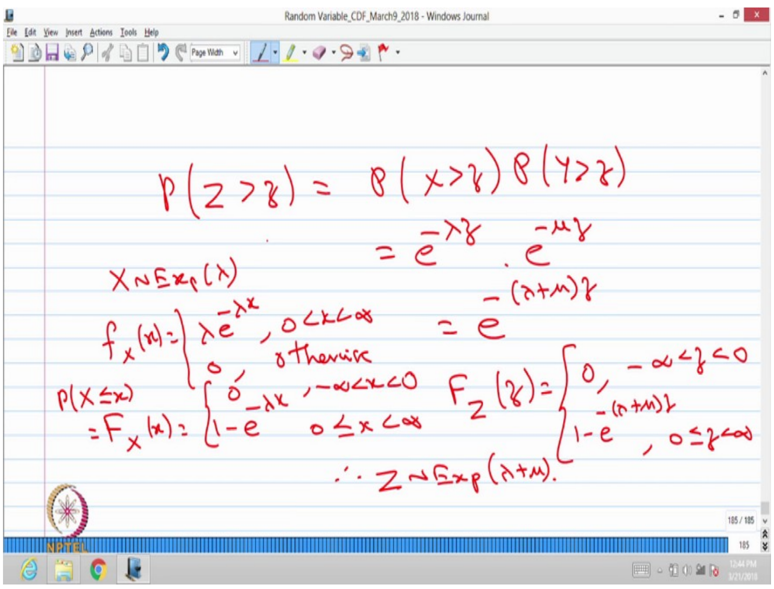
Instead of that I can easily able to find out the minimum of 2 random variables without using the previous theorem; that is let me make new random variable Z = min{X, Y}. Since X is a continuous type, Y is the continuous type, minimum{X, Y} that is also going to be a continuous type random variable. Therefore, I have to find the CDF of the random variable Z or probability density function of Z.

By using the previous theorem, I can get directly the probability density function of Z, but I am not going to do that. I am going to find out the distribution of Z, in the form of CDF. Since it is minimum, I will go for compliment CDF that is for z is greater than 0. The P{Z > z}. I am going for finding out the compliment CDF of the random variable Z, that is same as the probability of Z is nothing but min{X, Y}; that is greater than z.

Since min{X, Y} > z; that means, each random variable is also going to be greater than z. So, that is same P{X > z, Y > z}. If min{X, Y} > z, that means; both X and Y greater than z.

If I know the joint distribution, I can find out the probability of this. But I made the assumption; both the random variables are independent. Therefore, the joint distribution is same as the product of individual distributions. Therefore, the P{Z > z} = P{X > z}P{Y > z}.

(Refer Slide Time: 23:43)



I have already made the assumption X follows exponential distribution with the parameter λ. Therefore, I should know, what is the probability density function? Similarly, I should know, what is the CDF of exponential distribution? That is $1-e^{-\lambda x}$. It is 0 between - ∞ to 0 from 0 onwards, it is going to be $1-e^{-\lambda x}$.

So, I am going to use this result in the $P\{X > z\}$. The CDF is nothing, but $P\{X \leq x\}$. So, what I want is $P\{X > z\}$. So, that is same as $e^{-\lambda z}$. Similarly, Y is also exponential distributor with the parameter μ. The similar derivation makes; that is $e^{-\mu z}$. So, the conclusion is, it is $e^{-(\lambda+\mu)z}$, when z is going to be greater than 0.

So, I can write down things correctly that is CDF of Z that is going to be 0 when z is between - ∞ to 0 and 1 - $e^{-(\lambda+\mu)z}$, when z is starting from 0 to ∞. So, this is the distribution of Z. Distribution means, here it is a CDF.

In this page itself, you can compare; when X is exponential distribution the CDF is 0 from - ∞ to 0, then 0 to ∞ it is $1-e^{-\lambda x}$. It is in the same form. Therefore, one say two different variable having the same CDF, then you can conclude both the random variables are of the same distribution. Therefore, I can conclude Z is also same distribution of exponential distribution. The parameter, here it is - λ, here it is -$(\lambda+\mu)\wedge¿$ this is exponential distribution with the parameter λ. Therefore, Z is going to be exponential distribution with the parameter $(\lambda+\mu)$.

When two different random variables having the distribution of the similar form, then you can conclude both the random variables are having the similar distributions. So, here Z is going to be exponential distribution with the parameter λ + μ. So, the observation is whenever you have two independent exponential distributed random variable random variables, then the minimum of independent exponential distributed random variables is again exponential distribution.

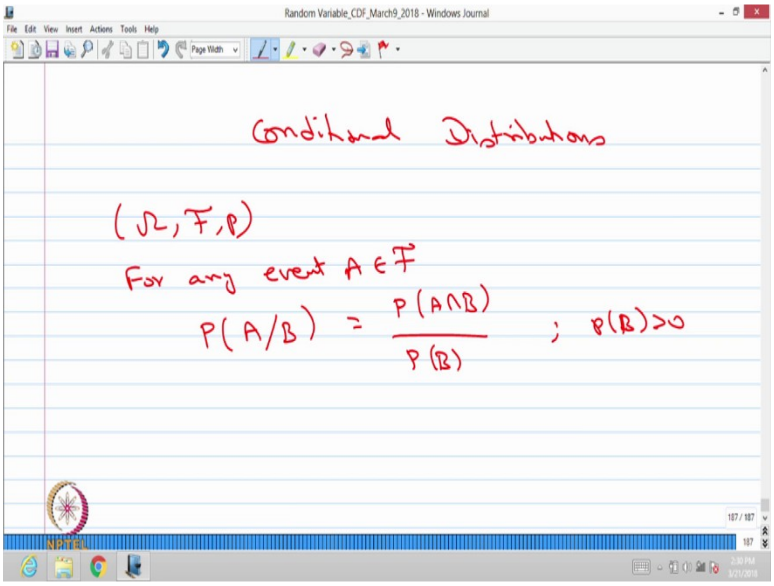
This concept can be extended for n dimensional random variable, n dimensional random variables; that means, if you have a mutually independent exponentially distributed n dimensional random variables, then the minimum of those mutually independent exponentially distributed random variables is again exponential distribution with the parameter is sum of the parameters of individual distribution. This is easy way of finding the minimum of two independent exponential distribution.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 33

In the last class in this module that is Functions of Several Random Variables. In the lecture 1 we discussed the distributions of functions of several random variables and in the second lecture we have discussed order statistics; that means, we have created n dimensional random variable, which is set of order statistics. And we discussed distribution of order statistics and we have discussed some simple examples also.

(Refer Slide Time: 00:44)

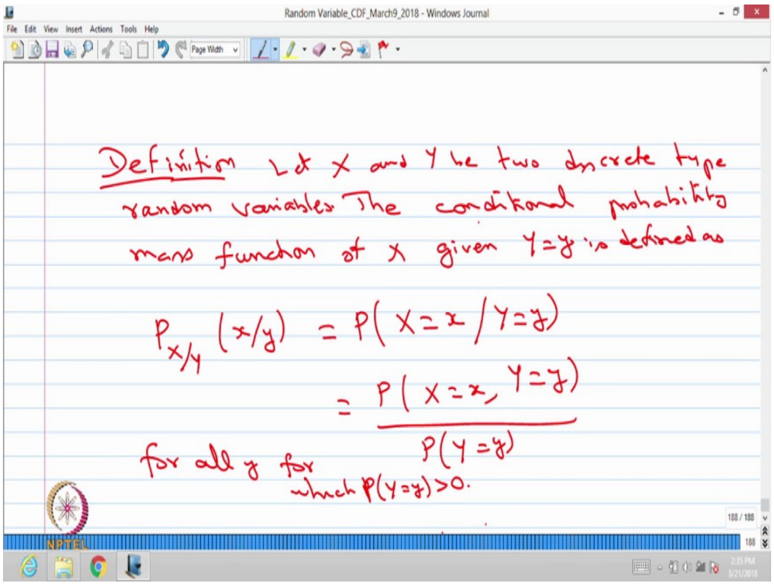


In this lecture we are going to discuss conditional distribution followed by random sum. So, these two topics which we are going to discuss in this lecture. Let us start with the conditional distribution.

Conditional distributions, we have already studied conditional of events, let me recall if you have a probability space (Ω, F, P) for any event A belonging to F the P{A/B}, B is also event that is same as the $\frac{P(A \cap B)}{P(B)}$ provided P(B) > 0. This is for conditional events, conditional probability of event A given event B that is $\frac{P(A \cap B)}{P(B)}$, provided P(B) > 0.

In the same concept we are going to introduce for the random variables whatever we have discussed in the module 1 for the events, we are going to discuss same thing for random variables. Therefore, conceptually it is same only thing is now we are going to discuss through the random variables. Let me start with the discrete type random variables first. The definition of conditional distribution.

(Refer Slide Time: 03:29)



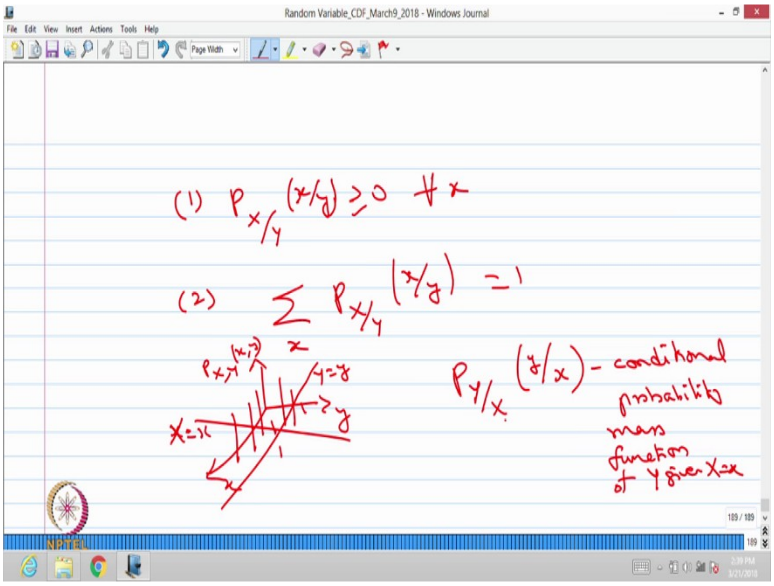
Then I give one example then I move into continuous type random variable and the conditional distribution of continuous type random variables, then one more example. Let's start with the definition.

Let X and Y be two discrete type random variables, the conditional probability mass function of the random variable X given the other random variable Y takes the value small y, that is defined as $P_{X/Y}$; that means, it is a conditional distribution of X given the other random variable Y takes a value y as a function of x and y, but we write it as a $P_{X/Y}(x/y)$it is not x divided by y; x slash y; that means, you have to treat y as a constant and the function is function of x the conditional probability mass function of X given Y takes a value small y here you gave to treat x as a variable and y as a constant.

That is same as the P{X = x/ Y = y} that is same as the conditional probability means $\frac{P\{X=x, Y=y\}}{P\{Y=y\}}$. This is for all y for which the P{Y = y}> 0.

So, the variable is x you have to treat y as a constant and this is defined whenever the P{Y = y} > 0. We call this as the conditional probability mass function of X given the other random variable takes a value Y is equal to small y. Since I use the word probability mass function one can easily verify this satisfies the properties of probability mass function that is this probability, the conditional probability mass function satisfies $P_{X/Y}(x/y) \geq 0$ for all x; x given y that probability value is always going to be greater than or equal to 0 for all x.

(Refer Slide Time: 07:03)



The second condition since it is a conditional probability mass function of X, the $\sum_x P_{X/Y}(x/y) = 1$; that means, this conditional probability mass function also satisfies these two conditions.

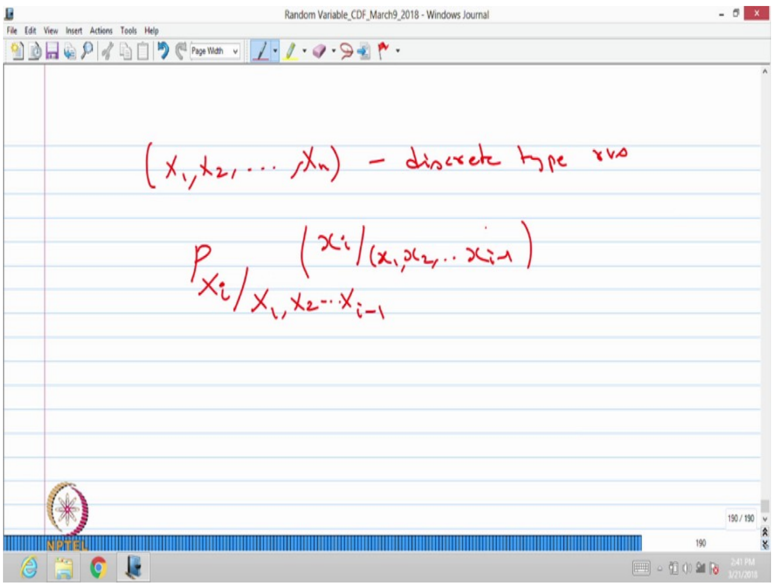
Since we are defining this random variable as X given Y takes a value small y therefore, we should use the word conditional probability mass function. One can visualize if it is two dimensional random variable with x y; the joint probability mass function for a different values of x, y. So, these different heights are nothing, but the joint probability mass function for the different values of x, y.

The conditional distribution can be visualized suppose we make Y = y; that means, you just think of this is going to be Y = y. Then look for what are all the possible joint probability mass function when Y = y. You collect those joint probability mass and you make it addition, then normalize it that becomes the conditional distribution. That means, this conditional probability mass function is nothing, but find out the joint probability mass function divided by P{Y = y} by normalizing, this value is going to be between 0 to 1.

The same way one can visualize the conditional distribution of Y/{X = x} that is conditional distribution of Y given X as function of y treating x as a constant. So, this is called conditional probability mass function of Y given X takes a value x. So, this also can be visualized by making X = x. So, this is a line, equal to x. So, you collect all the possibilities of the probabilities when X = x from the joint probability mass function, you normalize it therefore, the conditional distribution of Y given X = x you may get it.

So, this can be visualized only for two dimensional random variable and the same concept can be extended to any n dimensional random variable.

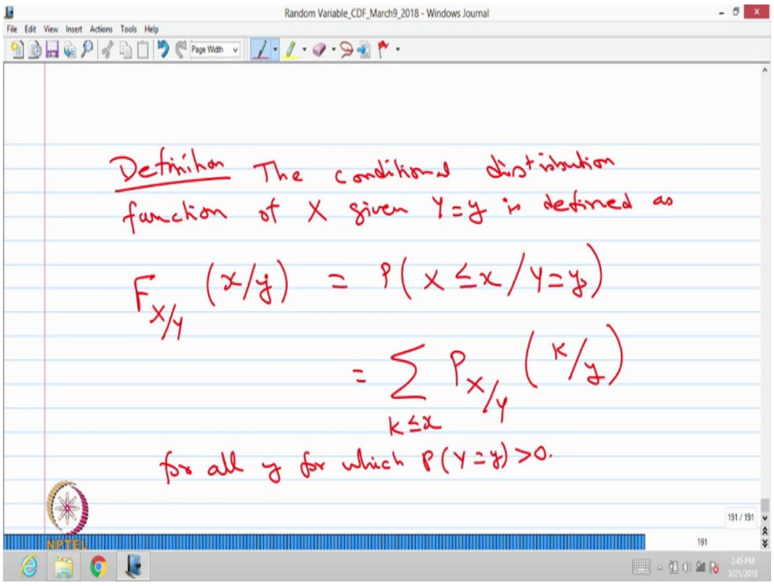
(Refer Slide Time: 10:59)



For example, suppose I have a discrete type random variable, n random variables all are going to be discrete type random variables. One can define conditional probability mass function of X_i given few X values. So, this is going to be a function of x_i given that all other values are already taken some values.

So, this is going to be a conditional distribution of X_i given X_1 takes value x_1, X_2 takes a value x_2 and so on X_{i-1}takes a value x_{i-1}. Now I am going for the simple example of how sorry, now I will go for the conditional distribution function.

(Refer Slide Time: 12:20)



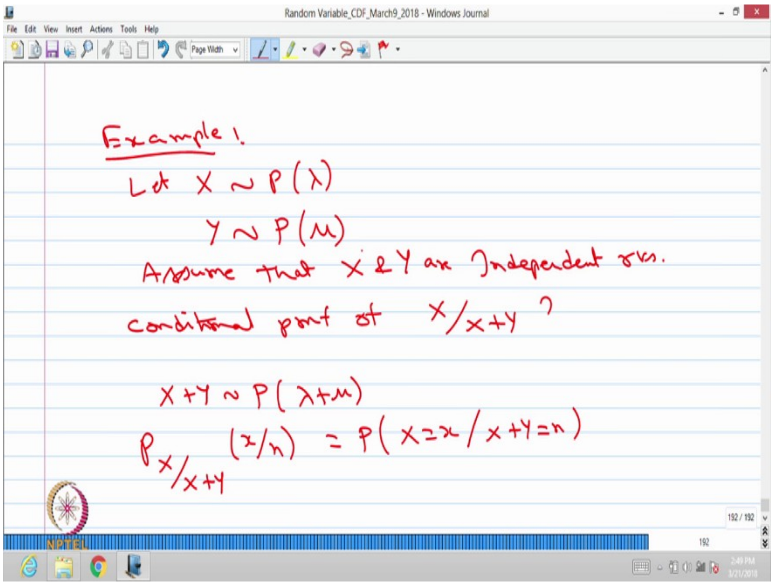
I am going to give the definition of conditional distribution of; conditional distribution function of the random variable X given the other random variable takes a value small y, that is defined as earlier we have given conditional probability mass function, now we are going to give conditional distribution.

That is the CDF therefore, it is a $F_{X/Y}$ in the conditional form therefore, X slash Y it is not X divided by Y. Whenever I use a slash; that means, the other random variable already taken some value. Again, this is also a function of x and y, but you have to treat y as a constant. That is same as the P{X ≤ x / Y = y}. That is same as already we have defined probability of X takes the value x given Y takes a value small y

Now, we are finding the conditional distribution function of the random variable X given Y takes the y. Therefore, we got X ≤ x that is same as from the conditional probability mass function of various values of; you can treat some k given y where k ≤ x. So, this is $\sum_{k \le x} P(X=k/Y=y)$ that will give P{X ≤ x / Y = y} that is called conditional distribution of X given Y takes a value small y. This is also true whenever for all y for which the P{Y = y} > 0; otherwise, the conditional probability mass function itself not well defined therefore, you cannot get the conditional distribution function.

So, in the first definition we have explained how to get the conditional probability mass function when two random variables are of the discrete type. In the second definition we have given when both the random variables are of the discrete type, how one can represent conditional distribution function of X given Y takes a value small y. We will go for one easy example in which you can discuss the conditional distribution again we take only two random variables. So, the same concept can be extended to many random variables.

(Refer Slide Time: 16:11)



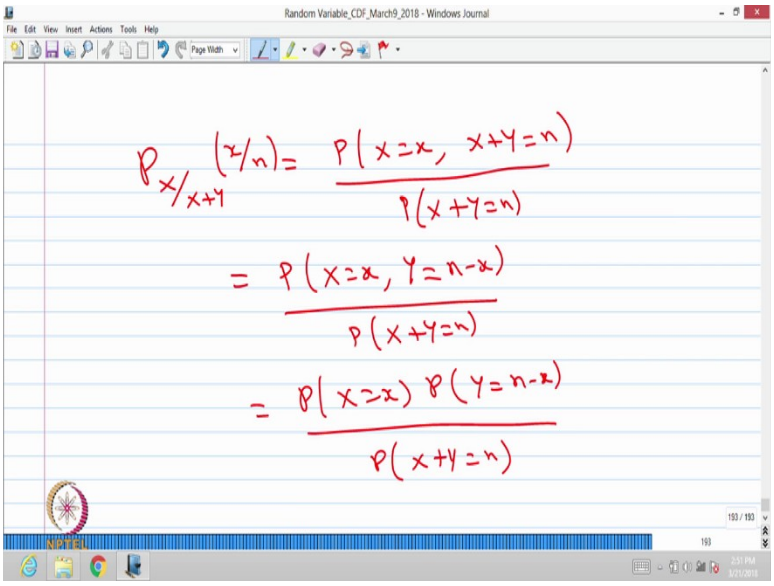
So, as example we give. So, example 1 let X be Poisson distributed with the parameter λ and Y be again Poisson distributed with the parameter μ and I make the assumption, assume that X and Y are independent random variables.

We will find out what is the conditional probability mass function of X/X+Y. Earlier in the definition we have discussed the conditional probability mass function of one random variable given another random variable, but here conditional distribution of X/X+Y. That means, first you should know what is the distribution of X+Y you know, then we have to go for finding the conditional probability mass function of X/X+Y.

Since X is a discrete type variable, Y is also discrete type random variable we know that X + Y is also going to be discrete type random variable therefore, we can go for finding a conditional probability mass function of X/X+Y that is a question. We know that X + Y is going to be again Poisson distributed with the parameter λ+ μ, we got this result from the earlier exercise; earlier examples of finding the distributions of functions of several variables; several random variables.

So, we know that when X is Poisson Y is also Poisson the summation is also going to be Poisson distributed by the reproductive property also. So, we know the distribution of X + Y is Poisson distribution, we will go for finding what is the $P_{X/X+Y}(x/n)$ we treat this value is going to be n, X + Y = n that is same as the P{X = x/ X+Y = n}.

(Refer Slide Time: 19:29)

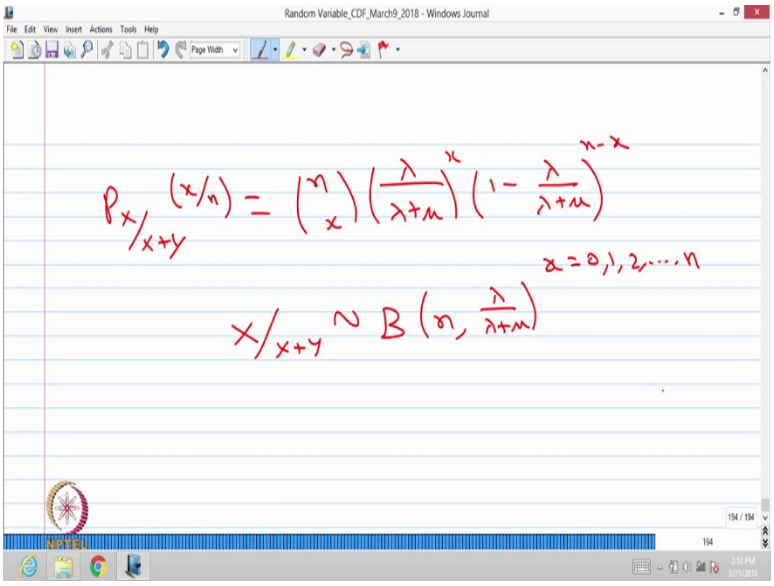


That is same as the left side is $P_{X/X+Y}(x/n)$ that is same as the $\frac{P(X=x, X+Y=n)}{P(X+Y=n)}$ by using the definition of conditional probability mass function. That is same as the $\frac{P(X=x)P(Y=n-x)}{P(X+Y=n)}$ since X = x the other one Y is going to take the value n – x.

Already we made the assumption the random variables X and Y are independent. Therefore, the numerator joint probability mass function is the product of a probability mass functions. Now, we can substitute that the probability mass function for the random variable X and Y similarly, X + Y. You can substitute the probability mass function of X we know that the random variable X is Poisson distributed you can substitute.

Similarly, Y is also Poisson distributed with the parameter μ we can substitute the probability mass function and in the denominator X+Y that is also Poisson distributed with the parameter λ + μ. So, you can substitute the probability mass function of the denominator also.

(Refer Slide Time: 21:36)



You can substitute and you can do the simplification you can get the answer it is

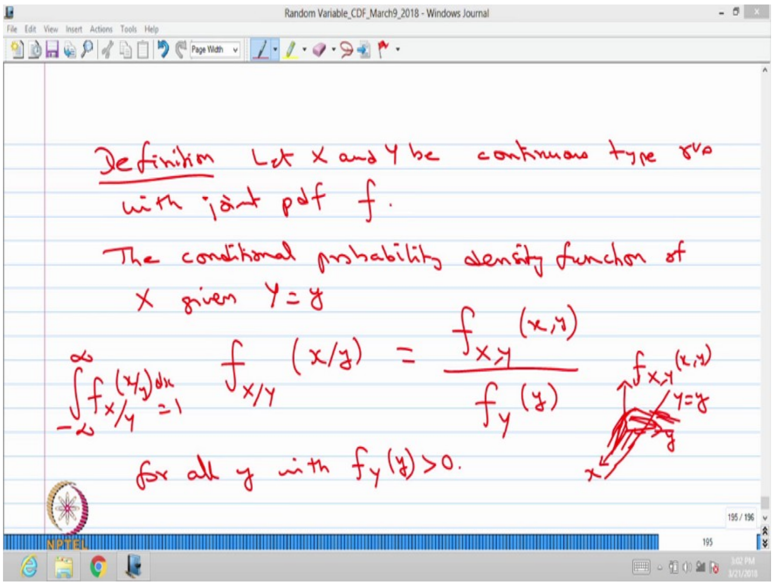
$$^{n}C_{x}\left(\frac{\lambda}{\lambda+\mu}\right)^{x}\left(1-\frac{\lambda}{\lambda+\mu}\right)^{n-x}.$$

This is the $P_{X/X+Y}(x/n)$, n to be treated as a constant. So, here the possible values of x are 0 or 1 or 2 and so on till n. By seeing the probability mass function of this conditional distribution of X/X+Y, you can conclude X/X+Y that follows binomial distribution with the parameters n, p. Here the p $=\frac{\lambda}{\lambda+\mu}$.

From the Poisson distribution $\lambda > 0$, $\mu > 0$ therefore, $\frac{\lambda}{\lambda+\mu}$ that is lies between open interval 0 to 1. Therefore, you can conclude this follows binomial distribution with the parameters n, $\frac{\lambda}{\lambda+\mu}$. It is a very important result from a Poisson distribution, the summation is also going to be a Poisson distribution if they are independent by the reproductive property, whereas the conditional distribution over one random variable given sum of these two random variables, that is binomial distribution.

Now, we will move into the conditional distribution for the continuous type random variables. Let me start with the definition of a conditional probability density function definition.

(Refer Slide Time: 24:09)



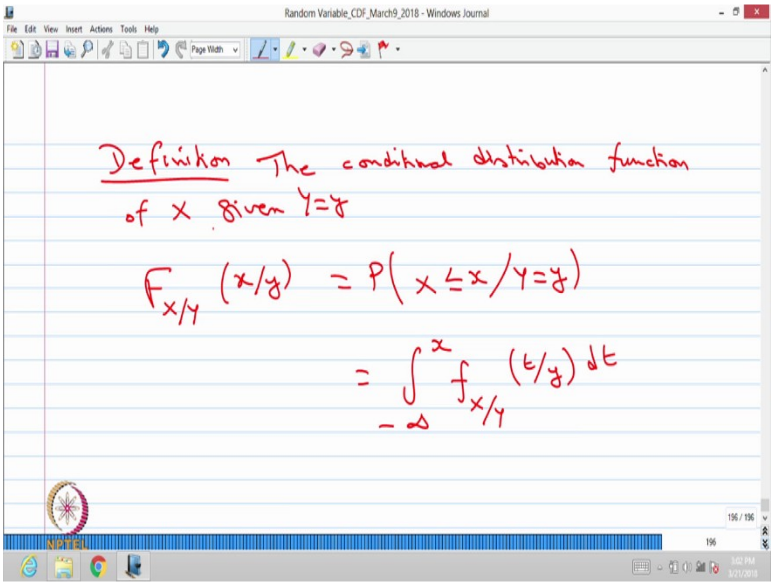
Let X and Y be continuous type random variables with joint probability density function; with the joint probability density function small f; the conditional probability density function of the random variable X given the other random variable takes a value y that is defined as $f_{X/Y}$. I am using f for probability density function, F for the CDF, small f, but in the suffix I am going to use the notation X/Y; that means, it is the conditional probability density function of X given Y.

Again, this is also going to be a function of x and y, but you have to treat y as a constant. This is same as $\frac{f_{X,Y}(x,y)}{f_Y(y)}$. For all y with the probability density function at that point y > 0.

You we know that the probability density function will be greater than or equal to 0, but when you define the conditional probability density function you have to make sure that the denominator does not vanish. That is $f_Y(y) > 0$, this ratio is going to be the probability density function. From the probability density function one can get the probability of any interval. So, here we are getting a probability density function.

Now, we are going for the same way conditional distribution as a next definition.

(Refer Slide Time: 26:50)



That is, the conditional distribution function of the random variable X given the other random variable takes a value small y that is F. This is in the conditional distribution that is $F_{X/Y}(x/y)$, that is same as the P(X ≤ x/Y = y).

That is same as since both are continuous type random variable, that is same as $\int_{-\infty}^{x} f_{X/Y}(t/y)dt$. That means, by integrating the conditional probability density function, you will get conditional distribution function. The way I have explained conditional probability mass function, one can visualize the conditional; the way I explained the conditional probability mass function, one can visualize conditional probability density function also. Always the joint probability density function is a surface over the x-y plane; that means, this function is always greater than or equal to 0 and double integration over the region in which it is going to be greater than 0, that volume has to be 1.

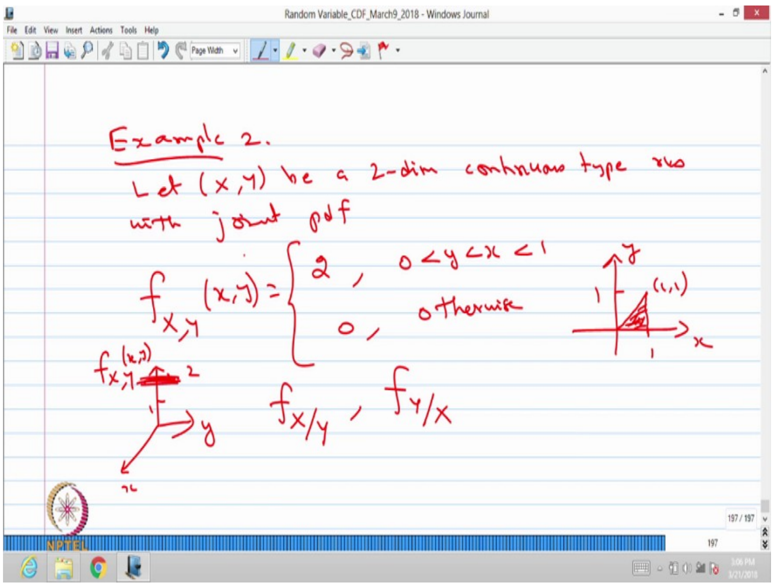
So that means, if you think of some surface over the x-y plane; that is the joint probability density function whenever you go for Y = y; that means, you are just cutting a one plane; that means, the surface and the plane Y = y that will make a one cut. So, you will get where the plane Y = y cutted the surface, you will get some sort of curve.

That curve by with a value of $f_Y(y)$, you are diminishing the curve or enlarging the curve based on the values going to be less than 1 or greater than 1. So, that the area below that area below the curve is going to be one; that means, this is a probability density function. So, $\int_{-\infty}^{\infty} f_{X/Y}(x/y)\,dx$ that is going to be 1; The conditional probability density function is nothing, but intersecting the surface with the plane Y = y by multiplying $1/f_Y(y)$ means you are either enlarging or diminishing the curve so that the area is going to be 1.

That means, the $1/f_Y(y)$ is a normalizing constant for the $f_{X/Y}(x/y)$. This satisfies the both the properties that is always going to be greater than or equal to 0 and $\int_{-\infty}^{\infty} f_{X/Y}(x/y)\,dx$is going to be 1, similar to the discrete type.

So, whenever you go for the conditional distribution you are going for again getting the same property of probability mass function or probability density function based on the random variable discrete or continuous. Whereas, here the $F_{X/Y}(x/y)$when both the random variables are of continuous type. This integration from minus infinity to x which is same as the usual way of one dimensional random variable instead of the probability density function you are using a conditional probability density function of X given Y.

(Refer Slide Time: 32:09)



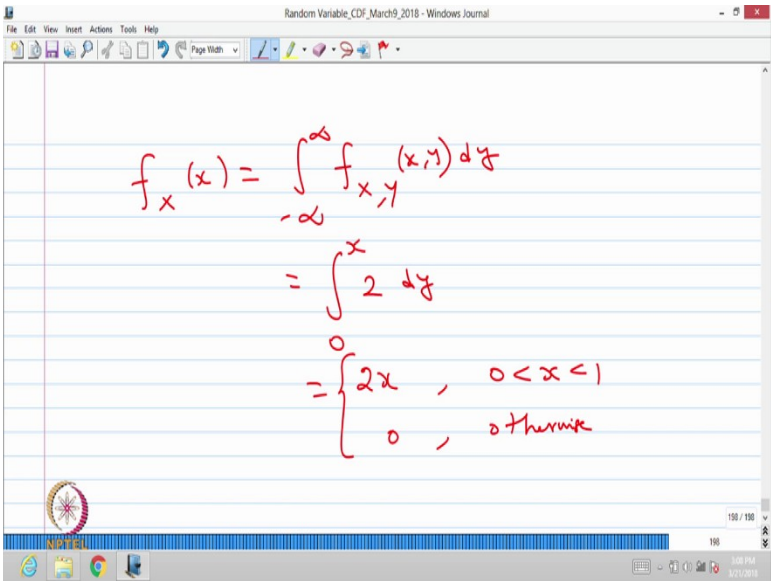
Now, we will go for another example to explain for the continuous type random variables. Already we have given one example for discrete type random variable now we will explain the same concept for continuous type random variables. Let X, Y be a two dimensional continuous type random variable with the joint probability density function is given by is a function of x and y that takes a value 2 when y is lies between 0 to x and x is lies between y to 1; 0 otherwise. This is a joint probability density function of two dimensional continuous type random variables. You can verify if you do the $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy$ will be 1.

But before that we will find out what is the region in which the joint probability density function is greater than 0 that is 2. So, we can make x versus y, you shade the region in which y is lies between 0 to x, x is lies between y to 1 so; that means, 1 1 that is (1,1). So, you draw a line that is y = x therefore, this is a shaded region. So, in this region the joint probability density function is 2.

That means x axis, y axis the joint probability density function is going to be at the height of 2 it has some sort of a plane at the height of 2 over the region xy in this triangle part. So, this is the joint probability density.

Our interest is to find out the $f_{X/Y}$ as well as the $f_{Y/X}$. For this problem we are interested to find the $f_{X/Y}$similarly $f_{Y/X}$. To do this first you should know what is the marginal distributions of X and Y then only you can go for it. For first one you should know the marginal distribution of Y, for the second problem we should know the marginal distribution of X. So, let us find both the marginal distribution of X and Y, then we will go for finding the $f_{X/Y}$and $f_{Y/X}$.

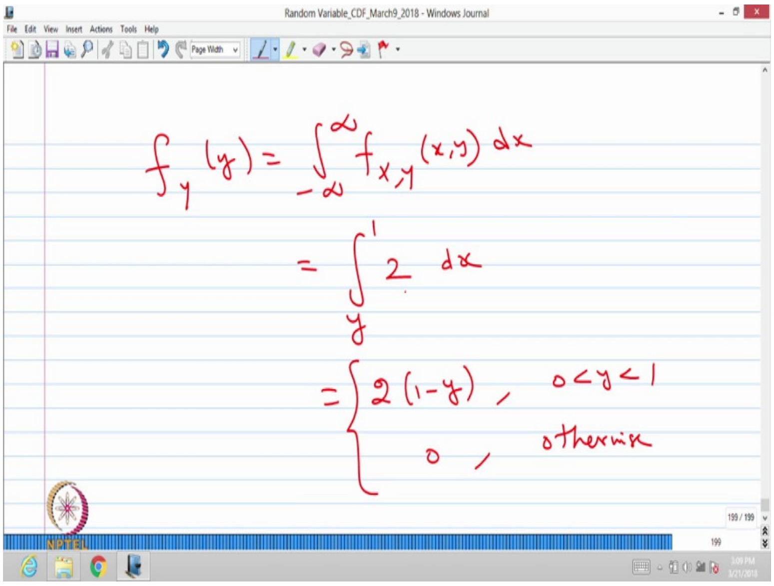
(Refer Slide Time: 35:56)



The marginal probability density function of X that is nothing but the $\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$. So, this is same as you should go and see the joint probability density function value is 2 between $0 < y < x < 1$ and you want integration with respect to y. Therefore, $\int_0^x 2\,dy$ therefore, when you simplify, you will get the answer is 2x and this is going to be 2x when x lies between 0 to 1.

So, the probability density function is going to be 2x when x is lies between 0 to 1; 0 otherwise. Similarly you can do the marginal distribution of Y.

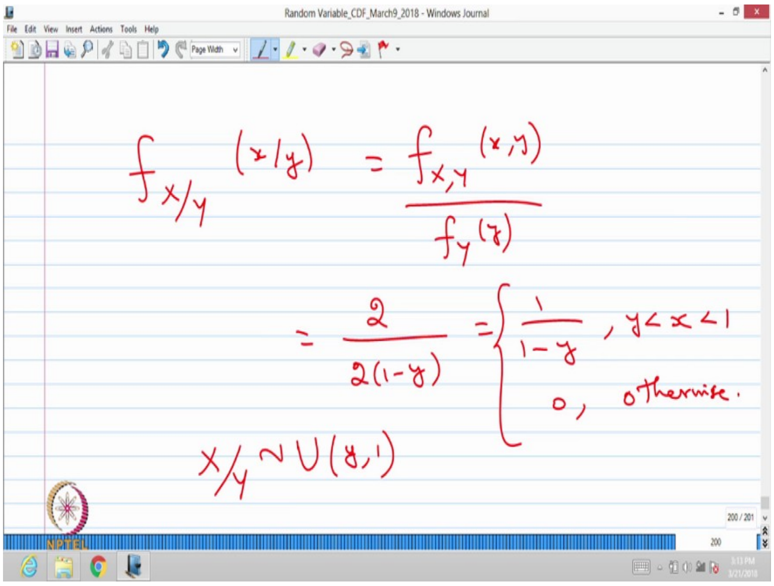
(Refer Slide Time: 37:12)



That is the probability density function of Y, that is going to be $\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$, that is same as. Now, you see the interval again it is 2 between 0 < y < x < 1. Therefore, $\int_{y}^{1} 2dx$. When you do the simplification, you will get the answer 2(1 - y), and this is going to be; the probability density function is 2(1 - y) when y is lies between 0 to 1

So, this is the marginal distributions of X and Y, now we will go for finding the conditional distributions.

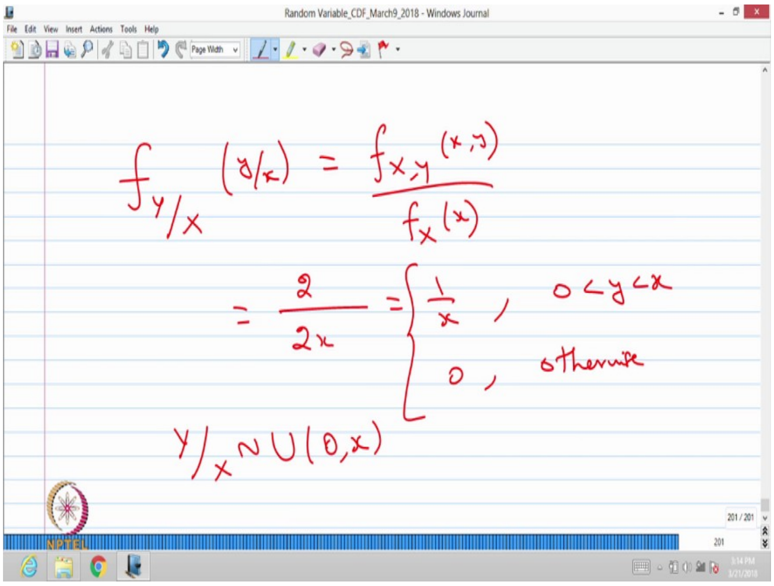
(Refer Slide Time: 38:27)



The first one we are going for finding the $f_{X/Y}(x/y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$. The joint probability density function is 2 between that interval and if the denominator we have already got the answer 2(1- y).

So, when you simplify you will get $\dfrac{1}{1-y}$. So, this is the $f_{X/Y}(x/y)$. So, you have to treat y as a constant here. So, the value is going to be $\dfrac{1}{1-y}$when x lies between y to 1. When x takes a value y to 1, the $f_{X/Y}(x/y)$ that is $\dfrac{1}{1-y}$; 0 otherwise.

(Refer Slide Time: 39:51)



Similarly, you can go for finding $f_{Y/X}(y/x)$by treating x as a constant. Again, I write the same definition $\frac{f_{X,Y}(x,y)}{f_X(x)}$. So, this is same as the joint is 2 and marginal is 2x. So, the simplification will give $\frac{1}{x}$. So, this is going to be $\frac{1}{x}$ when y lies between 0 to x; 0 otherwise. In this you have to treat x as a constant, y as a variable; y lies between 0 to x and the $f_{Y/X}(y/x) is \frac{1}{x}$, $f_{X/Y}(x/y)$ that is $\frac{1}{1-y}$, where x lies between y to 1.

Here you have to treat y as the constant. By seeing the probability density function, you can say $f_{X/Y}(x/y)$ is continuous type uniform distribution between the interval y to 1; you have to treat small y as a constant.

So, the conditional distribution of X given capital Y takes a value small y, that is continuous type uniform distribution with the parameters or with the interval with the interval y to 1. Similarly, here the $f_{Y/X}(y/x)$ that follows continuous type uniform distribution between the interval 0 to x, here you have to treat x as a constant; small x as a constant.

So, we started with the joint distribution; we started with the joint distribution and we are finding the conditional distribution after finding the marginal distributions of individual random variables. So, this is a very easy problem in which you are applying the definition and you can feel what could be the distribution of the conditional distributions.

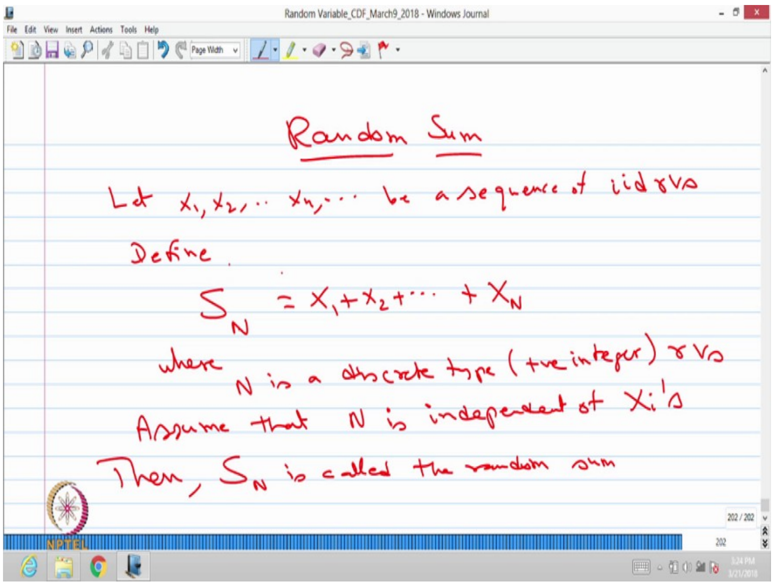
Introduction to Probability Theory and Stochastic Processes

Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 34

Now, we will move into the next topic that is Random sum.

(Refer Slide Time: 00:03)



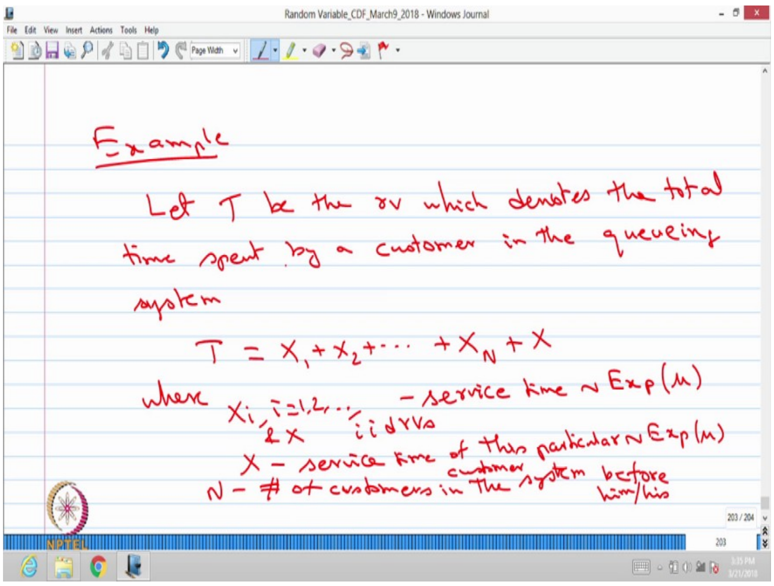
Let $X_1, X_2, \ldots, X_n$ be a sequence of; let me start with iid random variables; be a sequence of iid random variables. Define the new random variable as $S_N = X_1 + X_2 + \ldots + X_N$. I am going to define the random sum with the sequence of iid random variables defining the random variables S_N as a sum of N random variables

This is different from the earlier sum of the random variable, in the sense I am going to make N which I made it suffix, that is a discrete type, in particular it is type positive integer valued random variable. N is a discrete type in particular it is a positive integer valued random variable.

I am going to make the assumptions, assume that N is independent of X_i's. The way we create the sum of random variables, all those random variables are mutually independent and identically distributed. And how many random variables I am going to add that is also random, that is a positive integer valued discrete type random variable which is independent of X_i's.

This S_N that is called a random sum, then the S_N is called the random sum. You will come across many problems of this form, we will be adding many iid random variables and how many random variables we are going to add, that is also going to be a random variable. In that case the total number of random variables added that is a random sum. We will go for a one example for this. So, our interest is to find out the distribution of the random sum, once I know the distribution of the X_i's and N.

(Refer Slide Time: 03:56)



So let me start with the example, think of some queuing system in which the people are entering into the system, getting served and leave the place. There is the only one person to do the service; that means, the number of servers in the system is only one. Because of only one server in the system whoever comes when somebodies service is going on, he has to wait. He can think of queuing discipline is a first come first serve.

So, I am going to make a random variable, let T be the random variable which denotes the total time spent by a customer; customer means any person in the queuing system. Here I made a very simplest queuing system in which only one server in the system, there may be n number of people can wait. So, after the service is over only they can leave the system.

We are interested to find out the distribution of T if I supply the information about how much time taken for the customers service, and what is the distribution of a number of customers in the system. That means, I am going to represent this T as a random sum in the form of $X_1 + X_2 + \ldots + X_N + X$. Where each X_i's i from 1 to N are the service time the random variable denotes the service time for ith customer.

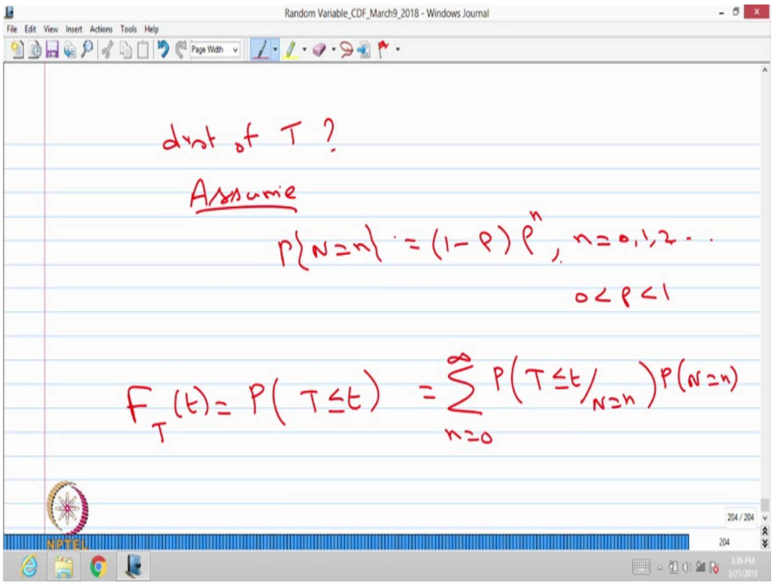
And I make the assumption the service time follows exponential distribution with the parameter μ. I will repeat the total time spent by a customer in the system that can be represented as $X_1 + X_2 + \ldots + X_N$+ X, each X_i's are iid random variables, where each X_i's are iid random variables with the common distribution which is exponential distribution with the parameter μ where X_i denotes the service time service time of any customer.

Whereas the X that is the service time of this particular customer, which is independent of all the previous service time. Therefore, all the X_i's and X are independent random variable, this also follows exponential distribution with the parameter μ. I am just giving another label for X to say that that is the service time for his own. Therefore, the T is going to be a random sum, because you never know how many customers are going to be before him when he enter in to the system.

So, N is the random variable. So, I can write the n as the N. We can write n as the N and X_i's are going to be iid random variables including the X. So, now, I can go for N is the number of customers in the system, which is independent of service times of all the customers. My interest is to find out the distribution of T. So, I have to supply what is the distribution of the number of customers in the system also. So, here the N means the number of customers in the system before him or his turn, that is the number of customers in the system when he enter into the system that is N.

So, in this case when N = 0; that means, nobody is in the system before he enters. When he enters nobody in the system; that means, T = X.

(Refer Slide Time: 10:16)



I am going to find out the distribution of T; for that you should know the distribution of X_i's as well as the distribution of N. So, we have already said the X_i's; the distribution of X_i's are exponential distribution with the parameter μ.

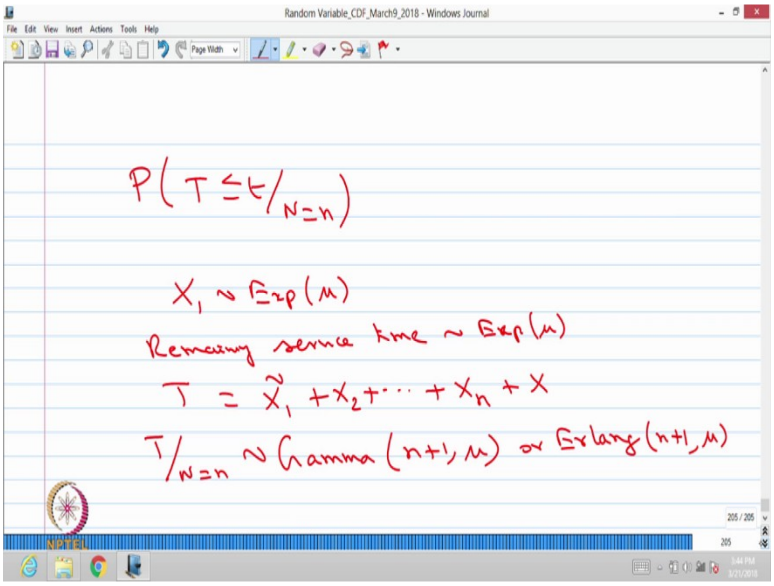
Now, I am going to make the assumption of, assume that the distribution of N that is N is the number of customers before he enters. So, that can be possible values are 0, 1, 2 and so on. Therefore, this is going to be; I just make the assumption that follows $(1-\rho)\rho^n$, where n can takes the value 0, 1 and so on; where the ρ is open interval 0 to 1. So, this is the probability mass function of the random variable N. When n takes the value 0; obviously, the total time spent by a customer that is going to be only his own service time that is the exponential distribution.

So, let us find out the probability density function of T as the function of t. It is easy to find out the conditional distribution of T given N takes the value; then go for finding the probability density function of T. That is suppose I want to find out the CDF of the T that is the $P(T \le t)$, that is same as the $\sum_{n=0}^{\infty} P(T \le t / N=n) P(N=n)$.

Left side we started with the CDF of the random variable T, that is same as P$(T \le t)$, that is same as first we compute the $\sum_{n=0}^{\infty} P(T \le t / N=n) P(N=n)$. We are getting the distribution of T by using a total probability theorem which we discussed in the beginning for the events, but now we are applying the same concept for the random variables.

That means that finding distribution of T by using the conditional distribution of T, by using the sort of a random sum concept then by using the total probability theorem we are getting the distribution.

(Refer Slide Time: 13:50)



But for that first you should find out the what is the conditional distribution of T given N takes a value small n. Once you fix the total number of customers before him that is going to be n.

The total time spent by a customer is going to be the service time of all the n people plus his own service, in that there is a small issue when someone enters there is a possibility, the service of the customer who is under service it may be keep going. And it may be the remaining service time for the customer who is under service plus the first service of the second customer, third customer and so on till the nth customer then plus his own service.

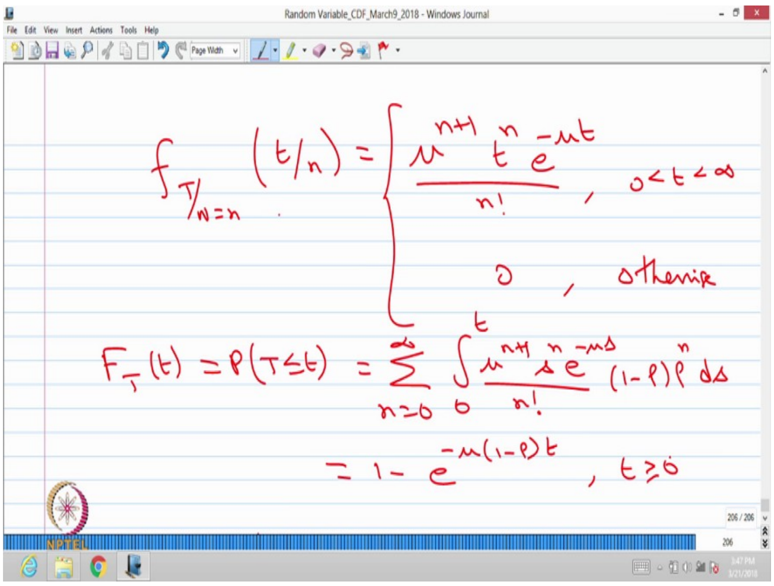
So, again I am going to use the logic of exponential distribution, since the first customer under service whose service is exponential distribution with the parameter μ. The remaining service time that is also going to follow exponential distribution, with the parameters μ by using the memoryless property. That is $P\{X > t+s/X > s\} = P\{X > t\}$; that means, if the past is erased and the conditional probability is again going to be the P{X > t}, which has same distribution of exponential distribution.

Therefore, the total time spent by a customer when already n people are in the system, that is going to be remaining service time of the customer under service plus service of n - 1 customer plus its own service. That is capital T is $X_1 \, sort\, of \, \tilde{\iota}$; tilda means remaining service time $+X_2+\ldots+X_n+X$.

Each service time is exponential distribution, identical, mutually independent with the X service is one followed by the other therefore, you can conclude for a fixed N. So, this is the distribution of T for fixed N that is gamma distributed with the parameters n + 1 with the second parameter μ. This is one of the important properties of a gamma distribution, the gamma distribution which has the parameters whenever it is a positive integer and the other parameter μ, that can be visualized as the sum of mutually independent exponentially distributed random variable with the parameter λ.

There is another name for this that is Erlang distribution, with the n +1 stages with each stage of exponential distribution with the parameter μ. So, this is a conditional distribution of T given N takes a value small n. Therefore, we know what is the $f_{T/N=n}(t/n)$.

(Refer Slide Time: 18:04)



That is $\frac{\mu^{n+1} t^n e^{-\mu t}}{n!}$ it is a gamma(n + 1), therefore, it is n!. So, this is valid when the t lies between 0 to infinity; 0 otherwise. This is basically a gamma distribution with the parameters n +1, μ. Now we are going back to finding out the $F_T(t) = P(T \leq t)$.

The above one is the conditional probability density function of T given N. Now, we need conditional distribution. So, therefore, it is $\sum_{n=0}^{\infty} \int_0^t \frac{\mu^{n+1} s^n e^{-\mu s}}{n!} (1-\rho) \rho^n ds$.

It is a good example of introducing the conditional distribution, random sum and the properties of exponential distribution and the gamma distribution. If you do the simplification one can get the answer that is $1-e^{-\mu(1-\rho)t}$.

So, this is a CDF of T for t ≥ 0; otherwise it is 0. It is a CDF of T as a function of t that is $1-e^{-\mu(1-\rho)t}$. With this we are completing the module of functions of several variable; several random variables starting with the distributions of functions of several random variables, then order statistics, then as a third lecture conditional distributions and finally, random sum.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

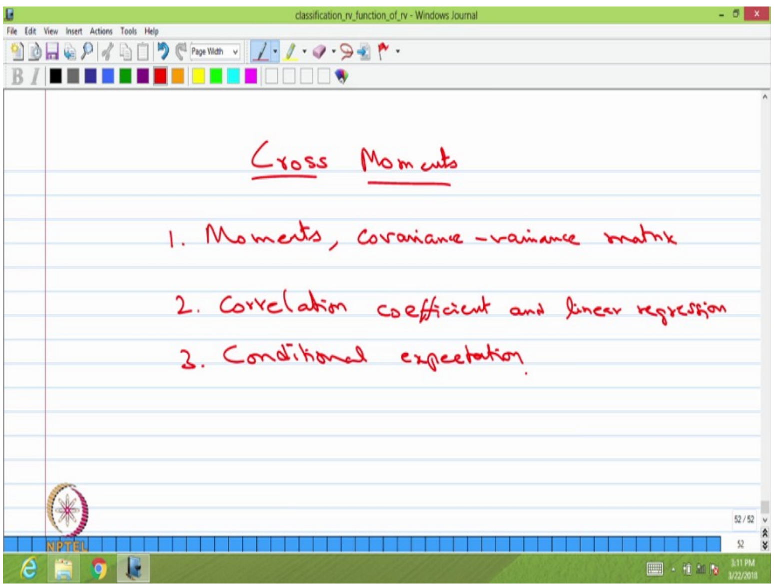
Module – 07
Cross Moments
Lecture – 35

In this model we are going to discuss Cross Moments. In the earlier models we have started with the basics of probability then we have discussed the single dimension random variable. Then third model we discussed the moments and equalities for a single dimension random variable, then in the fourth model we discussed the standard distributions, both the discrete and the continuous type.

And in the fifth model we discussed the two and high dimensional random variables or random vectors. In that we discussed the joint probability mass function, joint probability density function. In the same model we discussed the independent random variable source.

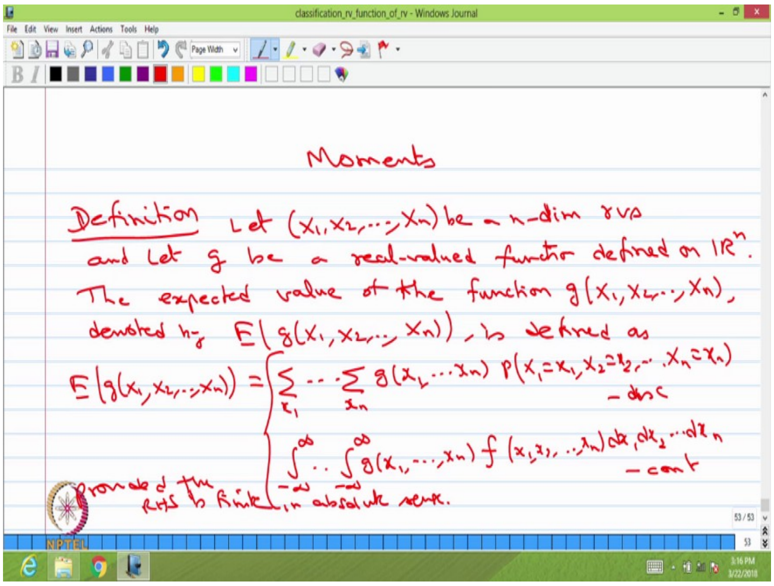
In the sixth model we discussed the functions of several random variables. In that we discussed the distribution of functions of several random variables and we discussed order statistics, then we discussed conditional distributions and random sum. In this model we are going to discuss cross moments.

(Refer Slide Time: 01:43)



In this section we are going to discuss the cross moments with the following lectures: the first lecture we are going to discuss the various moments. The first order moment, second order moment and so on, then we are going to discuss the covariance variance matrix that is in the first lecture. In the second lecture we are going to discuss the correlation coefficient and linear regression. In the third lecture we are going to discuss the conditional expectation.

(Refer Slide Time: 03:15)



Let us start with moments, before going to the moments of the first order, second order and the cross moments for multidimensional random variables, first let me give the definition of expected value of function of random vector that as a definition. So, this definition is going to be useful to compute the various moments of the several random variables.

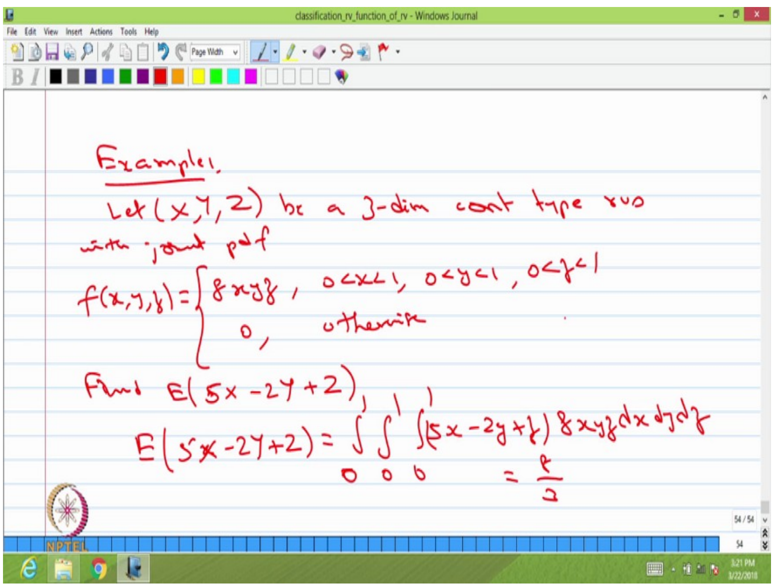
Let $(X_1, X_2, \ldots, X_n)$ be n dimensional random variables and let g be a real valued function defined on R^n; that means, it has n variables. The expected value of the function $(X_1, X_2, \ldots, X_n)$, that is denoted by E[g$(X_1, X_2, \ldots, X_n)$]. That is defined as E(g $(X_1, X_2, \ldots, X_n)$) that is equal to; suppose $X_1, X_2, \ldots, X_n$ are discrete type random variables, then this is going to be $\sum_{x_1} \cdots \sum_{x_n} g(x_1, x_2, \ldots, x_n) P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$ when all these random variables are of the discrete type.

If all such n random variables are of the continuous, then it is $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n)\, dx_1 \dots dx_n$ when all this random variables are of the continues type. Provided when the random variables are of the discrete type random variable the summation in absolute sense it is finite or when the random variables are of the continuous type in absolute sense the integration has to be finite quantity; that means, provided the right hand side is finite in absolute sense.

This is the same thing we have done it for only one dimensional random variable. When X is a random variable g is Borel measurable function g(X) is a random variable one can compute the R^1 can find the value that is expectation of g(X) with the summation, if they are if it is a discrete type random variable; integration, if it is a continues type random variable provided the summation or integration is going to be a finite quantity in absolute sense. The same provided condition is extended with the n dimensional random variable and we are finding the expected value of the function of random variables that is g.

When we say g is a real valued function, we have seen that this is going to be Borel measurable function in that we are finding the E(g$(X_1, X_2, \dots, X_n)$). Again, we are not finding the distribution of g$(X_1, X_2, \dots, X_n)$, we are directly computing the expected value of the random variable g$(X_1, X_2, \dots, X_n)$. To find the various moments you can use this definition nicely, so that you can get the various moments.

(Refer Slide Time: 09:46)

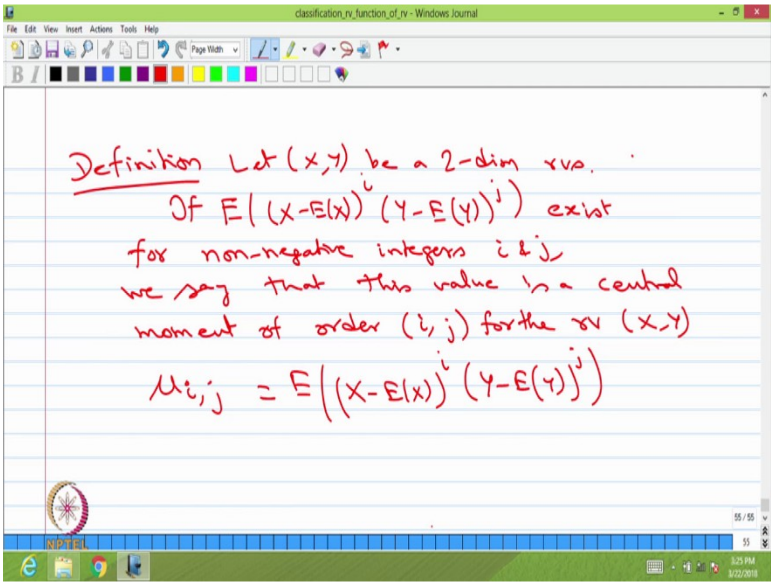


Before we go to the various moments let us give a one simple example of how to compute expected value of the function g$(X_1, X_2, \ldots, X_n)$ that is a simple example 1. Let (X, Y, Z) be a three dimensional continuous type random variable with joint probability density function is given by joint probability density function of X, Y, Z that takes a value 8xyz, when x takes a value 0 to 1; y takes a value 0 to 1 and z takes a value 0 to 1. It is nonzero 8xyz in these intervals; 0 otherwise. You can verify whether this is going to be joint probability density function by $\int_0^1 \int_0^1 \int_0^1 8xyzdxdydz$ that is going to be 1. Therefore, this is a joint probability density function of the random variable (X, Y, Z).

The question is find expectation of 5 X - 2 Y + Z, find E(5 X - 2 Y + Z) you can apply the previous definition. So, here the g(X_1, X_2, X_3) that is same as 5 X - 2 Y + Z. Since all are of the continuous type random variable make the $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (5x-2y+z) 8xyzdxdydz$ that is going to be the expected value.

That is E(5 X - 2 Y + Z) that is same as since x, y, z lies between 0 to 1, it is 0 to 1, 0 to 1, $\int_0^1 \int_0^1 \int_0^1 (5x-2y+z) 8xyzdxdydz$, you can do the integration and you can get the value that value is $\frac{8}{3}$. So, this is a very simple example in which we are finding the expectation of function of random vector.

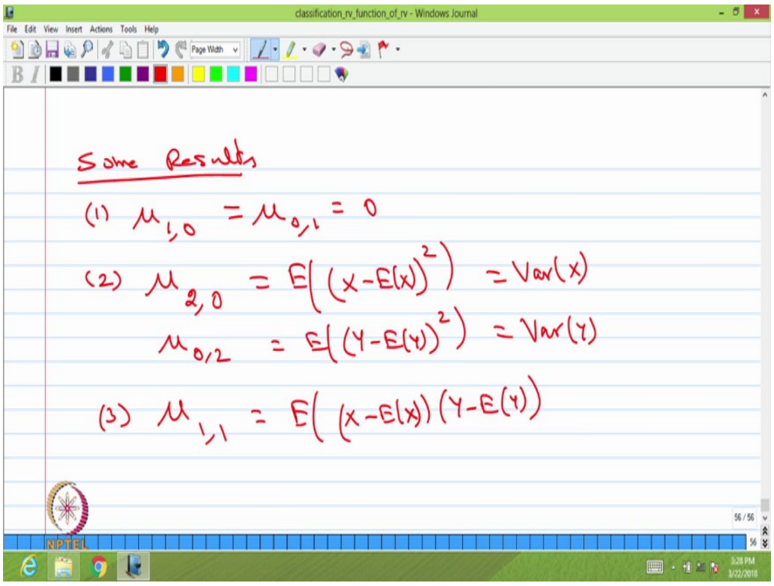
(Refer Slide Time: 13:58)



Now, we will move into the next definition we start with the two dimensional random variable, that (X, Y) be a 2 dimensional random variable, if the $E\left(\left(X-E(X)\right)^{i}\left(Y-E(Y)\right)^{j}\right)$; if this expectation exist. That means, in absolute sense this expectation value is going to be a finite for non negative integers i and j. We say that this value is a central moment of order i, j for the random variable X, Y, and it is written as, it is denoted by $\mu_{i,j}$ that is $E\left(\left(X-E(X)\right)^{i}\left(Y-E(Y)\right)^{j}\right)$.

After the expectation of this function of X and Y exist, we can conclude this value is going to be the central moment of order i, j for the random variable X, Y, that is denoted by $\mu_{i,j}$ here i, j both are non-negative integers. This is a special case of the definition which we have given the expected value of function of random vector, because this is a very special function that is one is $\left(X-E(X)\right)^{i}\left(Y-E(Y)\right)^{j}$.

(Refer Slide Time: 16:42)

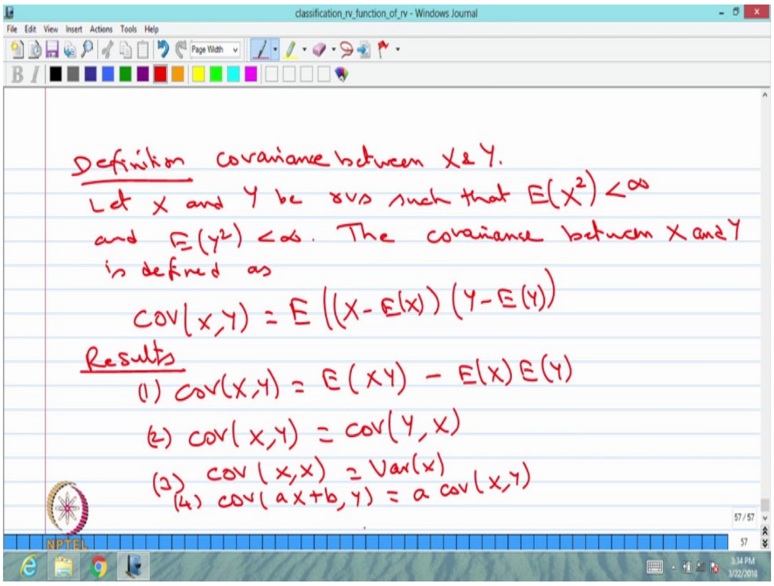


We can have a few results over this definition or remarks or some results. The first result is if you substitute i is equal to 1 and j is equal to 0 or you substitute i is equal to 0 and j is equal to 1 in the definition either you substitute i is equal 0 and j is equal to 1 or i is equal to 1 and j is equal to 0. In both the situation you will get the value is going to be 0.

Second result suppose you put the value i is equal to 2 and j is equal to 0 that is nothing, but the $E\left(\left(X-E\left(X\right)\right)^2\right)$and this is nothing, but variance of the random variable X.

Similarly, if you do $\mu_{0,2}$ that is same as $E\left(\left(Y-E\left(Y\right)\right)^2\right)$that is same as variance of Y. When you substitute i is equal to some integer positive integer and j is equal to some other positive integer, then you will get the central moment of order i comma j for the random variables X and Y or the random vector X, Y. As a special case there is a next result when i is equal to 1 and j is equal to 1 that is nothing, but $E\left(\left(X-E\left(X\right)\right)\left(Y-E\left(Y\right)\right)\right)$ which has the special name.

(Refer Slide Time: 19:09)



So, that I am going to make it as a next definition. That is definition, that is called a covariance between X and Y. Let X and Y be random variables defined on the same probability space such that, such that the second order moment for the random variable X exist and the second order moment for the random variable Y also exist.

Then we can define the covariance between X and the random variable Y that is defined as, the notation is cov(X, Y) that is same as that is $\mu_{1,1}$that is $E\left(\left(X-E\left(X\right)\right)\left(Y-E\left(Y\right)\right)\right)$, where E(X) is the mean of random variable X, E(Y) is the mean of random variable Y. First you multiply $\left(X-E\left(X\right)\right)\left(Y-E\left(Y\right)\right)$ then you find the expectation. So, this is a special case of the first definition, which we have given expected value of function of random vector.

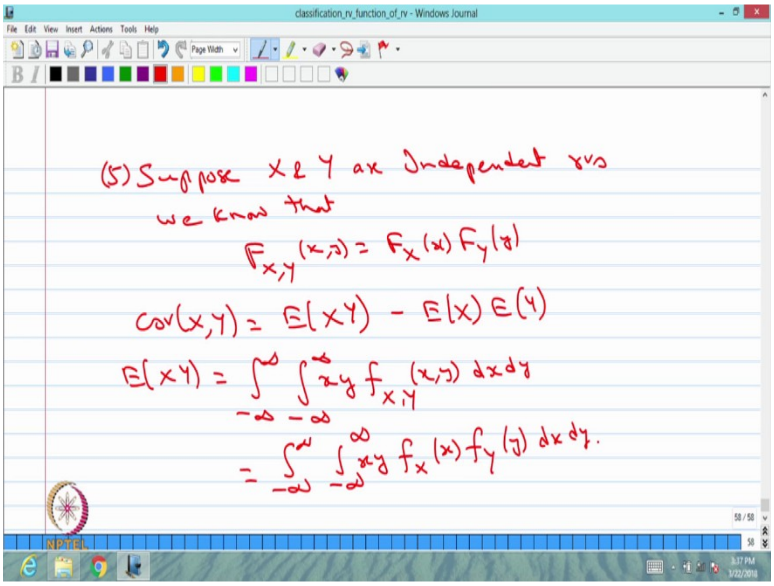
The previous definition, it is i is equal to 1 and j is equal to 1, which has a special name that is called the covariance are between these two random variables. This has the few important results that I make it as the points one by one. Either you can compute E¿¿ or if you do the expansion if you expand this that is XY, X E(Y). Similarly, E(X)Y, then E(X)E(Y) that is same as covariance of any two random variable is same as E(XY)-E(X)E(Y).

The second result if you find out the covariance(X, Y), which is same as covariance(Y, X); that means, by interchanging the role of X and Y, you will get the same value that is provided it exist the covariance(X, Y) which is same as covariance(Y, X).

Third result suppose I compute covariance of X with X itself you substitute you replace Y by X in the above definition; that means, it becomes E¿¿, that is same as E¿¿ that is same as variance of X.

The fourth property that is the covariance(aX + b, Y) that is same as if you expand by using the definition of covariance you will get a.cov(X, Y) and covariance(b, Y) that is going to be 0; covariance of any constant with one random variable that is going to be 0. Therefore, covariance of aX + b when a and b are constant; covariance of aX + b with Y when a and b are constant that is same as a.cov(X, Y).

(Refer Slide Time: 24:45)



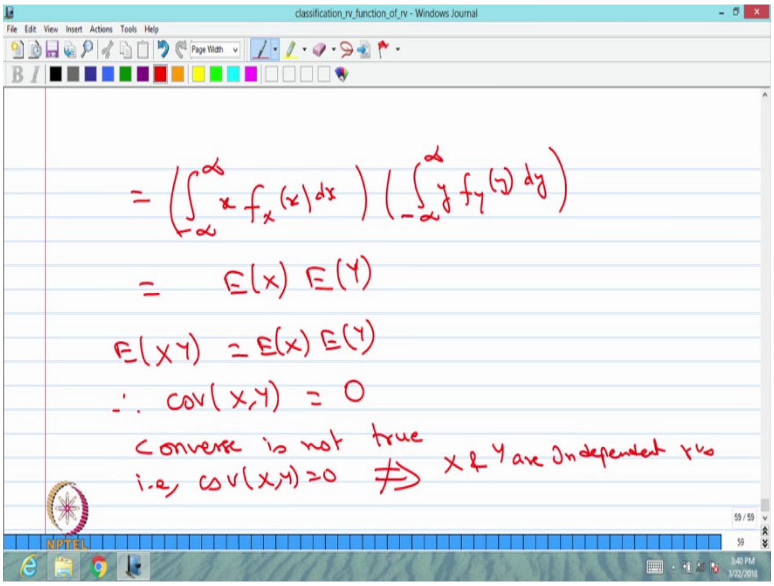
The next result number 5. Suppose the random variable X and Y are independent, suppose the random variable X and Y are independent; we know that the joint CDF is same as the product of two CDF's. If they are discrete type random variable then joint probability mass function is same as the product of probability of mass functions, if they are continuous type random variable then the joint probability density function is same as product of probability density functions of X and Y.

So, here we are computing the covariance(X, Y) that is same as E(XY)-E(X)E(Y). When two random variables are independent, if you compute the E(XY) that is same as suppose I assumed at both the random variables are of the continuous type that is nothing but

$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dxdy$. I am making the assumption both the random variables are of the continuous type.

Similar derivation I can do it for discrete type random variable also. Since these two random variables are independent, I can use this condition that is this is in the CDF and I can use the condition in the joint probability density function. Therefore, this is going to be $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X$¿¿¿.

(Refer Slide Time: 27:16)



Now, the same double integration can be written as $\int_{-\infty}^{\infty} x f_X(x)\,dx \int_{-\infty}^{\infty} y f_Y(y)\,dy$. This is because of the joint probability density function can be written as the product of probability density functions of X and Y. Therefore, I can keep $x f_X(x)$together similarly $y f_Y(y)$ together therefore, the integration becomes product of these two.

We know that that the first integration is expectation of X and the second integration of expectation of Y. That means, we have given the derivation for considering both the random variables are of the continuous type. Even if you do both the random variables are of the discrete type you will get the same conclusion that is E(XY) = E(X)E(Y). Therefore, if two random variables are independent, then the covariance between the random variables X and

Y that is same E(XY) - E(X)E(Y). But just now we got the result E(XY) is same as E(X)E(Y). Therefore, the covariance becomes 0.

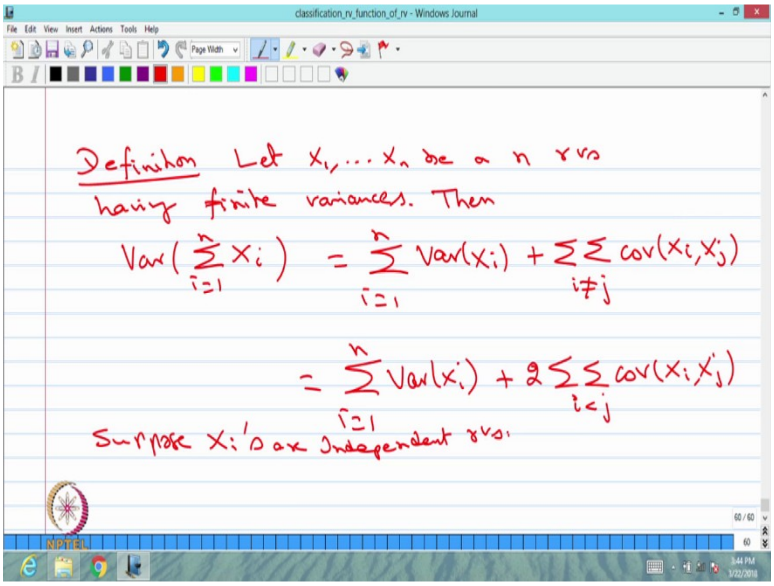
This is the fifth result; that means, if two random variables are independent, the covariance is going to be 0. Converse, that is the if the covariance is; the covariance of two random variables X, Y is 0, that does not imply two random variables are independent; that means, converse is not true. The covariance that is covariance between any two random variables is 0 that does not imply the random variable X and Y are independent random variables; does not imply. Whereas, X and Y are independent random variables, then covariance of random variable X, Y that is going to be 0.

Introduction to Probability Theory and Stochastic Processes
Prof. S Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 36

We are moving into the next important result of finding variance of sum of random variables.

(Refer Slide Time: 00:06)

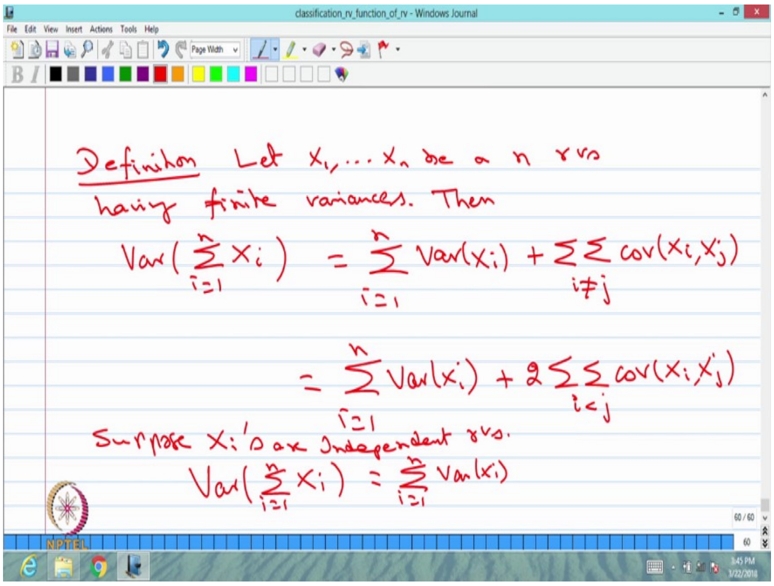


Next definition that is let $X_1, \ldots, X_n$ be n random variables having finite variances; then one can define variance($\sum_{i=1}^{n} X_i$¿¿. Let $X_1, X_2, \ldots, X_n$ be n random variables having finite variances; that means, first and second order moment exist and it is finite.

Then we are defining the sum of random variable that is going to be $\sum_{i=1}^{n} Var(X$¿¿$i) + \sum_{i \neq j} \sum Cov(X_i, X_j)$¿. That is a variance of sum of random variable is the variance of individual random variable with summation plus covariance of any two distinct random variable that is going to be the variance of sum of random variable.

This can be rewritten that is $\sum_{i=1}^{n} Var(X$¿¿$i)$.¿

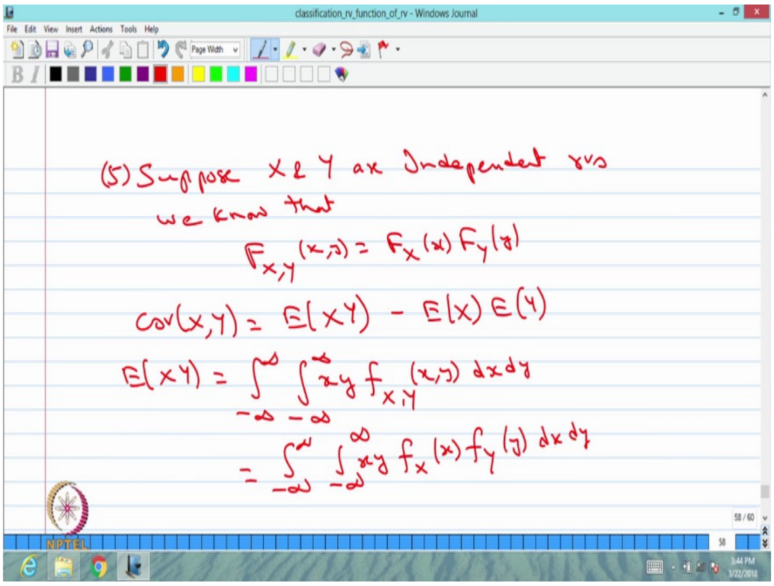
(Refer Slide Time: 02:25)



Since using the result number two, covariance(X, Y) that is same as covariance(Y, X); that means, the second summation can be rewritten in the form of 2 $\sum\sum_{i<j} Cov(X_i, X_j)$; now the condition is instead of i ≠ j, I can write i < j because we have put the 2 times. So, both the statements are one and the same.

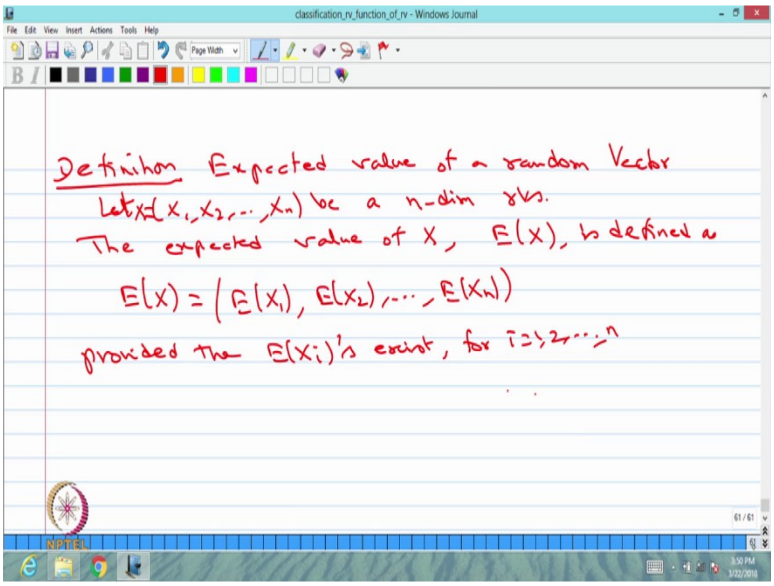
Here also we can go for one special case, suppose X_i's are independent random variables independent, when I say independent random variables; that means, they are mutually independent random variable. In that case we can use the previous result that is result number 5.

(Refer Slide Time: 03:35)



When any two random variables are independent then the covariance is going to be 0. Therefore, the second the summation whole thing; it will not come. Therefore, var($\sum_{i=1}^{n} X_i ¿¿=\sum_{i=1}^{n} Var(X ¿¿ i)¿$when X_i's are mutually independent random variables.

(Refer Slide Time: 04:18)

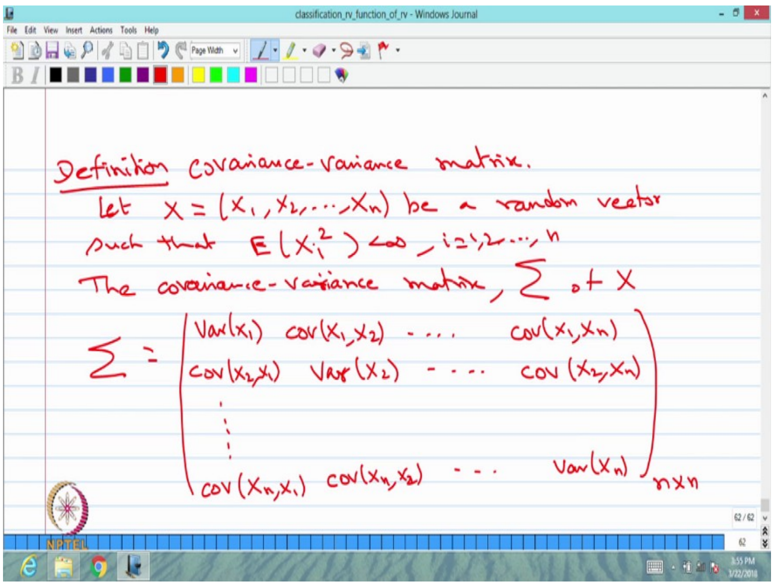


Now, we will give the definition of expected value of a random vector; we have already given the definition of expected value of function of random vector. Now we are going to give the definition of expected value of a random vector. Let $X_1, X_2, \ldots, X_n$ be n dimensional random variables.

The expected value of; let me denote this random vector in the form of X, let me denote $(X_1, X_2, \ldots, X_n)$ random vector with a notation X. So, I am going to define the expected value of the random vector X, which is denoted by E(X) is defined as; it is defined as; expectation of X is nothing but since $(X_1, X_2, \ldots, X_n)$ are vector the expected value of X is also going to be vector, whose elements are E(X_1¿$, E(X_2), \ldots, E(X$¿¿$n)$¿.

We are finding the expected value of a random vector therefore, that is also going to be a vector whose elements are expected value of individual random variables; provided the $E(X_i)'s$ exist for i is equal to 1, 2,…,n. So, as long as individual expectation exists one can define expected value of a random vector with the elements is $X_1, X_2, \ldots, X_n$. In the same way we are going to create a matrix whose elements are variance and the covariance between any two random variable that is a next definition.

(Refer Slide Time: 07:32)



The next definition is covariance variance matrix. Let X it is a random vector for the random variables $X_1, X_2, \ldots, X_n$be a random vector either you can say random vector or n dimensional random variables or random vector with the n random variables. Such that the $E(X_i^2)$ that is a finite for the random variable 1, 2 so on till n.

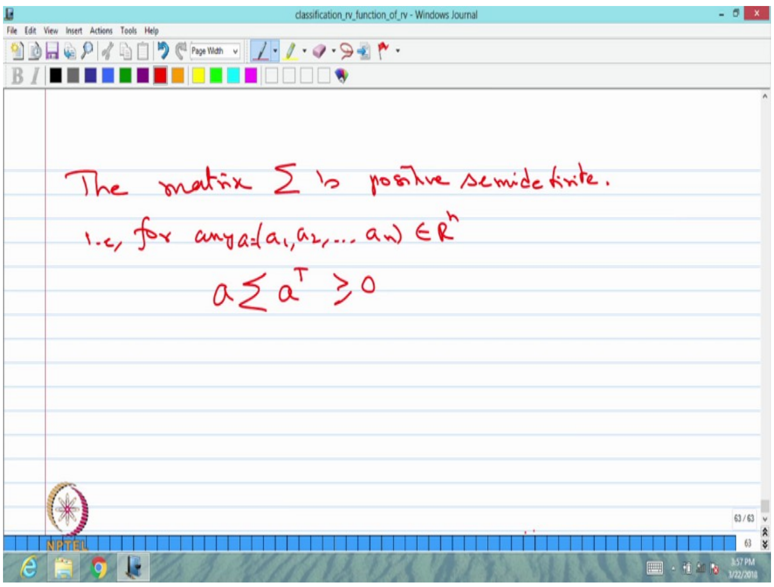
As long as the second order once exist, one can go for, one can go for defining covariance variance matrix. Then the covariance variance matrix that is denoted by in the big summation notation of the random variable X. That is defined as it is in the big summation notation it is a matrix whose elements are; the first element is var(X_1) and the first row second element it is a cov(X_1, X_2) like that so on the first row the last element that is cov(X_1, X_n).

Now, coming to the second row; second row is cov(X_2, X_1), second row second element that is diagonal element that is $var(X_2)$; like that you can keep writing. The last element in the second row, that is cov(X_2, X_n). Like that you can fill up the last row with the first column that is cov(X_n, X_1). The last row second column that is cov(X_n, X_2) so on the last row last element that is a diagonal element that is var(X_n).

So, this matrix is n cross n order; we are creating a covariance variance matrix for n dimensional random variable. Therefore, this matrix is always n cross n whose diagonal elements are covariance of individual random variables and the other elements are cov(X_i, X_j) for the i^{th} row and j^{th} column.

By using the property of covariance(X, Y) that is same as covariance(Y, X), you can conclude that this matrix is a symmetric matrix. All the diagonal elements are the variance that is nothing but cov(X_i, X_i); therefore, it becomes a var(X_i).

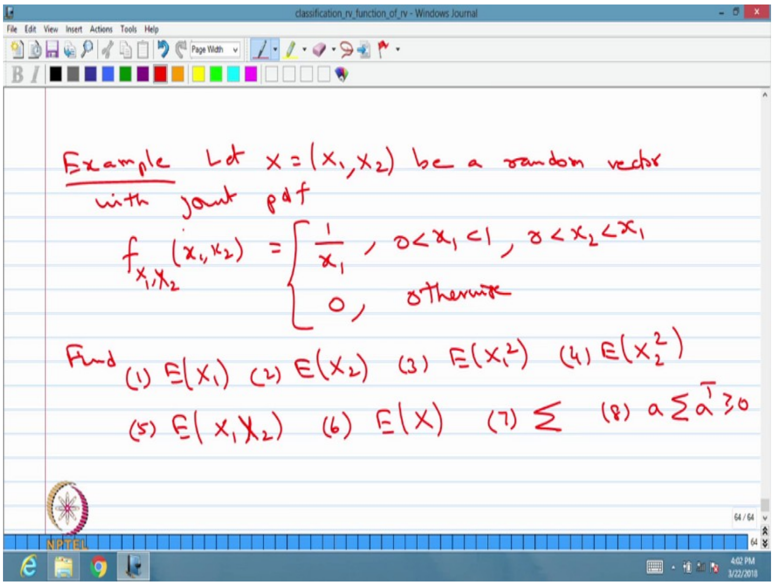
(Refer Slide Time: 11:53)



So, this matrix has the very important property, i.e., the matrix is positive semi definite. That means, for any $(a_1, a_2, \ldots, a_n)$ belonging to R^n the vector. Suppose I denote this as the vector a, suppose I denote this as the vector a whose elements are $(a_1, a_2, \ldots, a_n)$; belonging to R^n, $a\Sigma\, a^T \geq 0$. That is called the matrix is positive semi definite, the covariance variance matrix of random vector is always positive semi definite.

We will give a one simple example how to compute the expected value of a random vector and covariance variance matrix.

(Refer Slide Time: 13:23)

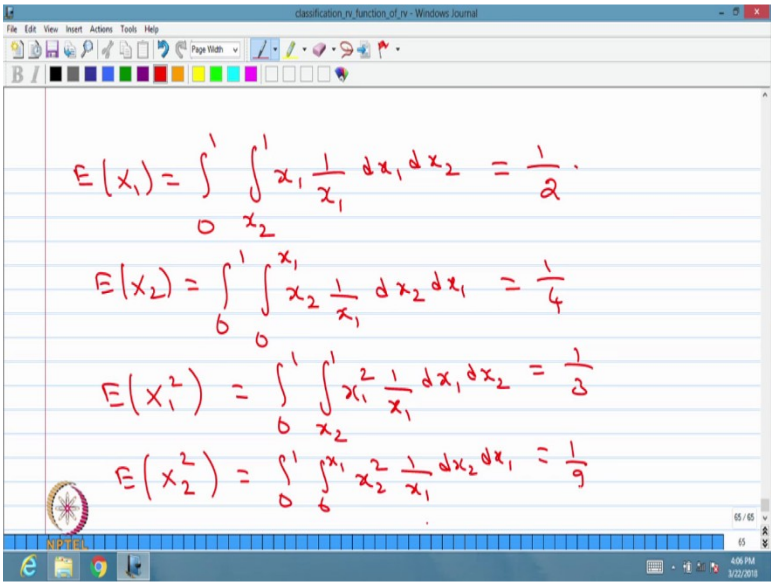


Example let X is the vector whose elements are (X_1, X_2); be a random vector with joint probability density function is given by; that means, both the random variables $X_1 \wedge X_2$ are continuous type random variable. Therefore, we are defining the joint probability density function of X_1, X_2; that is $1/x_1$; when x_1 takes a value 0 to 1 whereas x_2 takes the value 0 to x_1; 0 otherwise. So, this is the joint probability density function of the random vector (X_1, X_2). You can verify double integration of a joint probability density function has to be 1.

Let us find E(X_1), then we will go for finding E(X_2), then we can go for finding E(X_1^2), then you can go for E(X_2^2), then we can go for E($X_1 X_2$), then we can go for expected value of the vector X. We can find out what is the covariance variance matrix then we can go for verifying whether it satisfies $a\Sigma\, a^T \geq 0$.

The last one is the, verify $a\Sigma a^T \geq 0$. So, all those things we can do one by one.

(Refer Slide Time: 16:22)



First start with the E(X_1). E(X_1), you can find the probability density function of $X_1$from the joint probability density function, then you can go for the expectation or you can use the first definition that is expected value of function of a random vector. So, we are going to use that that is same as $\int_0^1 \int_{x_2}^1 x_1 \frac{1}{x_1} d x_1 d x_2$. You see that the joint probability density function is $1/x_1$; when $x_1$ranges from 0 to 1 whereas x_2 ranges from 0 to x_1.

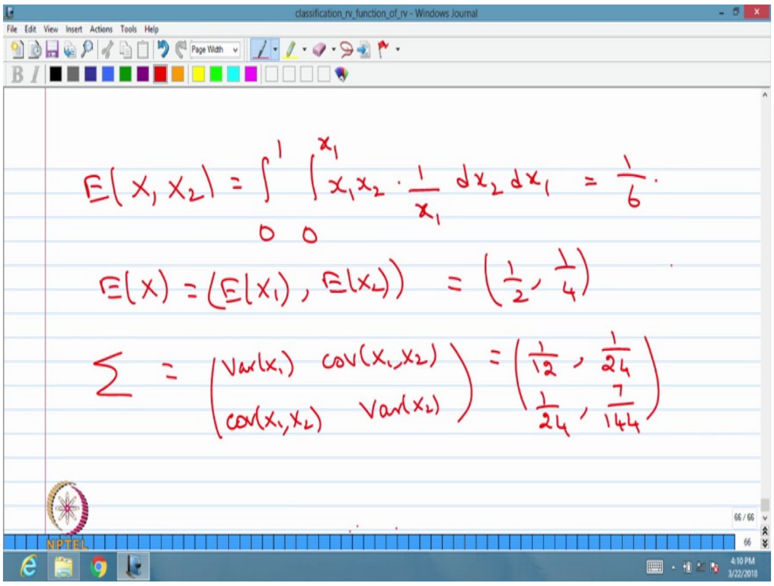
Therefore, E(X_1) = $\int_0^1 \int_{x_2}^1 x_1 \frac{1}{x_1} d x_1 d x_2$; $1/x_1$ is a joint probability density function of X_1, X_2. If you do the simplification we will get the answer 1/2; similarly you can go for finding

$$E(X_2) = \int_0^1 \int_0^{x_1} x_2 \frac{1}{x_1} d x_2 d x_1.$$

Again; I am using the expected value of function of a random vector; this is same as 1/4. Similarly, we can go for E(X_1^2). That is same as; same method what I have done it for E(X_1) that is $\int_0^1 \int_{x_2}^1 x_1^2 \frac{1}{x_1} d x_1 d x_2$or you can use the change of integration and you can change the order of integration still you can go for it and you can get the answer that is 1/3.

Similarly, you can go for $E\left(X_2^2\right) that is \int_0^1 \int_0^{x_1} x_2^2 \frac{1}{x_1} d x_2 d x_1$; I am using the same definition again, that is same as 1/9. So, till now we have got E(X_1), E(X_2), E(X_1^2), E(X_2^2).

(Refer Slide Time: 19:56)



Now, we will go for finding E($X_1 X_2$); the same technique that is E($X_1 X_2$) that is same as

$\int_0^1 \int_0^{x_1} x_1 x_2 \frac{1}{x_1} d x_2 d x_1$. If you do the simplification you will get the answer that is 1/6.

So, now we are going for the next result that is expectation of a random vector; that is a vector whose elements are E(X_1), E(X_2). This is a vector that is same as already we got the result E(X_1) = 1/2 , E(X_2) = 1/4 ; therefore, this vector is 1/2, 1/4.

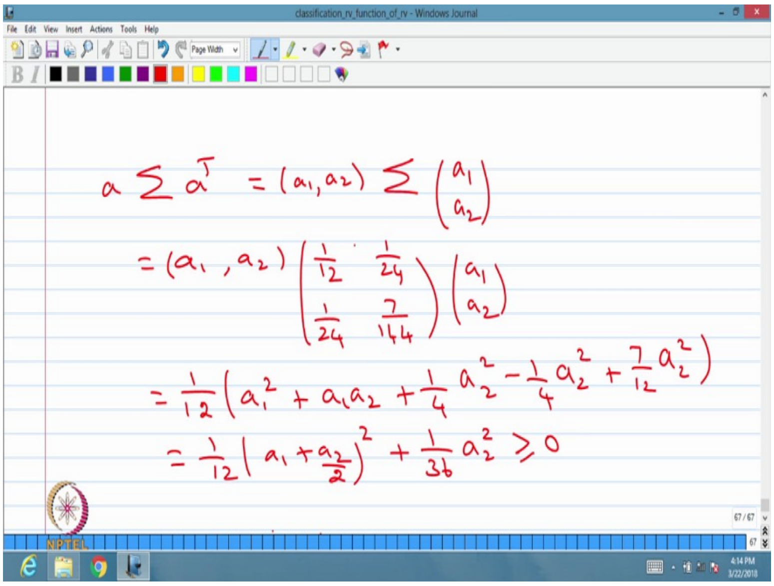
The next is finding the covariance variance matrix, since we have only 2 random variables which is going to be 2 cross 2 whose elements are $var\left(X_1\right)$, cov$\left(X_1, X_2\right)$, cov$\left(X_2, X_1\right)$; or $X_1 with X_2$ both are one and the same and var$\left(X_2\right)$. That is same as we got E(X_1) and E(X_2). So, the var(X_1) = E(X_1^2) - $E\left(X_1\right)^2$. So, if you do the simplification, you will get the answer 1/12.

To find the cov$\left(X_1, X_2\right)$you need E($X_1 X_2$) and E(X_1¿∧E¿¿). So all three we got it. So, substitute the values, that is 1/6 – 1/2 into 1/4. So, you simplify you will get the answer that is

1/24. Since cov(X_1, X_2) $is\, 1/24$ that is again cov(X_2, X_1)that is same as cov(X_1, X_2) that is 1/24. var(X_2) = E(X_2^2) - $E(X_2)^2$. So, you do the simplification you will get 7/144. So, this is the covariance variance matrix for the random vector X_1, X_2.

Now, we are going to verify whether this covariance variance matrix satisfies the condition $a\Sigma\, a^T \geq 0$.

(Refer Slide Time: 23:02)



Let us compute; so here we will go for a with the two elements, (a_1, a_2) and the matrix transpose that is (a_1, a_2) . So, substitute the, that is $(a ¿¿ 1, a_2)\begin{pmatrix} 1/12 & 1/24 \\ 1/24 & 7/144 \end{pmatrix}\binom{a_1}{a_2}¿$.

Do the simplification, do the simplification first you will get $\frac{1}{12}\left(a_1^2 + a_1\, a_2 + \frac{1}{4} a_2^2 - \frac{1}{4} a_2^2 + \frac{7}{12} a_2^2\right)$ and this is same as $\frac{1}{12}\left(a_1 + \frac{a_2}{2}\right)^2 + \frac{1}{36} a_2^2 \geq 0$. Therefore, we are concluding this particular covariance variance matrix also positive semi definite.

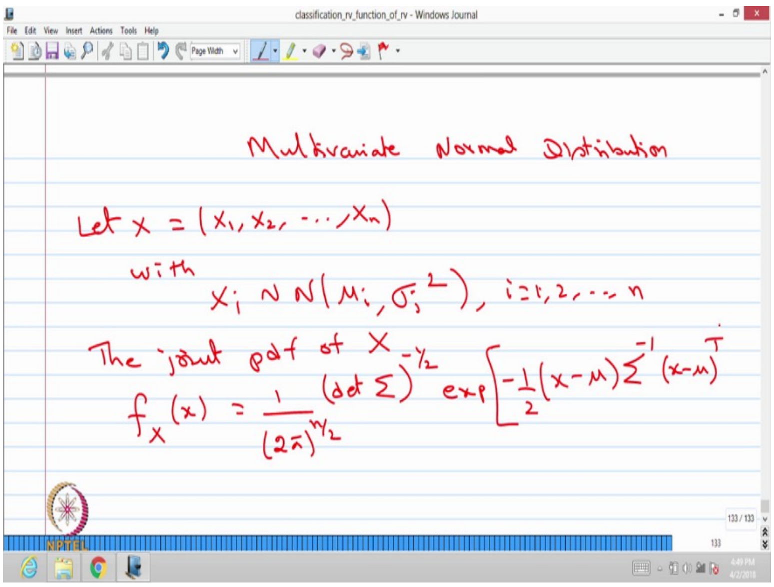
Introduction to Probability Theory and Stochastic Processes

Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 37

So, we have discussed covariance variance matrix then we have discussed as an example, we have discussed the discrete type random variable. Now we are going to discuss one continuous type random variables in which we can describe the covariance variance matrix in a nice way. That is one very important multidimensional random variable of continuous type that is called multivariate normal distribution which has a lot of applications.

(Refer Slide Time: 00:35)

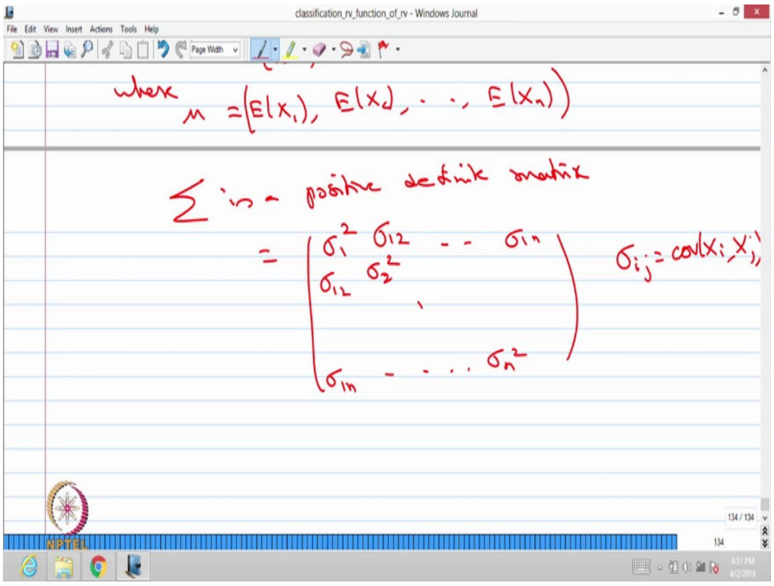


The way the central limit theorem which we are going to discuss later which has lot of applications in the real world problems, the same way the multivariate normal distribution also going to play important roles in many complicated problems in probability.

Let X be a vector whose elements are $(X_1, X_2, \dots, X_n)$. It is a n dimensional random variable of continuous type; with each random variable $X_i^{'}s$ follows a normal distribution with the mean μ_i and the variance σ_i^2; for i is equal to 1 to n. Then we call the random variable X as the multivariate normal distribution. Whenever each random variable is normal distributed at random variable, then the n dimensional random variable is going to be call it as a multivariate normal distributed random variable.

Then we can define the joint probability density function, the joint probability density function of the random vector capital X whose elements are $X_1, X_2, \ldots, X_n$; is given by X is the vector that is $\frac{1}{(2\pi)^{\frac{n}{2}}}(\det \Sigma)^{-1/2} e^{\frac{-1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T}$.

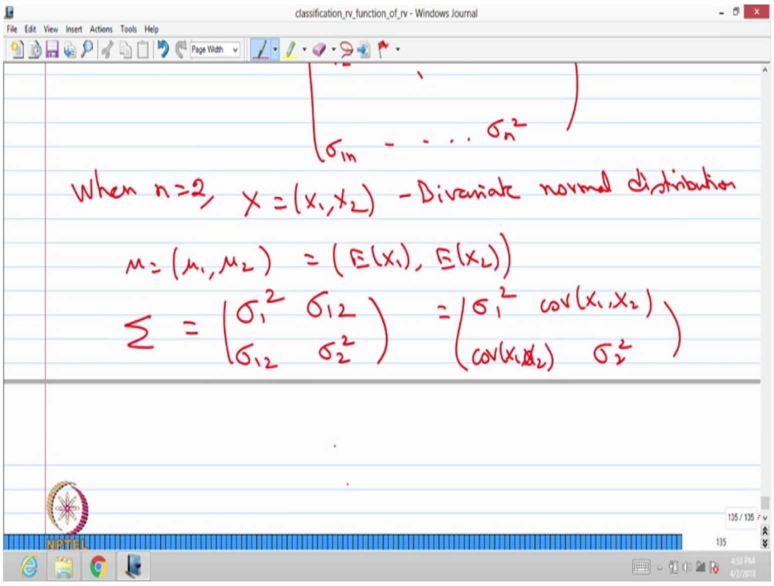
(Refer Slide Time: 03:09)



Where the μ is a vector. that is expectation of the vector; that means whose element are $(E(X_1), E(X_2), \ldots, E(X_n))$. μ is the vector whose elements are the individual expectations. And Σ is a positive definite matrix. It is the covariance variance matrix of the n dimensional random vector $(X¿¿1, X_2, \ldots, X_n)¿$. That means, whose elements are the diagonal elements are variance and off diagonal elements are covariance between any two random variables which is denoted by σ_{12}. Since σ_{12} is same as σ_{21}; so, both we write it as a σ_{12} and the last element is σ_{1n}.

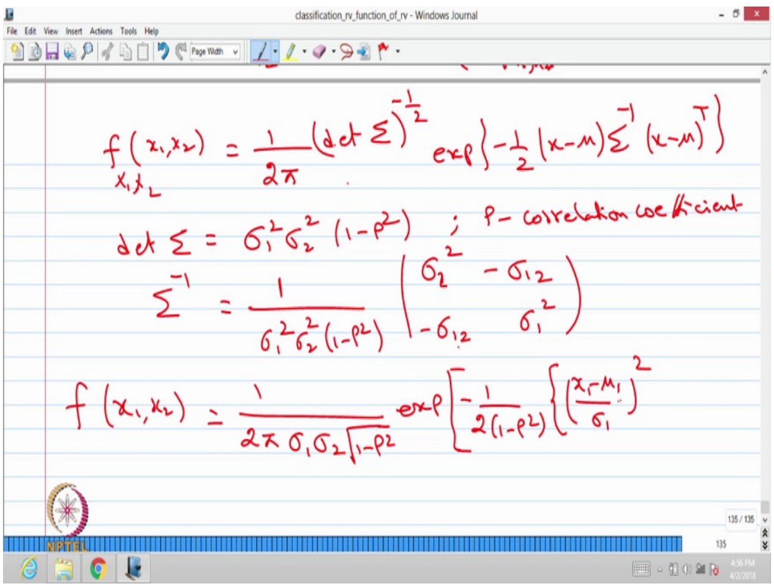
Here the first element is σ_{1n}, where σ_{ij} is nothing but the $cov(X_i, X_j)$. When i and j are same then it becomes var(X_i); variance of the random variable X_i. So, the joint probability density function can be written in the form where μ is the vector and Σ is the matrix and similarly X is also vector.

(Refer Slide Time: 05:18)



For example, when n is equal to 2, we call it as a bivariate X as elements $(X ¿ ¿ 1, X_2) ¿$ that is called a bivariate normal distributed random variable, we call it as a bivariate normal distribution. In that case the mu is going to be (μ_1, μ_2) where $\mu_1 = E(X_1)$ and $\mu_2 = E(X_2)$. And Σ that is covariance variance matrix is nothing but var$(X ¿ ¿ 1) ¿$, cov $(X ¿ ¿ 1, X_2), cov(X ¿ ¿ 1, X_2), var(X ¿ ¿ 2) ¿ ¿ ¿$; that is σ_1^2, cov $(X ¿ ¿ 1, X_2), cov(X ¿ ¿ 1, X_2), var(X ¿ ¿ 2) ¿ ¿ ¿$.

(Refer Slide Time: 06:43)

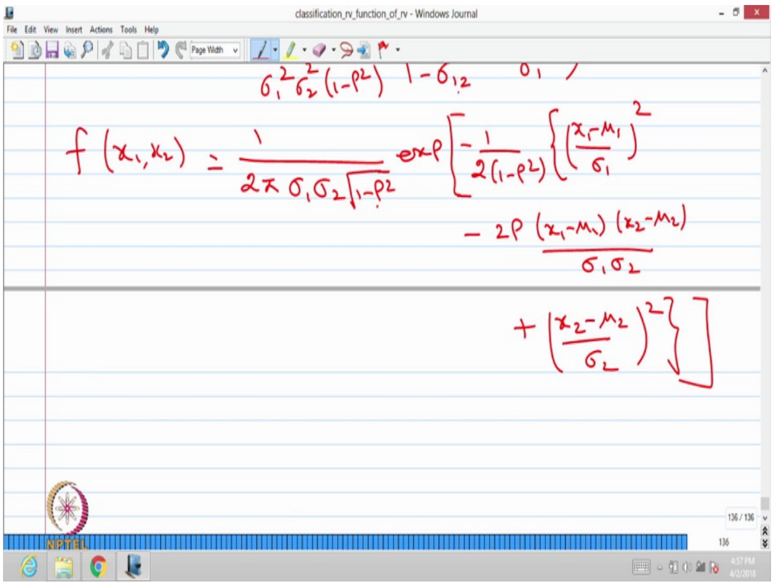


Now, the joint probability density function will be much simplified that is $f_{X_1,X_2}(x_1,x_2)$ that is nothing but $\frac{1}{(2\pi)^{\frac{n}{2}}}$, n is 2 here. Therefore, 2π and $(det\,\Sigma)^{\frac{-1}{2}} e^{\frac{-1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T}$.

So, we find out each quantity separately determinant of summation matrix; if you simplify you will get σ_1^2, σ_2^2, cov$(X$¿¿$1, X_2)$;¿ since we have only two elements we can make it as ρ therefore, it is going to be $(1-\rho^2)$. Similarly, if you find out the $\Sigma^{-1}=\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{pmatrix}\sigma_2^2 & -\sigma_{12}\\ -\sigma_{12} & \sigma_1^2\end{pmatrix}$.

Therefore, you substitute in the joint probability density function therefore, $f(x_1,x_2)$ that is going to be $\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right]$.

(Refer Slide Time: 09:33)

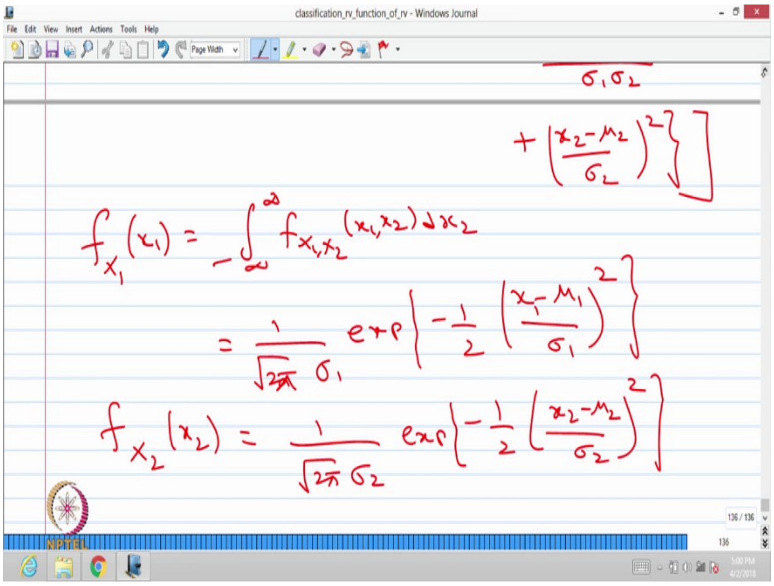


So, this is the joint probability density function of bivariate normal distribution in which each one is a normal distributed with the parameters μ_1, σ_1^2. Here there is another observation, we are not making the assumption of both the random variables are independent. If they are independent then the correlation coefficient becomes 0; then we would have the middle term.

So, this term will vanish; so, you will have first term as well as the third term. Similarly, when the ρ^2 becomes 0 then $\sqrt{1-\rho^2}$would not exist.

So, when they are independent random variable then easily you can write as the product of 2 probability density function of a normal distributed random variable. But immaterial of both the random variables are independent; we can get the marginal distribution of the random variable X_1.

(Refer Slide Time: 11:10)



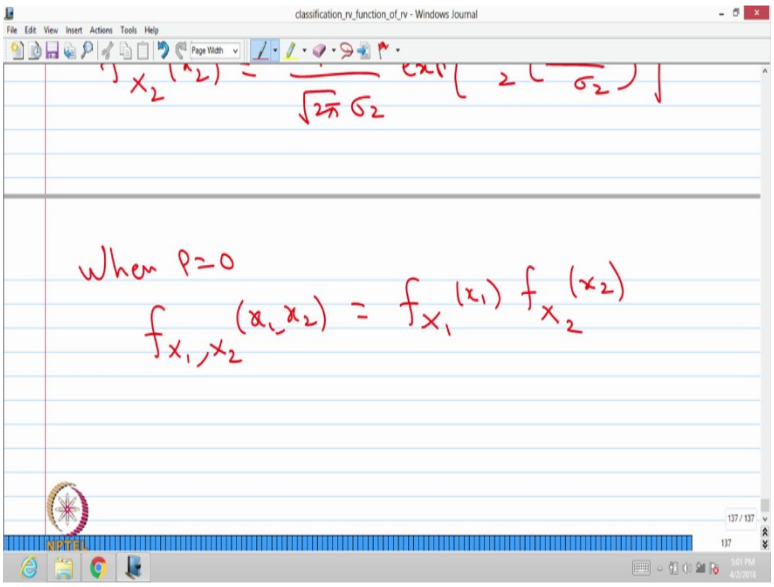
From the joint you can always get the marginal by integration with respect to the other variable; that is $\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\,dx_2$. If you simplify you will get the answer that is $\frac{1}{\sqrt{2\pi}\sigma_1}\exp\{\frac{-1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\}$. So, this is the marginal distribution of the random variable X_1.

Similarly, you can find the marginal distribution of X_2 by $\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\,dx_1$. So, that is going to be $\frac{1}{\sqrt{2\pi}\sigma_2}\exp\{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\}$. By seeing the probability density function, you can make out this is a normal distribution with the mean μ_1 and the variance σ_1^2 for the random variable X_1. Similarly, for the random variable X_2 it is also normal distributed with the mean

$\mu_2 \wedge \sigma_2^2$. Whereas, the joint one is given as this is the joint probability density function of bivariate normal distribution. So, this is a very good example of how the covariance variance matrix play a role.

So, we have discussed earlier in the discrete type, now we are describing the continuous type random variable. As an example for describing the covariance variance matrix, there is another important observation in the joint probability density function of normal distribution. If you substitute ρ = 0 that is a correlation coefficient that is 0, you will get the first term and the third term in that we can come to the conclusion you will get joint probability density function of X_1, X_2 is same as the product of probability density function of X_1, X_2.

(Refer Slide Time: 13:55)



That is when ρ = 0; the joint probability density function of bivariate normal distribution is land up the probability density function of normal distribution in the product.

This is a very important result in the sense when ρ = 0, that is a correlation coefficient is 0 we are getting independent relation. If they are independent then the joint probability density function is going to be the product of probability density functions.

Usually or in general the correlation coefficient is 0 that does not imply they are independent random variable whereas, independent random variable implies the correlation coefficient or covariance between any two random variables going to be 0; the converse is not true in general, but for the normal distribution the converse is also true. That means, the covariance between any two random variables or the correlation coefficient between those two random variables 0 implies those random variables are independent.

So, this is a very important result if and only if condition for the correlation coefficient 0 and independent are going to be satisfied only for normal distributed random variable.

Introduction to Probability Theory and Stochastic Processes

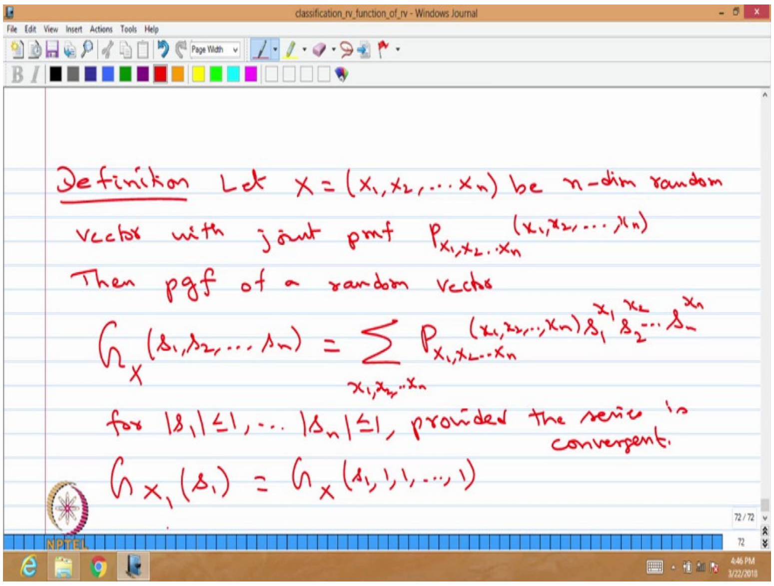
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 38

In Cross Moments, we will discuss the generating function for the random vector. Earlier we have discussed generating functions for the random variable that is a probability generating function for a random variable, then moment generating function for the random variable, then we have discussed the characteristic function for the random variable.

Now, we are going to discuss these generating functions for the random vector that is the first definition.

(Refer Slide Time: 00:40)



Let X be a random vector whose elements are $(X_1, X_2, \dots, X_n)$ be a n dimensional random vector with joint probability mass function that is $P_{X_1, X_2, \dots, X_n}(x_1, \dots, x_n)$. Then one can define the probability generating function of a random vector is defined as we use a notation G_X the X is a vector; whose elements are $s_1, s_2, \dots, s_n$.

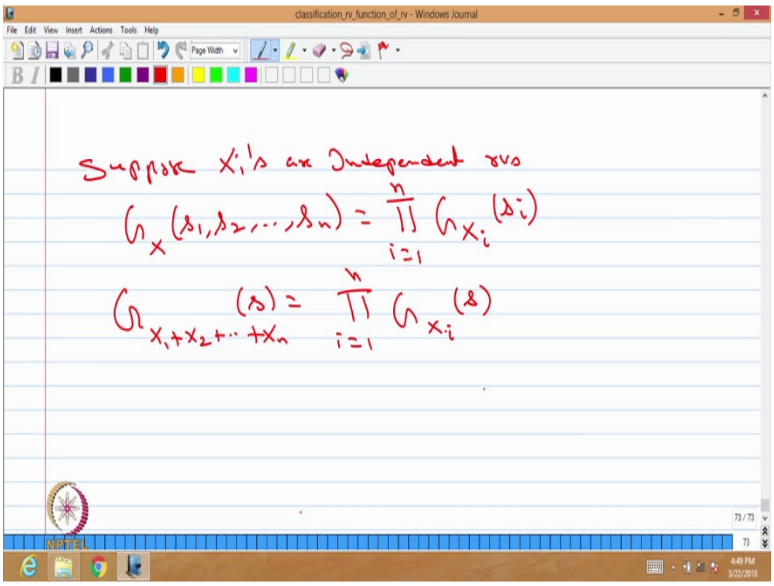
We have defined earlier the probability generating function of a random variable in that case we have $G_X(s)$, now we have a random vector therefore, it is a function of the n variants $s_1, s_2, \ldots, s_n$ that is same as $\sum_{x_1, x_2, \ldots, x_n} P_{X_1, X_2, \ldots, X_n}(x_1, \ldots, x_n) s_1^{x_1} s_2^{x_2} \ldots s_n^{x_n}$.

So, this is going to be a function of $s_1, s_2, \ldots, s_n$. This is called the probability generating function of a random vector. And this is for $¿ s_1 \vee \leq 1, |s_2| \leq 1, \ldots, |s_n| \leq 1$ provided; provided the series the right hand side series is convergent. So, as long as $¿ s_1 \vee \leq 1, |s_2| \leq 1, \ldots, |s_n| \leq 1$, *the* right hand side series converges then that is going to be call it as a probability generating function of a random vector.

From the probability generating function of a random vector, one can get the probability generating function of any one random variable also. That is suppose I want to find out the probability generating function of a random variable X_1; that is same as by substituting the probability generating function of a random vector in s_1 you keep it s_1; whereas, $s_2$onwards you substitute the value 1 till s_n, you will get the probability generating function of a random variable X_1.

Similarly, you can get the probability generating function of other random variables. In this way you can get the probability generating function of fewer random vector also by substituting all other variables as 1.

(Refer Slide Time: 05:14)

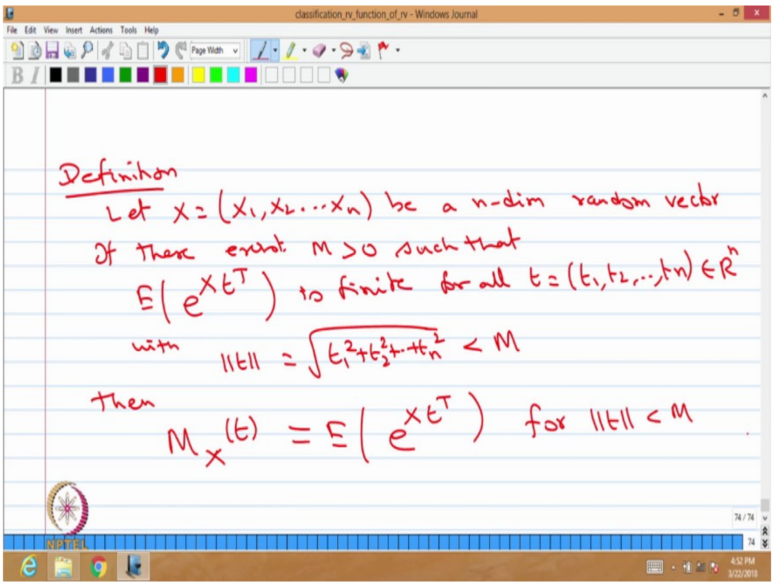


The next result, suppose X_i's are independent random vectors; suppose the random variables are independent, then you can replace the joint probability density function by replacing joint probability mass function by probability mass functions of X_i's; since they are independent random variable. Therefore, this summation will be simplified into the probability generating function of a random vector as a function of $s_1, s_2, \ldots, s_n$. That is going to be $\prod_{i=1}^{n} G_{X_i}(s_i)$.

Because the joint probability mass function can be replaced by the product of probability mass functions of individual random variable. So, the right hand side G_{X_i} that is a probability generating function of the random variable X_i. In this way, sometimes we may be interested to find out the probability generating function of sum of random variables as a function of s; that is same as if you apply the same concept that is same as; since they are independent random variable that is going to be $\prod_{i=1}^{n} G_{X_i}(s)$.

When X_i's are independent random variable the first one is a probability generating function of a random vector, the second expression is probability generating function of sum of random variables that is a one random variable with the variable s that is same as the product of probability generating functions of individual random variable.

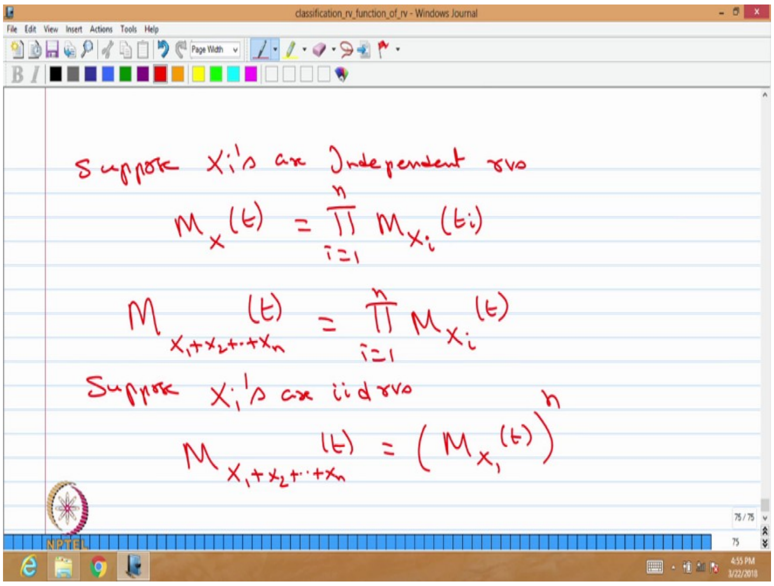
(Refer Slide Time: 07:51)



We will go further second definition that is moment generating function of a random vector. Let X be a vector whose elements are $(X_1, X_2, \ldots, X_n)$ be a n dimensional random vector; if there exist M > 0 such that the $E[e^{Xt^T}]$ that is finite for all t, where t is a vector whose elements are $t_1, t_2, \ldots, t_n$; that is belonging to R^n with the $\|t\| = \sqrt{t_1^2 + t_2^2 + \ldots + t_n^2}$ and that quantity is going to be less than M.

If this condition is satisfied, then one can define the joint moment generating function of a random vector with the notation M_X; X is a vector, t that is also vector whose elements are $t_1, t_2, \ldots, t_n$ that is nothing but $E[e^{Xt^T}]$ and this result is valid only for ||t|| < M. So, this is going to be called as a joint moment generating function of a random vector.

(Refer Slide Time: 10:51)

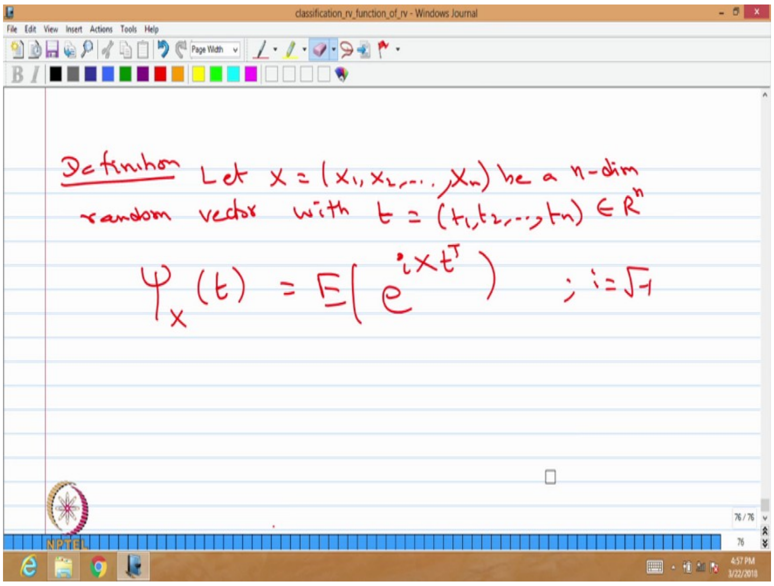


Here also we can go for suppose these random variables are independent; suppose these X_i's are independent random variables. Then it is going to be same as the moment generating function of a vector as a function of vector, that is going to be $\prod_{i=1}^{n} M_{X_i}(t_i)$.; t_i is not a vector, t_i is an element when this random variables are mutually independent.

Similarly, one can go for sum of random variables; similarly one can go for suppose I want to find out the MGF of sum of random variables has a function of t; here the t is a variable not the vector, but I am making the assumption X_i's are independent. Again this is going to be if you substitute in the definition and do the simplification, you will get $\prod_{i=1}^{n} M_{X_i}(t)$with the same variability. Suppose I will make additional condition; suppose X_i's are iid random variables; that means, independent and identically distributed random variables. In that case, the moment generating function of sum of random variables for the identical distributed random variable the MGF is also going to be identical.

Therefore, you find out the ¿¿because all the random variables are identical.

(Refer Slide Time: 13:31)



Now we will move to the third definition that is the joint characteristic function of a random vector. Again, the same way let X be a random vector whose elements are $(X_1, X_2, \ldots, X_n)$ be a n dimensional random vector with the t is a vector whose elements are $t_1, t_2, \ldots, t_n$ that is belonging to R^n. You can define the joint characteristic function of a random vector capital X denoted by φ_X, X is a vector as a function of t; t is also vector that is same as e$E[e^{iXt^T}]$ where i = $\sqrt{-1}$.

Here you do not need to any provided condition similar to the characteristic function of a single dimension random variable. So, this is a characteristic function of a; this is a joint characteristic function of a random vector. So, whatever we discussed the results on independent iid random variables and so on everything will be satisfied for here also.

Therefore, no need to do the simple results what we have done in further moments in getting function because it is just replacing t by it. Therefore, we leave it as it is with the definition.

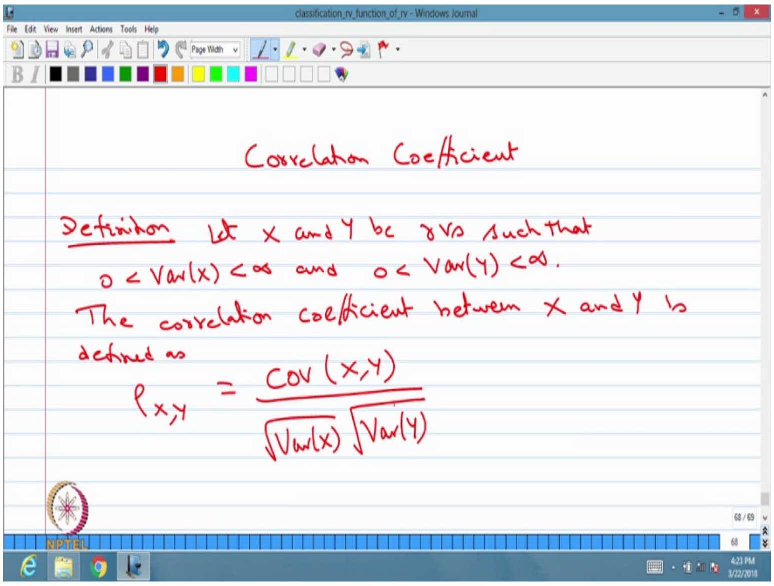
In the later, when we are solving the problems, we are going to use the joint probability generating function of random vector. Similarly, we will be using joint moment generating function of a random vector and also joint characteristic function of a random vector. Therefore, here I stop the stop it with the definitions and the simple results.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 39

Now, we discuss the correlation coefficient.

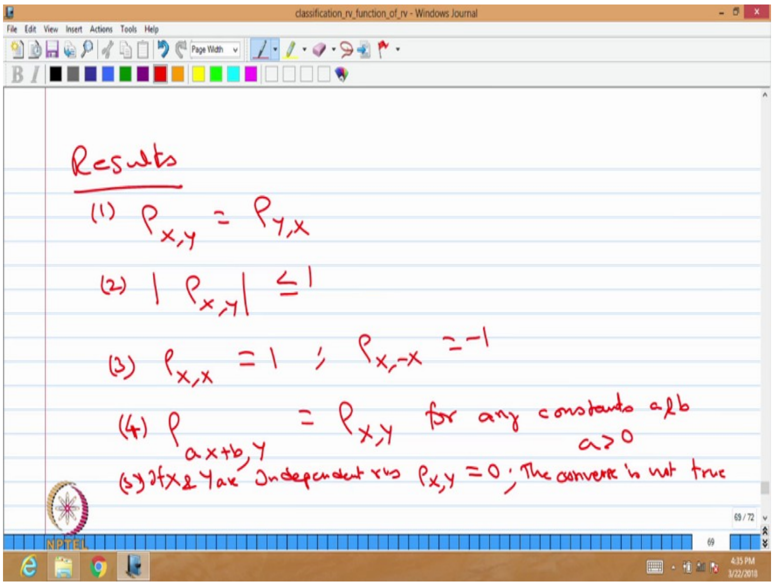
(Refer Slide Time: 00:09)



Let me start with the definition correlation coefficient. Definition, let X and Y be random variables such that 0 < Var(X) < ∞ and 0 < Var(Y) < ∞. That means, the variance exist which is not equal to 0 for both the random variables X and Y.

Then the correlation, then the correlation coefficient between the random variables X and Y is defined as with the notation ρ, you can use the notation suffix, or we do not need; if the suffix is there. That means, these two random variables if the suffix is not then the underlined random variables are two random variables, *then* $\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$. Since, these two comes in the denominator, we are making the restriction the variance has to be strictly greater than 0 for both X and Y.

(Refer Slide Time: 02:38)



This value is going to be call it as a correlation coefficient between the random variables X and Y. I am going to give few results over the correlation coefficient. The first result the $\rho_{X,Y} = \rho_{Y,X}$ by interchanging the role of X and Y it is going to be the same value. Since, the value is $\frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$, $|\rho_{X,Y}| \le 1$; that means, the value is lies between -1 to 1.

Third result if you compute the correlation coefficient between the same random variables; between same random variables. That means, $\rho_{X,X}$ that is nothing but $\frac{var(X)}{var(X)}$. So, you will get the answer 1. That means a correlation coefficient between the same random variable that becomes 1. Suppose you go for $\rho_{X,-X}$, then that quantity is going to be -1.

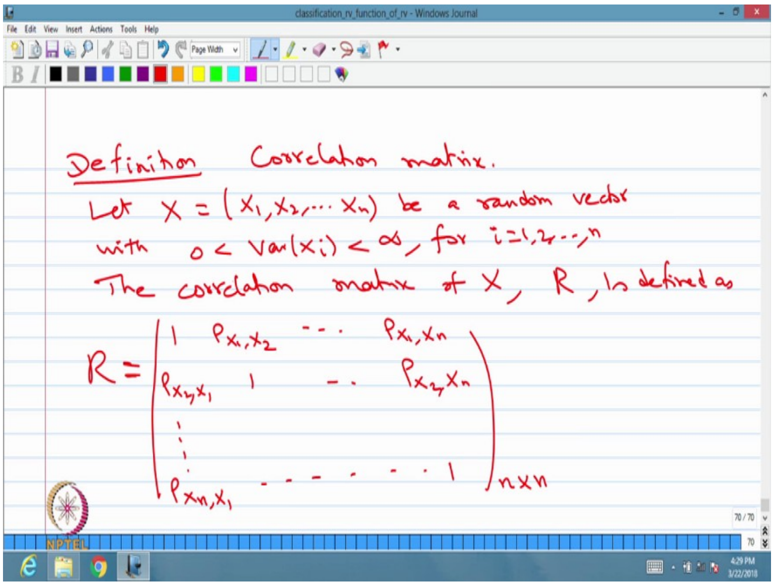
The fourth result, that is the $\rho_{aX+b,Y}$, where a and b are constant with the random variable Y that is going to be if we use the results of covariance between any two random variable, when one random variable is aX + b the other one is Y use that result, you will get it is $\rho_{X,Y}$ for any constants a comma b with a > 0 this can be proved. But we are not giving the proof of this result, but these results will be used again and again.

As a fifth result suppose the random variable X and Y are independent, if random variable X and Y are independent, then we can conclude the correlation coefficient between these two random variables is going to be 0. Because, the correlation coefficient is same as

$\frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$, and, we know that if two random variables are independent the cov(X,Y) is going to be 0. Therefore, the correlation coefficient is also going to be 0, when two random variables are independent.

Whereas, the converse is not true. The converse is not true that means, if the correlation coefficient is 0, for two random variables that does not imply those two random variables are independent.

(Refer Slide Time: 06:32)



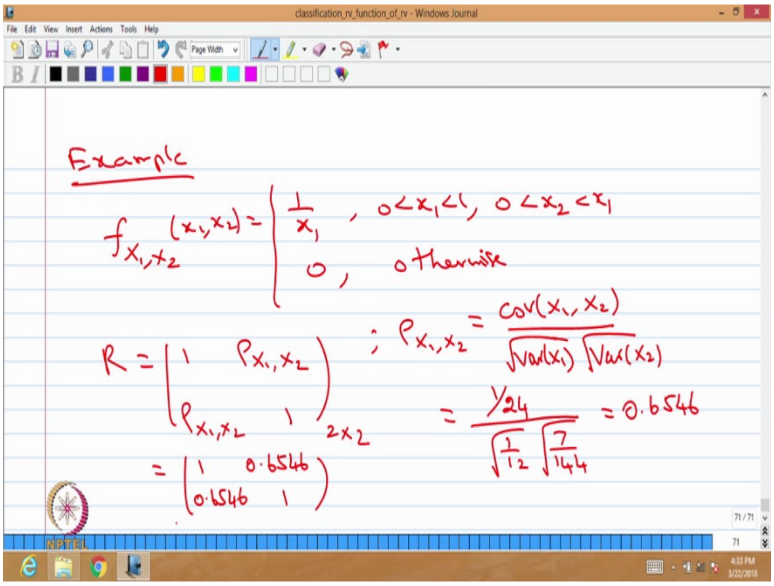
Now, we are going to define the matrix related to the correlation coefficient that is, now we are going to define a matrix which is related to the correlation coefficient that is correlation matrix. Let X is a random vector, whose elements are $X_1, X_2, \ldots, X_n$ be a random vector with individual variance of the random variable has to be, not equal to 0 and it is a finite quantity, for i is equal to 1, 2 and so on, all n random variables.

Then the correlation matrix of the random vector X, which is denoted by the matrix called R that is defined as R. Earlier, we have used the notation sum that is for the covariance variance matrix and here, we are using R for correlation matrix, whose elements are the diagonal element is 1 and, the second element is; first row second element is ρ_{X_1,X_2} and so on. The first row last element that is n-th element is ρ_{X_1,X_n}. Similarly, the second row that is a ρ_{X_2,X_1} and second row second element that is 1 and so on. The second row n-th column that is ρ_{X_2, X_n}, like that you can fill up n rows.

So, the n-th row first column that is ρ_{X_n, X_1}like that the last element will be 1. Again this is a n cross n order matrix, whose diagonal elements are 1 and other elements for the i-th row j-th column that is a ρ_{X_i, X_j}.

We will go for the example, but we will take the example what we have discussed for the covariance variance matrix. We will take the same example and find out the R matrix.

(Refer Slide Time: 10:06)



That is example is same as two dimensional random variable with the joint probability density function is $\frac{1}{x_1}$, when x_1 takes a value 0 to 1 and x_2 takes a value 0 to x_1; 0 otherwise. We have already computed; we have already computed the $E[X_1]$ and $E[X_2]$, $E[X_1^2]$, $E[X_2^2]$, then $E[X_1 X_2]$ using that we got the $\text{cov}(X_1, X_2)$ also and $\text{var}(X_1)$ as well as $\text{var}(X_2)$.

Therefore in the same example, the R matrix is going to be 2 cross 2 matrix; 1 and the second element is ρ_{X_1, X_2} and this is ρ_{X_2, X_1}, that is same as$X_1 \, with \, X_2$, because of symmetry and last element is 1 that is 2 cross 2 matrix. So, we have to compute what is the ρ_{X_1, X_2}.

So, ρ_{X_1, X_2}, that is $\frac{cov(X_1, X_2)}{\sqrt{var(X_1)}\sqrt{var(X_2)}}$. So, we already got the $cov(X_1, X_2)$ that is 1/24. Therefore, it is 1/24 divided by $var(X_1) = 1/$¿12.

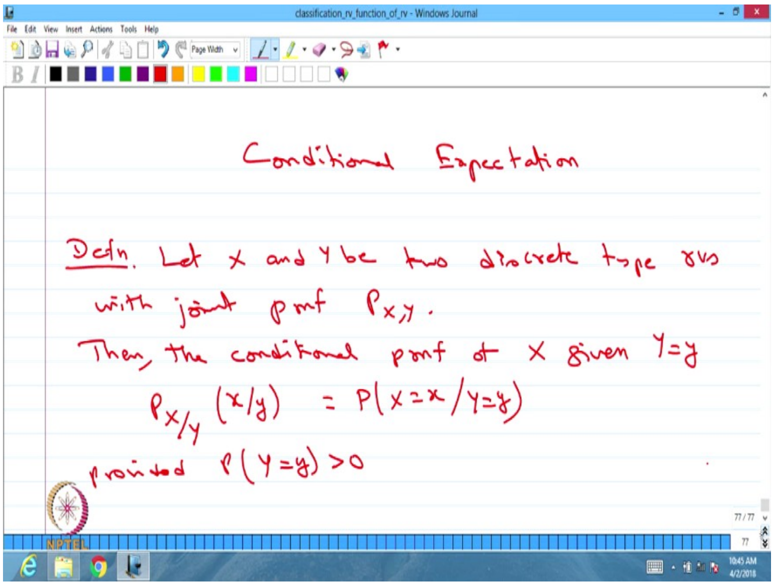
Therefore, it is 1/12, $var(X_2) =$¿7/144. Therefore, it is 7/144, you do the simplification you will get the answer that is 0.6546. Therefore, this R matrix is going to be 1, 0.6546, 06546, 1. That is going to be the correlation matrix for the same example which we have discussed earlier.

Introduction to Probability Theory and Stochastic Processes
Prof. S Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 40

In this lecture we have already discussed moments of functions of several random variables, covariance variance matrix, in the lecture 1 and in the lecture 2, we have discussed the correlation coefficient, in this lecture we are going to discuss about the conditional expectation.

(Refer Slide Time: 00:30)



This is a very important concept in probability, because the way we have discussed the conditional distribution, conditional expectation is also important; that means, after something happen, what is the distribution of the future events, we are finding the probability, then finding the distribution of the various random variables.

The same way we can go for computing the conditional expectation, for that first we should know what is the conditional distribution, then we can find out the expectation of that, is going to be call it as a conditional expectation. That means, first I should define what is a conditional distribution, then followed I can go for finding the expectation of that conditional distribution is a conditional expectation.

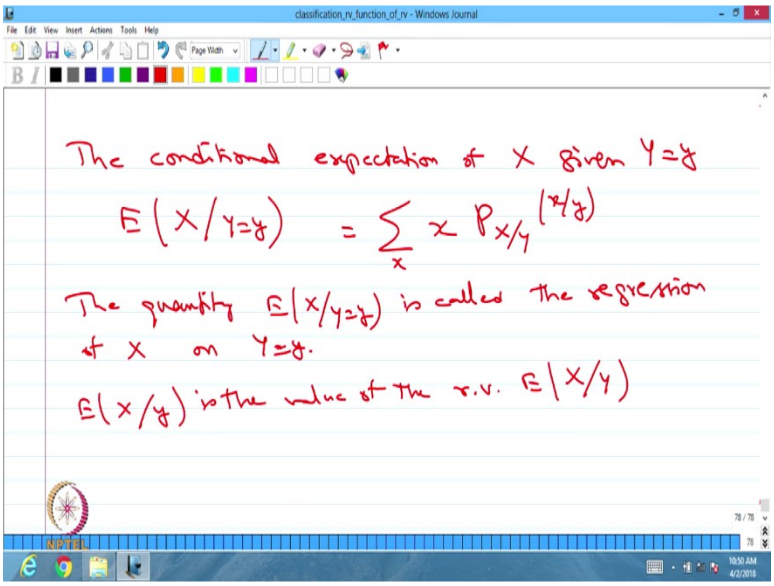
The provided conditional expectation in absolute sense it has been finite, the same thing play here, also to compute the conditional expectation, provided the expectation in absolute sense that is convergent or absolute sense it has a finite value.

So, let me give the definition of conditional expectation for the two dimensional discrete type random variable first, then I will go for the definition of conditional expectation of two dimensional continuous type random variables.

So, definition. First let me give the definition of conditional probability mass function. Let X and Y be two discrete type random variables with joint probability mass function is $P_{X,Y}$, then we have defined the conditional probability mass function of X given the other random variable takes value y is given by $P_{X/Y}(x/y) = P\{X=x/Y=y\}$.

This we have already defined provided, provided the P{Y = y} > 0, this is the conditional probability mass function of the random variable X given Y takes a value y.

(Refer Slide Time: 04:07)



Now, I can go for defining the conditional expectation, conditional expectation of the random variable X given the other random variable takes a value y, that is defined as $E(X/Y=y)$. Whenever we use a word slash; that means, it is a conditional or given that is same as $\sum_x x P_{X/Y}(x/y)$.

So, the definition says the $E(X/Y=y)$ means $\sum_{x} x\,P_{X/Y}(x/y)$ provided this right hand side summation in absolute sense, it is finite value. As long as the right hand side in the absolute sense, it is a finite value without absolute sense that summation quantity, is going to be called it as a conditional expectation of X given Y takes a value y.

There is another name for this, the quantity that is $E(X/Y=y)$, that is called the regression of the random variable X given Y takes a value small y. We will study this in detailed the regression of X given Y takes a value y in the statistic course, but here we are connecting the conditional expectation of X given Y that is nothing but the regression of X given Y takes a value y.

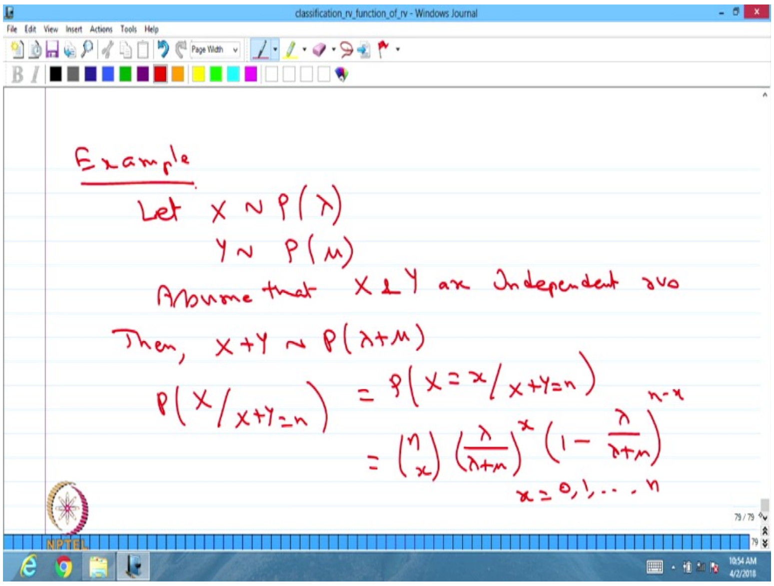
One more observation the conditional expectation X given Y is equal to y is a function of y; is a function of y; that means, the expectation of X given y is the value taken for the values of the different y. Therefore, E(X/y) is the value of the random variable E(X/Y). Since this is going to be a function of y and Y is a random variable.

Therefore, E(X/y) is the function of y that is nothing but the value of the random variable E(X/Y). The quantity E(X/Y=y) is called the regression of X on Y takes a value y. We will study this regression of X on Y takes a value y in detail in the statistic courses, but as far as this probability and stochastic process course is concerned, you can consider this as the this quantity as a regression of X on Y takes the value y.

And the other observation is the E(X/Y=y) is the function of y. And since Y is a random variable it is taking different values of the y therefore, E(X/y) is the value of the random variable E(X/Y). Note that expectation of a random variable is a constant whereas, the conditional expectation of one random variable given another random variable take some value, that is value of the random variable expectation of X given the random variable. So, there is a difference between expectation and the conditional expectation.

We will study some results over the conditional expectation, after I go for continuous type random variable definition and some more examples.

(Refer Slide Time: 09:16)

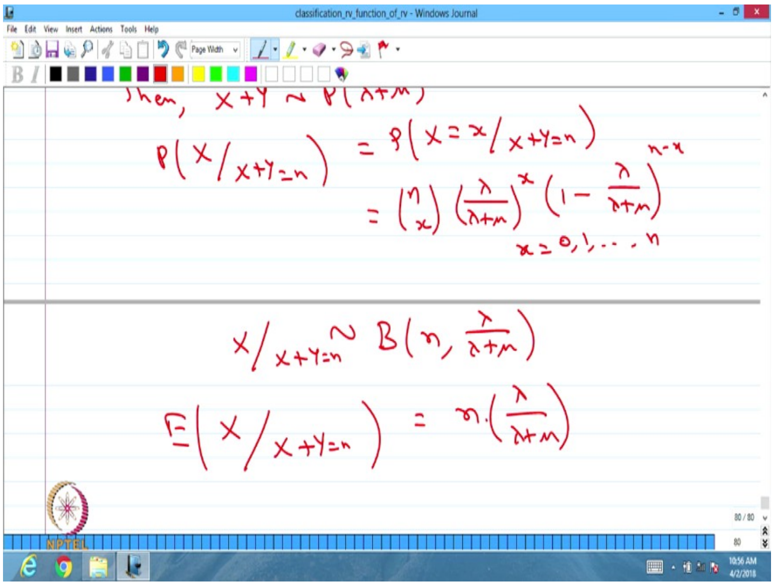


We will go for, example for conditional expectation of two dimensional discrete type random variables. Let X be Poisson distributed random variable with the parameter λ and let Y be again Poisson distributed random variable with the parameter μ, assume that X and Y are independent random variables, we have already proved the sum of two independent Poisson distributed random variable, also going to be Poisson distribution with the parameter is sum.

Also we have already proved the P(X/X+Y =n) that is nothing but P(X = x/ X+Y = n). This we have already proved that follows ${}^{n}C_{x}\left(\frac{\lambda}{\lambda+\mu}\right)^{x}\left(1-\frac{\lambda}{\lambda+\mu}\right)^{n-x}$, where x can take the value 0, 1 and so on till n; this is a conditional distribution of X given X + Y.

Since it is the probability mass function, the conditional probability mass function of X given X + Y that follows a binomial distribution.

(Refer Slide Time: 11:24)

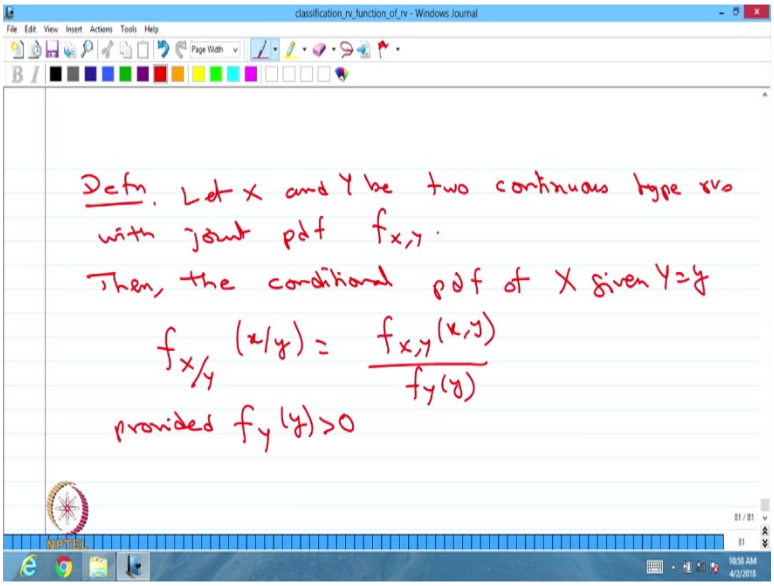


We can easily write X/X+Y = n follows binomial distribution with the parameters n, p; p = $\frac{\lambda}{\lambda+\mu}$.

Our interest is to find out the conditional expectation. So, the conditional expectation of X given X + Y = n. So, here also I can write is equal to n that is same as; since it is a binomial distribution you know that the expectation of binomial distributed random variable is product of the parameters, so, np. Since we know the mean of binomial distribution exist therefore, we are directly writing the E(X/X+Y = n) that is n$\frac{\lambda}{\lambda+\mu}$.

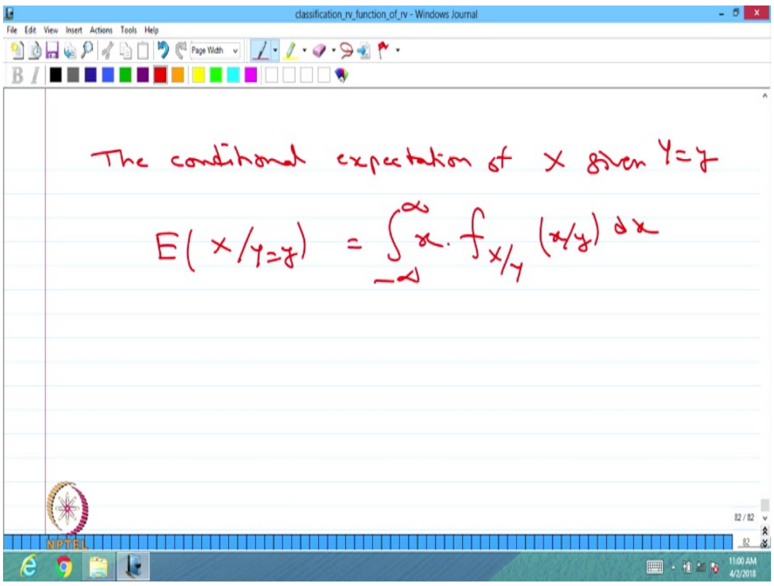
Like that for any two dimensional or n dimensional discrete type random variable one can first find the conditional probability mass function, from there you can find out the conditional expectation. Now, we will go for the conditional expectation for two dimensional continuous type random variables.

(Refer Slide Time: 13:00)



Let X and Y be two continuous type random variables with joint probability density function $f_{X,Y}$, then one can define the conditional probability density function of X given Y takes a value small y is $f_{X/Y}(x/y)=¿\frac{f_{X,Y}(x,y)}{f_Y(y)}$ provided, provided $f_Y(y)>0$. So, this is the conditional probability density function of X given Y takes a value y.

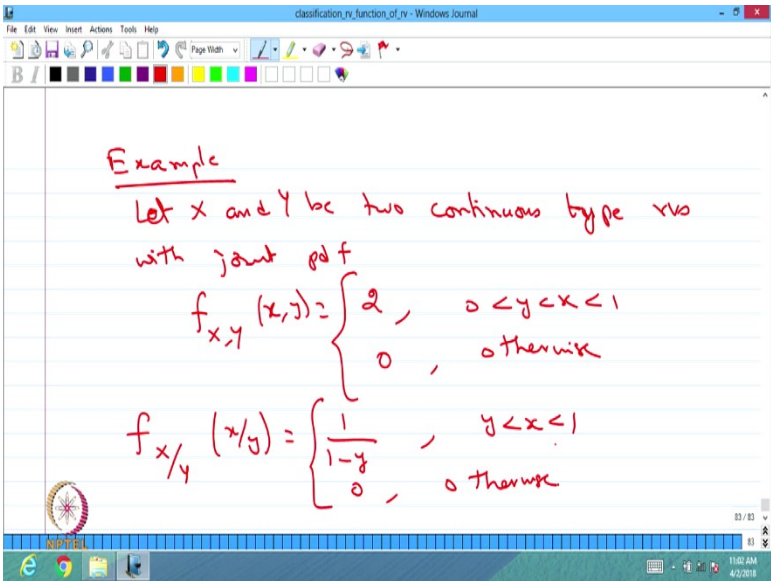
(Refer Slide Time: 14:34)



From here one can define the conditional expectation of the random variable X given Y takes a value y, that is defined as expectation of X given other random variable takes a value y that is nothing but $\int_{-\infty}^{\infty} x f_{X/Y}(x/y)\,dx$; provided the right hand side integration in absolute sense is the finite quantity.

So, this is the E(X/Y=y), when both the random variables are of the continuous type we will go for the simple example.

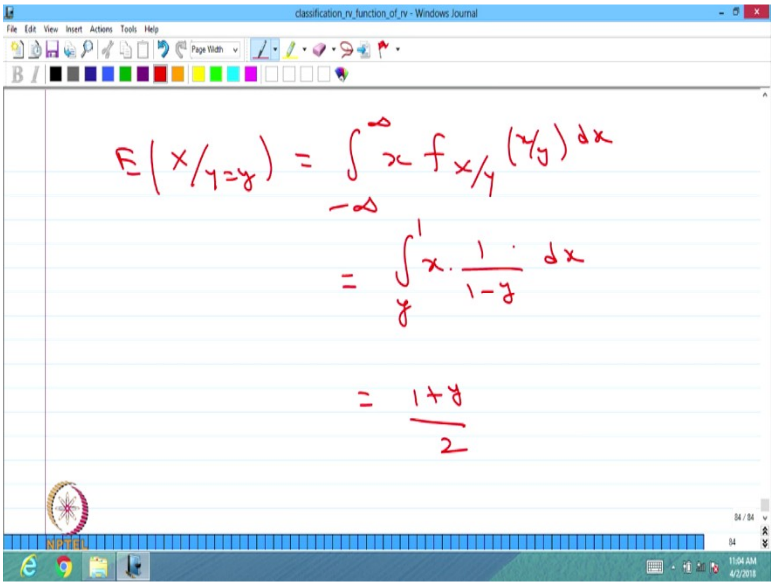
(Refer Slide Time: 15:55)



The example is let X and Y be two continuous type random variables with the joint probability density function that is given by 2, when y lies between 0 to x, x lies between y to 1; otherwise 0. This example we have already discussed when I discussed the conditional distribution.

For this problem we have already got the conditional, we have got the conditional probability density function, that is $\frac{1}{1-y}$, when x takes a value y to 1; otherwise it is 0. So, this is the conditional probability density function of X given Y, here y as we treated as a constant. So, this is a conditional distribution of X given Y.

(Refer Slide Time: 17:49)

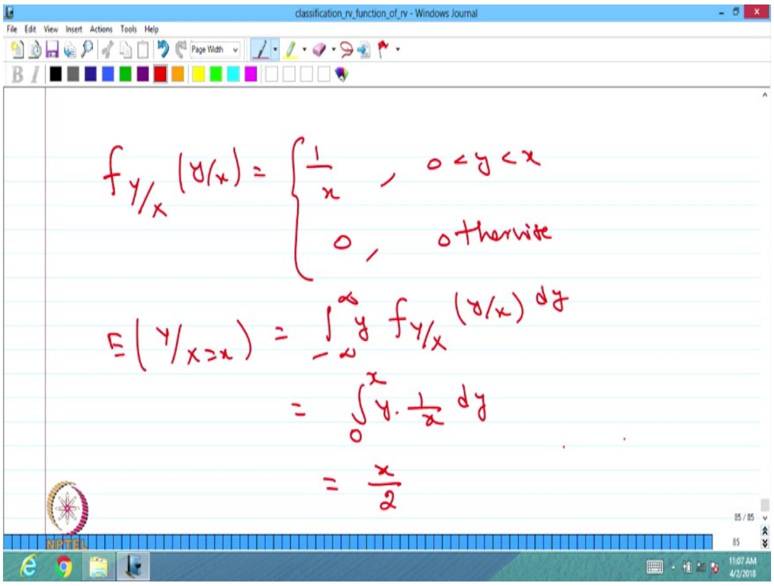


One can find the E(X/Y=y) that is nothing but by definition $\int_{-\infty}^{\infty} x f_{X/Y}(x/y)\,dx$, this is same as we know the $f_{X/Y}$¿) = $\frac{1}{1-y}$ between the interval y to 1. Therefore, it is going to be $\int_{y}^{1} x \frac{1}{1-y}\,dx$, if you do the simplification answer is $\frac{1+y}{2}$.

So, the conditional probability density function is $\frac{1}{1-y}$between the interval y to 1. That means, it is a uniform distribution between the interval y to 1 therefore, the mean of uniform distribution between the interval y to 1, that is $\frac{1+y}{2}$, that is same as the E(X/Y=y).

So, that is $\frac{1+y}{2}$that is going to; either you do by integration simplifying you will get this answer or by observing the conditional distribution is nothing but the uniform distribution.

(Refer Slide Time: 19:37)



Therefore, the conditional expectation is same as the expectation of uniform distribution during the interval y to 1. Similarly, one can get the conditional probability density function of Y given X in the same problem, you can get that is $\frac{1}{x}$ between y lies between 0 to x; otherwise it is 0. So, this is a conditional probability density function of Y given X.

So, from here you can get the E(Y/ X = x), that is same as $\int_{-\infty}^{\infty} y f_{Y/X}(y/x) d y$, by doing the simplification you will get $\int_0^x y \frac{1}{x} dy$.

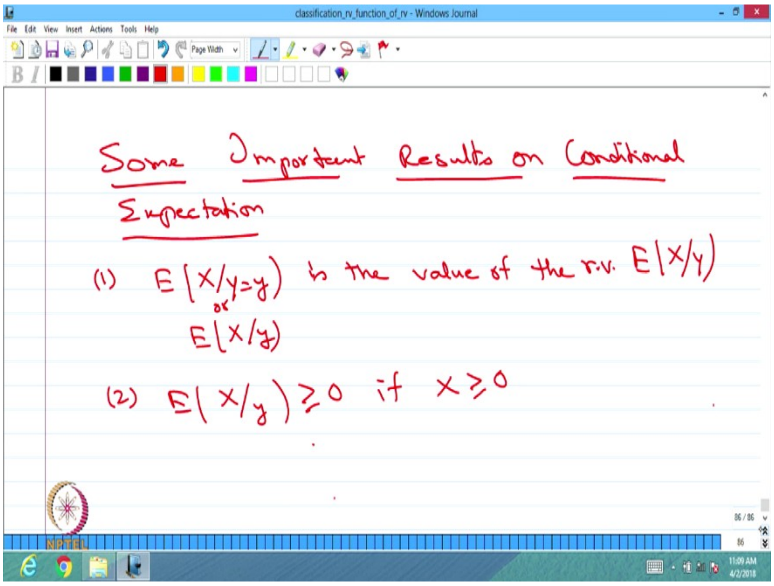
Again if you do the simplification you will get $\frac{x}{2}$, either by simplifying this integration you can get $\frac{x}{2}$, or here also you can observe again the conditional distribution of Y given X that is uniform distribution between the interval 0 to x therefore, conditional expectation of Y given X that is same as expectation of uniform distribution between the interval 0 to x that is $\frac{x}{2}$. So, till now we have discussed the conditional expectation for the two dimensional discrete type random variable with one example, two dimensional continuous type random variables with one example.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 41

Now, we are going to give few important results on the conditional expectation.

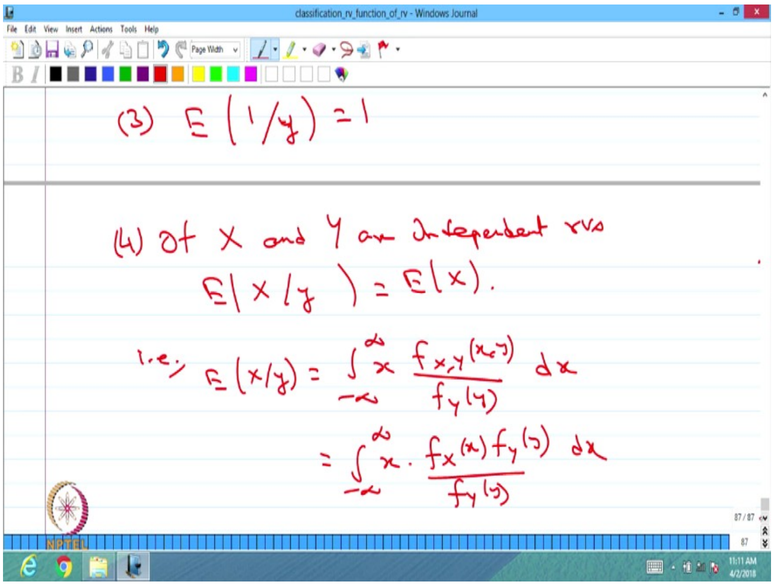
(Refer Slide Time: 00:10)



Some important results on conditional expectation, first important result E(X/Y=y) is the value of the random variable E(X/Y). So, sometimes we write this in the form of E(X/y), either we write like this or we write E(X/y), that is the value of the random variable E(X/Y).

The second result in the $E(X/y) \geq 0$, whenever $X \geq 0$ that means, whenever the random variable takes a values greater or equal to 0, or the $P\{X \geq 0\} = 1$ in that case the $E(X/Y) \geq 0$.

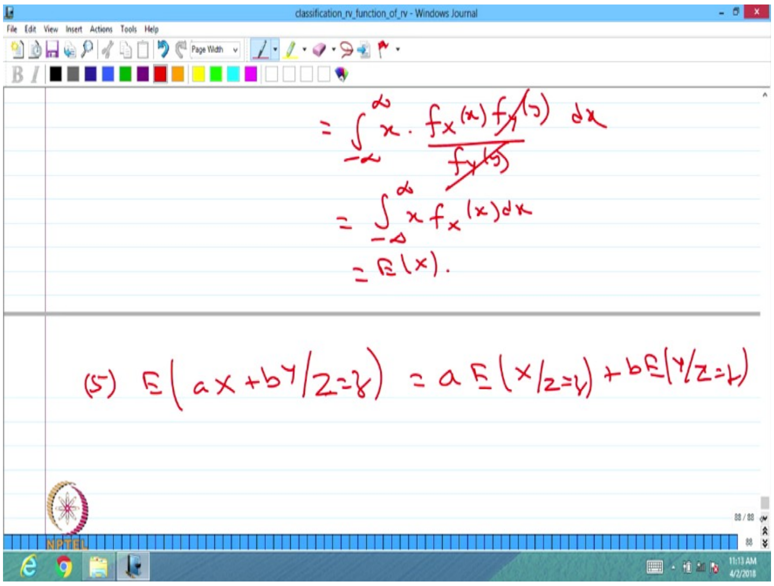
(Refer Slide Time: 02:19)



The third result, if you compute some constant given the other random variable, that is always going to be 1. Next one, if two random variables are independent; if two random variables are independent, then the conditional expectation same as the expectation of that random variable.

Because these two random variables are independent, you can prove it, we can prove it by considering both the random variables are discrete, both the random variables are continuous you can easily prove it. That means, that is suppose I consider both the random variables are of the continuous type; that means, E(X/y) that is nothing but $\int_{-\infty}^{\infty} x \frac{f_{X,Y}(x,y)}{f_Y(y)} dx$ since conditional is same as joint divided by marginal. Again, since these two random variables are independent, I can rewrite the $\int_{-\infty}^{\infty} x \frac{f_X(x) f_Y(y)}{f_Y(y)} dx$ and it cancels out.

(Refer Slide Time: 04:08)

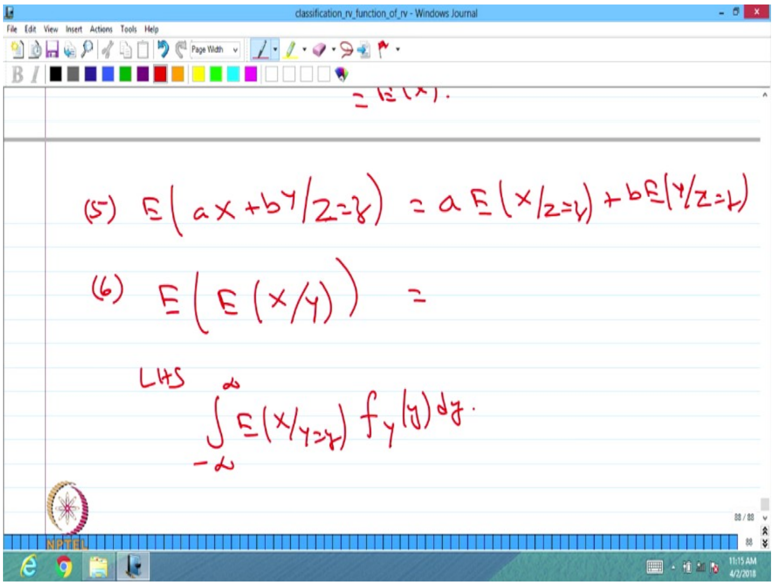


So, you will get $\int_{-\infty}^{\infty} x f_X(x) dx$, that you know that is same as E(X).

Here I have considered both the random variables are continuous type random variables, even if you consider both the random variables are of the discrete type again, you can prove it in the same way. So, if two random variables are independent then the conditional expectation is same as the original expectation.

Next result, suppose you go for finding the E(aX+bY/Z=z); that means, now I am considering three random variables X, Y, Z. So, expectation of some constant times one random variable plus constant times another random variable given the third random variable take some value, that is same as aE(X/Z = z)+bE(Y/Z = z). So, this is true for all a and b both are real this also can be proved.

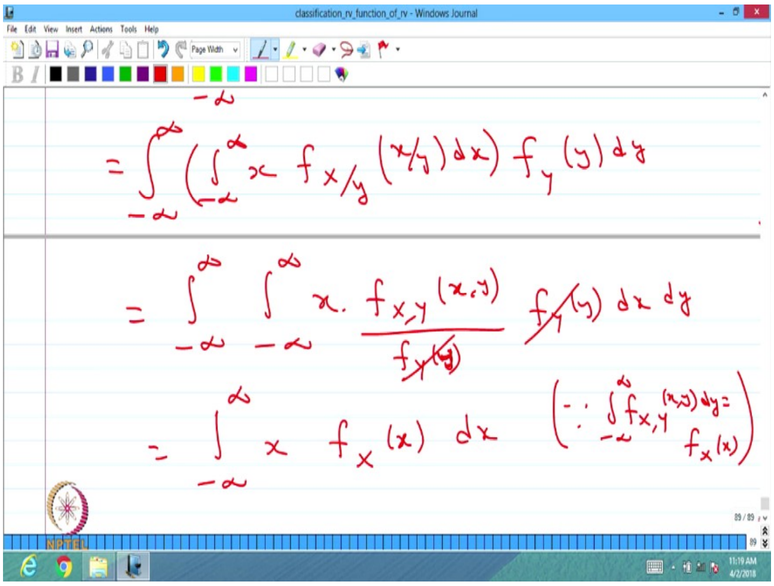
(Refer Slide Time: 05:53)



The sixth property I said E(X/y) is a value of the random variable. Therefore, you can go for finding E(E(X/Y)). E(X/Y) is a random variable. We are finding the expectation for the conditional expectation; that means, you can prove it for considering both the random variables are discrete and continuous.

So, let us prove it then we will write down what is the answer. So, the left hand side is going to be $\int_{-\infty}^{\infty} E(X/Y=y) f_Y(y)\, dy$ by considering both the random variables are of the continuous type.

(Refer Slide Time: 07:00)

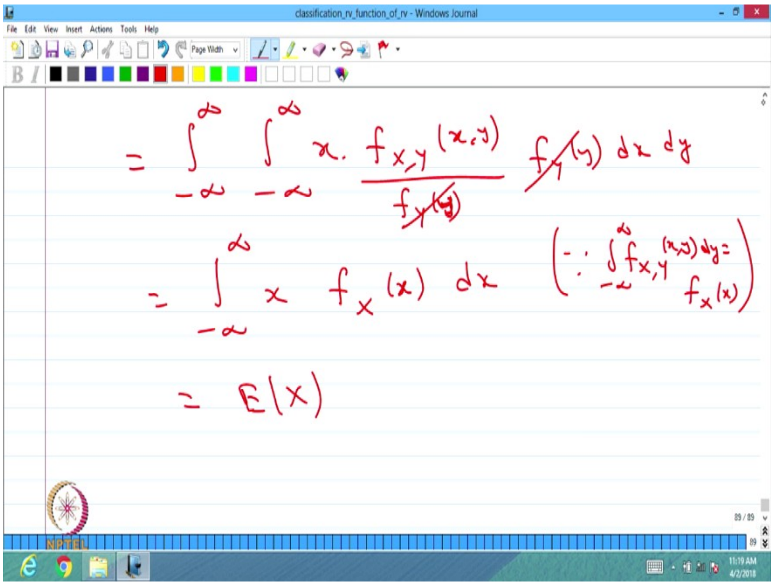


By using the definition of conditional expectation I can expand. So, the outer one is $\int_{-\infty}^{\infty} f_Y(y) dy$. Now, I am just expanding what is a E(X/Y=y) that is nothing but $\int_{-\infty}^{\infty} \frac{x f_{X,Y}(x,y)}{f_Y(y)} dx$so, this is inner integration and outer integration is $\int_{-\infty}^{\infty} f_Y(y) dy$.

That is same as, now, I can expand the conditional probability density function that is nothing but $\int_{-\infty}^{\infty} \frac{x f_{X,Y}(x,y)}{f_Y(y)} dx$ and I can rearrange $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{x f_{X,Y}(x,y)}{f_Y(y)} f_Y(y) dx dy$.

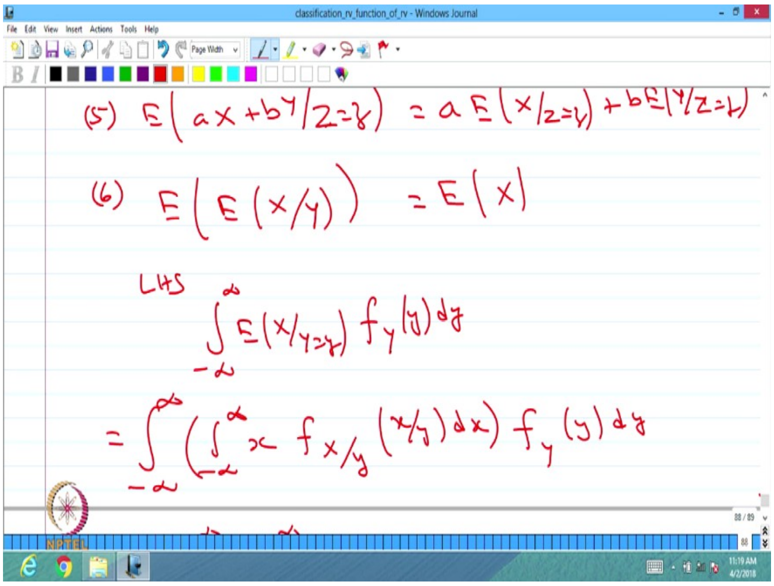
So, I can cancel these two and, also I can use $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x,y) dx dy$. So, I can take it out $\int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy dx$and $\int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$ is nothing but the probability density function of X. $\int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$ is nothing but the probability density function of X. Then $\int_{-\infty}^{\infty} x f_X(x) dx$ that is since $\int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$ that is going to be probability density function of X.

(Refer Slide Time: 09:26)



So, this is nothing but E(X) provided it exist.

(Refer Slide Time: 09:35)

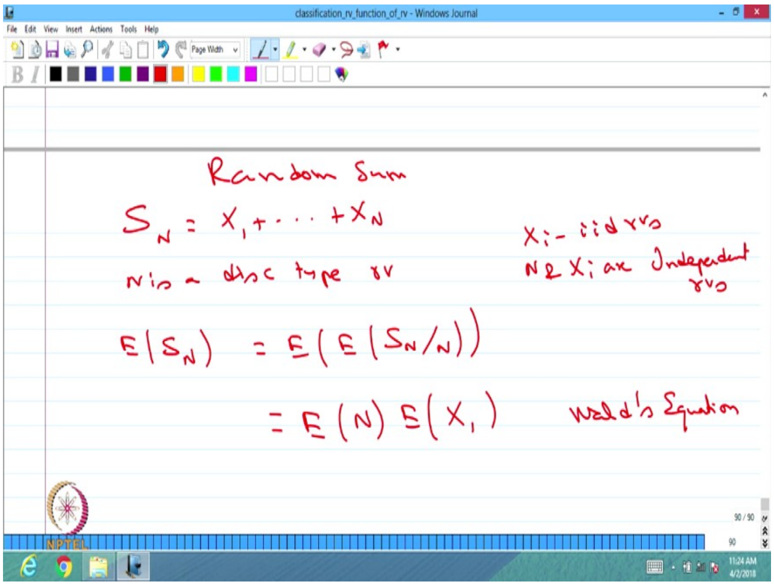


That means the E(E(X/Y)) = E(X). This is a very important result the left hand side involves the random variable Y whereas, the right hand side is free from the random variable Y. That means, if you want to compute the expectation of X, we can always find the another random variable for that random variable, you find out the conditional expectation of X with respect to the sort of dummy random variable. After finding the conditional expectation, you find out the expectation of the conditional expectation that is going to be the E(X).

That means you can choose any random variable Y, as long as the conditional expectation you can able to compute, then find out the expectation of conditional expectation is going to be the original expectation.

So, this is a very important result, whenever it is very difficult to find out the expectation of one random variable, but you can always relate that random variable with another random variable, then compute the conditional expectation. If that process is easy comparing to finding the expectation of the random variable, then you can use this result and find out the expectation of the random variable X, that is same as expectation of conditional expectation X given Y. So, we are going to use this property to compute the expectation for random sum.

(Refer Slide Time: 11:30)

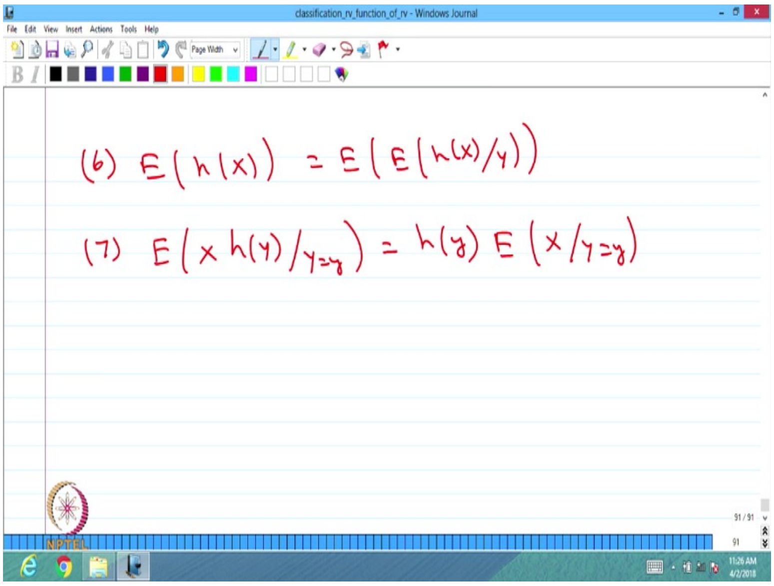


If you recall we have discussed the random some long back that is we have used S_N as the random sum, as a sum of N random variables, where N is discrete type random variable. N is a discrete type, in particular it is a positive integer valued random variable.

So, if you want to find out the E(S_N) that is nothing but E(E(S_N/N)). That means, first you fix the value of capital N, then find out the conditional expectation then, compute the expectation for that, that is nothing but since we made in the random sum problem we made the assumption of X_i's are independent, which will be X_i's are iid random variables; X_i are iid random variables as well as N and X_i's are independent. Therefore, you can compute that is equal to E(N)E(X_1) since all are iid random variable.

Therefore any one random variable E(X_1) multiplied E(N) that is going to be the expectation of the random sum. We have already discussed the random sum much earlier, in that time we have discussed only the distribution of the random sum. Now, we are discussing the E(S_N) = E(N)E(X_1). Since all the X_i's are independent random variable. This is called a Wald's equation. Using the earlier property we can find out the E(S_N) = E(N)E(X_1).

(Refer Slide Time: 14:06)

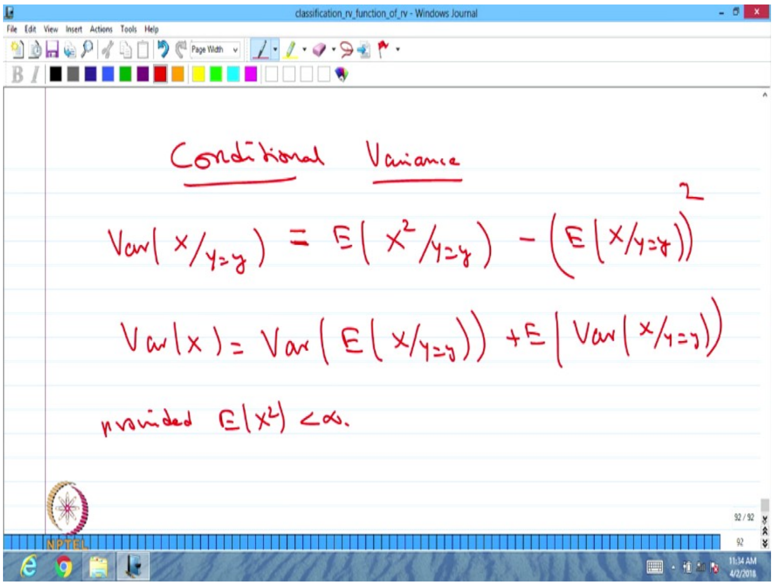


The next property, property number 6, so, this can be extended for a function of random variable E(h(X)), also can be computed in the same way E(E(h(X)/Y)). As long as h is a Borel measurable function, therefore, h(X) is a random variable. So, the E(h(X)) = E(E(h(X)/Y)). That means, we are extending the previous result with the function of random variable

In the 7th result E(Xh(Y)/Y=y) = h(y)E(X/Y=y). E(Xh(Y)/Y=y) = h(y)E(X/Y=y). That means, when you made it given Y takes a value y, the h(y) no more random variable. Y is a random variable h(Y) is a function of a random variable, where h is a Borel measurable function therefore, h(Y) is a random variable.

So, when Y takes a value y, h(Y) is no random it is going to be treated as a constant. So, constant is out while computing the expectation. So, h(y) times the conditional expectation. There are some more properties with the conditional expectation that is with respect to the sigma field and so on. So, for this course we stop it with the this many results.

(Refer Slide Time: 16:32)



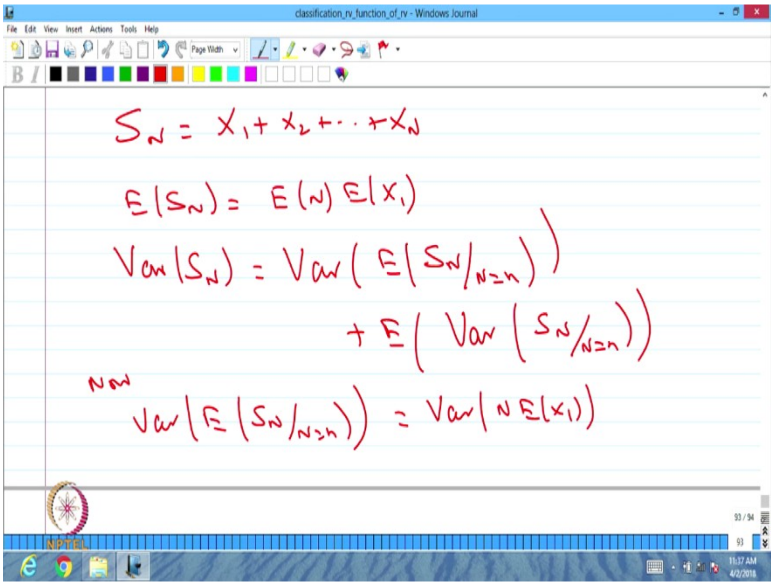
Now, we will go for the next topic is called conditional variance. The way we discussed the conditional distribution first then conditional expectation. So, now we are going to define the conditional variance that is nothing but variance of we are going to define, conditional variance as the Var(X/Y=y), that is same as, it is similar to the variance formula, but now it is a conditional variance that is same as $E(X^2/Y=y)$ - $(E(X/Y=y))^2$.

When you compute the $Var(X)=E(X^2)-E(X)^2$ the same way, one can define the conditional variance as a Var(X/Y=y) that is nothing but the $E(X^2/Y=y)$ - $(E(X/Y=y))^2$.

We have already given the result for conditional expectation of h(X). So, now, we are going to find out the var(X), using the conditional expectation and the conditional variance, that is the var($E(X/Y=y)$) + E(Var(X/Y = y)) provided the second order moment that is finite.

So, the way we are finding the E(X) using the conditional expectation, one can find the var(X) also, in terms of the conditional expectation and the conditional variance. That means, for any random variable, var(X) can be computed by finding the conditional expectation and the conditional variance first, then apply the variance and the expectation with the vice versa, then summation becomes the var(X).

(Refer Slide Time: 19:18)

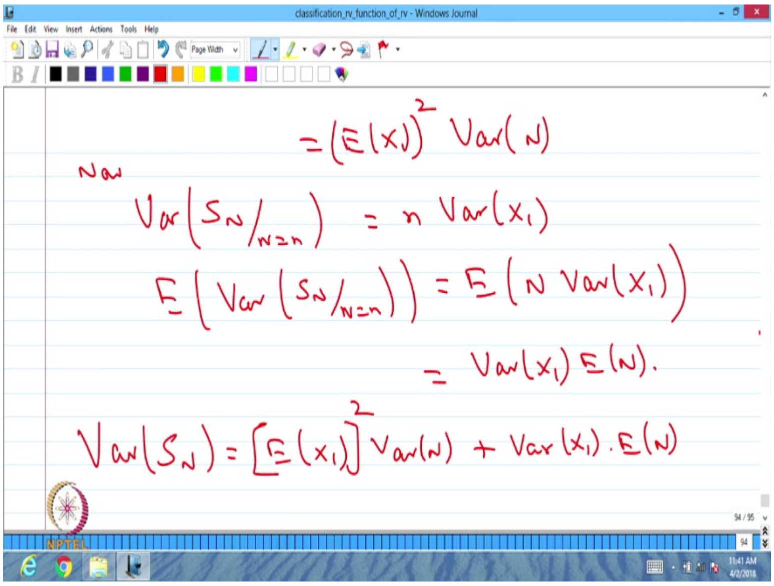


We are going to use this result to compute the variance of a random sum that is a random sum is defined, sum of iid random variable and, N is a discrete type positive integer random variable, which is independent of X_i's. So, we have already got it the $E(S_N) = E(N)E(X_1)$. Now, we are going to find out var(S_N) using the previous result that is going to be var(E($S_N/N=n$)) + E(var($S_N/N=n$)).

So, one can find the variance of random sum using the conditional expectation and the conditional variance, computing the expectation for the conditional variance and variance for the conditional expectation that sum quantity is going to be the variance of random sum.

We have already the results for var(E($S_N/N=n$)) that is nothing but the var(E($S_N/N=n$)) that is nothing but E($S_N/N=n$) that is nothing but N $E(X_1)$. Since they are identical, we can make it N $E(X_1)$.

(Refer Slide Time: 21:35)



That is same as that is same as since we are going for var(N $E(X_1)$). $E(X_1)$ you have to treat it as a constant. So, when you are computing the var(N), N is a random variable here and this is a constant therefore, $E(X_1)^2 var(N)$.

So, this is a one quantity for variance of conditional expectation. So, we got variance of expectation now, you have to go for conditional variance then you have to go for expectation. So, first we have to find out var($S_N/N=n$) that is going to be when N is fixed so, you are trying to find out the variance of finite N X_i's; that means, all are independent random variable therefore, it is going to be $N\,var(X_1)$.

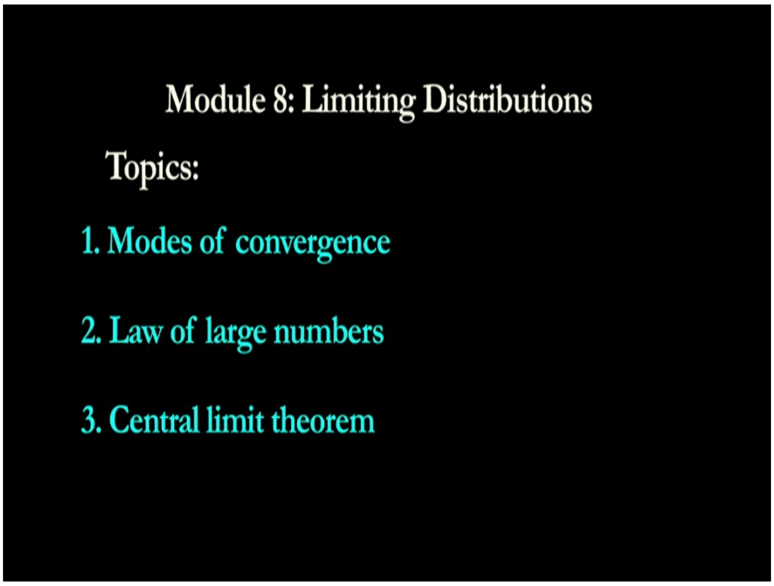
Therefore, the E(var($S_N/N=n$)), so, that is same as E($N\,var(X_1)$). $var(X_1)$is a constant, so, constant is out. So, that is $var(X_1)$E(N).

Therefore, adding both the quantities var(S_N) is going to be the earlier answer is it is $E(X_1)^2 var(N)+var(X_1)$E(N). So, this is going to be the variance of the random sum. So, once you know the distribution of X_i's and distribution of N, one can find the distribution of random sum. And if you know the mean and variance of X_i's and N, you can find out the mean and variance of the random sum also.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

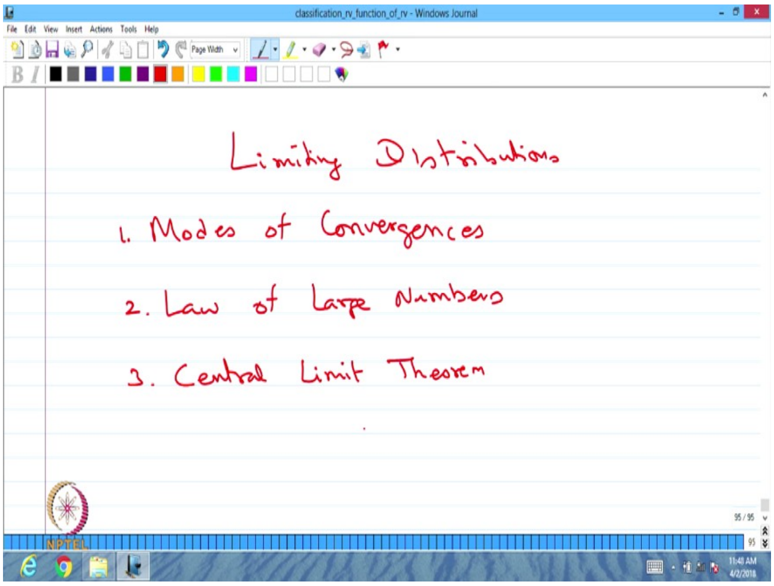
Module – 08
Lecture – 42
Limiting Distributions

(Refer Slide Time: 00:14)



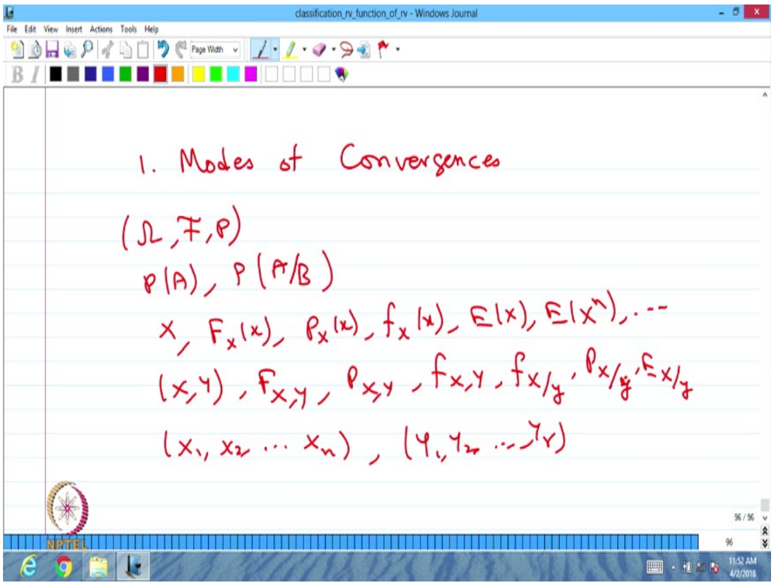
In this model, we are going to discuss limiting distributions.

(Refer Slide Time: 00:23)



So, we cover three important topics; first topic is called modes of convergences and the second topic we are going to discuss law of large numbers and the third topic which we are going to discuss, that is a very important topic, that is central limit theorem. In this model, we are going to discuss these three topics has three different lectures. In the first lecture, we are going to discuss modes of convergence; second lecture, we are going to discuss law of large numbers; third lecture, we are going to discuss central limit theorem.

(Refer Slide Time: 01:33)



The first lecture is on modes of convergences. Till now what we discussed the probability space, probability of an event, then we discussed a conditional probability of events, then we have introduced a random variable, then we discuss the CDF of the random variable. Then, based on the discrete type or continuous type random variable we discuss the probability mass function, probability density function.

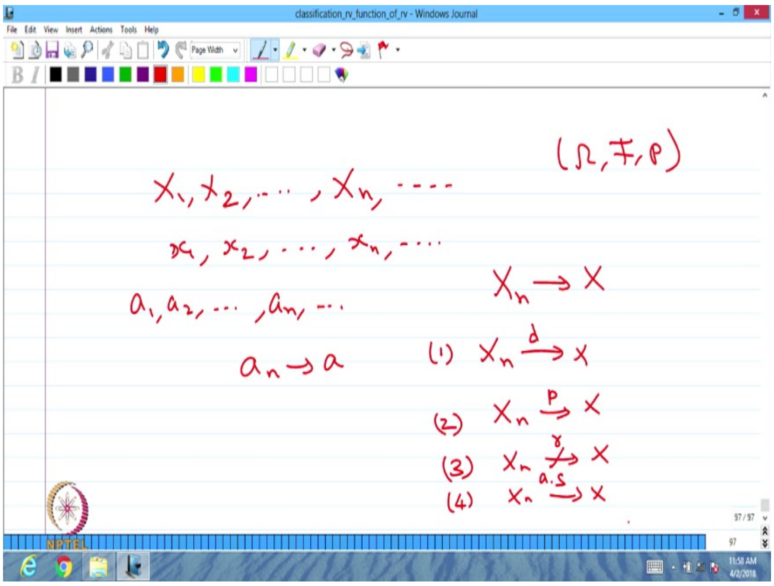
After we introduced one random variable in a probability space, we have discussed the distribution in the form of CDF, density function or mass function. Later, we said one random variable is not enough to solve some particular problems; you may need more than one random variable; to be defined in the same probability space to solve the given problem.

Then we introduce two random variables, then we discuss the CDF, we discussed the joint probability mass function, we discussed joint probability density function, then later we discuss the conditional probability density function, conditional probability mass function. For the first random variable, we discuss the mean, variance and so on, that I missed earlier second order moment in a third other moment and so on and here we discuss the conditional distribution, conditional probability, conditional expectation and so on.

Not only two random variables, the later we introduce n random variables in the same probability space, then we discuss the joint distribution of n dimensional random variable, after that we introduce the transformation from one dimensional to another n dimensional random variable or r dimensional random variable and so on.

Now, we are going to discuss not a one random variable, not two random variables, not n finite random variables we are going to discuss sequence of random variables. This is also possible when you solve a given problem you may need to know the sequence of random variable and what is their distribution as n tends to infinity?

(Refer Slide Time: 04:55)



The first question comes, whether if you have a sequence of random variables; whether if you have a sequence of random variable whether this sequence of random variable converge or not converges to one random variable. If it converges what is the distribution of that sequence of random variable that is going to be the topic of modes of convergence for the random variables.

For these first we can understand how the sequence of real numbers converges. Suppose you have real numbers $a_1, a_2, \ldots, a_n, \ldots$ whether this sequence of real numbers converges have discussed in any real analysis courses of a sequence and series of real numbers and so on.

The same concept the we are planning to introduce for the sequence of random variables, but the only difference is, the random variable takes real numbers with some distribution. That means, the X_1 random variable can take the real values x_1 with some distribution. Similarly, the random variable X_2 may takes a value, a real value x_2 with some distribution and so on.

Similarly, the random variables $X_n$can take the real number x_n with some distribution. If you know the distribution of each random variable $X_1, X_2, \ldots$ you have a sequence of random variable all are defined in the same probability space, that is very important. All are defined in the same probability space and if you know the distribution of this sequence of random variables whether this sequence of random variable converges to one random variable or not.

If it converges what is the distribution of that. That means, if I write $X_n \rightarrow X$this is notation. When I write $X_n \rightarrow X$; that means, I have a sequence of random variable $X_1, X_2, X_3, \ldots$ whether this sequence of random variable converges to the one random variable that is call it as X. If I know the distribution of X_i's what could be the distribution of X? That is the question.

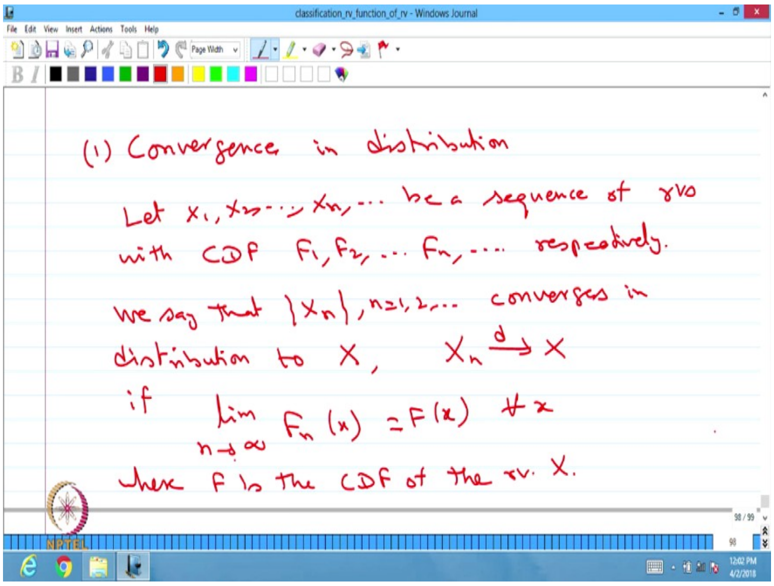
In sequence of real numbers converges to one real number that may be easy comparing to the sequence of random variable converges to the one random variable. Since, each random variable attached with some distribution for some random variables moments of first order may exist further moment may not exist and so on. Therefore, you cannot make only one way, you can conclude this sequence of random variable converges to one random variable X. There may be more than one ways, you can conclude that this sequence of random variable converges to one random variable that we call it as a modes of convergence.

So, we are going to discuss there are 4 modes of convergence. The first mode of convergence we write it as X_n converges to X in distribution by writing small letter d above the arrow I am going to give the definition one by one in detail with examples. Comparing to the sequence of real numbers how it converges to one random variable. Similarly, we are going to discuss how the sequence of random variable converges to one random variable in different ways. The different ways we say it as different modes.

The first mode of convergence that is called the convergence in distribution with the letter d above the arrow. The second one that is convergence in probability. The third one convergence in r-th moment, the forth one convergence in almost surely a.s. Above the arrow if you write d that means, convergence in distribution, above the arrow if I write small p that means, convergence in probability, above the arrow I put the slash with r; that means, it is a convergence in the r-th moment, the forth one convergence in almost surely.

So, these are all the four ways the sequence of random variable different in a same probability space converges to one random variable which we denote it as X that is also defined in the same probability space in four different modes of convergence.

(Refer Slide Time: 10:45)



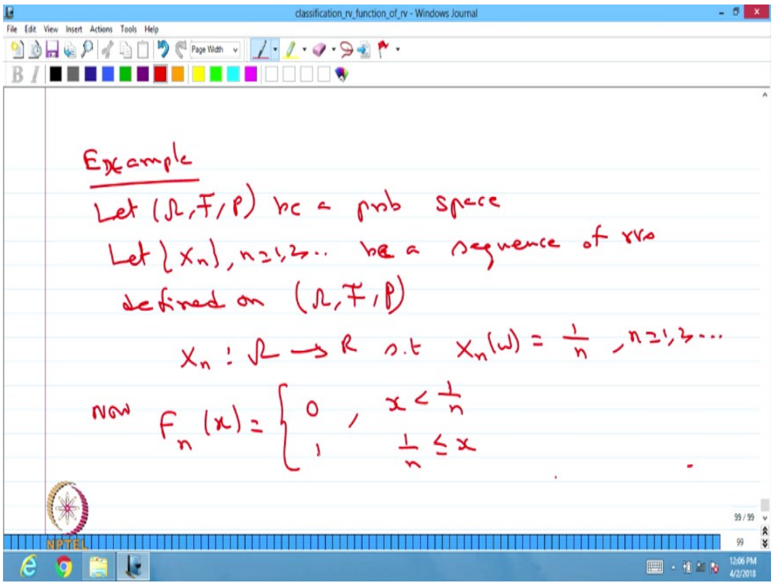
Let us start with the first mode of convergence that is convergence in distribution. In distribution let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables with CDF $F_1, F_2, \ldots, F_n, \ldots$ respectively. That means, the random variable X_1 has a CDF F_1 the random variable X_2 has a CDF $F_2$and so on. All these random variables defined in the same probability space.

We say that X_n the sequence n take the value 1, 2 and so on converges in distribution to the random variable denoted by X can be written as X_n converges to X in distribution if $\lim_{n\to\infty} F_n(x) = F(x)$ for all x where F is the CDF of the random variable X.

As long as the $\lim_{n\to\infty} F_n(x) = F(x)$ where F(x) is CDF of the random variable X that is valid for all x, this condition valid for all x. Then we can conclude the sequence of random variables convergence in distribution to the random variable X. Note that $F_1, F_2, \ldots, F_n, \ldots$ that is the

CDF of the sequence of random variables respectively and that converges to a function that is the CDF of the random variable X, then we can conclude this convergence in distribution to the random variable X.

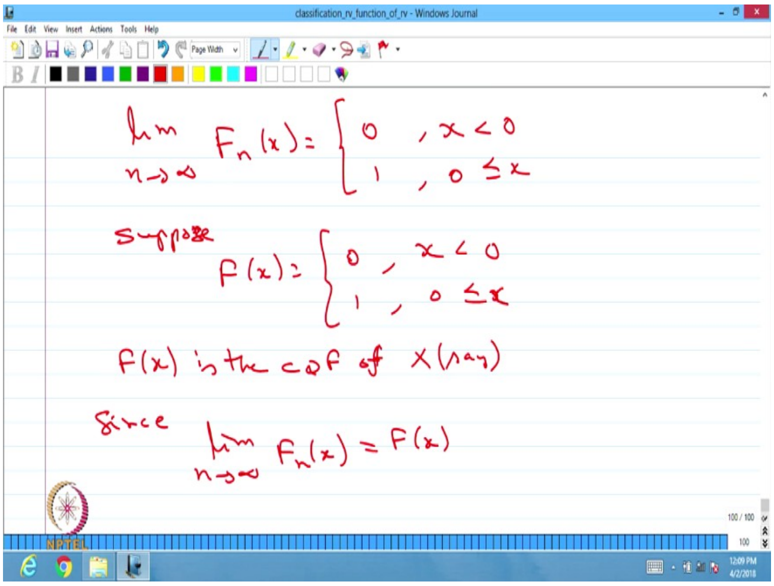
(Refer Slide Time: 14:37)



Let us give a one simple example through that we can understand the definition clearly. Example; let (Ω, F, P) be a probability space. Let X_n; n is equal to 1, 2, and so on be a sequence of random variables defined on the probability space (Ω, F, P), that is defined as; X_n is defined from Ω to R such that such that $X_n(w)=\frac{1}{n}$for n is equal to 1, 2 and so on. So, we are defining the sequence of random variable from Ω to R such that $X_n(w)=\frac{1}{n}$ for n is equal to 1, 2, and so on.

Now, we will find out what is CDF of this sequence of random variables. So, if you find out the CDF of the n-th random variable as a function of x this is going to take the value 0, when x < 1/n; from 1/n onwards; when x is going to be 1 by n onwards it is going to take the value 1. So, this is going to be the CDF of the sequence of random variable X_n's. So, here n takes a value 1, 2 and so on.

(Refer Slide Time: 17:15)

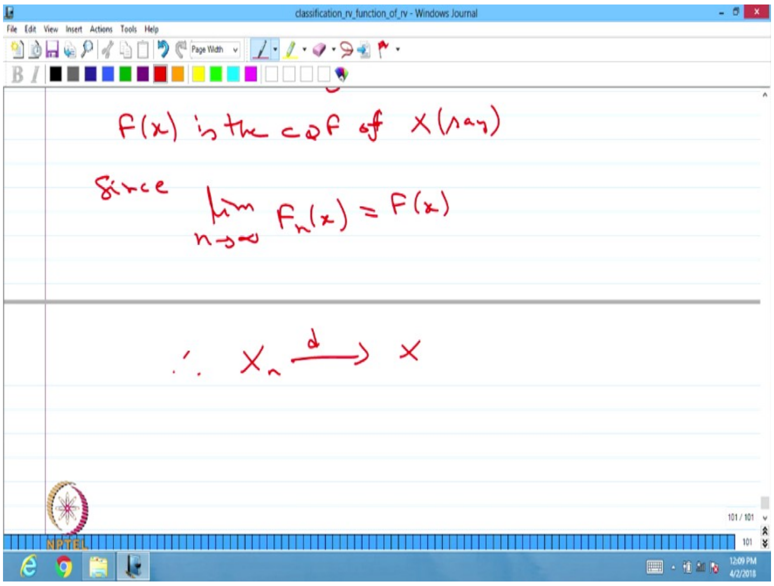


Now, let us go for finding the $\lim_{n \to \infty} F_n(x)$. What could be the value as $\lim_{n \to \infty} F_n(x)$. This is going to be 0 when x is lesser than 0 because as n tends to infinity of 1/n that becomes 0 and 1 from 0 onwards. This is a $\lim_{n \to \infty} F_n(x)$..

Suppose, I denote this as the F(x), suppose I make F(x) that takes a value 0, when x is lesser than 0, 1 from 0 onwards. Verify whether this is going to be the CDF of some random variable. It starts from 0 land up 1 and so on. It satisfies all the properties of a CDF therefore, this F(x) is the CDF of some random variables you denote it as X, say F(x) is the CDF of some random variable X. By seeing $\lim_{n \to \infty} F_n(x)$ that is same as F(x).

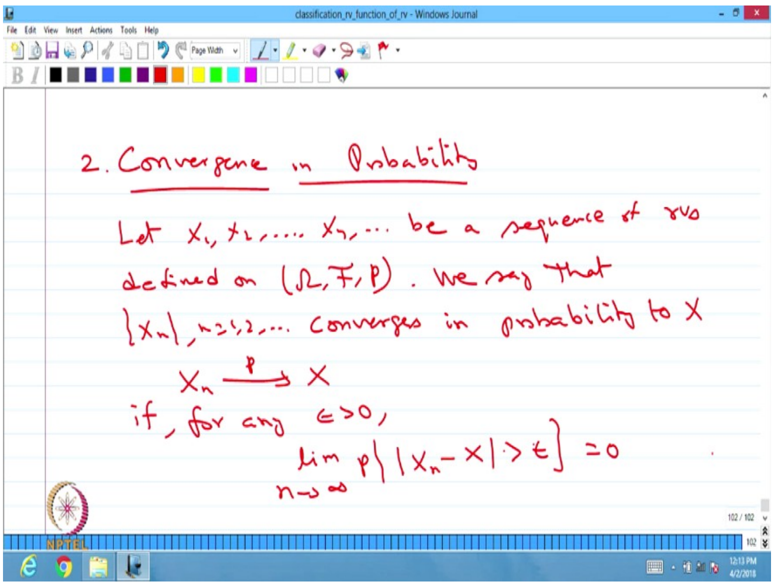
Since $\lim_{n \to \infty} F_n(x) = F(x)$, where F(x) is the CDF of some random variable X and the left hand side $F_n(x)$is the CDF of the sequence of random variable or $F_n(x)$ is the CDF of the random variable $X_n$as the limit n tends to infinity that is same as the CDF of the random variable X. Therefore, we can conclude, this is the condition is satisfied by the convergence in distribution.

(Refer Slide Time: 19:52)



Therefore, you can conclude the X_n converges to X in distribution. There is a possibility the sequence of random variable CDF's may converge to some function that may not be the CDF. So, as long as this sequence of CDF converges to some function that is also the CDF of some random variable then you can conclude the X_n converges to X in distribution. So, like that we have some more problems that we will discuss little later; that means, when we discuss other modes of convergence we can verify whether this satisfies convergence in distribution also.

(Refer Slide Time: 20:41)

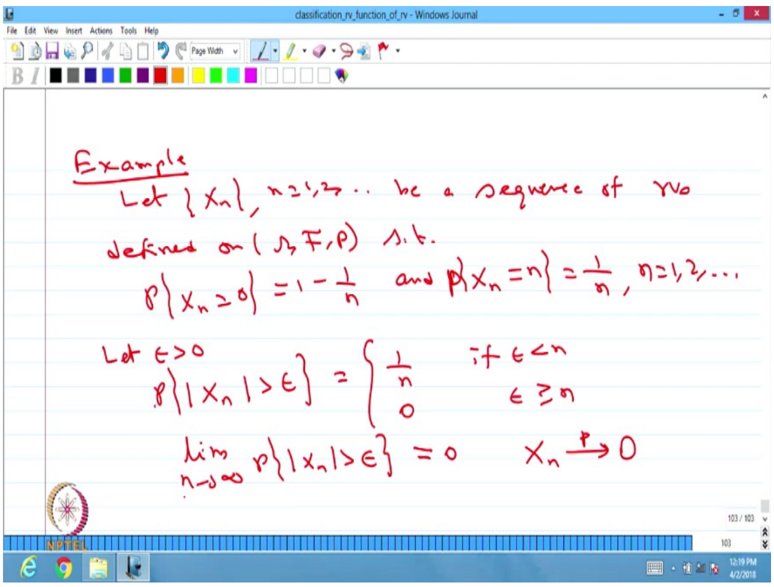


So, now will move into next mode of convergence that is a convergence in probability. Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables defined on the probability space (Ω, F, P). We say that this sequence of random variable converges in probability to the random variable X and write it as X_n converges to X in probability if for any ϵ, which is greater than 0, $\lim_{n\to\infty} P\{|X_n - X| > \epsilon\} = 0$. If this condition is satisfied for any $\epsilon > 0$, finding out the $\lim_{n\to\infty} P\{|X_n - X| > \epsilon\} = 0$, then we conclude this sequence of random variables converges in probability to the random variable X.

Note that to verify this sequence of random variable converges in probability you should know the random variable X, then finding of the probability after is verified then you can conclude this sequence of random variable convergences in probability. That means, you should know about the distribution of the random variable X or at least you should know how to compute the $P\{|X_n - X| > \epsilon\}$ for any $\epsilon > 0$. That means, beforehand you should have the distribution of the random variable X along with the distribution of the sequence of random variable X_n's then only you can conclude whether this sequence of random variable convergence in probability.

So, for this mode of convergence will go for the example, through that we will understand.

(Refer Slide Time: 24:14)



The example is, example let X_n be a sequence of random variables defined on (Ω, F, P) such that the $P\{X_n=0\}=1-\frac{1}{n}$ and the $P\{X_n=n\}=\frac{1}{n}$. So, this is true for all n, n is equal to 1, 2, and so on.

We have a sequence of random variable whose distribution is defined X_n takes a value 0 with the probability $1-\frac{1}{n}$ or X_n takes a value n with the probability $\frac{1}{n}$; that means, this sequence of random variables are of the discrete type which has the two points, 0 and n are the mass points either the mass is at 0 or n for the n-th random variable and you have sequence of random variable, n is equal to 1, 2 and so on.

In this example we can go for taking, let $\epsilon > 0$ you can go for finding $P\{|X_n| > \epsilon\}$. Finding out the $P\{|X_n| > \epsilon\}=\frac{1}{n}$ if $\epsilon < n$. If ϵ is going to be greater than or equal to n the $P\{|X_n| > \epsilon\} = 0$, this is for fixed $\epsilon > 0$.

Now, you can go for taking $\lim_{n\to\infty} P\{|X_n| > \epsilon\}$. As limit n tends to infinity this quantity is going to be; the right hand side is going to be 0. Since the $\lim_{n\to\infty} P\{|X_n| > \epsilon\} = 0$, therefore, you can treat the $|X_n-0|$ that is equivalent of concluding $X_n$converges to 0 in probability. 0 you can treat it as the random variables X takes a value 0 with the probability 1. You can make a X_n tends to X in probability, where X is a degenerated random variable or constant which takes a value 0 with the probability 1.

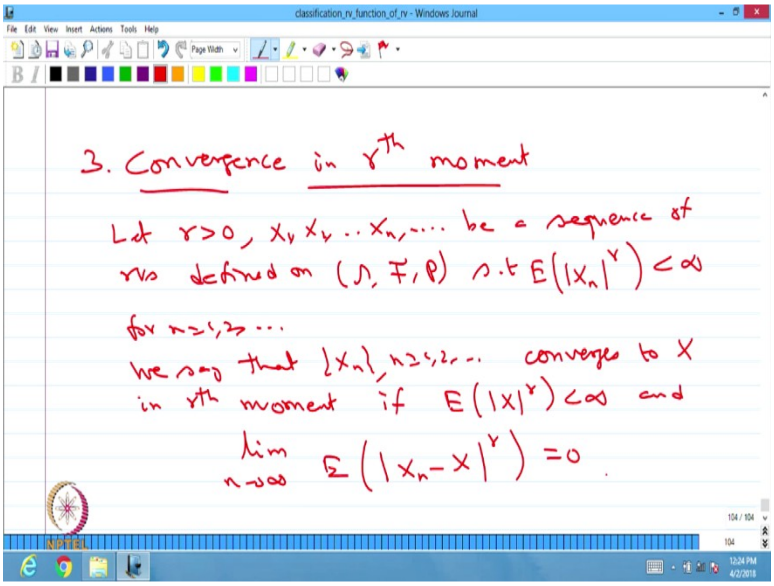
So, sometimes are the sequence of random variable converges to constant also. So, this is example in which we have given sequence of random variable converges to constant 0 in probability. Since it converges to 0 therefore, we are directly going for $P\{|X_n| > \epsilon\}$. Sometimes if you have a random variable X and whose distribution is known then you can go for finding out the $P\{|X_n-X| > \epsilon\}$, then whether the limit n tends to infinity, this quantity is going to be 0 or not accordingly you can conclude sequence of random variable converges to random variable X in probability or not. So, here it is a very easiest example in which we land up the sequence of random variable converges to 0 in probability.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 43

Now, we will discuss the third mode of convergence that is convergence in rth moment.

(Refer Slide Time: 00:09)



Let r > 0, $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables defined on (Ω, F, P) such that the rth moment for the sequence of random variable exist; the absolute sense expectation is going to be finite for n is equal to 1, 2 and so on.

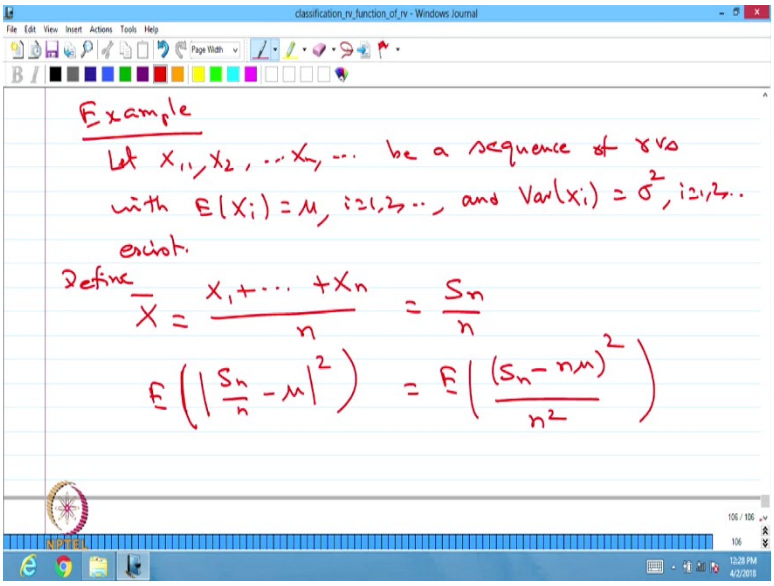
We say that this sequence of random variable converges to the random variable X in rth moment, if the expectation of the random variable X the rth moment that is finite and X is also defined in the same probability space.

And $\lim_{n\to\infty}$ ¿; $\lim_{n\to\infty} E[|X_n - X|^r]$ = 0. We are keeping r > 0 and we have a sequence of random variable defined in the probability space such that rth moment exist for those sequence of random variable.

We say that the sequence of random variable converges to X in rth moment provided the rth moment exist for the random variables X, which is defined in the same probability space and the $\lim_{n \to \infty} E[|X_n - X|^r]$, the rth moment is going to 0.

If these conditions are satisfied, then we can conclude the sequence of random variable converges to the random variable in the rth moment.

(Refer Slide Time: 03:27)

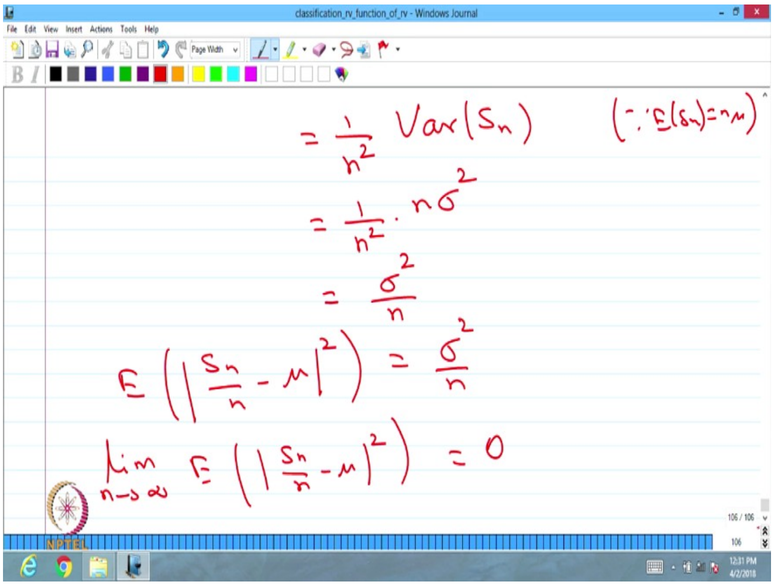


We can go for the example to explain this mode of convergence also. That is let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables with mean μ, identical μ and variance of each random variable that is σ^2 exist; both exist and is denoted by μ for the mean and variance is σ^2. I am going to define the new random variable that is $\frac{X_1 + X_2 + \ldots + X_n}{n}$; sum of random variable we usually use the notation S_n.

But now, I am going to divide that sum by n. So, I am going to use a notation called $\acute{X}$. $\acute{X}$ is the random variable; it is a function of n. I am not writing n in the left side, $\acute{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$ or we can use another notation we can use as a function of S. S n divided as a function of n. $\frac{S_n}{n}$. Either I can use $\acute{X}$ or $\frac{S_n}{n}$, both are one and the same that is sum of random variable.

So, I am going to define this random variable for different n will go for finding first $E[¿\frac{S_n}{n}-\mu|^2]$. In absolute sense, we will find what is the value of $[¿\frac{S_n}{n}-\mu|^2]$ that is same as expectation of since it is the ¿¿¿.

(Refer Slide Time: 06:42)



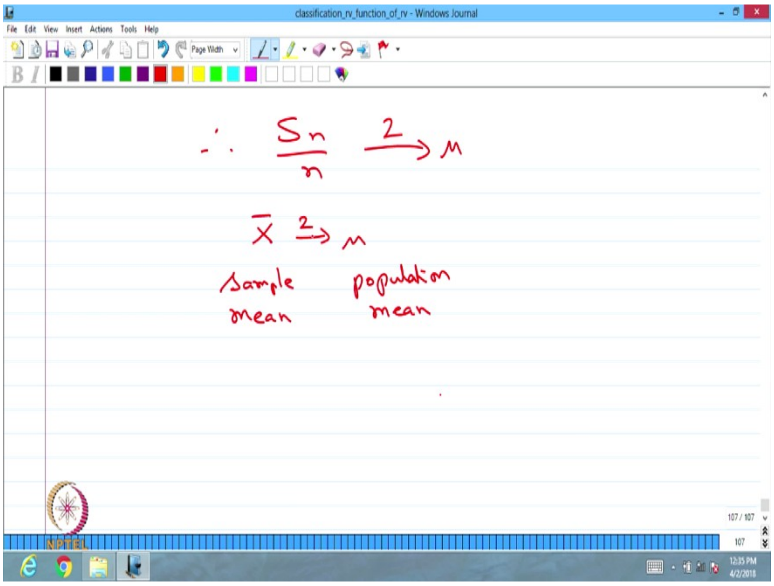
That is same as, I can take $\frac{1}{n^2}$outside, if you see E{ ¿¿}. If you find out the E[S_n] $E[S_n]=E[X_1+X_2+\ldots+X_n]$ all the E[X_i] = μ. Therefore, it is nμ.

So, since $E[S_n]=¿$nμ. So, the E{¿¿. I will make additional condition X_i's are sequence of independent random variable with the mean is μ and the variance is σ^2.

Therefore, the $Var(S_n)=Var(X_1+X_2+\ldots+X_n)$ and Var ¿s) are σ^2. Therefore, it is going to be n σ^2. Therefore, this is going to be $\frac{\sigma^2}{n}$. We got the $E[¿\frac{S_n}{n}-\mu|^2]=\frac{\sigma^2}{n}$.

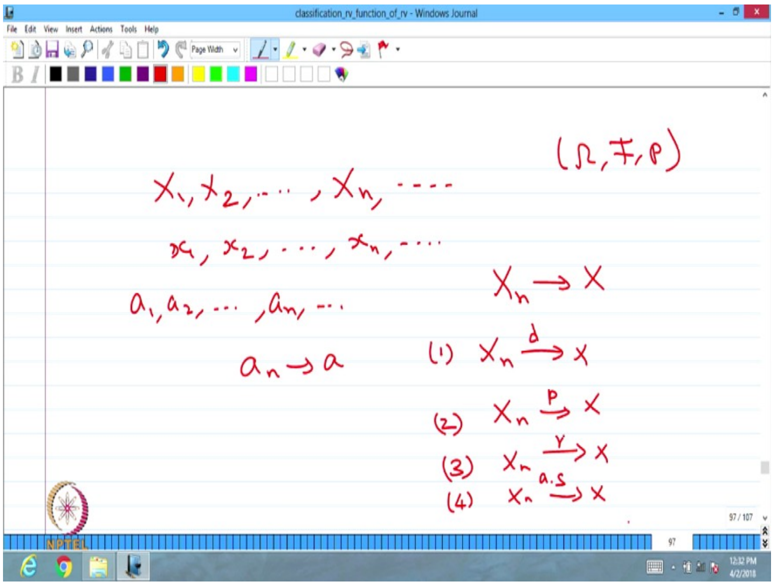
Now, let me apply $\lim_{n\to\infty} E[¿\frac{S_n}{n}-\mu|^2]$. Since, n is in the denominator this becomes 0.

(Refer Slide Time: 08:42)



So, this is the definition we wanted to convergence in rth moments; that means, I can conclude, $\frac{S_n}{n} \to \mu$ in second order moment. The left right hand side it is not a random variable, it takes a value μ; that means it is a random variable takes a value μ with the probability 1.

(Refer Slide Time: 09:12)

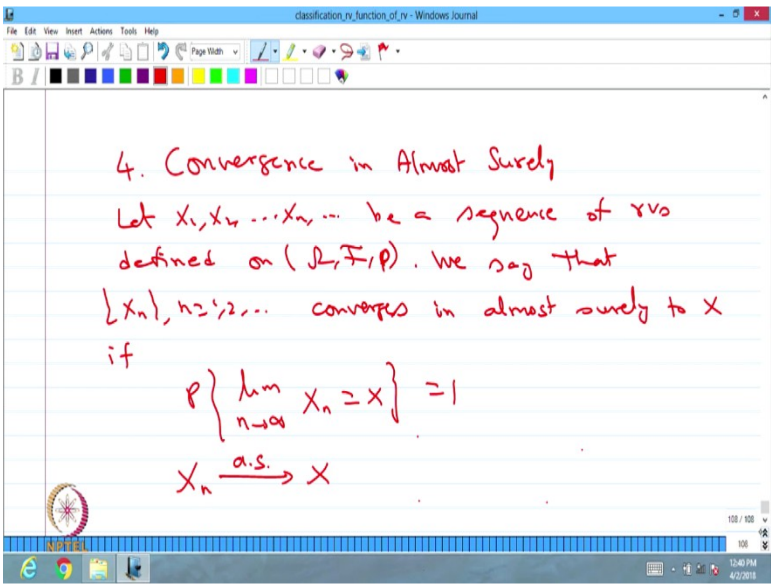


So, X_n converges to rth moment with number r. In notation, we write $\frac{S_n}{n} \to \mu$ in the second order moment. In the right hand side μ means it is a random variable takes a value μ with the probability 1. I have already written $\frac{S_n}{n} = \acute{X}$. So, $\acute{X} \to \mu$in second order moment.

In statistics we use the $\acute{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$that is from the n random variable we call it as a Sample mean. The sequence of random variable having the mean μ, variance σ^2. So, the μ is called Population mean. So, the conclusion is if you have a population with the mean μ and if you get a sample of size n, then the sample mean will converge to the population mean as n tends to infinity.

That means, for a large sample, for a large sample the sample mean will converge to the population mean and this converges takes place in the convergence in rth moment, where r = 2 here. Because we are finding the second order moment convergence is used. Therefore, sample mean converges to the population mean in second order moment.

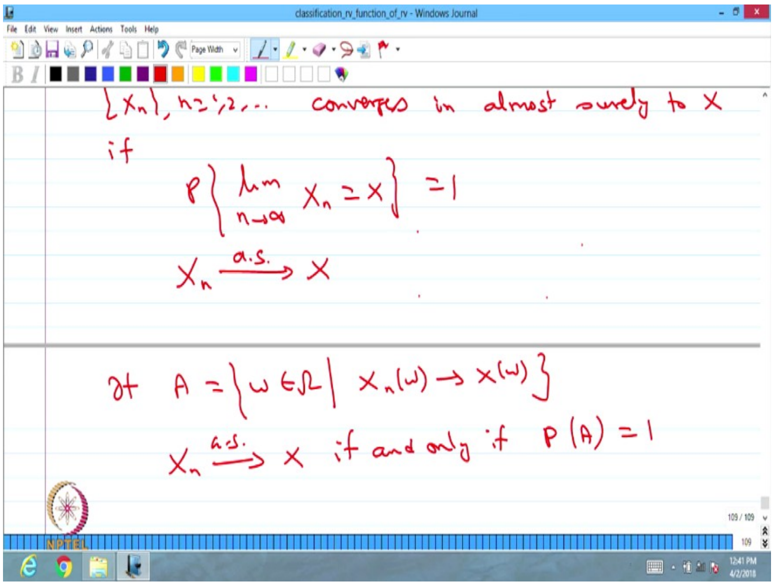
(Refer Slide Time: 11:18)



We move to the fourth mode of convergence; convergence in Almost Surely. Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables defined on the probability space (Ω, F, P). We say that the sequence converges; the sequence converges in almost surely to the random variable X, if the $P\left\{\lim_{n \to \infty} X_n = X\right\} = 1$.

So, here the random variable X is defined in the same probability space. The $P\left\{\lim_{n \to \infty} X_n = X\right\} = 1$, then one can conclude the sequence of random variable convergence to the random variable X in almost surely. So, we write it X_n converges to X above the arrow, you write a.s.; that means, this convergence in almost surely.

(Refer Slide Time: 13:42)

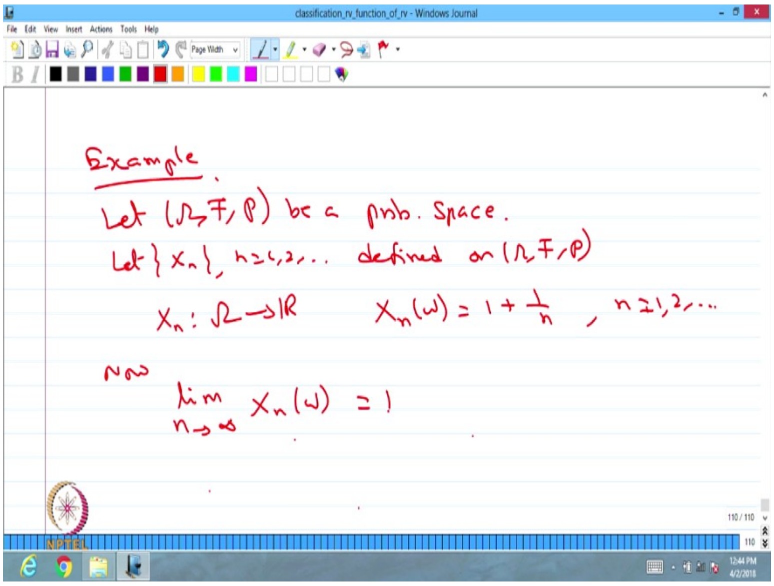


Suppose, we make event A is nothing but collection of w's belonging to Ω such that $X_n(w) \to X(w)$. In that case when I say $X_n$converges to X almost surely, if and only if the probability of the event A that is going to be 1. If I make event A that is nothing but as n tends to infinity $X_n(w) \to X(w)$ and collect those w's that is going to be the event A. If the probability of that event A is going to be 1; then, we can conclude the X_n converges to X almost surely.

In other words, negation of the event A that probability is going to be 0; that means, you collect all possible outcomes from omega which does not converge $X_n(w)$¿$X(w)$, you collect those possible outcomes whose measure is 0. Then, also we can conclude X_n converges to X almost surely; that means, the whole unit mass is attached with the collection of possible outcomes; those outcomes are the outcomes of $X_n(w) \to X(w)$.

Those outcomes satisfies $X_n(w) tends ¿ \vee converges ¿ X(w)$ as n tends to infinity. So, if you include those possible outcomes or w's those possible outcomes whose probability mass is 1 or whichever is not satisfy in this condition, those possible outcomes whose probability mass is 0. Then we can conclude X_n converges to X in almost surely. We will go for the example for this.

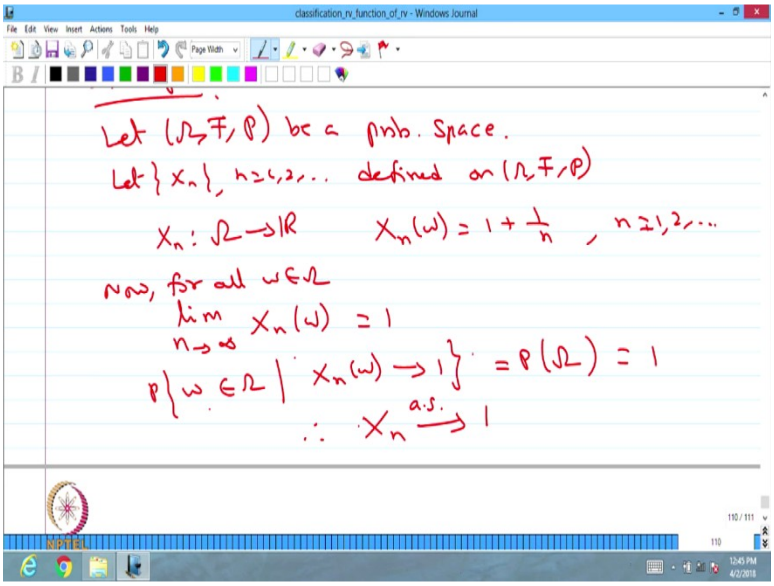
(Refer Slide Time: 16:07)



Let (Ω, F, P) be a probability space. Let sequence of random variable defined on the probability space (Ω, F, P) as X_n, X_n is mapping from Ω to real line such that $X_n(w)=1+\frac{1}{n}$ for n is equal to 1, 2 and so on.

Now, will go for computing $\lim_{n\to\infty} X_n(w)$ where w is belonging to Ω. $\lim_{n\to\infty} X_n(w)$ that is going to be 1.

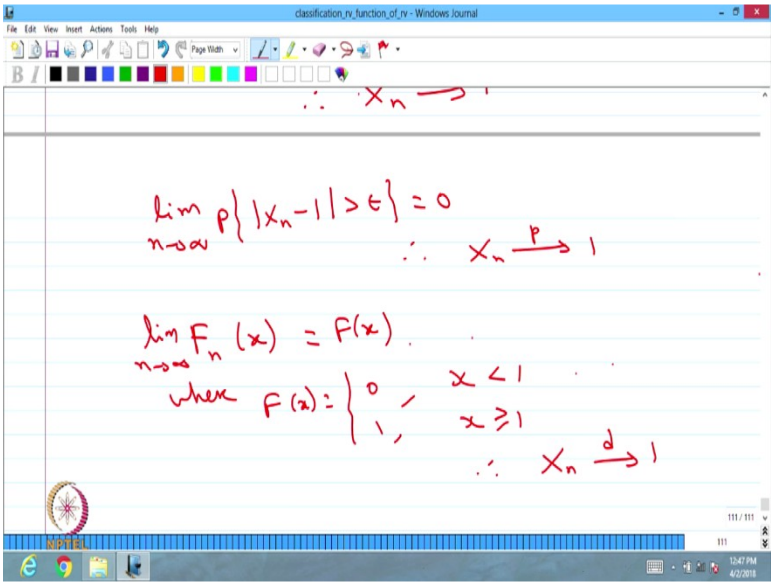
(Refer Slide Time: 17:25)



That means, the set of all w's belonging to **Ω** such that $X_n(w)$ will tends to 1. You collect those possible outcomes find out the probability of that, that is nothing but P(**Ω**) = 1. In that case, one can conclude X_n converges to 1 almost surely.

Because for all w belonging to the $\lim_{n\to\infty} X_n(w)$ = 1. Therefore, finding the probability of w belong into make a satisfying this condition that is $X_n(w)$ tends to 1 as n tends to infinity, that is going to be the whole possible outcomes that is nothing but the P(**Ω**) = 1. Therefore, it satisfies the condition for almost surely. Therefore, X_n converges to 1, almost surely. So, the same example can be considered for other mode of convergence also.

(Refer Slide Time: 20:01)

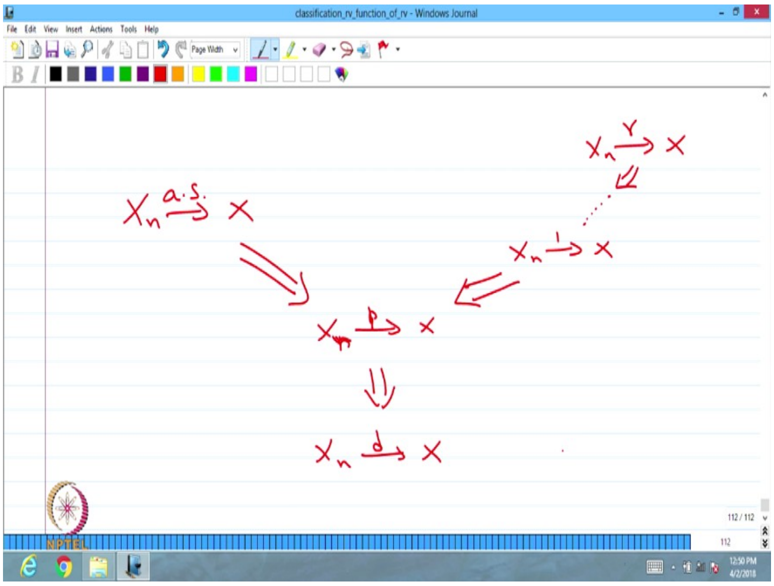


That means, one can prove $\lim_{n \to \infty} P\{\vee X_n - 1 \vee ¿ \epsilon\}$, that also we can prove this is going to be 0. This implies X_n converges to 1 in probability.

Similarly, for the same example, one can prove the sequence of random variable CDF's as $\lim_{n \to \infty} F_n(x)$ will converges to F(x), where F(x) is, where F(x) takes a value 0 till 1 and 1 onwards, it is going to be 1. Therefore, we can conclude the same example X_n converges to 1 in distribution.

That means, there are few a sequence of random variable may converge to more than one mode of convergence. So, that can be connected in 1 nice way that is suppose you have a sequence of random variable and the random variable X defined on the probability space (Ω, F, P) convergence in distribution and we have a convergence in probability.

(Refer Slide Time: 20:30)



Suppose, some sequence of random variable convergence in probability this implies convergence in distribution. The converse is not true it may be true for some few problems, in general not true.

Therefore, I am not writing the up and down. Converse is not true in general. Similarly, if you have a X_n converges to X in the first order moment, this implies the converges in probability; Converse is also not true in general here. Suppose, you have a convergence in rth moment that implies the r - 1 and so on till the first order moment.

But here, again converse is not true because of two reasons. The first order moment exist that does not mean that the second order moment is going to exist; even if it exists it does not imply the convergence takes place. Therefore, the converse is not true. Convergence in a rth moment exist, then you can go for tilde first order moment convergence that implies the convergence in probability that implies convergence in distribution.

Now, I am giving the connection with the convergence in almost surely that is X_n converges to X almost surely. This implies convergence in probability, you see the different direction. Almost surely convergence implies convergence in probability convergence in probability implies convergence in distribution; nowhere converse is true in general.

There are some additional condition we can mention for some sequence of random variable, then we can conclude the X_n converges in X in probability that convergence in almost surely also. But for that you have to make some additional condition. In general, the converse is not true for all the cases.

So, with this relation I am going to study the convergence or modes of convergence. In the next class we will discuss about the law of large numbers.

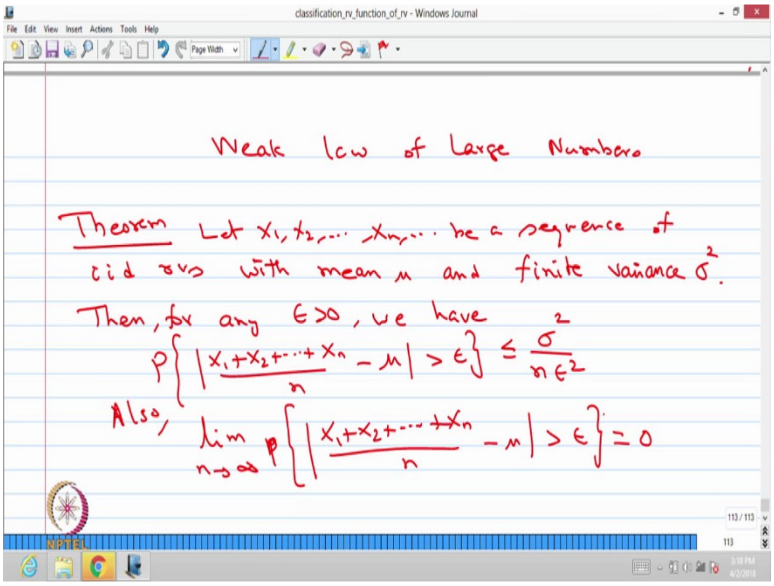
Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 44

So, we are in Limiting Distributions model. In this model we have already discussed modes of convergence. In that we have discussed four different modes of convergence. First is convergence in distribution, convergence in probability, convergence in rth moment, convergence almost surely.

In this lecture, we are going to discuss law of large numbers. In that we are going to discuss two types of law of large numbers; one is called Weak law of large numbers.

(Refer Slide Time: 00:43)



Then later, we are going to discuss strong law of large numbers.

Let me give the definition of weak law of large numbers. As a theorem, let $X_1, X_2, \dots, X_n, \dots$, be a sequence of iid random variables with mean μ and finite variance σ^2. That means, this sequence of random variable has at least second order moments.

Then for any ϵ greater than 0, we have $P\{\vee\frac{X_1+X_2+\ldots+X_n}{n}-\mu\vee ¿\epsilon\}$. This probability is always less than or equal to $\frac{\sigma^2}{n\epsilon^2}$. Also, $\lim_{n\to\infty}P\{\vee\frac{X_1+X_2+\ldots+X_n}{n}-\mu| ¿\epsilon|=0$ as n tends to infinity.

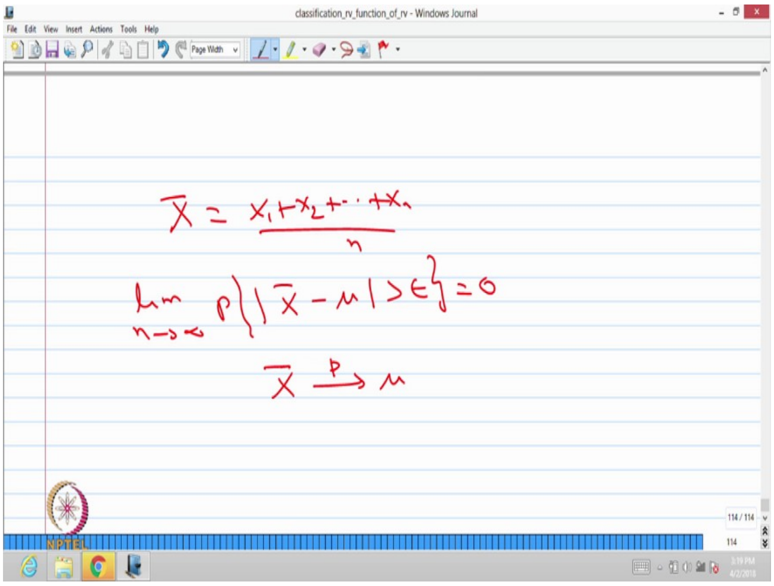
Then, we say that this sequence obeys weak law of large numbers. Here, large numbers means the sequence of random variables. When you have many random variables and if you create $P\{\vee\frac{X_1+X_2+\ldots+X_n}{n}-\mu| ¿\epsilon|$ will tends to 0 that is what this weak law of large numbers says.

For that you need a sequence of random variable should have a at least second order movement; that means, mean exist as well as the variance exist and we made the assumptions those random variables are iid random variables; that means, independent and identically distributed random variable.

Therefore, the mean is going to be same and the variance is going to be same. Then for any ϵ you can have $\lim_{n\to\infty}P\{\vee\acute{X}-\mu| ¿\epsilon|=0$

If this condition is satisfied, then we can conclude **t**his sequence of random variable obeys weak law of large numbers. Why it is called a weak law of large numbers? Because if you see the different modes of convergence, you can conclude, if you make a notation $\acute{X}=\frac{X_1+X_2+\ldots+X_n}{n}$.

(Refer Slide Time: 05:18)

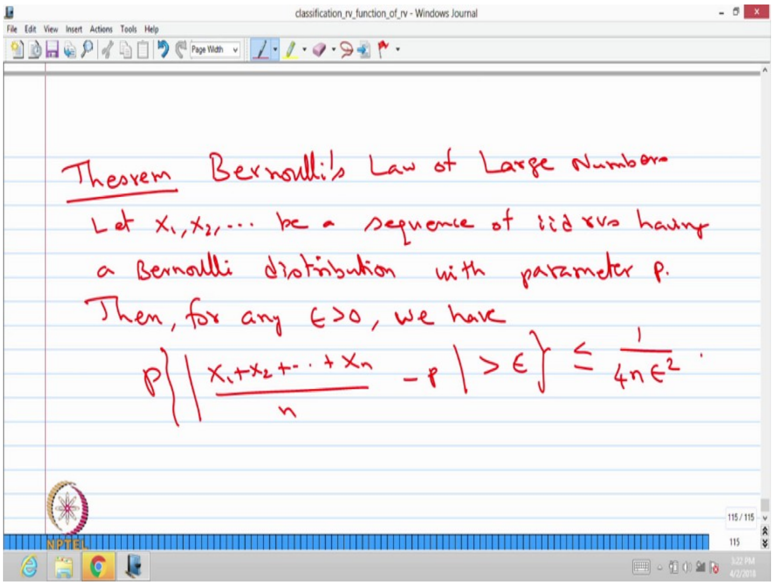


This result is nothing but $\lim_{n \to \infty} P\{\vee \acute{X} - \mu | ¿ \epsilon\} = 0$. If you see the definition of different modes of convergence, this is nothing but $\acute{X}$ *converges* ¿ μ in probability.

So, since here the convergence in probability we call it as this sequence of random variable obeys a weak law of large numbers. Whereas, when we are discussing a strong law of large numbers, those sequence of random variables convergence to some random variable and convergence in almost surely that is a strong law of large numbers; whereas, this one satisfies convergence in probability.

Therefore, it is called the weak law of large numbers. We are not going to give the proof of this where as we are going for one Bernoulli law of large numbers that is a special case of the weak law of large numbers for that we will provide the proof.

(Refer Slide Time: 06:50)

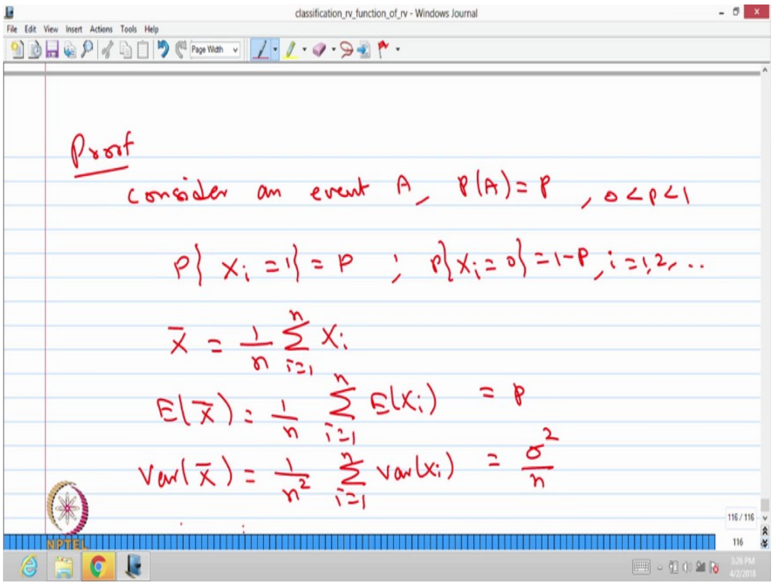


That is a next theorem that is Bernoulli's Law of Large Numbers.

Let $X_1, X_2, \ldots$ be a sequence of iid random variables having a Bernoulli distribution with parameter p. Then, for any $\epsilon > 0$, we have $P\{|\frac{X_1+X_2+\ldots+X_n}{n} - p| > \epsilon\}$. This probability of event is always less than or equal to $\frac{1}{4n\epsilon^2}$. This is a special case of the earlier theorem.

If you see the earlier theorem, the weak law of large numbers. We have a sequence of iid random variables with the mean μ and the finite variance, then we concluded as $\lim_{n\to\infty} \bar{X} = \mu$ and converges takes place in probability. In the Bernoulli's law of large numbers in addition to the previous theorem. We have introduced the distribution of each random variable that is Bernoulli distribution with the parameter p.

(Refer Slide Time: 09:30)



We are going to give the proof of this theorem as follows. Consider an event A whose probability is P(A) = p, where A is the success in each Bernoulli trial. A is the event in each Bernoulli trial and the P(A) = p, p is nothing but probability of success in each Bernoulli trail. Since, each X_i's are Bernoulli distributed; therefore, the $P\{X_i=1\}$; that probability is, that is P(A) that is a event that is probability is p and the p$P\{X_i=0\}$ that is 1 - p. This is for i is equal to 1, 2 and so on.

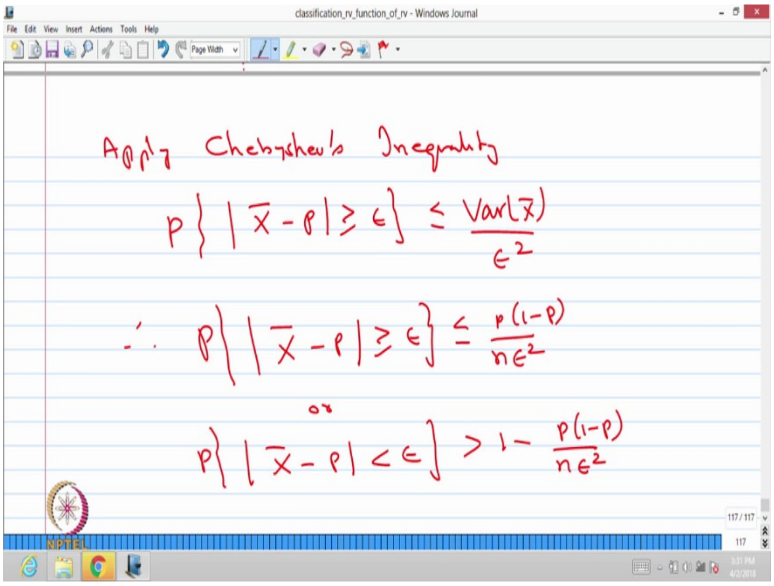
So, if you define a random variable $\acute{X}$ that is nothing but $\frac{1}{n}\sum_{i=1}^{n} X_i$. We can find μ and variance of this random variable. The mean of this random variable is $\frac{1}{n}\sum_{i=1}^{n} E[X¿¿i]¿$. Each one is Bernoulli distributed.

Therefore, the mean is going to be p; therefore, summation is n p. Therefore, $\frac{np}{n}$; therefore, it is going to be p. If you find out the var($\acute{X}$) that is nothing but $\frac{1}{n^2}$, all are iid random variables.

Therefore, it is $\sum_{i=1}^{n} var(X¿¿i)¿$. $var(X¿¿i)=\sigma^2¿$; therefore, it is going to be n$\sigma^2$when you make a summation. So, it is going to be $\frac{\sigma^2}{n}$. So, we got the mean and variance for the $\acute{X}$.

Now, we may not know the distribution of $\acute{X}$; whereas, we know the mean and variance of $\acute{X}$. Therefore, we can apply the Chebyshev's inequality.

(Refer Slide Time: 12:07)

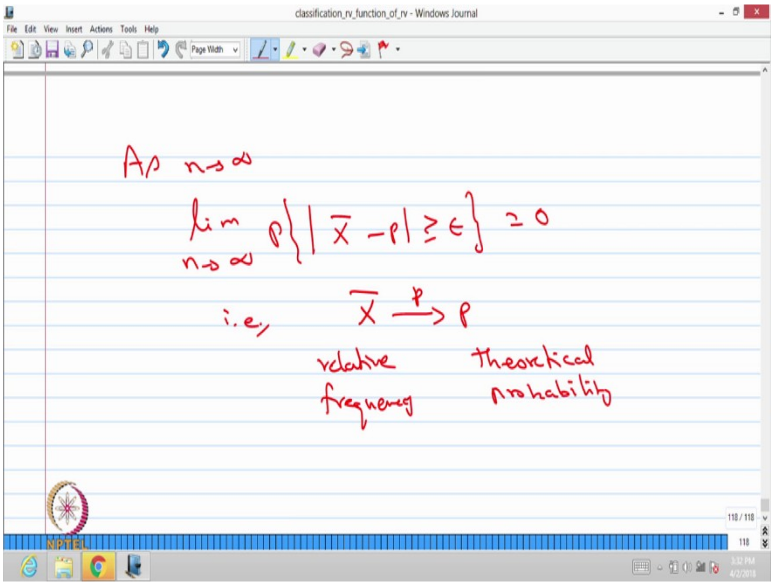


Apply Chebyshev's Inequality for the random variable $\acute{X}$. So, what the inequality says? The $P\{\vee \acute{X}-p|\geq \epsilon|\} \leq \frac{var(\acute{X})}{\epsilon^2}$.

So, just now we got $var(\acute{X})=\frac{\sigma^2}{n}$; therefore, $P\{\vee \acute{X}-p|\geq \epsilon|\} \leq \frac{\sigma^2}{n\epsilon^2}$. That means, the P $\{\vee \acute{X}-p|\geq \epsilon|\}$ has the upper bound $\frac{\sigma^2}{n\epsilon^2}$or we can write the $P\{\vee \acute{X}-p|¿\epsilon|\}$; that has the lower bound 1 - $\frac{\sigma^2}{n\epsilon^2}$. Since, $\sigma^2=p(1-p)$; therefore, it is p(1-p)/n.

Therefore, here also I can do the simplification, where variance of that is p(1- p). By applying the Chebyshev's Inequality $P\{\vee \acute{X}-p|\geq \epsilon|\} \leq \frac{var(\acute{X})}{\epsilon^2}$. We know that each X_i's are Bernoulli distributed and $var(\acute{X})=\frac{p(1-p)}{n}$ or the$P\{\vee \acute{X}-p|¿\epsilon|\}$> 1 - $\frac{\sigma^2}{n\epsilon^2}$.

(Refer Slide Time: 15:00)

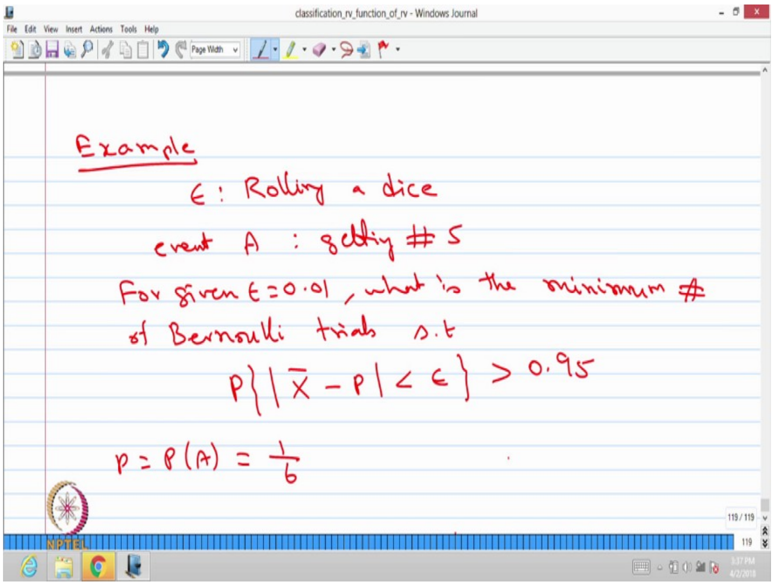


Here also we can go for as a n tends to infinity, the $\lim_{n \to \infty} P\{\vee \acute{X} - p| \geq \epsilon\} = 0$; that means, the $\acute{X}$ tends to p, convergence in probability. That means, for a larger n, for a Bernoulli distributed random variable $\acute{X}$, $\frac{1}{n} \sum_{i=1}^{n} X_i$ is nothing but the relative frequency.

So, the relative frequency converges to the theoretical probability; that is the theoretical probability. If you have independent Bernoulli trials for a finite n the relative frequency may deviate from the theoretical probability, but for a larger n that the relative frequency will converge to the theoretical probability and that convergence take place in probability.

Let us go for one simple problem, how to use the Bernoulli law of large numbers as an example.

(Refer Slide Time: 16:40)



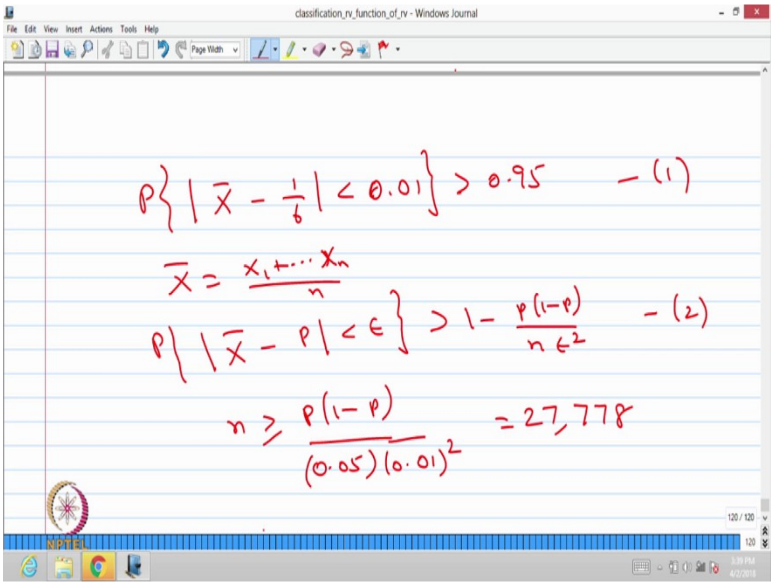
The random experiment is rolling a dice. For simplicity we assume that it is the fair dice. An event A is getting a number 5; getting number 5. Event A is nothing but getting a number 5. We are repeatedly rolling a dice countably infinite number of times and the question is for a given $\epsilon = 0.01$ what is the minimum number of Bernoulli trials such that the $P\{\vee \acute{X} - p|¿\epsilon\}>0.95$.

The random experiment is rolling a dice countably infinite number of times. The event A is getting a number 5 in each Bernoulli trial. The question is for a given ϵ, what is the minimum number of Bernoulli trials such that $P\{\vee \acute{X} - p|¿\epsilon\}>0.95$? That means, minimum how many number of times we have to roll a dice for getting minimum probability of 0.95 within the length of ϵ which is deviated from the p.

For this problem the p is nothing but the probability of success of a event A that is getting a number 5 in each Bernoulli trial that is 1 out of 6; I made it fair dice therefore, it is 1/6 that is p. So, we know p and we know the value of ϵ; therefore, you apply in the Bernoulli law of large numbers because we have iid random variables, each are having a Bernoulli distributed with a probability of success p is 1/6.

So, the question is sort of reverse problem, inverse problem finding the n such that this condition is satisfied.

(Refer Slide Time: 20:00)

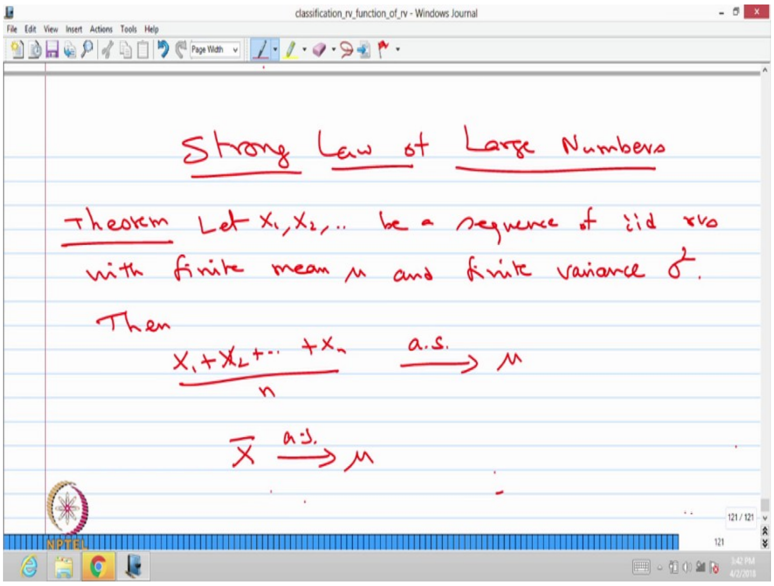


That means, $P\{|\bar{X} - \frac{1}{6}| < 0.01\} > 0.95$; if you simplify your $\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$; therefore, if you compare this with the definition of $P\{|\bar{X} - p| < \epsilon\} > 1 - \frac{p(1-p)}{n\epsilon^2}$.

So, now you compare equation number 1 with the 2; compare 1 and 2 we get $n \geq \frac{p(1-p)}{0.05(0.01)^2}$; where, p is 1 by 6. If you simplify you will get 27778, we are finding the nearest positive integer.

That means the n has to be minimum 27778 valid. That means, if you roll a dice minimum 27778 times we are going to attain the relative frequency deviation from 1/6 with the length of 0.01; probability of this event is going to be minimum point or the probability of this event is at least 0.95. So for that the number of trials needed is minimum 27778.

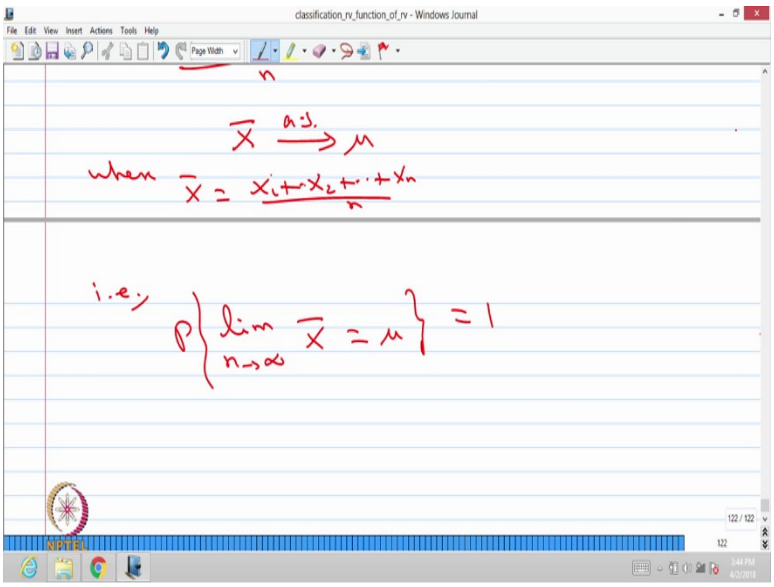
(Refer Slide Time: 22:33)



Now, we will move into the second law of large numbers that is a Strong Law of Large Numbers. Let me give the definition, in the form of theorem. Let $X_1, X_2, \ldots$ be a sequence of iid random variables with the finite mean mu and finite variance sigma square.

Then, X$\frac{X_1+X_2+\ldots+X_n}{n}$ will converge to μ almost surely. We can define $\frac{X_1+X_2+\ldots+X_n}{n}$ *as a* $\acute{X}$. So, $\acute{X}$, if I define $\acute{X}$ as this, $\acute{X}$ converges to μ almost surely.

(Refer Slide Time: 24:24)



Where, $\acute{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$; that means, if you have a sequence of random variables all are iid having at least second order movements, then adding all those random variables divided by n that is nothing but the average of n random variables converges to the mean of these random variable that is μ and that convergence takes place in almost surely.

That means, P$\{\lim_{n \to \infty} \acute{X} = \mu\}$ = 1. That means, if you collect the possible outcomes in which $\acute{X}(w)\, tends ¿ \mu ¿$and if you collect those possible outcomes, whose probability put together is going to be 1.

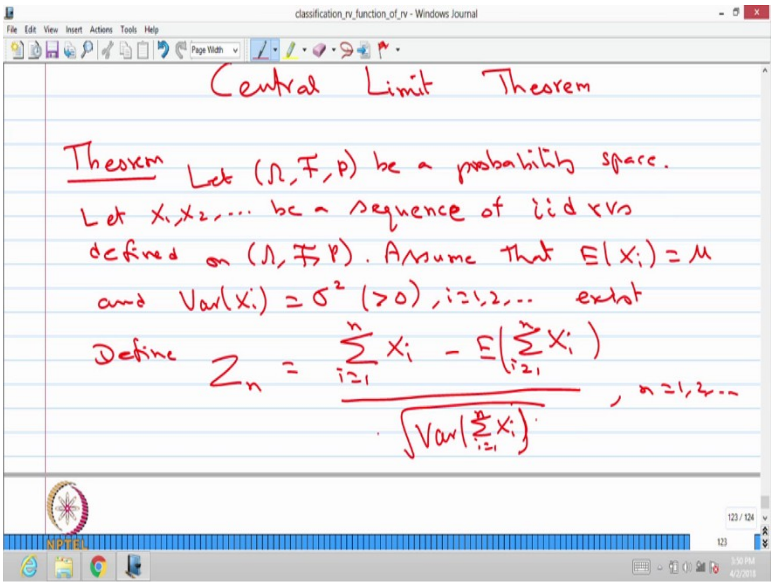
Then we can conclude $\acute{X}\, converges ¿ \mu ¿$almost surely. So, without proof we are giving the strong law of large numbers. And, why the word strong law of large number is here? The convergence in almost surely that is the strongest one; whereas, the weak law of large numbers involves convergence in probability. That is a weak law of large numbers whereas, the convergence in almost surely that is going to be calling it as a strong law of large numbers.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 45

Now, we will move into the third lecture that is called the Central limit theorem.

(Refer Slide Time: 00:06)



So, this is a very important result in probability that is central limit theorem which has the wide application in many real-world problems. Therefore, this theorem will be used again and again in many problems.

So, let me give the central limit theorem first, then I give the proof; then we will go for 1 or 2 examples of how to use the central limit theorem in the real-world problems. Let me give the theorem first. Even though, there are many versions over the central limit theorem, first we will get the easiest version; because it is an introduction to the probability theory and stochastic process course. If the course is advance probability theory course, then we can go for 2-3 levels of a central limit theorem.

So, here we will present only the simplest version of the central limit theorem; whereas, we will discuss how the complicated version in the central limit theorem after I give the proof of the simplest one. So, we will give the simplest version of the central limit theorem. Let (Ω, F,

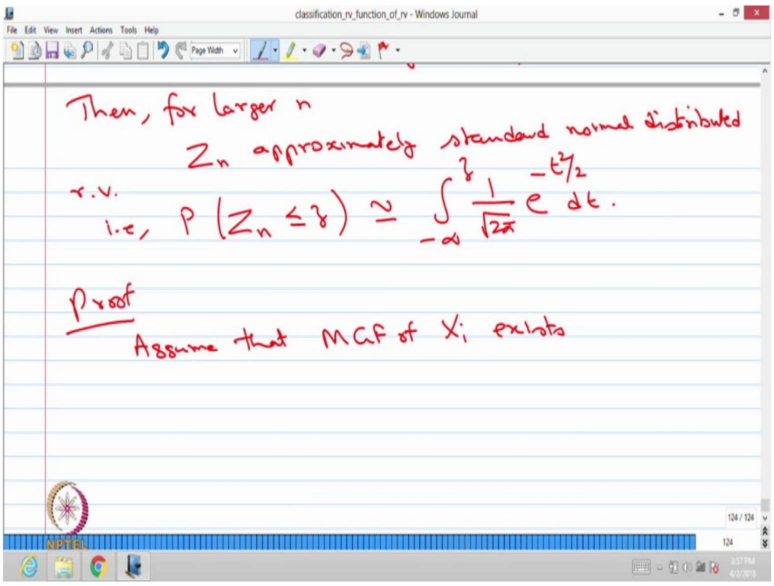
P) be a probability space, let $X_1, X_2, \ldots$ be a sequence of iid random variables defined on (Ω, F, P).

Assume that $E[X_i]=\mu$ and $var(X_i)=\sigma^2$ which is greater than 0 for i is equal to 1, 2,… exist; that means, we make sure that this sequence of random variables are iid as well as at least second order moment exist and the variance of each random variable is greater than 0.

Since I made it iid random variable, the $\sigma^2>$¿ 0 and also the finite quantity. And defining,

defining the new sequence of random variable I call it as a $Z_n=\frac{\sum_{i=1}^{n} X_i-\sum_{i=1}^{n} E[X_i]}{\sqrt{var(\sum_{i=1}^{n} X_i)}}$. I am defining a sequence of random variable for n is equal to 1 2 and so on.

(Refer Slide Time: 04:15)



What the central limit theorem says then, what the central limit theorem says, then for larger n, Z_n is approximately standard normal distributed random variable. Then for larger n, Z_n approximately a standard normal distributed random variable. That means, that is the

$$P(Z_n \le z) - ¿ \overline{\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt ¿}$$.

This is valid only for larger n that is very important and that to the CDF, CDF of a random variables Z approximately the $\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt$ that is nothing but the cdf of standard normal distribution. This is valid as long as X_i's are iid random variables defined and a probability space with at least second order moment exist and variance is greater than 0. And then, making a sum of random variables minus there mean divided by the standard deviation that is approximately a standard normal distributed random variable for larger n.

Indirectly whenever you have a normal distribution with the parameters mu and sigma square; by subtracting the mean divided by the standard deviation that becomes standard normal distribution. So, the same thing we are applying in the Z_n. The random variable is a sum of random variable that is a one random variable for fixed n minus their mean divided by the standard deviation; that means, this transformation is the transformation from normal distribution to the standard normal.

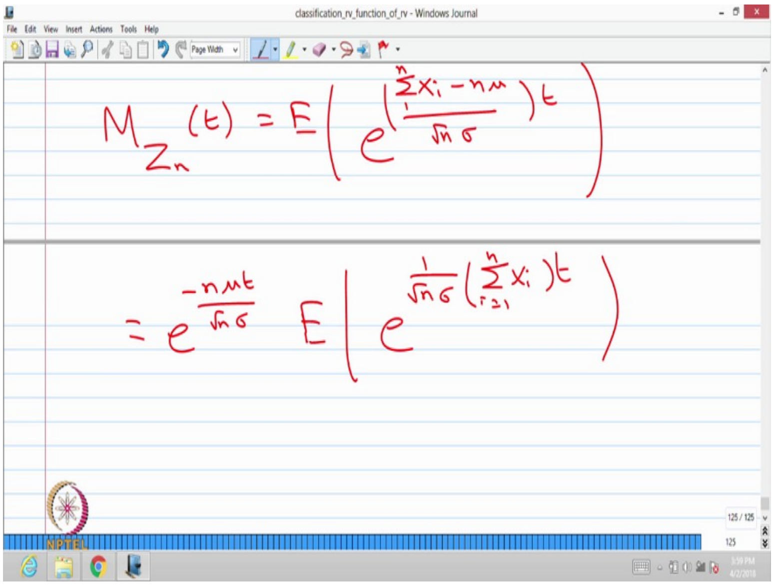
That means, indirectly when we say, when we say Z_n approximately a standard normal distributed random variable, indirectly what we are saying the sum of n random variable approximately a normal distributed random variable with mean, expectation of that random variable with the variance, variance of sum of random variable for larger n.

That is a meaning of Z_n approximately a standard normal distributed normal variable that is equivalent of a sum of random variable is approximately a normally distributed random variable with the mean is expectation of a sum of random variable and the variance is variance of sum of random variables. And here the assumptions are very important, it should be iid random variables with at least a second order moment exist.

Now, we will go for proof of this theorem. For the proof we will make the assumption that m g f of each X_i exist; even though for some random variable m g f may not exist and here we made the assumptions only at least second order moment exist, that does not mean that m g f for moment generating function of each X_i's exist.

We make the additional assumption of m g f exist, then later we will relax the m g f exist then we can give the proof of it. So, without loss of generality we assume that m g f of X_i's exist for all the random variable because all are iid random variables. With that assumption will give the proof, later we can relax this also.

(Refer Slide Time: 09:46)

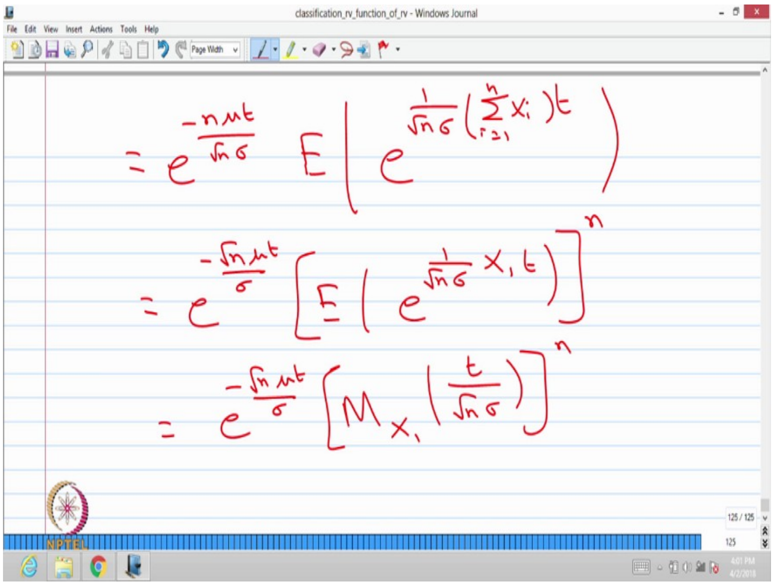


Let us go for finding out the m g f of Z_n as a function of t. Moment generating function for the random variables Z_n as a function of t that is nothing but $E[e^{(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma})t}]$. Since we made a all are iid random variables, their mean is going to be μ, n μ divided by variance of sum of random variables; each random variable variance is σ^2. Therefore, sum of random variables is $n\sigma^2$. Here, you need a square root of variance; therefore, $\sqrt{n}\sigma$.

So, this quantity is going to be the m g f of the random variable Z_n. This is possible as long as the m g f of X_i exist. Therefore, you made the assumptions m g f exist that is same as all the constant you can take it out. Therefore, it is going to be $e^{\frac{-n\mu t}{\sqrt{n}\sigma}} E[e^{(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}\sigma})t}]$

This is same as $e^{\frac{-\sqrt{n}\mu t}{\sigma}}\left(E\left[e^{\left(\frac{X_1 t}{\sqrt{n}\sigma}\right)}\right]\right)^n$.

(Refer Slide Time: 11:39)

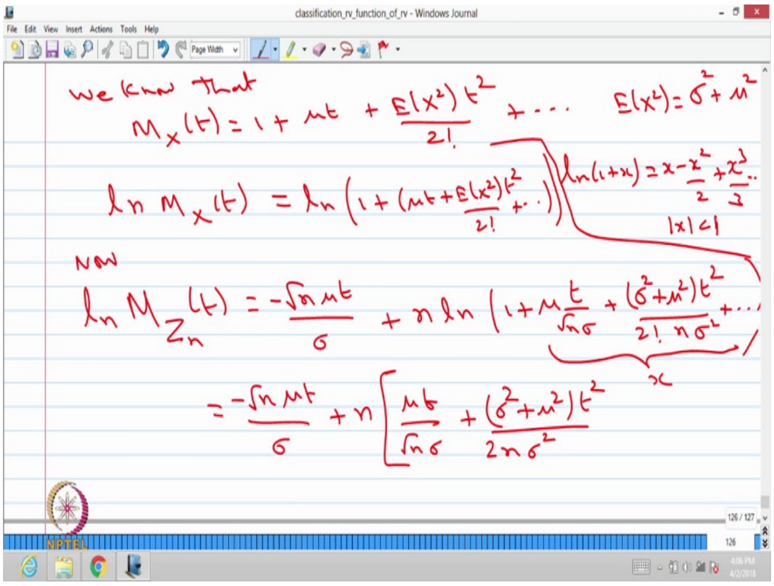


You can use the $E[e^{\sum_{i=1}^{n} X_i}]$ that is nothing but the all are iid random variable.

After getting the expectation you can raise it to the power n because all are independent as well as identical. That is same as $e^{\frac{-\sqrt{n}\mu t}{\sigma}} \left(M_{X_1}\left(\frac{t}{\sqrt{n}\sigma}\right) \right)^n$. This is nothing but m g f of the random variable X_1 instead of t, you can write $\frac{t}{\sqrt{n}\sigma}$, both are one and the same; whether you write $M_{\frac{X_1}{\sqrt{n}\sigma}}(t) \vee M_{X_1}\left(\frac{t}{\sqrt{n}\sigma}\right)$, both are one and the same; this power n because of identical.

Now, we need the expansion of m g f for any random variable; then, we can substitute that.

(Refer Slide Time: 13:18)



We know that $M_X(t) = 1 + \mu t + \frac{E[X^2]t^2}{2!} + \dots$. Again, you can write $E[X^2] = var(X) + E[X]^2$. Suppose var(x) is $\sigma^2, then \sigma^2 + \mu^2$. So, one can write $E[X^2] = \sigma^2 + \mu^2$; I am going to substitute little later by taking a logarithm of $M_X(t)$, I can use $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$ provided |x| < 1.

I can use this identity for the ln ($M_X(t)$) = ln($1 + \mu t + \frac{E[X^2]t^2}{2!} + \dots$).

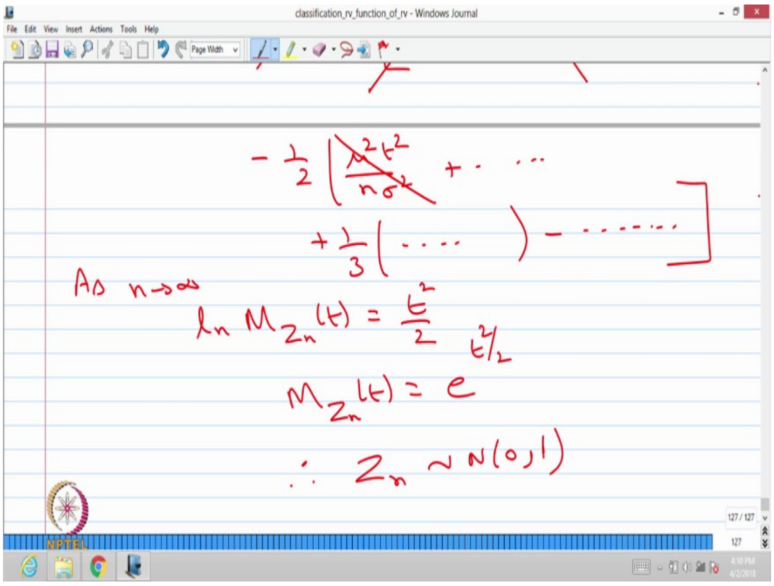
So, I can make it as the l n of 1 plus all the other term, I can make it as the sort of x as $\mu t + \frac{E[X^2]t^2}{2!} + \dots$. This I can keep it as a 1 + x. So, I have not substituted ln(1+x) now, I am just writing ln of the whole series as the 1 plus remaining terms as the x.

Now, I am going to apply the same logic for the $M_{Z_n}(t)$; that means, now ln($M_{Z_n}(t)$) that is going to be when you take a logarithm, it becomes $\frac{-\sqrt{n}\mu t}{\sigma}$. Then, the remaining terms with the power; therefore, it becomes n power n becomes n ln($1 + \mu\frac{t}{\sqrt{n}\sigma} + \frac{(\sigma^2+\mu^2)t^2}{2!\,n\sigma^2} + \dots$). Here, t is replaced by $\frac{t}{\sqrt{n}\sigma}$ plus $E[X^2] = \sigma^2 + \mu^2$. This is going to be

$\frac{-\sqrt{n}\,\mu t}{\sigma}+n\ln\left(1+\mu\frac{t}{\sqrt{n}\sigma}+\frac{(\sigma^2+\mu^2)t^2}{2!\,n\sigma^2}+\ldots\right)$.Now I am going to apply ln(1+x) that is n(x - $\frac{x^2}{2}+\frac{x^3}{3}-\ldots$). So, x is going to be this.

So, the first terms in the x = $\mu\frac{t}{\sqrt{n}\sigma}+\frac{(\sigma^2+\mu^2)t^2}{2!\,n\sigma^2}$. I am not going to write other terms of x limit as it is, whereas now I am going to write -$\frac{x^2}{2}$ terms that is - $\frac{1}{2}$.

(Refer Slide Time: 17:45)



In the x^2 also I am not going to write $x^2$of all the terms, I am going to write the $x^2$of only first term that is $\frac{\mu^2 t^2}{n\sigma^2}$; all the other terms I leave it as it is. There is a reason behind that; I am not going to write other terms of x^2.

Similarly, I am not going to write any terms for the x^3, only I write 1 by 3 all the other term as it is. Like that there are some more terms some more terms for the expansion of ln(1+x). This is going to be close bracket. The reason is as n tends to infinity, even though I use a word for larger n here, we are going for as n tends to infinity. The n in the numerator and many terms in the n in the denominator that canceled and all the other terms will be in the form of 1/n; not only that this one and this one cancel. Whereas, the $\frac{(\mu^2)t^2}{2!n\sigma^2}$ this one with the first term here that cancels.

So, the left out is $\frac{(\sigma^2)t^2}{2!n\sigma^2}$ that will be cancelled with n in the numerator. So, you will have only $\frac{(\sigma^2)t^2}{2\sigma^2}$, σ^2also cancel. So, you will left out with $\frac{t^2}{2}$. Even though we have many more terms as n tends to infinity all the other terms vanish. So, you will have as n tends to infinity $\ln(M_{Z_n}(t))$ = $\frac{t^2}{2}$ all the other term vanishes as n tends to infinity. Now, I am taking exponential both side; that means, $M_{Z_n}(t)=e^{\frac{t^2}{2}}$. If you recall the generating function for the standard distributions we have discuss for many discrete type random variables.

Similarly, we have discussed continuous type random variables m g f. So, if you compare the m g f of this with the m g f of standard distribution, you can conclude by using the uniqueness theorem of two different m g f s are same for all t, then both the random variables are identically distributed. So, you can conclude the Z_n is standard normal distribution. So, this is valid for n tends to infinity. That means, for larger n the Z_n approximately a standard normal distribution that is a proof. In this proof we have made assumption of m g f exist.

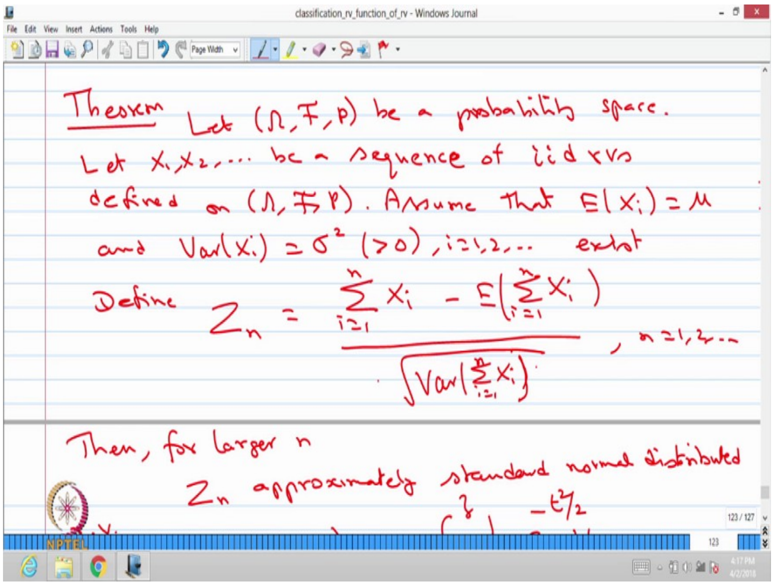
Now, we can see what could be the proof or how the proof goes when you do not have a assumption of m g f. The similar derivation I can go for characteristic function. So, the characteristic function of Z_n¿t) that is going to be E[e^{itZ_n}], where t is replaced by i t, where i is square root of minus 1. For that I do not need any assumption because the characteristic function exist for all the random variables; therefore, the characteristic function for Z_n exist. So, I can directly compute the characteristic function of Z_n.

In this result wherever the t, I have to replace by i times t that is going to be the derivation of characteristic function. So, if I do the same derivation everything goes in the same fashion because I keep iid random variables mean is mu variance is sigma square and so on. Therefore, wherever there is a t, it will be replaced by i t. So, that will be cancelled wherever there is a t square that is going to be minus t square because it is going to be $i^2t^2, i^2=-1$.

After you do the simplification till as n tend to infinity, you will get the answer $\frac{-t^2}{2}$ for the l n of characteristic function of Z_n; that means, the characteristic function of Z_n is going to be $e^{\frac{-t^2}{2}}$ that is there result for the characteristic function for standard normal distribution, then we can conclude also Z_n is approximately a standard normal distribution.

So, whether we made the assumption m g f or not the derivation is almost similar way to conclude that it is approximately a standard normal distribution. I said I am going to discuss the little higher versions of the central limit theorem.

(Refer Slide Time: 23:42)



Yes, see the theorem carefully I have made iid random variable. Suppose, if it is not identical distributed then you can find what are all the changes; that means, if each X_i's are not a identically distributed, then their mean will be μ_i's variance will be σ_i^2; that means, each one may have a different means. Still you can apply the theorem because Z_n is going to be sum of random variable minus their μ.

So, whatever the mean μ_i's, you add all them μ_i's, find out the summation of μ_i's; that is going to be the expectation. In this theorem, when they are identical it becomes n μ if they are not identical. Then it becomes $\mu_1+\mu_2+\ldots+\mu_n$. Similarly, the denominator here it is a $\sqrt{n}\,\sigma$, but if they are not identical, then you will have a $\sqrt{\sigma_1^2+\sigma_2^2+\ldots+\sigma_n^2}$.

Still the derivation goes, but we cannot apply the power n. We cannot apply the power n the way we have done it here because of identical we got power n. So, when you go for derivation for non identical distributed random variable you have individual m g f in the product form.

So, when you take a logarithm and so on, the expression will be huge. The process of derivation may be tedious, but still as n tends to infinity you can conclude the same result. The derivation may be very complicated when they are non-identical distributed, still we can go for it the same derivation.

One more observation, here we have used the independent random variable in finding the square root of variance of sum of random variables. Since all the random variables are independent the variance of sum of random variable is nothing but the individual variances summation. If they are not independent, then you have to go for adding the covariance of any two random variables.

So, since we mean the assumption there, independent random variable, we are finding the individual variance, then we are sum it up; that is going to be the variance of sum of random variables. Otherwise you have to use the covariance of any two random variables; that means, we can relax instead of they are independent random variable you can make the assumptions all the random variables covariance of any two random variables is 0, that is enough.

You do not need independent assumption. Independent is a strongest assumption comparing to the covariance of any two random variables are going to be 0 because the covariance of any two random variable 0, that does not imply they are independent. But if two random variables or some random variables are mutually independent, then the covariance of any two random variables are going to be 0.

So, here in this theorem, I made a strongest condition; therefore, this is the simplest version of central limit theorem. Whereas, we can go for covariance of any two random variables are 0 that is enough to use the central limit theorem. One more observation over this central limit theorem, why this is a used in many situations?

You see the theorem very carefully we have not used any distribution for random variables Xi's and we have used the only the mean and variance of random variables and assumption of independent nothing else. Because of that this theorem is used in many real-world problems;

that means, many real-world problems many random variables which we have created, those random variables we may not know the distribution of that. We may not know the distribution of those random variables, but we may know the mean and variance as a number.

We may know mean and variance of those random variables, even they are dependent or the dependency maybe very-very minimal or we can ignore the dependency or we can make the usage of those random variables or independent or in the lighter sense we can use the concept of covariance of those two random variables are 0 with that assumption we can use this theorem. So, the big advantage of this theorem is there is no assumption over the distribution or we do not need the distribution of each X_i's; we need only the mean and variance.

Therefore, we can use this theorem to find out the probability of event using a standard normal distribution by approximating this random variable as a standard normal distribution. That means, whatever be the distribution of those random variables. Once we sum it up by subtracting there mean divided by the standard deviation for larger n we can always approximate immaterial of whether it is a discrete type random variable or continuous type random variables, as long as they are independent random variable, that can be approximated with a normal distribution; by normalizing it can be approximated with a standard normal distribution. Therefore, we use this theorem quite a lot in many real-world problems.

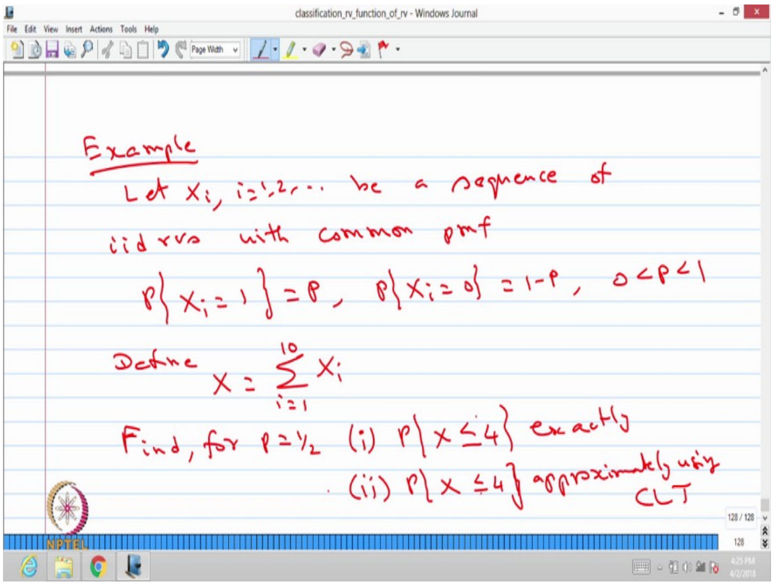
Now, let us go for a few examples how one can use the central limit theorem.

Introduction to Probability Theory and Stochastic Processes

Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 46

(Refer Slide Time: 00:03)

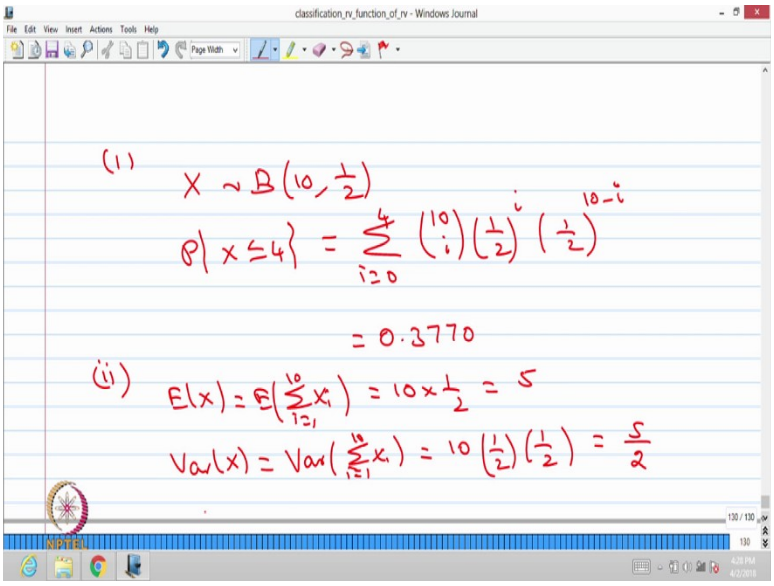


As a first example let us consider let X_i be a sequence of iid random variables with common probability mass function that is $P\{X_i=1\}$ = p and $P\{X_i=0\}$ = 1 - p, where p lies between 0 to 1. It is basically Bernoulli trials you have a sequence of random variable with the common probability mass function.

We define a random variable X which is sum of 10 random variables. The question is find for $p = \frac{1}{2}$; find for $p = \frac{1}{2}$, find $P\{X \le 4\}$ exactly. The second one $P\{X \le 4\}$ approximately using central limit theorem, CLT.

Let us go for finding this probability exactly, the first one, X is sum of first 10 random variables.

(Refer Slide Time: 02:15)

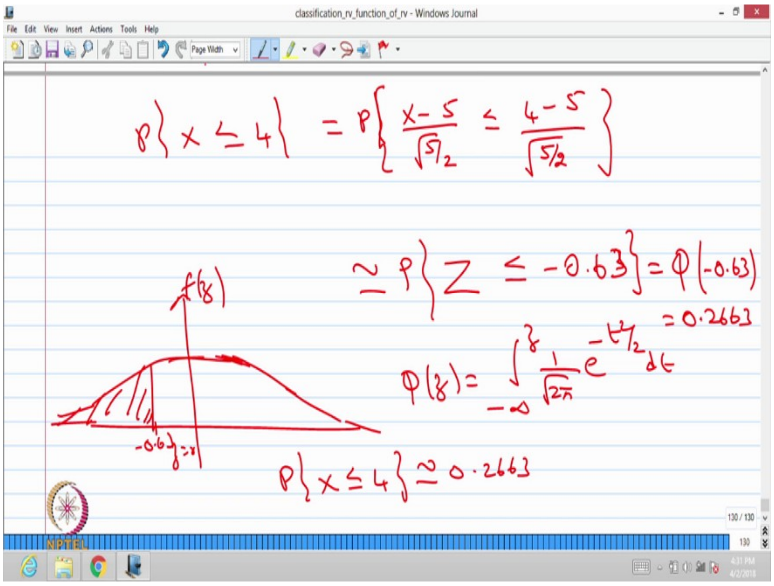


Each random variable is Bernoulli distributed. Therefore, the X is going to be Binomial distributed with the parameters 10, $\frac{1}{2}$. Each one is the Bernoulli distributed with the parameter $\frac{1}{2}$. Therefore, all are n such independent Bernoulli distributed random variable. Therefore, X is binomial distributed random variable with the parameters 10, $\frac{1}{2}$.

So, now you can find out the P{X ≤ 4} = $\sum_{i=0}^{4} {}^{10}C_i p^i (1-p)^{10-i}$. If you do the simplification, you will get the answer 0.3770. This is by finding the probability exactly because you know the distribution of X. The second part, finding the probability approximately using CLT; for that we can use the mean and variance of X without using the distribution of X.

So, if you want to find out the P{X ≤ 4} approximately, first we will find out what is the expectation of X that is nothing but expectation of sum of 10 random variables. So, the mean is going to be p. So, 10 p. So, 10 $\frac{1}{2}$, that is equal to 5. Similarly, you can find the variance of X, that is going to be variance of sum of random variables that is going to be 10 p(1-p); that is equal to $\frac{5}{2}$. So, to apply the central limit theorem, we need those random variables to be iid and mean and variance should be known, then you can find out the probability approximately using a central limit theorem.

(Refer Slide Time: 05:00)



So, the $P\{X \le 4\}$ = $P\{\frac{X-5}{\sqrt{\frac{5}{2}}} \le \frac{4-5}{\sqrt{\frac{5}{2}}}\}$. That is approximately probability of I am replacing $\frac{X-5}{\sqrt{\frac{5}{2}}}$ approximately with the standard normal distribution, less than or equal to this is by the central limit theorem. The first step $P\{X \le 4\}$ = $P\{\frac{X-5}{\sqrt{\frac{5}{2}}} \le \frac{4-5}{\sqrt{\frac{5}{2}}}\}$.

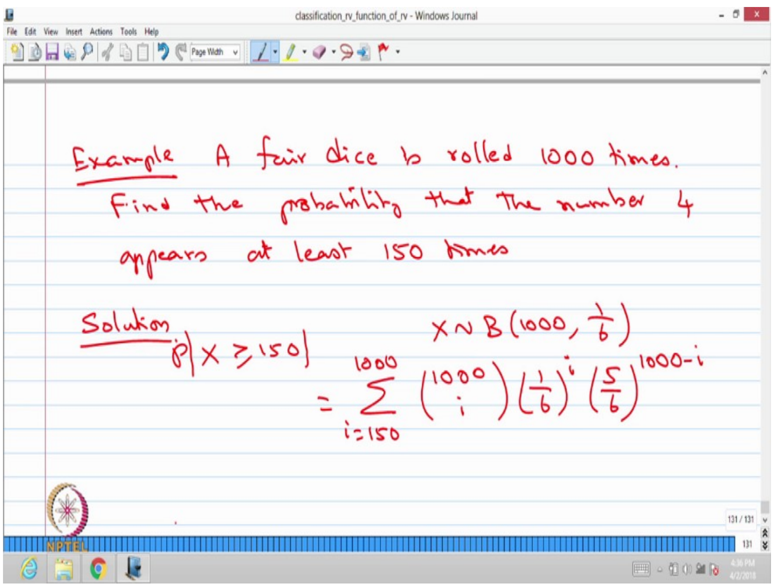
After subtracting the mean and the standard deviation that is approximately a standard normal. I am using this approximation symbol, that is less than or equal to $\frac{4-5}{\sqrt{\frac{5}{2}}}$ that is going to be -0.63. That is nothing but when you have a standard normal distribution probability density function z is equal to 0 and this is the probability density function of Z. So, -0.63

somewhere suppose this is a point -0.63. So, this probability that is nothing but we usually we make the notation φ(z) = $\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt$.

So, this is nothing but φ(-0.63). So, from the table 1 can find what is the φ(-0.63), when (z) = $\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt$. This value is going to be 0.2663. So, when you find out approximately P{X ≤ 4} that is approximately 0.2663 by using a central limit theorem; whereas, by getting the exact value you are getting 0.3770. You are getting exact answer because you know the distribution of this random variable X. Many times when you had the random variables you may not be able to easily get this distribution of the random variable. In that case in that case one can always apply the central limit theorem to find out the approximate probability, not the exact probability.

So, exact probability is possible only if you know the distribution; whereas, if you know mean and variance and all the random variables are independent, need not be identical still you can go for finding the approximate probability using the central limit theorem.

(Refer Slide Time: 08:44)



We will discuss one more problem. Example 2, A fair dice, a fair dice is rolled 1000 times. The question is to find the probability, find the probability that the number 4 appears at least 150 times. Find the probability that the number 4 appears at least 150 times.

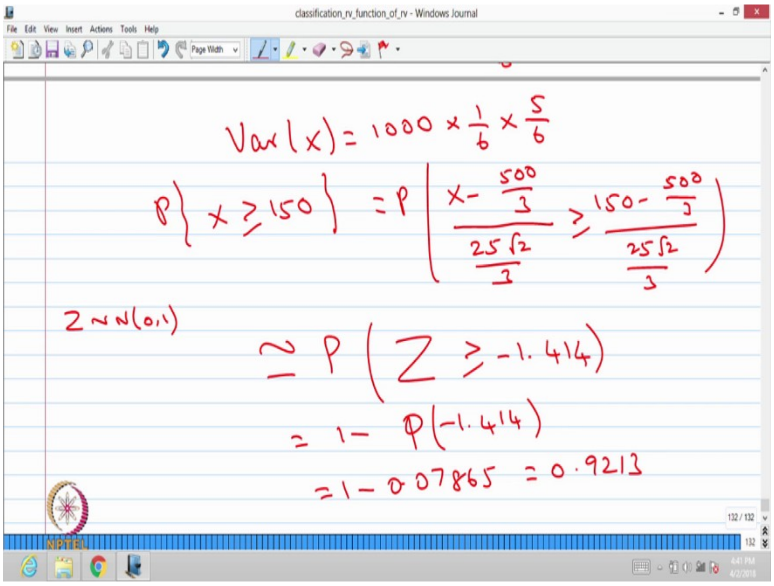
Here, we didn't say that use the central limit theorem, but we can apply the central limit theorem because of the difficulty while getting the answer for this or the computation is very difficult; therefore, we go for the central limit theorem.

Let X denotes the number of times the number 4 is obtained and the question is; what is the probability that X is going to be greater than or equal to 150? X is the number of times the number 4 is obtained and the question is; what is the $P\{X \geq 150\}$?

Since, a fair dice is rolled 1000 times independently; therefore, you can conclude X follows sum of 1000 Bernoulli distributed random variable with the parameter of getting a number 4 probability is $\frac{1}{6}$. Therefore, the parameter is 1000, $\frac{1}{6}$. The n is 1000 and the probability of success p is $\frac{1}{6}$. By computing $P\{X \geq 150\}$ is nothing but by using this formula that is $\sum_{i=150}^{1000} {}^{1000}C_i (\frac{1}{6})^i \left(\frac{5}{6}\right)^{1000-i}$. To get this result, you need factorials which involve which comes from the ${}^{1000}C_i$ which is very difficult or the negation is $1 - \sum_{i=0}^{140} {}^{1000}C_i (\frac{1}{6})^i \left(\frac{5}{6}\right)^{1000-i}$ and both are tedious calculation.

Therefore, we are going for getting the approximate probability instead of exact probability. That means we would not use the binomial distribution for the random variable X; whereas, we will use only the mean and variance of the random variable X, not the distribution. Because if we use the distribution, you can get the exact probability; but computationally it is not feasible. So, we apply the central limit theorem. Apply central limit theorem.

(Refer Slide Time: 12:51)



For that you need a mean of that random variable. Mean since it is a binomial distribution, the mean is going to be n p or even you can use the mean of each Bernoulli distributed random variable then you can just multiply by 1000. So, that is going to be 1000 $\frac{1}{6}$. This is going to be the mean and you can get the variance of X that is going to be 1000 p(1-p). So, this is going to be the variance of X.

Now, I can apply the CLT that is P{X ≥ 150} = $P\{\frac{X-\frac{500}{3}}{\frac{25\sqrt{2}}{3}} \geq \frac{150-\frac{500}{3}}{\frac{25\sqrt{2}}{3}}$. This is same as approximately probability of Z; where Z is a standard normal distribution greater than or equal to. So, this quantity is -1.414; that is same as 1 - φ(- 1.414); where the φ is defined in the same way in the earlier problem.

So, if you see the table for the φ(- 1.414), you will get the answer that is 1 - 0.07864. So, this is 0.9213. In this problem, we are not able to get the exact probability even though we know the distribution. Even though we know the distribution of X is binomial distribution computationally it is not possible, but by applying the central limit theorem we are able to find out the $P\{X \geq 150\}$. That is the advantage with the central limit theorem.

As long as at least a second order moment exist and the random variables are independent, need not be identical and sum of random variable for larger n approximately normal distribution. In other words, sum of random variables their mean divided by the standard deviation will be approximately standard normal distribution by using central limit theorem.

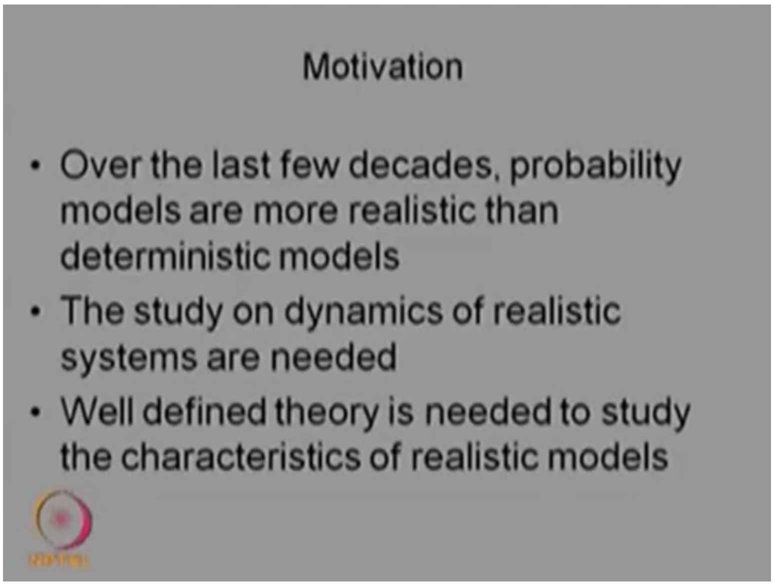
So, with this we are completing this model of limiting distributions with the three lectures; first lecture is modes of convergence. Second one is law of large numbers; in that we discuss the weak law of as well as the strong law of large numbers. Then finally, we discuss the central limit theorem with the proof and two examples.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Module - 09
Introduction to Stochastic Processes (SPs)
Lecture – 47

Our Lecture is Stochastic processes. So before we move into the stochastic process, I am going to give what is the motivation behind the stochastic process.

(Refer Slide Time: 00:27)



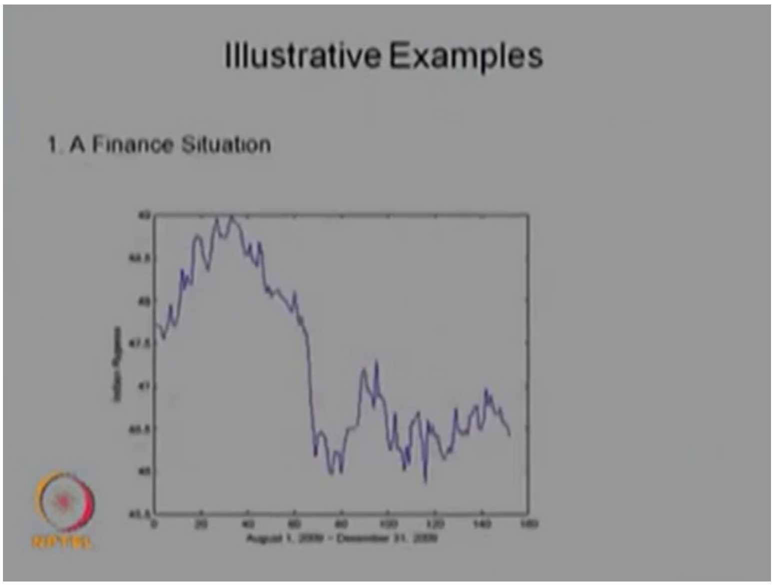
When we see the last few decades problems, more of the probability models are not deterministic that means you need more probability theory to understand the stochastic to understand the system then only you can study the dynamics of the model. If you see the, if you want to study the dynamics of the system, then you need more probably theory.

So the simple probability theory may not be enough to study the more study on the realistic system. The way this realistic system behaves in a very dynamical way, it is not easy to capture everything through the probabilistic or usual probability models. That means you need more than the probability theory to understand the system or to study the system in a well behaved way. For that one of the important thing is stochastic process.

It deals about the collection of a random variable, so that you can study the dynamics of the system in a better way. Even though I am giving very light way of saying the collection of random variable first we should know how the random variable can be defined, so that you can study the collection of random variable in a better way. So for that we are going to spend few examples through that how the more realistic models need more probability theory other than the usual probability theory.

So that the stochastic process definition and those things I am going to cover at a later part.

(Refer Slide Time: 02:09)



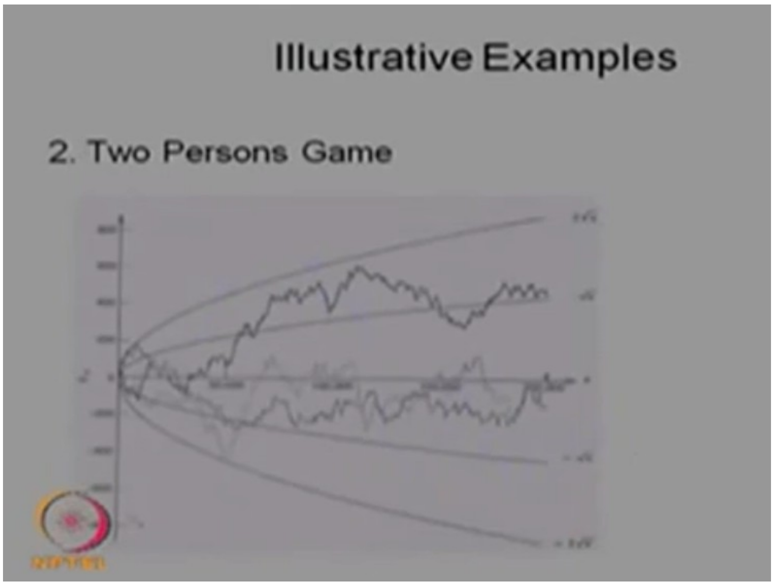
First let us see the first example, that comes in the finance situation. This is the actual data which captured over the period of time from August 01, 2009 to December 31, 2009 of what is the current price of one US dollar in Indian rupees. So if you see the graph you can make out August 1st, 2009, the price of one US dollar was Rs. 47.57 or Rs. 47.58.

And if you see the dynamics over the years, over the days from August 01st, 2009 to till December 31st, 2009 it keeps on changing and it takes some values higher and after that it goes down and it fluctuates and so on. So this is the actual data which we have captured. And from that our interest will be what could be the US dollar price after sometime. If I know till today what is the price, my interest would be what would be the price after 1 or 2 days or after one month or after 6 months.

That means, I should know how the dynamics keep moving over the days and what is the hidden probabilistic distribution is capturing over the time, so that I can identify what is the distribution behind that. Therefore, I can study the future prediction; I can study the dynamics of this particular model in a much better way. That means I need what is the background or what is the hidden distribution playing or hidden distribution which causes the dynamics of the system.

After identifying what is the distribution, my interest could be, what could be the some other moment over the time. That means what would be the average value or what could be the second order moment if it exists and so on, that can be obtained if I know the actual distribution in the underlying model.

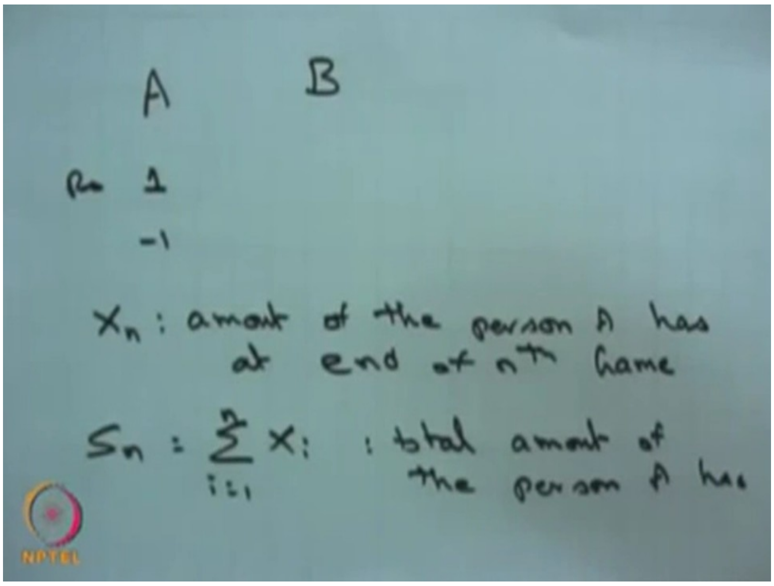
(Refer Slide Time: 04:31)



If I see the second example, I am just changing into another model in which there are two people playing a game. The person A and person B.

Whenever the person A wins he gets rupees one.

(Refer Slide Time: 04:47)



Suppose the person B wins then he will get the one rupee and at the same time person A loses one rupee in the same way and the play is keep going. Suppose you make the random variable as $X_n$is the amount of the person A has at the end of nth game. If you make out the random variable X_n for the person A has the amount at the end of the nth game.

The way the game going on, the value of the X_n keep changing and if you make out another random variable. $S_n = \sum_{i=1}^{n} Xi$.

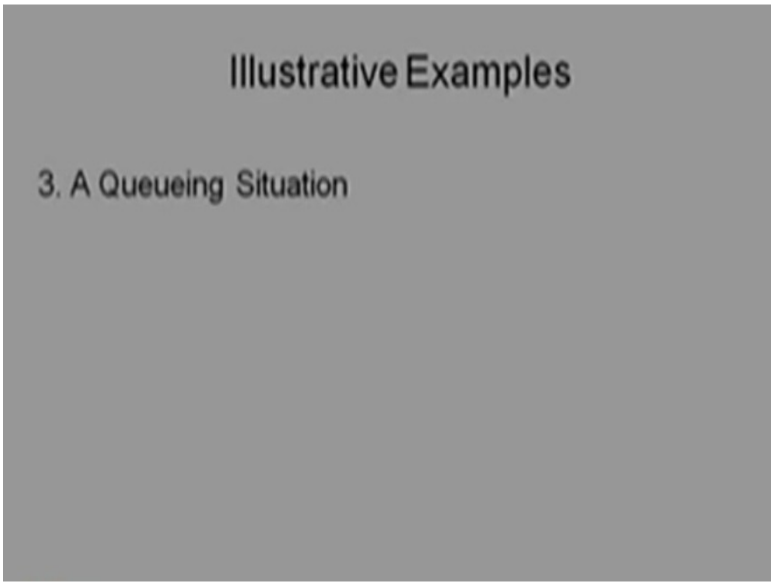
This gives what is the total amount of the person A has. The diagram in which the S_n gives, what is the way the dynamics goes and over the n. And if you see the diagram you can make out the whole dynamics goes.

How the game is going on in the first few games, accordingly it changes the positive side or it goes to the negative side. And if the n goes large then the dynamics of the S_n over the n will keep changing over the time and you will get the realization of the S_n over the time. And here I have given 3 different realizations and this diagram is taken out from the book by U.N. Bhatt. The title of the book is elements of applied stochastic process.

So this is one of the motivations behind stochastic process. And from these our interest will be after the, what is the distribution of Sn at any n and also as n tends to infinity what could be the distribution of S_n. That means you need the distribution of the random variable and also you need what could be the distribution as n tends to infinity or the limiting distribution of S_n.

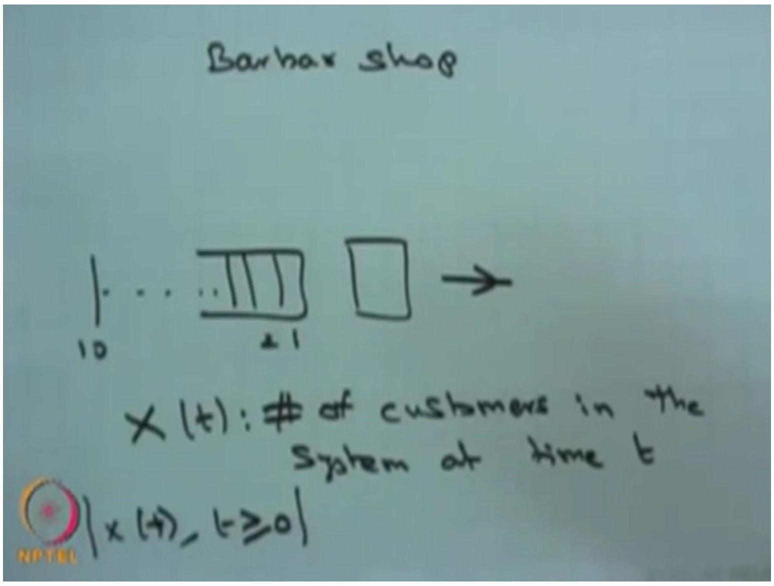
If you know the distribution, then you can get all other moments for different n as well as the asymptotic behavior of the random variable S_n.

(Refer Slide Time: 07:34)



Next I will move into another example in which it is the queuing situation. The queuing situation, here I have taken it as, taken a simple example, that is a barber shop example.

(Refer Slide Time: 07:46)



In which there is only one barber shop person and who does the service for the people whoever entering into the barber shop. And there are only a limiting capacity in which there is a maximum 10 people can stay in their barber shop and one person will be under service.

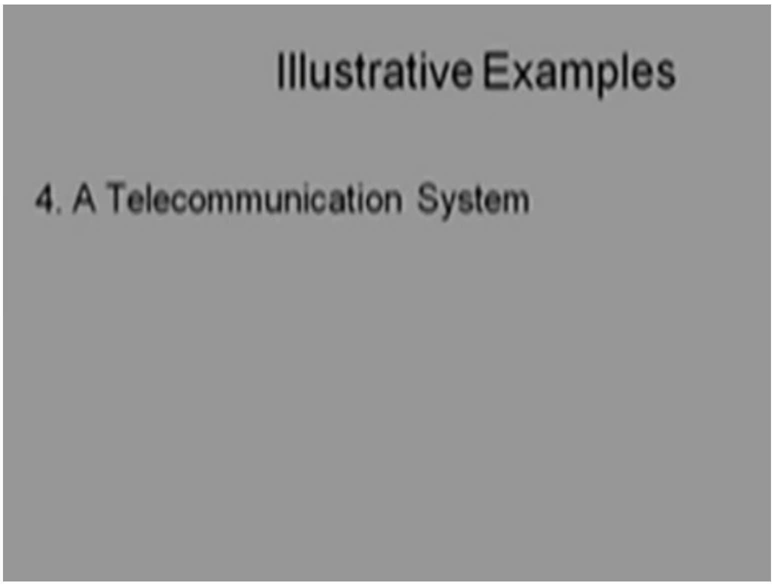
Once the service is over and the system will be, the customer can leave the system. At any time, maximum 10 people can be in the barber shop and only one person is doing the service for the customer who enter into the barber shop. Suppose you take the random variable as X(t) is the number of customers in the barber shop or in the system at time t the way the dynamics goes; the possible values of X(t) will be starting from 0 to n.

To study this system, you need what is the way the people or the customers are entering into the system and what is the way the service is going on for the customers and what is the discipline in which the customers are getting served also. Our interest will be, suppose we have the capacity of 10, what could be the waiting time, whenever the customers are entering into the system.

My interest will be one is how to reduce the waiting time on average, this is the customers point of view. As the barber shop point of view how much I can get more revenue, that means how I can increase the capacity of the systems so that I can make more profit over the time. That means if I know the dynamics of X(t) over t where t is varying from 0 to infinity, I can understand the system over the time as well as I can whatever the probabilistic measures or whatever the other measures average number of customers or average waiting time and so on.

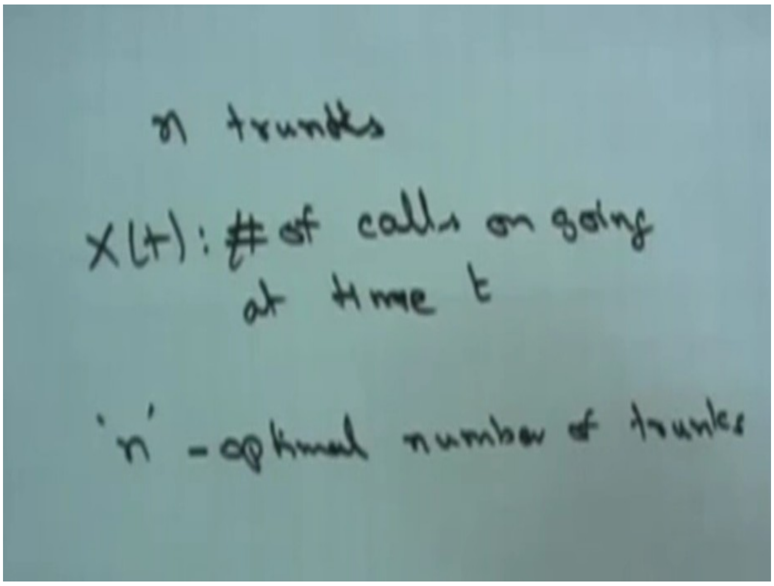
I can find out using this type of random variable. So later we are going to say, this is going to be one of the stochastic process for this example.

(Refer Slide Time: 10:34)



Next I am going to consider the fourth example as the telecommunication system.

(Refer Slide Time: 10:42)



Suppose you think of a system in which you have n trunks are there. Trunks are nothing but it is a maximum number of calls will be allowed at anytime. Whenever a call entering into the system and you have given one trunk to the call and at the end of the call is over the trunk will be back. So you have a telecommunication system in which n trunks are available at any time not at any time just n trunks available.

Suppose I make a random variable X(t) as the number of calls on going, at time t, see here also the dynamics of X(t) is going to be keep changing from 0 to n over the time and my interest will be how I can do the service such a way that more calls will be entertained as well as how I can find out the optimal n such a way that what is the optimal number of trunks such that I can minimize the waiting time or I can maximize the revenue.

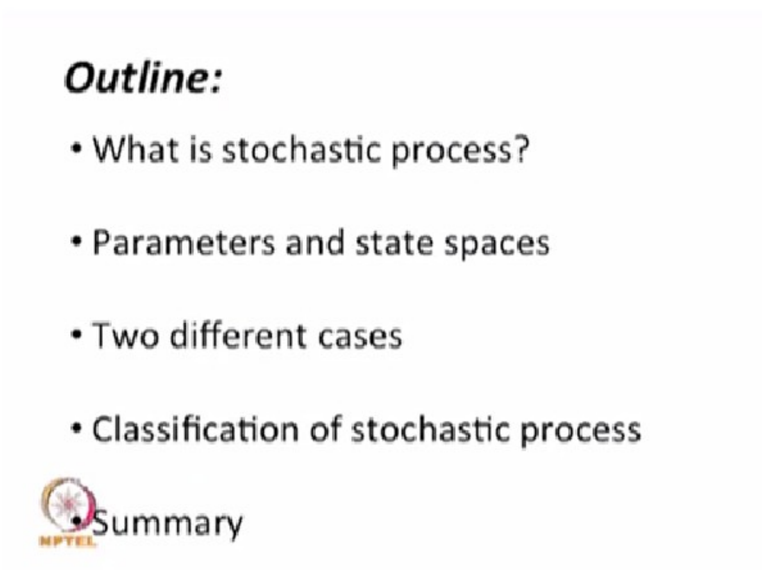
So this is also one of the problems which we come across in the daily life and so on. So my interest is to introduce the stochastic process so that I can study this type of system in a better way.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Module - 09
Introduction to Stochastic processes (SPs)
Lecture – 48

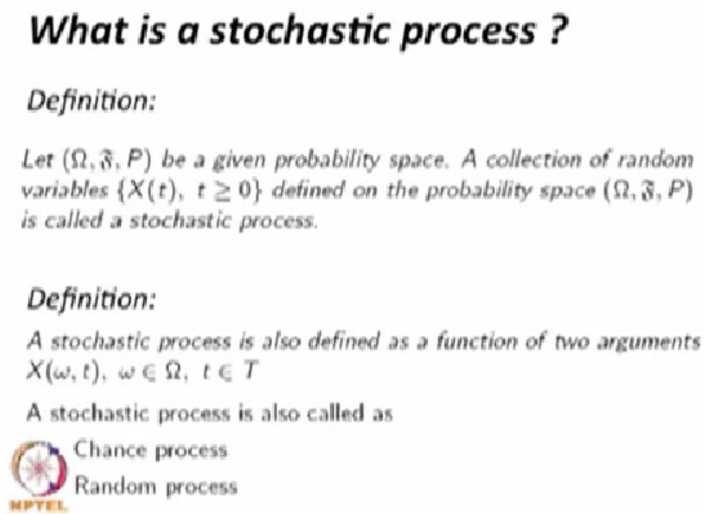
In this model what we are going to discuss is stochastic process, then we are going to discuss the classification of a stochastic process followed by a few simple examples which arises in the real-world problem.

(Refer Slide Time: 00:15)



So the content of this lecture is going to be as I said let me first give that definition of stochastic processes then I will explain how to create or how to develop this stochastic process. What is the meaning of a parameter and state space then I am going to give what are all the approaches in which stochastic processes can be described and the classification of stochastic processes based on the parameter and the state space. Then, at the end of this lecture we are going to discuss some of the few simple stochastic processes and the summary of the lecture one and there are few reference book also listed for this course preparation. What is stochastic process?

(Refer Slide Time: 01:01)



Let me give the definition. Let, (Ω, F, P) be a given probability space. That means you know what is a random experiment? From the random experiment you know what the Ω and from the collection of possible outcomes you got that sigma algebra that is capital F and you have probability measure also. Therefore, this triplet is going to be the probability space and you have a given probability space.

From the given probability space, you have the collection of random variables that is X(t) where t is belonging to capital T defined on the probability space that is (Ω, F, P) that is called a stochastic process. That means you have probability space from the probability space you have collected random variables with the t belonging to capital T and this collection is going to be called it as a stochastic process.

Now the question whether we can create a only one stochastic process or how to create a stochastic process from the sigma algebra that means suppose you have a Ω from the Ω you can always create a sigma algebra that is a F that is a collection of a subsets of Ω satisfying the condition. If you take a few elements, then the union of elements is also belonging to one of the elements and if you take any one of the elements in the F then the compliments is also belonging to F.

So, if these conditions are going to be satisfied then that collection of subsets of Ω is going to be called it as sigma algebra. So, from the Ω we have created a random variable that is $X(t_1)$ that is nothing but a real valued function which is defined from Ω to R such that it satisfies the condition $X(t_1)$ of inverse of $(-\infty,x]$ that is belonging to F for all x belonging to R.

That means whatever be the x belong to F if the inverse images from $-\infty$ to some point x if that is belonging to F then that real valued function is going to be called it as random variable. Like that if you make a different random variable for different t where all the t_i's belong to so I can go for t_i, so all the t_i's are belonging to T. So, that means if I have a collection of random variables for the different values of t then that collection is going to called it as a stochastic process.

Now, the question is whether we can create only one stochastic process from a given probability space or more than one stochastic process can be created from the same probability space. The answer is yes you can always create more than one random variable from the same probability space that means for a different collection of a T you can have a different stochastic process.

More than one stochastic process can be created from one probability space. Now, the next question if I change the sigma algebra what happens? If I change the sigma algebra F then I may land up collecting some other stochastic process in which those real valued function is going to be random variable for that particular Ω and F and P for a given probability space the stochastic process is going to be changed for a different collection of a t belonging to T.

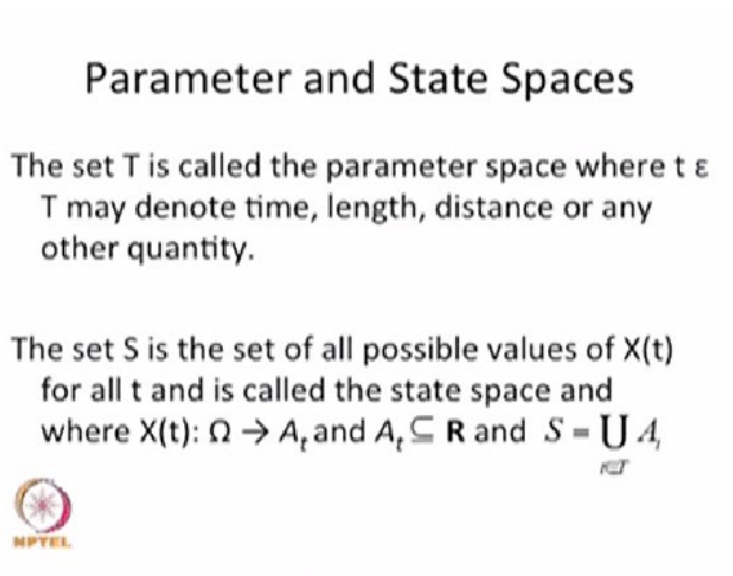
That means once you know the F then you will have some collection of random variables that will form a stochastic process. If you change another F, then you may get the different stochastic process and also for a given probability space you can have more than one stochastic process by the way you define a collection of random variables. The way you have a T accordingly you will have a different stochastic process.

Now the way I have given the collection of random variable I can say it in a different way that is a stochastic process also defined as a function of two arguments that is X (w, t) where w is belonging to Ω and t is belonging to T that means the same way I can define the collection of random variable as a collection of (w, t) where w is belonging to Ω and t belonging to T and this is also going to be form it as a stochastic process.

That means always the w is belonging to Ω that means w is belonging to the possible outcomes and the t is belonging to capital T and this is going to set the stochastic process. The other names for the stochastic processes are going to be John process. There are some others use the word John's process. There are some others they use the notation that is called a random process.

So either the stochastic processes can be called it as John's process or the random process also. Now, what we are going to see once you have a collection of random variables so based on the values of X(t) and the values of the different values of t we are going to define what the parameter space and what is state space. What is the meaning of parameter space?

(Refer Slide Time: 08:46)

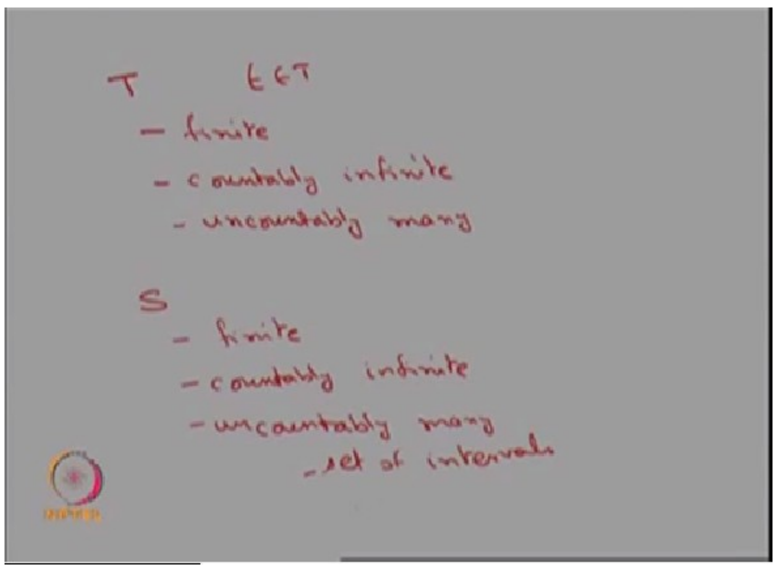


The set we use the notation T that is called the parameter space. The set T is called the parameter space and it is usually represented as the time most of the time or it can be represented as the length or it can be represented as the distance and so on. So, we usually we go for t as a time. So, the set T is called the parameter space. Similarly, I can define the state space as the set S that is nothing but all possible values of X(t) for all t.

So, this set is called the state space. X(t) is a random variable from omega into A_t where A_t is a subset of R then the A_t are going to be the elements of it is going to be contained in the real line then the S is nothing but union of t belonging to T all the A_t that is going to form a state space. That means for fixed t you will have a collection of a possible values that is going to be the A_t and for variable t you collect all the union and that possible values of X(t) is going to form a set and that set is called the state space.

Similarly, the all possible values of a t belonging to T and that set is going to be call it as a parameter space. So, based on the parameter space and the state space we can go for classification. Now, I can explain what are all possible values of S can take.

(Refer Slide Time: 11:10)



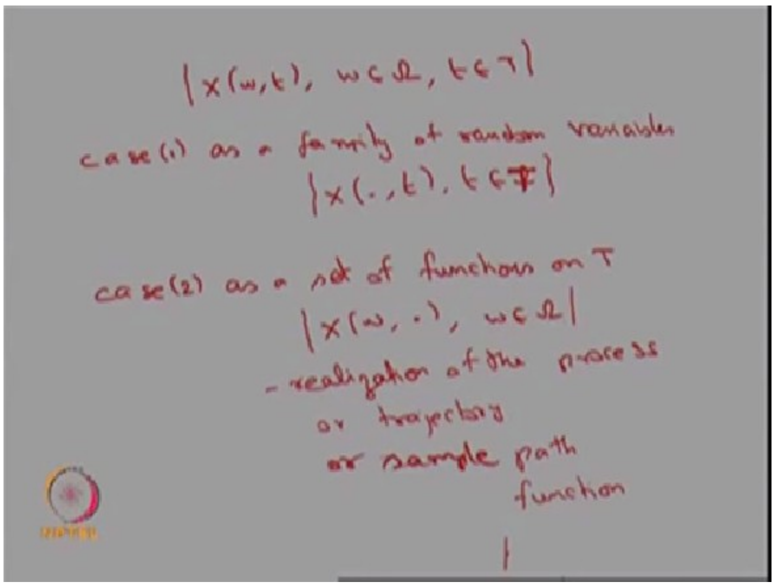
So, this T is going to be the collection of t therefore this can be a finite that means countably finite or it could be countably infinite also or it could be uncountable many elements of t. So, that set can be a finite set or it could be countably infinite or it could be uncountably many elements also. T can also be multidimensional set. Similarly, the state space the S that can be a same way it could be a finite or it could be accountably infinite or it could be uncountably many elements.

So, since the state ways are going to be the collection of all possible values of X(t) and X(t) is a real valued function and then it is going to be random variable therefor these elements are going to be always the real numbers. So, either it could be a finite element or it could be accountably infinite elements and it is going to be uncountably many elements that means it could be a set of intervals on a real line or it could be the whole real line itself.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 49

(Refer Slide Time: 00:07)



So, based on the values of the way I have explained the random variable or the stochastic processes is going to be X(w, t) where w is belonging to Ω and t is belonging to T.

There are two approaches can define the stochastic process. The first one that is a we name it as a case one. I can say it as the collection of random variables as a family. Family of random variables as a X(., t) where t is belonging to T. So, this is the way I can create random variable and this is the easier approach in the sense once I know the different t or fixed t it is going to be random variable.

And I have collected a family of random variable for different values of t. Therefore, this is the way we can create the stochastic process. This is the easier approach also. The next one that is the case two, that is nothing but as a set of functions on t that is nothing but a collection of X(w,.) for w is belonging to Ω that means I have made a function on T and once I fix one w I will have a one function and if I fix another w where w is nothing but a possible outcomes.

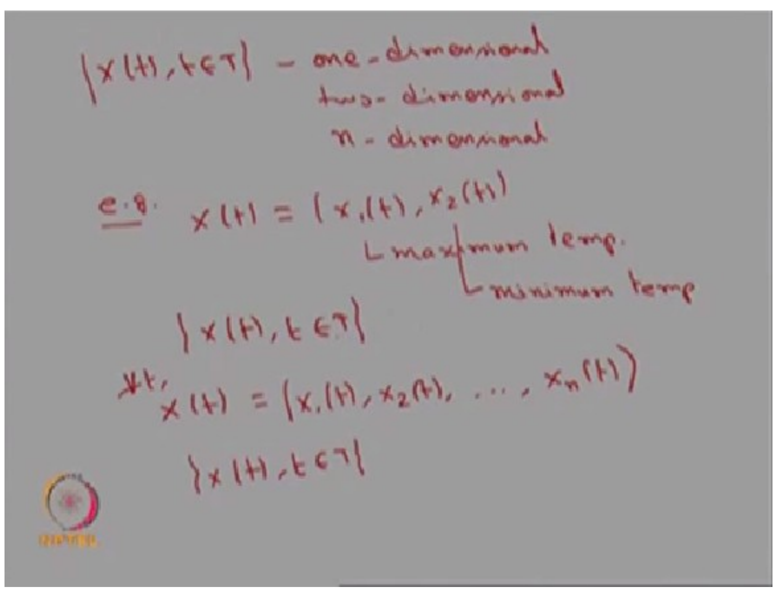
Therefore, if I have a different possible outcome that is going to be create a different stochastic process where as therefore I can create a stochastic process of X(w, t) either fixing a t or fixing the w accordingly I can have a two different ways of creating a stochastic process. And the case two the way I have made a collection of random variables by fixing the w then I made a function of functions on t.

Therefore, this is going to be the method of realization of the process or it is going to be called it as a trajectory or it can be called it as a sample path or we called it as a sample function also. So, these are all the different ways the case two can be called that means once you know the one possible outcome therefore you are tracing the stochastic processes along the one possible outcome.

Therefore, that is going to be called it as a realization of the process or the trajectory of the sample path. So, the conclusion is we can always define a stochastic process as a collection of random variables for different value of t or we can go for a collection of functions on t for different values of the possible outcomes that is w belonging to omega.

So, these are all the two approaches in creating the stochastic process. Not only we can go for making one dimensional random variable or one dimensional stochastic process.

(Refer Slide Time: 03:50)



So, you can create a stochastic process it could be a one dimensional or it could be two dimensional or it could be a n dimensional also. So, first we have discussed what is stochastic process and how to create a stochastic process whether it existed so on then we have given the parameters based on the state space then we have given what are all the ways we can create the two different approaches you can create the stochastic process.

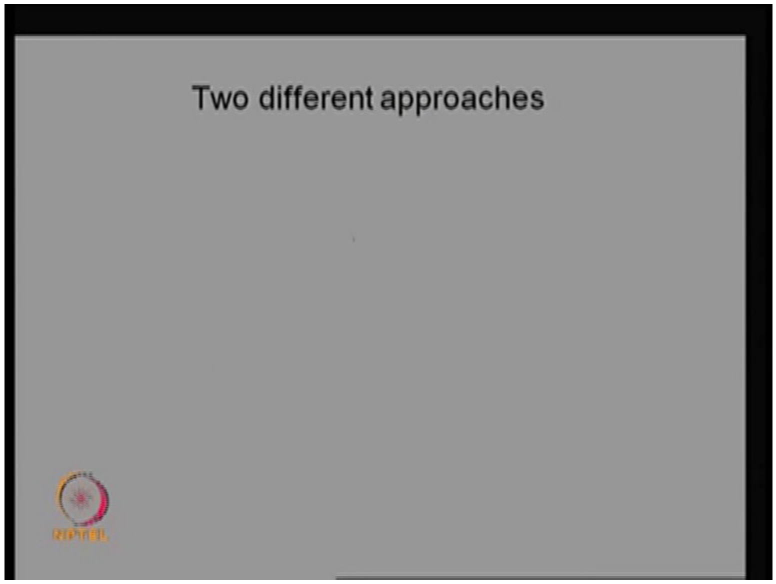
Now, we are discussing what is the dimension of the stochastic process? Whether the default it would be a one dimensional or it could be two dimensional or it could be n dimensional? Let me give a one simple example in which it is going to be two dimension that means I have a random variable X(t) that is going to be $\{X_1(t), X_2(t)\}$ in which $X_1(t)$is nothing but the maximum temperature and the $X_2(t)$could be minimum temperature.

The maximum and minimum temperature possible of a place at any time t and this together is going to be a one random variable that means this is a random vector which consist of two random variable $X_1(t)$ and $X_2(t)$that means for fixed t you have a one random vector X(t) and therefore you have random vector for over the t and this random vector will form a stochastic process therefore, this is going to be a two dimensional stochastic process.

Therefore, in general, you can define a n dimensional stochastic process with the for fixed for every t you have a random vector at some t that is going to be a $X_1(t), X_2(t)$and so on. It is going to be the n th element is the$X_n(t)$ that is going to be n different in which each one is going to be random variable for fixed t and this is going to be random vector for fixed t and this is going to be n dimensional stochastic process in which each one is going to be one dimensional random variable for fixed t.

So, that means you can go for making one dimensional random variable then you have a collection of random variables form one dimensional stochastic process or you can have a two dimensional like that you can have a n dimensional stochastic process. In the course what we are going to discuss always it is one dimensional stochastic process.

(Refer Slide Time: 06:55)

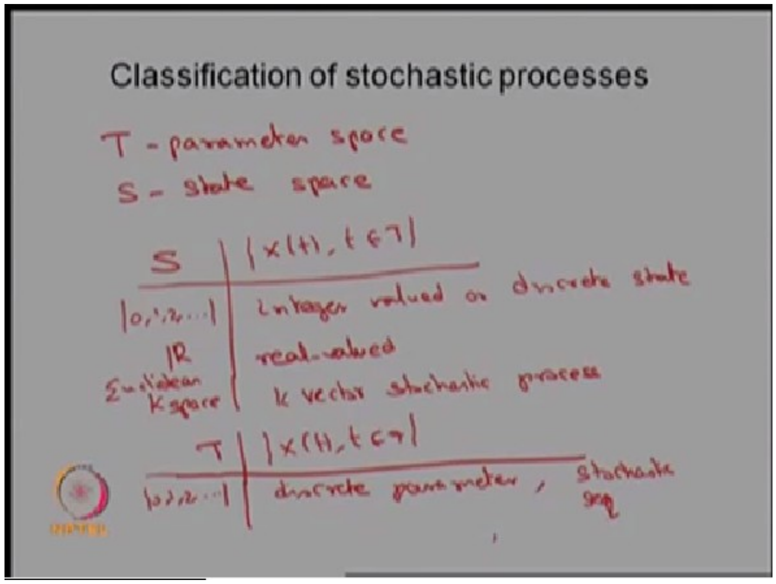


We can always create a complex valued stochastic process also in the form of X(t). If I need it here the X(t) is going to be $X_1(t)+i\,X_2(t)$where i is nothing but the complex quantities square root of minus one. That means the $X_1(t)$ is a real valued random variable for fixed t and $X_2(t)$is also a real valued the random variable for fixed t. The way I have made the X(t).

This is going to be complex valued random variable for fixed t therefore the X(t) over the t that is going to be form a complex valued stochastic process because for fixed t, X(t) is going to be complex valued random variable. The corresponding stochastic process is called complex valued stochastic process with the one dimensional form like that you can go for the multidimensional complex valued stochastic process also.

But in this course here what we are not really discussing is only the real valued one dimensional random variable most of the times. Sometimes we are discussing real valued two dimensional or n dimensional stochastic process that too with the real valued random variable not the complex valued. So, now we are going for classification of stochastic process.

(Refer Slide Time: 08:35)



The way, I have explained the parameter space T. The T is parameter space and S is going to be the state place that is nothing but the collection of a possible values of X of t and the possible values of a t belonging to the T that form a parameter space. Some books, they use the notation parameter Z also and S is going to be the state space. Now based on these we are going to classify the stochastic process. Suppose let us start with the S.

Suppose, the possible values of S and what is the name of the stochastic process if S is going to take the only countably infinite or countably finite values? Then it is going to be called as the corresponding stochastic process is going to be called as a integer valued stochastic process or we can call it as a discrete state stochastic process. So, whenever the possible values of S is going to be accountably finite or accountably infinite then we say it is a integer valued stochastic process or a discrete state stochastic process.

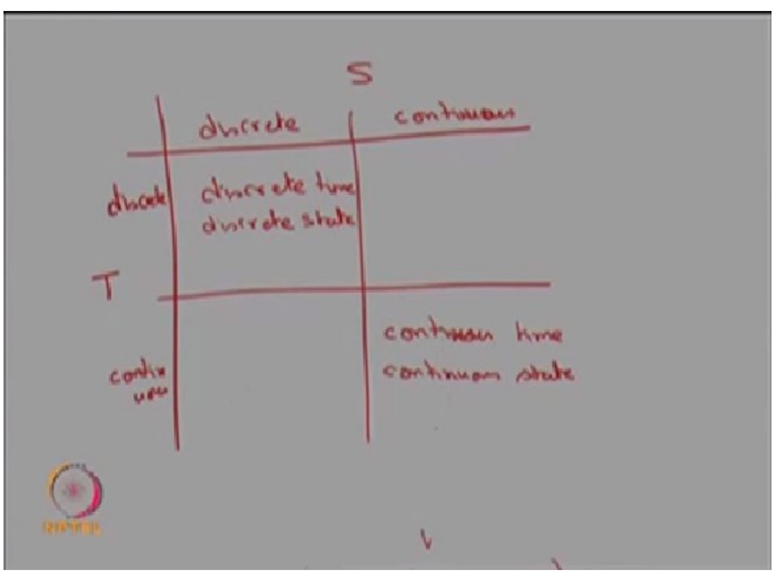
Suppose, the possible values of S is going to be the real values then we call it as a real valued stochastic process. Suppose, if it takes euclidean space with the k dimensional euclidean k dimensional space then we call it as a k vector space, k vector stochastic process, that means each random variable going to have a one dimensional random variable and like that you have a k random variables of our fixed t.

Therefore, you have a k vector stochastic process therefore it is going to be call it as a k vector stochastic process in which each element is going to be one dimensional random variable for fixed t. So, the collection that the k tuple value stochastic process is going to be call it as k vector stochastic process. Similarly, you can go for based on the T what is the name of the stochastic process for different values of t.

That means if it is going to take the value countably finite or countably infinite or it is going to take a only the integer values then we say it is a discrete parameter stochastic process or there is another name it is called the stochastic sequence also whenever the possible values of the T is going to be countably finite or accountably infinite then we call the corresponding stochastic process as the stochastic sequence or it is a discrete parameter stochastic process.

Otherwise, if it takes uncountable many values in the T then it is going to be call it as a continuous parameter or it going to be called it as a stochastic process itself. Therefore, based on that discretize it uses the world sequence or if it is going to be uncountable many values of T then it is going to be called it as a stochastic process. So, based on the classification I can go for making a one table in which the possible values of S will take a column and the possible values of T will make row.

(Refer Slide Time: 12:54)



So, either it could be a countably finite or countably infinite data uses the word discrete. If the possible values of T is going to be uncountably many either it is set of all intervals or it would be whole real line itself or it is going to be a union of many intervals in that case it is going to be call it as a continuous parameter.

Similarly, if the possible values of S are going to be countably finite or countably infinite then the state space is going to be call it as a discrete. Similarly, if it is going to be uncountably many values then it is going to be called it as continuous. So, accordingly you can classify the stochastic process into the four type in which if the T is going to be discrete as well as S is going to be a discrete.

Then it is going to be a discrete time or discrete parameter both are one and the same. So, discrete time, discrete state stochastic process. Similarly, if the T is a discrete and the state space is continuous then we can call it as a discrete time continuous state stochastic process. Similarly, this is going to be a continuous time discrete state stochastic process and this is going to be a continuous time, continuous state stochastic process.

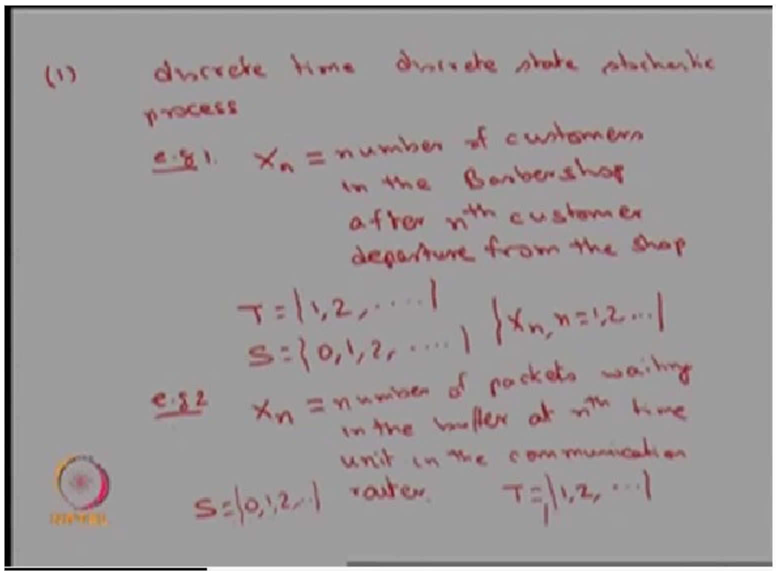
That means based on the possible values of a T and the possible values of S any stochastic process can be classified into the four types in which it is going to be a discrete discrete or continuous continuous or discrete continuous, continuous continuous based on the time and the state space.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 50

So let us see some simple example based on the possible values of T and S.

(Refer Slide Time: 0:08)



So, the first one is going to be a discrete time or you can use a discrete parameter also. Discrete time, discrete state stochastic process that means the possible values of S as well as the possible values of T has to be either it has to be of countably finite or countably infinite elements in it. Let see the one simple example. Let us create a random variable X_n, that is nothing but the number of customers in the Barber shop after nth customer departure from the shop.

So, here suffix n that will form a parameter space therefore the T can be a possible value of n. That means whenever one customer leave the system how many are in the system after he leaves. So the possible values of T will be the first customers when he leaves out when he is not there he want to find out and so. Therefore, the possible values of T is going to be 1, 2 or three therefore this is the number of making the number of customers in the system.

Whereas the possible values of X_n or possible value of n that is going to be the –there is a possibility no customers in the system when someone leaves. So, there is a possibility zero when someone leaves only one customer in the system when it is going to be 1 or 2 and so on. Therefore, there is a possibility it could be finite also. So the S can be countably finite or in this case I have made the it is countably infinite.

Therefore, the T as well as T is going to be form of the discrete therefore the corresponding stochastic process X_n for possible values of n is going to be 1, 2 and so on and this is going to be a discrete time, discrete state to stochastic process. You please note that here the parameter space T is not the time. The parameter space, forming the 1, 2, 3 these are all the customers, the nth customers. Therefore, n can be 1, 2 and so on.

Therefore, it usually the T is time whereas sometimes it would be a distance or length or the number or whatever the other quantities. So here the typical situation in which the parameter space is not considering the time. Therefore, this is going to be a random variable because you never know how many customers are going to be in the system after the nth customers leaves.

Therefore, this is going to be a random variable. Obviously it is a real value function satisfying all the properties of the definition and you can see the probability space for this and from the probability space you have to create the random variable and therefore this random variable is going to be the –this collection of random variable over the n that is going to be the discrete time and discrete state.

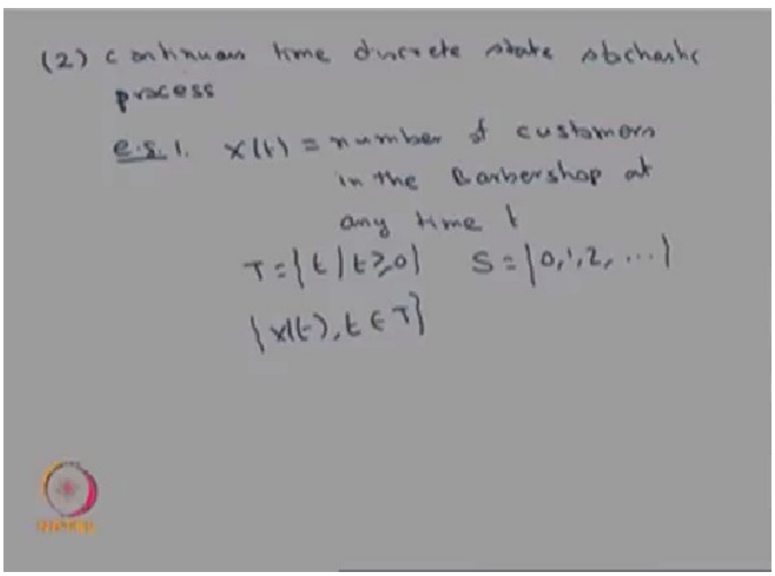
Therefore, this random variable here you can create it with the help of a case one by making for fixed n, what is the random variable then you make a collection of random variable. So, we can create this stochastic process by using the case one or the approach one which the easier one. I can go for creating one more stochastic process for this discrete time and the discrete state stochastic process that comes under daily communication problems X_n is going to be number of packets waiting in the buffer at nth time unit in the communication router.

That means there is a communication router in which the packets are coming for transmission. So after the transmission is over in the buffer the packets are leave the router. So at any time you don not how many packets are waiting in the buffer for the transmission. So there is a possibility no packets will be there at some time point and there is a possibility there are many more packets may be waiting for the transmission in the buffer.

So, the possible values of S that is going to be –there is a possibility no packets in the buffer or one or so on and similarly the possible values of T that is also we are marking the nth time unit. Therefore, the time unit could view first time unit or second time unit and so on. Therefore, here the S is going to be the discrete as well as the T is going to be discrete therefore this collection of random variables $X_n$for possible values of n that is also going to form a discrete time, discrete state stochastic process.

Because of the possible both the values are going to be of discrete type discussing the simple stochastic process based on the parameter ways and the state ways and we have seen the discrete time discrete state stochastic process. The first one now we are seeing the second one.

(Refer Slide Time: 07:06)



That is continuous time discrete state stochastic process. That means the possible values of parameter space is going to be a uncountably many values therefore we get the continuous time and the possible values of the state based that is going to be a countably finite or countably infinite therefore you get the discrete state. So, you will see the few simple example of this type. The first example that is X(t) that is going to be the number of customers in the Barber shop at any time t that is difference.

In the earlier, example we have seen the number of customers in the barber shop for the nth customers departure now we are seeing the number of customers in the barber shop at any time t. Therefore, we are looking at how many customers at any time t in the barber shop. Therefore, the possible values T that is going to be a collection of t such that the $t \geq 0$.

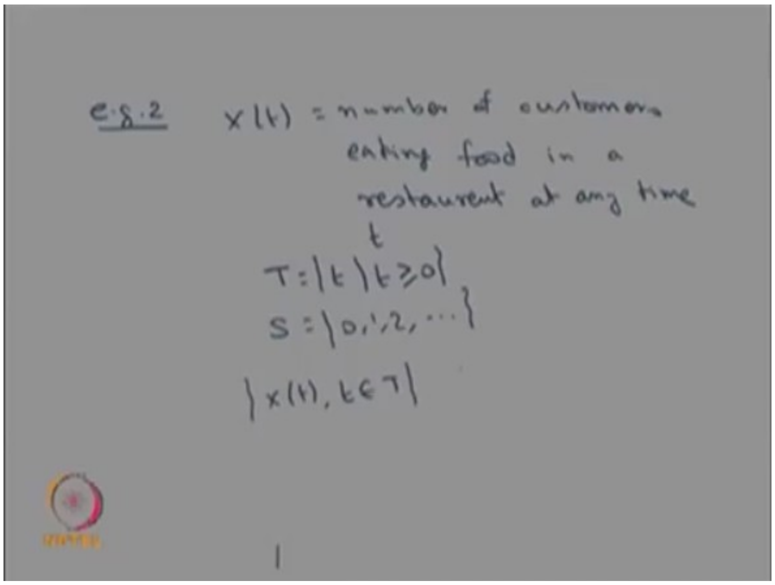
And the possible values of S that is going to be still it is a number of customers therefore the possible values are 0, 1, 2 or it can be when there is a possibility it can be countably finite also. So whether the state place is going to be countably finite or countably infinite we classify as a discrete state. Therefore, this is a typical example of continuous time discrete state stochastic process and the collection of random variables is going to be $X(t)$ for all possibly values of t.

So, this is going to form a real value the stochastic process which for each t it is going to be a random variable. So, this is going to be a real value the stochastic process of one dimensional type and the t is belonging to the T that is going to be the time that is the default one and it is going to be a countably many therefore it is going to be continuous parameter. So it is going to be call it as a continuous parameter discrete state stochastic process also.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 51

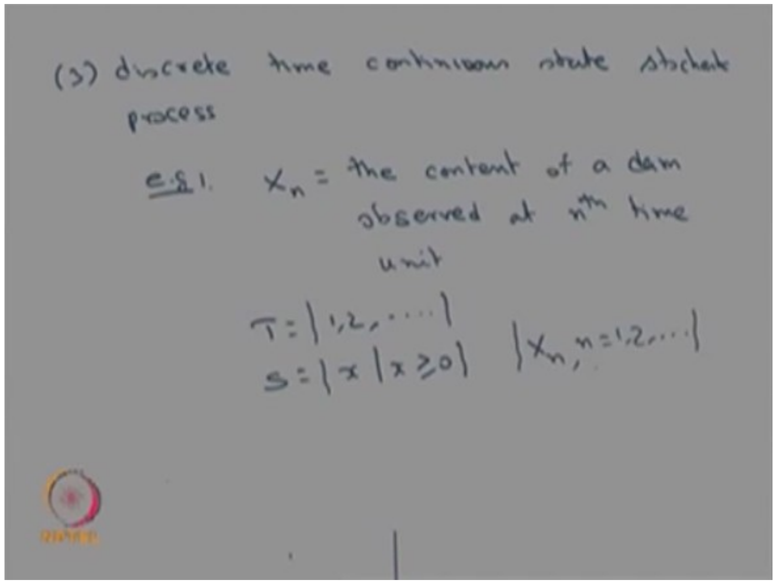
(Refer Slide Time: 00:03)



The example two. In the example two, when we make X(t) that is going to be number of customers eating food in a restaurant at any time t. Therefore, you are observing the system you are observing the restaurant. How many customers are taking their food? Therefore, the possible values of parameter T is going to be $t \geq 0$. And the possible values of S still it is a count therefore the possible values are accountably finite or countably infinite.

Therefore, this collection of random variable over the t that is going to be a continuous time or continuous parameter discrete state stochastic process. This is a very typical example so it could be countably finite or countably infinite also.

(Refer Slide Time: 01:33)



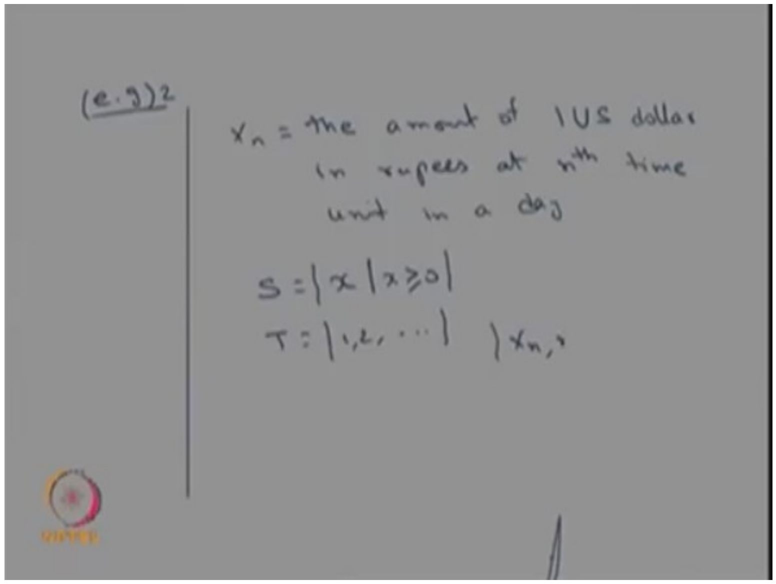
Now let us see the third type that is discrete time continuous state stochastic process that means we need to have the T value has to be countably finite or countably infinite. Whereas the possible values of state space has to be countably many of that type. So, let us create example for that. The X_n is nothing but it is a random variable that denotes the content of a dam or water reservoir observed at nth time unit.

So, here the time unit could be every one hour or that could be because you are seeing what is the content of dam or water reservoir. It could be everyday fixed time of everyday or it could be fixed time of weekly once so that is going to be the time. So that is going to be the time period. So at the end of each nth time unit we are observing what is the content of dam. So that is nothing but it is real quantity.

Therefore, T is going to be you are observing only at that time unit. So, either it would be one or daily once or weekly once or so on. So, therefore I can make a one to one correspondence with the countably finite or countably infinite numbers so that they will form a parameter space and the S this is going to be possible value of X_n for all possible values of n. So, this is water content of dam that is going to be the real quantity that is going to be for some x where $x \geq 0$.

So, that means parameter space is going to be a discrete whereas the state space is going to be a continuous therefore these two stochastic process X_n for possible values of n is going to be 1, 2, 3, 4 and so on and this is going to form discrete time continuous state stochastic process.

(Refer Slide Time: 04:17)

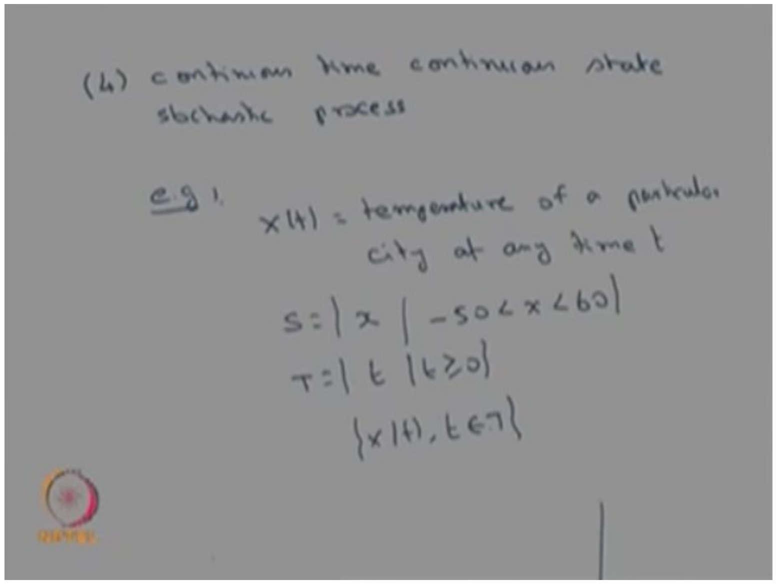


Let me give one more example for the same type that is example two that is nothing but example two X_n is nothing but the amount of 1 US dollar in rupees at n th time unit in a day. That means I am just observing what is the value of 1 US dollar in rupees in a day for the n th time unit. It could be every five minutes or it would be every minute or it could be every hour of any particular day and that is going to form a random variable and that collection is going to form a stochastic process.

In, this the possible values of X are going to be since it is the amount of 1 US dollar in rupees it could be some fraction also. Therefore, you do not want to take it as the integer number. It could be real numbers therefore it is going to possible values of $x \geq 0$ and the T that is going to the time unit either it is every minute of every once in five minute or once in ten minutes or every one hour or so on.

So, this is going to form a countably finite or countably infinite one and this stochastic process will form a discrete time and continuous state stochastic processes.

(Refer Slide Time: 06:14)

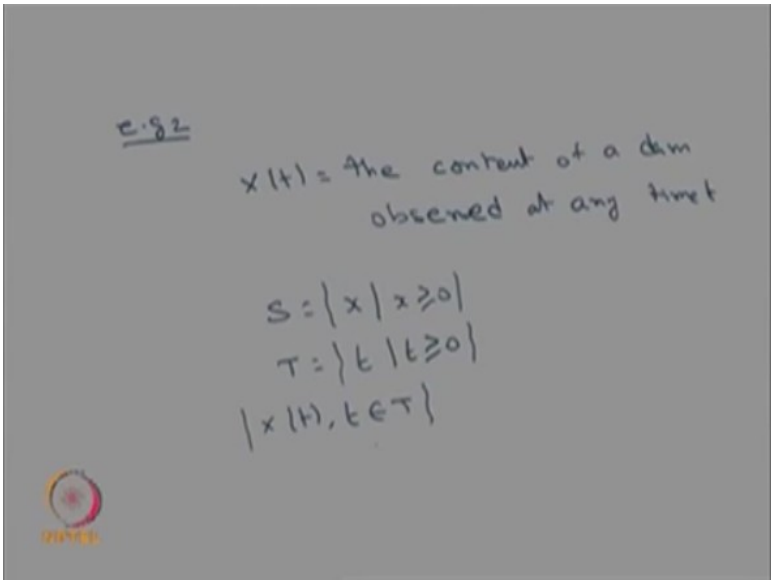


Let me go for the fourth type that is the fourth classification of the Stochastic Process that is continuous time, continuous state stochastic process that means the possible values of the parameter is going to be uncountably many therefore you get the continuous time or continuous parameter and the possible values of the state base that is going to be uncountably many therefore you get the continuous state stochastic process.

The examples or the first one X(t) is going to be temperature of a particular city at any time t. So whenever I use anytime t you can take any value therefore the possible value of S is going to be the temperature so you can think of temperature suppose a particular city lies between -50 to 60 degree Celsius So this quantity is going to be the Celsius of -50 to + 60 and the parameter space t is going to be observed over the time.

Therefore, this time $t \geq 0$. Therefore, the parameter space is continuous one and the state space is continuous one therefore this collection or random variable form continuous time continuous state stochastic process.

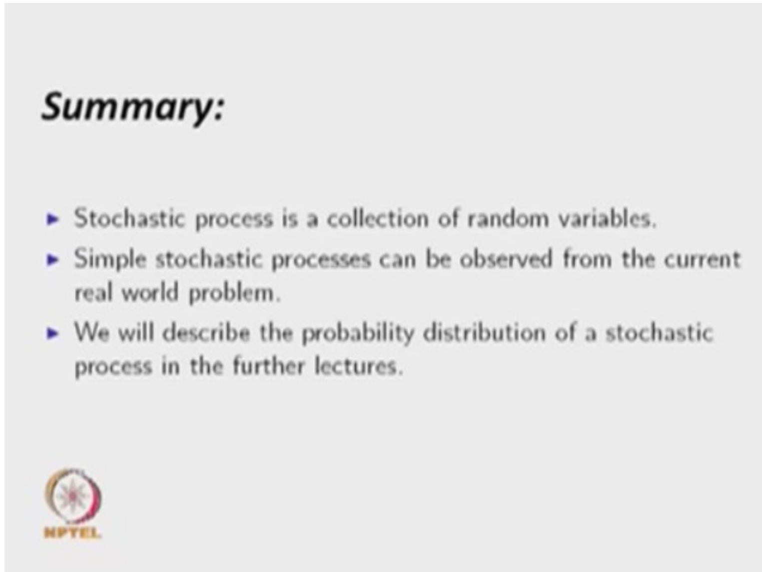
(Refer Slide Time: 08:23)



Let me give another example of fourth type, that is example 2. X(t) is the content of dam observed at any time t. so the content of dam reservoir that is going to be the real quantity therefore S is going to be a collection of x such that x ≥ 0. And you are observing over the time therefore that is also collection of t such that t ≥ 0.

Therefore, this will form a stochastic process in which it will be under that classification of continuous time, continuous space stochastic process and this can be created with the help of the first approach that means for fixed t find out what is the random variable and you collect the random variable over the all possible values of t. Therefore, this is going to be of the continuous time and continuous state stochastic process.

(Refer Slide Time: 09:53)



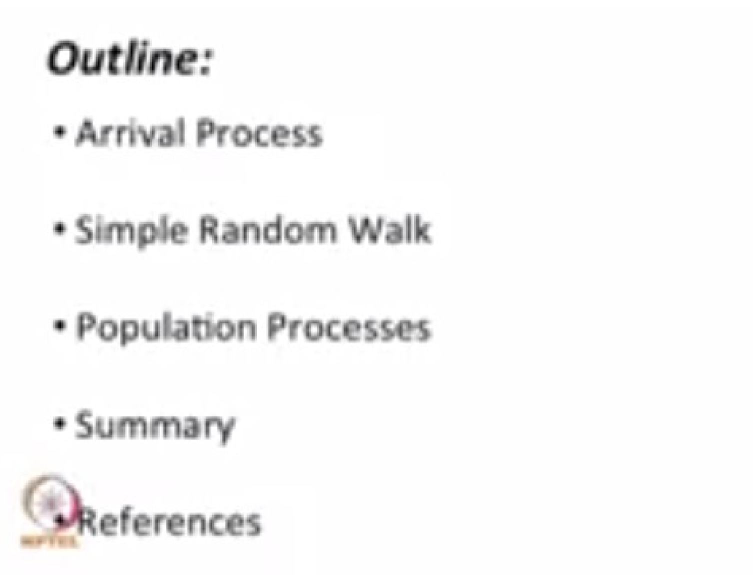
So in this lecture what we have seen what is the meaning of stochastic process or how to create the stochastic process so that is nothing but it is going to be a collection of random variables. So, we have defined stochastic process as well how to create then later we have given what is parameter space and what is state space and we have given the classification of stochastic process based on the parameter space and state space and also some of the real world problems from that we can create stochastic process.

And that stochastic processes are the simple stochastic process and there are many more stochastic processes can be created with the help of the definition and so on.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

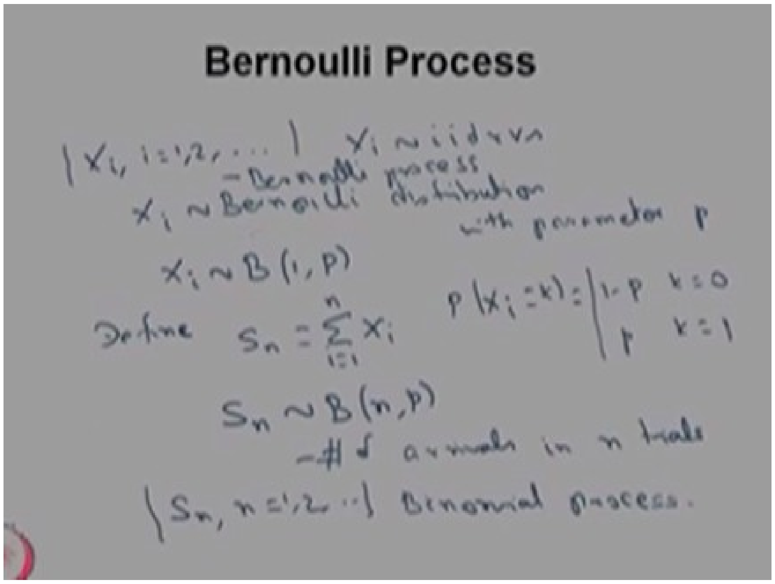
Lecture – 52

(Refer Slide Time: 00:04)



In this lecture, we are going to discuss some simple stochastic process starting with the discrete time arrival process that is the Bernoulli process and continuous time arrival process that is a Poisson process. Followed by that we are going to discuss the simple random walk, then we are going to discuss a one simple population process, which arises in the branching process, then we are going to discuss the Gaussian process. So, with that the lecture 2 will be over.

(Refer Slide Time: 00:39)



What is Bernoulli Process? Bernoulli process can be created by the sequence of a random variable. Suppose you think of a random variable X_i, where, ‘i’ takes the value one, two and so on, therefore this is going to be a collection of random variable and each random variable are X_i, and you can think of X_i are going to be an iid random variables. And each is coming from the Bernoulli trials.

That means each random variable is a Bernoulli distributed, each random variable is a Bernoulli distribution and with a parameter p. So the same thing can be written in the notation form X_i takes the, X_i’s are in the notation, it is the B(1, p). That means it is a binomial distribution with the parameters 1 and p that is same as each X_i are Bernoulli distributed with the parameter 1 and p.

So now I can, so this is going to be a stochastic process or we can say it is a stochastic sequence.

Now I can define another random variable, for every n, $S_n = \sum_{i=1}^{n} X_i$. Suppose you think X_i is going to be the outcome of the i-th trial, so the X_i can take the value zero or one that means with the probability, X_i can take the value k. If k =0 with the probability 1-p and k can take the value 1 with the probability p.

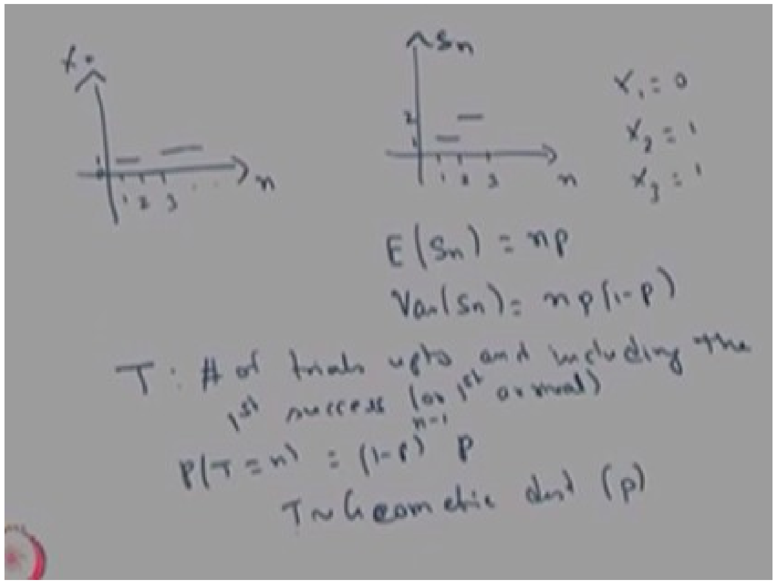
Therefore, since each X_i's are iid random variable, you can come to the conclusion Sn is nothing but binomial distribution with the parameters (n, p). Suppose you assume that X_i is going to be number of whether the arrival occurs in the i-th trial or not. If X_i takes the value zero, that means no arrival takes place in the i-th trial. If Xi takes the value one that corresponding to the i-th trial, there is at arrival.

So the Sn represents, Sn denotes the number of arrivals in n trials. So now you can create a stochastic process with Sn, where n takes the value one, two and so on, therefore this is going to be a binomial process. So the X_i takes the value zero or one with the probability 1- p and p, each one is going to be Bernoulli distributed, therefore this is going to be a Bernoulli process. This X_i are going to form Bernoulli process.

The way you have created $S_n = \sum_{i=1}^{n} X_i$ and each S_n is going to be a binomial distribution with the parameters n and p. Therefore, this S_n, that sequence of S_n for n is equal to 1, 2, 3,… is a binomial process. Therefore, since you have collected arrivals over the possible values of one, two and so on, therefore this is going to be a one of the discrete time arrival process.

So similarly we are going to explain what is the continuous time arrival process, whereas here binomial process. This is going to be a discrete time arrival process.

(Refer Slide Time: 05:01)



Suppose you like to see the trace of S_n, so before you go to the trace of S_n, we can go for what is a trace or sample path of X_i. For different values of n is equal to one, n is equal to two, n is equal to three and so on, if you see each X_i takes a value zero or one, therefore it can take the value zero or X_1 can take the value one or X_2 can take the value zero or this can take the value one, again it can take the value one and zero.

So the possible values of X_i's are going to be zero and one, therefore each X_i's can take the value zero in the horizontal line or it can take the one till you get the next trial. Similarly, if you make the sample path or the trace of S_n, since S_n is going to be sum of first n random variable, therefore based on the X_i takes the value, suppose X_1 takes the value zero and suppose X_2 takes the value one and suppose X_3 takes the value one and so on.

So since $X_1 = 0$, therefore S_1 is zero. Then that S_2 is same as the $X_1 + X_2$, therefore it takes the value one. And $S_3 = X_1 + X_2 + X_3$, therefore that is going to be again, you are adding the values therefore it is going to be a two, therefore this is one and this is two. So based on the X_4, it is going to be zero or one, either it can take the value two itself or it can go to the three.

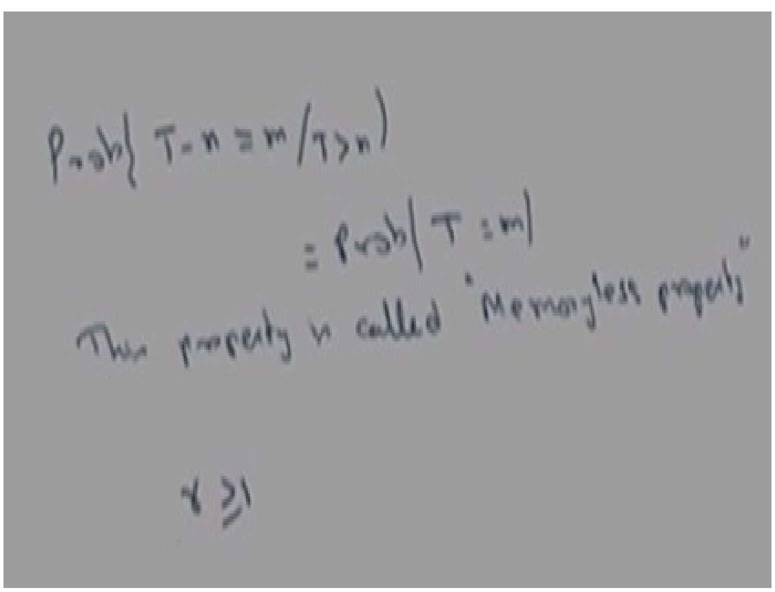
Therefore, if you see the sample path of S_n, it is going to be either incremented by one or it takes the same value till the next n. Therefore, not only you can find out the S_n, you can, not only you can find out the sample path of S_n, you can get the mean and variance because each S_n is going to be a binomial distribution with the parameters n and p, therefore the expectation of S_n is going to be np and the variance of S_n is going to be np(1 – p).

So you can able to see the sample path of X_i's as well as S_n over the different values of n. In discrete time sample paths are sequences. I can also define the new random variable T is nothing but number of trials upto and including the first success, that means suppose it takes a value n that means for subsequent n minus one trial so I got the failures or no arrival takes place in subsequent n minus one trial and at the n-th trial I get the first arrival.

That means the T is a random variable to denote how many trials to get the first success or the first arrival. So if it is going to take the first arrival in the n-th trial then the probability of T takes value n that is same as (1- p)(n- 1)p because all the trials are independent and subsequent n- 1 trial gives no arrival and the n-th trial you get the first arrival. Therefore, this is going to follow a geometric distribution with the parameter p.

So since you know the distribution of T, you can find out the mean and variance because the mean of a geometric distribution is going to be 1 / p and the variance of T is going to be $\frac{1-p}{p^2}$.

(Refer Slide Time: 09:22)



Similarly, I can go for finding out what is a probability that till n-th trial, I did not get the first or I did not get the first arrival. So if n plus m th trial, if I am getting the first arrival, what is the probability that it is going to take after n trials, it gets the first arrival that probability you can able to get that is same as the probability of the T takes value m. So this property is called memory less property.

Since T is geometrically distributed and the geometric distribution satisfies the memory less property that can be visualized in this example, $P\{T - n = m / T > n\} = P\{T = m\}$. That means the right hand side result is independent of n and it is same as the distribution of, that means the residual arrival, number of arrivals that is same as the original arrival distribution.

Therefore, this satisfies the memory less property. So this is the geometric distribution satisfies the memory less property in the discrete time and there is another distribution satisfies the memory less property in the continuous time that is the exponential distribution. So the way I have related the binomial distribution from the Bernoulli process, then I get the binomial process, also I was able to create the geometric distribution.

You can create the or you can develop the Pascal distribution or negative exponential distribution. The way I have defined the T is going to be the number of trials to get the first success or first arrival, instead of that if I make another random variable to go for, how many trials are needed to get the r-th success, where r can take the value greater than or equal to 1.

If it is the r-th first success is going to happen in the n-th trial, if r is greater than one, then I can go for defining what is the negative binomial distribution for that particular random variable. If $r = 1$, then that is landed to be the same random variable capital T.

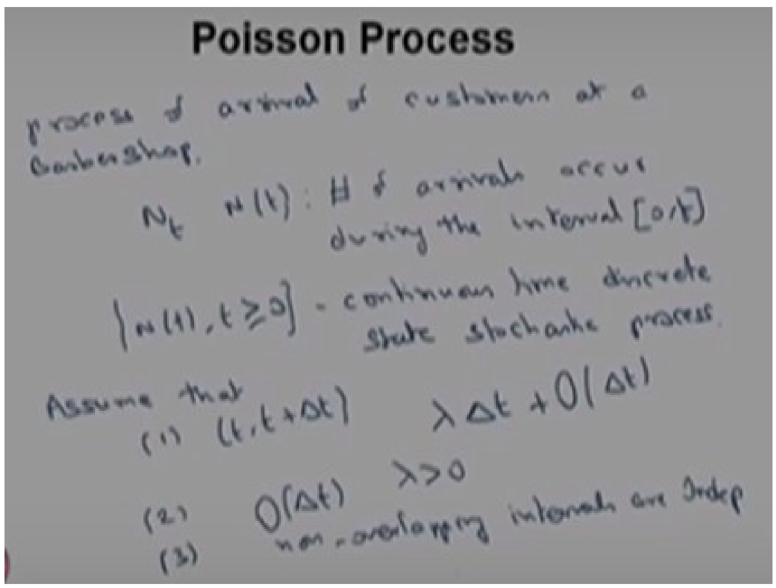
Introduction to Probability Theory and Stochastic Processes

Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 53

So, till now we have discussed what is the discrete time arrival process. Now we are going to discuss the continuous time arrival process that is a Poisson process.

(Refer Slide Time: 00:13)



In this lecture, I am going to develop what is the Poisson process and how we can get the Poisson process from the scratch. Suppose you consider the process of arrival of customers at a barber shop. So this is the same example, we have discussed in the beginning of this course also. So over the time, how many arrivals is going to take place? That is going to be a random variable. So let $N_t$or in some books they uses N (t).

So the N (t) denotes number of arrivals occur during the interval zero to the closed interval [0,t]. That means we are defining a random variable N (t) that denotes the number of arrival occurs during the interval [0,t]. For fixed t, N (t) is going to be a random variable. Therefore, N (t) over the time, because t ≥ 0, this is going to be a since the possible values of t that is the parameter space is going to 0 to ∞, therefore this is going to under the classification of a continuous parameter or continuous time.

And the possible values of N (t) for different values of t that is going to be takes the value zero or one or two, therefore it is going to be a countably infinite. Therefore, this is going to be a continuous time or continuous parameter discrete state stochastic process. So, this is the N (t) over the $t \geq 0$ that is going to be a continuous time discrete state stochastic process. Now, we are going to develop the theory behind Poisson process.

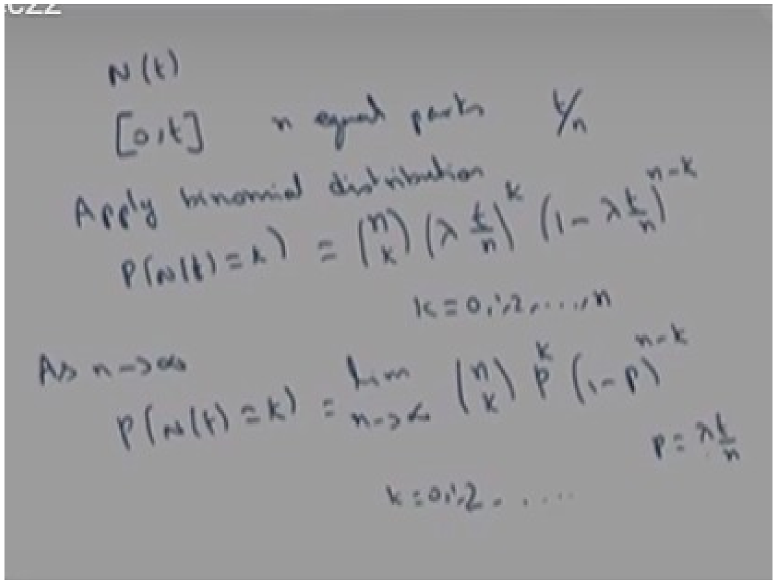
To create the Poisson process, you need few assumptions so that you can able to develop the Poisson process. The first assumption, in a small negligible interval, if the interval is $[t,t+\Delta t]$, then the probability of one arrival is going to be $\lambda \Delta t + O(\Delta t)$. The probability of one arrival occurs during the interval $[t,t+\Delta t]$ is going to be $\lambda \Delta t + O(\Delta t)$.
Here the $\lambda > 0$ and we are going to discuss what is λ and so on in the later, after explaining the Poisson process. So here the λ is going to be a constant and which takes the value greater than zero. And the $O(\Delta t)$ means as Δt tends to zero the $O(\Delta t)$ that is going to be tends to zero, as Δt tends to zero. So this is the first assumption.

The second assumption, the probability of more than one arrival is going to be an $O(\Delta t)$, in a same interval $[t,t+\Delta t]$, more than one arrival in this small negligible interval that probability is an order of Δt. That means, as Δt tends to zero, these values are going to tends to zero. Then the third assumption, occurrence of arrivals in non-overlapping intervals are mutually independent. Non-overlapping intervals are independent. So this is very important assumption.

That means, what is the probability that the arrival occurs in non-overlapping intervals that probability is same as the product of a probability of an arrival occurs in the each interval. Therefore, it is going to satisfy the independent property, occurrence of events in non-overlapping intervals are mutually independent. Therefore, the probability is going to be probability of intersection of all those things is same as the probability of individual probability and their product.

So with these three assumptions, we are going to develop the Poisson process.

(Refer Slide Time: 05:32)



So what I am going to do, since I started with the random variable N (t) is the number of arrivals in the interval [0,t], I am going to partition the interval [0,t] into n equal parts. I am going to partition the interval [0, t] into n equal parts. Since I made it the interval [0, t] into n equal parts, then each will be of the length t/n. And since I made the assumption the non-overlapping intervals are independent.

And the probability of one arrival is $\lambda\Delta t$ and the probability of more than one arrival is order of Δt and so on, therefore I can apply binomial distribution. The way I have partitioned the interval [0,t] into n pieces, therefore this is going to be , of a n intervals of interval length t/n. Therefore, I can say what is the probability that I can able to find out, what is the probability that k arrivals takes place in the interval, n intervals of each length t/n.

What is the probability that k arrivals takes place, therefore the possible values of k is going to be 0 to n and I can able to find out by using the binomial distribution, what is the probability that N(t) takes the value k. Since non overlapping intervals are independent and each probability of one arrival is $\lambda\Delta t$, where Δ t is a t/n. So each interval behaves as a Bernoulli trial, whether the arrival occurs or there is no arrival and like that you have n such independent trials.

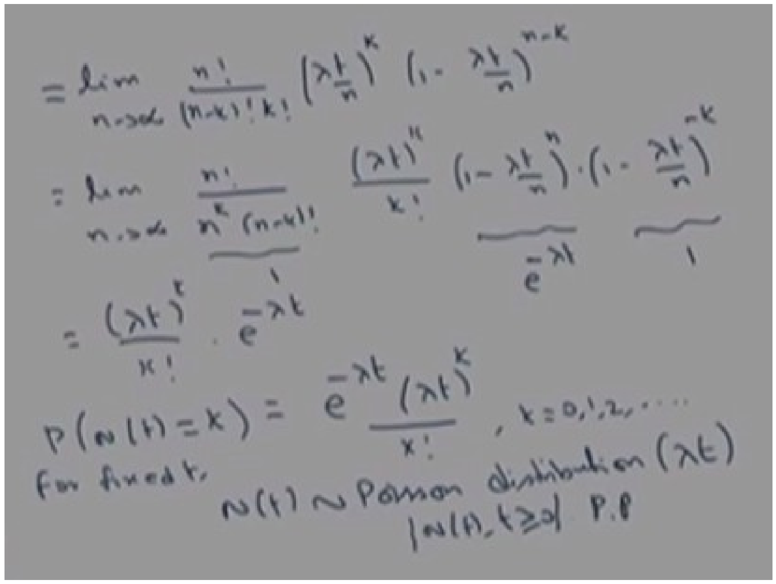
Therefore, the sum of n independent Bernoulli trials land up a binomial trials. Therefore, by using the binomial distribution, I can able to get what is a probability that N (t) takes a value k that is what is the possible ${}^{n}C_{k}$ way. And what is the probability of arrival takes place in one interval that is lambda times, this interval length is t/n, $\left(\frac{\lambda t}{n}\right)^{k}$. And what is the probability of no arrival takes place in each interval that is $\left(1-\frac{\lambda t}{n}\right)^{n-k}$.

So this is the way, I can able to get what is the probability that k arrival takes place in the interval [0,t] by partitioning, n intervals, so this is the probability. But the way I made a partition n equal parts, so now I have to go for what is the result as n tends to infinity. That means my interest is, what could be the result, if n tends to infinity, what is the probability that N (t) takes a value k, as n tends to infinity.

Therefore, the running index for n is going to be 0, 1, 2 and so on. What is the probability of N (t) takes a value k. That means in the right hand side, I had to go for finding out, as n tends to infinity what is the result for the right hand side, what is the probability of N (t) takes the value k. We take n tends to infinity because we need to study the limiting behavior of the stochastic process.

So that is same as limit n tends to infinity of ${}^{n}C_{k}$, I can make it as p^{k}, where p is going to be $\frac{\lambda t}{n}$ and $(1-p)^{n-k}$. Now I have to find out what is the result for limit n tends to infinity of this expression ${}^{n}C_{k}\,p^{k}(1-p)^{n-k}$, where p is going to be $\frac{\lambda t}{n}$.

(Refer Slide Time: 09:51)



If I do this simple calculation, let me explain, the limit n tends to infinity that is same as limit tends to infinity of ${}^{n}C_{k}$, I can make it as a $\frac{n!}{(n-k)!\,k!}$and that is $\left(\frac{\lambda t}{n}\right)^k$ and that is $\left(1-\frac{\lambda t}{n}\right)^{n-k}$ and that is same as the limit n tends to infinity of n factorial. And here this power n- k, I can take it outside. And n - k factorial and $(\lambda t)^k$and divided by k! . So this k! I take it inside.

And the power $\left(1-\frac{\lambda t}{n}\right)^{n-k}$, I split it into $\left(1-\frac{\lambda t}{n}\right)^{n}$and $\left(1-\frac{\lambda t}{n}\right)^{k}$. So, now I can look as n tends to infinity this is nothing to do with n. Therefore, $\frac{(\lambda t)^k}{k!}$ will come out. So this result is going to be $\frac{(\lambda t)^k}{k!}$. And this will land up as n tends to infinity. This is going to be $e^{-\lambda t}$ and this will land up one and this is also land up one as n tends to infinity. Therefore, I may land up it is e power minus lambda t. $e^{-\lambda t}$.

Hence the final answer of what is the probability that k arrival takes place in the interval [0, t] that is going to be $e^{-\lambda t}\frac{(\lambda t)^k}{k!}$. And the possible values of k can be 0, 1, 2 and so on. For fixed t, if you see this is same as, for fixed t it is going to be a random variable. For all possible values of t, it is going to be a stochastic process.

So for fixed t, the N (t) is a random variable and that probability mass function is $e^{-\lambda t}\frac{(\lambda t)^k}{k!}$. So λ is a constant. For fixed t, lambda into t, that is going to be a constant. Therefore, the right hand side look like the probability mass function of the Poisson distribution. Therefore, for fixed t, the N (t) is Poisson distribution. The random variable N (t) for fixed t, it is going to be a Poisson distribution with the parameter λt. λ is a constant.

And for fixed t, t is a constant. So λt, again this is going to be a constant. Therefore, for fixed t, it is going to be a Poisson distribution with the parameter λt. Therefore, for possible values of t, the N (t) is going to form a stochastic process. And since for fixed t, it is going to be a Poisson distribution, the collection of a random variable and each random variable is a Poisson distribution.

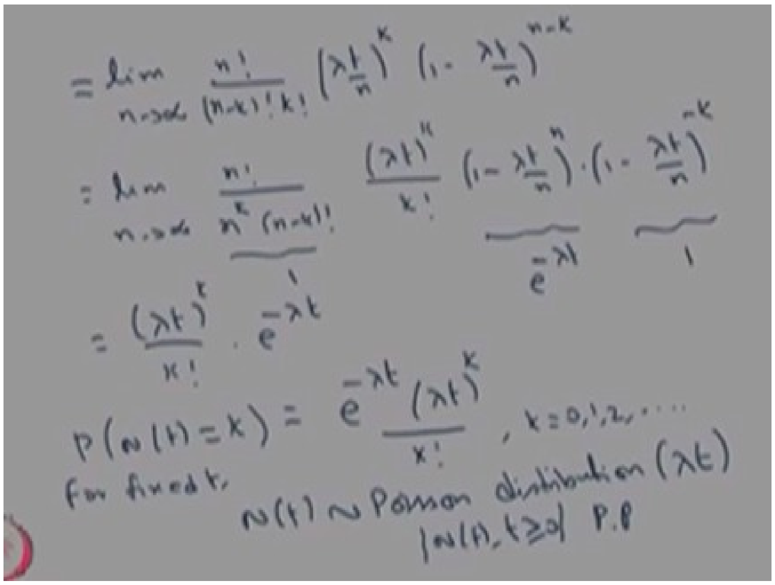
Therefore, this is going to be call it as the Poisson process. The way I have, we have explained earlier, each random variable is a Bernoulli distributed random variable, the collection of random variables is a Bernoulli process, similarly each Sn is going to be a binomial distribution, therefore the collection is going to be a binomial process. The same way, for fixed t, it is going to be a Poisson distribution, therefore the collection is going to be call it as the Poisson process.

So now we have developed N (t) is going to be a Poisson process.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 54

(Refer Slide Time: 00:01)

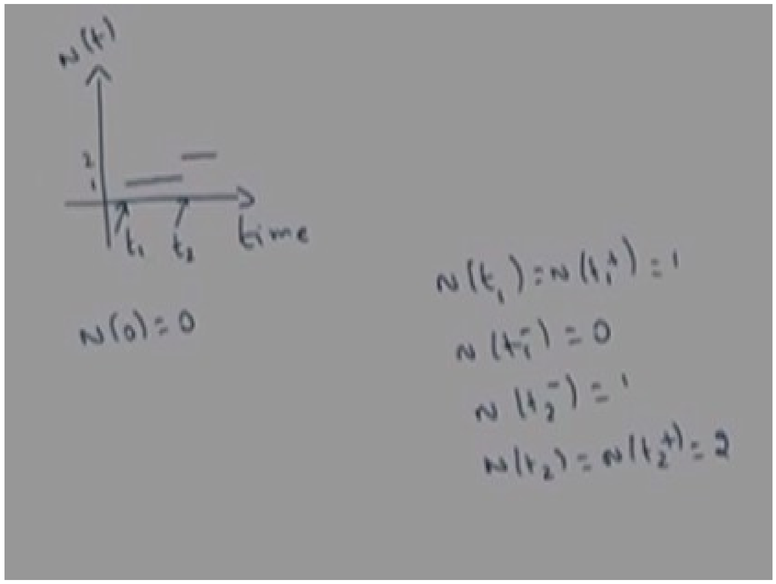


Here the λ is a constant and there is another name for the default Poisson process is called the homogenous Poisson process because there is another one called non homogenous Poisson process in which the λ need not be a constant. It can be a function of time t also. Therefore, the one we have derived now, it is a homogenous Poisson process in which the λ is a constant, which is greater than zero.

When λ is going to be a function of t, the corresponding Poisson process is called a non-homogenous Poisson process. So this is the one particular and very important continuous time or continuous parameter discrete state stochastic process that is a Poisson process or this is also we can say, this is going to be a very important continuous time arrival process that is a Poisson process.

The way we are counting N(t) is going to be a number of arrivals over the interval [0,t] or number of occurrence of the event over the t, the way you are counting over the time. Poisson process is an example of counting process. So, the N(t) is also called a counting process. So the Poisson process is also called as the counting process.

(Refer Slide Time: 01:24)



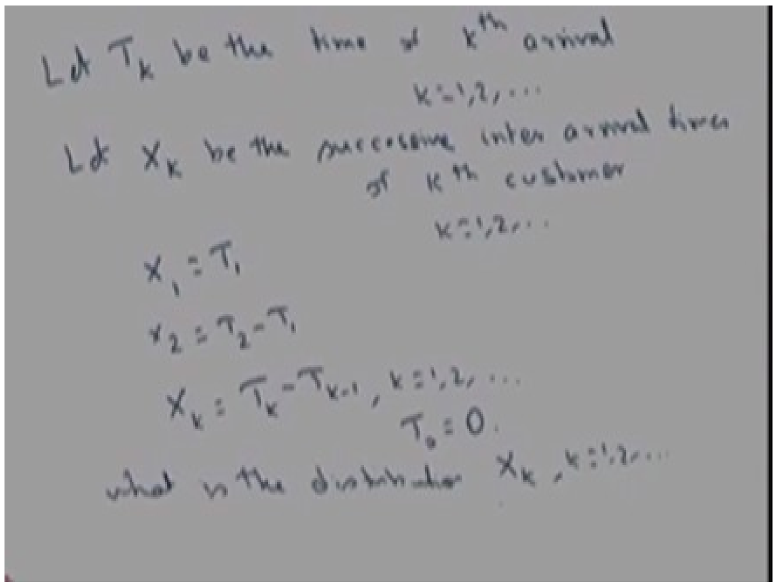
I can go for giving the sample path of N(t) over the time, what is the different values of N(t) is going to take. Obviously, N (0) = 0. Whenever some arrival occurs in some time, then the arrival is going to occur, therefore suppose the arrival occurs at this time, I make it as the up arrow. Then the value of N(t) is going to be incremented by one, till the next arrival comes. Suppose the next arrival takes place at this time point then the N(t) value is going to be one, till that time and it is going to be a right continuous function.

That means the time point in which the first arrival occurs, suppose you make it as t_1, so the N(t_1-) is going to be zero and the t_1 as well as N (t_1+) that is going to be one. Whereas the left limit N (t_1-) that is going to be zero. Suppose, the second arrival occurs at some point t_2, then the N (t_2-) that is the left limit at the time point t_2 that is going to be one.

And the N (t_2) that is same as N (t_2+) that is going to be two. So therefore, it is incremented by one, so the value is going to be two. So this is random time in which the arrival is going to occur and the way we have made the assumption in a very small interval only one, maximum only one arrival can occur. Therefore, the N(t) is going to be a non-decreasing right continuous and increased by jump of size one at the time epoch of arrival.

So whenever you see the sample path of the Poisson process, it is always going to be a non-decreasing right continuous and increased by a jumps of size one at the time epoch of arrivals. Now I am going to relate another random variable which involves in the Poisson process or I am going to discuss another stochastic process which involved in the Poisson process.

(Refer Slide Time: 04:00)



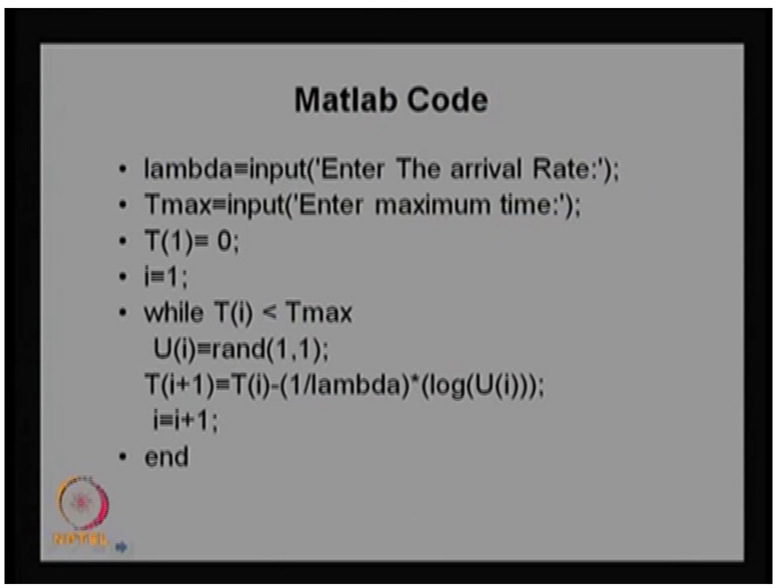
So for that, I am going to define the new random variable as let T_k be the time of k th arrival. So k can take the value one or two and so on. So therefore T be the random variable, takes what is the time point in which the k-th arrival occurs. That means the way I have given the sample path in the previous slide, the t1 and t 2, the small t1 and t2 are the different values of the T_k.

I am going to define another random variable $X_k$be the successive inter arrival times of k-th customer. So now the k can take the value one, two and so on. So the T_k be the time point, whereas the X_k be the inter arrival time. That means the X_1 is nothing but $T_1-T_0$and obviously T_0 is zero, therefore X_1 is same as T_1. And $X_2$is nothing but T_2-T_1. That means what is the inter arrival time for the second arrival, that inter arrival time is what time the first arrival occurs, that is the T_1.

And what time the second arrival occurs, that difference is going to be the inter arrival of the second customer. So this is the way I can define X_k is going to be, T_k – T_{k-1}. So now the running index for k can take the value one and so on. Obviously T_0 is going to be zero. Our interest is to find out what is the distribution of X_k for all k =1, 2 and so on. Is it feasible to find out the distribution of X_k? It is possible.

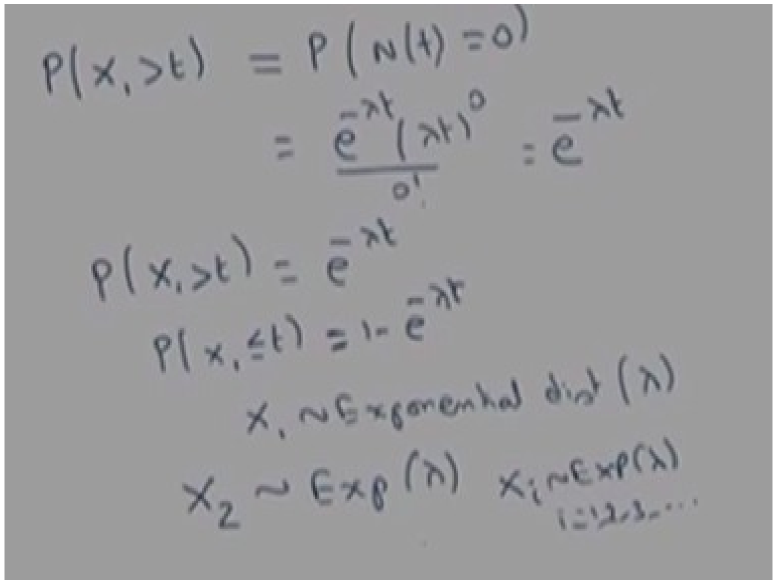
First we can start with k equal to one, what could be the distribution of X_1. Then once we get the X_1 distribution, the same analysis can be repeated to get the distribution of X_2 and X_3 and so on because the scenario which we are going to take it for finding out the distribution of X_1, that is the same as for X_2 and so on.

(Refer Slide Time: 06:58)



So now our interest is to find out what is the distribution of X_1. First we will try to find out that X_1only. Now we will find out the distribution of X_1.

(Refer Slide Time: 07:08)



Since X_1 is a continuous random variable, we can go for finding out what is the compliment cdf of X_1. So this is the compliment cdf of X_1. That is nothing but what is the probability that the first arrival occurs after time t. That is same as what is the probability that till time t, no customer enter into the system.

The left hand side is unknown, whereas the right hand side is the known one. So we are relating two different random variable. So here this is what is the probability that the first arrival occurs after time t. That is same as what is the probability that no arrival takes place during the interval [0,t]. But we know what is the probability of N(t) is equal to zero because just now we have made it.

For each t this is going to be a Poisson distribution with the parameter λt. Therefore, $P\{N(t)=0\}=\frac{e^{-\lambda t}(\lambda t)^0}{0!}$. And this is same as $e^{-\lambda t}$. So the left hand side is the unknown. The unknown is what is the P{ $X_1>t$ } that is same as $e^{-\lambda t}$.

Therefore, we can get what is the $P\{X_1\leq t\}=1-e^{-\lambda t}$.

So this is going to be a, what is a cdf for the random variable X_1. And the cdf of X_1 is same as the cdf exponential distribution with the parameter λt. Therefore, we can come to the conclusion,

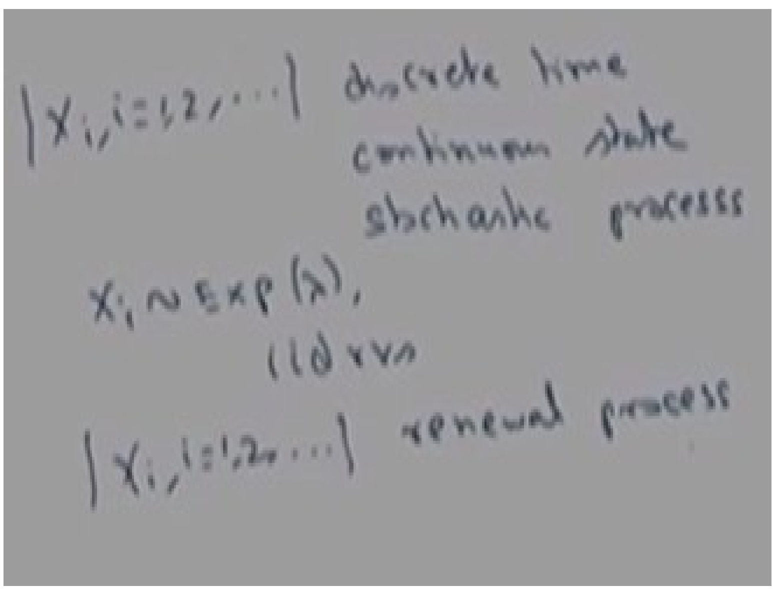
X_1 is going to be a exponentially distributed. The X_1 is exponentially distributed with the parameter λ.

So the unknown distribution X_1, first we are trying to find out what is the compliment cdf of X_1 and that land up to $e^{-\lambda t}$. Therefore, the cdf of X_1 is going to be $1-e^{-\lambda t}$.From this we conclude the X_1 is going to be exponential distribution with the parameter λ, where λ > 0.

The way we have compute the, the way we get the distribution of X_1, similarly one can show X_2 that is the inter arrival time of the second customer entry into the system, that is also can be proved, it is exponential distribution with the parameter λ. Not only X_2, we can go for the further all the $X_i's$, so we can able to prove all the $X_i's$ are going to be exponential distribution with the parameter λ for 'i' takes the value one, two and so on.

Not only that, we can able to prove all the $X_i's$ are independent random variable also and identical with the exponential, each one is exponential distribution with the parameter λ.

(Refer Slide Time: 10:32)



Therefore, the way we land up relating Poisson process with the inter arrival time, so this $X_i's$ will form a discrete time or discrete parameter, continuous state stochastic process in which each random variable X_i is going to be an exponential distribution with the parameter λ and all the

$X_i's$ are iid random variable also. And this each $X_i's$ are nothing but inter renewal time. Therefore, this is going to be, call it as renewal process.

We are going to discuss the renewal process in detailed later of this course. But here, I am just explaining how you will land, create the renewal process from the Poisson process. And the N(t) is the Poisson process for different values of t, whereas the inter arrival time that is the time in which the renewal takes place or the arrival takes place. Therefore, the renewals will form a stochastic process and that corresponding process is called a renewal process.

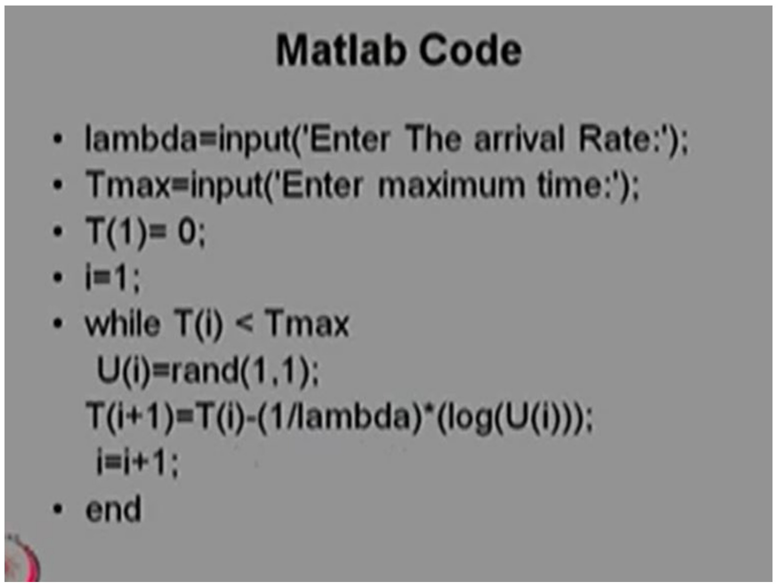
Therefore, this is going to be a one particular type of renewal process in which the renewal takes place of an exponentially distributed time intervals and all the times are iid random variables also.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 55

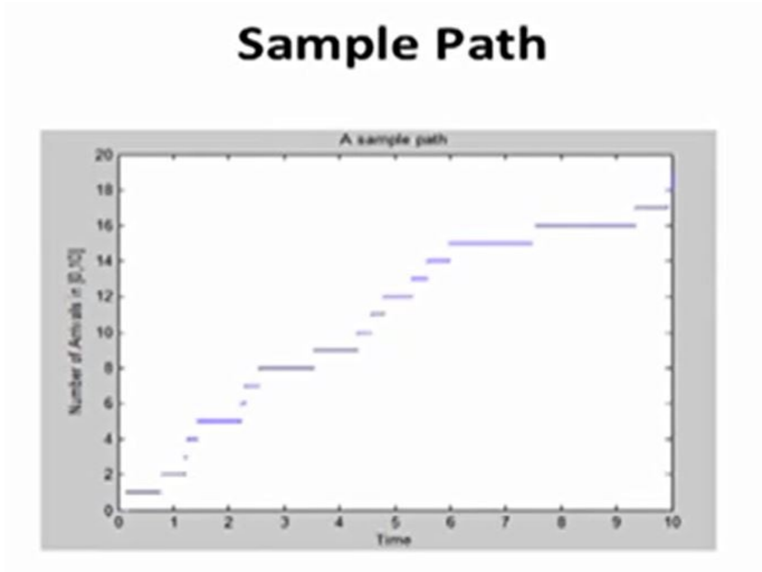
Now I am going to explain how we can create the sample path of the Poisson process using the MatLab code.

(Refer Slide Time: 00:08)



So since I said the Poisson process is related with the inter arrival times are exponential distribution, so I can start with the time 0 there is no customer in the system and I can go for what is a maximum time I need the sample path then I can keep on creating the random variable. From the random variable I can generate the exponentially distributed the time event then I can shift the time event bye T (i + 1) by adding the next exponentially distributed time event, then I can go for plotting the sample path.

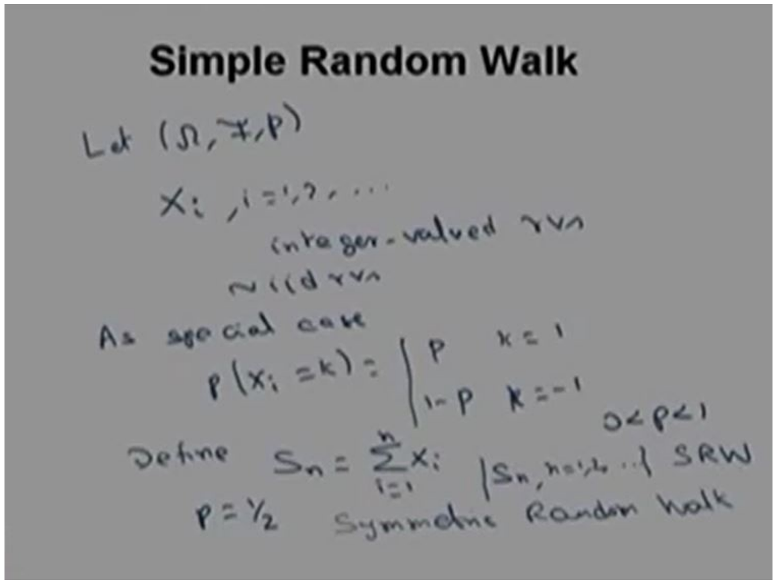
(Refer Slide Time: 00:41)



So this is the one sample path in which over the time from 0 to 10, the number of arrivals occurs in the interval 0 to 10 in the form of that means there is one arrival occurs at this time, therefore the N (t) value is incremented by 1 and it is taking the same value and when the second arrival occurs, then the increment is taken by 2 and so on.

So and if you see carefully the sample path you can find out the increment is always by 1 over the time and there is no 2 arrival or more than 1 arrival in a very small interval of time and you can able to see the inter arrival time that is going to be a exponentially distributed with a parameter λ whatever the λ I chosen in this sample path. So this is the way the sample path of the Poisson process looks like. Now we are going to discuss the third type of stochastic process that is a simple random walk.

(Refer Slide Time: 01:55)



So how we can create the simple random walk let me explain. You have a probability space, from the given probability space, you define a sequence of random variable X_i's and those random variables are integer valued random variables. Each X_i's are integer valued random variable not only that all the X_i's are iid random variables also.

All the X_i's are iid random variables and each one is integer valued discrete type random variable. As a special case, I can go for the random variable X_i takes a value 1 or -1 with the probability p and 1 - p. This is a special type of random walk in general I am going to define the, in general random walk also as a special case, I will go for the random variable X_i takes a value 1 with a probability p.

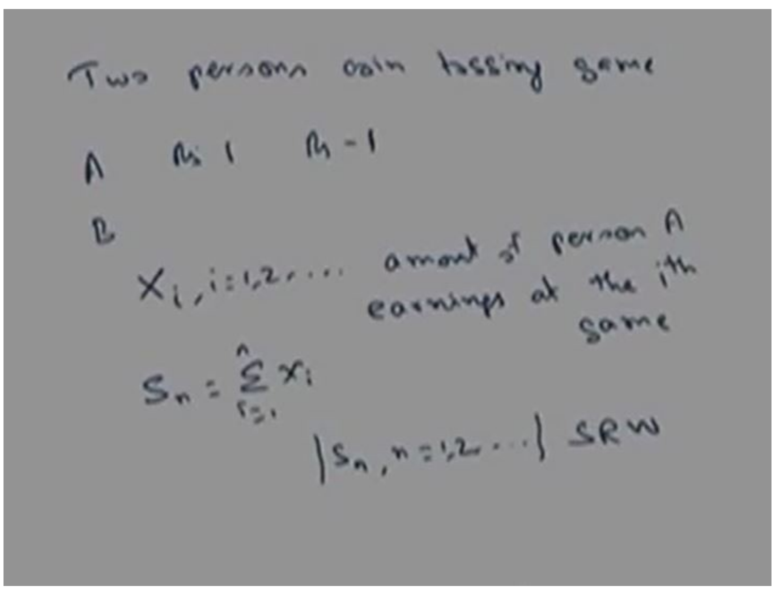
And X_i takes a value -1 with the probability 1 – p, where the p can take the value 0 to 1. Now I am going to define the random variable $S_n = \sum_{i=1}^{n} X_i$ that is going to form the random variable S_n and the Stochastic process Sn or the Stochastic sequence S_n for different values of n this will form a simple random walk. The S_n is going to form a simple random walk, why it is simple?

Because it is going to take an integer valued random variable and each value are going to take, each random variable is going to take the value 1 or -1, therefore this is going to be called as a simple random walk. In general, the k can take the any integers accordingly you will land up having S_n's are going to be a random walk.

And I am going to give the another special case, when $p = \frac{1}{2}$ that means each Xi random variable takes a value 1 with a probability $\frac{1}{2}$or -1 with a probability $\frac{1}{2}$, then that random walk is going to be call it as a Symmetric random walk. Why it is symmetric? Because with the probability $\frac{1}{2}$, it takes a forward one step or with the probability $\frac{1}{2}$it takes a backward one step.

Therefore, that type of a random walk is called symmetric random walk. In general, if we take a value 1 or -1 then it is called a simple random walk. If k can take any integers, then it is going to be call it as a generalized random walk.

(Refer Slide Time: 05:18)



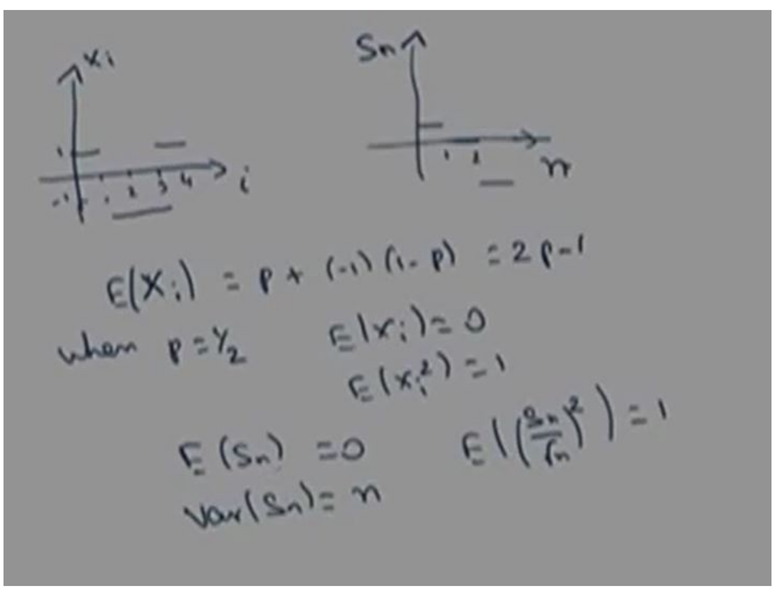
So this random walk can be created in a simple example of two persons coin tossing game also, this simple random walk can be explained by the example two persons coin tossing example in which you have a person A and B, if at the end of the coin tossing, if he is going to head then he is going to win Rs.1 or if at the end of the nth coin tossing, if it is going to get the tail then he is going to be lose in this game. If A wins, then B gives Rs.1 to A and if A loses, then A gives Rs.1 to B.

So accordingly, I can go for creating a random variable X_n or X_i for i is equal to 1, 2 and so on. Therefore, X_i denotes what is the amount of the person A earning at the ith game. Similarly, we can construct a stochastic process for player B and calculate the measures of interest. I can go for creating a random variable $S_n = \sum_{i=1}^{n} X_i$. Therefore, the S_n denotes what is the amount earned by the person A at the end of nth game, that's what a total amount.

So the X_i denotes how much he is going to earn at the end of each game, whereas the S_n is going to be the total amount earned by the person A at the end of a first n games. Therefore, this S_n is going to form a simple random walk, where X_i's are going to take a integer valued with a value 1 and -1 with the probability p it is going to take the value 1 or it is going to take the value -1 with the probability 1 - p.

So, I am just relating the simple random walk with the simple scenario of a two person's coin tossing game.

(Refer Slide Time: 08:01)



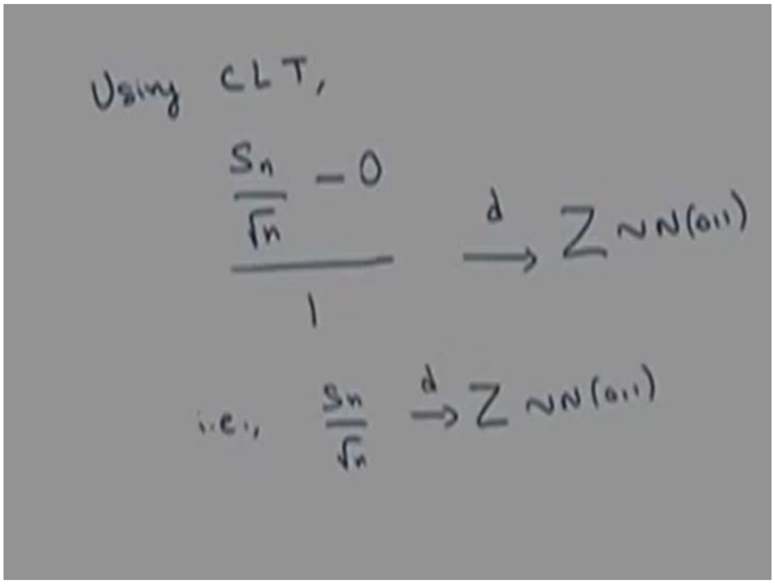
If you see the sample Path of the S_n, first I can go for what is a sample path of each X_i's, each X_i's can take the value 1 or -1, therefore it is going to take the value 1 or -1. Therefore, if X_1 takes a value 1, it is 1, if X_2 takes a value -1 it is like this if X_3 takes the value -1 then it is here. If X_4 takes the value 1 then it is like this, so this is a sample path of X_i over the i.

The way I have given the X_i's now I can go for writing what is the possible values of n and what is a possible values of S_n, so since X_1 is equal to 1, therefore S_1 is going to be 1 and X_2 is going to be -1, therefore it takes a value 1 - 1 therefore it is going to be 0 and X_3 is going to be -1, therefore S_3 is going to be -1 and X_4 is going to be 1, therefore it is going to be again 0.

So this is the way the sample path goes over the n. So, this is a one sample path for the possible values of X_i takes a value 1 and -1 accordingly. I have drawn the sample path of S_n over the n. Since X_i's are going to take the value 1 and -1 and with the probability p and with the probability 1 – p, it takes the value -1, I can go for finding out what is the expectation of X_i that is nothing but E[X_i¿ is equal to p + (- 1)(1 - p), therefore this is nothing but 2p - 1.

So, when I go for discussing the symmetric random walk when the p is equal to $\frac{1}{2}$then the expectation of each X_i is going to be 0 and also I can able to find out what is $E(X i^2)$ that is going to be 1, not only that when p is equal to $\frac{1}{2}$, I can able to find out what is the expectation of S_n that is going to be 0 and the variance of S_n is going to be n and I can go for writing what is an $E(\left(\frac{S_n}{\sqrt{n}}\right)^2)$, that is going to be 1. So the way I have got the result for expectation of X_i's and expectation of S_n, I can go for what is the limiting distribution of S_n.

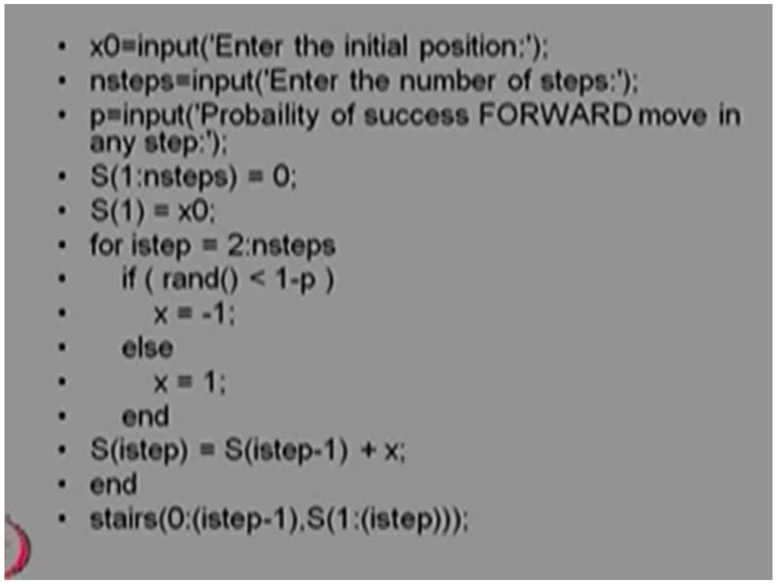
(Refer Slide Time: 11:02)



So using central limit theorem, I know what it is a mean for each Sn and I know what is the variance of each Sn also, therefore using a CLT I can able to conclude $\frac{S_n}{\sqrt{n}}$- the mean of this random variable is 0 divided by the standard deviation is going to be 1 and this has a n tends to infinity.

This will be a standard normal distribution, where Z is going to be a standard normal distribution has n tends to infinity and this convergence is via distribution that means I can able to conclude the distribution of $\frac{S_n}{\sqrt{n}}$as n tends to infinity in distribution this sequence of random variable will converge to the standard normal in distribution. I can go for creating what is a sample path of the simple random walk by using the MatLab code.

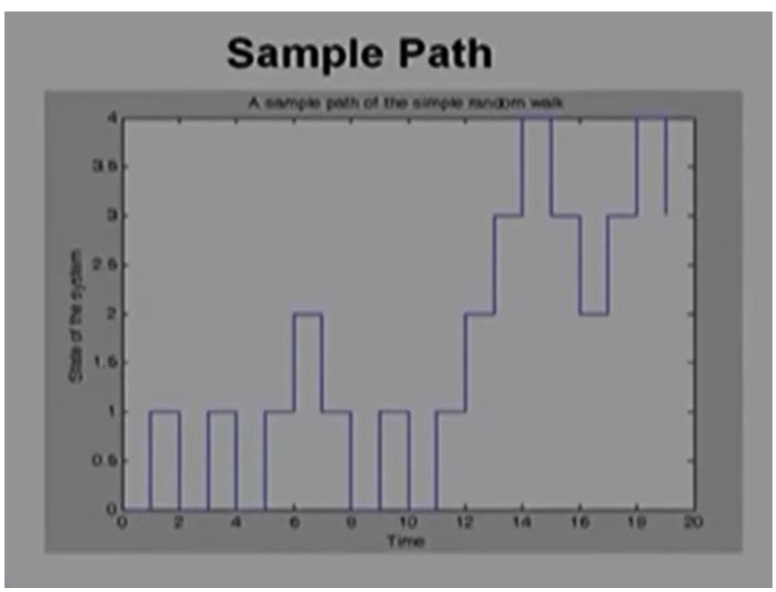
(Refer Slide Time: 12:11)



So for that I have to fix what is the initial position and what is a maximum number of steps I would like to go for finding the sample path and what is a probability of success in each, what is a forward move probability accordingly it is going to take the value 1 with the probability p and it is going to take the value -1 with the probability 1 – p.

So I am giving the value of p only and then I am just going for the possible values of S_n by adding the 1 or -1 accordingly, I am just writing the sample path of S_i's.

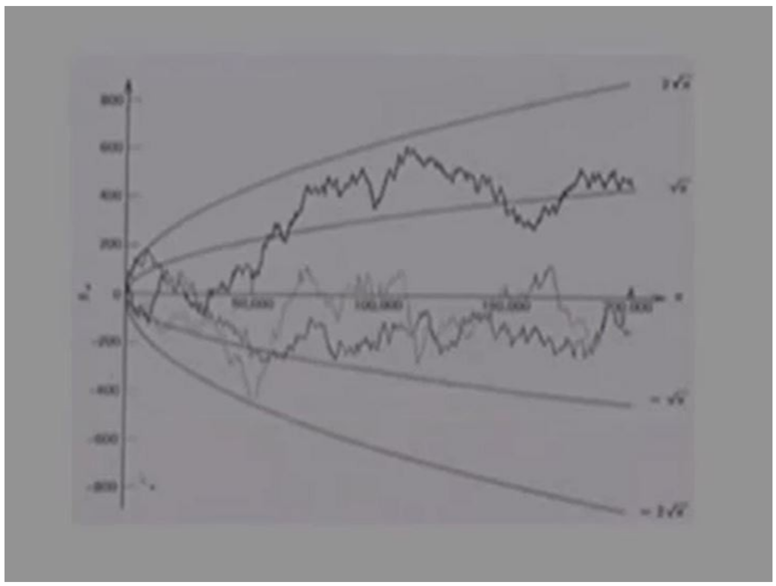
(Refer Slide Time: 12:48)



So if you see the sample path over the time 0 to 20 and each X_i's are going to take the value 1 or -1 accordingly, the S_n is going to take the same value or incremented by 1 or decremented by -1 according to the values of X_i's. Therefore, this is going to be the one sample path, which is depicted using the MatLab code.

So this is the earlier, I have shown the same graph.

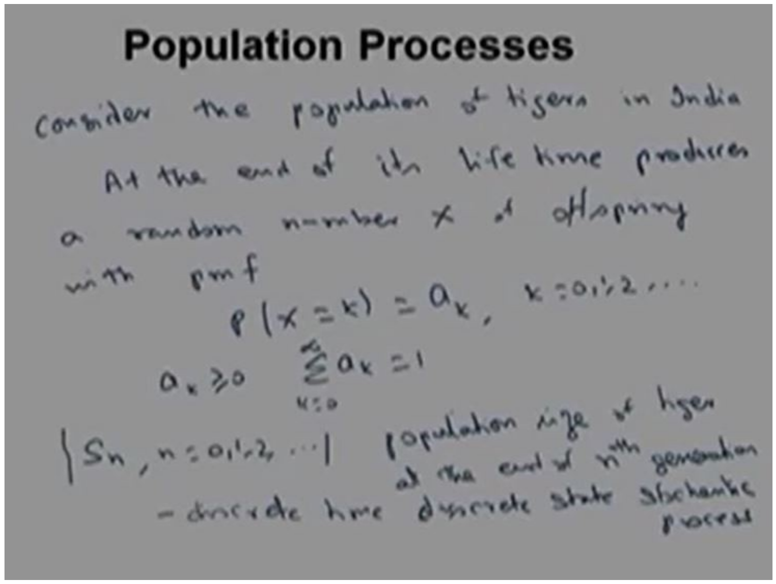
(Refer Slide Time: 13:27)



This is the S_n as n tends to infinity. Here you can see the different sample path for as n tends to infinity, you can find out what is the distribution of $\frac{S_n}{\sqrt{n}}$, as n tends to infinity also and this figures it has a three different sample path and one can observe what is the amount of a person A has n tends to infinity that depends on whether he is going to take the positive value or he is going to have the negative value depends on the 1st few games.

That it can be observed from this diagram, the first few results whether he is going to gain by 1 Rupee or he is going to lose by 1 Rupee. Accordingly, the possible values of Sn will go as n tends to infinity. Now we are going to discuss the fourth simple Stochastic process that comes in the population model. Now we will see the fourth simple stochastic process arises in the population model.

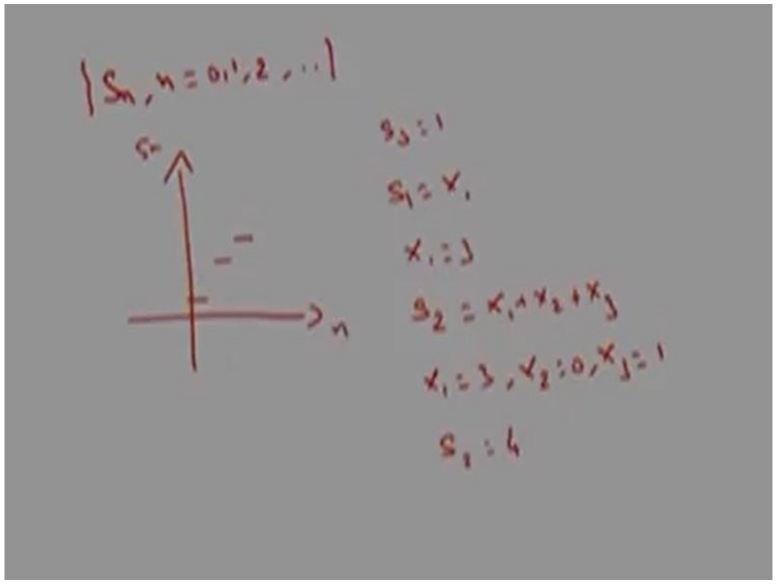
(Refer Slide Time: 14:30)



Consider a population of Tigers in India. For over the time, this is going to form a stochastic process, so I am going to make the assumption. At the end of its lifetime, it produces a random amount, random number X of off spring with the Probability mass function, that is the probability of X takes the value k that is a_k , where it satisfies.

a_k are going to greater than or equal to 0 and the summation is going to be 1 and also I am making the assumption, all the offspring's act independently of each other and at the end of their lifetime, individually can have a pregnancy accordance with the probability mass function, the same probability of X_i's takes the value k, with this S_n.

(Refer Slide Time: 15:31)



With this, S_n will form a discrete time and discrete state stochastic process, where S_n is the population size of tiger, at the end of the nth generation and if you see the sample path of S_n over the different generation, suppose you make it $S_0=$¿0 and suppose you make it $S_1=X_1$ and suppose X_1 takes the value 3 and then the second generation $S_2=X_1+X_2+X_3$and suppose you make it X_1 takes the value 3 and X_2 takes the value 0 and X_3 takes a value 1, then we have a S_2 is going to take the value 4.

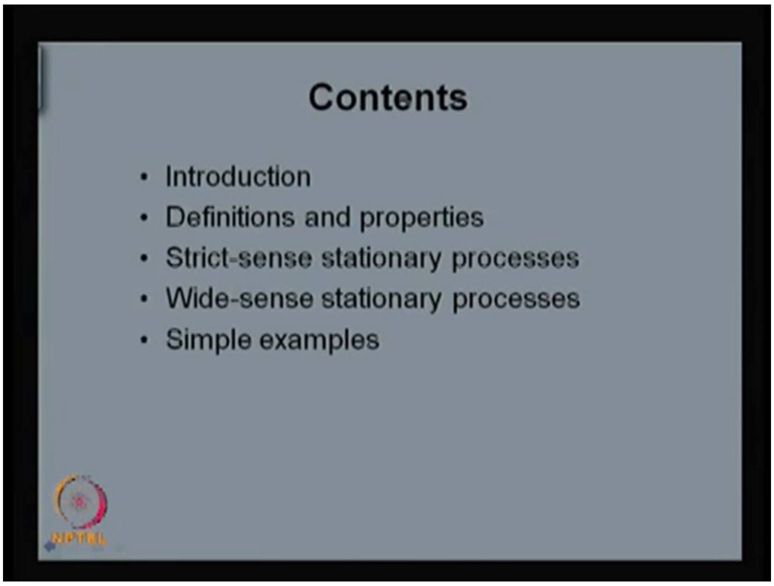
So, if you see the sample path of S_n over the n it is going to take the value 1, then it is going to take the value 3, then it is going to take the value 4 and so on. And this is the sample path of the population size of a Tiger over the nth generation and this is going to form a discrete time, discrete state stochastic process.

Thank you.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 56

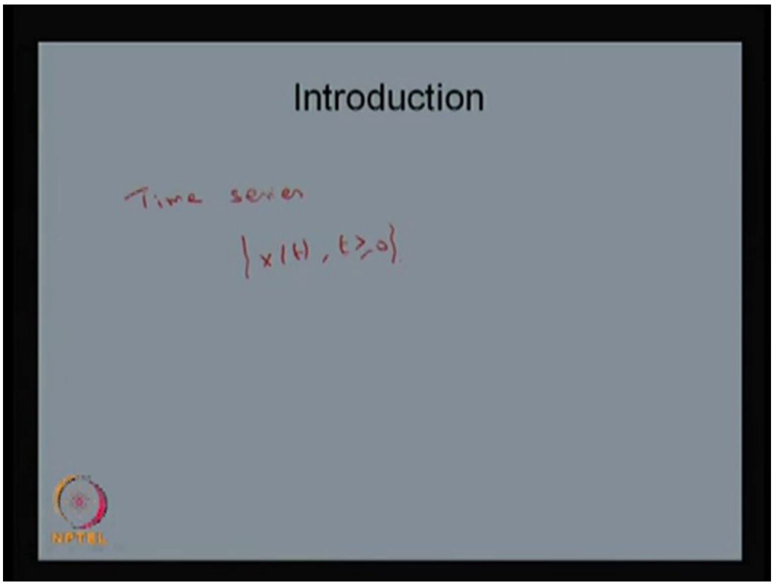
(Refer Slide Time: 00:01)



In this talk I am going to cover the introduction of the stationary process, and the few definitions and the properties of the stationary process. Then there are two important stationary processes; one is the strict sense stationary process, the second one is the wide sense stationary process. After this I am going to give few simple examples of stationary processes. Introduction, a stationary process is a stochastic process whose probabilistic laws remain unchanged through shifting times or in space.

Stationarity is a key concept in the time series analysis as it allows powerful techniques for modelling and forecasting to be developed. What is the meaning of time series? Time series is a set of data ordered in time usually recorded at regular interval of, regular in time interval. In probability theory a time series if you make out the, a time series is a collection of random variables indexed by time. Time series is a special case of stochastic processes.

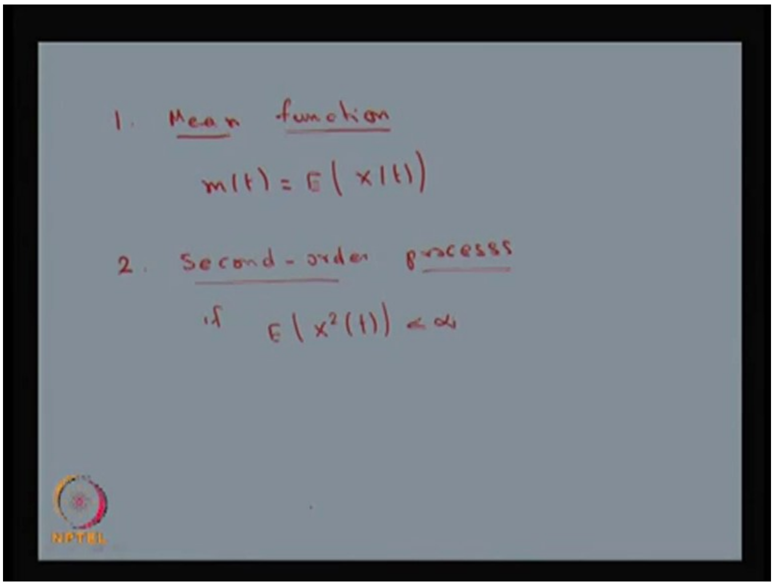
(Refer Slide Time: 01:20)



One of the main features of time series is the inter dependency of observation over time. This inter dependency needs to be accounted in the time series data modelling to improve temporal behavior and forecast of future movement. So, basically the stationary is used as a tool in time series analysis when the raw data are often transformed to become stationary; that means, if you collect the raw data and that raw data need not be satisfying the time series.

It need not satisfies a stationary property, but using the stationary property the time series of that raw data is a transformed so that you can model as well as you can forecast for the future moment by using the stationarity property. There are different forms of a stationarity depending on which of the statistical properties of the time series are restricted. The most widely used form of stationarity are strict sense stationarity and weak sense stationarity. So, basically before we go to the 2 types of, two important types of stationary property that is a weak sense stationary property. And strict sense stationarity property we will just see few definitions followed by these two important stationary properties.

(Refer Slide Time: 03:28)

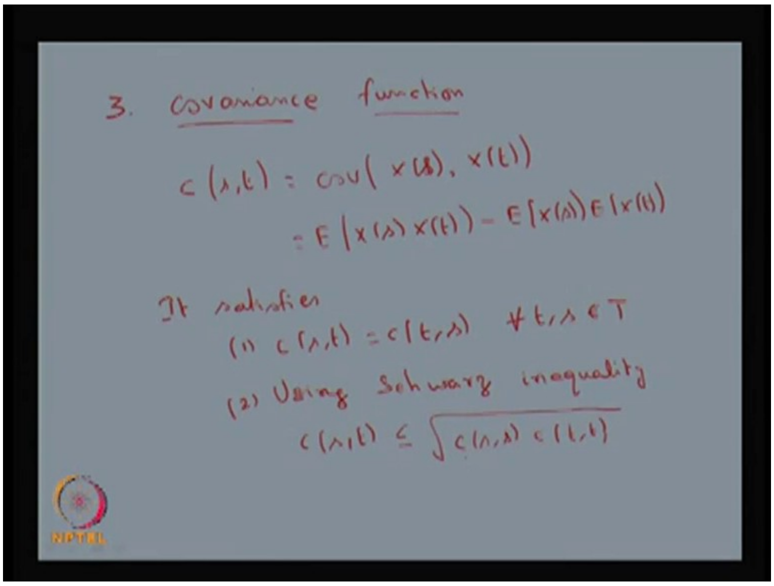


The first one is the mean function. Mean function is defined with the notation m(t), that is nothing but expectation of the random variable X(t). So, here the stochastic process is the collection of random variable X(t) over the $t \in T$ and you are defining the mean function as the function of t that is expectation of random variable X(t). Sometimes this is going to be a function of t, sometimes it is going to be independent of t, according to the function of t or independent of t we can classify the stochastic process later.

So, this definition is going to be very important, that is mean function. The second one it is a second order stochastic process. When we say a stochastic process is going to be a second order stochastic process, if it satisfies the condition the second order moment it's going to be finite for all t, if this condition is satisfied; that means, if random variables with the finite second order moment then that corresponding stochastic process is called a second order stochastic process. That means, there is a possibility the stochastic process may not satisfy the second order moment may be infinite or it would not exist. In that case, it is not going to be call it as a second order process.

So, whenever you collect the random variables from a stochastic process and satisfying the second order moments are going to be finite for all t, then we see that stochastic process is going to be a second order process.

(Refer Slide Time: 05:51)



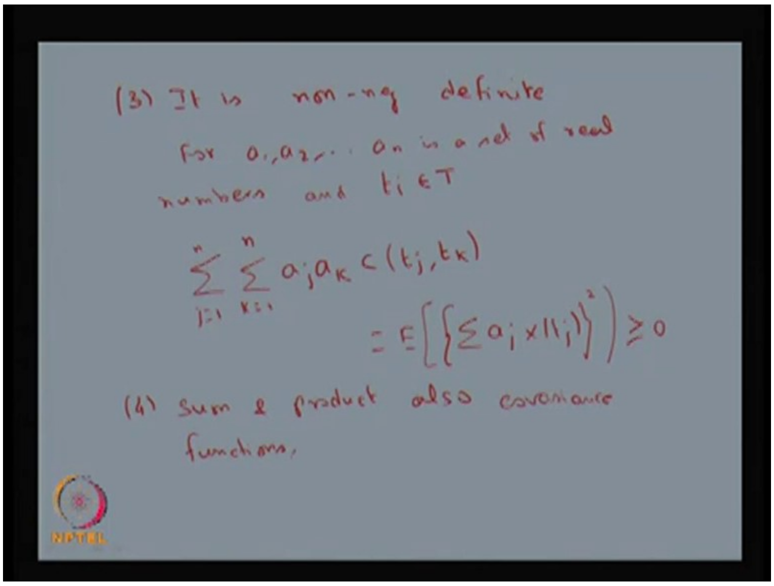
The third definition is covariance function. How to define the covariance function? Covariance function in notation it is c(s, t), that is nothing but covariance of two random variables X(s), X(t). Since it is a collection of random variable, so, for each t you will have one random variable so; that means, you have here you have taken two s and t and you got the corresponding random variable, and you are finding the covariance of these two random variables.

That is nothing but the E[X(s)X(t)]-E[X(s)]E[X(t)]; obviously, since you are finding the covariance of any two random variable; obviously, this stochastic process must be a second order stochastic process so that are the second order moments exist. And you are able to find out the covariance of this one; that means, the existence of the second order moment is going to be finite that is assume it to be that is assumed. And therefore, you are getting the covariance of these two random variables. So, using that you are defining c(s, t), that is a covariance.

Since it is expectation of the product minus expectation of the individual one, it is going to satisfy, it satisfies the first condition the c(s, t) is same as c(t, s) for all t, s belonging to T. Where capital T is a parameter space from the parameter space if you take any two t and s, then if you find out the covariance function of s, t is same as c(t, s).

The second property using Schwartz inequality, you can now always able to say the upper bound is going to be $\sqrt{c(s,s)c(t,t)}$. This is going to exist because the second order moments was finite therefore, c(s, s), that is nothing but the variance of X(s) and this is going to be the variance of X(t). And therefore, this is nothing but the product of the variance and the square root; so, this is going to be a finite quantity. Therefore, this has the upper bound of c(s, t) has the upper bound the $\sqrt{Var\left(X\left(s\right)\right)Var\left(X\left(t\right)\right)}$

(Refer Slide Time: 09:25)



The third property, it is the covariance matrix, non-negative definite also; that means, for $a_1, a_2, \ldots, a_n$ that is a set of real numbers, and if you take t_i's belonging to T and if we find this $\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k c\left(t_j, t_k\right)$ is nothing but the $E[\left|\sum a_j X\left(t_j\right)\right|^2]$ this expectation quantity is always going to be greater or equal to 0 since it is a whole square.

So, the expectation of whole square quantity is always greater or equal to 0 for all the set of all real numbers $a_1, a_2, \ldots, a_n$ and the $t_i^{'}s$ are belonging to, and this is nothing but the expectation of this quantity. And that quantity is always going to be greater or equal to 0; so, you can conclude the covariance of function is going to be a non-negative definite.

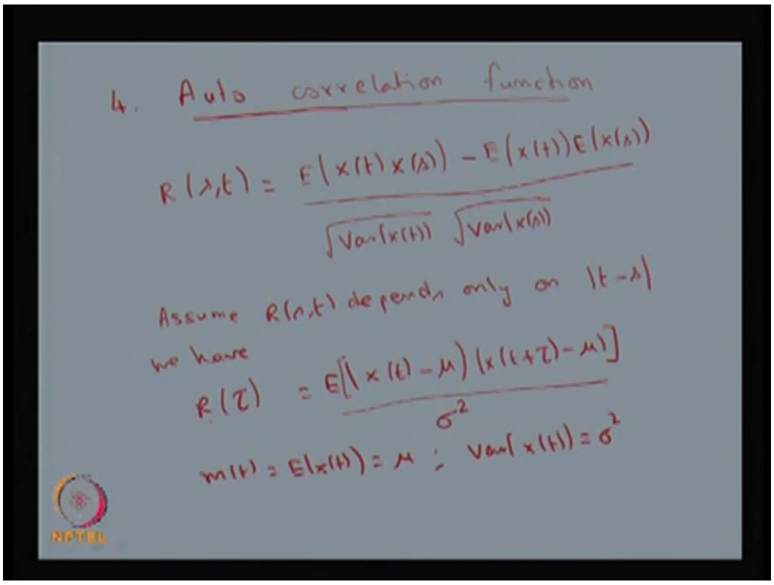
The 4th property if the sum as well as the product of any two covariance functions, also covariance functions. The sum and products also going to be the covariance function. These property needs elaboration. However, we assume these for this course. So, these 4 properties are going to be used later whenever you would like to cross check whether the covariance function is going to be satisfied or how to find out the covariance function so these properties will be used.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 57

Now we are moving into the 4th definition that is, Auto correlation, Auto correlation function.

(Refer Slide Time: 00:08)



The way we have defined the covariance function, now we are defining the auto correlation function. It is defined with the notation R(s, t) that is nothing but or we can write it in the terms of $\frac{E[X(t)X(s)]-E[X(t)]E[X(s)]}{\sqrt{Var(X(t))}\sqrt{Var(X(s))}}$.

So, the numerator can be written cov(X(t), X(s)), $\frac{Cav(X(t),X(s))}{\sqrt{Var(X(t))}\sqrt{Var(X(s))}}$ .So, this is going to be used in with the notation R(s, t), and this is going to be auto correlation function for the random variable X(t) and X(s). So, it basically describes the correlation between values of process at the different time points s and t.

Sometimes we assume the, we assume R(s, t) depends only on |t-s|. In the later case when you are discussing the stationary process it is going to depends only on the interval length not the actual time therefore, the R(s, t) is going to be depend only on the length of the |t-s| not the actual s and t.

Therefore, by assuming R(s, t) is going to only depends on t - s, we can have R of instead of two variables I can use the one variable as the R(τ); that is nothing but the $\frac{E[X(t)-\mu]E[X(t+\tau)-\mu]}{\sigma^2}$.

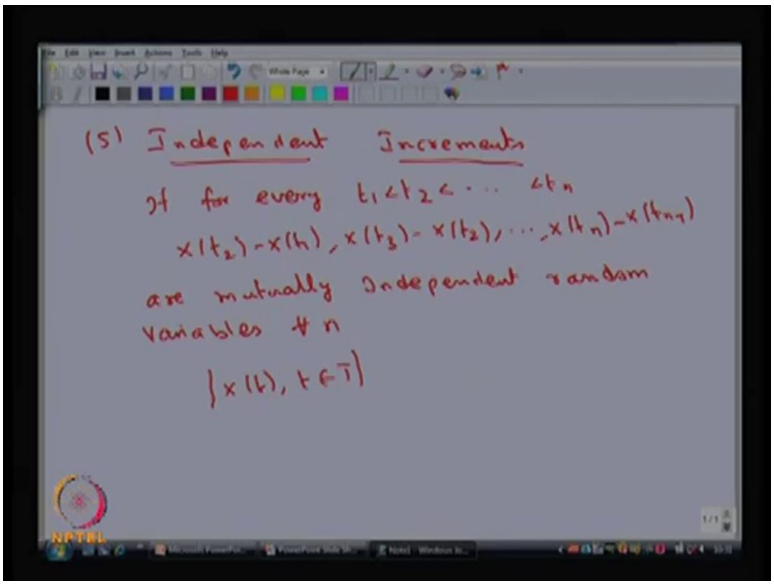
So, here I have made the one more assumption, the m(t) that is nothing but the expectation of X(t) that is going to be μ. And variance of X(t) is going to be σ^2, with that assumption only the R(τ) is going to be expectation of this productor divided by σ^2, where the variance of X(t) is going to be not a function of t it is a constant that is σ^2. And similarly, the mean function expectation of X(t) is going to be μ that is also independent of t.

Therefore, I can simplify this R(s, t), the product expectation minus individual expectation that can be simplified as expectation of this product and so, basically this is evaluated at X(t) and X(t +τ) and that difference is going to be τ. And this is also going to be even function; that means, it has R(τ) is same as a R(-τ). And this auto correlation function is used in time series analysis as well as signal processing.

In the signal processing, we assume that the signal the corresponding time series satisfying the stationary property therefore, the stationarity property implies the auto correlation function is going to be depends only on the length of the interval not the actual time. Therefore, these R(τ) will be used in the signal processing as well as in general time series analysis also.

The fifth definition, so be we are covering the different definitions which we want.

(Refer Slide Time: 05:10)



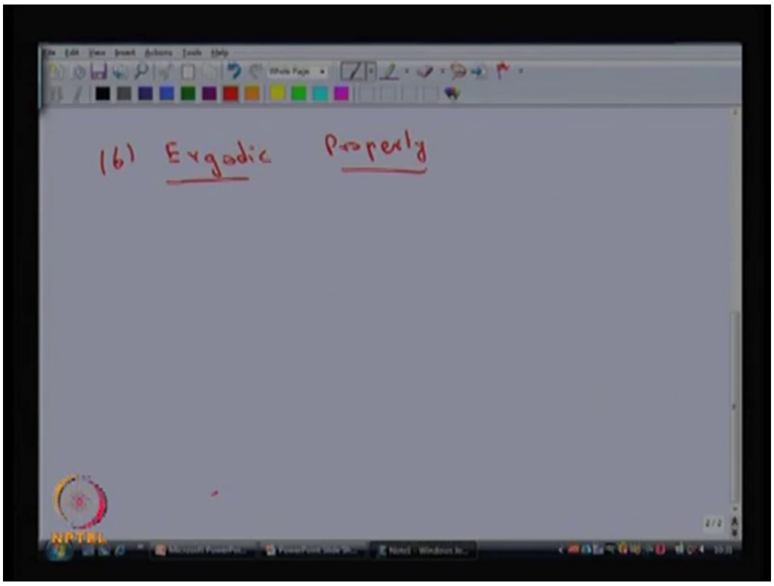
The fifth definition, first we started with the mean function, second we started with the second order stochastic process, then third we have given the covariance function and the 4th we have given the auto correlation function.

Now, we are giving the fifth definition that is independent increments. If for every $t_1<t_2<\ldots<t_n$, the random variables $X(t_2)-X(t_1), X(t_3)-X(t_2), \ldots, X(t_n)-X(t_{n-1})$are mutually independent, mutually independent random variables for all n then we say; the corresponding stochastic process is having independent increment property.

So, whenever you take few t_i's $t_1, t_2, \ldots, t_n$ and the increments that is $X(t_2)-X(t_1), X(t_3)-X(t_2), \ldots, X(t_n)-X(t_{n-1})$. So, these are all going to be the increment and each one is a random variable therefore, the increment is also going to be a random variable. And you have n such random variables and suppose these n random variables are mutually independent random variable for all n. So, this is fixed for one n, like that if you go for all n if this property is satisfied then we can conclude the corresponding stochastic process having the property of independent increments.

So, the independent increment that does not imply some other properties, but here what we are saying is the increment satisfies the mutually independent property. That means, if you find out the CDF of the joint CDF of this random variable that is same as the product of the individual CDF, that is property satisfied by all the n then you can conclude that stochastic process has the independent increment.

(Refer Slide Time: 08:30)

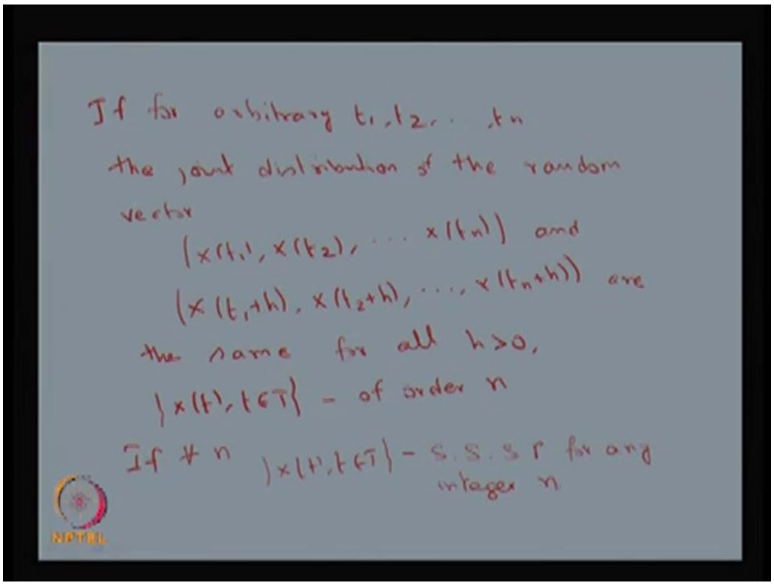


The next property or the next to definition is Ergodic Property. What is the meaning of Ergodic property? It says the time average of a function along a realization or sample exists, almost everywhere and is related to the space average. What it means? Whenever the system or the stochastic process is Ergodic the time average is the same for all almost initial points that is, the process evolved for a longer time forgets it's initial state.

So, statistical sampling can be performed at one instant across a group of identical processes or sampled over time on a single process, with no change in the measured result. We will discuss the Ergodic property for the Markov process in detail later, but this Ergodic property is going to be very important, when you study the Markov property or when you study the stationarity property therefore, these Ergodic properties always goes along with the stationarity property or goes along with the Markov property therefore, the stochastic property is going to behave in a different way, and that we are going to discuss later.

The most important stationary process that is a Strict sense stationary process.

(Refer Slide Time: 10:15)



First let me start with the strict sense stationary process of order n, then I will define the strict sense stationary process for all order n or the exist strict stationary process itself. If for arbitrary$t_1, t_2, \ldots, t_n$, the joint distribution of the random vector that is $(X(t_1), X(t_2), \ldots, X(t_n))$.

And the another random vector that is $(X(t_1+h), X(t_2+h), \ldots, X(t_n+h))$ are the same; for all h which is greater than 0 then we say, the stochastic process is a strict sense stationary of order n because here we restricted with the n random variable.

So, we take n random variable taken at the points $t_1, t_2, \ldots, t_n$and a find out the joint distribution of $(X(t_1), X(t_2), \ldots, X(t_n))$. So, you can find out what is a joint distribution of this n random variable. Also you find the joint distribution of n random variable shifted by h. That means, earlier the random variable $X(t_1)$, now you have a random variable $X(t_1+h)$, with the same shift h you do it with the t_2 therefore, the random variable $X(t_2+h)$.

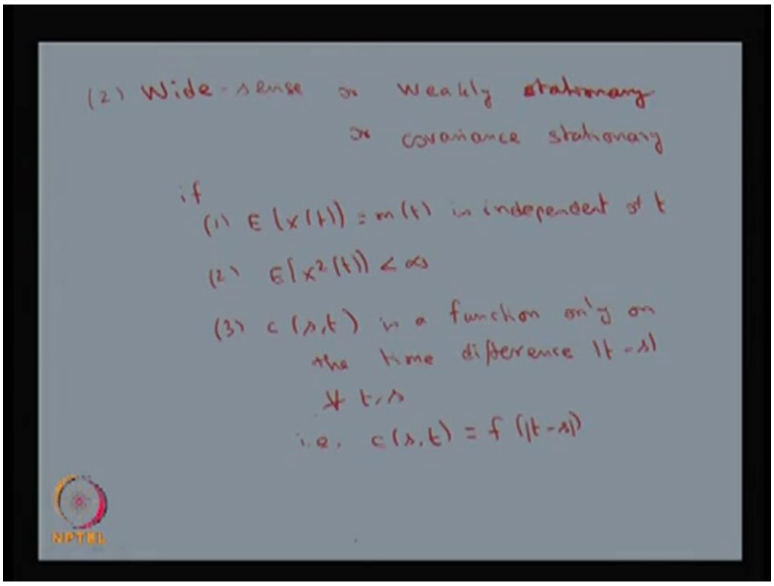
Similarly, the n th random variable is $X(t_n)$ earlier, now you have a random variable $X(t_n+h)$. So, you have another random vector with n random variables and find out the joint distribution of that. If the joint distribution of this first n random variable as well as the joint distribution of the shifted by h that random variable. If both the distributions are same; that means, they are identically distributed, the joint distributions are going to be identical, then you can conclude this stochastic process is a strict sense stochastic process of order n because you use the n random variable.

If this is going to be satisfied, the order property is going to be satisfied for all n; then you can conclude the stochastic process is going to be a strict sense stationary process for any integer n. This is going to be a strict sense stationary process for any integer n. So, we start to cross check the joint distribution of n random variable. So, if it is satisfying only with the maximum sum integer, then it is going to be a strict sense stationary process of order that n if it is going to be satisfied for all n, then for any integer n, then it is going to be call it as just strict sense stationary process.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 58

(Refer Slide Time: 00:07)



The next definition is a wide sense. The next one is a wide sense or weakly stationary or is another word covariance stationary process. When we say, the given stochastic process is going to be a wide sense or weekly stationary or covariance stationary, if it is going to satisfy the following properties, following conditions; mean function that is m(t) is independent of t.

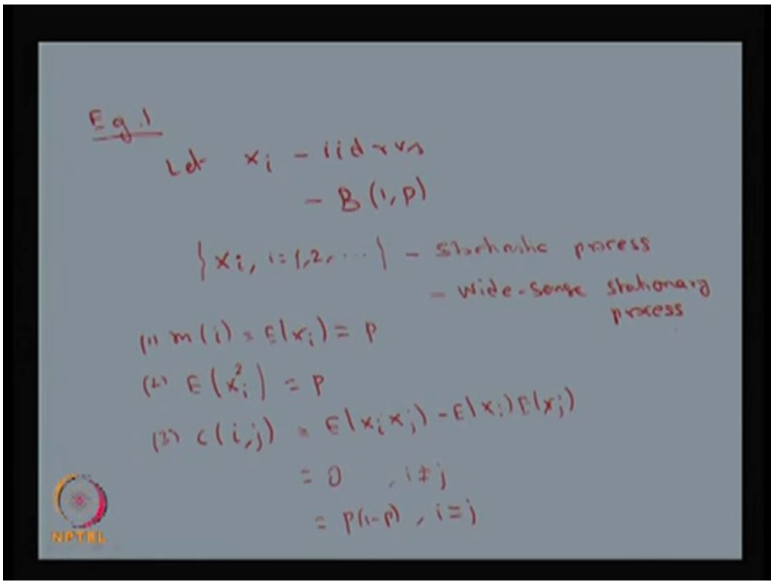
The second condition, the second order moment is going to be finite. Basically, the stochastic process is going to be, second order moments are going to be finite. Third condition if you find the covariance function c(s, t) that function depends only on the time difference, t - s for all t, s. If you find the covariance function for any two random variables X(s) and X(t) then that is always going to be a function of the only the difference t - s not the actual t or actual s; that is, inverts in mathematically you can write c(s, t) that is going to be a function of t minus s in absolute.

If these three properties are going to be satisfied by any stochastic process, then we say that stochastic process is going to be a wide sense or weakly stationary or covariance stationary. This is entirely different from the strict sense stationary. The strict sense stationary you are finding the joint distribution of n random variables, then find the joint distribution of n random variable shifted by h and for all h > 0, and for all n if that property is satisfied then we say that is a strict sense stationary process whereas, here we check the mean function is going to be independent of t.

And the second order moment is going to be a finite value and the covariance function is going to be a function of only the difference of t - s therefore, any stochastic process satisfying the strict sense stationary process, strict sense stationary property that does not imply the wide sense stationary property. As well as the wide sense stationary process need not satisfy all the strict sense property therefore, you cannot imply one stationary process that does not imply the wide sense and the wide sense stationary process that does not imply the strict sense stationary process.

So, in the strict sense process, what we are saying is it is a stochastic process whose joint distribution does not change; when shifting time or space. As a result, the parameters such as the mean and variance if they exist, also do not change over the time or push or the strict sense stationary process. Now I am going to give few examples for the stationary process, maybe it could be strict sense stationary process or it could be a wide sense stationary process.

(Refer Slide Time: 04:43)



The first example let X_i is going to be iid random variables, independent, identically distributed random variable and assume that each one is going to be Bernoulli distributed random variable with the parameter p, a notation it's a binomial distribution with the parameters. So, 1 and p that is same as each X_i's are Bernoulli distributed random variables with the parameter p.

Now, I am creating a stochastic process with those such iid random variable, in which each random variable is a Bernoulli distributed random variable therefore, this is going to be a stochastic process. Now you can verify whether it is going to be a strict sense stationary process or a wide sense stationary process. The assumption is all the random variables are mutually independent and each random variable is identically distributed, which is a Bernoulli distributed.

So, this is just for examples sake we have taken, and if you find out the mean function for each random variable that is going to be expectation of X_i and that is going to be the mean of Bernoulli distribution is going to be p which is independent of i. The second condition if you find out what is the second order moment of, the second order moment is going to be, $1^2 p+0(1-p)$. So, therefore, that is also going to be p.

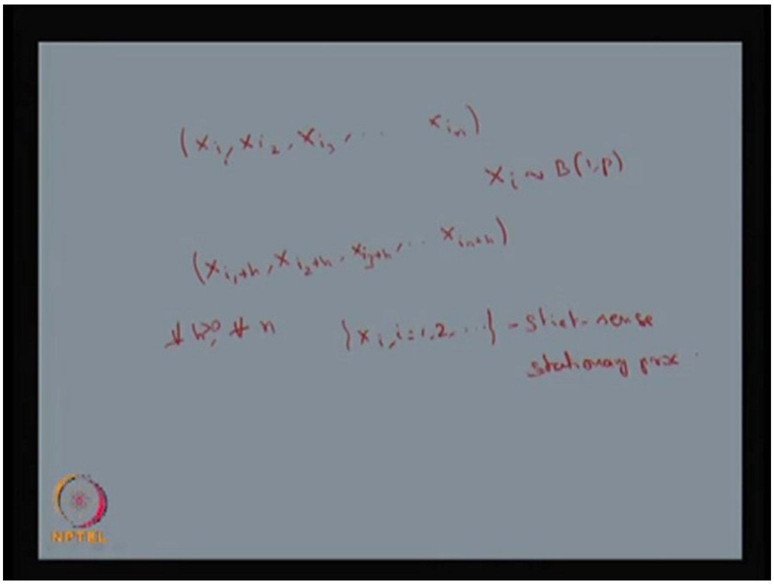
So, if you find out c(i, j) instead of c(s. t), you have I, j that is nothing but $E[X¿¿iX_j]-E[X_i]E[X_j]¿$. If you find out this quantity this is nothing but X_i and X_j and since they are independent random variable therefore, $E[X¿¿iX_j]¿$ is nothing but the $E[X_i]E[X_j]$that is same as this one therefore, this is going to be 0.

For all i is not equal to j, for i is equal to j that is nothing but $E\left[X_i^2\right]-E\left[X_i\right]^2$ that is nothing but the variance. And the variance of the random variable Bernoulli distribution that is going to be npq therefore, that is going to be p(1-p) for i is equal to j. And this value is independent of, these values is going to be a function of i - j you can make out therefore, since these all three properties of the weakly stationary property or wide sense stationary property is satisfied therefore, this is going to be a wide sense stationary process.

In fact, even if the random variables are simply i.i.d. then too we can check that the process is wide sense stationary, for illustration purpose we have discussed Bernoulli process. More examples on continuous time stochastic processes are discussed in the problem sheet.

Now, we can cross check whether this is going to be a strict sense stationary process also.

(Refer Slide Time: 09:43)



If you find out the joint distribution of a suppose, you take a few random variables $(X_{i_1}, X_{i_2}, \ldots, X_{i_n})$. So, this is the n such random variable and each random variable is Bernoulli distributed with the parameter p and all are independent therefore, the joint distribution is going to be the product of individual distribution.

And if you shift these i_1 with some number h and $(X_{i_1+h}, X_{i_2+h}, \ldots, X_{i_n+h})$. You shift those random variable with the h if you find out the joint distribution and since each one is independent random variable therefore, the joint distribution by shifted by h that is also going to be the product of those n random variables product therefore, the distributions are again going to be identical; because they are, because each random variable is identical as well as mutually independent therefore, the joint distribution is going to be product. And all are going to be identical therefore it is going to be power m of the distribution.

So, this is going to be satisfy the strict sense property that is a joint distribution of this random variable and joint distribution of this random variable are going to be same for all h as well as for all n also. Since, it is satisfied for all $h > 0$ and for all any integer n therefore, this is going to be the same collection of a random variables the stochastic process is going to be a strict sense stationary process.

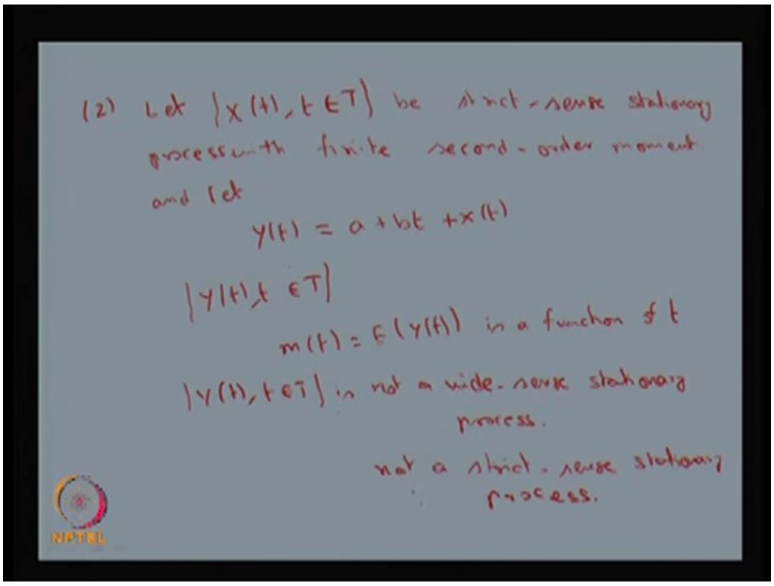
So, this is the cooked up example in which this stochastic process is going to be a strict sense stationary process, as well as a wide sense stationary process, but there are many situation in which stochastic process may be a strict sense; not the wide sense and the some stochastic process may be a wide sense stationary process not the strict sense stationary process. And how this particular stochastic process become a strict sense and a wide sense because of each random variable is a mutually independent as well as identical therefore, it is going to be a strict sense stationary process as well as a wide sense stationary process.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 59

I will have another example in which it is going to be, only the it would not be a strict sense itself.

(Refer Slide Time: 00:10)



Let me start with the example in which this stochastic process is a strict sense stationary process. The given X(t) is a strict sense stationary process with a finite second order moment. So, you do not want the finite a second order moment for the strict sense stationary process, but I have taken as an example. The given X(t) is going to be a strict sense stationary process along with the finite second order moment.

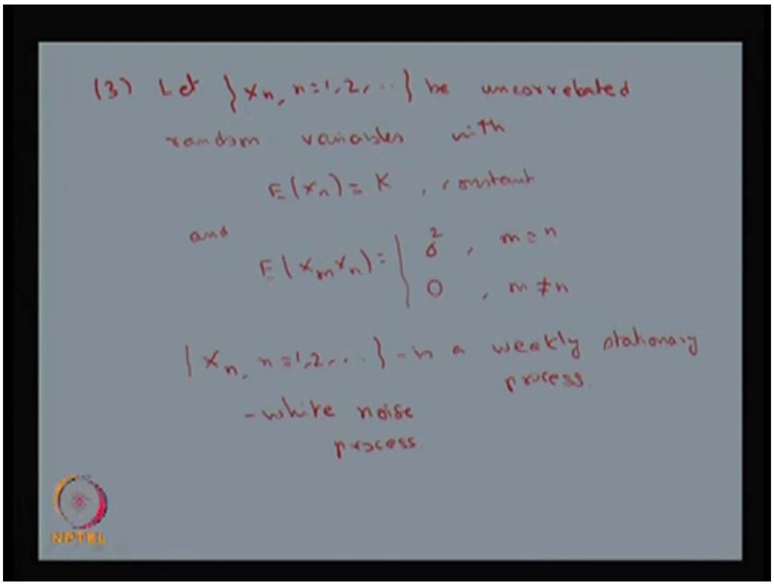
Now, I am going to define another stochastic process with the random variable Y(t); that is a +bt+ X(t). So, this is going to be a stochastic process. This is a stochastic process Y(t). Now we want to check whether the Y(t) is going to be a stationary strict sense stationary process or not, as well as whether this is going to be a wide sense stationary or not. The X(t) is the strict sense stationary process, suppose you find out the mean for this random variable, the mean for this random variable, if you find out the mean for the Y(t) where a and b are constant therefore, this is going to be a function of t.

Since a and b are constant the mean of Y(t) is a function of t therefore, this is a function of t, since it is not satisfying the first property of, the first condition to become a wide sense stationary process. Therefore, the Y(t) is not a wide sense stationary process. We started with the strict sense stationary process, and we created a new stochastic process Y(t), that is a + bt +X(t) where a and b are constant. Now if you find out the mean of Y(t) mean function that is going to be a function of t. That is nothing but that depends on t therefore, Y(t) is not going to be a wide sense stationary process whereas, X(t) is a strict sense.

Now, similarly you can cross check whether the joint distribution of a Y(t), and shifted by h, t shifted by h you can conclude, this is also not going to be a since it is a function of t, since the mean is going to be a function of t and the Y(t) also involves the function of t as well as a X(t). Even though X(t) is a strict sense stationary process, the way you made a + bt+ X(t), you will land up the joint distributions are going to be different by the t with the shifted t + h it would not be satisfied. Therefore, you can conclude Y(t) is not a strict sense stationary process also.

That means from this example we can conclude whenever you have a strict sense stationary process, if you make a +bt+ X(t) definitely the Y(t) is not going to be a wide sense stationary process and as well as a strict sense stationary process. We go for the third example.

(Refer Slide Time: 04:51)



In this third example, let me start with the stochastic process be here this each random variable are uncorrelated random variables with the mean of each random variable is going to be some constant K, which may be, assume it to be 0 in some situation. So, in general you keep the mean of each random variable is going to be some constant K.

And you make $E[X_m X_n]$, that is going to be it is variance for m = n, and all other quantity you make it 0. Not only this each random variables are uncorrelated random variable; that means, if you find out the correlation coefficient that is going to be 0, and the mean is going to be constant and the expectation of the product of any two random variables if they are different it is 0 and obviously if they are same.

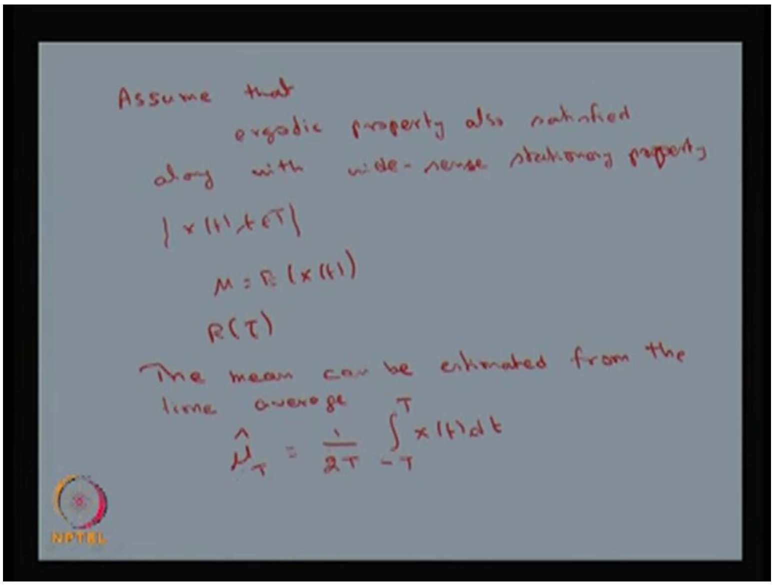
Since you make the assumption therefore, this is going to be a variance σ^2. If you cross check all the properties of, all the conditions of the wide sense stationary property starting with the mean function and second order moment exists that is finite and covariance function of any two random variables is going to be a function of only the difference. There all those three conditions are going to be satisfied therefore, you can come to the conclusion, I am not working out here this is going to be a weakly stationary process or wide sense stationary process. Or it is going to be call it as a covariance stationary process also.

And this stochastic process is also called white noise process. This is very important in the signal processing you keep the uncorrelated random variable with this assumption, the mean is going to be a constant which may be 0. And the product of expectation is going to be these values, and this is going to be a weakly stationary process in the sense it satisfies all three conditions of that weak sense or wide sense stationary process, and this stochastic process is called a white noise process.

Note that this stochastic process, we did not make the distribution of each random variable X_n what is the distribution of X_n is not defined here, without that we give all the assumptions of the mean and variance. Therefore, this is going to be very useful in the time series analysis as well as the signal processing. And this particular stochastic process is called the white noise. And sometimes we make the assumption the X_n's are going to be normally distributed random variable also.

But in general, we would not to define; we would not to give what is the assumption; what is the distribution of X_n. Without that with this stochastic process is going to be call it as a white noise process.

(Refer Slide Time: 08:59)



Addition to the wide sense stationary process one can assume, one can assume that ergodic property also satisfied along with wide sense stationary property. For illustration purpose we have discussed a Bernoulli process. That means, the given stochastic process is a wide sense stationary process. As well as ergodic property is also satisfied.

In that case, the mean function is going to be some independent of t, that you can make it as the mu. And auto correlation function is going to be R(τ) only because it is a wide sense stationary process. Therefore, the mean is independent of t and the auto correlation function is going to be a function with the only τ. And we have ergodic property therefore, you can find the mean can be estimated from the time average so, this is possible only if the ergodic property satisfied.

So, the mean can be estimated with the up arrow; that means, the estimator estimation of a mean, that is same as $\frac{1}{2T}\int_{-T}^{T} X(t)\,dt$. So, this is possible as long as the stochastic process, so, in general I define t belonging to T, that t is different from this T. So, here you have the time interval of length 2T within that 2T if you find out the time average, and that time average quantity is going to be the estimation for the mean; that means, if μ_T converges in the squared mean to μ as T tends to infinity then the process is going to be a mean ergodic. That stochastic process is going to be call it as a mean ergodic process.

Similarly, one can estimate other higher order moments also provided the process is ergodic with respect to those moments. So, here I have made the ergodic with respect to the mean therefore, you are estimating the mean with the ergodic property. Similarly, if this given stochastic process is satisfying the ergodic property with the higher order moment, then those measures also can be estimated in the same way. So, here the μ_T converges in mean, in squared mean to μ as a T tends to infinity. So, that is the conclusion we are getting from the ergodic property along with the wide sense stationary property.

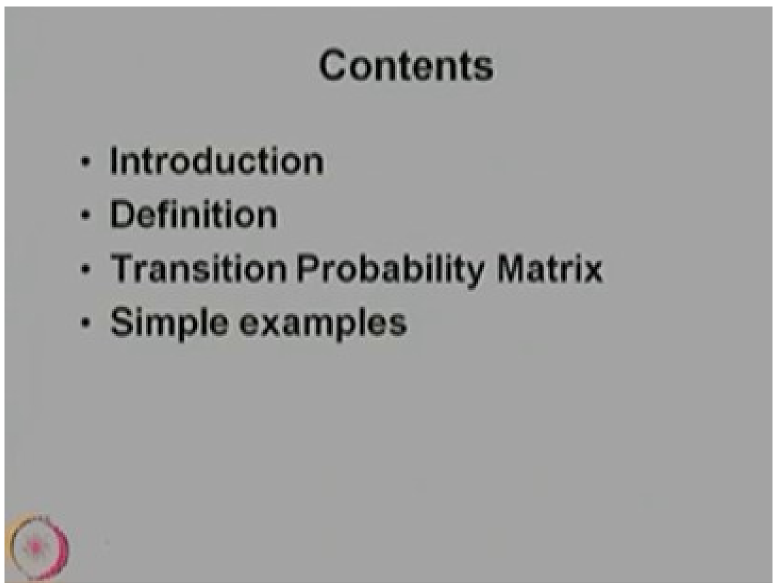
With this let me stop the todays lecture.

Thanks.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

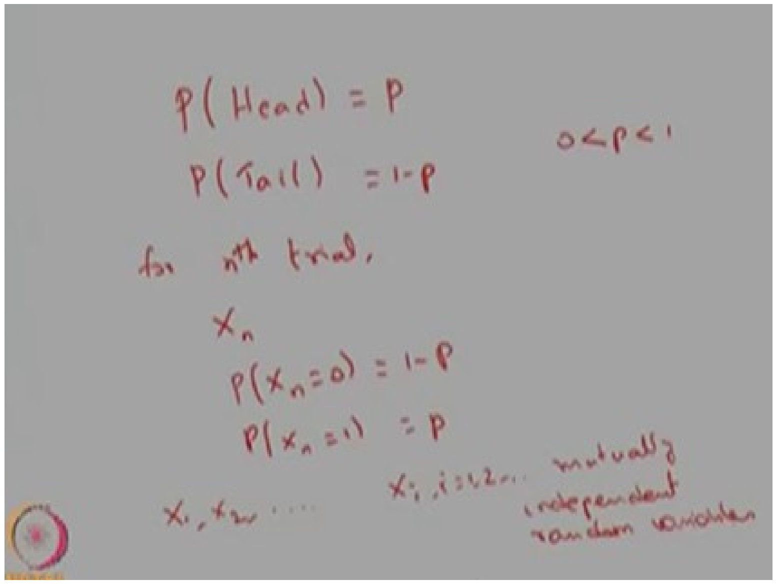
Module – 10
Discrete-time Markov Chains (DTMCs)
Lecture – 60

(Refer Slide Time: 00:19)



This stochastic process in this, we are going to discuss the Discrete time Markov Chain and this is a lecture 1, in this lecture, I am going to discuss the introduction about the discrete time Markov chain, then followed by the definition and the important one concept called one step transition probability matrix and few simple examples also. Consider a random experiment of tossing a coin infinitely many times. Each trial there are two possible outcomes namely head or tail.

(Refer Slide Time: 01:03)



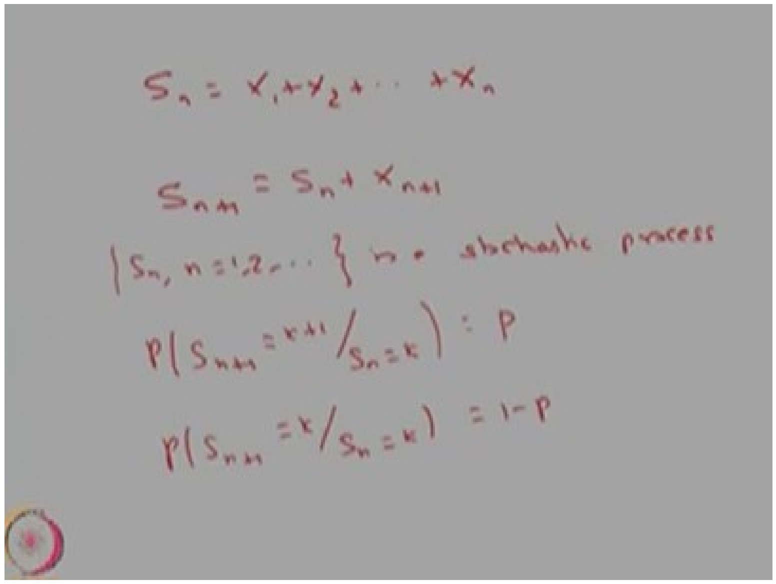
Assume that the probability of head, that probability you assume that is p and the probability of tail occurring in each trial that you assume it as 1 - p, you assume that the p lies between 0 to 1, denote for the nth trail, because you are tossing a coin infinitely many times for the nth trial, you denote the a random variable X_n is the random variable whose values are 0 or 1 with the probability.

The probability of X_n takes a value 0 that is same as in the nth trial you are getting the tail that probabilities 1 - p and the probability of X_n takes a value 1 that probability is make it as p for the head appears and already you assume that and the probability is lies between 0 to 1, thus you have a sequence of random variable $X_1, X_2, \ldots$and this will form a stochastic process.

And assume that all the X_i's are mutually independent random variable, so this is a random experiment in which we are tossing a coin infinitely many times and for any nth trial, you define the random variable X_n with the probability it takes a value 0 the probability 1 - p and takes a value 1 with the probability p and that is equivalent of appearing head at the probability p and occurring the tail with the probability 1 – p.

Now I am going to define another random variable that is a partial sum of first n random variables, n $X_i{}'s$.

(Refer Slide Time: 03:23)



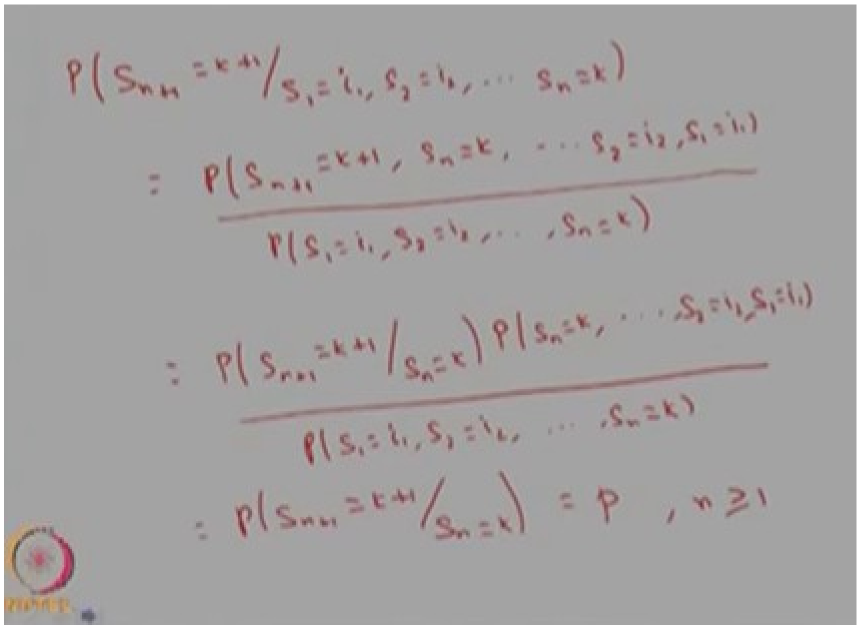
So the Sn will be sum of a first n random variables therefore the sum S_n gives the number of heads appears in the first n trials, it can be observed that $S_{n+1}=S_n+X_{n+1}$, since S_n is the partial sum of a first n trials outcome so the $S_{n+1}=S_n+X_{n+1}$. you can also observe that since $Sn=\sum_{i=1}^{n} Xi$ and $S_n+X_{n+1}$and also all the X_i's are mutually independent variables.

Sn is independent with X_{n+1}, that means here the S_{n+1}^{th} random variable is a the combination of a two independent random variables whereas the Sn is the till nth trial how many heads you appear plus whether it is a head or tail accordingly these values is going to be 0 or 1, therefore if you see the sample path of S_{n+1}, it will be incremented by 1 if X_{n+1} takes a value 1 or it would have been the same value earlier if this X_{n+1} takes a value 0.

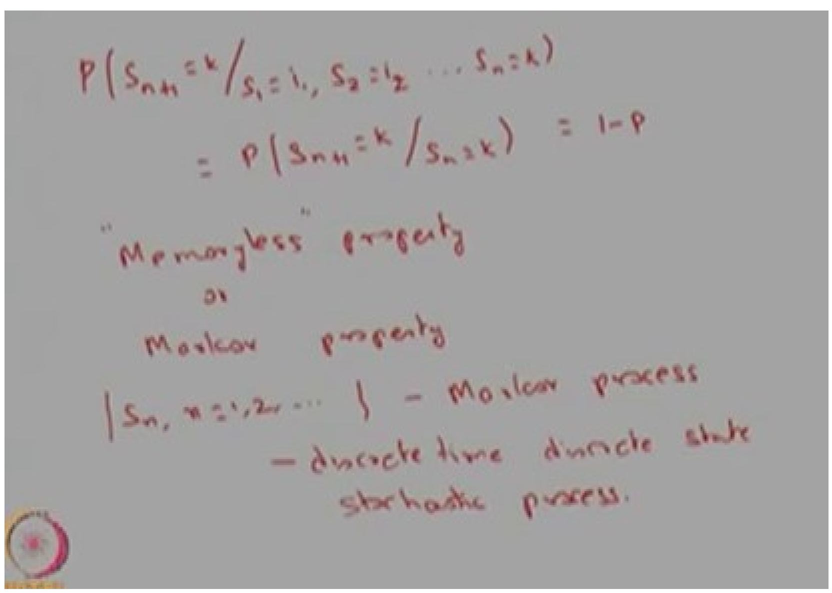
And also you can observe that S_{n+1} is a depends on S_n and only on it, it is not a depends on S_{n-1}or S_{n-2} so on, because it is accumulated the number of trials values over the n, therefore S_{n+1} is depends on S_n and only on it, the S_n for different values of n this will form a stochastic process and now you can come to the conclusion the probability of this is a stochastic process.

The probability of S_{n+1} suppose this values is k + 1 given that S_n was k that means the S_{n+1} value would have been 1, therefore the appearance of a the head appears in the (n + 1) th trial and probability is going to p. similarly you can make out suppose S_{n+1}value will be k such that S_n is also k then that is possible with a (n + 1) th trial you got the tail, therefore that probability is 1 – p, this is satisfied for all n. So you can make out this is satisfied for all n is greater than or equal to, even I can go for n is greater than or equal to 1.

(Refer Slide Time: 06:45)



(Refer Slide Time: 06:55)



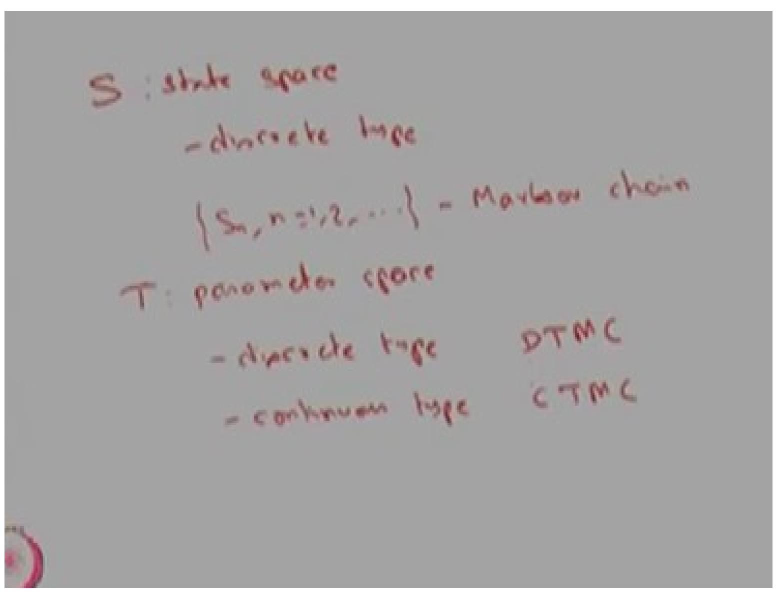
Not only these similarly, I can come to the conclusion the P($S_{n+1}=k/S_1=i_1, S_2=i_2, \dots, S_n=k$¿ that is also can be proved the P($S_{n+1}=k/S_n=k$¿that is same as what is the probability that value was the same k in the subsequent trials that is possible of appearing a tail in the (n + 1) th trial.

Therefore, that the appearance of the tail in the (n + 1) th trial the probability is 1 - p or I can use the notation q, that means the probability of (n + 1) th trial that distribution given that I know the value till the nth trial that is same as the distribution of (n + 1) th trial, given with the only the distribution nth distributions not the earlier distributions and this property is called memoryless property.

The stochastic process the S_n satisfies the memoryless property or the other word called Markov property, the distribution of (n + 1) given that the distribution of first random variable, second random variable, the nth random variable that is same as the conditional distribution of (n + 1) th random variable given that with the nth random variable only and this property is called a memoryless or Markov property.

The stochastic process the S_n satisfying the Markov property or memoryless property is called Markov process, the stochastic process satisfying the memoryless property or Markov property is called Markov process, in this example the stochastic process S_n is the discrete time, discrete state stochastic process. Now I can give based on the state space and the parameter space I can classify the Markov process or I can give the name of the Markov process in an easy way based on the state space as well as the parameter space.

(Refer Slide Time: 09:57)



So when the state space S, the S is the state space this is nothing but the collection of all possible values of the stochastic process, if this is of the discrete type that means the collection of a elements in the state space S is going to be a finite or countably infinite then we say the states spaces of the discrete type.

So whenever the stochastic process satisfying the Markov property then the stochastic process is called the Markov process or you can say whenever the state space is a discrete then you can say the corresponding stochastic process we can call it as a Markov chain, whenever the state spaces are discrete.

Now based on the parameter space T, parameter space is nothing but the possible values of T whether it is going to be a finite or countably infinite then it is going to be a discrete parameter space or discrete time or it is going to be uncountably many values then it is going to be called it as a continuous type, so whenever T is going to be a discrete type then the Markov chain is going to be called it as discrete time Markov chain.

Whenever the parameters which is going to be of the continuous type that means the possible values of T is going to be uncountably many then we say continuous time Markov chain, so in this example the S_n, the possible values of S_n is also going to the state space is going to be discrete type and the parameter space is also going to be discrete type therefore the given example the sun is going to be the discrete time Markov chain.

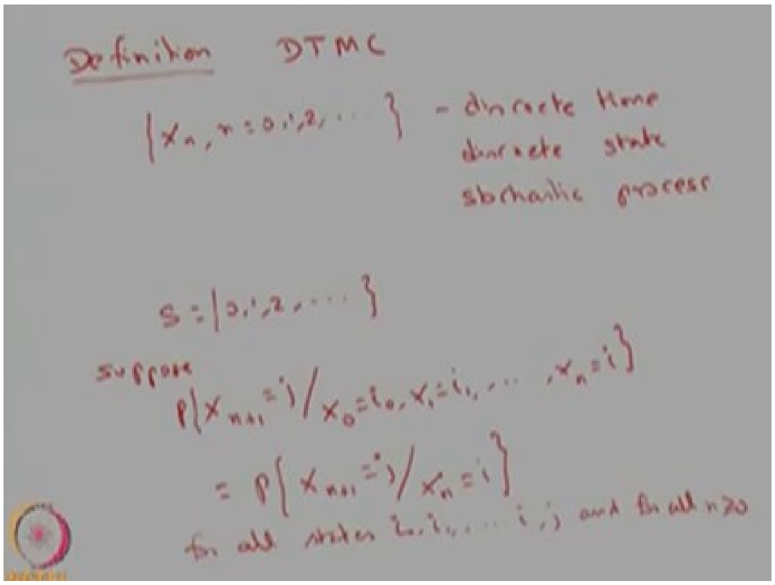
So in this module, we are going to study the discrete time Markov chain then next model, model five we are going to discuss the continuous time Markov chain, so in general whenever the stochastic process satisfying the Markov property it will be called Markov process. So based on the state space the Markov process called as Markov chain and the based on the parameter space it is called discrete time Markov chain or continuous time Markov chain accordingly, discrete type or continuous type.

Introduction to Probability Theory and Stochastic Processes

Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 61

(Refer Slide Time: 00:12)

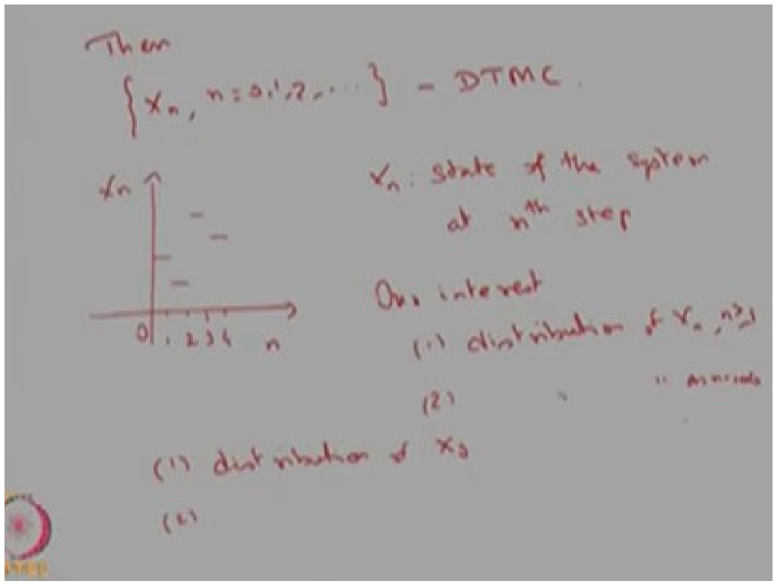


Discrete time Markov chain, I am going to give the formal definition of a discrete time Markov chain. Formal definition of discrete time Markov chain in notation. We in short we call it as a DTMC, consider a discrete time discrete state stochastic process, consider here discrete time this is a discrete time discrete state stochastic process, assume that X_n takes a finite or countable number of possible values unless otherwise mentioned this set of possible values will be denoted by the set of non-negative integers S ={ 0, 1, 2,…}.

Unless otherwise measured you can mention you can always assume that the states space S ={ 0, 1, 2,…}, even if you take a other values also you can always make a one to one correspondence and make the state space is going to be a S ={ 0, 1, 2,…}. Suppose the $P\{X_{n+1}=j/X_0=i_0, X_1=i_1, \ldots, X_n=i\}=P\{X_{n+1}=j/X_n=i\}$, for all states - for all states i_0 whatever be the value of $i_0, i_1, ..$ and j and also for all n ≥ 0, if this property is satisfied by for all states $i_0, i_1, .. i, j$ as well as for all n ≥ 0. Then this stochastic process that is a discrete time discrete state stochastic process is going to be known as a discrete time Markov chain.

So basically this is a Markov property and the Markov property is satisfied by all the states as well as all the random variables so if this Markov property is satisfied by any stochastic process then it is called a Markov process and since it is the time space is discrete and parameter space is discrete therefore it is called a discrete time Markov chain.

(Refer Slide Time: 03:38)



Then we see this stochastic process $\{X_n, n=0,1,2,...\}$ and so on is the discrete time Markov chain, we can just have a look of how the sample path look like for different value of n and the y axis is X_n, suppose at n is equal to 0, it started with some value at X and n equal to 1, it would have been a different value and n is equal to two it would have been the different value.

So these values are the either it could be a finite value or countably infinite number of values therefore the state space is going to be discrete and the parameter space is going to be discrete, so like that it is taking a different value over the n, so this is going to be the sample path or trace of the stochastic process X_n, suppose you assume that X_n is the state of the system at nth step or nth time point and this X_n satisfies the previous this Markov property.

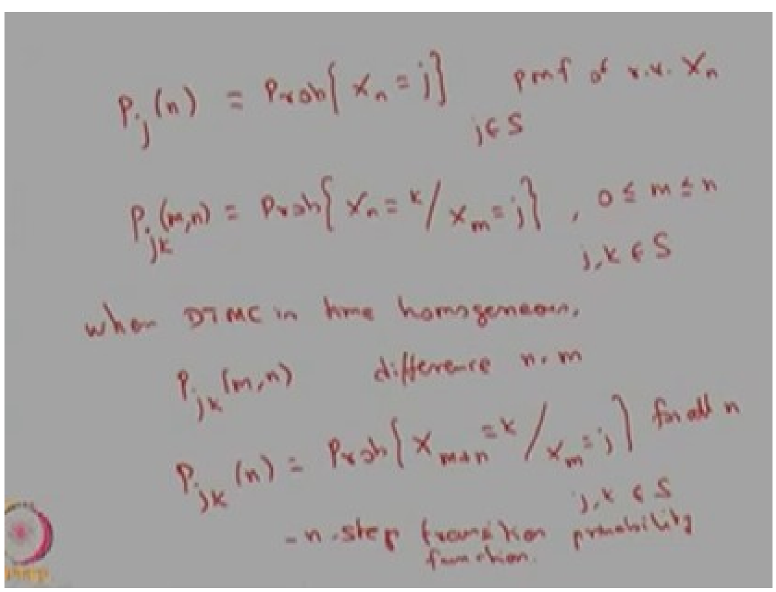
Then the stochastic process is going to be call it as a discrete time Markov chain and our interest will be suppose the stochastic process satisfies the Markov property our interest will be to know the two things one is what is the distribution of X_n for n ≥ 1, you know where the system starts so X_0 you - you know, your interest will be what could be the distribution of X_n that is nothing but what is the $P\{X_n = j\}$.

And also what could be the distribution of X_n as n tends to infinity, as n tends to infinity our interest will be finding out the distribution of X_n, so at any finite n as well as n tends to infinity that will be of our interest, to compute this you need two things, one is you need what is the distribution of X_0 that is a initial distribution vector, where the system starts at the zero th step, what is the distribution of X_0.

And also second things of your interest will be, what is the transition distribution or how the transition takes place, what is the distribution of a transition from any nth step to (n + 1)th step for all n, so if you know the two things the initial distribution vector as well as the distribution of the transition from nth step to (n + 1)th step, using these two quantity you can find out what is the distribution of X_n for any n as well as you can find out the distribution of X_n as n tends to infinity.

For that I am going to define few conditional probability distribution as well as the marginal distribution for the random variable X_n through that we are going to find out the distribution of X_n for any n as well as n tends to infinity.

(Refer Slide Time: 07:53)



So the first one, I am going to define the probability mass function as the $P_j(n)$ that is nothing but what is the $P\{X_n=j\}$, so this is the probability mass function of the random variable X_n, what is the $P\{X_n=j\}$ that I am going to denote it as the $P_j(n)$, where here the j is belonging to the state space S, this is the probability of the mass function of the random variable X_n.

Similarly, I am going to define the conditional probability mass function as a $P_{jk}(n)$, that is nothing but what is the probability that X_n takes a value k given that X sorry, I need two suffixes, two index for here with the two variables (m, n) what is the $P\{X_n=k/X_m=j\}$,, obviously the m is lies between 0 to n, whatever m and every n and the j, k is belonging to S.

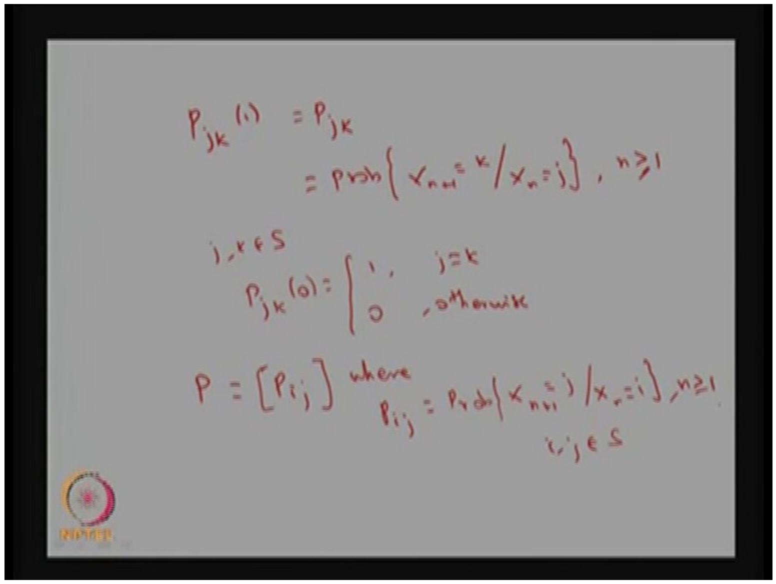
so this is a conditional probability distribution of the random variable X_n with X_m and the m th step the system was in the state j, and the n th step the system is in the state k and this is the conditional probability with the two arguments (m, n) so this is the probability that the system makes a transition from the state j at step m to the state k at step n, this is called transition probability function of the discrete time Markov chain.

When the DTMC is a time homogeneous, this is very important when DTMC is a time homogeneous that means it satisfies the time invariant property, that means the $P_{jk}(m,n)$ depends only on the time difference n - m, whenever the DTMC is a time homogeneous that means in the time invariant so the actual time is not a matter only the time difference is the importance therefore this is going to depends only on the time difference n - m.

In this case I don't want to the two arguments m, n I can go for writing $P_{jk}(n)$, that is nothing but what is the $P\{X_{m+n}=k/X_m=j\}$,for all n and here j, k belonging to S, so the m does not matter only the interval or the interval length of step n is matter, so that means the system was in the state j and it is a transition into the state k in n steps.

Because the DTMC is a time homogeneous so the X_m to X_{m+n}, it is valid for all m, for all n we are finding out for the n step transition, this is called n step because the DTMC is a time homogeneous and this is called n step transition probability function, this is a n step transition probability function, using this we can define the one step transition probability.

(Refer Slide Time: 12:46)

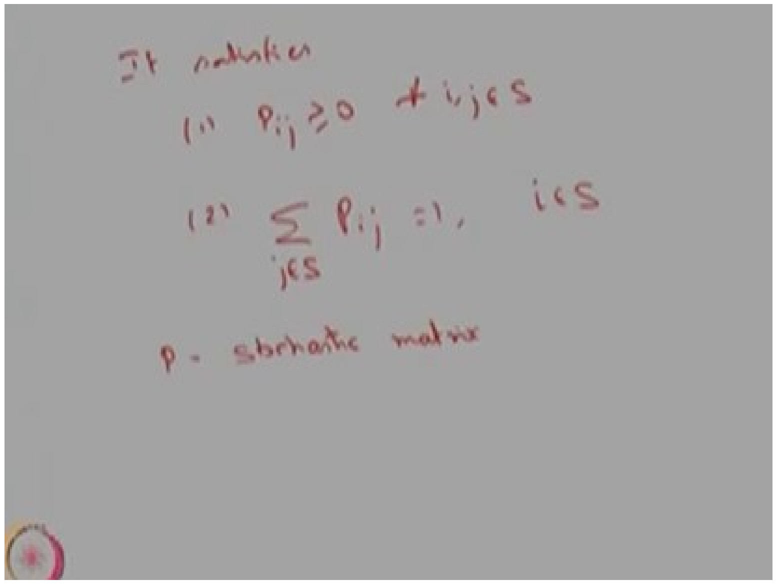


That is denoted by $P_{jk}(1)$or we can avoid the bracket 1 also so you can write it as the P_{jk} that is nothing but what is the probability that the $P\{X_{n+1}=k/X_n=j\}$,, for all n ≥ 1, obviously for j, k belonging to S, if you find out the zero step transition probability that values is going to be 1 for j equal to k, otherwise it is going to be 0.

This one step transition probability I can make it in the matrix form as the P is the matrix and that consists of P_{jk}, where the $P_{jk}$is nothing but one step transition probability elements of $X_{n+1}$is equal to j given that X_n is equal to i, here i, j belonging to the states space S, you should remember that states space S is consist of finite elements or countably infinite number of elements.

Accordingly, this matrix is going to be either when S is going to be finite elements then the P matrix is going to be a square matrix, since the $P_{jk}$is the one step transition probability of a system moving from the state i to j in one step and since it is a time homogeneous this is valid for all n this is valid for all n ≥ 1.

(Refer Slide Time: 14:56)

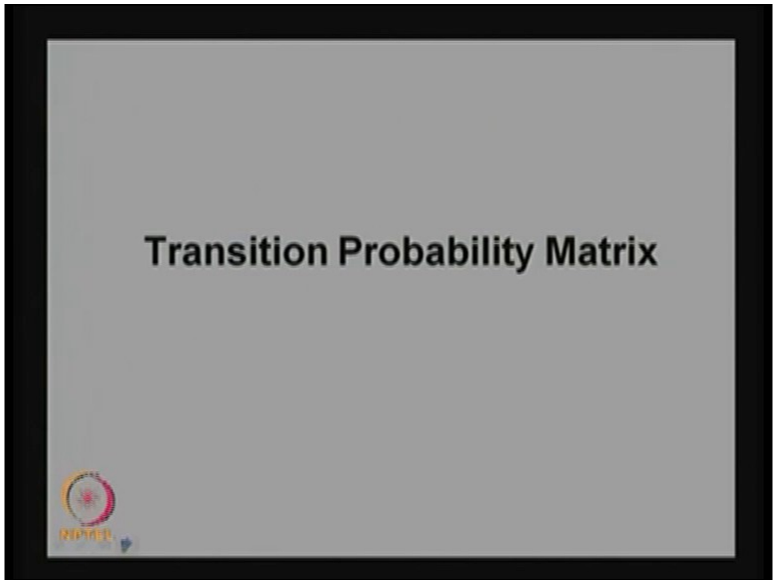


And this satisfies the one step transition probability matrix satisfies two properties the each entity will be greater than or equal to zero for all i, j belonging to S, because these are all only the conditional probability of system moving from state i to j in one step, therefore either it will be a zero or greater than zero for all possible values of i, j.

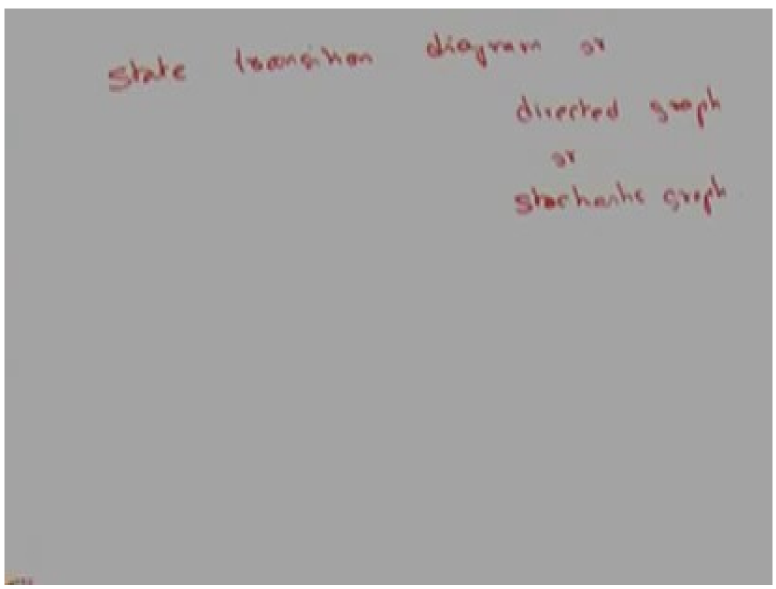
The second condition if you make the $\sum_{i \epsilon S} P_{ij} = 1$, i belonging to S, that means the row sum is going to be 1, because it is a conditional probability of system moving from one state to another states, if you add all the other possible probabilities then that is going to be 1, since these one step transition probability matrix satisfies these two properties and this matrix is P is known as the stochastic matrix.

Because of satisfying these two conditions the matrix one step transition matrix is also called stochastic matrix. Now I am going to explain what is the pictorial way of viewing the one step transition probability matrix or the stochastic matrix.

(Refer Slide Time: 16:48)



(Refer Slide Time: 16:53)



That is provided by state transition diagram or the other words it is called a directed graph the DTMC can be viewed as a directed graph such that the state space S is a set of vertices or nodes and the transition probabilities that is a one-step transition probabilities are the weights of the directed arcs between these vertices or nodes, since the weights are positive and the sum of the arc weights from the each node is unity this directed graph is also called stochastic graph.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 62

(Refer Slide Time: 00:01)

Example 1

A factory has two machines and one repair crew. Assume that probability of anyone machine breaking down on a given day is α. Assume that if the repair crew is working on a machine, the probability that will complete the repairs in a day is β. For simplicity, ignore the probability of a repair completion or a breakdown taking place except at the end of a day. Let X_n be the number of machines in operation at the end of the nth day, Assume the behavior of X_n can be modeled as a Markov chain.

I am going to explain the Discrete time Markov Chain with the three simple examples the first example is as follows.

A factory has a two machines and one repair crew, assume that probability of anyone machine breaking down a given day is α, so the α is the probability, assume that if the repair crew is working on a machine the probability that they will complete the repairs in two more day is β, for simplicity ignore the probability of a repair completion or a breakdown taking place except at the end of a day.

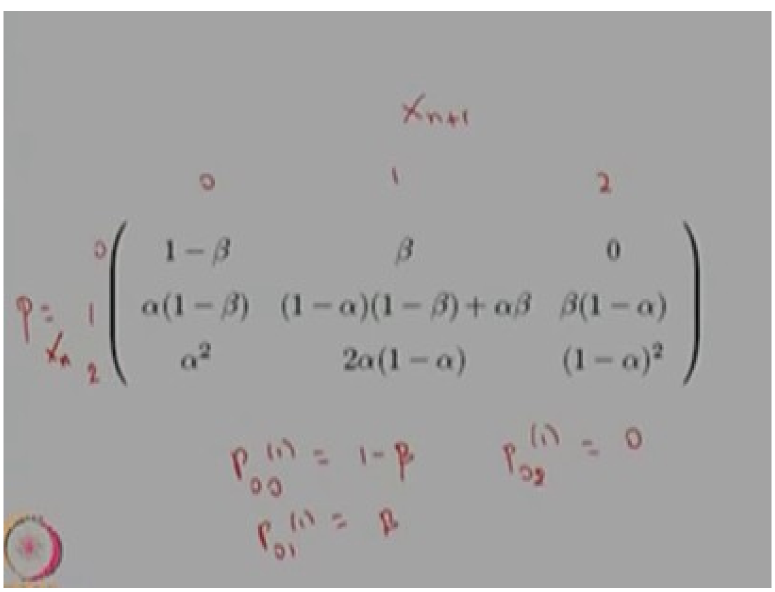
That means we observe the system at the end of the day how many working machines in the system, let X_n be the number of machines in operation at the end of the nth day, assume that the behaviour of $X_n$can be modeled as a Markov chain, so based on the information available here the machine can be breakdown and we have only one repair person and the probability of he can do the repair in a day that probability is β.

And 1 - β is the probability that he cannot be able to complete the repair of a machine in a day and the random variable X_n is it denotes how many machines are in the operation at the end of the day therefore the possible values of X_n, since we have a two machines the possible values of X_n will be 0 1 or 2, so this will form a state space S, so the S consists of the element 0 1 and 2 and the X_n over the n it is going to form a discrete time Markov chain.

Because it is a discrete time discrete state stochastic process and also the based on the crew the number of machines are working in any day depends on how many machines are working on the previous day and how many things are under repair and so on, so the dynamics of the number of machines in operation depends only on the number of machines working in the previous day not all the previous earlier days.

Therefore, the memoryless property is satisfied by the stochastic process X_n, therefore this is called a discrete time Markov chain, our interest is to find what is the one step transition probability matrix with the assumption that the X_n is the time homogeneous also, since it is a time homogeneous DTMC therefore we are trying to find out what is the one step transition probability matrix for the given time homogeneous DTMC.

(Refer Slide Time: 03:30)



So this is the one step transition probability matrix P and the possible states are 0 1 and 2 and suppose the system was in the state 0, 1 or 2 in the n th step where the system will be in the (n +

1) th step, therefore this is the possible values of $X_{n+1}$and this is a possible values of X_n and this one step transition probability matrix will give suppose the system was in the state and the n th step what is the probability that it will be in these states in the (n + 1) th step.

So the first index will give what is the probability that 0, 0 in one step that means the n th step number of working machines are 0 and what is the probability that in the (n + 1) th step also 0 machines are able working condition that means all are under repair all two machines are under repair and the probability of a crew is going to be not repair that is going to be 1 - β therefore the probability is 1 - β.

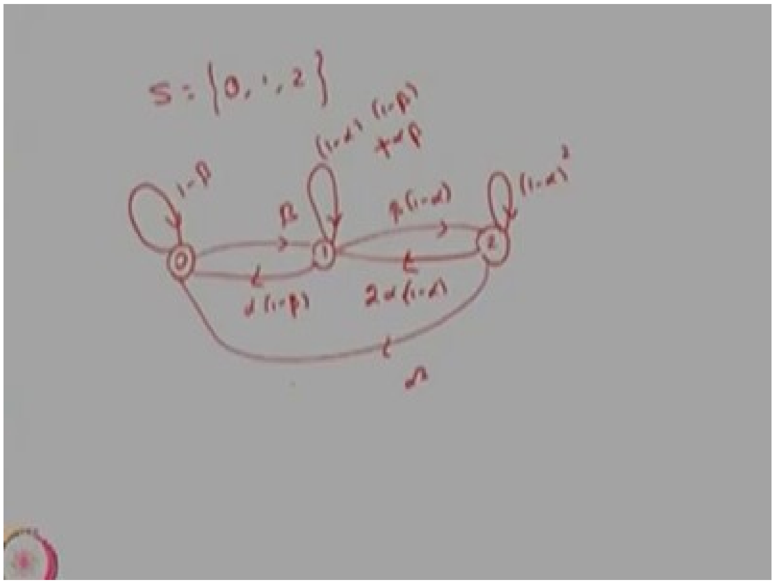
In one step what is the probability that the number of a working machines going from 0 to 1 that is because of the crew is finishing the repair in a day and that probabilities β and since he can do the only 1 repair in a day, therefore the possibility of repairing more than 1 machine in a day it is not possible it's a rare event and the probability is going to be 0, therefore P{ 0 ,2} is going to be 0.

Similarly, now we can visualize what is the probability that number of working machines is 1 in the n th step and what is the probability that in the 0 machines will be working in the (n + 1) th step, that is possible with the two independent things the 1 machine can be the failed and the other machine cannot be finishing the repair therefore the crew is not finishing the repair that probability is 1 - β multiplied by one machine is going to be repaired.

Therefore, the total number of machines working will be 0 in the (n + 1) th step, that is α(1 - β) and similarly you can evaluate the other element also and for example the system is going from the state 2 to 0 that is nothing but at n th step 2 machines are in the working condition and (n + 1) th step, 0 machines are the working condition that means both the machines got failed in the same day.

Therefore, that probability is α*α that is the probability both the machines got failed the same day therefore in the next day the number of working machine is going to be from 2 to 0, like that you can visualize the other elements also, the same one step transition probability matrix can be visualized with the state transition diagram.

(Refer Slide Time: 07:27)



And state transition diagram you have to make with this state space has a vertices or the nodes and the weights of the directed arcs are nothing but the one step transition probability of system is moving from one state to other states, those are going to be the weights, if the probabilities are 0's, then no need to draw the directed arc from that particular node to the destination node.

So first you start with a nodes as the possible values of the state space so you list out all the state space as a node now by seeing the one step transition probability matrix you should make the arc from 0 to 0 self-loop is allowed, if the probability is going to be greater than 0, so you should draw the self-loop from 0 to 0 with the arc value 1 - β and you should draw the arc from 0 to 1 with the arc weight β and you should not draw any arc from 0 to 2.

Because that probability is going to be 0, therefore 0 to 0 that probability is 1 - β and 0 to 1 it is going to be β and there is no arc from 0 to 2, because that probability is 0 and similarly, now you can go for filling the second row so 1 to 0 is α(1 - β), 1 to 1 is (1 - α)(1 - β)+ αβ and 1 to 2, so you have all three probabilities are greater than 0.

Therefore 1 to 0 that arc is $\alpha(1 - \beta)$, and 1 to 1 is $(1 - \alpha)(1 - \beta)+ \alpha\beta$ 1 to 2 is $\beta(1 - \alpha)$, similarly you can draw the arc for 2 to 0 that is $\alpha*\alpha$ to 1 and so on, therefore 2 to 0 that is $\alpha^2$and 2 to 1 is $2\alpha(1-\alpha)$ and 2 to 2, that is $(1-\alpha)^2$. So this state transition diagram is a pictorial view of one step transition probability matrix.

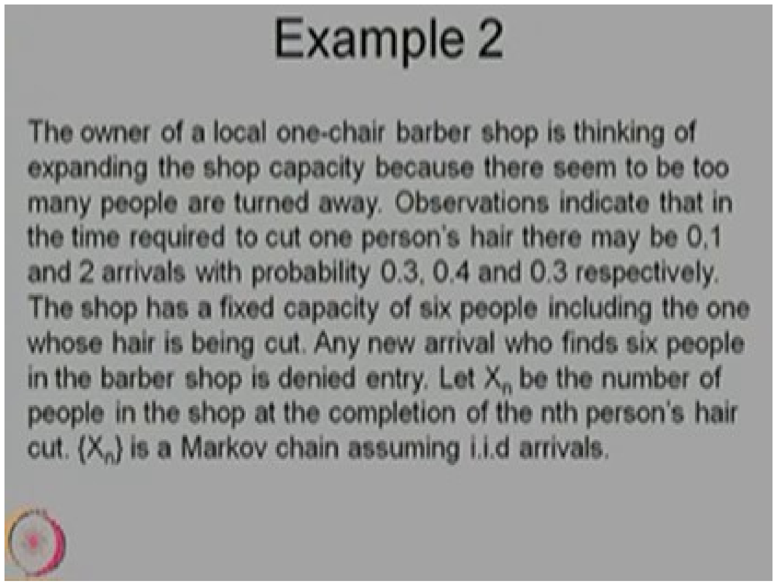
This is nothing to do with the initial probability distribution it gives only information about whenever the DTMC is a time homogeneous, suppose the system start from one particular state what is the probability that the system will move into another states with probability and it won't give more than that information.

But this much information is useful when you are going to study the properties of the discrete time Markov chain as well as when you are want to find out the limiting distribution that is the distribution of the X_n has n tends to infinity the diagram will be very useful to conclude whether the limiting distribution exists or not, if it exists whether it is going to be unique or not and so on, so those things can be visualized easily by seeing the state transition diagram.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 63

(Refer Slide Time: 00:01)



Now I am moving into the second example.
In this example I have taken the barbershop example which I have discussed in the module 1 and also the same example so the owner of a local one share barber shop is thinking of expanding the shop capacity because there seems to be too many people are turned away observation indicate that in the time required to cut one person's hair there may be 0, 1 and 2 arrivals with the probability 0.3, 0.4 and 0.3 respectively.

So this information is very important that means, during one person's haircut what is the probability that no people turned up with the probability 0.3 and one people may turned up with the probability 0.4 and there is a possibility two arrivals is possible during the one person's haircut with the probability 0.3 therefore the summation of probability is going to be 1.

So during the one person's haircut these are all the only 3 possibilities are possible with the zero arrival or one arrival or two arrival, the shop has a fixed capacity of 6 people including the one whose hair is being cut that means a maximum 6 people can be allowed in the system, so 5 people can wait maximum and 1 people under the service, any new arrival who find 6 people in the barbershop is denied entry that is the meaning of a capacity of the system is finite with the size 6.

Now I am going to define the random variable, let X_n be the number of people in the shop at the completion of the nth person's haircut this is very different random variable or this is very different steps for stochastic process usually the parameter space is a time but here the parameter space is the number of people in the shop the n is the parameter space is the person who leaves after the haircut.

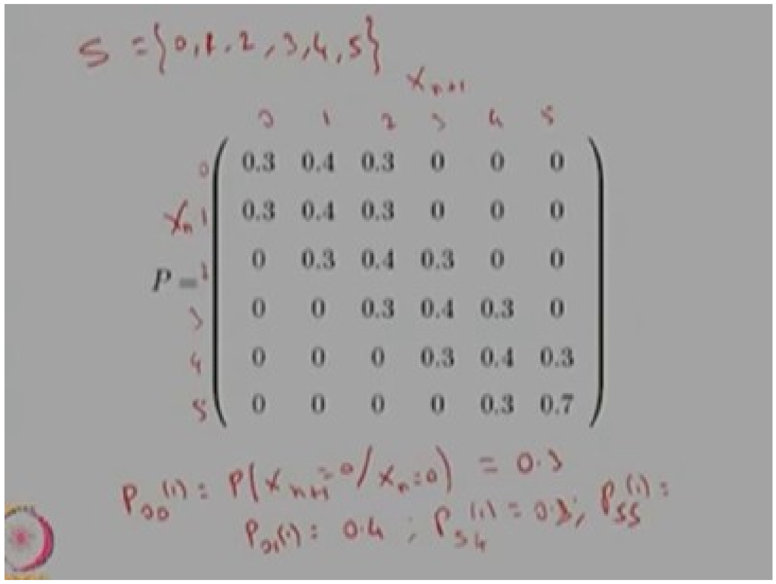
So it's a nth person who leaves the system that becomes the parameter space where as the random variable is how many people in the system when the nth person leaves the system that means you should not count that person when you are finding the values of X that means this number is counted at the departure time point so when the nth person leaves how many people in the system.

The system is a maximum 6 people allowed therefore he cannot see more than 5 people in the system when he leaves, so because of this constraint because of during the 1 persons arrival either 0 or 1 or 2 arrivals can takes place and so on based on this information the stochastic process X_n is going to be a discrete time discrete state stochastic process as well as the Markov property satisfied.

That means the probability of X_{n+1} take some value given that all the previous values are known that is same as the conditional probability distribution of X_{n+1} take some value given that X_n was some value, so also the future distribution given that present as well as the past information is same as the future distribution given the present, not the whole past information so this Markov properties will be satisfied by this stochastic process.

Therefore, this X_n will form a discrete time Markov chain, obviously it is the time homogeneous discrete time Markov chain also so in this example our interest is to find out what is the one step transition probability matrix.

(Refer Slide Time: 04:26)



This is going to be the one step transition probability matrix and the possible states S is going to be 0, 1, 2, 3, 4 or 5, because the capacity of the system is 6 and whenever the nth persons leaves the first person second person third person leaves how many people are in the system, therefore the maximum will be 5 and there is a possibility when he leaves and no one in the system also.

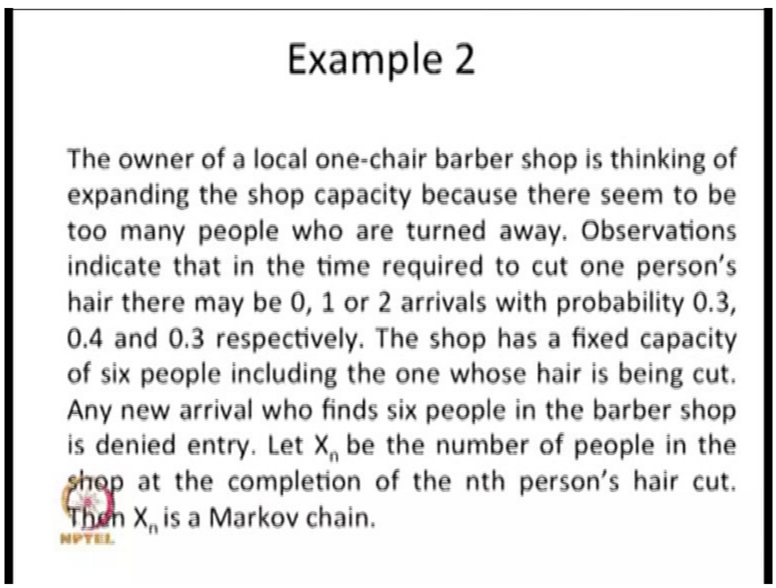
And this is the one step process transition probability matrix and this is also going to be a square matrix because it is going to be a countably finite number of elements and this is 0, 1, 2, 3, 4, 5, now we can discuss what is the probability that 0, 0 in one step that is nothing but when the n th person leaves no one in the system, when the (n + 1) th person leaves no one in the system.

What is the probability for that? That is the P{ X_{n+1} = 0/ X_n = 0}, so one step transition probability its independent of n, because it's a time homogeneous, it's a one step transition probability matrix, so this is possible at some person leaves whatever be the n nobody in the system when the next person leaves nobody in the system.

So that is possible by when some person leaves the system was empty for some time you don't know how much time it was empty then the (n + 1) th person enter in to the system and during his haircut no one turned up or no arrival takes place during his or (n +1)th haircut is going on, therefore when he leaves no one in the system, so we are not bothering when he entered into the system and so on.

Our interest is how many numbers of people in the system when the (n + 1)th person leaves and this probability is (n + 1)th person leaves is 0 people in the system and given that when the nth person leaves also 0 person in the system, so that that is possible with a explanation I have given no one enter into the system during the (n + 1)th person's haircut.

(Refer Slide Time: 07:26)



And the information is provided indicate that time required to haircut one-person haircut there may be a 0, 1 or 2 arrivals with the probability 0.3, so no arrival takes place during the one person's haircut is a 0.3, therefore this probability is possible with the probability 0.3, whereas P 0 one of one step, the same way you can write P$\{X_{n+1} = 1/ X_n = 0\}$, that is possible then the nth person leaves no one in the system.

When the (n + 1) th person leaves one person in the system that means during his haircut one person enter into the system that is possible with the probability 0.4, similarly from 0 to 2 in one step that is going to be 0.3 with the probability two arrival takes place during the (n + 1) th persons haircut. Now the second row, second row what is the probability that when the nth person leaves one person in the system when (n + 1) th percent leaves 0 person in the system.

That is possible, during the (n + 1) th the person leaves - haircut there is no one in the system no - arrival takes place, therefore the probability is 0.3 and from 1 to 1, that is possible with one person arrived during the (n + 1) th person haircut, therefore that probability is 0.4, and going from the state 1 to 2 that is possible two persons arrived during the (n + 1)th person haircut.

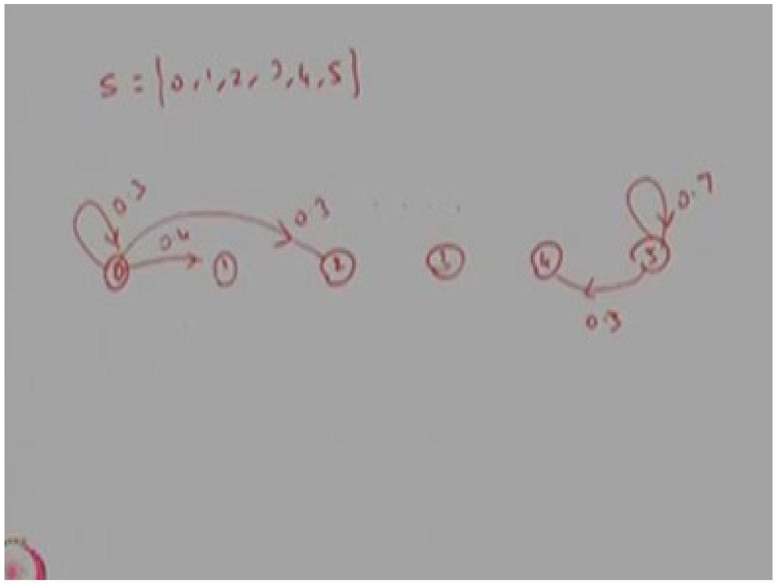
Whereas from 2 to 0 that is not possible because when nth person leaves two person in the system therefore (n + 1)th person in the leaves definitely he will see one person in the system because of no arrival and one arrival two arrival, therefore it will be shifted by one column and it will be keep continuing till the end.

Whereas the last one what is the probability that the 5 people in the system when the n th person leave and 4 people in the system when the (n + 1) th person leave that is same as no arrival takes place during the (n + 1)th haircut going on so therefore this is going to be 0.3.

Whereas P 5 to 5 in one step that is possible if the combination of one person arrived the system or two person arrived the system this system size is going to be maximum 6, therefore when (n + 1)th person haircut is going on, if one person arrives then he will be entered, if two person arrives then he cannot be accommodated therefore he will he won't join the system therefore the system the number of customers in the system in the Xn that is going to take the value 5.

And the combination of a 0.4 as well as 0.3, therefore this probability of system is moving from 5 to 5 is a 0.7, because of 0.4 + 0.3. Now I can give the state transition diagram for this example.

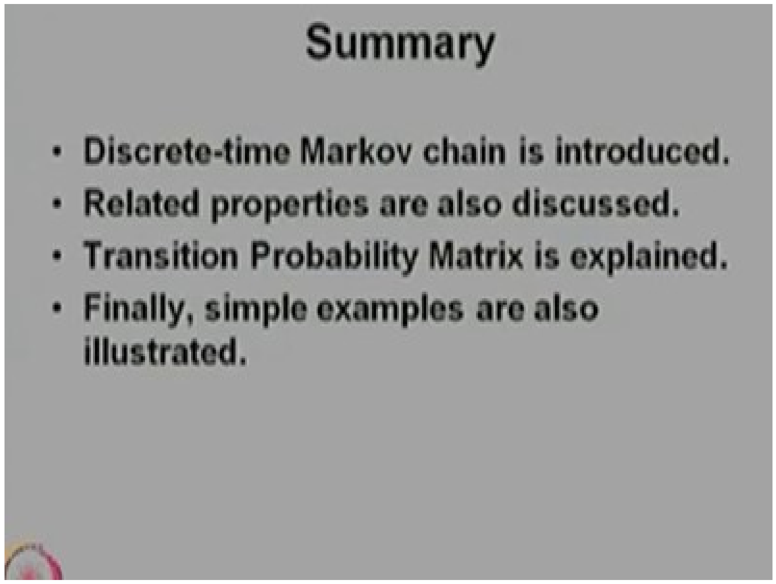
(Refer Slide Time: 11:25)



Because S is going to be 0, 1, 2, 3, 4, 5, therefore the nodes are going to be 0 1 2 3 4 and 5 and possible values from the one step transition probability matrix I can make out so 0 to 0 that probability is 0.3 and 0 to 1 is 0.4 and 0 to 2 is 0.3, similarly, I can fill up the all other things and 5 to 5 that is very important and 5 to 4 that is possible with the probability 0.3 and 5 to 5 is possible with the probability 0.7.

So this is the state transition diagram I did not complete the state transition diagram you have to fill up all the arcs with the weights going from one arc to other arc with the positive probability wherever there is a probability zero we should not draw the arc for it. So in this lecture, we have discussed the discrete time Markov chain.

(Refer Slide Time: 12:42)

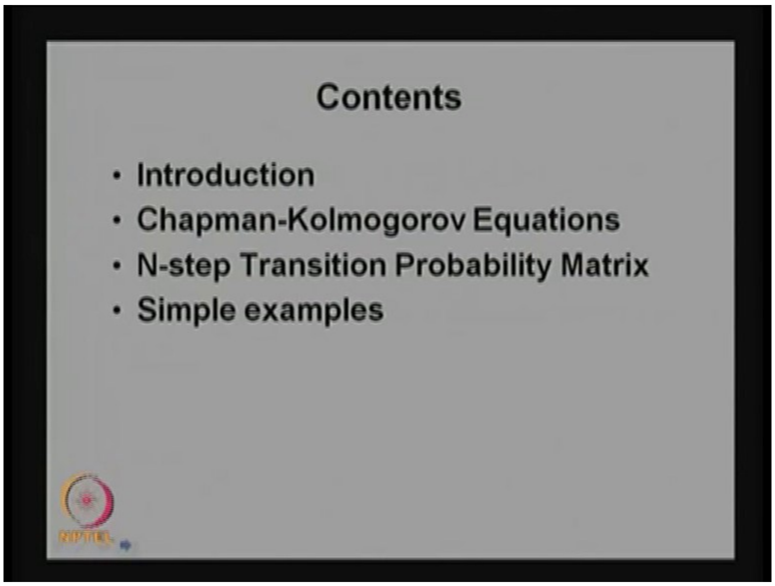


Then we have given the few important properties followed by we have explained the one step transition probability matrix and also we have given two simple examples with this, the lecture one is over for the module 4. Thanks.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

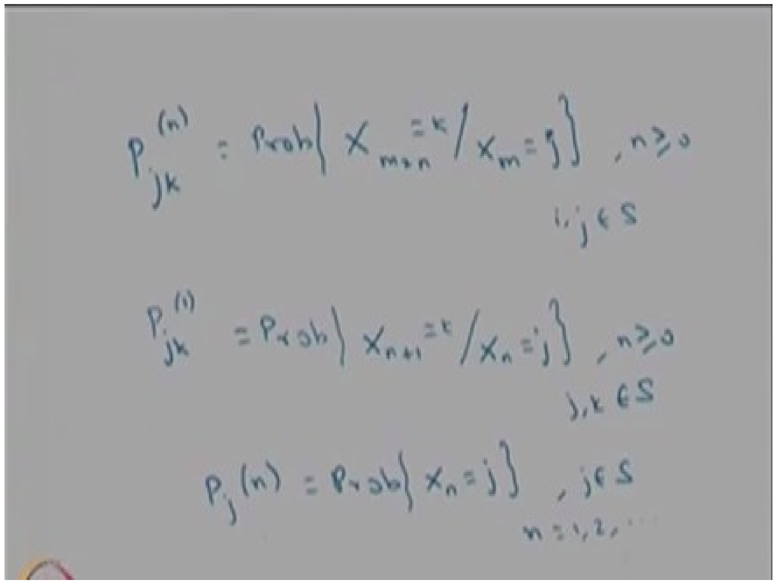
Lecture – 64

(Refer Slide Time: 00:01)



This is Discrete-time Markov Chain. And this is the Lecture 2. And in this lecture we are going to discuss about the Chapman-Kolmogorov Equations then we are going to discuss N-step Transition Probability Matrix and we are going to discuss a few more examples in this lecture.

(Refer Slide time: 00:32)



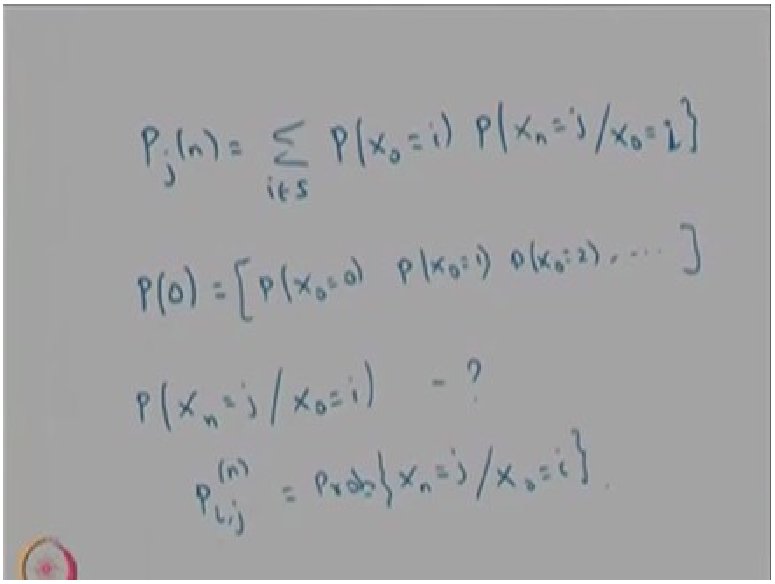
In the last class, we have discussed that transition probability of j to k in n steps *has* $P\{X_{m+n} = k/ X_m = j\}$ for m is greater than equal to 0 and j belonging to S. Since the underlying DTMC is a time homogenous this is the N-step Transition Probability of system is moving from the state j to k in n steps. So this we denoted as the conditional probability of P_{jk} in n-step Transition Probability where i,j is belonging to S where S is the state space and n can take the value greater than or equal to 0.

Also we have discussed in the last class what is the one step transition probability of P, we can write it within the bracket one or we can remove the bracket one in the Stochastic also that I nothing but what is the probability that the system will be in the state k in (n+1)th step given that it was in the state j in the nth step. Here also j, k belonging to S. So this is the one step transition probability.

So our interest is to find out what is the distribution of X_n. Another basic sequence of random variable X_n is time homogenous a DTMC our interest is to find out the distribution of X_n. So it has the probability mass function the $P_j(n)$that is nothing but what is the probability that the system will be in the state j at the nth step. So if the j is belonging to S and the n can be 1 or 2 and so on because you know the distribution of X_n is equal to 0 that means you know the initial probability vector of X_0.

So our interest is to find out what is the distribution of X_n for n = 1,2,3 and so on. So how we are going to find out this distribution?

(Refer Slide Time: 03:13)



So this distribution can be written using the $P_j(n)$is nothing but the submission over i belonging to S such that the system was in the state i at 0th step and multiplied by what is the probability that it will be in the state j given that it was in the state i at 0th step. So this is nothing but what is the probability that the system will be in the state j the nth step that is same as what are all the possible ways the system would have been moved from the state i from the 0th step to the state j in the nth step.

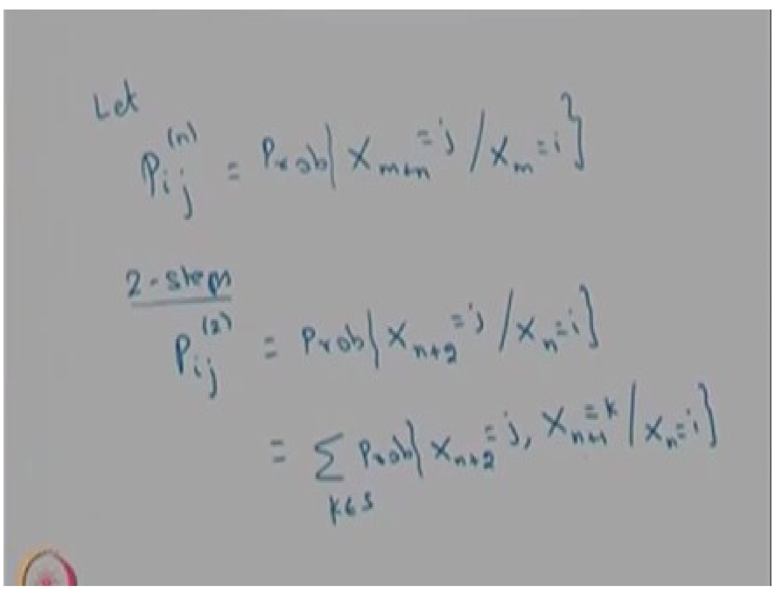
So this is the product of one marginal distribution and one conditional distribution for all possible values of i that gives the distribution of a X_n in the nth step. So for that you need to compute this distribution of X_n you need n-Step Transition Probability as well as the initial distribution vector or initial probability vector or the distribution of X_0. So the distribution of X_0 can be as a vector P(0), this vector P(0) it consists of the element what is the probability that X_0 takes the value 0.

What is the probability that X_0 takes the value 1? What is the probability that X_0 takes value 2 and so on? So this is the initial probability vector. Why we have taken the state 0, 1, 2 and so on unless otherwise as mentioned the set of state space that is going to be the possible values of 0, 1, 2 and so on unless otherwise it is assume it you can take always these values. So this is the initial probability vector or initial distribution vector.

So what we need, what is the n-Step Transition Probability of the system will be in the state j given that it was in the state i at the 0th step. This is what we want to find out. What is the additional probability mass function of n-Step Transition Probability vector? So that we can write it in the form of $P_{jk}^{(n)}$that is nothing but the probability of the system will be in the state j given that the system was in the state i at the 0th step.

That is that we need to compute the n-Step Transition Probabilities that is that $P_{jk}^{(n)}$. So this can be computed by using the method called Chapman-Kolmogorov Equations. So this Chapman-Kolmogorov Equations provide a method for computing this n-Step Transition probabilities. So how we are going to derive this Chapman-Kolmogorov Equation that I am going to do it now. So we are going to derive the Chapman-Kolmogorov Equations for the time homogenous Discrete time Markov Chain.

(Refer Slide Time: 06:47)



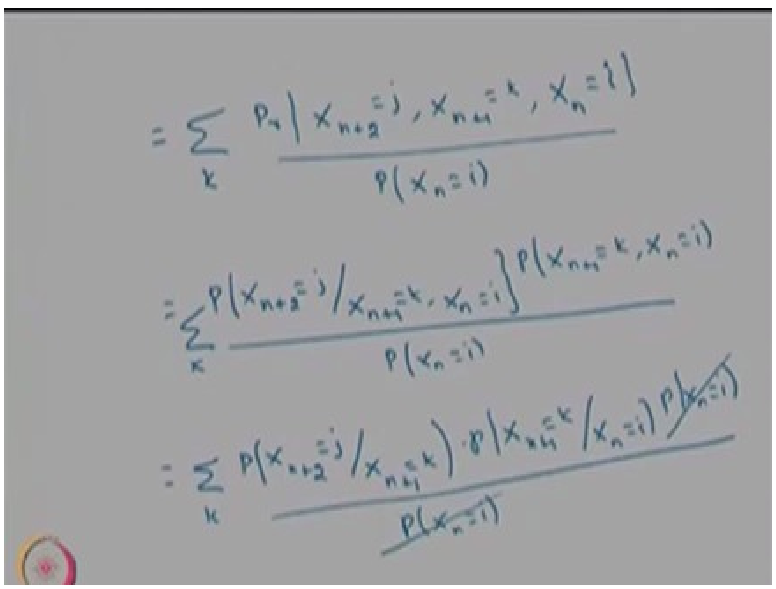
So let the $P_{jk}^{(n)}$that is nothing but what is the $P\{X_{m+n}=j/X_m=i\}$. Since the Discrete-Markov Chain is the time homogenous so this is the transition probability of system moving from the state i to j from the mth step to (m + n)th step.

Therefore, this transition is the nth transition probability matrix for the time homogenous Discrete-Markov Chain. Let us start with the 2-Steps. The 2-Step is nothing but what is the probability that system is moving from the state i to j in two steps. So X_{n+2} takes the value j given that X_n was i, this is for all n it is true because discrete DTMC is time homogeneous.

So this probability you can write it as-- this 2-Step transition probability of system moving from i to j the state i to the state j in 2-Steps that that you can write it as what are all the possible ways the system is moving from the state i to j by including one more state in the first step the state is k given that the system was in the state i in the nth step for all possible values of k belonging to S.

I can write this conditional 2-Step, conditional probability mass function from the nth step to (n+2)th step that is same as I can include one more possible state of k in the (n+ 1)th step.

(Refer Slide Time: 09:09)



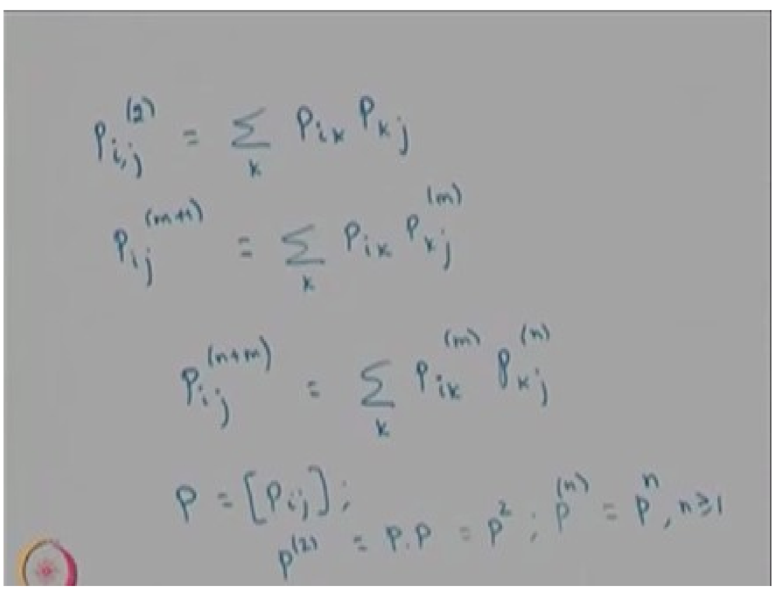
Now I can expand this as that is same as $\sum_k \frac{P\{X_{n+2}=j, X_{n+1}=k, X_n=i\}}{P\{X_n=i\}}$. The numerator join distribution of these three random variables that I can write it as in the form of conditional.

What is the conditional probability that $\sum_k \frac{P\{X_{n+2}=j/X_{n+1}=k, X_n=i\} P\{X_{n+1}=k/X_n=i\}}{P\{X_n=i\}}$. So basically, I am writing the numerator joint distribution of these three random variables as the product of conditional distribution with the marginal distribution of those two random variables.

Since the X_i's are time homogenous Markov Chain this conditional distribution by using the Markov property is same as the conditional distribution ofX_{n+2} takes the values j given that only the latest value is important the latest value is needed not the previous history therefore-- because of the memory less property X_n takes the value i is removed therefore this conditional-- distribution is the conditional distribution only $X_{n+1}$with X_{n+2}.

And similarly, I can apply the join distribution of this two random variable X_{n+1} and X_n I can again write it as the probability of X_{n+1}takes the value k given that X_n takes the value i and the probability of X_n takes the value i whole divided by probability of X_n and takes the value i. So this and this get canceled; so this is nothing but the conditional probability; this is nothing but the one step transition probability of system moving from k to j and the second term is a one-step transition probability of system is moving from i to k.

(Refer Slide Time: 12:11)



Therefore, the left hand side we have what is the 2-Step transition probability of i to j is same as all possible values of k. What is the one step transition probability of system is moving from i to k and one step transition probability of k to j? So this product will give the 2-Step transition probability of system is moving from the state i to j that is same as what is the possible values of k the system is moving from state i to k and k to j. So this is for the 2-Step.

Similarly, by using the induction method one can prove i to j of (m+1) steps that is same as what is the possible values of k the system is moving from one step from i to k and m steps from k to j. This is the 2-Step. So this is one step from i to k and one step from k to j by induction I can prove the (m+1) step will be i to k and the k to j in m step. Similarly, I can make it the other way round also. It is i to k in m steps and k to j in one step also.

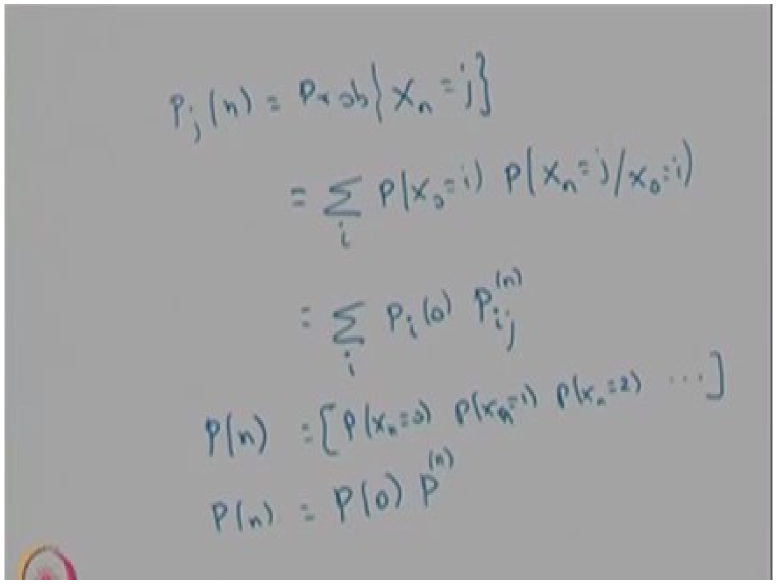
That combination also land up a the (m+1) step the system is moving from i to j. In general, we can make the conclusion the system is moving from i to j in (n+m) steps that is same as the possible values of k of probability of system is moving from i to k in the m steps and by n step the system is moving from k to j that will give for all possible values of k that will give the possibilities of system is moving from i to j in (n+ m) steps.

So this equation is known as Chapman-Kolmogorov Equation for the time homogenous Discrete-time Markov Chain. So whenever you have a Stochastic Processes time homogenous Discrete-Markov Chain then that satisfies this equation and this equation is known as the Chapman-Kolmogorov Equations. In the matrix form you can write the capital P is the matrix which consist of the element of one step transition probabilities.

In that case if you make m=1 and n=1 then the matrix of $P^{(2)}$ that is a matrix form of 2-Step transition probability that is nothing but if you put n=1 and m=1 you will get P into P and that is going to be P^2. So the right-hand side $P^{(2)}$ means it is a 2-Step transition probability matrix and the right hand side the P square that is the square of the P matrix where P is the one step transition probability matrix.

So in this form in general you can make the n-Step Transition Probability matrix is nothing but P(n) for n is greater than or equal to 1 for n=1 it is obvious for n= 2 onwards the P power n that is same as the n-Step Transition Probability matrix.

(Refer Slide Time: 16:25)



Hence, so now we got the n-Step Transition Probability is nothing but the P^n where P is the one step transition probability matrix therefore in matrix form I can given the P(n) is nothing but in the matrix form of the distribution of X_n or this is nothing but the vector which consist of a the nth step where the system will be. So this is nothing but what is the probability that in the nth step the system will be in the state 0 or in the nth step, the system will be in the state one and in the nth step the system will be 2 and so on. This is the vector. So P(n) you can find out in the matrix form by using the above equation; it is going to be P(0) that is also vector initial probability vector multiplied by $P^{(n)}$ that is the n-Step Transition Probability matrix. But the n-Step transition Probability matrix is nothing but the P^n therefore this is same as the P(0) P^n.

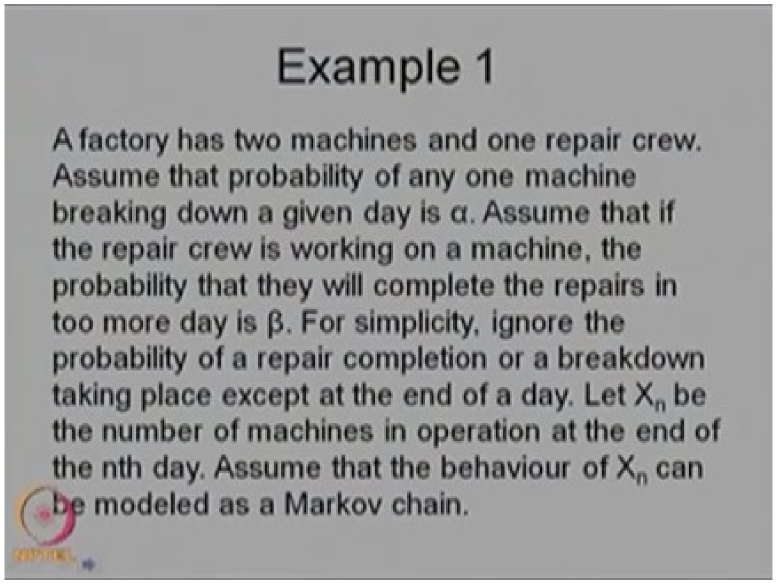
In the last slide, we got $P^{(n)}$that is the n-Step Transition Probability matrix is same as the one step transition probability with the power n therefore this is going to be the distribution of Xn in the vector form that is same as the P(0) $P^n$where the P is nothing but the one step transition probability matrix. That means if you want to find out the distribution of X_n for any n you need only the initial probability vector and one step transition probability matrix.

Because the Discrete-time Markov Chain is the time homogenous we need only the one step transition probability matrix and the initial probability vector that gives to find out the distribution of X_n for any n. So with the help of one step transition probability and the initial probability vector you can find the distribution of X_n for any n.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

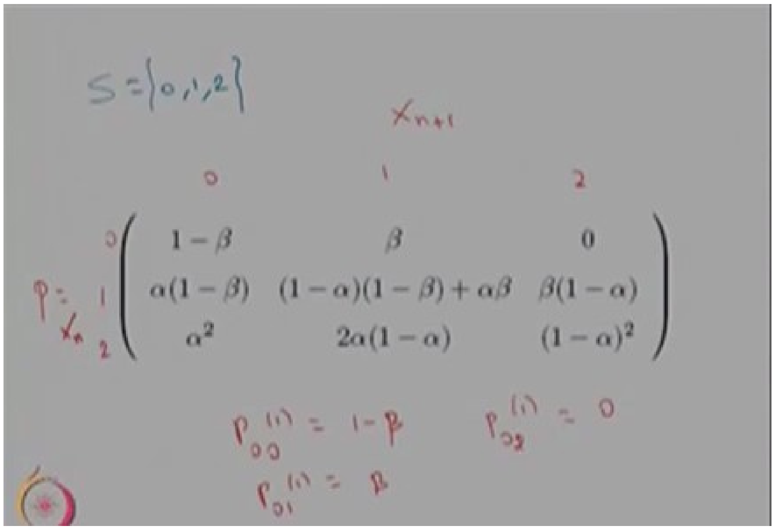
Lecture – 65

(Refer Slide Time: 00:01)



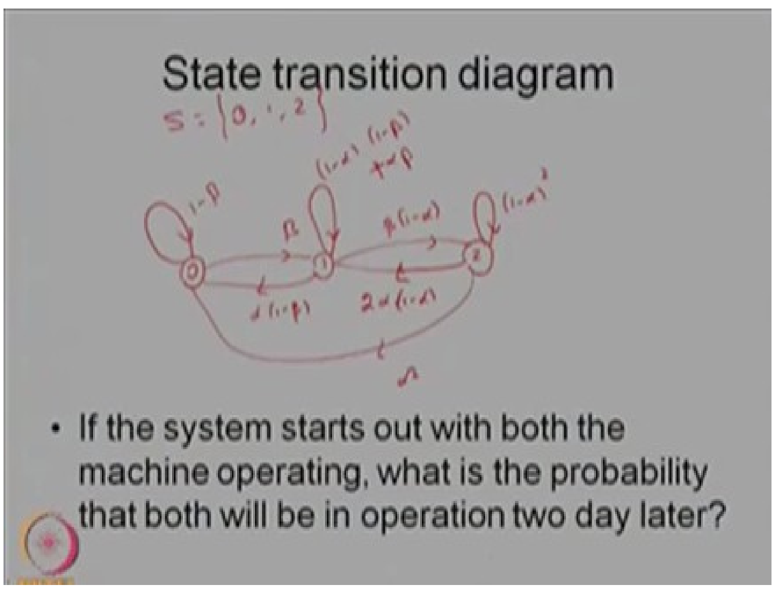
Now you are moving to simple examples using the n-Step Transition Probability matrix and the one step transition probability vector how to find the distribution of X_n for some simple example. The first example which I have discussed in the lecture one, this is a very simple example in which the underlying stochastic processes is a time homogenous Discrete-time Markov Chain with the state space S is 0, 1 and 2.

(Refer Slide Time: 00:40)



So this is the state space and the information which we have based on that we can make a one step transition probability matrix that is nothing but what is the possible probability in which the system is moving from state i to j in one step that you can fill it up. So this exercise you have done it in the lecture one.

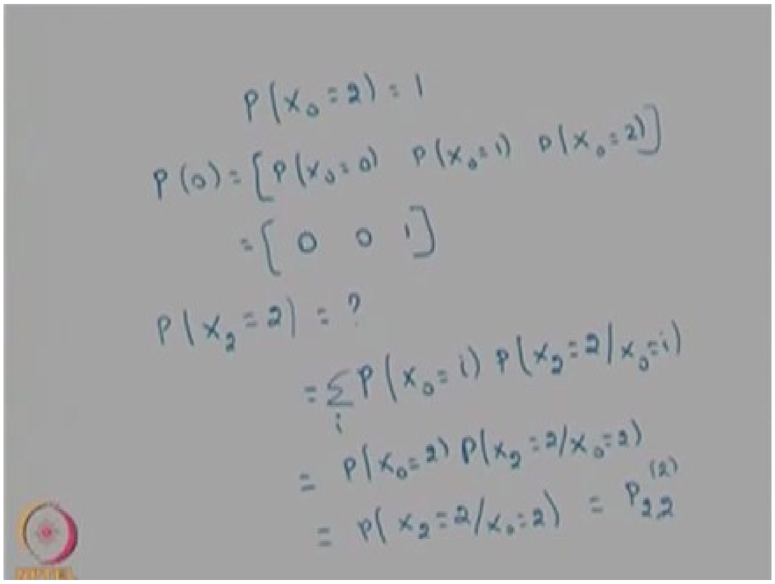
(Refer Slide Time: 01:13)



Now our interest is to find out and also we have made the State transition diagram which is equivalent to the one step transition probability matrix and we have got this state transition diagram. Now the question is if the system starts out with both the machines operating, what is the probability that both will be in operation two days later?

So if you recall what is the random variable X_n be the number of machines in operation at the end of the nth day. So the random variable is how many machines are in the operation at the end of nth day. So here the clue is at the time 0 or the 0th step both the machines are operating therefore P$\{X_0 = 2\} = 1$. So the given information with the probability 1 both the machines are working at 0th step.

(Refer Slide Time: 02:24)



So this can be converted into the P$\{X_0 = 2\} = 1$. Or you can make it in the initial probability distribution or initial probability vector as a that is the probability that at X_0 the system was in the state 0 at the 0th step the system was in the state one so this is the initial probability vector. At a time 0 the system was in this state two therefore that probability is one and all other probabilities are 0. So this is the given information about the initial probability vector.

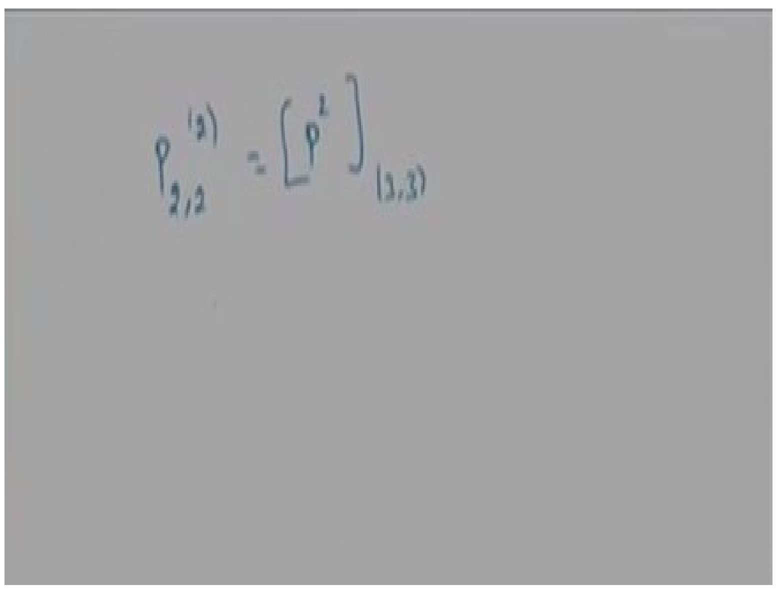
Now the question is what is the probability that both will be in operation two days later? That means, what is the probability that we can convert this into what is P$\{X_2 = 2/X_0 = 2\}$. So this is what you have to find out that is the conditional probability system, if the system starts with the both machines what is the probability that both will be operation two days later.

So not even this a conditional probability the question is what is the probability that; what is the probability that the system will be in the state. So to find this you can make what is the probability that-- the given information is there X_2 is equal to 2 given that X_0 is equal to i for all possible values of i.

This is same as since the initial probability vector is going to be (0, 0, 1) one so this is land up; P{X_0 = 2}P{X_2 = 2/X_0 = 2} and all other probabilities are 0 therefore 0 into anything is going to be 0 therefore it is same as P{X_0 = 2} and to the conditional probability.

And P{X_0 = 2} = 1, therefore this is same as what is the P{X_2 = 2/X_0 = 2}. So this is a same as what is the probability that {2, 2} in two steps. This is nothing but the system was in the state two at 0^{th} step and the will system being the state two after the two steps. So this is the two step transition probability of system moving from the state 2 to 2.

(Refer Slide Time: 05:45)

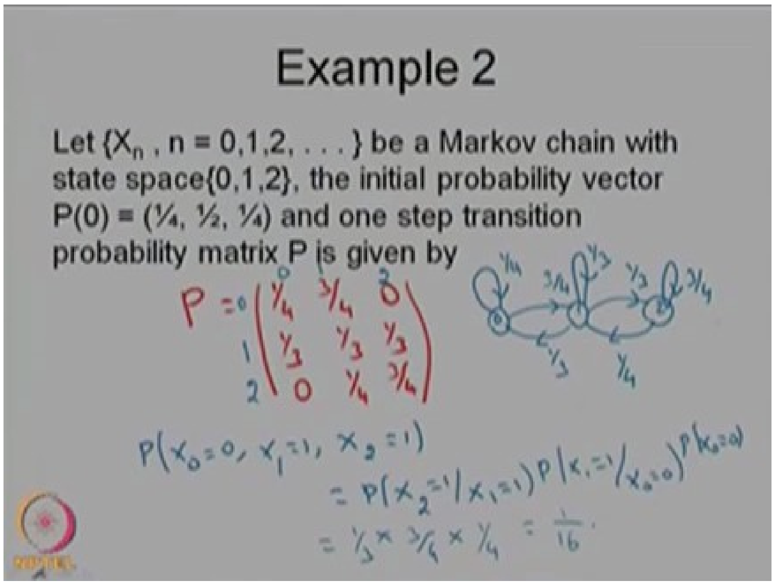


This is same as you find out the P^2 and from the P^2 this nothing but the {2, 2} that is going to be the last element, out of that nine elements the third row, third column element that is going to be the element for this probability. So what do you have to find out is that find out the P^2; find out the P^2 so we have provided the P.

So this is the P matrix so from the P matrix you will found out the P^2 so the P^2 is also going to be a three cross three matrix. So from the P^2 three cross three matrix you take the third row third element third column element and that is going to be the probability for two step transition of system moving from the state 2 to 2 that is going to be the answer for the given question.

What is the probability that both will be in operation in two days later? Similar to this you can find out the probability for any day or any finite day by finding the P^n matrix then pick the corresponding element and that is going to be the corresponding probability.

(Refer Slide Time: 07:29)



Now, we moving to the next example. This is abstract example in which the X_n be the discrete-time Markov Chain the default discrete-time Markov Chain is always it is a time homogenous. So this is the time homogenous discrete-time Markov Chain with the state space 0, 1 and 2. And also it is provided the initial probability vector that is P(0) that is the vector that is one fourth of one fourth. So the submission is going to be one therefore this is the initial probability vector.

That means the system can start from the nth 0th step with the probability 1/4 from the state 0. From, the state one which the probability is 1/2 and with the probability 1/4 it can start from the state 2. And also it is provided the one step transition probability matrix. From the one step transition probability you can draw the state transition diagram also because the state space is 0, 1, 2 therefore the nodes are 0, 1 and 2.

And this is the one step transition probability therefore 0 to 0 that probability one step the system is moving from the state 0 to 0 that is 1/4 and the system is moving from the state 0 to 1 in one step that is 3/4 and there is no probability going from the state 0 to 2 therefore you should draw the arch. From one, the one step transition probability of 1 to 1 is 1/3 and this is 1/3 and similarly this is 1/3.

From the state two, 2 to 0 is 0 and 0 to 1 is 1/4 and 2 to 2 is 3/4. This diagram is very important to study the further properties of the states therefore we are drawing the state transition diagram for the Discrete-time Markov Chain. So this is the one step transition probability matrix and this is the state transition diagram. Our interest is to find out the few quantities that is that what is the $P\{X_0 = 0, X_1 = 1, X_2 = 1\}$.

What is the probability that the system was-- it is a joint distribution of these three random variable $\{X_0 = 0, X_1 = 1, X_2 = 1\}$. So this is same as the joint distribution the same as you can write in the product of the conditional distribution and a conditional distribution again you can write it using the Markov property the conditional probability of only one step.

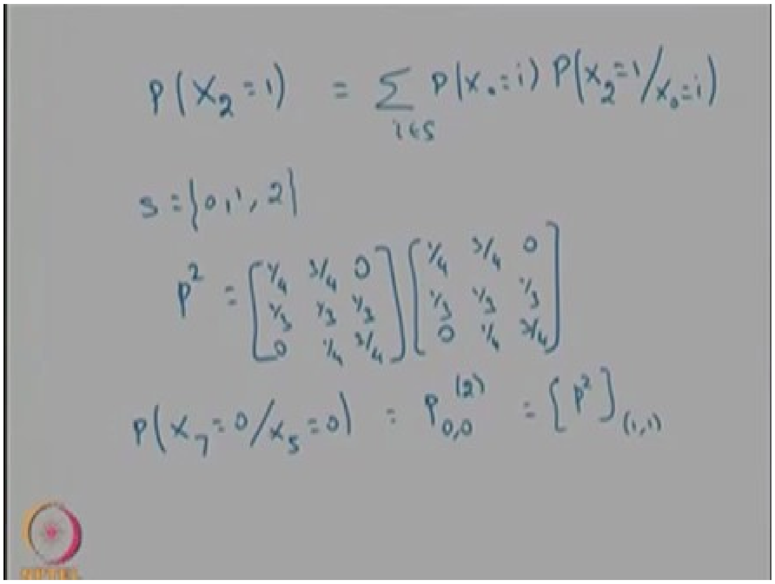
Therefore, this is going to be by using the probability theory, you apply the joint distribution is same as the product of conditional distribution by using the Markov property you reducing into the another conditional distribution so this is same as what is a $P\{X_2 = 1/ X_1 = 1\}P\{X_1 = 1/X_0 = 0\}P\{X_0 = 0\}$.

So this is the first term is nothing but the one step transition of system is moving from one to one and this nothing but the system is moving from the state 0 to 1 and this is the initial probability-- you take the probability from the initial probability vector of X_0 is equal to 0. Now, we are going to label the one step transition probability matrix with the 0, 1, 2 and 0, 1, 2 from this, you can find out this is the one step transition probability of system moving from 1 to 1.

So 1 to 1 is 1/3 into-- this is the system probability of system moving from 0 to 1 so 0 to 1 is 3/4 and a system started from the state 0 in the 0th step so that you can take it from the first element that is 1/4. So if you do the simplification we will get 1/16. So this is the joint distribution of the system was in state 0 at 0th step; the system was in the state one at the first step and the system was in the state one at the second step their probabilities 1/16.

Similarly, you can find out the other probabilities also that is suppose our interest is what is the probability that at the end of second step the system will be in the state one.

(Refer Slide Time: 12:38)



That is nothing but what are all the possible states in which the system would have been started from the state i and what is the two step transition of system is moving from the state i to 1. So the i is belonging to S, so here the S is a 0, 1, 2 so already we have given the initial probability vector that is one fourth and one fourth using this.

And you need two step transition probability that means you need to find out what is the P square. So the P square will give the two steps transition probability matrix therefore the P is provided to you so the P is a 1/4, 3/4 and 0; 1/3, 1/3, 1/3; 0, 1/4, 3/4, so this is the P. So you multiply the same thing /3, 1/3, 1/3; 0, 1/4, 3/4.

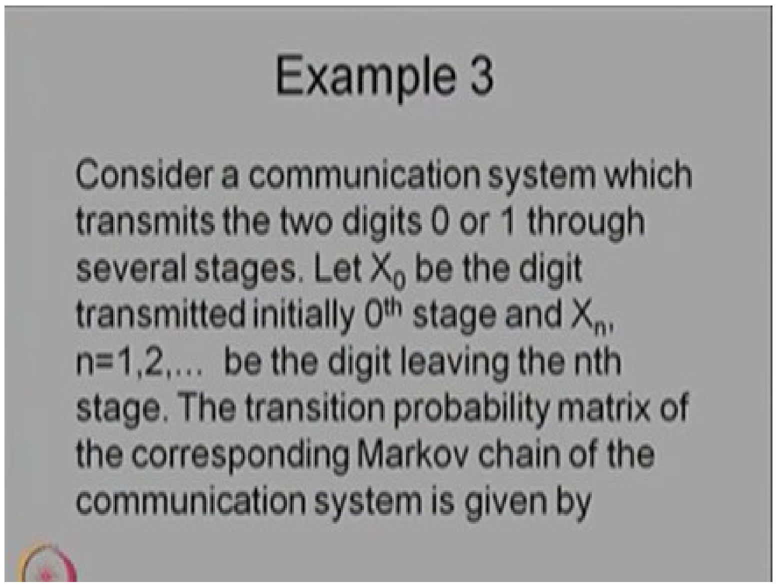
You find out the P^2 so from the P^2you pick out the element of X_0 is equal to for all possible i then multiply this and this that multiplication will give you probability of X_2 is equal to 1. So I am not doing the simplification. So once you know the P^2you can find out probability of X_2 is equal to 1. Similarly, one can compute the other conditional probabilities also.

Suppose our interest find out what is the P$\{X_7 = 0/X_5 = 0\}$. This is same as what is the probability that the system was in the state 0 with the fifth step given that what is the probability that this system will be in 0 in this seventh step, that is same as What is the probability of 0, 0 in two steps, that means you find out the $P^2$from the $P^2$the 0, 0 is nothing but you take the first row first column element and that is going to be the probability of $P\{X_7 = 0/X_5 = 0\}$. Similarly, you can find out all other different conditional probability and what do you have to do is always you have to convert because of the given DTMC is a time homogenous, so you convert that into find out the n-Step Transition Probability and the n-Step Transition Probability same as the P^n, so you pick the corresponding element to find out the conditional probability.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 66

(Refer Slide Time: 00:01)



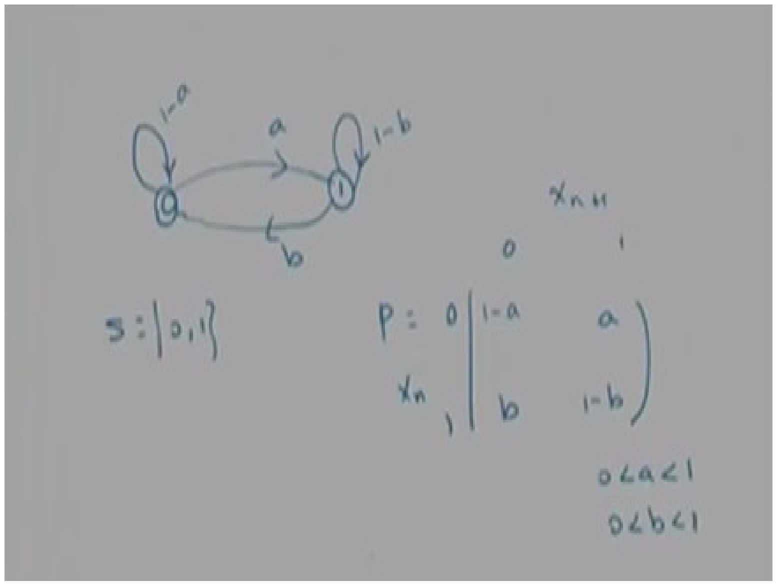
This example talks about the communication system in which whenever the transmission takes place with the digits 0 and 1 in the several stages. Now we are going to define the random variable X_0 be the digit transmitted initially that is 0th step. Either the transmission digital be 0 or 1 therefore only two possibilities can be takes place at any nth step transmission either 0 or 1, like that we are making the transmission over the different stages.

Therefore, this X_n over the n will form a stochastic process because we never know which digit is transmitted in the nth stage. So nth stage is going to be one random variable and you have a collection of random variables over the stages therefore it is a sequence of random variables so this going to form a is Stochastic process. And this stochastic process is nothing but a discrete time discrete stage stochastic process.

Because the possible values of X_n is going to be 0 or 1 therefore the state space is 0 or 1 and it is a discrete state stochastic process. The way the subsequent transmission takes place depends only on the last transmission not the previous stages therefore you can assume that this follows

Markov properly. Therefore, this stochastic process is going to be call it as a Discrete-time Markov Chain. Now our interest is to find out-- so now I will provide what is the one step transition probability for the Markov Chain or let me give the transition diagram for that.

(Refer Slide Time: 01:55)



So state transition diagram the possible states are 0 or 1 because the state space is 0 and 1. And the probability that in the next step also the transmission is 0 with the probability 1-a. This is a conditional probability of the nth stage the transmission was 0 the (n+1)th stage is also the transmission 0 with the probability 1-a.

The one step transition probability of system is moving from 0 to 1 that probabilities is a. That means the nth stage the transmission was the digit 0 the (n+1)th stage the transmission will be digit 1 with the probability a. Similarly, I am going to apply the one step transition probability of 1 to 1 that is 1-b and this is a b that means this 1 to 0 that probability is a 1 to 0 is a b and 1 to 1 is 1-b.

Obviously, this a is lies between 0 to 1 and b also lies between 0 and 1.

(Refer Slide Time: 03:30)

$$
\begin{array}{cc}
 & \begin{array}{cc} \quad 0 & \qquad\qquad 1 \end{array} \\
P^{(n)} = \begin{array}{c} 0 \\ \\ 1 \end{array} & \begin{bmatrix} \dfrac{b+a(1-a-b)^n}{a+b} & \dfrac{a-a(1-a-b)^n}{a+b} \\ \dfrac{b-b(1-a-b)^n}{a+b} & \dfrac{a+b(1-a-b)^n}{a+b} \end{bmatrix}
\end{array}
$$

$$for \ |1-a-b| < 1$$

So this is the a is the probability that the system is transmitting from the nth stage with the digit 0 and the (n+1)th stage with the digit 1 that probability is a therefore the negation is 1-a because there is the system can transmit either 0 or 1. So once you say that one step transition probability of 0 to 1 is a then 0 to 0 will be 1-a. Similarly, 1 to 0 is given as a probability b and the other digit transmission will be one therefore it is going to be 1 to 1 will be 1-b. So this is the state transition diagram.

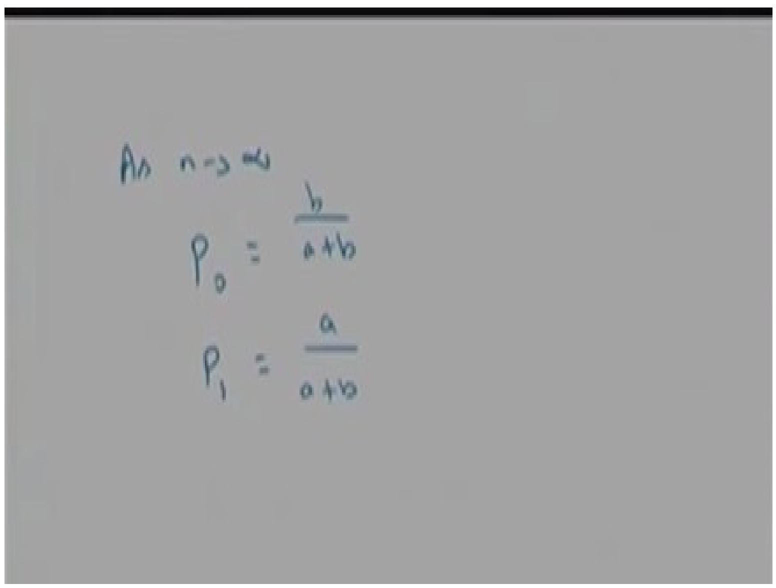
And this is the one step transition probability for a given time homogenous discrete-time Markov Chain. Our interest is to find out what is the distribution of X_n for n. For that you need what is the n-Step Transition Probability matrix. Since the one step transition probability matrix is given you can find out P^2, $P^3$and so on. By induction method you can find out the P^mbut using the P^m you can find out the P^{m+n}.

Therefore, you can come to the conclusion what is the n-Step Transition Probability of system is moving from 0 to 1 and 0 to 0 and so on. So this is nothing but I am just giving the only the result $\frac{b+a(1-a-b)^n}{a+b}$ and this is nothing but $\frac{a-a(1-a-b)^n}{a+b}$.Similarly, if you find out the n-Step transition probability of system moving from 1 to 0 that is $\frac{b-b(1-a-b)^n}{a+b}$.

This is nothing but $\frac{a+b(1-a-b)^n}{a+b}$. So here I am just giving the n-Step Transition Probability in matrix form. By given P you should find out P^2, P^q by induction you can find out the P^n. And this is valid provided 1-a-b which is less than 1. Because we are finding the P^n matrix, so here it needs some determinant also unless otherwise the absolute of 1-a-b which is less than 1, this result is not valid.

So provided this condition, the $P^{(n)}$ that is the matrix. So that is same as P^n also. $P^{(n)}$ is same as P^n.

(Refer Slide Time: 06:50)

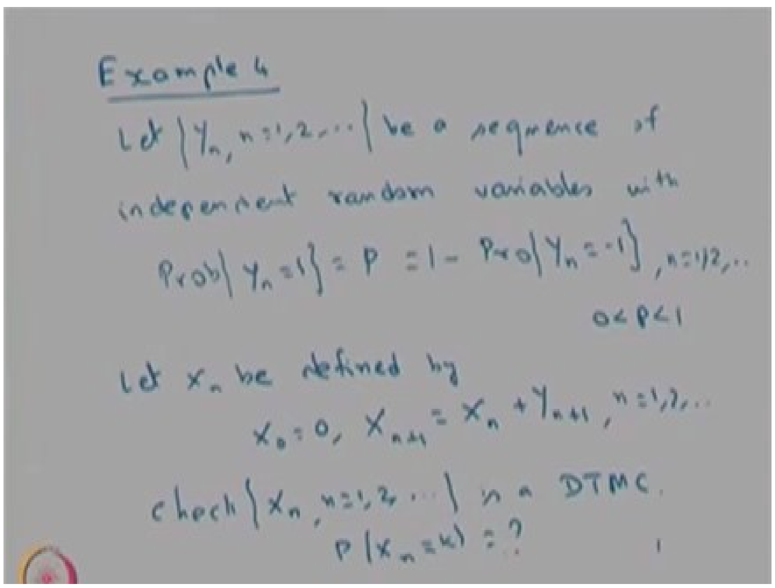


So as n tends to infinity you can come to the conclusion what is the probability that the system will be in the state 0 that is same as $\frac{b}{a+b}$. And similarly, what is the probability that the system will be in the state 1 as entrance to infinite that will be $\frac{a}{a+b}$. This can be visualized from the state transition diagram easily. Whenever the system is keep moving into the state 0 or 1 with the probability a, b and with the self-loop on 1-a and 1-b the subsequent stages the system will be any one of these two states.

So with the proportion of $\frac{b}{a+b}$ the system will be in the state 1. Similarly, with the proportion $\frac{a}{a+b}$ the system will be in the state 1 with the proportion $\frac{b}{a+b}$ the system will be in the state 0 in a long run. The interpretation of as n tends to infinite this probability is nothing but in a long run with this proportion the system will be in the state 0 or 1.

So this state transition diagram will be useful to study the long run distribution or where the system will be as n tends to infinity to study those things the state transition diagram will be useful.

(Refer Slide Time: 08:42)



Now we will move to the next problem that is Example 4. Let it is a sequence of random variable be a sequence of independent random variables with condition the P{Y_n = 1} = p = 1 – P{Y_n = -1}.

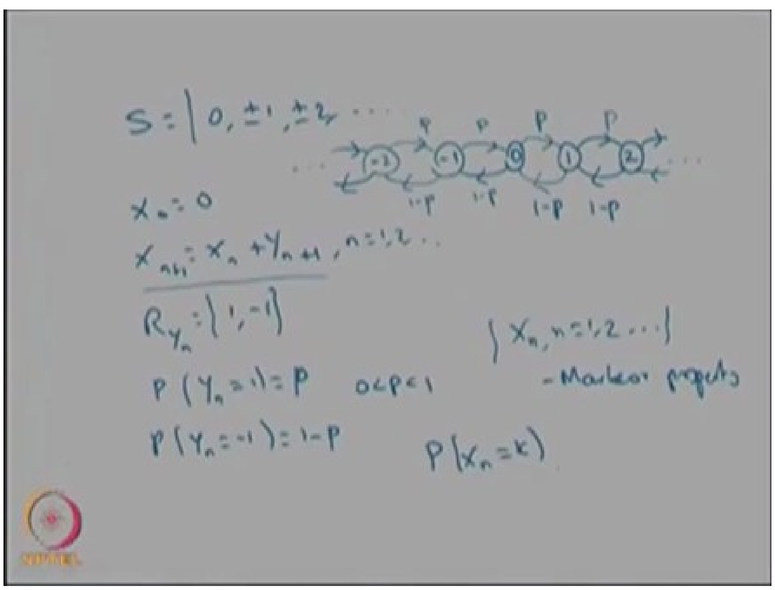
We have a stochastic process and each random variable is independent random variable and the probability mass function is provided with this situation the P{Y_n = 1} = p.

You can assume that p takes the value 0 to 1. That is same as 1 – P{Y_n = -1} for all n. Now I am going to define another random variable. Let {X_n} be defined by X_0 is equal to 0 whereas X_{n+1} onwards that is going to be X_n + $Y_{n+1}$for n is equal to 1, 2 and so on. So we are defining another random variable X_n with the X_0 = 0 and X_{n+1}= X_n + Y_{n+1}. Now the question is check { X_n} that stochastic process is the DTMC.

If it is a DTMC also find out what is the P{X_n = k}. We started with one stochastic process and we defined another stochastic process with the earlier stochastic process and check whether the given the new stochastic process is a Discrete-time Markov Chain that is the default one that is time homogenous Discrete-time Markov Chain. If so then what is the probability of X_n takes the value k that is nothing but find out the distribution of X_n.

So how to find out this the given or the X_n is going to be the DTMC. Since Y_n takes the value 1 with the probability p and Y_n takes the value -1 with the probability 1-p you can make out the possible values of Y_n is going to be 0 or +1 or -1; +2 or -2 and so on.

(Refer Slide Time: 12:18)



Because the relation is a X_0 = 0 and the X_{n+1}= X_n + $Y_{n+1}$and the range of Y_n is 1, -1 therefore the range of X_n that inform a state space and X_0 is equal to 0 therefore X_{n+1}= X_n + Y_{n+1} so n takes a value 1 and so on.

Therefore, the possible values of X_n will be 0, +1 or -1; +2 or -2 and so on therefore that will form a state space. Now the given clue is that P{Y_n = 1} = p and P{Y_n = -1} = 1- p and the probability p is lies between 0 to 1. So using this information you can make a state space of the X_n that is going to be 1, 2 and so on -1, -2 and so on.

Now we can fill up what is the one step transition of system is moving from 0 to 1 that means the X_1 – 0 to 1 suppose you substitute 0 here then suppose it takes the value 1 then the system can move from the state 0 to 1 in one step. Suppose you put the value X_n is equal to 0 suppose you put X_n is equal to 0 and Y_{n+1} takes the value 1 with the probability p then the X_{n+1} value will be one with the probability p.

Now you can go for what is the state transition probability of 1to 0. Suppose the X_n = 1; suppose the Y_{n+1} = -1 then the X_{n+1} = 0. So the one step transition of a system moving from 1 to 0 because of happening P{Y_{n+1} = -1} = 1-p. So whenever the system is moving from one step forward that probability will be the probability p and one step backward that probability will be 1-p.

So this is the way it goes forward step and this is the way it goes to the backward step so you can fill up all other probabilities forward probability with the probability p and the backward probability with the 1-p. Also, we can come to the conclusion, the way we have written X_{n+1}= X_n + Y_{n+1} and all the Y_i's are independent random variable the X_{n+1}going to take the value depends only on X_n not the previous X_{n-1} or X_{n-2} and so on.

Therefore, the conditional distribution of X_{n+1} given that X_n, X_{n-1} till X_0 that is same as the conditional distribution of X_{n+1} given X_n. That means the X_n is going to satisfy the Markov property because of this relation because of X_{n+1} is equal to X_n plus independent random variable. Therefore, the X_n, n= 1, 2, 3 and so on. This Stochastic process is going to satisfy the Markov property therefore this discrete time discrete state stochastic process is going to be the Discrete-time Markov Chain because of the Markov property satisfied.

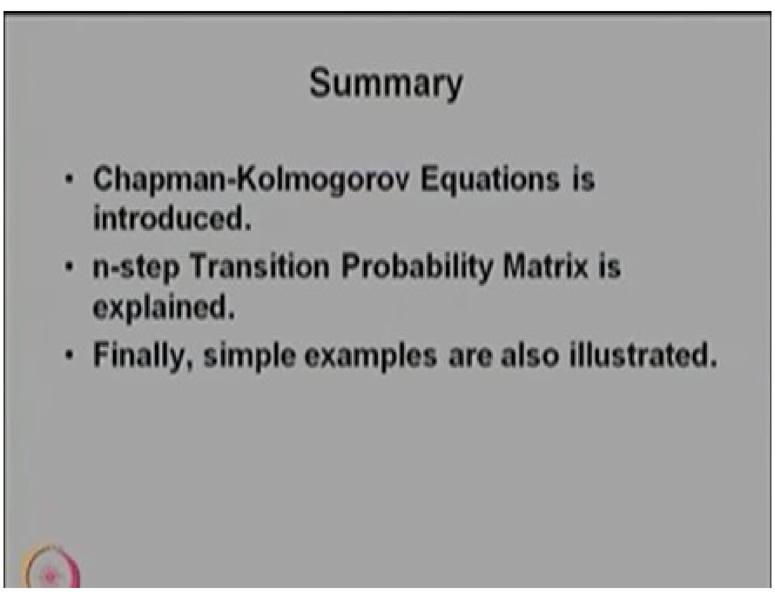
Once it is Markov property satisfied by using the Chapman-Kolmogorov Equation you can find out what is the distribution of X_n takes the value k, that is nothing but where it started a time 0 and what is the conditional distribution of n-Step Transition Probability and the n-Step Transition Probability is nothing but the elemental from the $P^n$and From here you can find out the one step transition probability matrix from the one step transition probability matrix you can find out a P, P^2, $P^3$and so on.

And you can find out the P^n and that element is going to be the n-Step Transition Probability using that you can find out the distribution. And since we do not know the value of where p is lies between 0 to 1. I am not going to discuss the computational aspect of finding out the distribution.

This is left as an exercise and the final answer is provided. The difference between the earlier example and this example is the state space is going to be a countable infinite. Therefore, the P is not going to be easy matrix it is going to be a matrix with the many elements in it therefore finding out $P^2$and P^n is going to be little complicated than the usual square matrix.

So hence the conclusion is a by knowing the initial probability vector and the one step transition probability matrix or the state transition diagram we can get the distribution of X_n for any n. There is a small mistake the running index for X_{n+1}= X_n + Y_{n+1} that is you starting from 0, 1, 2 and similarly the previous slide X_{n+1}= X_n + $Y_{n+1}$and the n is running from 0, 1, 2 and so on.

(Refer Slide Time: 19:36)



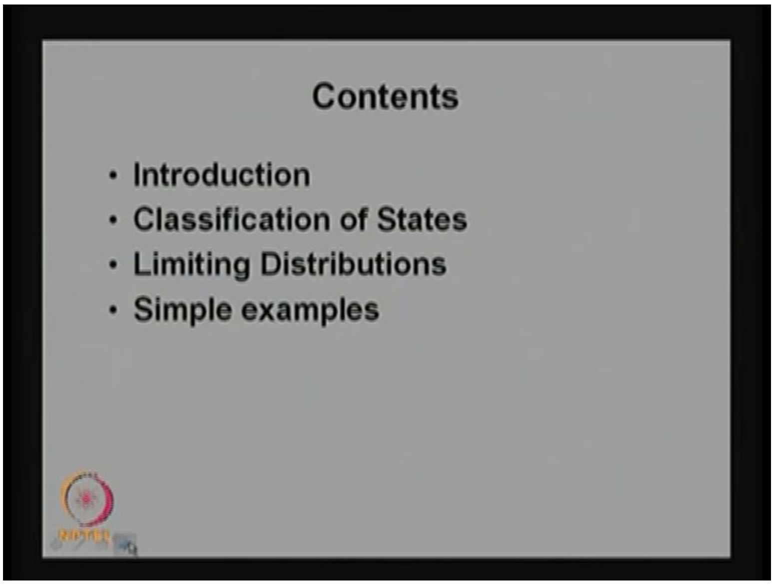
So in this lecture we have discussed Chapman-Kolmogorov Equation and also we have discussed the n-Step Transition Probability Matrix. So the n-Step Transition Probability Matrix can be computed from the one step transition probability matrix with the power of that n. And also we have discussed four simple examples for explaining the Chapman-Kolmogorov Equation and the n-Step Transition Probability Matrix.

Thanks

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 67

(Refer Slide Time: 00:01)

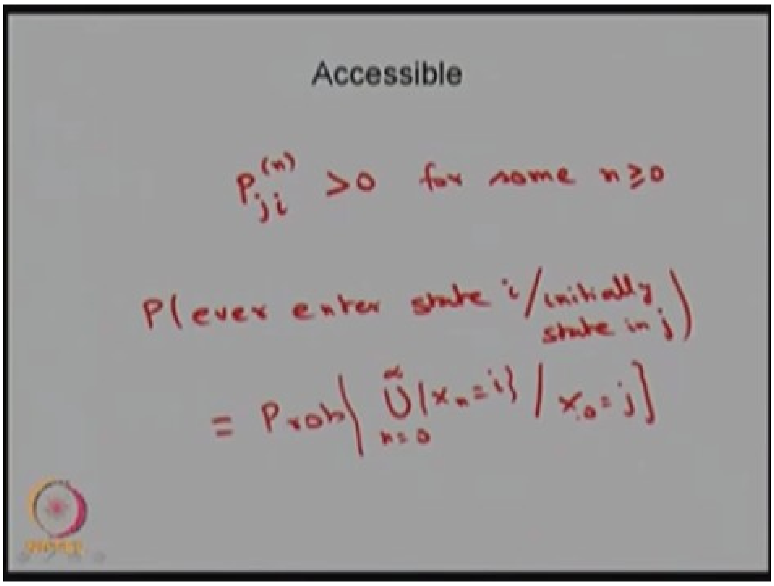


In this we are discussing Discrete time Markov Chain and we have finished already two lectures on this model and this is a third lecture classification of a states and limiting distributions. In this lecture, I am going to give the information about the classification of the states for the Time Homogeneous or Discrete time Markov Chain then I am going to give the definition of a limiting distribution then I am going to discuss a few simple examples.

So that we can understand the classification of a states as well as the limiting distribution. Why do you need classification of states? Whenever we study the Time Homogeneous or Discrete Time Markov Chain our interest is to find out the limiting distribution of the random variable X_n. To study the limiting distribution or stationary distribution later, we are going to use word called an equilibrium distribution.

All those things, you need the classification of a state without the classification of a states we cannot come to the conclusion whether the limiting distribution exist whether that is going to be unique and so on. So for that we need a classification of states. Before moving into the classification of states we need some concepts so that using those concepts we can classify the states.

(Refer Slide Time: 01:47)

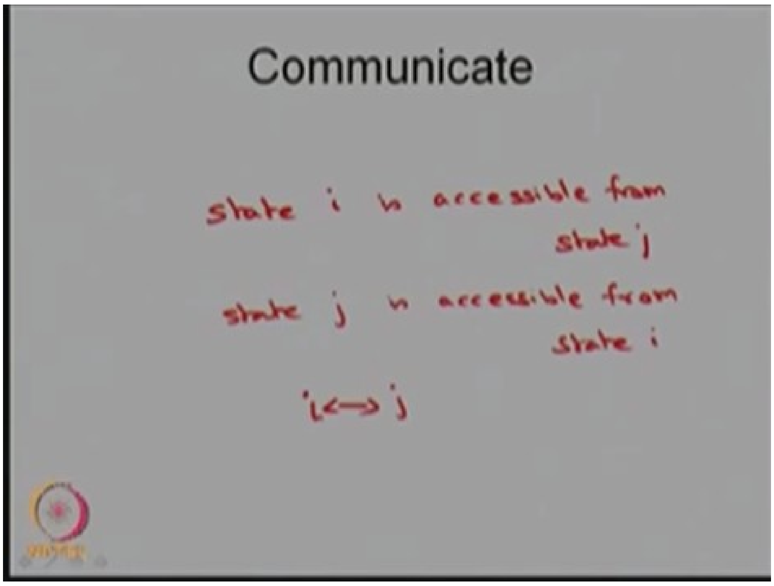


The first concept is called accessible. When we say the state i is said to be accessible from the state j whenever the $P_{ji}^{(n)} > 0$ for some n which is greater than or equal to 0. We are including n is equal 0 for the safer side whenever we say the state i is said to be accessible from the state j if the $P_{ji}^{(n)} > 0$.

That means this is the transition probability from the n step transition probability matrix and if that element is going to be greater than 0 then we say the state i is said to be accessible from the state j. Using this, we can write down what is the probability that ever enter state i given that initially the system is in the state j.

You can find out what is the probability of the system ever enter to the state i given that initially it was in the state j that is nothing but the union of all the events corresponding to the X_n takes a value i given that it was in the state j initially. You can find out what is the probability that ever entering the state i given that initially the system is in the state j that is the $P\{\cup\{X_n = i\} / X_0 = j\}$.

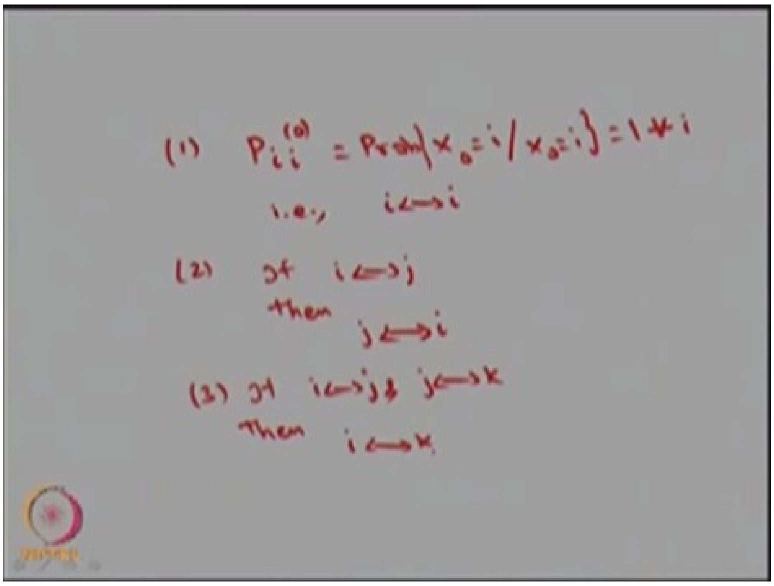
(Refer Slide Time: 04:04)



Now I am going to define, now I am going to give the next concept called communicate using the accessible. Two states are said to be communicate that means the state i is accessible from the state j as well as the state j that is accessible from the state i. Whenever the state i is communicate with the state j that means the state i is accessible from the state j as well as the state j is accessible from state i.

In notation, we can use the notation i arrow in both side with the j. State i is communicating with the state j means that state i is accessible from state j as well as the state j is accessible from state i. Since I use the concept of access to define the communicate it is going to satisfy few properties. The first property any state communicates with itself.

(Refer Slide Time: 05:17)



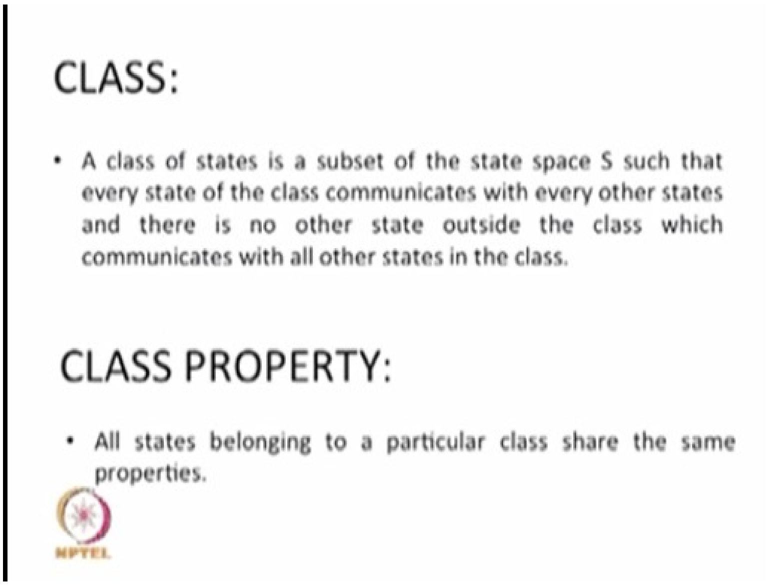
That means that the $P_{ii}^{(0)}$ that is nothing but what is the probability that the X_0 is equal to i given that X_0 as equal to I that is going to be one for all i. Any state communicates with itself. The second one that means that in notation i communicates with the i itself the second property.

If state i communicate with the state j then the state j communicate with the state i also that means it is a symmetric property that means if i communicates with the j then j communicates with the i so the communicates satisfies the symmetric property. The third one if i communicate with the j and j communicate with the k then we can conclude the i communicates with k. This relation is called a transitive.

So the communicate that properties satisfies itself and it satisfies the symmetric property as well as the transitive property that is if i communicates with the state j and state j communicates with the state k then the state i communicates with the state k. Communication is an equivalence relation on the set of states and hence this relation partitions the set of states into communicating classes.

I am not giving the proof here so one can easily prove using the one step transition probability and the n step transition probability matrix and the accessible concept one can prove these three properties.

(Refer Slide Time: 07:28)



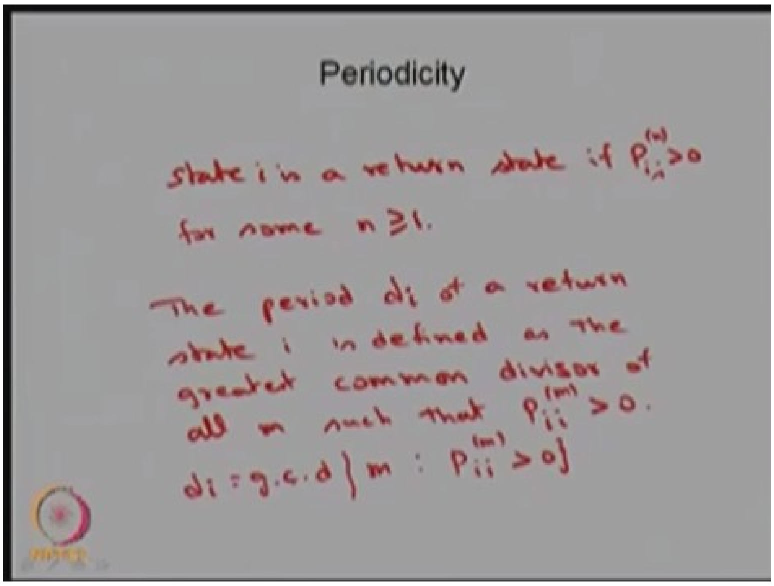
Now I am going to define the next concept called the class property. What is class property? Here class of state is a subset of the state space S such that every state of the class communicates with every other states and there is no other state outside the class which communicates with all other states in the class. Instead that time Homogenous Discrete Time Markov chain.

Since it is a Discrete Time Markov Chain you have a state space the state space maybe a finite number of elements or countably finite number of elements so that is state space S. In the state space S, you are going to create a subset that is going to be call it as a class if within the subset of that collection it satisfies the communicate that means each state inside the class has to be communicate with each other state.

And also it has to satisfies the second property there is no other state outside the class which communicates with all other states in the class that means you can start with one element then you can include one more element once this property is satisfied that means you cannot make a including one more state and make it as a class then you have to stop the framing the class. So the subset will be created by including one more state, one more state, one more state, in the state space as long as this property is satisfied.

So once the second property violates that means you should stop with creating that subset and that is going to be the class. We are going to discuss this how to create the class I have to play some examples so that I am going to do it later. Next concept is periodicity. The definition of periodicity goes like this.

(Refer Slide Time: 09:31)



The state i is return state if the $P_{ii}^{(n)}$ which is greater than 0 for some n which is greater than or equal to 1 first time defining what is the meaning of return state. Here any state is going to be called it as a return state if the probability of starting from the state i coming to the same state in the nth step if that is greater than 0 then we say it is a return state.

Now I am going to define the periodicity only for the return state. The period in notation it is d_i, i is for the state of a return state is defined as the greatest common divisor of all m such that $P_{ii}^{(m)}$ which is greater than 0. So the period of return state is going to be a integer and that integer is computed by using the greatest common divisor of all the possible m such that $P_{ii}^{(m)}$ should be greater than 0.

That means you find out how many steps you will take to come to the same state if you start from the state i. You collect all the possible number of steps you will come back to the state with the positive probability and you find out the greatest common divisor of those integers those positive integers. Then that number is going to be the period or periodicity of the return state or the period of the state that means you can write down in short d_i is the greatest common divisor collection of m such that the $P_{ii}^{(m)}$ of m should be greater than 0.

If the greatest common divisor of collection of m such that greater than 0 if this d_i is going to be 1 then we say that state is aperiodic state. Otherwise if it is greater than 1 and whatever be the integer you are going to get and that is going to be period of the state i. If the period is going to be 1 then we call it as the aperiodic state.

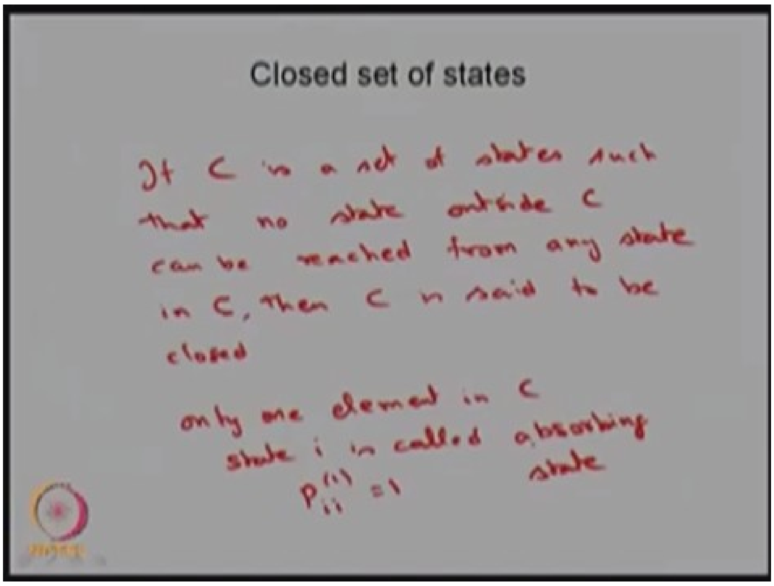
Note that whenever you have a class in which we have a more than one states if one state has the period some number than the other states of the same class also going to have the same period that can be proved easily. So within the class all the states will be having the same period.

Introduction to Probability Theory and Stochastic Processes

Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 68

(Refer Slide Time: 00:05)



Now I am going for the next concept called a closed set of states. If C is set of states such that no state outside C can be reached from any state in C then we say the collection or the set C is said to be closed. So whenever you create a collection of states and that set we call it as a C. If it satisfies this property then we say that set is called closed set.

So we can combine the class property with the closed set property if both the properties are satisfied they communicates with each other as well as the closed property satisfies then you can say that a closed communicating class. So any subset in the states space S if it satisfies each element within the set is communicating each other and satisfies this property then we say that collection is going to be a closed communicating class.

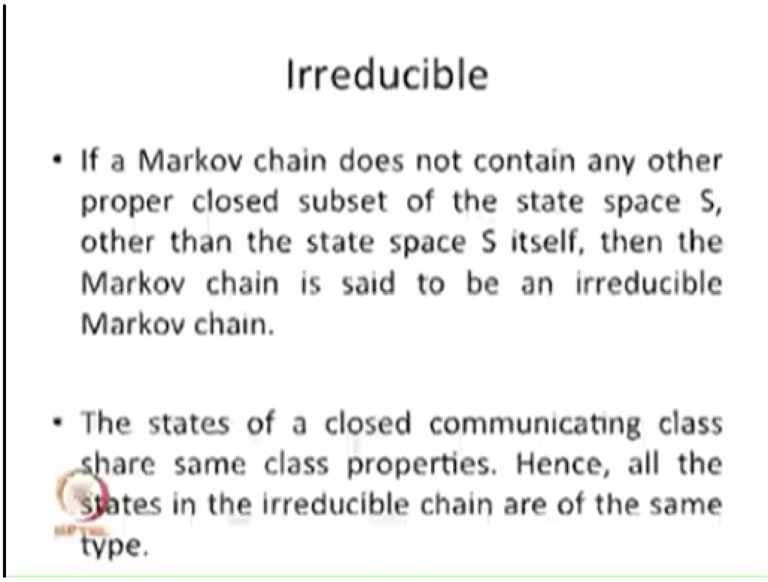
There is a possibility in a set you can have a more than one element, one than one states in the collection. The class may have only one element or it may be more than one element. If any closed set or the closed communicating class has only one element that means you cannot include one more state and to make it as the closed or communicating class then that closed set is called or that state is called only one element in C then the state i is called absorbing states.

The state i is said to be absorbing state then it is going to form a closed communicating class which has only one element in that class. There is a possibility more than one element is also possible in the closed communicating class. So we can define the absorbing state through the closed communicating class or we can make it in the same absorbing state using the definition $P_{ii}^{(1)}=1$.

That means if you see the one step transition probability matrix the diagonal element of that corresponding state the corresponding row, the element is going to be one that means the system starting from the state i and in one step the system moves into the same state i that probability is one. If this probability is one, then we say that state is going to be absorbing state.

In another way around you can go for defining the absorbing state by a closed communicating class has only one element also. So there are two ways we can say the absorbing state. Using these concepts, I am going to develop the next concept called the Irreducible Markov Chain. We are discussing a Time Homogenous Discrete Time Markov Chain whereas this concept called the irreducible that is valid for the Discrete Time Markov Chain as well as the Continuous Time Markov Chain so that we are going to discuss later.

(Refer Slide Time: 04:29)



Now I am defining the Irreducibility for a Time Homogeneous Discrete Time Markov Chain.

If the Markov Chain since the irreducible concept comes for the Discrete Time Markov Chain and a Continuous Time Markov Chain, we use the word called the Markov Chain that is valid for both. If the Markov Chain does not contain any other proper closed subset other than the state space S, then the Markov Chain or in short we can use the word MC or Markov Chain then the Markov Chain is called irreducible Markov Chain.

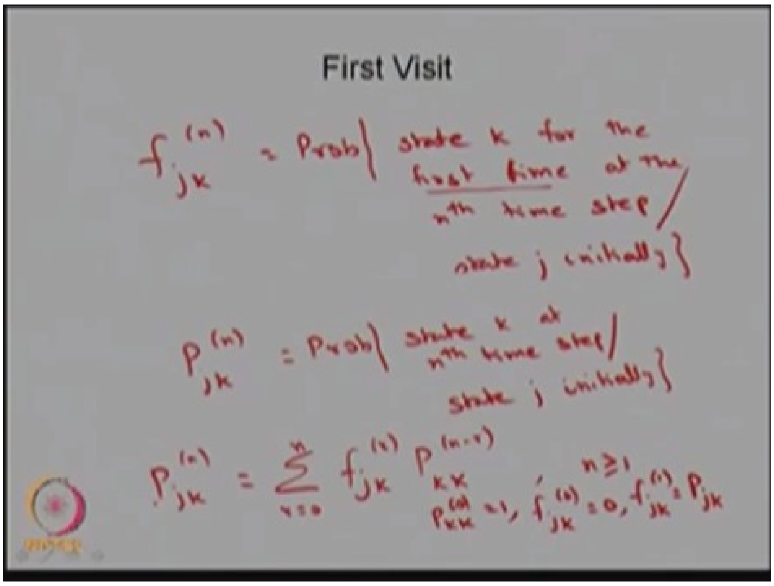
Whenever the state space cannot be partitioned into more than one closed set the proper set that means you can have only one closed set and that is same as the S. All the elements in the state space is going to form a only one closed set in that case that Markov Chain is going to be call it as a Irreducible. Irreducible means you cannot partition the state space.

If more than one proper closed subsets are possible from the state space, then that Markov Chain is going to be call it as a reducible Markov Chain. If more than one or we can able to make the partition of the state space into more than one closed set as well as the few transition states and so on that I am going to discuss later. So whenever if you are able to partition the state space then that is going to be a reducible Markov Chain.

If you are not able to partition the state space and the whole state space is going to be a only one closed proper closed subset then that Markov Chain is going to be call it as a Irreducible Markov Chain. In this case all the states belonging to that class is going to form one class and since it is going to have a only one class all the states is going to have if one state as the period something then all the other states also going to have the same period because you are not able to partition so you have only one class.

Therefore, one state has the period or some number or some integer then that same period will be for the all other states also. So the Markov Chain which are not irreducible or said to be reducible or Non Irreducible Markov Chain.

(Refer Slide Time: 08:26)



Now I am going to give the next concept called the first visit. We did not come to the classification of a state before that we are developing a few concepts using these concepts we are going to classify the states. The next concept is called a first visit what is the meaning of a first visit? I am going to define the probability mass function as the $f_{jk}^{(n)}$ that means that what is the probability that the system reaches the state k for the first time that is important.

For the first time at the nth time step given that the system starts the state j initially. This is a conditional probability mass function of a system moving from the state j to k. And system reaching the state k at nth time step for the first time that is important. So this is the first time the system reaches the state k at the nth step exactly at the nth step and this conditional probability mass function that I am going to write as the $f_{jk}^{(n)}$.

This is different from the $P_{jk}^{(n)}$. This is also conditional probability whereas this probability is defined what is the probability that the system reaches the state k at the nth time step. Given that it was in the state j initially. This is also conditional probability the only difference is the first time that means that is a possibility the system here the $P_{jk}^{(n)}$ means that is a possibility the system would have come to the state k before nth step also.

So that probability is included whereas as the $f_{jk}^{(n)}$ means that this is the only the nth step it reaches the state k. Therefore, the way I have given the first time conditional this probability and this is not necessarily the first time this is also conditional probability I can relate the f_{jk}