


Stochastic Processes
Prof. Dr. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology Delhi

Module - 1
Probability Theory Refresher
Lecture - 2
Introduction to Stochastic Processes (Contd.)

This is the continuation of module one probability theory refresher and this is the lecture two. So, the lecture one we have covered what is the motivation behind the stochastic process, then we have given few examples and, followed by that, we have explained what are all the minimum things necessary to study the stochastic process. We started with the random experiment and events, then the probability space and to create the probability space we need a sigma algebra. Then, after creating the probability space, then we have defined the conditional probability, then we discuss independent of events, then we have list out few standard discrete random variable as well as standard continuous random variable, even though we have discuss only three or four discrete and continuous random variable, there are more, but whenever the problem comes we will discuss those standard distributions when we come across those distribution. And the lecture two we are going to continue whatever we have discuss in the lecture one, basically the probability theory refresher.

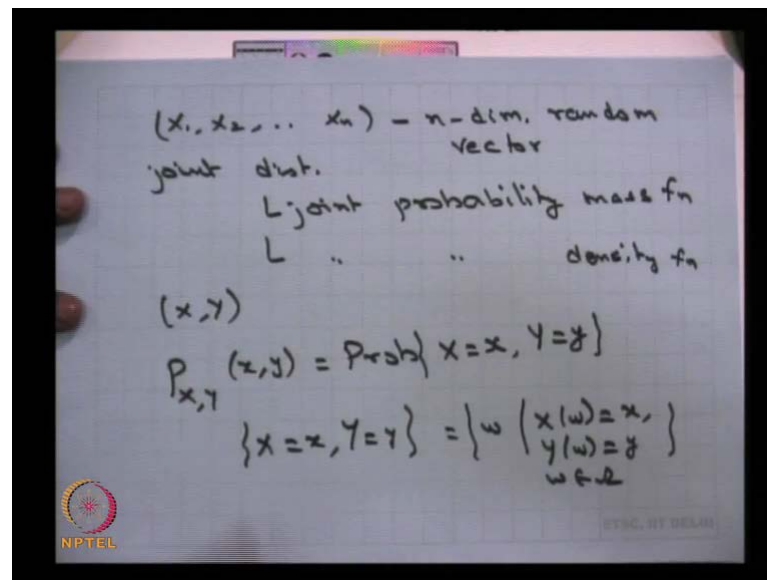
(Refer Slide Time: 01:47)

Contents	
•	Joint Distribution of Random Variables
•	Independent Random Variables
•	Covariance and Correlation Coefficient
•	Conditional Distribution
•	Conditional Expectation
•	Generating Functions
•	Convergence of Sequence of Random Variables
•	Law of Large Numbers
•	Central Limit Theorem



In these we are going to give a brief about what is joint distribution and if the random variables are independent what is the behavior of a joint distribution and so on. Then we are going to discuss covariance and correlation coefficients, after that we are going to discuss the conditional distribution, then followed conditional expectation also and we are going to list out a few generating functions, probability generating function, moment generating function and also the characteristic function. Then at the end of the probability theory part we are going to discuss how the sequence of random variable converges to some random variable, and for that we are going to discuss a law of large numbers also, and at the end of the lecture two, we are going to complete with the central limit theorem.

(Refer Slide Time: 02:51)



So, let me start with the joint distribution of random variable. So, suppose you have a random variables X_1, X_2, \dots, X_n , we say this random variable is a n dimensional random vector - means each random variable X_1, X_2, X_n are the random variables, either it could be a discrete random variable or continuous random variable and you are going to make it as a together in the vector form and each one is going to be a random variable then it is called a n dimensional random vector. Once you have a together as a vector form, then we can go for giving the joint distribution. So, the joint distribution we can discuss in two ways, either it is a joint probability mass function or we can define as a joint probability density function.

Suppose you take an example of two dimensional random variable X comma Y , if both the random variables are discrete, random variable X is discrete as well as the random variable Y is discrete, then you can define what is the joint probability mass function of this two dimensional discrete random variable as probability of x comma y . Here the small x small y are the variables, and this X comma Y denotes the two dimensional random variable, this is nothing but, what is the probability that X takes the value, random variable X takes the values small x and the random variable Y takes the value small y and based on the possible values of x and y you have the probability of this; that means, you land up creating what is the event which corresponding to X equal to x and Y equal to small y . That means if you not getting any possible outcomes, that gives some possible values of x comma y , then it may be the empty set. Otherwise, you land with the different possible possible we can collect the, that means, X is equal to x and Y is equal to y . For all possible values of x comma y , you may relate with what is the event in which X of w gives the value x as well as Y of w gives the value y , where w is belonging to ω . That means you collected a possible outcomes w , such that it satisfies both the conditions, where w is belonging to ω ; that means, this is going to be the event therefore, this is the probability of event and you know by using the axiomatic definition, the probability of event is always greater than or equal to zero, and the probability of ω is equal to one. And, if you take mutual exclusive events then the probability of union of events is going to be the summation of probability.

Therefore, this is the way you can define, when the random variables x and y both are discrete type then, you can give the joint probability mass functions. Therefore, here this is the joint probability mass function and this satisfies all the values are always greater than or equal to zero, for any x and y . And, if we make the summation over x as well as summation over y , then that is going to be one. Therefore, it satisfies the property of always greater than or equal to zero for any x and y , and summation, double summation over X and Y is equal to one. Therefore, this is going to be the joint probability mass function corresponding to the discrete type random variable.

(Refer Slide Time: 06:34)

Handwritten notes on a whiteboard:

(x, y) - 2 dim. cont. type r.v.

$f_{x,y}(x, y) =$

$F_{x,y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{x,y}(r, s) dr ds$

joint pdf

Ex. 1. (x, y)

$P_{x,y}(i, j) = \frac{1}{2^{i+j}}, i=1, 2, \dots, j=1, 2, \dots$

$P_x(i) = \sum_j \frac{1}{2^{i+j}} = \frac{1}{2^i}$

NPTel logo is visible in the bottom left corner of the whiteboard image.

Suppose the random variable x and y are, each random variable is a continuous type random variable therefore, you land up with two dimensional continuous type random variable or random vector. In that case, you can define what is the joint probability density function is of the form f of X comma Y small x and small y , that is going to be what is, so, you can have a joint probability density function and you can relate with this joint probability density function with the c d f by, what is the c d f of the random variable, that is nothing, but what is the integration from minus infinity to x and what is the integration from minus infinity to y of the integrant is going to be r comma s , where r is with respect to x that is dr and this is ds .

That means, since you have a continuous type random variable therefore, you may land up with the continuous function with the two variables x comma y , by using the fundamental theorem of algebra you can always land up a unique integrant and that is going to be the density function for this two dimensional continuous random variable. And you can able to write the left hand side continuous function in x and y can be written in the form of integration from minus infinity to x and minus infinity to y of this integrant and ds , where this function is going to be call it as a joint probability density function for the random variable x comma y .

So, there is a relation between a joint probability density function with the with the c d f you can I can always able I can give a few examples. the example one suppose you have

if both the random variables x and y are of a discrete type, then I can give a simple example of i and j , that is going to be one divided by two to the power $i + j$ for i is belonging to one two and so on, and j can take the value one two and so on and this is going to be the joint probability, joint probability mass function for two dimensional discrete type random variable. If you make a summation over x and summation over y then that is going to be one. And you can get by summing at over only j , you will get the probability mass function for the random variable x . Similarly, if you make the summation over i with the joint probability mass function, then you may get the probability mass function for the random variable y .

(Refer Slide Time: 09:36)

$$P_y(j) = \sum_i P_{x,y}(i,j)$$

$$(x_1, x_2, \dots, x_r)$$

$$P_{x_1}(x) = \sum \sum \sum \sum P_{x_1, x_2, \dots, x_r}(\dots)$$

$$f_x(x) = \int_{-\infty}^{\infty} f_{x,y}(x,y) dy$$

$$f_y(y) = \int_{-\infty}^{\infty} f_{x,y}(x,y) dx$$

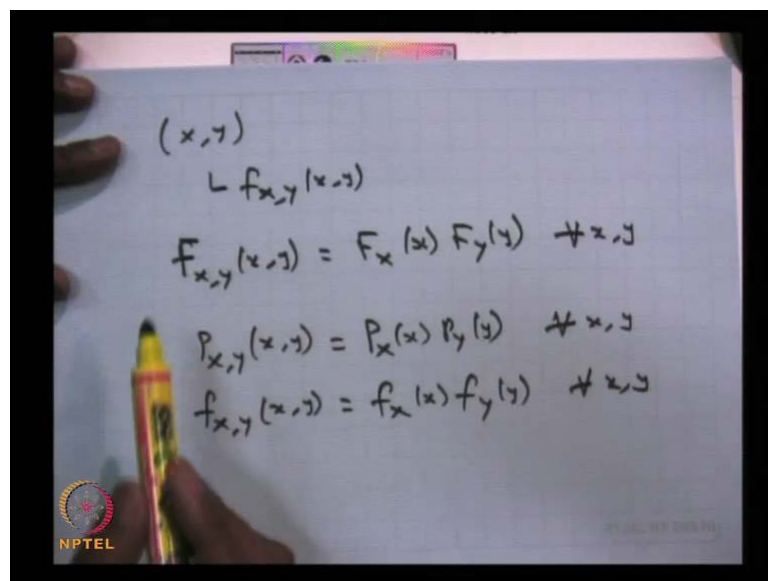
That means, from the joint probability density function by the joint probability mass function by summing over the other random variable you can get the distribution of this single random variable.

Suppose, if you have a n dimensional random variable, suppose you for n is equal to five you have a five dimensional random variable. Suppose all the random variables are of the discrete type, then you may have a, what is the you may have the joint probability mass function of this five dimensional discrete random variable, and by summing over each random variable, and if you want to find out what is the probability mass function for the one random variable, you can always get make the summation over other random variables with respect to the other random variables you can get what is the probability

mass function for by summing into over the other variables you will get the marginal distribution of the random variable X_1 similarly you can get the marginal distribution of x_2 or any other random variable.

Similarly, suppose you have a joint probability density function, from that you can get you can get the marginal distribution of one random variable by integrating with the other random variable. By integrating with the other random variable, so this is a joint probability density function. From the joint probability density function by integrating with the other random variable, you will get the probability probability density function for this random variable y . Similarly, you can get the probability density function for y is... So, you can by integrating with respect to r so, that means, you are just finding out the marginal distribution of the random variable Y and here you are finding the marginal distribution of X .

(Refer Slide Time: 12:09)



Handwritten notes on a whiteboard showing formulas for joint and marginal distributions of independent random variables:

$$(x, y)$$

$$\hookrightarrow f_{x,y}(x, y)$$

$$F_{x,y}(x, y) = F_x(x) F_y(y) \quad \forall x, y$$

$$P_{x,y}(x, y) = P_x(x) P_y(y) \quad \forall x, y$$

$$f_{x,y}(x, y) = f_x(x) f_y(y) \quad \forall x, y$$

The NPTEL logo is visible in the bottom left corner of the whiteboard image.

Now, I am going to discuss what is the meaning of independent random variable. Suppose you have a two random variables x and y , and you know what is the joint probability density function or joint probability mass function based on the random variable, both are discrete or continuous, then if both the random variables are independent, then the c d f of this random variable random vector is same as the product of c d f's of individual random variable, whether it is a discrete random variable or continuous random variable and this is valid for all x comma y . That means, if you have

a two random variables and this satisfied for all x, y ; that means, the joint c d f is same as the product of c d f, this is basically if and only if condition, if this condition is satisfied, then both the random variables are call it as a independent random variable. So, this suppose this random variables are both are discrete, then you can come down from the c d f into the joint probability mass function.

The joint probability mass function you can write it as the product of individual probability mass function for all x, y . If both the random variables are discrete, then you have a joint probability density function, the same joint probability density function will be the product of individual probability density function; that means, based on the random variable is discrete or continuous, you can cross check whether this property is satisfied. So, if this property is satisfied then you can conclude the random variables are independent. Similarly if the random variables are independent, then this property is going to be satisfied.

So, whether it is a discrete or continuous you can always check in the c d f level also. If the c d f, joint c d f and ma individual c d f has satisfies this property, then you can conclude both the random variables are going to be independent random variable, and this logic can be extended for the any n random variables. So, instead of two random variables you can go for having a n random variables. Then finding out what is the joint c d f, if the joint c d f of n dimensional random variable is going to be the product of individual random variable, then you can conclude both all n random variables are mutually independent random variables.

Now, we are moving into the next concept. There are some moments we can find out from the random variable. The way you are computing, suppose you have a random variable x , you can able to find out the expectation of x if it exists; that means, if the random variable x is there, you can always write expectation of x is from minus infinity to infinity x times d of F of x where capital F is the c d f of the random variable.

So, whether the random variable is a discrete or continuous or mixed type, if this integration is going to be exists, then you can able to give expectation is equal to this much. If the integration is does not converges or that means, if the integration diverges, then you cannot go for writing expectation of x .

(Refer Slide Time: 14:40)

The image shows a piece of paper with handwritten mathematical formulas. At the top, the formula for the expectation of a continuous random variable X is given as $E(X) = \int_{-\infty}^{\infty} x dF_X(x) = \int_{-\infty}^{\infty} x f_X(x) dx$. Below this, it states 'provided $E(|X|) < \infty$ '. Then, it lists three properties: 1. If $X \geq 0$, then $E(X) \geq 0$. 2. Linearity: $E(aX + b) = aE(X) + b$. 3. Monotonicity: If $X \geq Y$, then $E(X) \geq E(Y)$. An NPTEL logo is visible in the bottom left corner of the paper.

$$E(X) = \int_{-\infty}^{\infty} x dF_X(x) = \int_{-\infty}^{\infty} x f_X(x) dx$$

provided $E(|X|) < \infty$

$$X \geq 0 \quad E(X) \geq 0$$
$$E(aX + b) = aE(X) + b$$
$$X \geq Y \quad E(X) \geq E(Y)$$

Suppose the random variable is a continuous random variable, then the c d f is going to be the continuous function therefore, this is same as this is same as minus infinity to infinity x times f of x dx if the random variable is a continuous random variable. In that case also, we have to cross check whether this integration is going to be, see provided provided it says, expectation of absolute x is converges. This is because absolute convergence implies convergence. That means, whenever you replaced x by absolute of x and you find out if this provided condition is satisfied, then without absolute, whatever the quantity you are going to get in the if it is in the continuous random variable the integration is converge whatever the value you are going to get that is going to be the expectation of the random variable.

So, the expectation of the random variable has expectation as a few property: this is going to be a always a constant, this is not a random variable and the expectation of x , if the random variable is greater than or equal to 0, then the expectation of x is always greater than or equal to 0, and the expectation of x has the linear property, that is same as a times expectation of x plus constant expectation of a constant is equal to 0. If we have two random variables, then the expectation of x is greater than or equal to expectation of y .

(Refer Slide Time: 16:53)

Handwritten notes on a whiteboard:

$$(x, y)$$
$$X \quad \text{var}(x) = E(x^2) - (E(x))^2$$
$$\text{cov}(x, y) = E(xy) - E(x)E(y)$$
$$E(xy) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{x,y}(x, y) dx dy$$

if x & y are Independent r.v.s.

$$E(xy) = E(x)E(y)$$
$$\Rightarrow \text{cov}(x, y) = 0$$

NPTEL logo is visible in the bottom left corner.

So now we are going to discuss, since we have a more random variables, we are going to discuss what is the what is other than expectation we can go for finding out the variance of the random variable also. Variance is nothing but, the second order moment, that is E of x square minus E of x whole square. So, here also as long as E of x square that means, expectation in absolute x if that is converges, then you can able to get expectation of E of x square and once you have a second order moment is exist; obviously, the all the previous order moment exists, but that does not imply the further moment exist. So, now I am going to define what is the covariance of x comma y . So, covariance of x comma y is nothing, but expectation of x into y minus expectation of x into expectation of x provided the expectation exists.

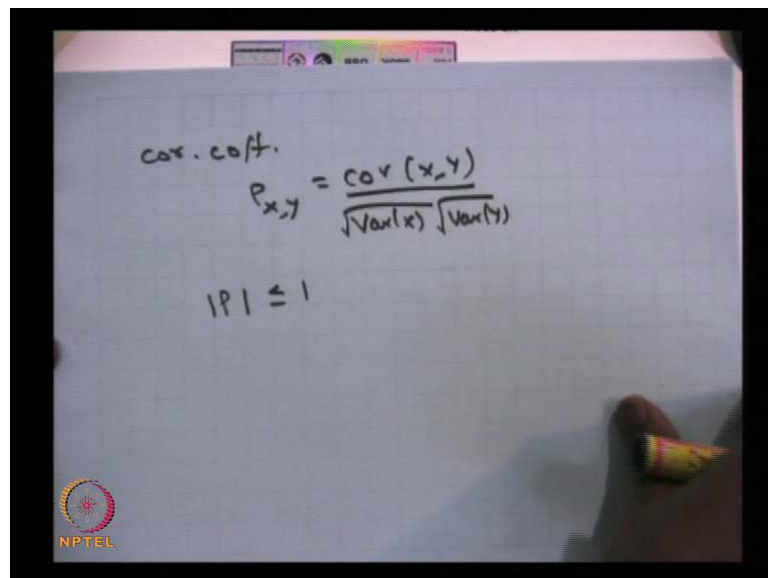
So, here it is the expectation of x into y , that means, you have to find out what is expectation of x into y . Based on the random variable is the discrete or continuous, you can able to use functions of random variable method and getting the expectation and note that, even you don't know the distribution of the x into y , you can always find out the expectation of x into y .

Let me give a one situation, if both the random variables are a continuous then, the expectation of x into y is going to be x into y and the joint probability density function of f of x , f of y ; that means, this is going to be the value of x x y and what is the joint distribution of this; that means, you are not finding out what is a distribution of x y , but

you can still you can find out the expectation of $x y$ by possible values and corresponding joint distribution you can get the expectation, and here also provided the absolute sense exists, then without absolute sense it is going to be E of x into y . Suppose x and y are independent random variable independent independent random variables then the expectation of x into y , the way I have given the situation with the both the random variables are continuous, this integration will be splitted into the two parts, such way that the f of x comma y , this is going to be sorry this minus infinity to infinity.

So, this is going to be the product of individual probability mass function therefore, this is going to be integration will be splitted into two single integration, minus infinity to infinity x times f of x and minus infinity to infinity y times f of y dy . Therefore that is nothing, but the expectation of x into expectation of y . That means, if two random variables are independent, then implies the covariance of the two random variables is going to be 0, but the covariance of x comma y equal to 0, that does not imply the random variables are independent. So, this is going to be the, not if and only if the random variables are independent then you will come to the conclusion of covariance of x comma y equal to 0, not the converse.

.(Refer Slide Time: 20:23)



Handwritten notes on a whiteboard:

$$\text{cor. coeff.}$$

$$\rho_{x,y} = \frac{\text{cov}(x,y)}{\sqrt{\text{Var}(x)} \sqrt{\text{Var}(y)}}$$

$$|\rho| \leq 1$$

The NPTEL logo is visible in the bottom left corner of the whiteboard image.

Now we are going to define the another measure, that is a correlation coefficient. That is nothing but, the correlation coefficient is nothing but with the letter rho rho of x comma y ; that means, I am trying to find out what is the correlation coefficient between the

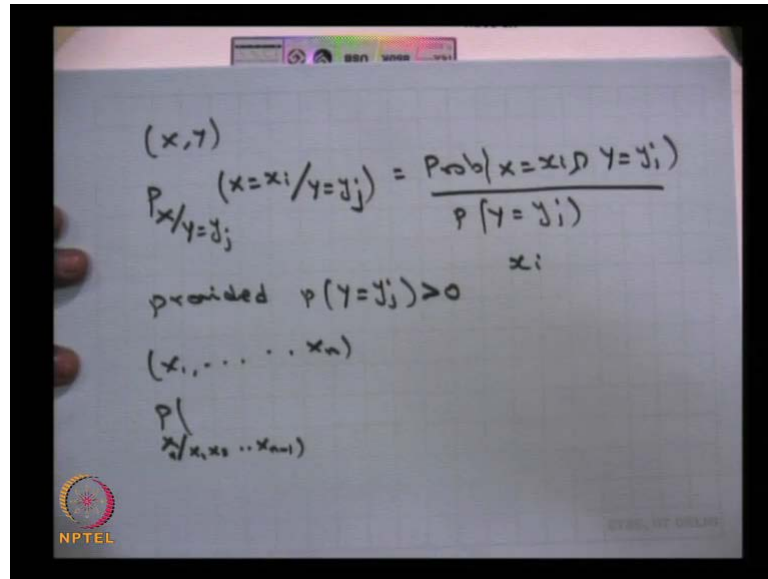
random variables x and y , that is nothing, but the covariance of x and y divided by the square root of variance of x into square root of variance of y .

That means to have an existence of the covariance correlation coefficient, you should have a random variable that should have at least a second order moment. So, unless otherwise the second order moment does not exist, you cannot find out the correlation coefficient between these two random variables x and y , because you are using the variance as well as the covariance. Therefore, if the random variables are independent, then obviously the ρ equal to 0, because the numerator is going to be 0.

And since you are dividing the covariance divided by the square root of variance in x as well as y , this quantity in absolute is always less than equal to 1 or the ρ lies between minus 1 to 1, and the way the correlation coefficient value lies between -1 to 1, that conclude it is a positive correlated and the values lies between -1 to 0, it gives a negatively correlated, and if the value is positive one or minus one then you can conclude the random variables x and y are linearly correlated, based on the value is positive side or the negative side then, you can conclude it is positively correlated or negatively correlated.

So, other than the value minus one and one, you cannot conclude what is the relation between the random variable. Only if it is one and minus one, then you can conclude the random variables are correlated in the linear way. Now, I am going to discuss conditional distribution, because these are all the concepts are needed when you are start defining some of the properties in the stochastic process. So, therefore, I am just discussing what is conditional distribution.

(Refer Slide Time: 22:32)



(x, y)
 $P_{x/y=y_j}(x=x_i/y=y_j) = \frac{P(x=x_i \text{ and } y=y_j)}{P(y=y_j)}$
 $\text{provided } P(y=y_j) > 0$
 (x_1, \dots, x_n)
 $P(x_1, x_2, \dots, x_n)$

Suppose you have a two dimensional random variable x comma y , you can define, suppose I make the one more assumption both are discrete type random variable then I can define, what is a conditional distribution of the random variable x given that y takes some value y_j , and here x takes a value x_i given that y takes a value y_j , that is nothing but, what is the I can compute by finding what is the probability that x takes the value x_i intersection with the y takes the value y_j divided by what is the probability that the y takes a value y_j and here the running index is for all x_i 's and this is for fixed y_j .

Therefore the provided condition, provided the probability of y takes a value y_j has to be strictly greater than 0. That means, you are making a the way you made a conditional conditional probability over the event, the same way we are making this is going to be the event y is equal to y_j . So, as long as the probability of the event corresponding to y is equal to y_j is strictly greater than 0; that means, it is not a impossible event with the probability 0. It is the event which has the positive probability. If this happens already, then what is the probability of the random variable, x takes the value of x_i .

That means still our interest is to find out the distribution of x only, the random variable x with the provided or given situation that the other random variable y takes the value y_j . That means, from the omega, you land up having a one reduced sample space that corresponding to y is equal to y_j and from the reduced sample space, you are trying to find out what is the distribution of the random variable x for all possible values of x_i .

So, this we call it as a conditional distribution of x , given the other random variable and this logic can be extended for more random variables. That means, if you have n discrete random variables, then you can always define, suppose you have a x_1, x_2, \dots, x_n and suppose all are discrete random variable, then you can always define what is the probability distribution of x given, what is the distribution of x_n given you know the distribution of x_1, x_2 , till x_{n-1} . That means still it is one dimensional random variable of x_n given that, already the random variable x_1 to x_{n-1} takes some particular value. Similarly you can go for what is the joint distribution of a few random variables, given that all other random variables already taken some value.

(Refer Slide Time: 25:44)

Handwritten notes on a whiteboard:

$$(x, y) - 2 \text{ dim. cont. type.}$$

$$f_{x/y=y} = \frac{f_{x,y}(x,y)}{f_y(y)}$$

provided $f_y(y) > 0$

$x/y=y$ $x/y - r.v$
 \hookrightarrow dist (CDF)

NPTEL logo is visible in the bottom left corner.

Now, I can go for defining the same way I can go for defining what is the conditional distribution of two dimensional continuous type random variable. That means, you can define what is the probability density function of a random variable x , given that y takes the value y ; that means, that is x given y , this is nothing but, what is the joint probability density function of x with y and divided by what is the marginal distribution of y and here also the provided condition, f of y is strictly greater than 0. That means where ever, there is a density which is greater than 0 and with that given situation, you can find out the distribution of the random variable x with the given Y takes the value small y . That is nothing but, what is the ratio in which the joint distribution with the marginal distribution.

Once you know the conditional distribution, this is also sort of another random variable this; that means, X given Y takes the value y , so, that I can use it as the word X small y , this is also is a random variable. Therefore, you can find out what is the distribution therefore, this distribution is called a conditional distribution and you can find out what is the c d f of that random variable. So, the way you find out the c d f of the any discrete random variable by summing what is the mass or by integrating the probability density function till that point, you will get the c d f of this conditional distribution. Before I go to the. So, one more thing that is the conditional expectation.

(Refer Slide Time: 27:35)

Handwritten notes on a whiteboard:

$$X/y - \text{r.v.}$$

$$E(X/y) = \int_{-\infty}^{\infty} x f_{X/y}(x/y) dx$$

\hookrightarrow is a fn of y
 \hookrightarrow is a r.v.

$$E(E(X/y)) = E(X)$$

X & Y are indep. r.v.s.
 $E(X/y) = E(X)$

NPTEL logo is visible in the bottom left corner of the whiteboard.

So, since I said X given y is a random variable, I can go for finding out what is the expectation of X given y . So, this is called the conditional expectation. That means, the X given y is the still it is a random variable, but it is a conditional distribution. Therefore, finding out the expectation for that, that is called the conditional expectation.

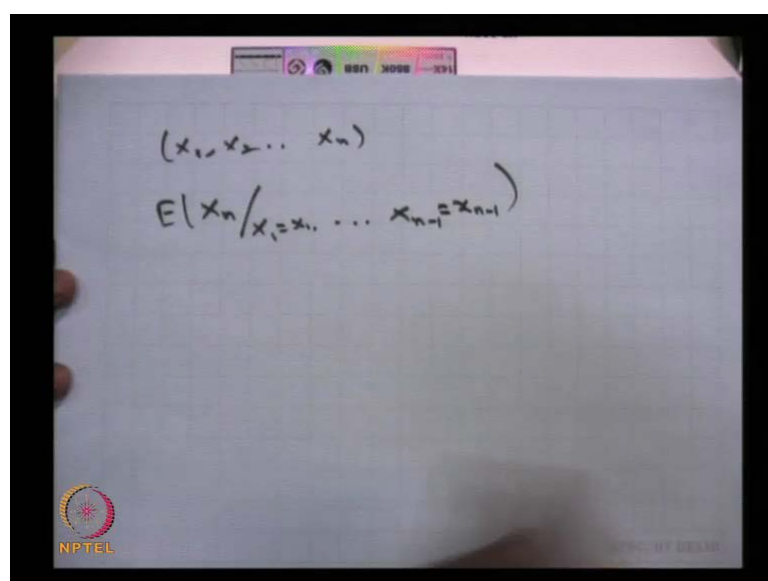
Suppose, I treat both the random variables are continuous case, then the conditional expectation is nothing but, minus infinity to infinity x times $f_{X/y}$ of x comma x given y integration with respect to x . That means, by treating x and y are continuous random variable, I can able to define the conditional expectation is this, provided this expectation exists. That means in absolute sense, if this integration converges then without absolute, whatever the value you are going to get that is going to be the conditional expectation of the random variable. And if you note that, since the y is the y

also can take any value therefore, this is a function of y . Not only, this is the function of y , the expectation conditional expectation is a random variable also. That means x given y is a random variable. The expectation of x given y is a function of y and vary y is a random variable, it takes a different values small y therefore, expectation of x given y is also a random variable. That means, you can able to find out what is the expectation of expectation x given y .

If you compute that, it is going to be expectation of X . This is a very important property, in which you are relating two different random variable with the conditional sense and if you are trying to find out the expectation of that, that is going to be the original expectation. That means the usage of this concept, instead of finding out the expectation of one random variable, if it is easy to find out the conditional expectation, then you find out the expectation of conditional expectation that is same as the original expectation.

Suppose, you have two random variables are independent random variables, then you know that there is no the there is no dependency over the random variable x and y , therefore, the expectation of x given y that is same as the expectation of x . So, this can be validated here also, because this expectation of x given y is going to be expectation of x , as in the expectation of x is a constant and the expectation of a constant is a constant that is same as this same constant.

(Refer Slide Time: 31:03)



$$(x_1, x_2, \dots, x_n)$$

$$E(x_n / x_1 = x_1, \dots, x_{n-1} = x_{n-1})$$

So, that can be cross checked. So, here I have given expectation of x given y in the integration form. If both the random variables are continuous, then accordingly you have to use initially the joint probability mass function, then conditional probability mass function to get that conditional expectation. And this conditional expectation is very much important to give one important property called martingale property in the stochastic process, in which you are going to discuss not only two random variables, you are going to discuss you have a n random variables and you can try to find out what is the conditional expectation of one random variable given that the other random variable takes some values already.

. So, so there we are going to find out what is the conditional expectation of n dimensional random variable with the given that remaining n minus one random variable takes already some value. So, so here I have given only with the two random variables how to compute the conditional expectation, but as such you are going to find out that conditional expectation of n random variables with n minus 1 random variables, already taken some value.

(Refer Slide Time: 32:06)

1) (x, y)
 \hookrightarrow discrete
 $P_{x,y}(x, y) = \frac{1}{2^{x+y}}$
 $x = 1, 2, \dots$
 $y = 1, 2, \dots$

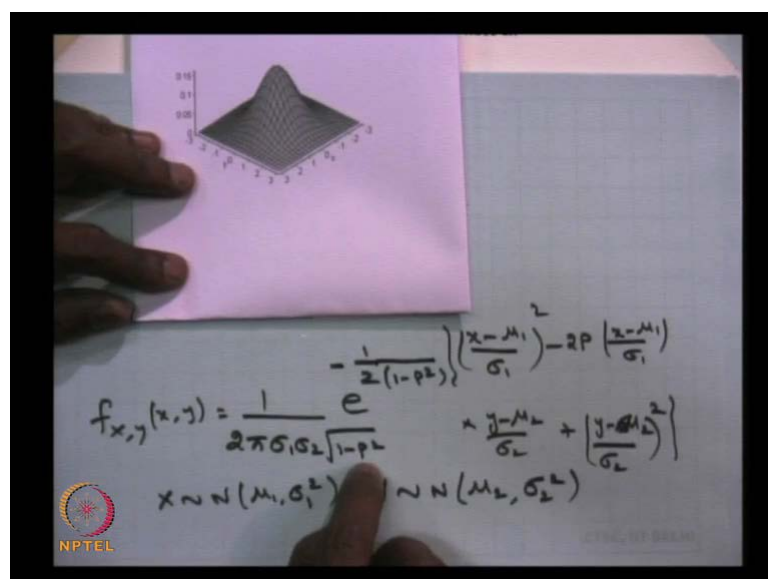
2) — cont
 $f_{x,y}(x, y) = \lambda \mu e^{-\lambda x - \mu y}$
 $x > 0$
 $y > 0$
 $\lambda > 0, \mu > 0$
 $f_x(x) = \lambda e^{-\lambda x}$
 $f_y(y) = \mu e^{-\mu y}$

So before I go to the another concept, let me just give few examples in which I have already given if both the random variables are of discrete type, I have given a example of joint probability mass function as 1 divided by 2 power x plus y and x takes a value 1, 2 and so on and y takes a value 1, 2 and so on. This is the joint probability mass function

example. And suppose you have a random variables are of the continuous type, then I can give one simple example of the joint probability density function of two dimensional continuous type random variable as joint probability density function $\lambda \times \mu e^{-\lambda x - \mu y}$, where x can take the value greater than 0, y can take the value greater than 0 and λ is strictly greater than 0, as well as μ greater than 0. So, this is going to be the joint probability density function of a two dimensional continuous type random variable. You can cross check this is going to be joint, because it is going to be always take greater than or equal to 0 values for all x and y and if you make a double integration over minus infinity to infinity over x and y then that is going to be 1. And you can verify the other one.

If you find out the marginal distribution of this random variable, you may land up the marginal distribution of this random variable is going to be $\lambda e^{-\lambda x}$. And similarly if you find out the marginal distribution of the same one, you will get $\mu e^{-\mu y}$ and if you cross check the product is going to be the joint probability density function, then you can conclude, this both the random variables are independent random variable. Similarly, you can find out what is the marginal distribution of the random variable x , similarly marginal distribution of y , if you cross check the similar independent similar property of independent, then that is satisfied. Therefore, you can conclude here the random variables x and y both are discrete as well as both are independent random variable also.

(Refer Slide Time: 34:52)



So, the advantage with the independent random variable, always you can find out from the joint you can find out marginals, but if you have a marginals you cannot find out the joint unless otherwise they are the independent random variable. Therefore the independent random variable makes easier to find out the joint distribution with the provided marginal distribution. And here is the one simple example of here is the simple example of bivariate normal distribution, in which the both the random variables x and y are normally distributed.

Therefore, the together joint distribution is going to be of the form, let me write the joint probability density function of two dimensional normal distribution random variable as $\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$ into e power minus half times of $\frac{1}{1-\rho^2}$ multiplied by $x - \mu_1$ by σ_1 whole square minus 2ρ times $x - \mu_1$ by σ_1 times $y - \mu_2$ by σ_2 plus $(y - \mu_2)^2$ by σ_2^2 whole square. So, here, if you find out the marginal distribution of the random variable x and the marginal distribution of y , you can conclude x is going to be normally distributed with the mean μ_1 and the variance σ_1^2 and similarly you can come to the conclusion y is also normally distributed with the mean μ_2 and the variance σ_2^2 .

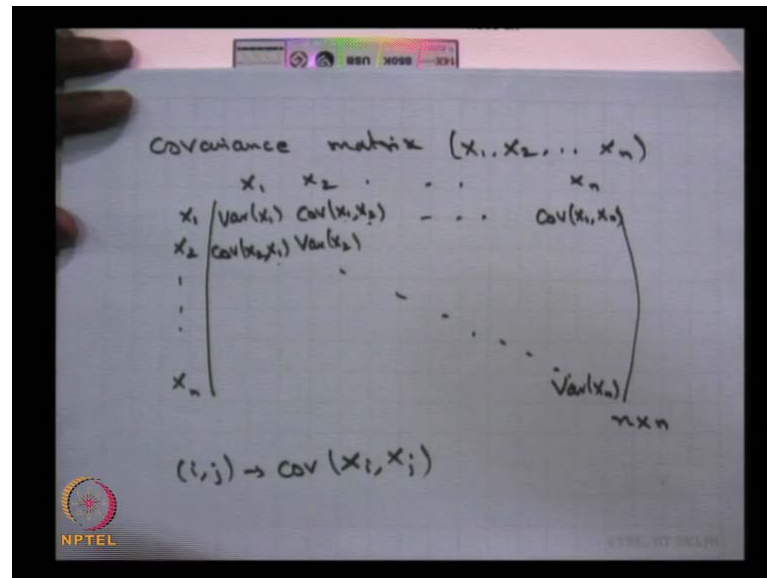
That means, if you make the plot for the joint probability density function, that will be of this shape one is the x and one is y , and this is going to be the joint probability density function for a fixed values of μ_1 and μ_2 and σ_1 and σ_2 . And, this is going to be the joint probability density function and here ρ is nothing, but the correlation coefficient; that means, what is the way the random variable x and y are correlated that comes into the picture when you are giving the joint probability density function of this random variable and they are not independent random variable unless otherwise the ρ is going to be 0.

So, if the ρ is going to 0, then it gets simplified and you can able to verify the joint probability density function will be the product of two probability density function and each one is going to be a probability density function of a normal distribution with the mean μ_1 and the variance σ_1^2 and μ_2 and σ_2^2 .

So, this bivariate normal distribution is very important one when you discuss the multinomial normal distribution. So, only we can able to give the joint probability density

function of the bivariate. So, the multivariate you can able to visualize how the joint probability density function will look like and what is the way the other factors will come into the picture.

(Refer Slide Time: 38:21)



The image shows a handwritten definition of a covariance matrix on a piece of paper. At the top, it says "Covariance matrix (x_1, x_2, \dots, x_n) ". Below this, a matrix is drawn with x_1, x_2, \dots, x_n as column headers and x_1, x_2, \dots, x_n as row headers. The diagonal elements are labeled $Var(x_1), Cov(x_1, x_2), \dots, Cov(x_1, x_n)$ for the first row, and $Cov(x_2, x_1), Var(x_2)$ for the second row, with ellipses indicating the pattern. The bottom-right element is labeled $Var(x_n)$. A bracket on the right side of the matrix indicates it is an $n \times n$ matrix. Below the matrix, the notation $(i, j) \rightarrow Cov(x_i, x_j)$ is written. In the bottom left corner, there is a small NPTEL logo.

So, other than correlation and other than covariance correlation and correlation coefficient, we need the other called covariance matrix also, because in the stochastic process we are going to consider a n dimensional random variable as well as the sequence of random variables, so you should know, how to define the covariance matrix of n dimensional random variable. That means, suppose if you have a n random variables x_1 to x_n , then you can define the covariance matrix as, you just make a row wise x_1 to x_n and column also you make a x_1 to x_n , now you can fill up, this is going to be n cross n matrix, in which each entity is going to be covariance of, so that means, the matrix entity of i comma j is nothing but, what is the covariance of that random variable x_i with x_j .

You know that the way I have given the definition, covariance of x_i and x_j , if i and j are same, then that is nothing but E of x square minus E of x whole square. Therefore, that is nothing but the variance of that random variable therefore, this is going to be variance of x_1 and this is going to be the variance of x_2 . Therefore all the diagonal elements are going to be variants of x_i 's, where as other than the diagonal elements you can fill it up, this is going to be a covariance of x_1 with x_2 and the last like that, the last element will

be covariance of x_1 with x_n . Similarly, second row first column will be covariance of x_2 with x_1 . And you can use the other property, the covariance of x_i with x_j is same as covariance of x_j with x_i also, because you are trying to find out expectation of $(x_i - \mu_i)(x_j - \mu_j)$ minus expectation of $(x_j - \mu_j)(x_i - \mu_i)$. Therefore, the both the covariance of x_2 with x_1 is same as x_1 with x_2 .

So, it is going to be a, whatever the value you are going to get, it is going to be the symmetric matrix and the all the diagonal elements are going to be the variance. So, the way I have given the two dimensional normal distribution that is a bivariate normal, suppose you have a n dimensional random vector in which each random variable is a normal distribution, then you need what is the covariance matrix for that. Then only, you can find out what is the, then only you can able to write what is the joint probability density function of n dimensional random variable. Followed by this, now we are going to discuss few generating functions.

(Refer Slide Time: 41:13)

1. prob. generating fn.
 X - discrete
 $x_i = 0, 1, 2, \dots$

$$G_X(z) = \sum_i z^i P(X=i)$$

$$G_X(1) = 1$$

$X \sim B(n, p)$	$G_X(z) = (1-p + pz)^n$
$X \sim P(\lambda)$	$G_X(z) = e^{\lambda(z-1)}$

NPTEL

So, the first one is called probability generating function. So, this is possible only with a random variable is a discrete random variable and the possible values of x 's has to takes 0 or 1 or 2 like that; that means, if the possible values of the random variable x takes the value only 0, 1, 2 and so on, then you can able to define what is the probability generating function for the random variable x as with the notation $G_X(z)$, that is a probability generating function for the random variable x as a function of z , that is

nothing but summation z^i and what is the probability x takes the value i , for all possible values of i . That means, if the discrete random variable takes the only countably finite value, then the probability generating function is a polynomial. If the discrete random variable takes a countably infinite values, then it is going to be the series.

So, this series is going to be always converges and you can able to find out what is the value at 1 that is going to be 1. And since it is going to be z^i , by differentiating you can get, there is a easy formula or there is a relation between the moment of order n with the probability function in the derivative of n th derivative and substituting z is equal to 1.

And you suppose x is going to be a binomial distribution, with the parameters n and p then you can find out what is a probability generating function for the random variable x . That is going to be $1 - p + p \times z^n$, because ah the binomial distribution has the possible values are going to be 0 to n therefore, you will get the polynomial of degree n .

(Refer Slide Time: 43:57)

2. Moment Generating function

$$M_X(t) = E(e^{xt})$$

provided expectation exist

$$e^{xt} = 1 + \frac{xt}{1!} + \frac{(xt)^2}{2!} + \dots$$

$$M_X(0) = 1$$

$X \sim B(n, p)$	$M_X(t) = (1 - p + pe^t)^n$
$X \sim P(\lambda)$	$M_X(t) = e^{\lambda(e^t - 1)}$
$X \sim N(\mu, \sigma^2)$	$M_X(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$

NPTTEL

Suppose x is going to be a Poisson distribution with the parameter λ , because this is also a discrete random variable and the possible values are going to be countably infinite, where as, here the possible values are going to be countably finite. So, here also you can find out what is the probability mass function sorry what is the probability generating function for random variable x and that is going to be $e^{\lambda z}$

minus 1. So, like that you can find out probability generating function for only of discrete type random variable with the possible values has to be a countably finite or countably infinite with the 0, 1, 2 and so on.

The next generating function which I am going to explain that is, moment generating function. The way we use the word moment generating function, it will use the moments of all order n ; that means, it uses the first order moment, second order moment and third order moment and you can define the moment generating function for the random variable x as the function of t . That is nothing but expectation of e^{xt} provided the expectation exists. That is very important. That means, since I am using the expectation of a function of a random variable and that too this function is e^{xt} , you can expand e^{xt} as $1 + xt + \frac{(xt)^2}{2!} + \frac{(xt)^3}{3!} + \dots$ so on.

Therefore, that is nothing but the moment generating function for the random variable x is nothing but expectation of this expansion. That means, expectation of $1 + xt + \frac{(xt)^2}{2!} + \dots$ That means, if the moment of all order n exist, then you can able to get what is the moment generating function for the random variable x . That is, the provided condition is important as long as the right hand side expectation exist, you can able to give the moment generating function for the random variable x .

So, here also many properties are there, I am just giving one property. $M_x(0)$ is going to be 1 and there are some property which relate with the moment of order n with the derivative of a moment generating function and I can give one simple example if x is going to be binomial distribution with the parameters n and p , then the moment generating function for the random variable x that is going to be $(1 - p + pe^t)^n$.

Similarly, if x is going to be a binomial, if x is going to be a Poisson with the parameter λ , then you may get the moment generating function is going to be $e^{\lambda(e^t - 1)}$. And you can go for continuous random variable also, if x is going to be a normal distribution with the parameters μ and σ^2 , then the moment generating function is going to be $e^{t\mu + \frac{1}{2}\sigma^2 t^2}$. So, this is very important moment generating function because we are going to

use this moment generating function of normal distribution in the stochastic process part also. There is some important property over the moment generating function.

(Refer Slide Time: 46:52)

$$(X_1, \dots, X_n)$$

$$X_i \rightarrow \text{i.i.d. r.v.s} \quad i=1, 2, \dots, n$$

$$X \stackrel{d}{=} Y \quad F_X(x) = F_Y(y)$$

$$S_n = \sum_{i=1}^n X_i$$

$$M_{S_n}(t) = \prod_{i=1}^n M_{X_i}(t) = (M_X(t))^n$$

Suppose you have a n random variables and all n X_i 's random variables are i.i.d.; that means, independent identically distributed random variable; that means, the distribution of when you say the random variable x and y are identically distributed, that means the c.d.f of x and the c.d.f of y are same. For all x and y both the values are going to be same, then you can conclude the both the random variables are going to be identically distributed. So, here I am saying the n random variables are i.i.d random variable; that means, not only identical they are mutually independent also.

If this is the situation and my interest is to find out what is the m.g.f of sum of n random variables that is S_n , so the moment generating function for the random variable S_n is going to be the product of the m.g.f of individual random variable. Since, they are identical, the m.g.f also going to be identical. Therefore, this is same as you find out the m.g.f of any one random variable then make the power. So, this independent random variables having the property when you are trying to find out the m.g.f of the sum of the random variable that is same as the product of m.g.f of individual random variables.

Here there is a one more property over the m.g.f. Suppose, you find out the m.g.f of some unknown random variable and that matches with the m.g.f of any standard random variables, then you can conclude the particular unknown random variable also distributed

in the same way. That means, the way you are able to use the c d f's are same, then the corresponding random variables are identical. Same way if the m g f of two different random variables are same, then you can conclude the random variables also identically distributed.

(Refer Slide Time: 49:02)

Handwritten notes on a whiteboard:

$$\begin{aligned} & \text{3. characteristic function} \\ & \Phi_x(t) = E(e^{itx}) \\ & i = \sqrt{-1} \\ & = \int_{-\infty}^{\infty} e^{itx} dF_x(x) = \int_{-\infty}^{\infty} e^{itx} f_x(x) dx \\ & \Phi_x(-it) = M_x(t) \\ & S_n = \sum_{i=1}^n x_i \quad \Phi_{S_n}(t) = (\Phi_x(t))^n \end{aligned}$$

The whiteboard also features an NPTEL logo in the bottom left corner.

Third, we are going to consider the another generating function that is called characteristic function. This is important than the other two generating function, because the probability generating function will exist only for the discrete random variable and the moment generating function will exist only if the moments of all order n exists, where as the characteristic function exists for any random variable, whether the random variable is a discrete or the moments of all order n exist or not, immaterial of that the characteristic function exist for all the random variable.

That I am using the notation ψ suffix x as the function of t , that is going to be expectation of e power i times x t . Here the i is the complex number that is a square root of minus 1. So, this plays a very important role, such that this expectation is going to be always exist, whether the moment exists or not .Therefore, the characteristic function always exist. You can able to give the interpretation of e power, this is same as minus infinity to infinity e power i times t x d of c d f of that random variable. So that means, whether the random variable is a discrete or continuous or mixed, you are integrating with respect to the c d f of the integrant function is e power i times t x , where i is the

complex quantity. And if you find out the absolute, this absolute this is going to be using the usual complex functions, you can make out this is going to be always less than or equal to 1 in the absolute sense. Therefore, this integration is exist and this integration is nothing but the Riemann Stieltjes integration.

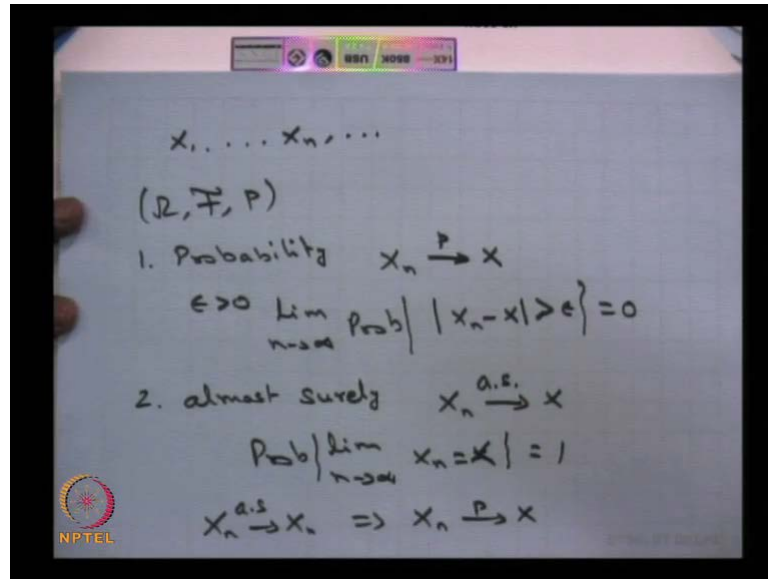
And if the function is going to be the if the random variable is a continuous then you can able to write this is same as minus infinity to infinity e^{itx} of the density function integration with respect to x ; that means, this is nothing but the Fourier transform of f .

And here we have this f is going to be the probability density function and you are integrating the probability density function along with e^{itx} and this quantity is going to be always converges, where as the moment generating function without the term i , the expectation may exist or may not exist. Therefore, the $m g f$ may exist or may not exist for some random variable.

And I can relate with the characteristic function with the $m g f$ with the form ψ_x of minus i times t , that is same as $m g f$ of the random variable t . That means, I can able to say what is the $m g f$ of the random variable x , that is same as the characteristic function of minus i times t , where i is the complex quantity. And here also the property of the summation of, suppose I am trying to find out what is a characteristic function of sum of n random variables and each all the random variables are $i i d$ random variable, then the characteristic function of s_n is same as, when x_i 's are $i i d$ random variable, then the characteristic function of each random variable power n .

And this also has the property of uniqueness; that means, if two random variables characteristic functions are same, then, you can conclude both the random variables are identically distributed. So, as a conclusion we have discuss three different functions. First one is probability generating function and the second one is a moment generating function and a third one is a characteristic function. And, we are going to use all those all those functions and all other properties of joint probability density function distribution everything we are going to use it in the at the time of stochastic process discussion.

(Refer Slide Time: 53:17)



Next we are going to discuss, what is the how to define or how we can explain the sequence of random variable converges to one random variable. Till now, we started with the one random variable, then using the function of a random variable you can always land up another random variable or from the scratch you can create another random variable, because random variable is nothing but real valid functions satisfying that one particular property inverse images also belonging to \mathcal{F} . Therefore, you can create many more or a countably infinite random variables or uncountably many random variables also over the same probability space. That means, you have a one probability space and in the one single probability space you can always create either countably infinite or uncountably many random variables and once you are able to create many random variables, now our issue is what could be the convergence of sequence of random variable.

That means, if you know the distribution of each random variable and what could be the distribution of the random variable x_n , as n tends to infinity. So, in this we are going to discuss different modes of convergence. That is the first one is called convergence in probability. That means, if I say a sequence of random variable x_n converges to the random variable some x in probability, that means if I take any epsilon greater than 0, then limit n tends to infinity of probability of absolute of x_n minus x which is greater than epsilon is 0. If this property is satisfied for any epsilon greater than 0, then I can

conclude the sequence of random variable converges to one particular random variable x in probability. That means, this is the convergence in probability sense.

That means you collected possible outcomes that find out the difference of x_n minus x which is in the absolute greater than epsilon; that means, you find out what is the what are all the possible event which is away from the length of two epsilon, you collect all possible outcomes and that possible outcomes is that probability is going to be 0. Then it is convergence in probability; that means, you are not doing the convergence in the real analysis. The way you do, you are trying to find out the event then you are finding the probability therefore, this is called the convergence in probability.

The second one, it is convergence almost surely. So, this is the second mode of convergence, this the notation is x_n converges to x a dot s dot; that means, the sequence of random variable as n tends to infinity converges to the random variable x as n tends to infinity that almost surely, provided the probability of limit n tends to infinity of x_n equal to x or x_n is equal to capital x , that is going to be 1. That means, first you are trying to find out what is the event for n tends to infinity the x_n takes the value x . That means, you are collecting the few possible outcomes that as n tends to infinity what is a event which will give x_n same as the x , then that event probability is going to be 1. If this condition is satisfied then we say it is going to be almost surely.

I can relate with the almost surely, with the if any sequence of random variable converges almost surely that implies, x_n converges to x in probability also. This is a third mode of convergence; that means, if the sequence of random variable $c d f$ converges to the $c d f$ of the random variable x , then you can say that the sequence of random variable converges to the random variable in distribution. And I can conclude, the sequence of random variable converges in almost surely implies in probability that implies in distribution, whereas, the converse is not true. And, when I categorize this into the law of large numbers as a weak law of large numbers and strong law of large numbers, if the mean of x_n that converges to μ in probability, then we say it says it is a weak law of large numbers. Similarly if the convergence in the almost surely, then we conclude this is going to satisfies the strong law of large numbers.

The final one that a central limit theorem... You have a sequence of random variable with each are i i d random variables and you know the mean and the variance and if you

define the s_n is in the form then s_n minus $n\mu$ divided by σ times square root of n , that converges to standard normal distribution in convergence in distribution.

That means whatever be the random variable you have, as long as they are i i d random variable and even these things can be relaxed, the sequence of random variable the summation will converges to the normal distribution or their mean divided by the standard deviation will converges to the standard normal distribution. With this, I complete the review of theory of probability in the two lectures. Then the next lecture onwards I will start the stochastic process.

Thank you.