Mathematical Aspects of Biomedical Electronic System Design

Professor. Chandramani Singh

Department of Electronic Systems Engineering

Indian Institute of Science, Bangalore

Week – 12

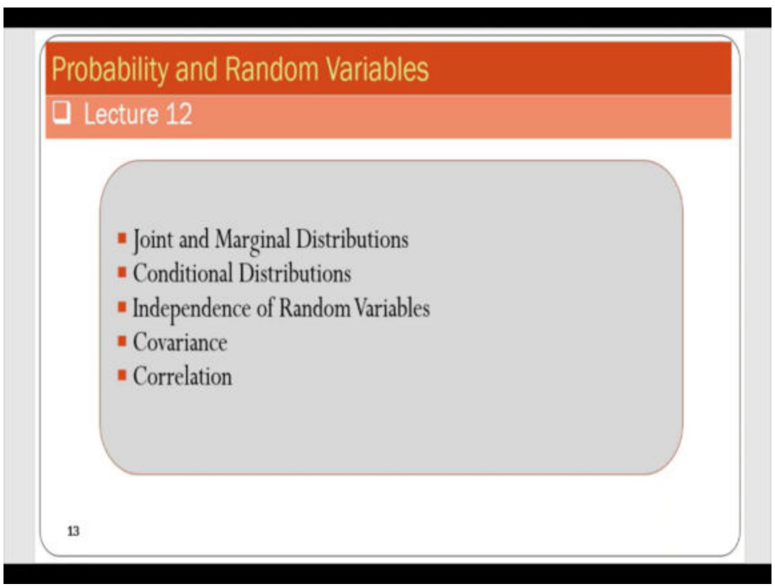
Probability Distribution and Biomedical Systems Design

Lecture – 38

Joint and Marginal Probability Distribution

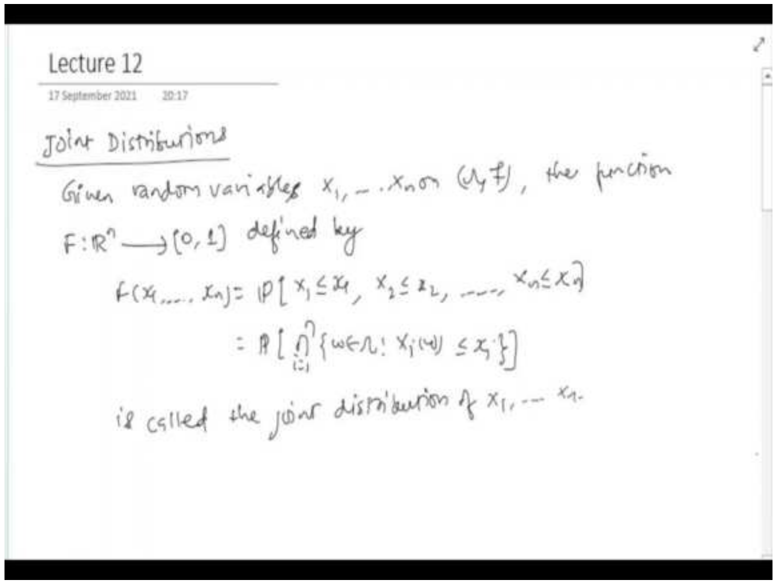
Hello everyone, welcome to today's lecture of the course mathematical aspects of biomedical electronic system design.

(Refer Slide Time: 00:35)



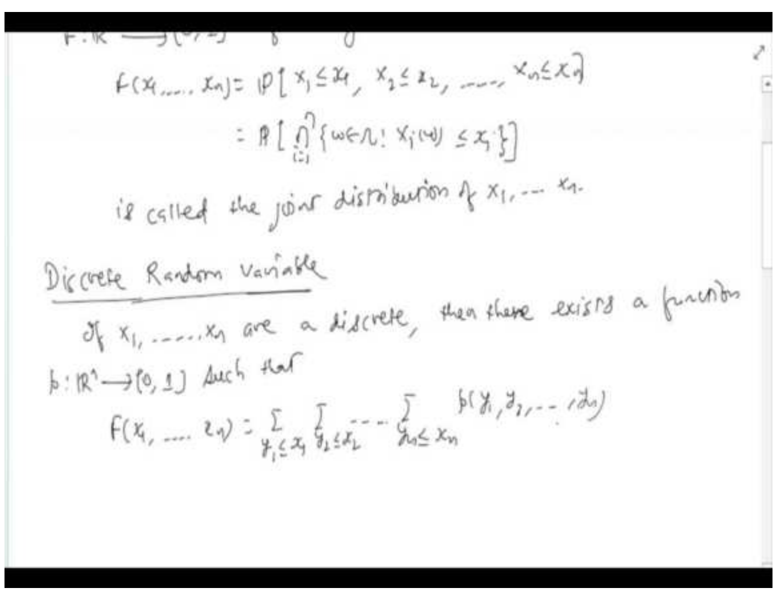
This is the last lecture on the module probability and random variables. In today's lecture, we will cover joint and marginal distributions, conditional distributions independence of random variables, covariance and correlation.

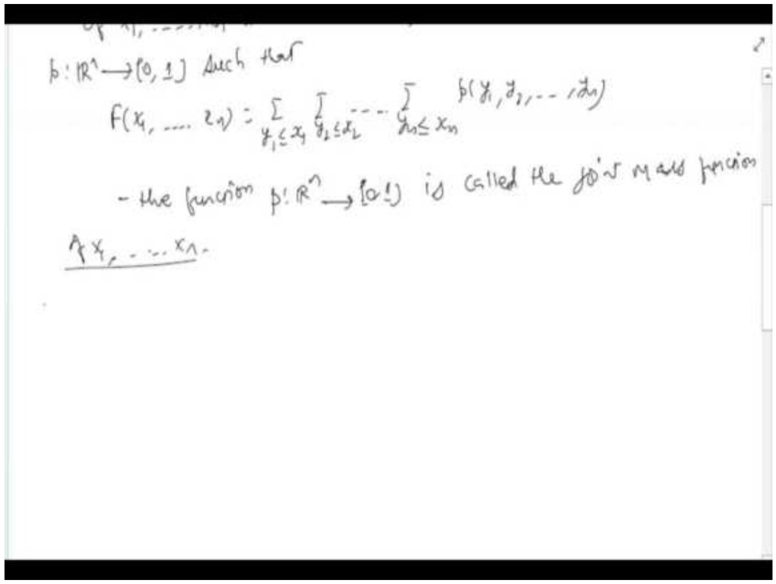
(Refer Slide Time: 00:51)



So, let us start today's lecture, we will start with the notion of joint distribution. Suppose, we have n random variables on the same space $X_1 \ldots X_n$ on space (Ω, F) then the function this is from $R^n \rightarrow [0, 1]$ that is it takes n dimensional real vectors and gives an output that is a number between 0, 1 defined by $F(x_1 \ldots x_n) = P[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n]$ that is probability of the event that intersection of the event $X_i(\Omega) \leq x_i$. This function is called a joint distribution of X_1 to Xn is called the joint distribution of $X_1 \ldots X_n$.

(Refer Slide Time: 02:37)



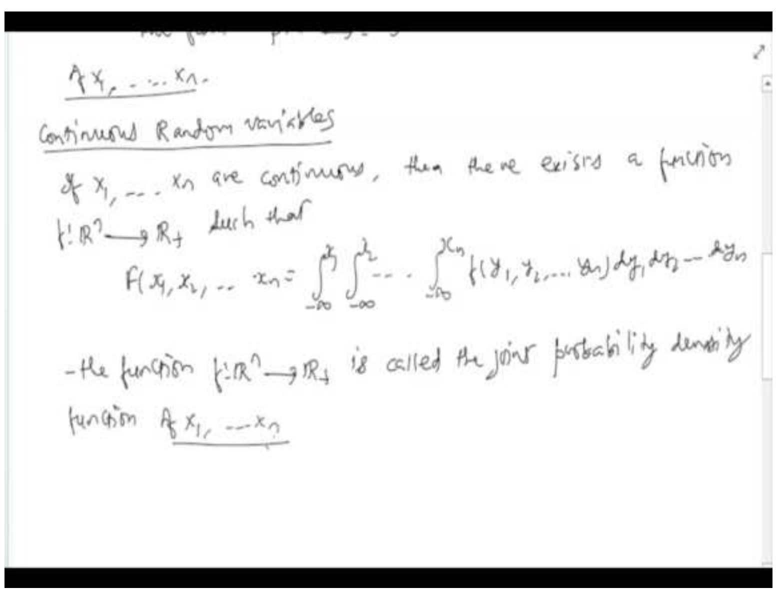


To proceed further let us consider the case of discrete and continuous random variable separately discrete random variables. So, now, we assume that the variables X_1 to X_n are all discrete if X_1 … X_n are all discrete then there exists a function $p: \mathbb{R}^n \to [0,1]$ such that the joint distribution that is

$$F(x_1, \dots, x_n) = \sum_{y_1 \le x_1} \sum_{y_2 \le x_2} \dots \dots \dots \sum_{y_n \le x_n} p(y_1, y_2, \dots \dots y_n)$$

this function p is called the joint probability mass function. The function p which as we saw from $R^n \to [0, 1]$is called the joint mass function of X_1 to X_n.

(Refer Slide Time: 04:14)

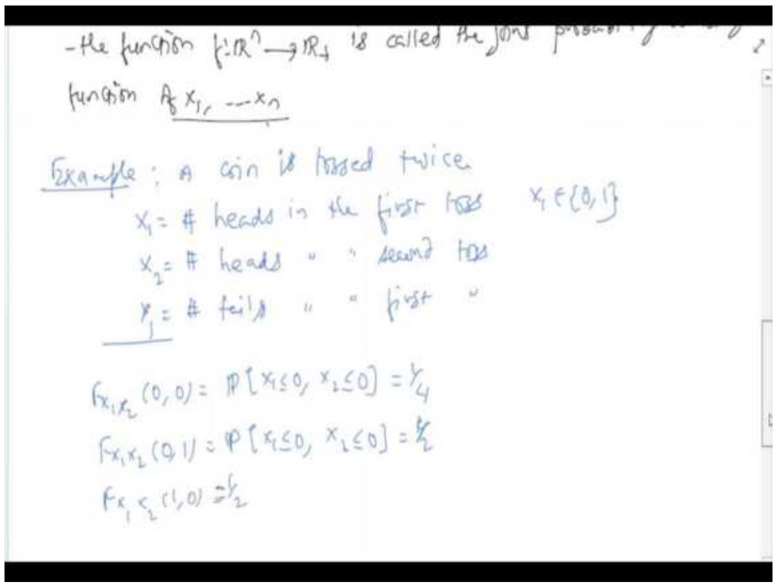


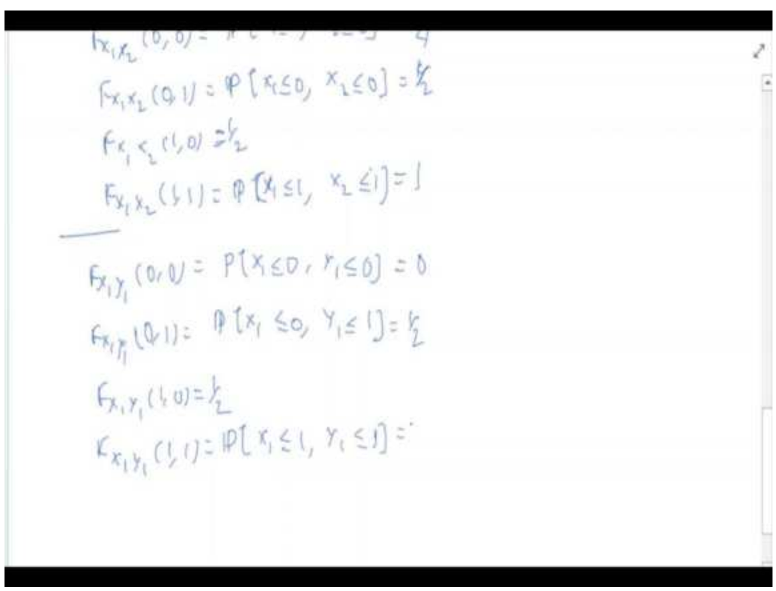
Now considered continuous random variables, if X_1 to X_n are all continuous then also there exist a function f: $\mathbb{R}^n \rightarrow \mathbb{R}_+$. So, this can take values larger than 1 such that the joint distribution of

$$F(x_1, x_2 \ldots x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \ldots . \int_{-\infty}^{x_n} f(y_1, y_2 \ldots y_n) dy_1, dy_2 \ldots dy_n$$

and this function small f is called the joint probability density function. The function f: $\mathbb{R}^n \rightarrow \mathbb{R}_+$ this is called the joint probability density function of X_1 to X_n.

(Refer Slide Time: 06:01)





Let us, consider a simple example to illustrate a few of these concepts. Consider the following experiment a coin is tossed twice and let us define random variables X_1 to be number of heads in the first flip in the first toss. Clearly X_1 can take value 0 or 1, X_2 is number of heads in the second toss and Y_1 is number of tails in the first toss.

In this case we can consider pairs of random variables X_1 X_2, X_1 Y_1, X_2 Y_1 etcetera and we can talk of their joint distributions. For instance, if we consider X_1 and X_2 the joint distribution of X_1 X_2, $F_{X_1X_2}$ can be obtained as follows. So,

$$F_{X_1X_2}(0,0) = P[X_1 \leq 0, X_2 \leq 0] = \frac{1}{4}.$$

$$F_{X_1X_2}(0,1) = P[X_1 \leq 0, X_2 \leq 0] = \frac{1}{2}.$$

Similarly,

$$F_{X_1X_2}(1,0) = \frac{1}{2}.$$

$$F_{X_1X_2}(1,1) = P[X_1 \leq 1, X_2 \leq 1] = 1.$$

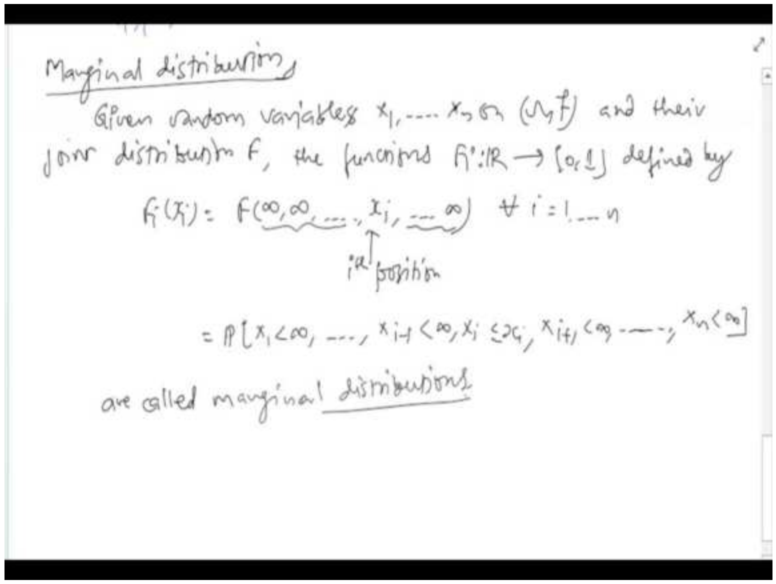
Similarly, if we consider X_1, Y_1 then joint distribution of X_1, Y_1 can be obtained as follows. So,

$$F_{X_1Y_1}(0,0) = P[X_1 \leq 0, Y_1 \leq 0] = 0,$$

that is no head in the first toss and no tail also in the first toss, this is impossible event. So, the probability of this event will be 0. $F_{X_1Y_2}(0,1)$ can be obtained at probability of no head in the first toss and 0 or 1 tail in the first toss.

Clearly intersection of these two events will be even that there is no head in the first toss and that probability will be half. Similarly, the joint distribution evaluated at 1, 0 will also be half the joint distribution evaluated at 1, 1 will be the probability that there is 0 or 1 head in the first toss and there is 0 or 1 tail in the first toss. Clearly the probability of this event will be 1. So, this is how we can evaluate joint distributions.

(Refer Slide Time: 09:27)



Next we will move to marginal distributions. Suppose we are given n random variables on the same space and there is one distribution is F. So, given random variables $X_1 \dots X_n$ on same probability space and the joint distribution is F then the functions F_i: $\mathbb{R} \to [0, 1]$ defined by

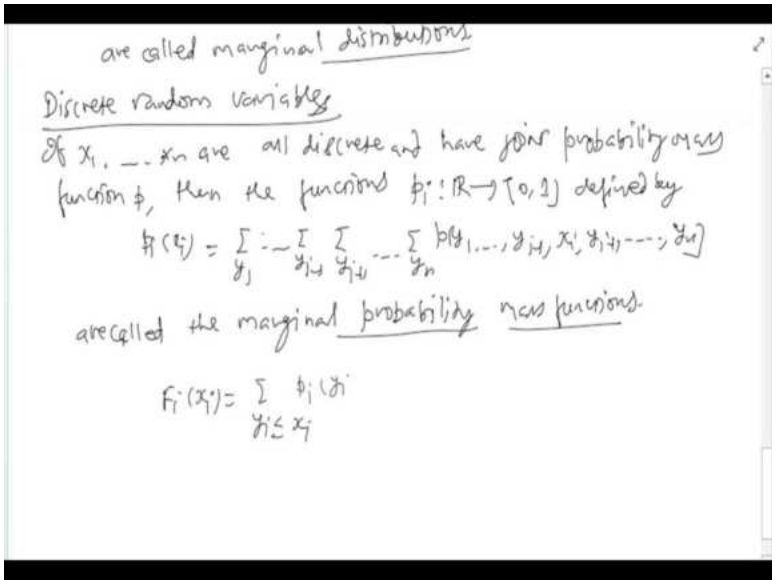
$$F_i(x_i) = F(\infty, \infty, \dots, x_i, \dots, \infty)\ \forall\ i = 1, \dots, n.$$

In fact, we can see that the right-hand side equals

$$= P[X_1 < \infty, \dots, X_{i-1} < \infty, X_i \le x_i, X_{i+1} < \infty, \dots, X_n < \infty].$$

These functions are called marginal distributions are called marginal distributions. So, F_i marginal distribution of X_i.

(Refer Slide Time: 11:37)



We can again consider the cases of discrete and continuous random variables separately. So, let us just look at discrete random variables. So, imagine that X_1 to X_n and are all discrete and have joint probability mass function. We define the table P, then the functions p_i and p_i: $\mathbb{R} \rightarrow [0, 1]$ define by

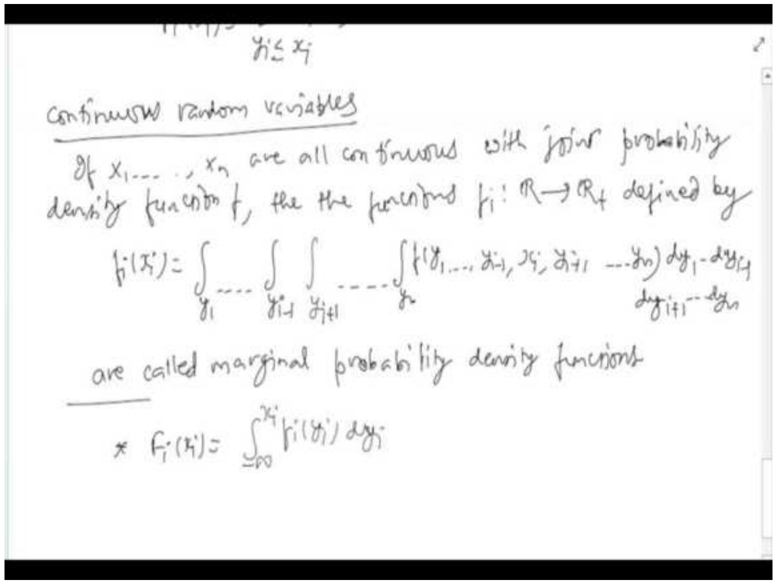
$$p_i(x_i) = \sum_{y_1} \cdots \sum_{y_{i-1}} \sum_{y_{i+1}} \cdots \sum_{y_n} p[y_1, \ldots, y_{i-1}, x_i, y_{i+1}, \ldots, y_n],$$

these functions are called the marginal probability mass functions.

We can define marginal distributions also from marginal probability mass function in fact in terms of that

$$F_i(x_i) = \sum_{y_i \le x_i} p_i y_i.$$

(Refer Slide Time: 13:37)



Let us, now turn to the case of continuous random variables. Suppose X_1 to X_n are all continuous these are all continuous with joint probability density function f, then we can define functions f_i: $\mathbb{R} \to \mathbb{R}_+$ define by

$$f_i(x_i) = \int_{y_1} \ldots \int_{y_{i-1}} \int_{y_{i+1}} \ldots \int_{y_n} f[y_1, \ldots, y_{i-1}, x_i, y_{i+1}, \ldots, y_n] dy_1 \ldots dy_{i-1} dy_{i+1} \ldots dy_n,$$

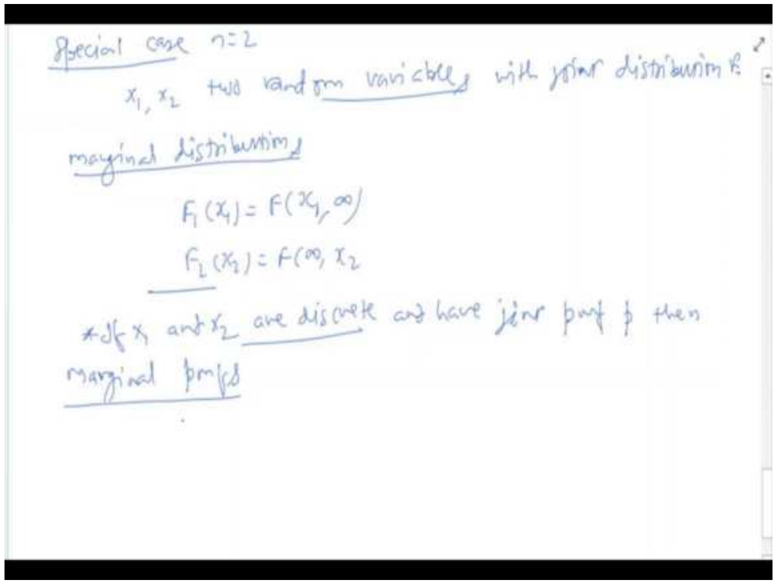
these functions are called marginal probability density functions.

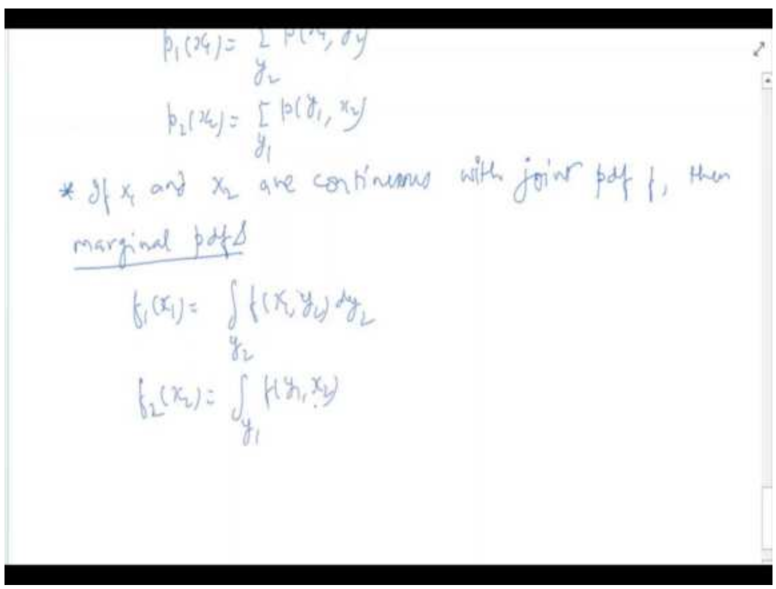
In case of continuous random variables, we can obtain marginal distributions also from marginal probability density functions. For instance, it turns out that the

$$F_i(x_i) = \int_{-\infty}^{x_i} f_i(y_i)\, dy_i.$$

The concept of marginal distributions as defined above for n random variables might look complex, but can be easily motivated by considering the case of two random variables.

(Refer Slide Time: 16:14)





So, let us see the special case of n = 2 to better understand the idea of marginal distributions. So, in this case if X_1, X_2 are two random variables with joint distribution capital F then we can write the marginal distributions F_1 and F_2 as follows $F_1(x_1) = F(x_1,\infty)$, $F_2(x_2) = F(\infty,x_2)$. If x_1 and x_2 are discrete and have joint probability mass function joint pmf p, then we can obtain the marginal pmfs p_1 and p_2 as follows. So,

$$p_1(x_1) = \sum_{y_2} p(x_1, y_2),$$

$$p_2(x_2) = \sum_{y_1} p(y_1, x_2).$$

In either case we can compute the marginal distribution as summation of marginal probability mass functions on the other hand if X_1, X_2 are continuous with joint probability density function f, then we can compute the marginal probability density functions f_1 and f_2 as follows marginal density functions

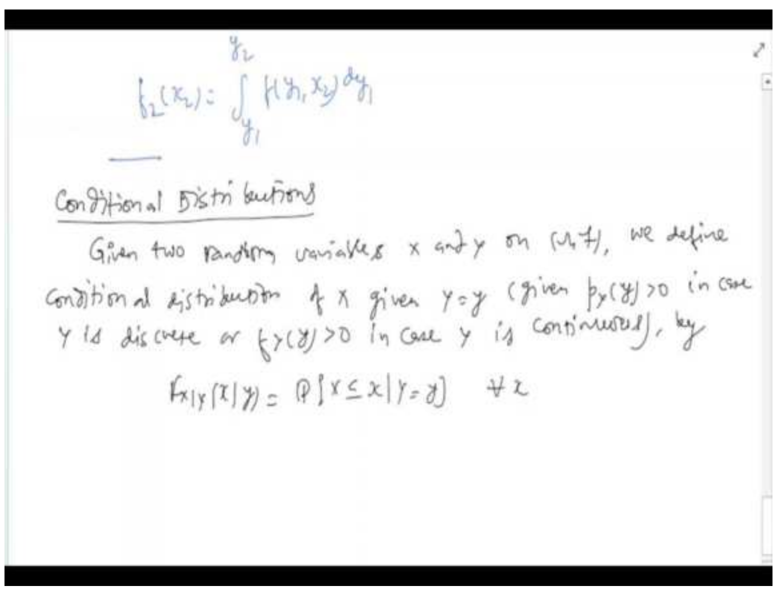
$$f_1(x_1) = \int_{y_2} f(x_1, y_2) dy_2.$$

Similarly,

$$f_2(x_2) = \int_{y_1} f(y_1, x_2) dy_1.$$

Again for both x_1 and x_2 their marginal distributions can be computed by integrating the marginal probability density functions as described above. So, this was about marginal distributions.

(Refer Slide Time: 19:22)

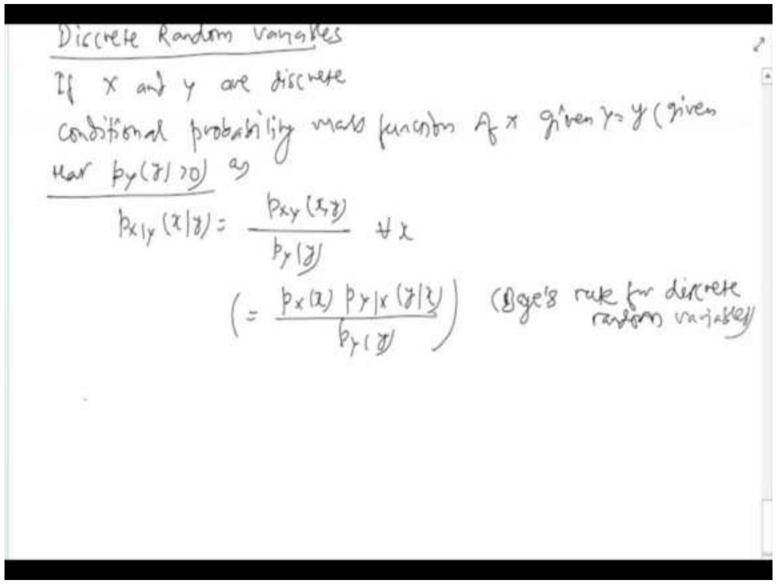


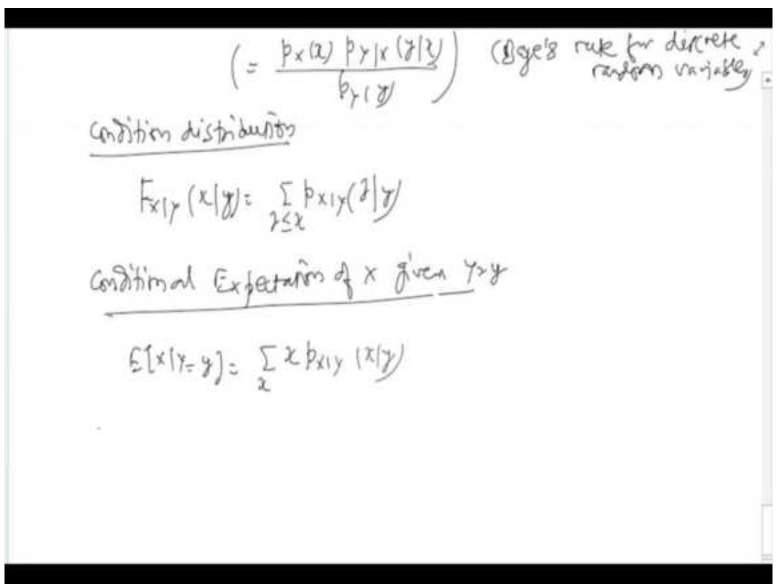
Next let us, see the notion of conditional distributions. Here we will restrict attention to two random variables given two random variables X and Y on same space say (Ω, F). We define conditional distribution have X given Y = y and this is defined only if $p_y(y) > 0$ in case Y is discrete or $f_y(y) > 0$ in case Y is continuous.

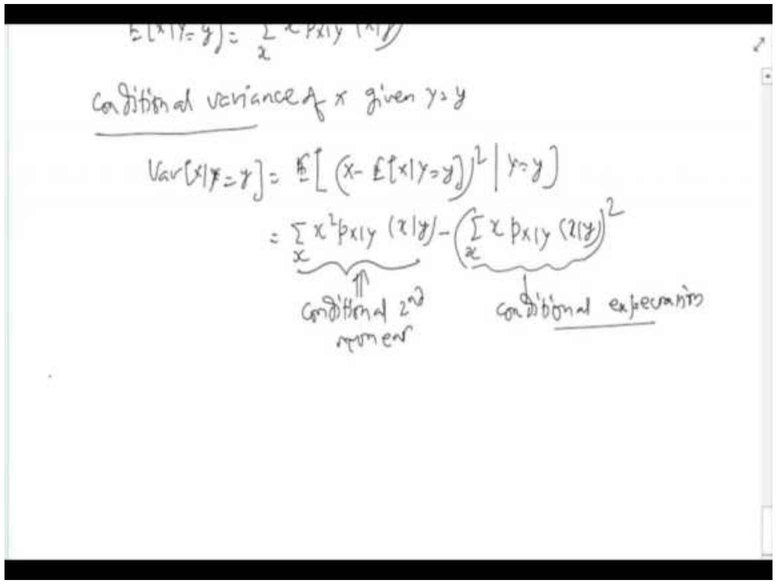
In such cases, we can define the conditional distribution of X by

$$F_{X|Y}(x|y) = P[X \leq x | Y = y] \quad \forall x.$$

(Refer Slide Time: 21:01)







Let us, again consider the cases of discrete and continuous random variables separately. So, first let us consider the case of discrete random variables. If X and Y are discrete, then we can define conditional probability mass function have X given Y = y and of course, this is given that $p_y(y|>0)$ as

$$p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)} \quad \forall x,$$

that is a joint probability mass function divided by marginal probability mass function of y for all x.

Notice that the right hand side can also be written using Bayes rule as follows, it can be written as

$$= \frac{p_X(x)p_{Y|X}(y|x)}{p_Y(y)},$$

this is Bayes rule for discrete random variables. From conditional probability mass function, we can compute the conditional distribution of X as follows

$$F_{X|Y}(x|y) = \sum_{z \le x} p_{X|Y}(z|y).$$

Further we can define conditional expectation of X given Y = y denote as

$$E[X|Y = y] = \sum_{x} x\, p_{X|Y}(x|y),$$

we can also define conditional variance of X given Y = y this is denoted as

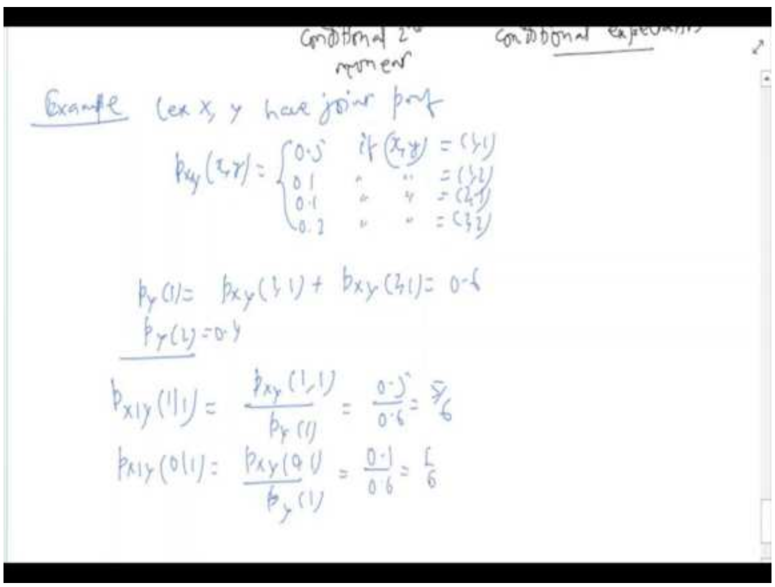
$$Var[X|Y=y] = E[(X - E[X|Y=y])^2|Y=y]$$

and this turns out to be

$$= \sum_x x^2\, p_{X|Y}(x|y) - \left(\sum_x x\, p_{X|Y}(x|y)\right)^2.$$

Notice that the first quantity is conditional second moment, second moment or second order moment whereas, the second quantity here is square of conditional expectation. This is conditional expectation.

(Refer Slide Time: 25:10)



Let us, see an example. Suppose we have two random variables X and Y discrete random variables that have joint probability mass function

$$p_{X|Y}(x|y) = \begin{cases} 0.5 & if(x,y) = (1,1) \\ 0.1 & if(x,y) = (1,2) \\ 0.1 & if(x,y) = (2,1) \\ 0.3 & if(x,y) = (2,2) \end{cases}.$$

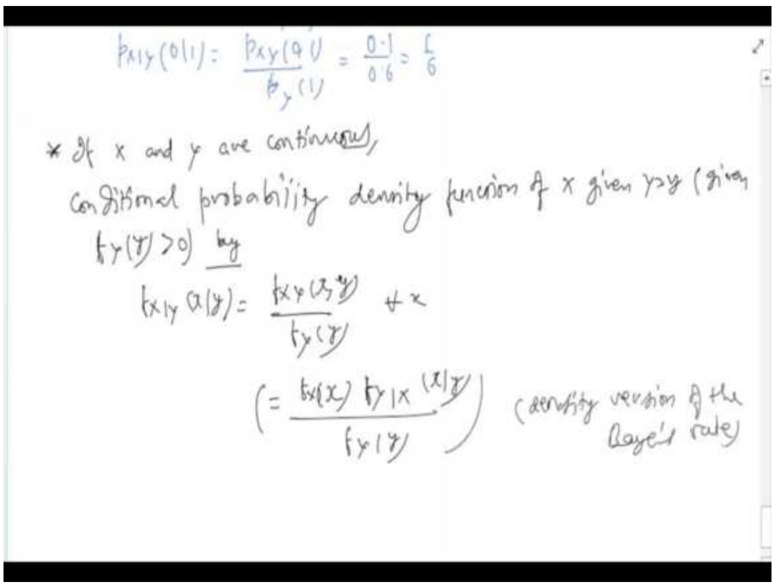
From the joint probability mass function we can compute the marginal probability mass function of y as follows $p_Y(1) = p_{XY}(1, 1) + p_{XY}(2, 1) = 0.6$. Similarly, $p_Y(2) = 0.4$, having computed the marginal probability mass functions we can compute the conditional probability mass function of X|Y in particular

$$p_{X|Y}(1|1) = \frac{p_{X|Y}(1,1)}{p_Y(1)} = \frac{0.5}{0.6} = \frac{5}{6}.$$

We could similarly compute

$$p_{X|Y}(0|1) = \frac{p_{X|Y}(0,1)}{p_Y(1)} = \frac{0.1}{0.6} = \frac{1}{6}.$$

(Refer Slide Time: 27:00)



Next we turn our attention to continuous random variables. If X and Y are continuous, we define conditional probability density function of X given Y = y additional probability density function of X given Y = y of course, it is defined only when $f_y(y > 0)$, this is defined

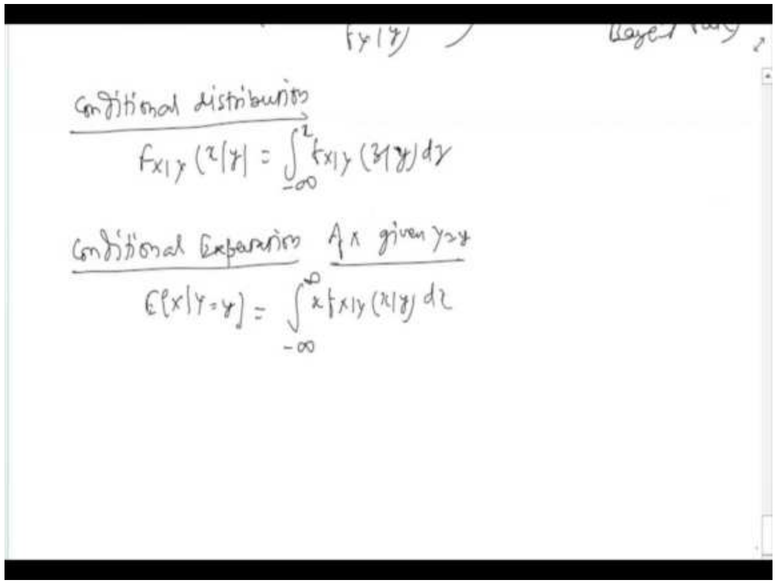
$$f_{X|Y}(x|y) = \frac{f_{X|Y}(x,y)}{f_Y(y)}.$$

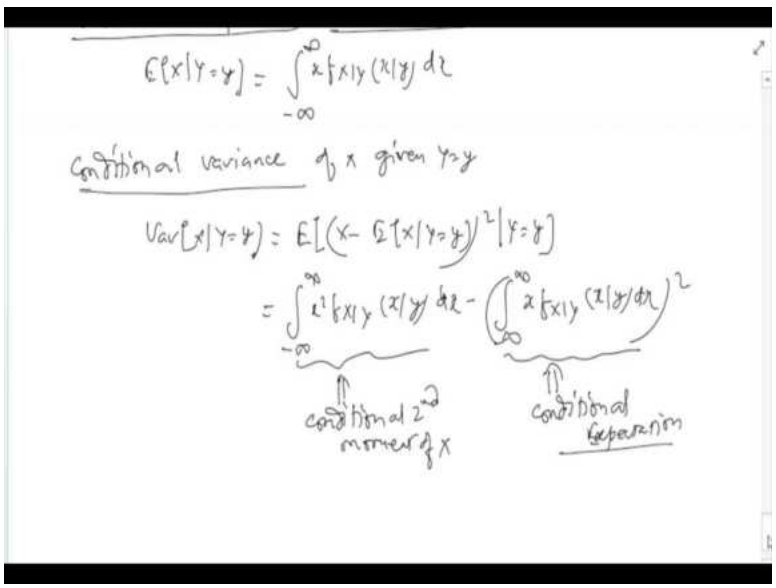
Again, using the Bayes rule, it can be seen that the right hand side expression here equals

$$= \frac{f_X(x) f_{Y|X}(x|y)}{f_Y(y)}.$$

So, here what we have to use is density version of the Bayes rule having obtained the conditional probability density function.

(Refer Slide Time: 28:39)





We can obtain the conditional distribution of X as follows; conditional distribution that is

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(z|y)\, dz.$$

We can also define conditional expectation of X given Y = y. This is denoted as

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx.$$

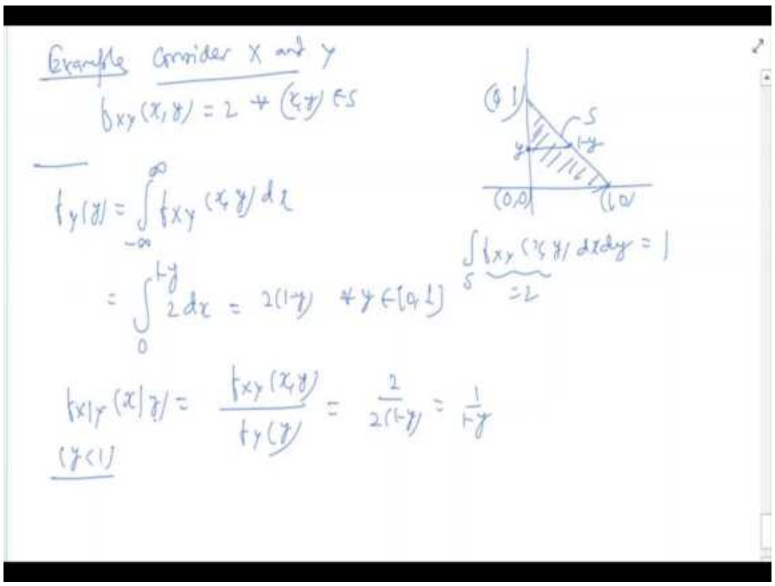
We can also define conditional variance of X given Y = y. This is denoted as

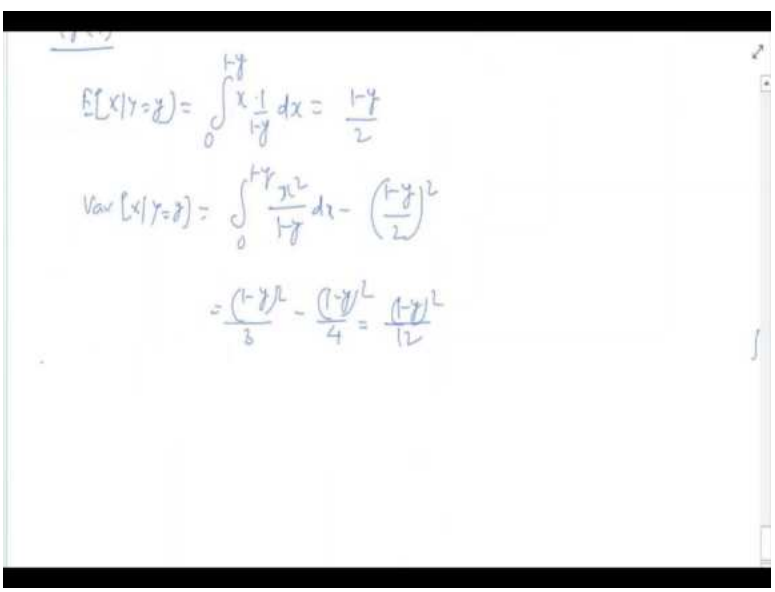
$$Var[X|Y=y] = E[(X - E[X|Y=y])^2|Y=y]$$

$$= \int_{-\infty}^{\infty} x^2 f_{X|Y}(x|y)\,dx - \left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx\right)^2,$$

as in discrete case, the first term is conditional second moment of X conditional second order moment of X and the second term is the square of the conditional expectation. So, this is conditional expectation.

(Refer Slide Time: 31:05)





Let us, illustrate a few of these notions again with an example. For this example, let us consider two random variables X and Y and region S defined as follows. So, this region, we call S is points 0, 1 and 1. So X, Y pair takes values in this region and probability of X, Y pair taking any value in this region is same. So, joint density function of x and y is a constant, which is 2 for all x, y in S the reason to pick 2 here is we need to satisfy the following condition

$$\int_{S} f_{XY}(x, y) dx dy = 1$$

and this region itself has area half.

So, we need to pick this to be equal to 2 to make this whole integral 1. So, this is the joint probability density function of X and Y. From joint probability density function, we can compute the marginal density of Y as follows. So we want to compute $f_Y(y)$. Notice that for fixed y x can range from 0 to 1 - y.

So, $f_Y(y)$, which will be

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

$$= \int_{0}^{1-y} 2dx = 2(1 - y), \ \ \forall y \in [0,1]$$

having defined the marginal probability density function of X we can compute the conditional probability density function of X given Y that is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{2}{2(1 - y)} = \frac{1}{1 - y}.$$

In this definition, we assume that y < 1 as when y = 1, the marginal density of y at 1 is 0. So, this conditional density is not defined. Having defined the conditional density of X we now turn to conditional expectation of X given Y = y and this is simply the expectation with respect to the above defined conditional density.

So, it will be

$$E[X|Y = y] = \int_{0}^{1-y} x . \frac{1}{1 - y} dx = \frac{1 - y}{2}.$$

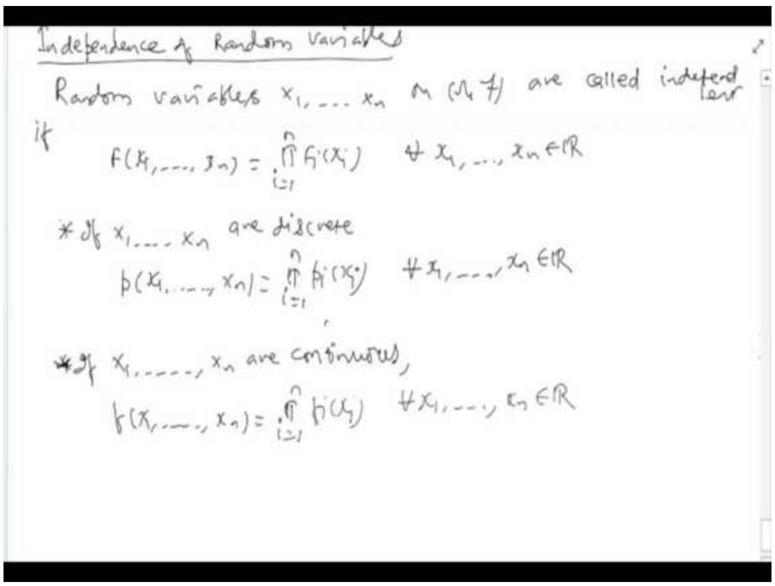
We can also write conditional variance of X given Y = y, this will be

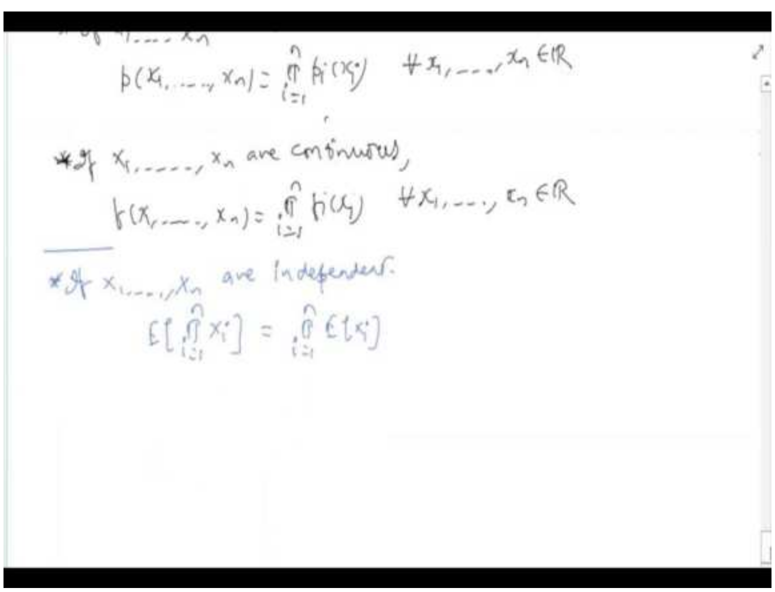
$$Var[X|Y = y] = \int_0^{1-y} \frac{x^2}{1-y}\, dx - \left(\frac{1-y}{2}\right)^2$$

and this turns out to be equal to

$$\frac{(1-y)^2}{3} - \frac{(1-y)^2}{4} = \frac{(1-y)^2}{12}.$$

(Refer Slide Time: 34:38)





Now, we will see the notion of independence of random variables, random variables X_1 to X_n defined on same space are called independent if there is one distribution equals the product of marginal distribution that is if

$$F(x_1, \dots, x_n) = \prod_{i=1}^{n} F_i(x_i) \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

If the random variables X_1 to X_n are discrete the equivalent definition of independence is that these random variables are independent if their joint probability mass function is a product of marginal probability mass functions that is

$$p(x_1, \dots, x_n) = \prod_{i=1}^{n} p_i(x_i) \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

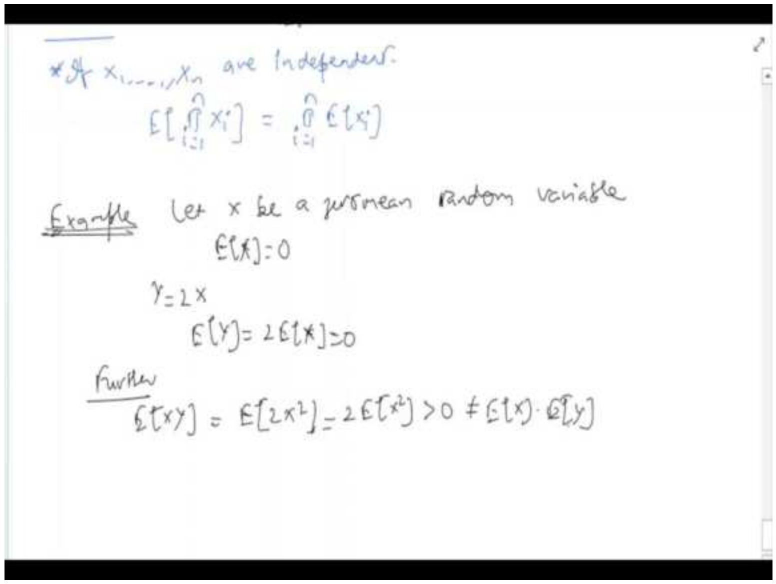
Similarly, if these random variables X_1 to X_n are continuous then also we have an alternate definition of independence that is these random variables are independent if their joint density is product of marginal densities this is

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_i(x_i) \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

Here is a property of independent random variables. If $X_1, \dots, X_n$ are independent then expectation of their product that is

$$E\left[\prod_{i=1}^{n} x_i\right] = \prod_{i=1}^{n} E[x_i].$$

(Refer Slide Time: 37:12)



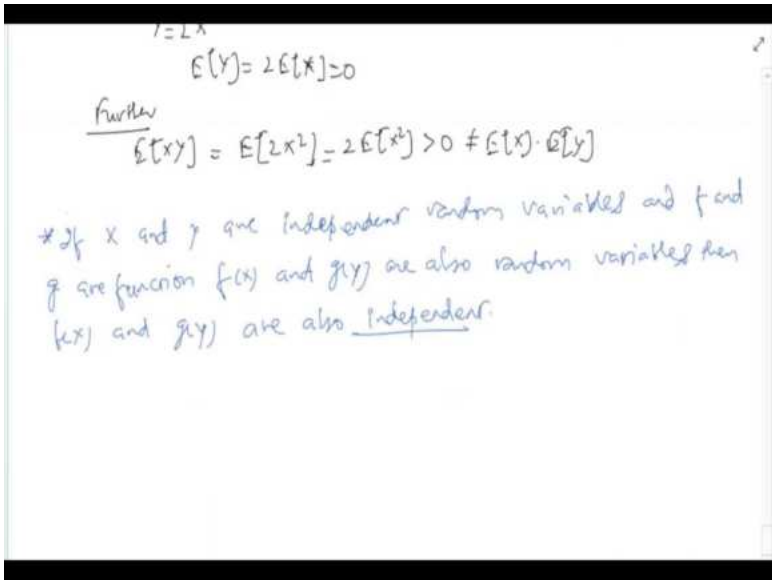
Let us see an example to see that independence is required for such a condition to hold. Example let X be a 0 mean random variable that is E[X] = 0. Suppose Y is also a random variable on the same space and Y = 2X, then, E[Y] = 2E[X] = 0.

However, if we consider product of X and Y and take its expectation expected value of X, Y, this will be

$$E[XY] = E[2X^2] = 2E[X^2] > 0 \neq E[X].E[Y],$$

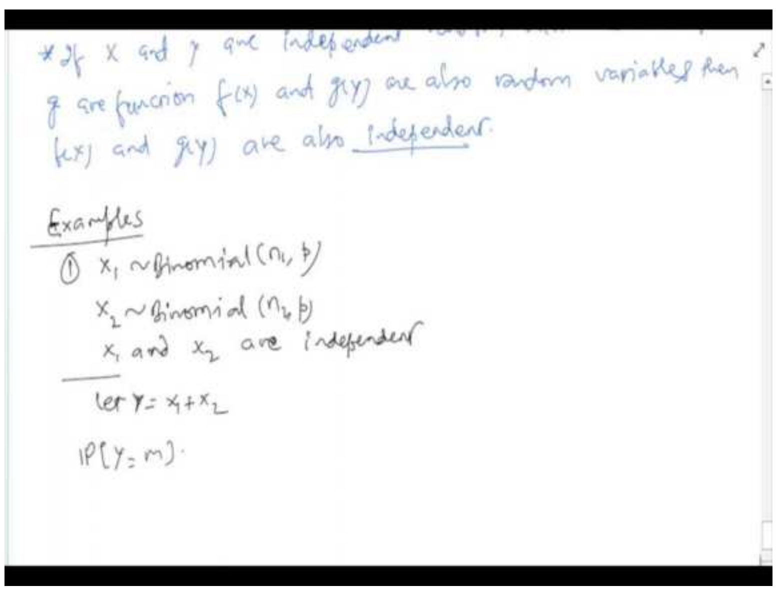
this illustrates that independence of random variables is required for the above relation to hold.

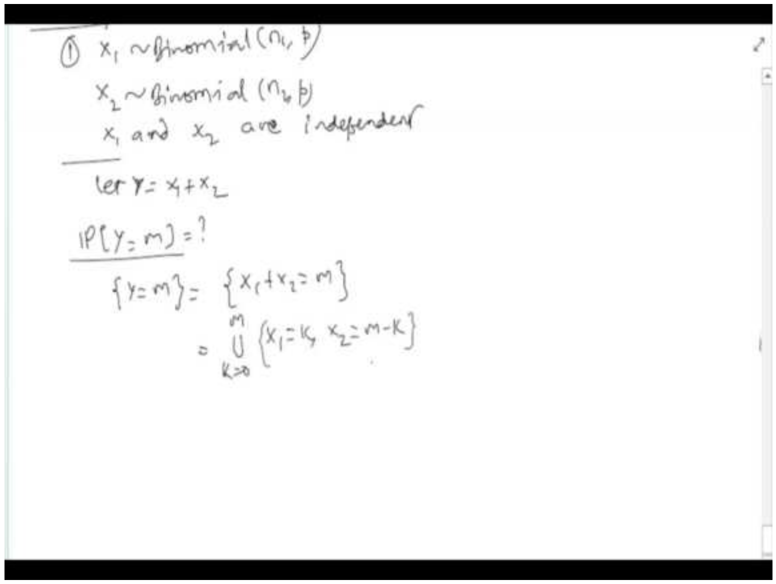
(Refer Slide Time: 38:25)

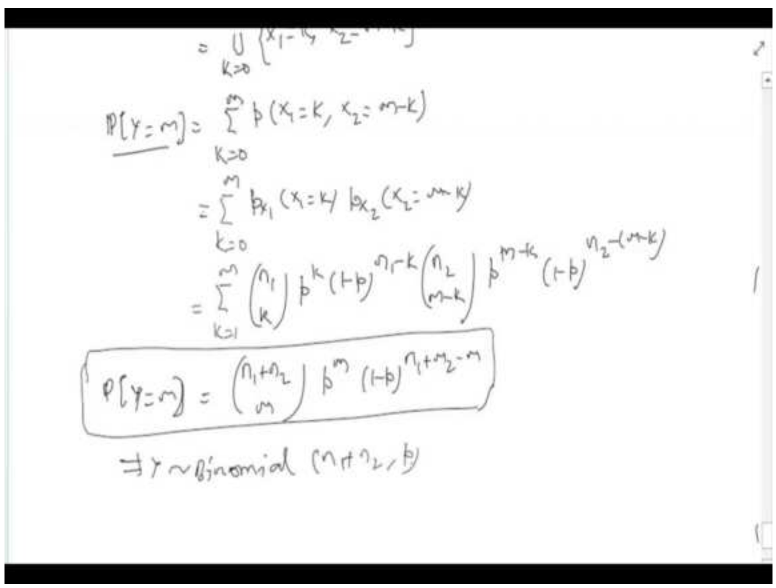


Let us, see 1 more fact. If X and Y are independent random variables and f and g are functions such that f(X) and g(Y) are also random variables. Then, f(X) and the g(Y) are also independent.

(Refer Slide Time: 39:24)







Let us, now see a couple of more examples that illustrate how the property of independence of random variables can be used to compute the probability mass function or probability density function of some of those random variables. Examples as the first example it is assumed that X_1 and X_2 are two independent binomial random variables X_1 is binomial which parameter n_1 and p and X_2 a binomial which parameters n_2 to p. X_1 and X_2 are independent.

Let $Y = X_1 + X_2$, we are interested in probability mass function of Y that is we are interested in probability that Y will be some number m. Notice that $\{Y = m\} = \{X_1 + X_2 = m\}$ and this later event is $= \bigcup_{k=0}^{m} \{X_1 = k, X_2 = m - k\}$.

If any of these events happen, the above event would also happen, then using law of total probability we can write

$$P(Y = m) = \sum_{k=0}^{m} p(X_1 = k, X_2 = m - k)$$

at this point, we can exploit the fact that X_1 and X_2 are independent to write the right-hand side as

$$= \sum_{k=0}^{m} p_{X_1}\big(X_1 = k\big)\, p_{X_2}\big(X_2 = m - k\big).$$

Since both X_1 and X_2 are binomial random variables, the terms inside the summation it become equal to

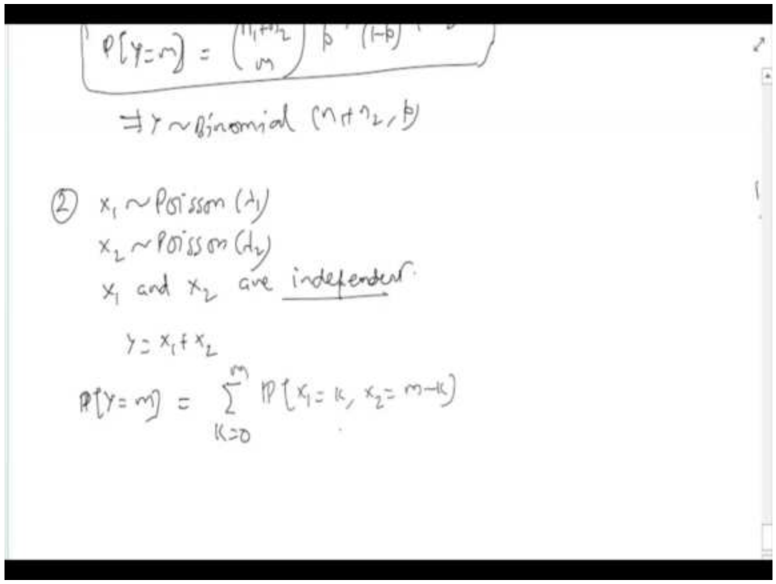
$$= \sum_{k=0}^{m} \binom{n_1}{k} p^k (1-p)^{n-k} \binom{n_2}{m-k} p^{m-k} (1-p)^{n-(m-k)}.$$

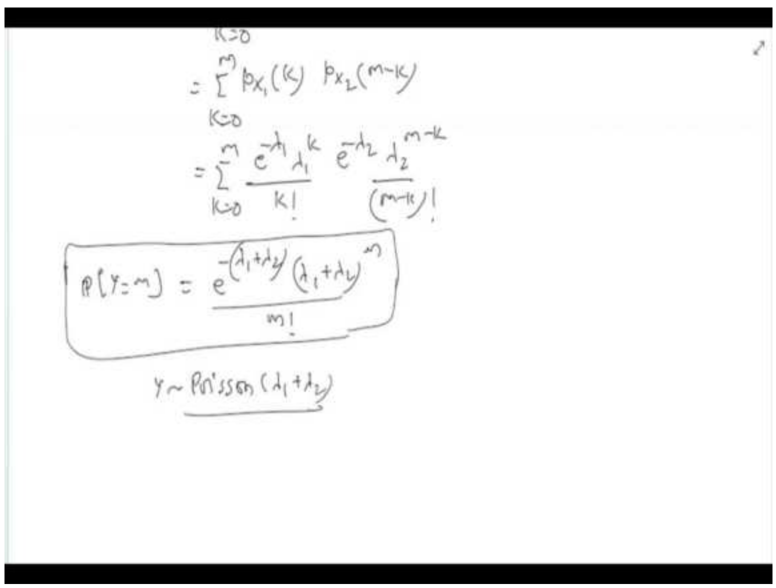
In writing this I assume that m is smaller than both n_1 and n_2, but similar calculation can be for other values of m also. If we do the calculations correctly, the right-hand side turns out to be

$$P(Y = m) = \binom{n_1 + n_2}{m} p^m (1-p)^{n_1+n_2-m}.$$

So, looking at the form of probability mass function of Y, we can infer that Y is also binomial random variable with parameters ($n_1 + n_2$, p). So, sum of two independent binomial random variables is also a binomial random variable.

(Refer Slide Time: 42:48)





Let us, now take a second example and now we assume that X_1 and X_2 are independent Poisson random variables. So, that is X_1 is Poisson with parameter λ_1 and X_2 is Poisson with parameter λ_2 X_1 and X_2 are independent and as before, we defined $Y = X_1 + X_2$ and we are interested in probability mass function of Y. That is, we are interested in P(Y = m) for an integer m and as in the binomial case, I can write it as

$$P(Y = m) = \sum_{k=0}^{m} P[X_1 = k, X_2 = m - k].$$

Again, exploiting the fact that X_1 and X_2 are independent, we can write this summation as

$$= \sum_{k=0}^{m} p_{X_1}(k)\, p_{X_2}(m-k)$$

and since X_1 and X_2 both are Poisson, we can further the right-hand side at

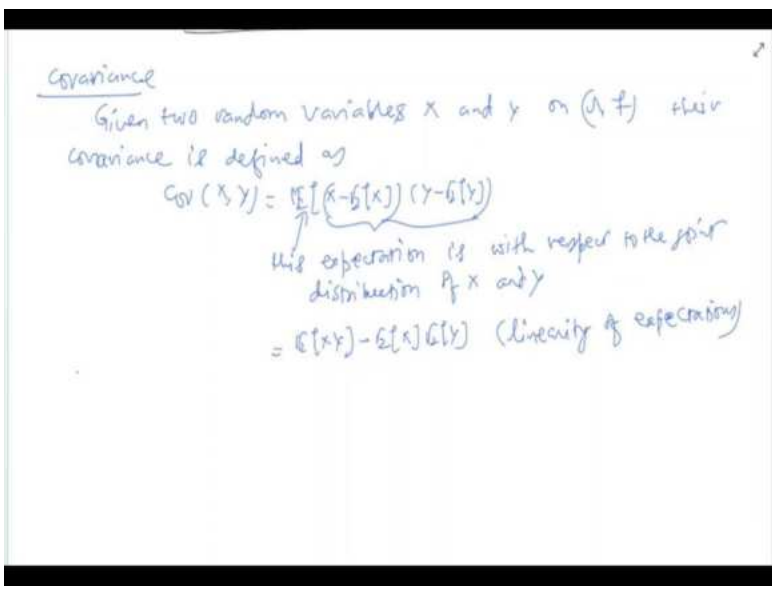
$$= \sum_{k=0}^{m} \frac{e^{-\lambda_1}{\lambda_1}^{k}}{k!} \frac{e^{-\lambda_2}{\lambda_2}^{m-k}}{(m-k)!}.$$

And this summation turns out to be equal to

$$= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^m}{m!}.$$

Again looking at the probability mass function of Y, we can infer that Y is also a Poisson random variable Y is Poisson which parameter $(\lambda_1 + \lambda_2)$. So, we see that sum of two Poisson random variables if they are independent, is also Poisson random variable with the parameter of this new variable the sum of the parameters of the original value.

(Refer Slide Time: 44:55)

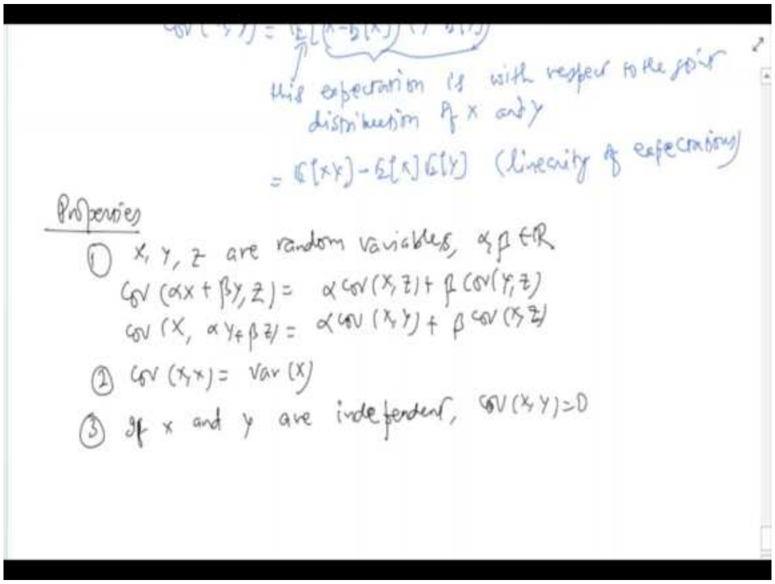


Now, let us see the notion of covariance. Given two random variables X and Y on same space, their covariance is defined as COV, this is the notation $COV(X, Y) = E[(X - E[X])(Y - E[Y])]$.

Notice that the product here depends on both X and Y. So, this is expectation with respect to joint distribution, this expectation is with respect to the joint distribution of X and Y.

Expression turns out to be $= E[XY] - E[X]E[Y]$. The second expression is obtained using linearity of expectation. As above the first expectation here is taking using a joint distribution of X and Y.

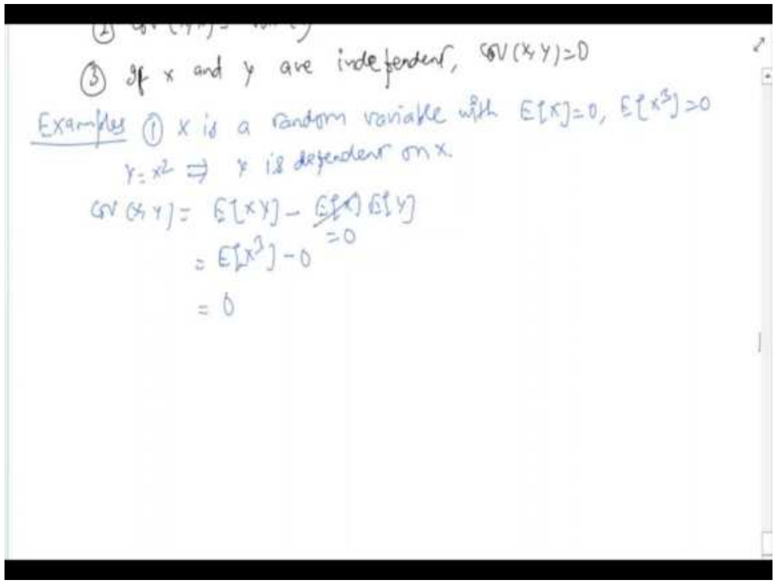
(Refer Slide Time: 46:44)



Let us, see a few properties of covariance. First properties is if X Y and Z are random variables, and α,β are 2 numbers, then, $COV(\alpha X + \beta Y, Z) = \alpha COV(X, Z) + \beta COV(Y, Z)$, further the $COV(X, \alpha Y + \beta Z) = \alpha COV(X, Y) + \beta COV(X, Z)$.

Second property is $COV(X, X) = Var(X)$. Another property is if X and Y are independent random variables, then their covariance $COV(X, Y) = 0$.

(Refer Slide Time: 48:03)



Let us, see a couple of examples to understand the notion of covariance. The first example illustrates that X and Y need not be independent for their covariance to be 0. Suppose, X is a random variable its expectation 0 and the third moment of X is also 0. Let Y = X^2 then certainly X and Y are not independent.

Now, let us compute the covariance of X and Y,

$$COV(X,Y) = E[XY] - E[X]E[Y] = E[X^3] - 0 = 0$$

Notice that Y = X^2, so the first term becomes $E[X^3]$. As far as second term goes, $E[X]$ = 0, though $E[Y]$ which is $E[X^2]$ is not 0, the second term becomes 0. From the hypothesis, the first term is also 0. So, we see that covariance of X and Y is 0, even though they are not independent.

(Refer Slide Time: 49:28)

$= 0$

② Let X be a random variable, uniformly distributed over $\{1,2,3\}$. Let Y be another random variable (independent of X), uniformly distributed over $\{-1,1\}$.

Let $Z = XY$

Now consider X and Z

* X and Z are dependent.

$$F_{XZ}(1,1) = \mathbb{P}[X \le 1, Z \le 1]$$
$$= \mathbb{P}[X \le 1, XY \le 1]$$
$$= \mathbb{P}[X \le 1]$$

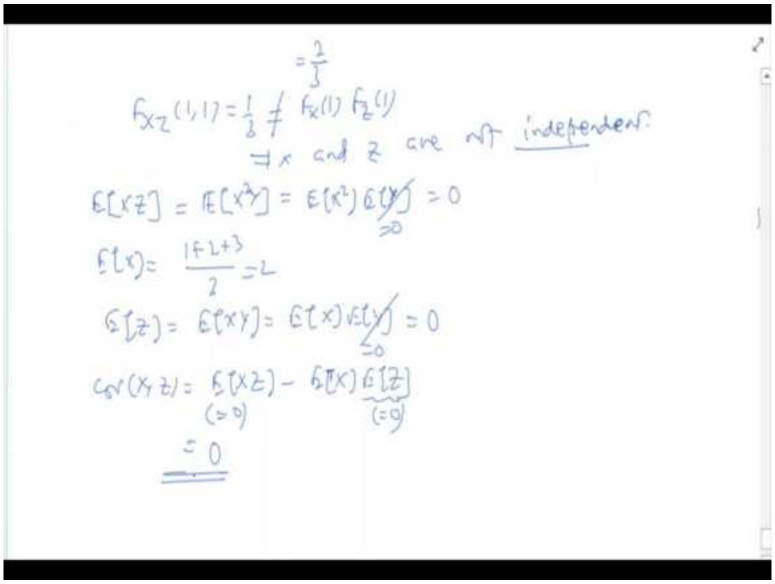
$$= \mathbb{P}[X \le 1, XY \le 1]$$
$$= \mathbb{P}[X \le 1] = \frac{1}{3}$$

$$F_X(1) = \mathbb{P}[X \le 1] = \frac{1}{3}$$

$$F_Z(1) = \mathbb{P}[Z \le 1]$$
$$= \mathbb{P}[XY \le 1]$$
$$= \mathbb{P}[X \le 1] + \mathbb{P}[X > 1, Y = -1]$$
$$= \frac{1}{3} + \frac{\cancel{2}}{3} \cdot \frac{1}{\cancel{2}}$$
$$= \frac{2}{3}$$

$$F_{XZ}(1,1) = \frac{1}{3} \neq F_X(1)\, F_Z(1)$$

$\therefore X$ and Z are not independent.



We take one more example that also shows the same fact. Let X be a random variable that is uniformly distributed over {1, 2, 3}. Let Y be another random variable be another random variable and this Y is independent of X and this is uniformly distributed over {-1, 1}. Let us define the third random variable Z = XY.

Now, consider X and Z, we will first show that X and Z are not independent. Towards this let us consider the joint distribution of X and Z evaluated (1, 1). So,

$$F_{XZ}(1,1) = P[X \le 1, Z \le 1]$$

$$= P[X \le 1, XY \le 1]$$

$$= P[X \le 1] = \frac{1}{3}.$$

Clearly, the marginal distribution of X at 1, $F_X(1) = P[X \le 1] = \frac{1}{3}$.

Now, see the marginal distribution of Z at 1 that is

$$F_Z(1) = P[Z \le 1]$$

$$= P[XY \le 1]$$

$$= P[X \le 1] + P[X > 1, Y = -1]$$

$$= \frac{1}{3} + \frac{2}{3}.\frac{1}{2} = \frac{2}{3}.$$

Notice that,

$$F_{XZ}(1,1) = \frac{1}{3} \neq F_X(1)F_Z(1).$$

So, X and Z are not independent. But as we show below covariance of X and Z is 0. To see this let us first observe what is expectation of (XZ).

$$E[XY] = E[X^2Y] = E[X^2]E[Y] = 0.$$

Expectation of X alone is

$$E[X] = \frac{1+2+3}{3} = 2.$$

Expectation of Z alone will be

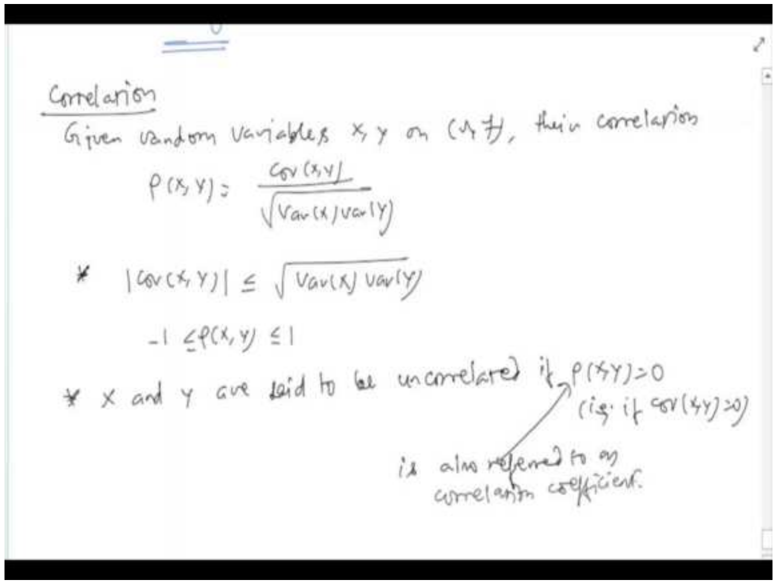
$$E[Z] = E[XY] = E[X]E[Y] = 0.$$

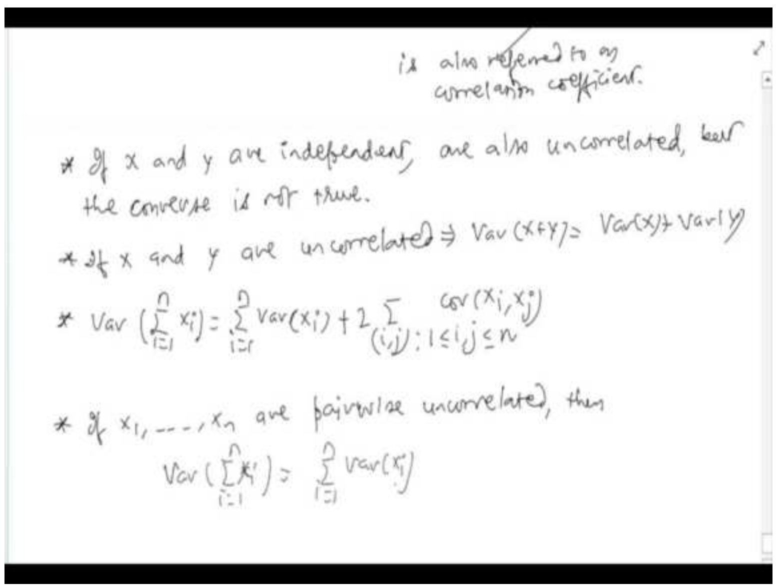
Now, we can see that

$$COV(X,Z) = E[XZ] - E[X]E[Z] = 0.$$

This is because the first term is 0 and here the last term is also Z. So, again we see that covariance of two random variables can be 0 even though they are not independent.

(Refer Slide Time: 54:15)





Finally, let us see the notion of correlation of random variables. Correlation given two random variables X and Y on same probability space there correlation is defined as

$$\rho(X, Y) = \frac{COV(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

for any two random variables X and Y, the absolute value of covariance turns out to be

$$|COV(X, Y)| \leq \sqrt{Var(X)Var(Y)}.$$

This implies that correlation of the random variables X and Y is always between -1 and +1. Further, the random variables X and Y are said to be uncorrelated, if the correlation is 0, which is same as saying that their covariance is 0, if COV(X,Y) = 0, this quantity $\rho(X, Y)$ is also known as correlation coefficient, also referred to as correlation coefficient.

Let us, see a few more facts about correlation. If the two random variables X and Y are independent, then they are also uncorrelated. We have already seen that in this case, X and Y have 0 covariance that means they have 0 correlations, they are also uncorrelated, but converse is not true.

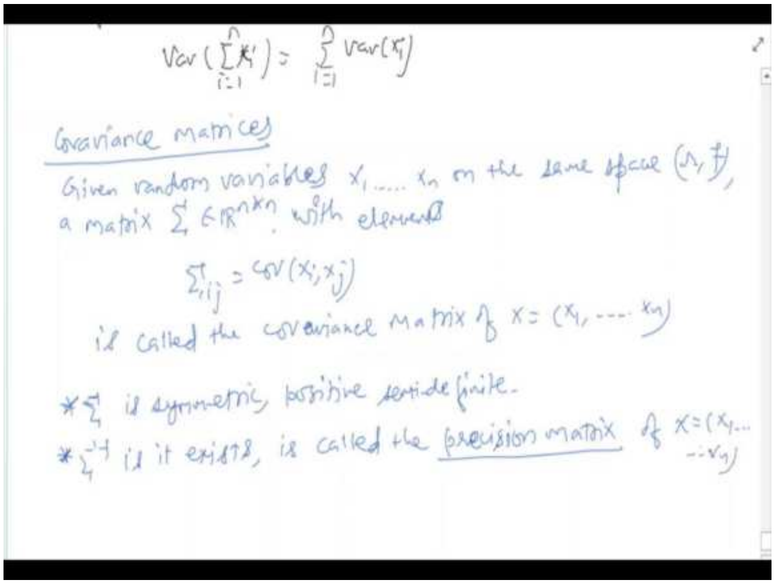
That is if the correlation between two random variables is 0 that does not mean that there would be independent, we have already seen examples where two random variables at 0 covariance that means they were uncorrelated, but they were not independent. Next fact is if X and Y are uncorrelated, then variance of their summation is simply sum of their variances.

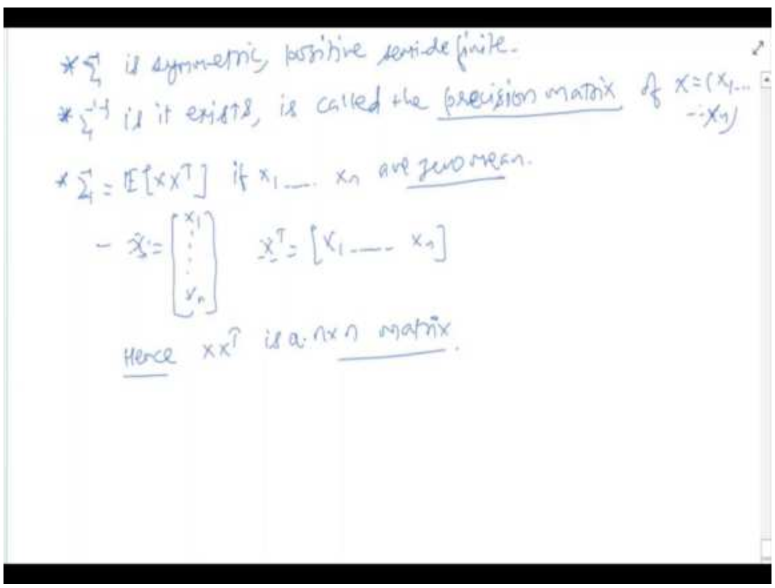
The next fact is generalization of this fact it says

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2 \sum_{(i,j):1 \le i,j \le n} COV(X_i, X_j).$$

So, this is the expression of variance of sum of random variables, this is general expression this does not need the random variables to be independent. Clearly, if the random variables are independent, in that case, they are uncorrelated, the second summation here becomes 0 and variance of summation becomes sum of variances. So, in fact, if X_1, …., X_n are pairwise uncorrelated, that is, the pairwise correlations are 0. All the terms in the second summation above become 0. So, variants of sum of these random variables become sum of the variances.

(Refer Slide Time: 59:01)





We end with discussing the notion of covariance matrices given random variables $X_1 \ldots X_n$ on the same space a matrix, $\sum \in \mathbb{R}^{n \times n}$ with elements $\sum_{ij} = COV(X_i, X_j)$ that is it is a matrix which elements are covariance's of pairs of random variables, this matrix is called covariance matrix of the random vector $X = X_1 \ldots X_n$.

Let us, see a few facts about covariance matrices first facts is covariance matrix $\sum$ is symmetric positive semi definite matrix, symmetric positive semi definite. Next the inverse of covariance matrix if the covariance matrix is invertible that if the inverse exists this inverse is called the precision matrix of the random vector X it is X = X1 … Xn.

Finally, the covariance matrix turns out to be $\sum = E[XX^T]$ if $X_1 \ldots X_n$ are all 0 mean notice that here in writing this

$$X = \begin{bmatrix} X_1 \\ . \\ . \\ . \\ X_n \end{bmatrix}.$$

So,

$$X^T = [X_1 \ldots \quad \ldots \quad X_n],$$

this implies that XX^T is a $n \times n$ matrix random matrix and above fact we are talking of expectation of this matrix. This brings us to the end of today's lecture. Thank you everyone.