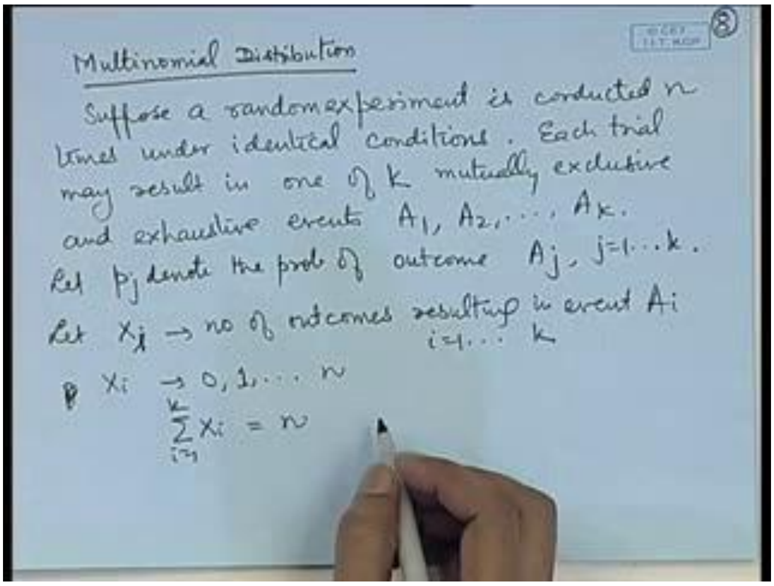
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 42
Additive Properties of Distributions – II

We will consider a few multivariate distributions, which are quite commonly used, one of them is a generalization of binomial distribution

(Refer Slide Time: 00:35)

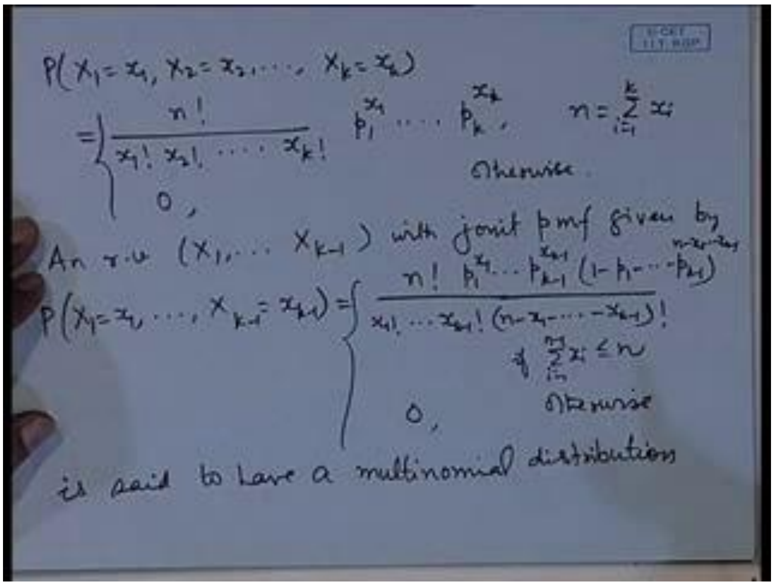


The so called multinomial distribution, in the binomial distribution we are considering a sequence of Bernoullian trails in which each trail of the experiment results in two options one is called success and another is called a failures; that is two types of outcomes are possible. However, there are a variety of trails in which we may be interested in categorizing not only in two, but in k type of outcomes So, for example, if you are looking at tossing of a die then you have the phases coming up 1, 2, 3, 4, 5, 6.

If you are looking at drawing a card from a pack of cards then it could be any of the fours, you would say heart, spade, club or diamond or if you are looking at say whether what is the number on the that is 1, 2, 3 up to 13 So, there are a variety of experiments where the possible outcomes can be more than one. So, if their probability of ending up in outcome one is say p 1, ending in outcome two is p 2, getting an outcome k is p k and then if we conduct a certain number of trail say n, so out of that say x 1 is the number of outcomes resulting in first type, x 2 is the number of trails resulting in the second type of outcome etcetera.

What is the distribution of that, so that is called a multinomial distribution, so suppose a random experiment is conducted n times under identical conditions. Each trail may result in one of k mutually exclusive and exhaustive events. Let us call them say A 1, A 2, A k; let p j denote the probability of outcome A j for j is equal to 1 to k. So, let us consider say X 1 number of outcomes resulting in event say A i; X i denotes this for i is equal to 1 to k. Then what are the possible values of X i; X i's can take value 0, 1 to n subject to the condition that sigma X i is equal to n because n is the total number of trails.

(Refer Slide Time: 04:38)

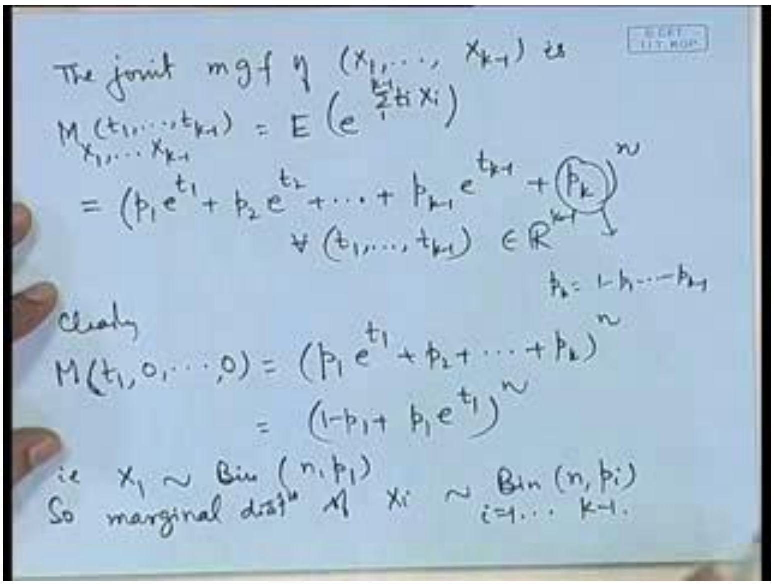


So, if we write down probability of say x 1 is equal to x 1, x 2 is equal to say x 2, x k is equal to x k then this is equal to n factorial divided by x 1 factorial x 2 factorial and so on x k factorial, p 1 to the power x 1, p 2 to the power x 2 p k to the power x k where n is equal to sigma X i; i is equal to 1 to k, it is equal to 0 otherwise.

Now if you look at these distribution here, if x 1, x 2, x k minus on that is any k minus one of these variables are given, the last one can be determined in terms of n minus the sum of the remaining once So, if we consider the joint distribution of a random variable say X 1, X 2, X k minus 1 with joint probability mass function given by this, so we consider it as probability of X 1 is equal to x 1, X 2 is equal to x 2, X k minus 1 is equal to x k minus 1; that is n factorial divided by x 1 factorial, x 2 factorial, x k minus 1 factorial; n minus x 1 minus x 2 minus x k minus 1 factorial; p 1 to the power x 1 and so on; p k 2 minus 1 to the power x k minus 1 and then 1 minus p 1 minus p 2 minus p k minus 1 to the power n minus x 1 minus x 2 minus x k minus 1.

If sigma x i; i is equal to 1 to n minus 1 is less than or equal to n 0 otherwise; this is said to have a multinomial distribution. So in fact if you look at these two, they are the same so, but formally we define a multinomial distribution to be k minus 1 dimensional because the last value is determined automatically; like in the binomial distribution, we talk about the distribution of the number of successes, we do not say that distribution of number of successes and failures.

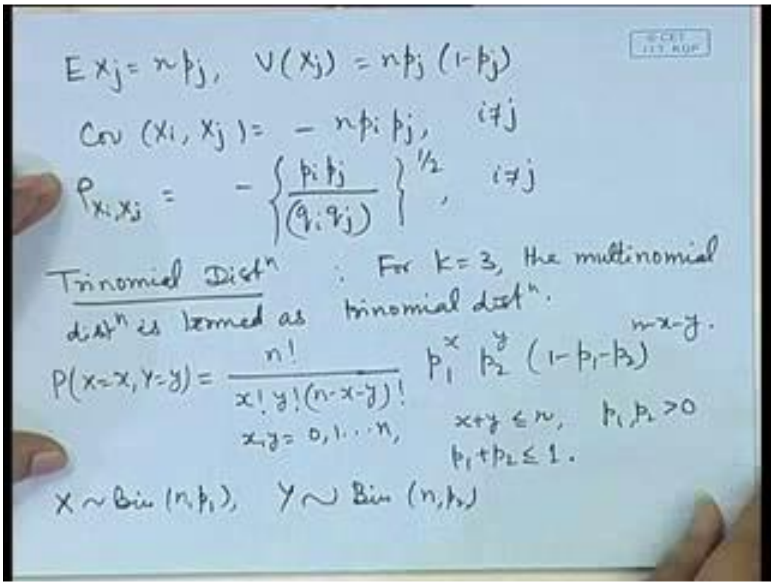
(Refer Slide Time: 07:32)



Now from a multinomial distribution, we can consider the joint mgf, the joint moment generating function of x 1, x 2, x k minus 1. So, it is evaluated at the point t 1, t 2, t k minus 1; that is expectation of e to the power sigma t i; X i from 1 to k minus 1. So, if we look at this distribution here now the sum of this is a multinomial that is p 1 plus p 2 plus up to 1 minus p 1 minus p 2 minus p k minus 1 to the power n. So, the sum of this over all this combinations is actually a multinomial sum, so if we want to calculate this term; it will become p 1; e to the power t 1 plus p 2 e to the power t 2 and so on plus p k minus 1 e to the power t k minus 1 plus p k to the power n and this is valid for all t 1, t 2, t k minus 1 belonging to R k minus 1.

Now easily you can see that suppose I substitute t 2, t 3 up to t k minus 1 is equal to 0; I will get p 1 e to the power t 1 plus p k to the power n which will become actually the; so this p k is actually 1 minus p 1 minus p 2 minus p k minus 1. So, clearly we can see that M t 1, 0, 0, 0 is equal to p 1 e to the power t 1 plus p 2 and so on plus p k to the power n that we can write as 1 minus p 1 plus p 1 e to the power t 1 to the power n; that is X 1 follows binomial n p 1; that means, the marginal distributions of X i's are binomial n, p i for i is equal to 1 to k minus 1.

(Refer Slide Time: 10:09)

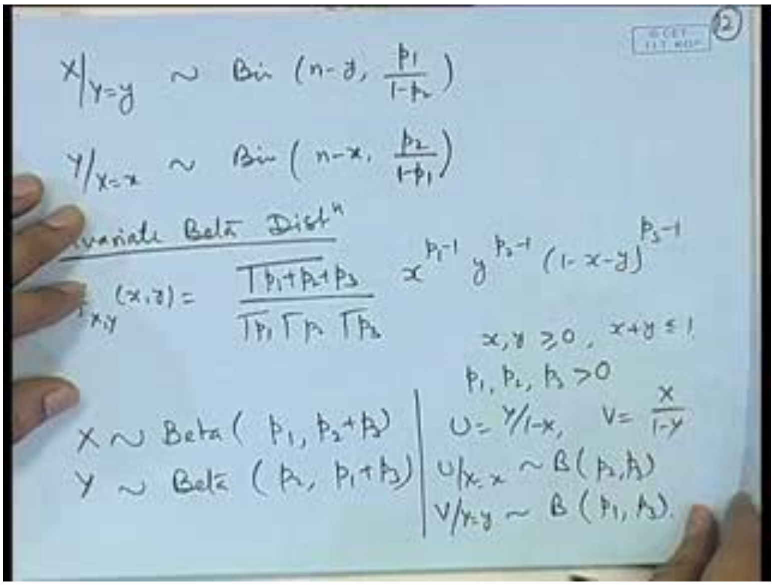


In particular we can talk about expectations so naturally expectation of X j will be n p j variance of X j will be n p j; 1 minus p j. We can also talk about the covariance terms between X i and X j that will be minus n p i p j for i not equal to j and therefore, correlation coefficient between X i and X j can be calculated to be minus p i p j divided by q i q j where q i q j denotes 1 minus p i and 1 minus p j etcetera to the power half for i not equal to j.

So, the correlation coefficient between two of this can also be calculated; in particular if in the multinomial distribution, I consider k equal to 3. So, I will have two of these variables that is x 1, x 2 that distribution is called a trinomial distribution, So, that is a state forward generalization of binomial distribution to the case when we are having three categories as the outcomes. So, it is called trinomial distribution that is for k equal to three the multinomial distribution is termed as trinomial distribution So, if I say trinomial distribution we can write the probability mass function as n factorial divided by x factorial, y factorial, n minus x minus y factorial p 1 to the power x, p 2 to the power y; 1 minus p 1 minus p 2 to the power n minus x minus y.

Here x and y can take values 0, 1 to n subject to the condition that x plus y will be less than or equal to n and of course, p 1, p 2 are greater than 0 subject to the condition that p 1 plus p 2 is less than or equal to 1.

(Refer Slide Time: 12:56)



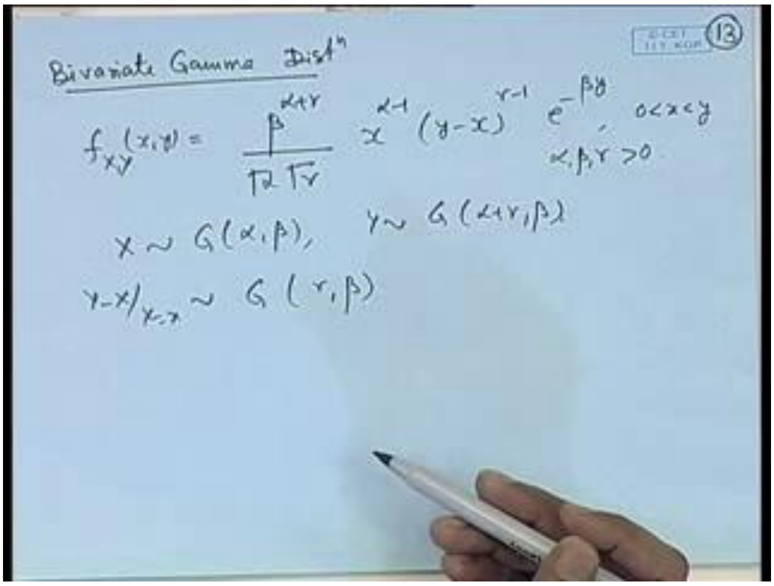
So, here if I look at the marginal distribution of x that will be binomial n p 1 if I look at the marginal distribution of y that will be binomial n p 2. Not only that if we look at the conditional distributions of X given Y and y given x, then conditional distribution of X given Y is binomial n minus y p 1 by 1 minus p 2 and Y given X has binomial n minus x; p 2 by 1 minus p 1. Of course, when we write this p 1 by 1 minus p 2 and p 2 by 1 minus p 1, we are assuming that the number is between 0 to 1.

So, this is one particular bivariate distribution, which is trinomial distribution and a general multivariable distribution that is a multinomial distribution, so it is a generalization of the univariate binomial distribution. We have a couple of more generalizations for example, we have done beta distribution; so a beta distribution can be generalized as a bivariate beta distribution in the following way; f x y as gamma p 1 plus p 2 plus p 3 divided by gamma of p 1, gamma of p 2, gamma of p 3; x to the power p 1 minus 1, y to the power p 2 minus 1, 1 minus x minus y to the power p 3 minus 1, where x and y are greater than or equal to 0 and x plus y is less than or equal to 1; p 1 p 2 p 3 must be positive.

Here, we can see that the marginal distribution of x is beta with parameters p 1 and p 2 plus 3; in a similar way marginal distribution of y can be calculated it is beta with parameters p 2 and p 1 plus p 3. The conditional distributions are also beta with a little a scaling for example, if I consider U is equal to Y divided by 1 minus X or V is equal to X divided by 1 minus Y, then u given X is equal to x follows beta distribution with parameters p 2, p 3; if we consider V given y that follows beta distribution with parameters p 1 and p 3. In no way this is a unique generalization of a beta distribution to do two dimension, we can generalize in different ways also.

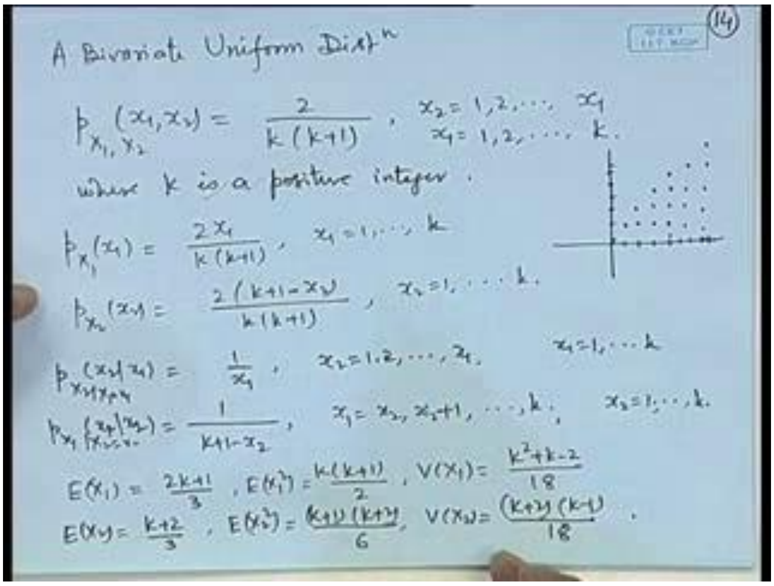
What we are trying to see here is that the marginal distributions also have beta distributions, so then in that case we are calling it has a bivariate beta distribution.

(Refer Slide Time: 16:02)



A very similar thing is done for gamma distributions; so a bivariate gamma distribution; it can be defined as f x y; beta to the power alpha plus gamma divided by gamma alpha gamma, gamma x to the power alpha minus 1, y minus x to the power gamma minus 1, e to the power minus beta y; where 0 less than x less than y alpha, beta, gamma greater than 0. Here if we see the marginals X follows gamma alpha beta and Y follows gamma alpha plus gamma and beta, also Y minus X given x, this follows gamma gamma beta. So, this is another generalization of a univariate gamma distribution to a bivariate gamma distribution. We consider a bivariate uniform distribution; it is a discrete uniform distribution.

(Refer Slide Time: 17:14)

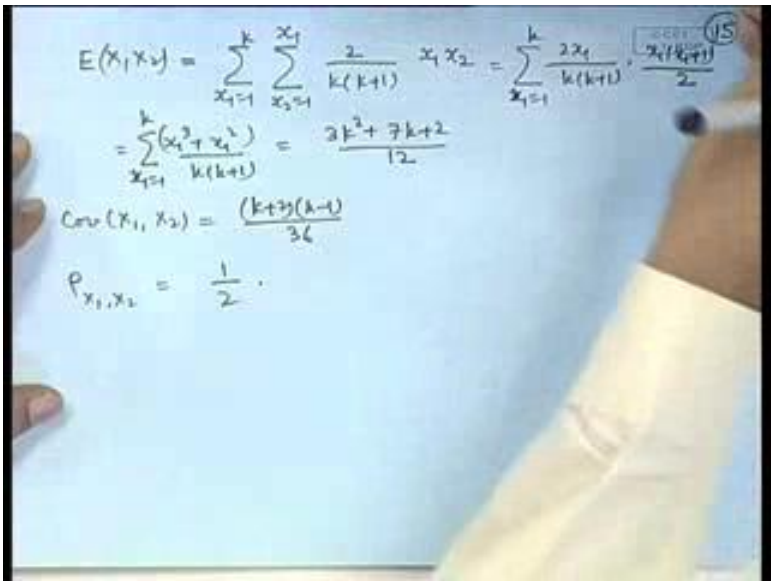


So, consider probability of X 1 is equal to X 1, X 2 is equal to small x 2 has 2 by k into k plus 1 where x 2 takes values 1 2 x 1 and x 1 takes values 1 to k. For example, if we consider the points 1, 2, 3 up to k then if I take x 1 is equal to 1, then x 2 will take value 1, if I take x 1 is equal to 2; then x 2 can take values 1 and 2 that is 1 and 2, if I take x 1 is equal to 3 then x 1 can take value; x 2 can take values 1, 2 and 3. So, you can see that this distribution is a discrete uniform distribution with probabilities concentrated on this diagonals; this you can say it as the half of the square actually. We can easily see that the marginal distributions if we sum over x 2 from 1 to x 1, this will give 2 x 1 by k into k plus 1 for x 1 is equal to 1 to k.

So, the marginal distribution of x 1 is obtained like this. In a similar way; if we sum over x 2, sum over x 1 from x 2 to k then we get the marginal distribution of X 2 is 2 into k plus 1 minus x 2 divided by k into k plus 1; for x 2 is equal to 1 to k. Important thing here to notice here is that if I take the conditional distributions of x 2 given x 1, it is a discrete uniform distribution on 1 to up to x 1. Similarly, the conditional distribution of X 1 given X 2 is a discrete uniform distribution univariate 1 by k plus 1 minus x 2 that is range from x 2 to k. So, this generalization of a bivariate of a discrete uniform distribution to two dimensions has the conditional distributions has discrete uniforms, but the marginal's are not uniform.

We can calculate certain moments about this distribution, which can be obtained from the marginal distributions like expectation of x 1, expectation of x 1 is square, variance of x 1 and similar characteristics for the x 2 variable.

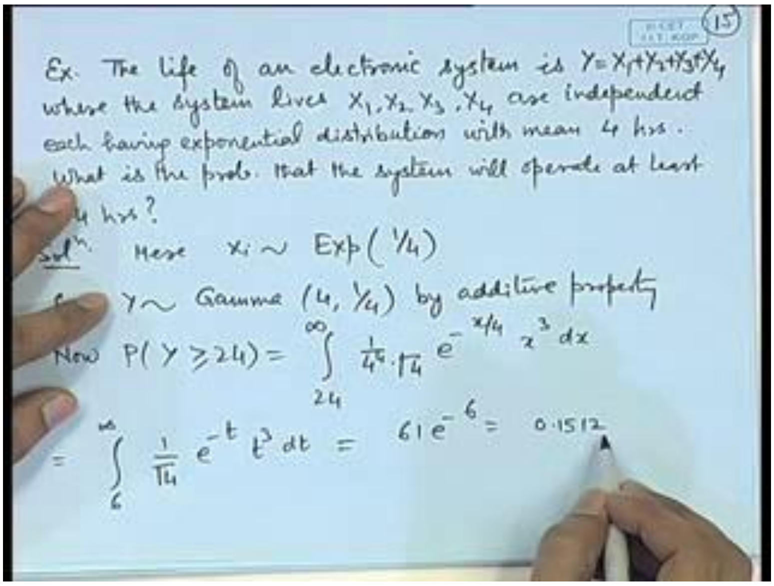
(Refer Slide Time: 20:19)



We can also look at the product moment here that is expectation of x 1 x 2 which is calculated from the joint distribution of x 1, x 2 and after certain simplification it turns out to be 3 k square plus 7 k plus 2 by 12. So, covariance of X 1, X 2 that is expectation of X 1, X 2 minus expectation of x 1 into expectation of x 2. So, after simplification this quantity turns out to be k plus 2 into k minus 1 by 36.

So, now if we divide covariance by the product of the square root of the variances, we get simply half because it is k plus 2 into k minus 1 by 18 both of this. So, the value turns out to be half, so the correlation coefficient between the random variables X 1 and X 2 is half, here we look at some application of the additive properties here.

(Refer Slide Time: 21:20)

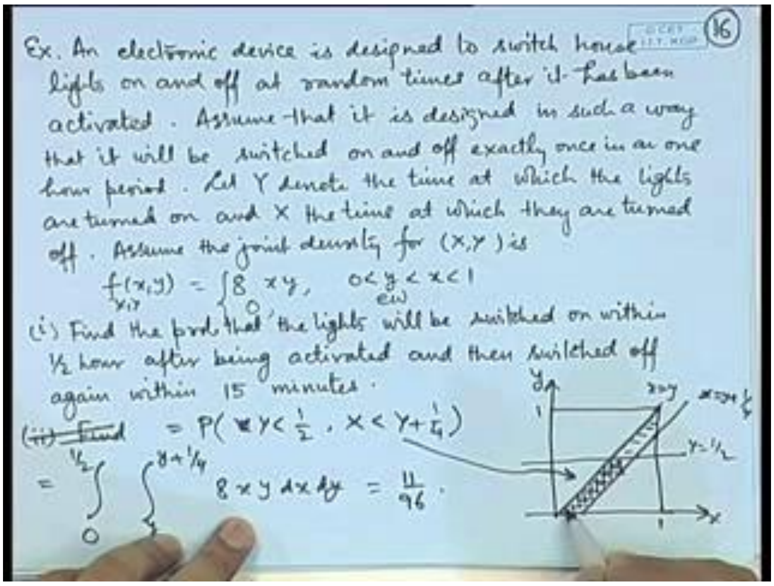


Let us consider here, suppose the life of an electronic system is described as the sum of four independent exponential lives. So, Y is X 1 plus X 2 plus X 3 plus X 4 and each of the exercise is exponential with mean life 4 hours.

So, what is the probability that the system will operate at least 24 hours; that means, we are interested to find out what is the probability of Y greater than or equal to 24. Now here we can use the additive property of the exponential distribution, we have proved that the sums of independent exponentials are following a gamma distribution. So, here each exercise exponential with parameter lambda is equal to 1 by 4; here mean is 4 that is parameter lambda will be 1 by mean because mean is 1 by lambda. So, y will follow gamma distribution with parameters 4 and 1 by 4. Now the density of a gamma distribution with parameters 4 and 1 by 4 is given by 1 by 4 to the power 4 gamma 4; e to the power minus x by 4, x to the power 4 minus 1.

So, we integrate from 24 to infinity, so after some simplification this value turns out to be 61 e to the power minus 6 that is 0.1512 that is there is a almost 15 percent of the chance that the system will be operating for at least 1 day.

(Refer Slide Time: 23:04)

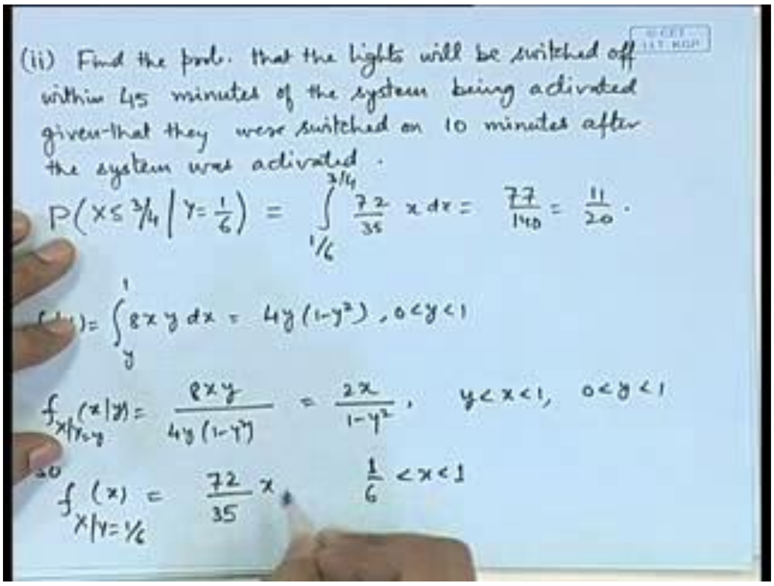


Now let us look at one more application of bivariate distributions, so consider an electronic device so it is designed in such a way to switch house lights on and off at random times after it has been activated. Assume that it has been designed in such a way that it will be switched on and off exactly once in a 1 hour period. Let Y denote the time at which the lights are turned on and X the time at which they are turned off; that means, firstly, it will be switched on and then there will be switched off and the joint density function for x y is given by 8 x y for y less than x are and of course, since we are considering only 1 hour period so both will lie between 0 to 1.

And it is 0 elsewhere; what is the probability that the lights will be switched on within half hour after being activated and then switched off again within 15 minutes; that means, what is the probability that Y is less than half and X is less than Y plus 1 by 4 because within 15 minutes of getting on, it should be switched off. So, X must be less than y plus 1 by 4; now to determine this probability we look at the region of integration of the density. So you see here the density is defined for in a unity square, the density is defined for y less than x that is this region.

Now, here Y less than half means that we are in the bottom region and X is less than y plus 1 by 4. Now x is greater than y; that means, here and the line x is equal to y is equal to 1 plus 4 is this line. So, we are basically in this zone, so we have to integrate the joint density 8 x y over this region so the limits of integration our for x y to y plus 1 by 4 and for y it will be 0 to half. So, this turns out to be 11 by 96.

(Refer Slide Time: 25:19)

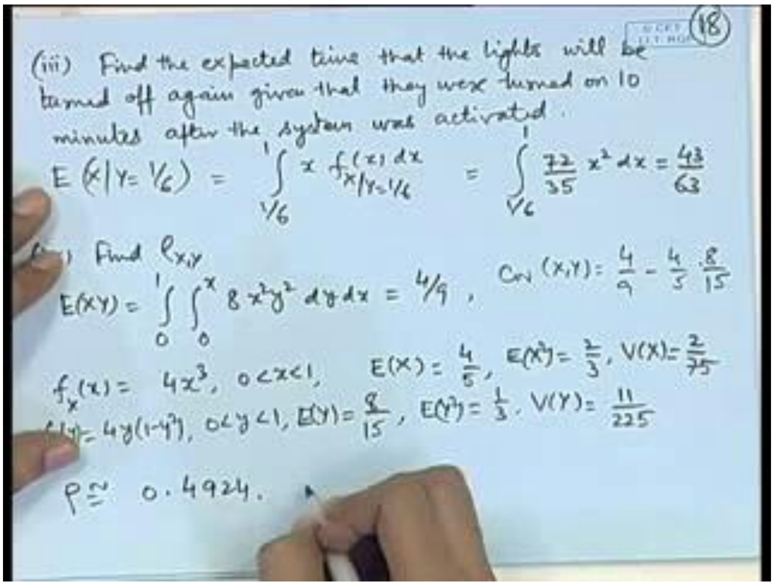


In the same problem, Let us look at what is the probability that the lights will be switched off within 45 minutes of the system being activated given that they were switched on 10 minutes after the system was activated; that means, what is the probability that X is less than or equal to 3 by 4; given that Y is equal to 1 by 6, so we need the conditional distribution of X given Y.

Firstly, we look at the marginal distribution of y, so here the joint distribution is 8 x y, we will integrate with respect to x. Now the region of integration for x is from y to 1. So, after integrating we get 4 y into 1 minus y square and therefore, the conditional distribution of x given y is evaluated as the ratio of the joint distribution divided by the marginal distribution of y, which turns out to be 2 x by 1 minus y square for x lying between y and 1; where y is a value fixed between 0 to 1. So, the conditional distribution of x given y is equal to 1 by 6 is easily evaluated by substituting y is equal to 1 by 6 here and we get 72 by 35 x; for x lying between 1 by 6 to 1.

Therefore, this conditional probability is obtained by evaluating the integral of this density over the region 1 by 6 to 3 by 4 and it is evaluated to be 11 by 20.

(Refer Slide Time: 27:02)



Find the expected time that the lights will be turned off again given that they were turned on 10 minutes after the system was activated; that means, what is the expected value of the distribution that we obtained just now; given Y is equal to 1 by 6; so the density is given here over this region so we calculate the expected value as integral of x into the density from 1 by 6 to 1 d x and it is evaluated to be 43 by 63.

Finally in this problem; what is the correlation coefficient between the random variables X and Y? So, in order to evaluate the correlation, we need the covariance term and the expectations of x and y and the variances of x and y. In order to evaluate the covariance term we need the product moment, so here expectation of x y is x in to y into the joint density that is 8 x y y is integrated from 0 to x and x is integrated from 0 to 1 which is 4 by 9.

The marginal distribution of x is obtained by integrating with respect to y from 0 to x which is simply 4 x cube So, expectation of x turns out to be 4 by 5 variance of x turns out to be 2 by 75. The marginal distribution of y was evaluated here as 4 y into 1 minus y square, so we can evaluate the mean and the variance of y also. Therefore, we can find the covariance between X, Y as 4 by 9 minus 4 by 5 into 8 by 15 and the correlation turns out to be this divided by the square root of the variances of x and y; which is 0.49 approximately; that is nearly half. So, in this particular problem the timings of switching on and off of the lights is having correlation nearly half year.

We have in general discussed the joint distributions, where we initially considered two variables, and then we considered multivariate that is k dimensional or n dimensional random variables. We have discussed the concept of marginal distributions, conditional distributions, the concept of correlations and we have also discussed certain additive properties of the distributions So, now in the next lecture; we will see that if we consider transformations of the random vectors, how to obtain the distributions of that, so in the next lectures, we will be covering that topic.

Thank you.