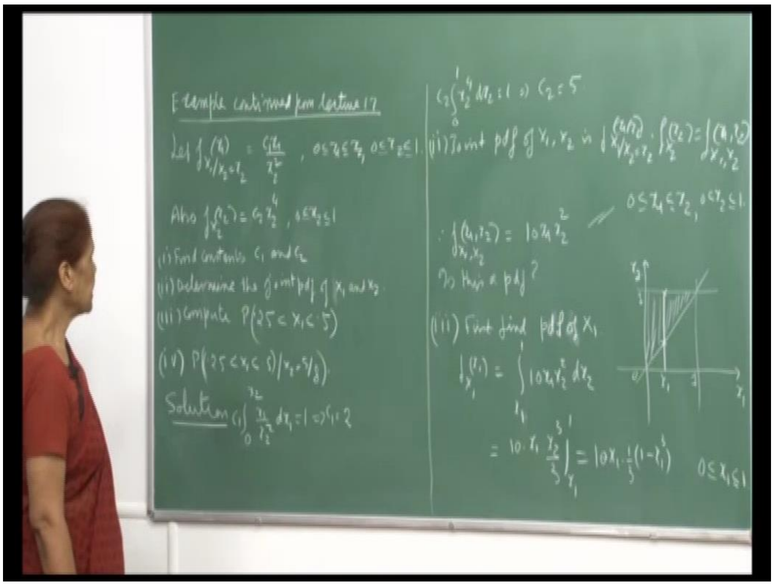
Introduction to Probability Theory and its Applications

Prof. Prabha Sharma
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 18
Conditional Expectation Best Predictor

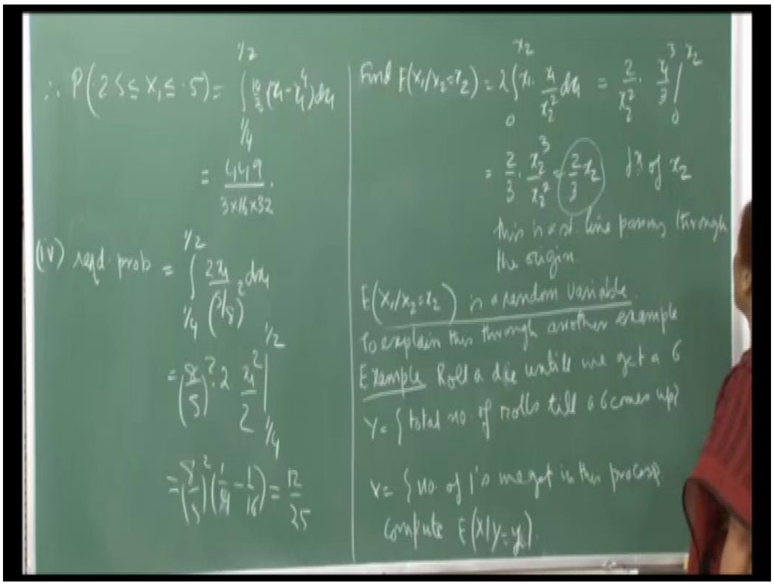
(Refer Slide Time: 00:15)



I will continue with the example, I was discussing at the end of lecture seventeen. So, this was the conditional p d f of x 1 2 given equal to x 2, and you are given the marginal of x 2 also. Then we had to find constant c 1 and c 2; of course, the criteria's that they should be, they are p d f. So, the integral in the specified region must be one. So, we made these computations in the in last lecture. So, by saying this integral 0 to x 2, because given n and x 2 when you draw this, your x 1 will be varying around 0 to x 2, because this is the region of integration. So, 0 to x 2 integral, this equal to 1 implies c 1 is 2. Similarly by integrating this from a 0 to 1, because of this marginal of x 2, then this comes out to this gives us c 2 equal to 5. So, the joint p d f of x 1 and x 2, will be conditional of x 1 given x 2 into the marginal of x 2. So, the product of the 2, which we already have now with us, since we have computed c 1 and c 2. So, that will be the joint p d f of x 1 and x 2, and that we also computed as 10 x 1 2 square, where the range for x 1 from 0 to x 2 and x 2 varies from 0 to 1. Now, is this a p d f surely, because it is a product of 2, this we have verified as a p d f, this we have verified as p d f, so the product must also be a p d f. So, there is no need to verify this again, though if you want you can integrate, respect to x 1 from here to here, and for x 2 0 to 1, and you can show that this is a integrate, so double integral will come out equal to be 1.

So, now we have to find in that number three, is compute the probability of x 1 between point 2 five and point five. So, to do this, I need to compute the marginal of x 1 marginal p d f of x 1, and that will be taking the joint p d f, and here you will be integrating respect to x 2, so given an x 1 this is the line, and therefore, your x 2 will vary from x 1 to 1. So, this is the range for x 2. So, x 1 to 1 you integrate the joint p d f, to obtain the marginal p d f of x 1, and so this comes out to be. So, you integrate in respect to x 2. So, x 2 cube by 3 10 x 1, so this is the a expression for the marginal. So, once you have the marginal for x 1, then you want integrate again the marginal from 1 by 4 to 1 by 2, to obtain the probability, and the number that I get is this. So, maybe you need to simply this further, and get the right answer. Then fourth one required you to find the conditional probability, of x 1 given x 2 is 5 by 8, so this is for again 0.25 to 0.5.

(Refer Slide Time: 03:28)



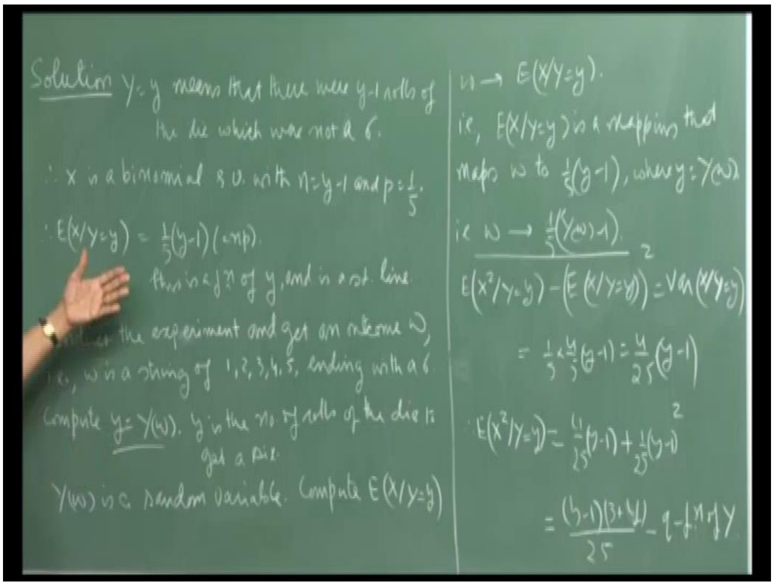
So, the conditional density function we already have; which is c x 1 upon x 2 square, but x 2 is given to 5 by 8. So, we have to integrate this from 1 by 4 to half 2 x 1 upon 5 by 8 whole square d x 1. So, here this is the simple these things x 1 square by 2 which is 1 by 4 to half, and what you get is, 12 by 25, but now since we have also talked of conditional expectation, I thought we will include that part also here. So, for example, you are asked to find the expected value of x 1 given x 2 is equal to x 2. So, you will find out the expectation here; that means, it will be the joint p d f it will be the conditional p d f of x 1 given x 2 equal to x 2, which is 2 times x 1 upon x 2 square, and since you are finding the expectation here with respect to x 1, so it will be 1 into this. So, this is the integral, and of course, 1 varies from 0 to x 2. And you are integrating with respect to x 1. So, 2 by x 2 square come outside, and then this will be 1 square, so which is x 1 cube by 3 and 0 to x 2.

So, you simplify and this is the expression 2 by 3 x 2 which is a function of x 2. So, which is to be expected, because in this integral you are integrating with respect to x 1, and the limits of from 0 to x 2, and since x 2 is given to be 2. So, this will turn out to always be a function of x 2, when you are finding expectation of x 1, given x equal to x 2, which I mentioned yesterday also we defined. Now this is a straight line passing through the origin. So, the expectation that comes out is a function of x 2, and this is a straight line passing through the origin. So, that means, here the relationship is linear; that means, the expectation of x 1 given x 2 equal to x 2, is a linear function of x 2. then therefore, follows that expected value of x 1 given x 2 X 2, will be 2 by 3 x 2, will be a random variable. See the whole idea here is, that this was repeated earlier also, what we are trying to say here is, that for a particular value of x 2 this is what you get, the expectation value of x 1, slash conditional x 2 equal to small x 2.

Now, as values of x 2 vary, then this becomes the expected value of x 1, condition on the random variable 2. So, the notation that we called it 2 by 3 X 2, so this is a random variable. So, for the different particular values of x 2, we will get the expected value from this formula, this is the whole idea right. And therefore, since this is now a random variable, we can again talk about the expected value of here, and I had done this even earlier, in the last lecture also the same thing. So, therefore, the expectation of x 1 condition on x 2, will be just the 2 by 3 into a expectation of x 2, which will by the formula would be, you know integral 2 by 3 0 to 1 2 f x 2 x 2 d x, and this will turn out to be the expected value of x 1. So, when you take first expectation of x 1 condition on x 2, and that comes out to be a function of x 2, then you take expectation again, and then you will get the expectation of x 1. So, this is idea being repeated in the last lecture, and I have asked you to verify. So; that means, you will have to compute the marginal of x 1, and then which I think you already here; yes that say conditional p d f. So, you will have to compute the marginal of x 1, and then find out the expected value of x 1 independently, and verify that this it comes out to be the same, what will get from here. Further I will continue with this concept, and take another example to make these things clear, try to do it.

So, this was an example I got from the net; roll a die, until we get a 6. So, this is the experiment. You continue roll a die, until you get a 6. And let y be the random variable which is equal to the total number of rolls, till a 6 comes up. So, you continue rolling the die till a 6 chose up and then you stop. So, and x is the number of, once we get in this process. So, when you are rolling a die, you keep noting also the number of times 1 appears, and then of the experiment stops the moment a 6 comes up. So, x is the number of once you gets in this process. Now, you are asked to compute expectation of f x given y is equal to y; that means, the number of times you have to roll the die, is small y, and then this process you want to compute the expectation of x, given that you had to roll the die y number of times.

(Refer Slide Time: 08:38)

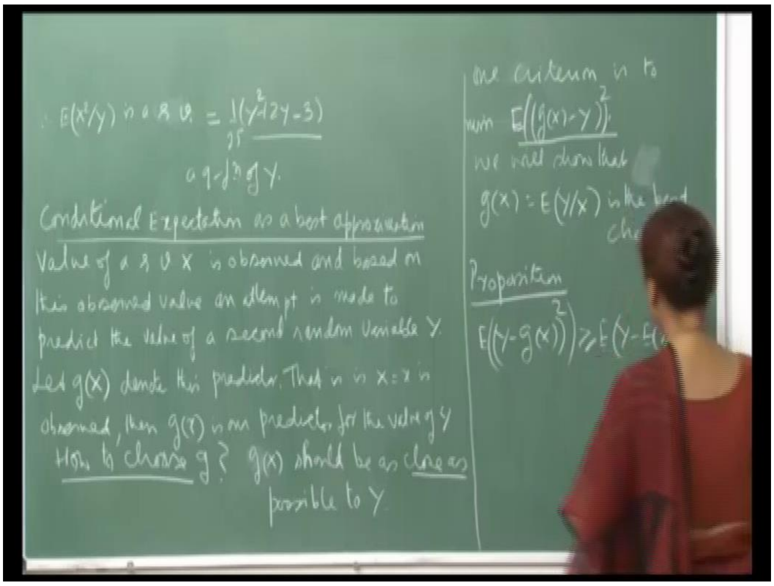


So, Y equal to small y means; that there were y minus 1 rolls the die, which one not a 6. All the numbers are appeared, but the 6 did not appear, because the experiment will stop the moment 6 appears. So, y minus 1 rolls did not show a 6. Now, therefore, you see an x is the number of once that comes up, before the experiment ends. So; that means, in this showing that you get, the y minus 1 rolls that you made of the die; x is the number of ones that show up. So, when you can treat x as a binomial random variable, with the number of experiment as a y minus 1, and the probability of occurrence of a 1 is 1 by 5, because remember up to y minus 1 rolls 6 is not appearing. So, and the other five numbers are equally likely. So, the probability of 1 showing up, is 1 by 5, and the number of experiments that you make is y minus 1. So, we will treat occurrence of 1 is success, and occurrence of 2 3 4 5 as a failure, and therefore, x can be treated as a binomial random variable, with n equal to y minus 1 and p equal to 1 by 5. So, you see the whole idea, is that how you can compute a certain quantity that is required, by just realizing how the experiment has been conducted. So, therefore, immediately you can say expected value of x when y is given to be small y, is slimily because for a binomial random variable, the expected value is n p. So, n here is y minus 1 and p is 1 by 5. So, without any hassle, we get to the expected value; the conditional expected value of x given y is equal to y.

Now, this is again a function of y; that is as y takes values possible values 1 2 3 and so on, the points 1 by 5 y minus 1 will lie on this straight line, expected value of x by y; that means, conditional expectation of x given y capital Y is a random variable. Now conduct the experiment and get an outcome omega; that is omega is string of 1 2 3 4 5, ending with a 6. So, like we conduct the experiment we keep rolling a die, till we get a 6, and we record the outcomes at each roll of the die. So, it will be string of these numbers, ending with the 6, so omega is that. Then you compute how many numbers are there in the string, so that means, how many times we have to roll the die. So, now you can say that y small y is the value of Y at omega, because omega represents a string, and y counts the number of elements in this string. So, this is actually our relationship small y is Y at omega. So, y is number of rolls of the die to get a 6. So, y omega is a random variable, surely because this can go on depending. I mean when the 6 shows that is not a certain event, so there is a chance element here, and therefore, y omega is a random variable, and then you compute the expectation of x given y equal to y, which we did right now. so; that means, you are relating omega with this, because given a omega you computed the y, and then you compute the expected value of x given y equal to y.

So, that is expectation of x given y equal to y is a mapping; that maps omega to 1 by 5 y minus 1, and remember I defined a random variable also is a mapping, because as I said random variables associate, with the sample base real numbers. So, here also what you have to happening is that, this expected value of x given y equal to y is a mapping that maps omega to this, where y is capital Y omega. So, therefore, omega is mapped to 1 by 5 y omega minus 1, which is a random variable. So, now when you write capital Y omega. So, the idea was to explain to, again in a different way why we are saying that this is a random variable, though it should be clear, because as values of y change. There is a probability associated with what value y takes, and therefore, this is again a random variable. You can go ahead and makes some more computations; for example, if you look at variance. So, again the conditional variance of x given y equal to small y, I know because I have said that x is a binomial random variable, so this number I know, and for this number also I know. So, if you want to compute expectation of x, is square given y equal to y, then this minus this is equal to the variance. And therefore, I can say that, yes and this is equal to n p q, by the binomial formula, so p q and n. So, 1 by 5 into 4 by 5 into y minus 1; so 4 by 25 y minus 1 is the variance, and expectation x given y equal to y, we already computed as 1 by 5 of y minus 1. So, therefore, expectation of x square given y equal to y, is variance plus this, which is 4 by 25 my y minus 1 plus 1 by 5 y minus whole square, which is 1 by 25 into y minus 1 whole square. So, when you simplify this expression you get this y minus 1. So, this is a quadratic function of. I should say here quadratic function of small y.

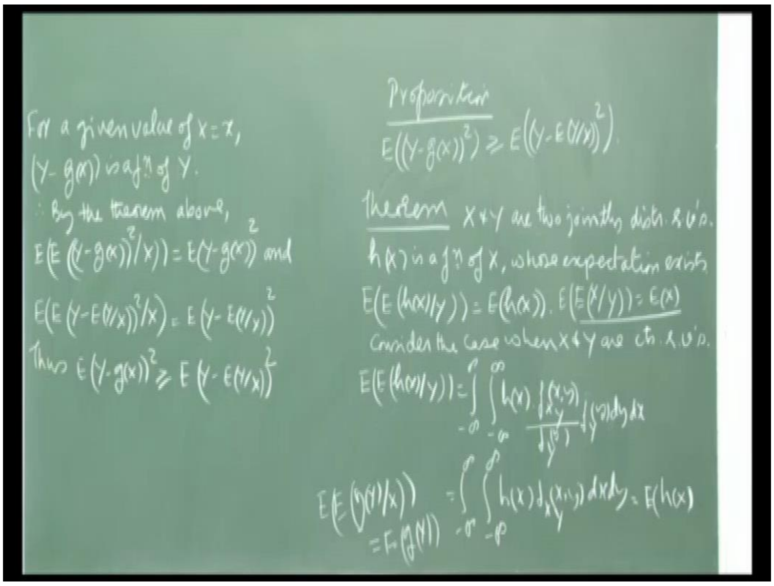
(Refer Slide Time: 15:00)



So, the random variable represented by expectation x is square given capital Y, is a random variable, which is this. So, this becomes a quadratic function of y. So, I hope this gives you a little inside into, what we mean by. And now another roll that a conditional expectation place, is as a best approximation. Conditional expectation as a best approximation, and I thought we should talk about this to, give you some more feeling, and this is value of the random variable is observed. So, suppose a value of random variable is observed, and based on this observed value, an attempt is made, to predict the value of a set second random variable y. see sometimes maybe easier to observed value of the certain random variable, and then if based on that observed value, you can makes some attempt to predict value of another random variable, that helps you, because then you do not have to conduct another experiment, to obtain value of y, but of course, the value that you predict from knowing the value of x, may not be the exact one. So, let g x denotes this predictor. So, suppose, so some function of the random variable x. So, this denotes this predictor, so that is x equal to x is observed, and then g x. So, g of small x, is our predictor for the value of y, so for one value of y, and then you observed another value of capital X and g of, that value of x will give you another a predictor for the value of y, so this is the idea.

Now, how to choose g, because you have to have some concept, as to what is the g that is acceptable to you; of course, the quality that g must be possesses, that it should be as close as possible to y. So, now, whole idea is, what do you mean by closeness here, and how can we define. And of course, later on also, when we talk of limits and so on, and conversions and probability in law, these things become more clear. But right now, that we say that our criteria criterion is to minimize, expectation of g x minus y whole square. So, this is my criterion, then I want to choose that g, which minimizes this expectation g x minus y whole square. So, expectation of this should be as small as possible. Now, we will show that g x equal to expectation y condition on x is a best choice. So, this is a whole idea, and therefore, you see another roll that the conditional expectation place, and the proposition is. So, we want to show, that expectation y minus g x whole square, is greater than or equal to expectation y minus expectation y given x whole square. So, then this will establish, that this is a smallest value of this, and therefore, the best choice for g is expectation y given a conditional for expectation y given x. So, we will just prove this proposition for you.

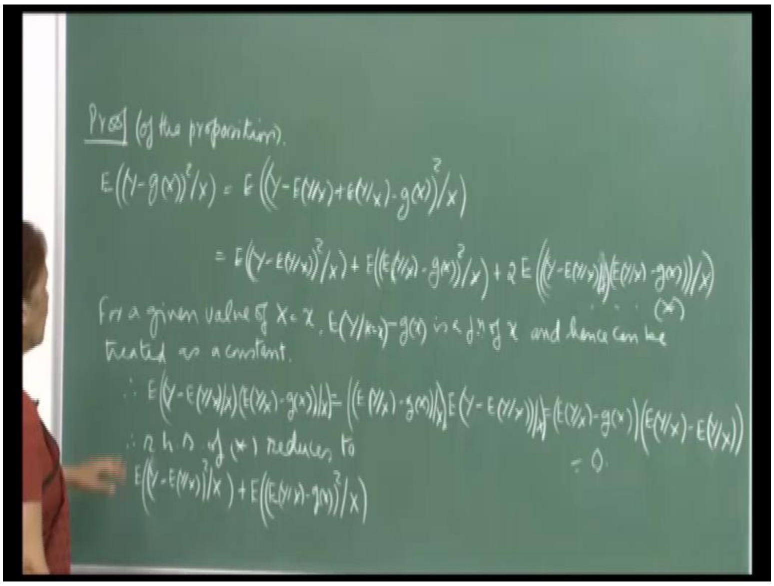
(Refer Slide Time: 18:37)



So, before we proved this proposition, I will like to state a theorem, and fact that proof is also straightforward. So, the theorem says that if x and y are 2 jointly distributed random variables, and h x is a function of x, whose expectation exists. So, then if you have the conditional expectation of f x given y and then we take the expectation again, this will come out to be expectation of h x. Now this is on the same lines as if you say that expectation of x given y, and then you take the expectation again. So, this is E x, remember we have already shown this result. So, on the same lines we are trying to show that, expected value of h x conditioned on y, and when you take the expectation again it will come out to be expectation of h x provided of course, the expectation of h x exists.

So, I have just written down for the continuous case. I have just written down the expression for x expectation of the expectation of conditional of h x on y, then it will be minus infinity to infinity minus infinity to infinity h x. the conditional p d f would be f x y x y divided by f y, because this condition on y, and then since you this turn out to be a function y. So, when I take the expectation again it will be f y y d y d x, and this you can see that, this will cancel out, and it will get expectation of h x. Now, similarly the result can be stated, that if you take a function g of y and then condition on x, then just reverse the rolls of x and y, and then you will get here a expected value of g y; that means, your expected g of y condition on x, this will turn out to be expected value of g y. provided of course, a expectation g y exists. So, this is this and then proof, in case x and y are discrete random variables, you can just imitate the proof for the continuous case. So, once this theorem is there, we will be using it in to prove this proposition. So, let me now consider the proof of the proposition.

(Refer Slide Time: 20:59)

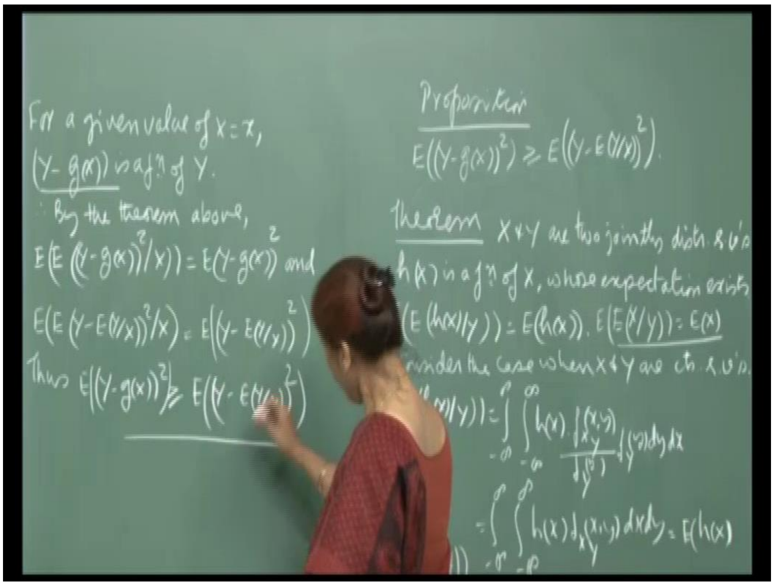


So, this is I am considering y minus g x whole square, condition on x, and because my proposition I am taking condition on x. So, now, add and subtract expected value y by x you was in the bracket. So, then this is expected value of y by x plus expected value of y by x minus g x whole square condition on x, open up the brackets. Then this will be y minus E x y by x whole square, condition on x expectation of that plus the square of this expected value of y by x minus 0 whole square. Then condition on x expectation plus twice a product of these 2 terms. So, say the condition on x separately, and then the expectation of that. So, twice E of that. So, then let me call this, equation as a star, denoted by star

Now, see for given value of x equal to x, this will be a function of x. remember we have shown you that the expectation of y condition on x. So, for a fixed value of x it will be a function of x. So, for the purpose of taking this expectation, because that expectation with respect to, ya this will be ya. So, therefore, I can treat this as a constant, and so this comes out to be expected sign. So, this is a function of x and hence can be treated as a constant. So, because I will be computing it for different values of x, this thing, and for every fixed value of x, this will be a constant. So, I can take it out of my expectation. So, therefore, y minus E of y by x condition on x into expectation of this, I can take this outside, and then this will be expectation y minus expectation y by x condition on x.

Now, here again, you see when you bring x inside, it will be y by x y condition on x and this does not change, because this is already on condition on x. So, this will become expectation of y by x. I mean this of course is outside. So, this portion will be expectation y by x minus expectation y by x, which is 0. So, therefore, the in the star expression this portion has no contribution, this is equal to 0. So, the right hand side, this reduces to simply the, expectation of y minus expectation y by x whole square condition on x, this expectation of expectation y by x minus g x whole square condition on x. So, this is what you have.

(Refer Slide Time: 23:52)

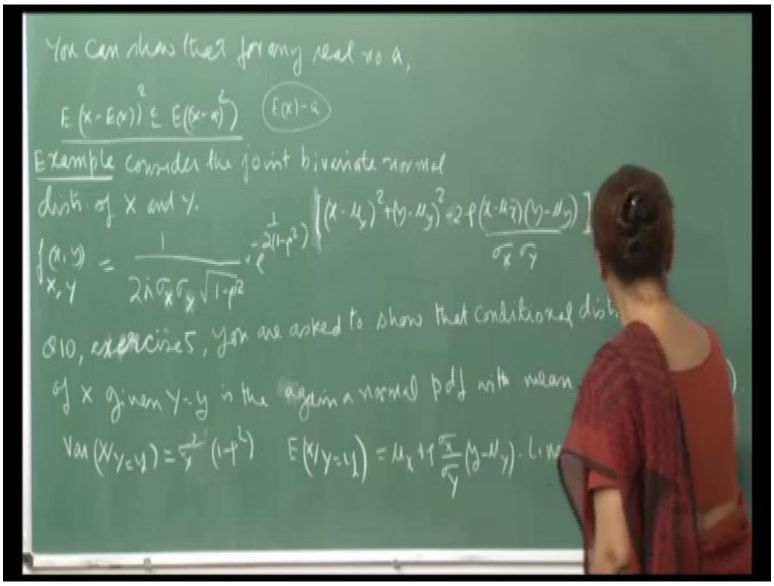


And since for a given value of x y minus g x is a function of y. Therefore, by the theorem above; this will be a function of y only, because this will be a constant for every fixed value of x. and so by our theorem, expected value of expected value y minus g x whole square condition on x. So, for different values of x, this will be a function of y only. So, for each fixed value of x, it will turn out to be expected value y minus g x whole square, and then for the whole different values of x to the whole thing will become y minus g x, expectation of y minus g x whole square. See it is a same concept scalding over. And similarly expectation of y minus expectation y by x whole square condition on x, will turn out to be expectation y minus expectation y by x whole square. I mean this whole square and then the expectation; this is what we are doing. So, you get that when you you know you have this because this is equal to zero. So, I got yes, where is the greater. So, the right hand side, when I take the expectation, I get; see we started out from here, you started out from here, and then this got 0, then I took expectation of both the sides.

So, this became expectation of y minus g x whole square, and this became expectation of y minus expectation of y by x whole square plus expectation y by x minus g x whole square. So, the conditional things disappeared, and since this expectation of squares are non-negative. So, therefore, I can this, because this portion; the expected value of y by x minus g x whole square, this will non negative, y by x minus g x whole square this will be non-negative. So, therefore, my equality will be convert into an inequality, and this what you have. So, this is what we are trying to prove, at you know for any function of g x for any function g of x, if you take this expectation y minus g x whole square. Here again if you want to put this, then this will always greater than or equal to expectation of y minus expectation y by x whole square, expectation y by, and this whole square expectation of that. I should first say y minus expectation y by x whole square and expectation of that. So, this is what we wanted to prove.

So, in order words you see the process, that you are observing value of x, and then computing the expectation y given x, and in the earlier lecture when I had considered, define this conditioned expectation, and conditional expectation and taken this discrete case. So, I showed you how, you change keep changing values of x, then you will compute the expectation y given x. So, this in a way we are treating as a good approximation for the value of y, and when we are when our criteria criterion is, in terms of minimizing this expression, then; obviously, no other function of x, will qualify to be the best predictor, except for the expectation y given x. So, in this way we are treating or showing that this can be also looked upon as a very good approximation, for the value of x. So, for the different values of x will compute this, I mean we will observe possible values of x. All possible values compute this, and this will give us the different values for of y. so; that means, you can compute this expectation, for. Well, yes there is, still a few questions which I are not answered, and hopefully we will continue, looking at this again.

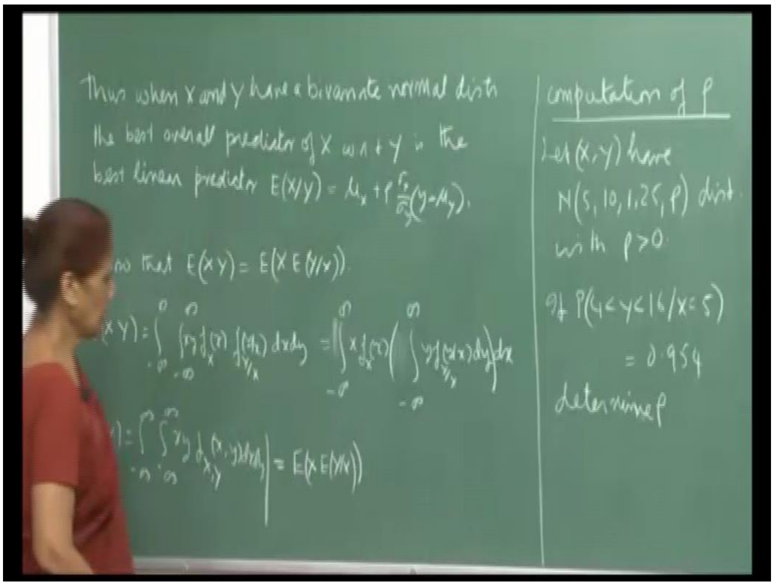
(Refer Slide Time: 28:11)



Now, take this example, you can show that for any real number a. In fact, what we have shown here, can be also. you know This you can show very easily, that expectation of x minus E x whole square, is always less than or equal to expectation of x minus a whole square, for any real number a. and you know the same thing here, you will write minus E plus E minus a, open up the square and so on, and then again one part, when you look at this thing E x minus a. So, this is the constant, and therefore, when you are taking the expectation of the product term, this will come out, and see the second factor will be expectation of x minus E x, which will be 0. So, for the same reasoning, you can show that this is always a less than or equal to this. So, you can do this for, any real number a.

Now, look at the, example of bivariate normal distribution of x and y, and we have already looked about this p d f, which is you know in terms of sigma x sigma y and rho. So, these are correlated, because; I mean I am just taking the general case, x and y are not necessarily independent. So, this expression, in exercise 5 of and question 10, I asked you to show that conditional distribution of x, given y equal to y, is again a normal p d f with mean this, and variance equal to rho square. Now why am I writing rho square, this should be sigma square. sorry So, this will be sigma square into 1 minus rho square. So, this is your mean, and that is your variance. So, therefore, expectation of given y equal to y, will be this mean, because this is for capital Y equal to small y. So, this is mu x plus rho into sigma x upon sigma y y minus mu y, which is again a linear function of y.

(Refer Slide Time: 30:28)



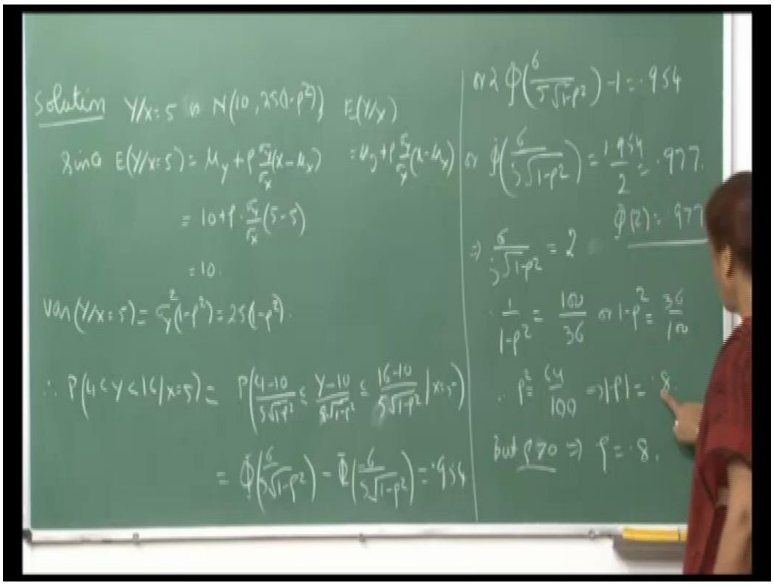
So, what we have found is that, when x and y have a bivariate normal distribution, the best overall predictor, because according to our proposition, of x with respect to y, turns out to be, mu x expected value of x condition on y, is mu x plus rho sigma x upon sigma y into capital Y minus mu y, and this turns out to be linear in y. So, that is all we can say that the best overall predictor, in case x and y are bivariate normally distributed, then the best overall bivariate of with x respect to y, is this expression which is the conditional expectation of x given y. It turns out to be this, and this is linear in y. So, this is all what we can say, or to talk of this linear product linear predictor and so on; that is another thing, that we have to go about it in a different way, and I have to first defined what we mean by linear predictor and so on. So, right now all we are saying, is that the best overall predictor, in case x and y are both normally distributed; that means, they have a bivariate normal distribution. Then the best overall predictor, would be a linear in y. And similarly if you are talking of the best overall predictor of y with respect to, and that will be linear in x.

Now, let us just further continue talking about a conditional expectation. So, now, we will show, that expectation x y is expectation x into expectation of x expectation of y given x. So, this is what a simple calculation we will do. Now here of course, I have not written out all the steps. So for example, when you want to write expectation x y, so actually the starting expression will be x y into the joint d x d y. So, this is the thing, but as we have seen that the joint p d f of and y, can be written as a product of marginal of x into the conditional of y given x. So, therefore, this is what I am writing here. So, expectation x y. I am writing as x y f x x into f of y given x, d x d y.

Now, the thing is that. So, y integral y into f y by x conditional this, conditional p d f of y given x. So, this I can separate out, and I can write this as y, because this is not of, when I integrate this is given value of x is fixed here; f y given x. So, X into small x so I integrate this respect to y. And then, so I can just separate out this double integral into minus infinity to the infinity x f x, and then this will come out to be a function of x. So, then that whole thing I will integrate as a function of a d x, and so you can immediately see that this is expectation y given x, and then x times f x this gives you expectation of x into expectation of y by x. So, simple calculation, but again just emphasizing the fact that, this is a random variable, and therefore, becomes a function of x, and so you compute this expectation again, and you get this. So, you know you were using a conditional expectation to compute expectation.

Now, let us go through this exercise for computing rho, and of course, for different situations you can use different techniques to handle it. So, now here if you are given x and y, have a bivariate normal distribution, so this is mu x mu y sigma x square sigma y square and rho. So, this is the bivariate normal distribution, and you are given that a rho is greater than zero. So, you have to find the conditional expectation of y between 4 and 15 given x is equal to 5. I am sorry I mean you are given that this probability is equal to 0.954 you have to determine rho. So, rho is the unknown here, and therefore, you want to determine that. Now, from our this result, that x given y is this. So, what will be this thing.

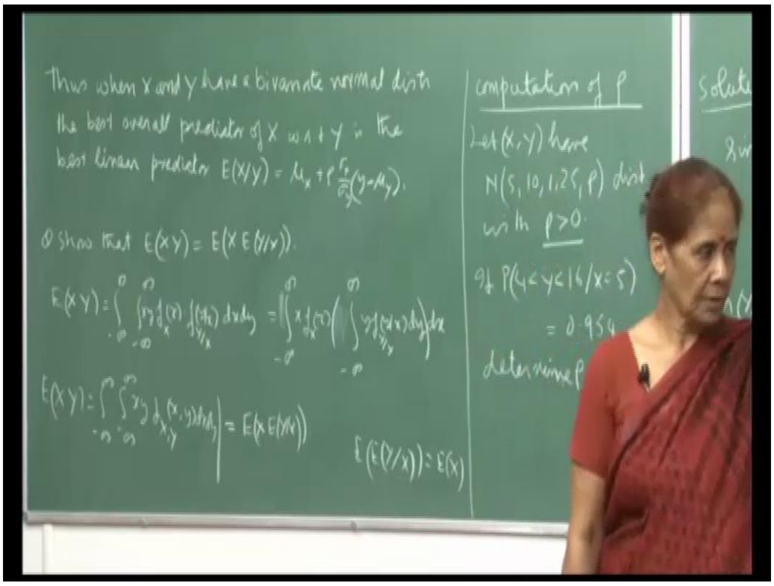
(Refer Slide Time: 35:21)



So, from here only you can write here, that expectation of y given x, this will be mu y. we just replace x y by y mu y plus rho sigma y by sigma x, and then this will be x minus mu x. So, this is the formula which, I have written it already. So, this is mu y plus rho times y upon sigma x into x minus mu x; that will be this. And then since x is given to be 5 and mu x is also 5. So, therefore, this part is 0, and so expectation y when x is equal to 5 is 10. So, the conditional expectation y or the mean is 10. So, y by x is normally distributed, with mean 10, and the formula for the variance would be. So, variance formula for the variance y given x equal to 5, is rho y square 1 minus rho square, which is 25 times 1 minus rho square. So, this is the variance, therefore, this is what I have written here. So, this is normally distributed, with mean 10 and variance 25 into 1 minus rho square. So, therefore, I will standardize, usual thing that we do. So, now, computing this probability, I will standardize the a variate here, which means I will subtract 10, and divide by the. So, divide by the standard deviation. So, standard deviation is 5 under root 1 minus rho square; this is what you have. So, this becomes a 6 by 5 under root 1 minus rho square. So, this is the c d f for the, probability less than or equal to this.

So, therefore, I mean the cumulative density function, for the standard normal. So, this minus this is minus 6 this, and again from symmetry of the standard normal distribution, around the origin. We can write this as twice phi of 6 upon 5 under root 1 minus rho square minus 1, because this can be written as 1 minus of phi of 6 upon 5 under root 1 minus rho square minus sign outside. So, therefore, this is what you get, and you are given this is equal to 0.954 or a 6 upon this, is equal to 1.954 by 2 which is 0.955. So, if you look up these standard normal tables, the value of z which corresponds to 0.977 is 2; that mean, phi of 2 is 0.977. So, this you can obtained from the tables, for the standard normal and. So, this means that this number 6 by 5 under root 1 minus rho square should be equal to 2. So, from here, you do a simpler arithmetic square everything, so you get 136 a 1 minus rho square is 36 upon hundred. So, rho square is this. So, the absolute rho is 0.8, but you are given a rho to be positive, and therefore, we will take a positive value from here, which gives you rho is 0.8. So; that means, x and y are positively correlated. If rho was minus 0.8, then we would say that and y are negatively. So, the relationship between this. Of course, throughout this we have also been able to establish, that your correlation coefficient or the covariance can measure effectively, linear relationship between the variables, but it fails to show you a quadratic relationship and so on, so we saw that.

(Refer Slide Time: 39:32)



So, this comes to sort of our treatment of a conditional expectation, and how this can be used, for computing various things. In fact, for computing your actual this thing, because we have shown other that this result also expectation, is expectation x and so on. And then here this result we have obtained for you, and then I have also shown you the roll of a conditional expectation as a best predictor, so this is it. Let me now discuss exercise 6 with you, which is related to what all we have discussed in the last three lectures.

(Refer Slide Time: 40:12)

1. A fair die is rolled at random three times, Let the random variable X_i be equal to the no. that appear on the i^{th} trial for i=1,2,3. Let $Y=\max(X_1, X_2, X_3)$. Find distribution function and probability density function of Y.

2. Show that correlation coefficient $\rho(X,Y)$, where X and Y are two random variable is $= \dfrac{Var(X)+Var(Y)-Var(X-Y)}{2\sqrt{Var(X).Var(Y)}}$.

3. X and Y have the joint probability density function $f(x, y) = 1,\ -x<y<x,\ 0<x<1.$
   $0,$ elsewhere.
   Draw the graph of $E(Y/X=x)$ as a function of x and also that of $E(X/Y=y)$ as a function of y.

4. Suppose the conditional probability density function of X_2, given $X_1=x_1>0$ is
   $$f_{(X_2/X_1=x_1)}(x_2/X_1) = \frac{(x_1+x_2)}{(x_1+1)}e^{-x_2}, x_2>0$$
   $$= 0, \text{ Otherwise.}$$
   And that $f_{X_1}(x_1) = \dfrac{(x_1+1)}{2}e^{-x_1}, x_1>0$
   $$= 0, \quad \text{Otherwise.}$$
   Find $E(X_1/X_2=x_2)$, $E(X_1)$ and Correlation (X_1, X_2).

5. Let (X_1, X_2) be a two- dimensional discrete random variables with joint probability function

	$P_{x_1,x_2}(X_1,X_2)$	
X_1 \ X_2	1	2
1	1/8	1/16
2	1/1	1/16
3	1/8	1/8
4	3/16	1/4

(i) Compute Correlation (X_1, X_2) (ii) Are X_1 and X_2 independent random variables?

So, a fair die is rolled three times. There should be a full stop instead of comma. Let a random variable x i be equal to the number that appears on the i th trial, for i varying from 1 to 2 3, and then we defined y as the max of x 1 x 2 x 3, so the largest numbers. So, up to on the die has been rolled 3 times, then you record the numbers that were that show up, and then you take the maximum one. So, y is the value of the maximum of x 1 x 2 and x 3. Find distribution function and probability density function of y. So, the distribution function means; cumulative distribution function, and probability density function of y. So, this is question one. Question seven consider the joint probability density function of x comma y, given by f x y of x comma y. So, here it should be x and y as suffixes are the bigger the capital one and small x y is equal to 2 minus x minus y x between 0 and 1 and y between 0 and 1 0 otherwise.

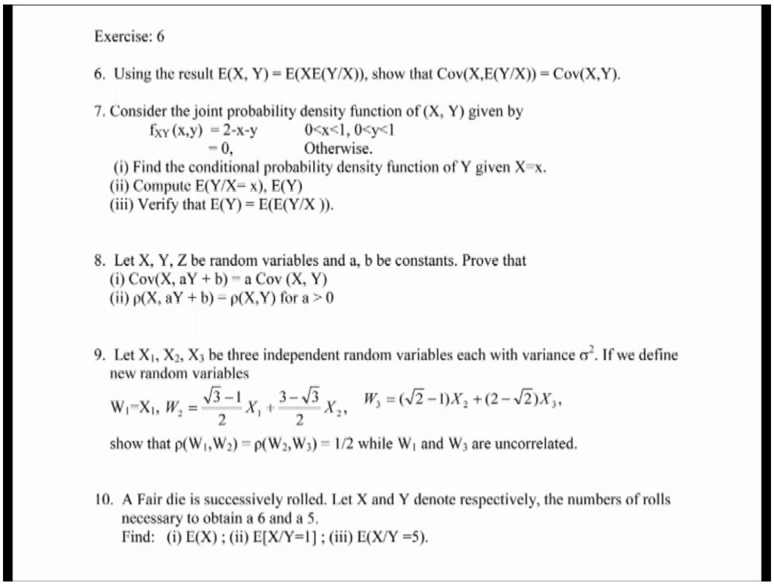
Now, here a part three, I want you to find. So, part 1 says, find the conditional probability density function of y, given X equal to small x, and then in part 2, I want you to find the expected value of y given capital x equal to x and E y. Then three I want you to find out E y, though. I want you to show that E y is actually equal to E conditional expectation of y given x. So, what we have been doing in a lecture also. We have been verifying, you know we have been computing E y independently, by first computing the marginal of y, and then its expectation, and secondly, by breaking it up into first the conditional probability conditional expectation of y given x, and then we take the expectation again. So, you have to say that the 2 processes redo the same answer.

Now, question two I already discussed with you in the lecture. So, you can have an alternate expression for the correlation coefficient, when x and y are given to be 2 random variables. Question three, is that x and y have the joint probability density function f x y equal to one. So, y lying between minus x and x and x between 0 and 1, so please be careful when you draw the boundaries for y, because you see 1 boundaries y equal to minus x and other is y equal x. So; that means, your y varies from. I hope you can just make sure that. So, you can, see along the x axis, you will have 1 line, y equal to and the other will be in the in the fourth quadrant y equal to minus x, and therefore, your y will be vary from minus x to x. So, this is what you have to careful.

And then you have to draw the graph of. So, given the joint density function, you will compute the conditional, because you have to draw the graph of expectation of y given x equal to x, as a function of x. So, you know how to do it. You have to compute the conditional p d f, and then compute the expectation y given x equal to x, and also you have to draw the graph of. So, find out both the conditional p d fs; conditional p d f y given x, and conditional p d f x given y as a function of y. So, you get some feeling about the. Question four. So, suppose the conditional probability density function of x 2 given x 1 equal to x 1, is given by this function. This is the conditional p d f, so x 2 positive, and its 0 otherwise. So, the reason on which it is defined. So, x 1 equal to x. and that f x 1 a marginal of x 1, is also given by this function, where x 1 is positive. So, both the variables are supposed to be positive, take positive values, and so again I want you to find out expectation of x 1 given x 2 equal to x 2, and then also find expectation x 1, and correlation x 1, x 2. So, you should be able to do it, because you have all the tools and this thing.

Question five; x 1 comma x 2 to be a 2 dimensional discrete random variable about are 2 discrete random variables with joint probability functions. So, now, you have to compute correlation x 1 x 2, and are x 1 x and 2 independent random variables. So, remember even if your correlation coefficient is 0, it will not necessarily imply that x 1 and x 2 are independent. So, to verify you will have to compute the you know show that for, or find at least 1 pair of values of x 1 and x 2 for which the probability x 1 equal to that number and x 2, equal to a particular number, is not equal to the product of individual probabilities. If you can show that, then you can conclude that x 1 and question 4 not independent, but otherwise you will have to go on verifying for all possible pairs, which means you have eight pairs. So, for eight pairs if you can show that the probability of the product, is equal to the product of the individual probabilities, then you can conclude that they are independent. For the discrete case, this is the only way you can do it.

(Refer Slide Time: 46:19)



Exercise: 6

6. Using the result $E(X, Y) = E(XE(Y/X))$, show that $Cov(X,E(Y/X)) = Cov(X,Y)$.

7. Consider the joint probability density function of (X, Y) given by

$$f_{XY}(x,y) = 2-x-y \qquad 0<x<1,\ 0<y<1$$
$$= 0, \qquad \text{Otherwise.}$$

(i) Find the conditional probability density function of Y given X=x.
(ii) Compute $E(Y/X= x)$, $E(Y)$
(iii) Verify that $E(Y) = E(E(Y/X))$.

8. Let X, Y, Z be random variables and a, b be constants. Prove that
(i) $Cov(X, aY + b) = a\, Cov(X, Y)$
(ii) $\rho(X, aY + b) = \rho(X,Y)$ for $a > 0$

9. Let X_1, X_2, X_3 be three independent random variables each with variance σ^2. If we define new random variables

$$W_1=X_1,\ W_2 = \frac{\sqrt{3}-1}{2}X_1 + \frac{3-\sqrt{3}}{2}X_2,\quad W_3 = (\sqrt{2}-1)X_2 + (2-\sqrt{2})X_3,$$

show that $\rho(W_1,W_2) = \rho(W_2,W_3) = 1/2$ while W_1 and W_3 are uncorrelated.

10. A Fair die is successively rolled. Let X and Y denote respectively, the numbers of rolls necessary to obtain a 6 and a 5.
Find: (i) E(X) ; (ii) E[X/Y=1] ; (iii) E(X/Y =5).

Question 6; using the result that we just obtained this result for I just obtained it for you, that expectation or I am sorry here there should be no comma. So, expectation of x y. So, p is removed the comma expectation x y is equal to expectation x into expectation y given x. So, I just proved this result for you. Now using this result, show that covariance, again here this is x, expectation y given x is covariance x, y. So, the comma in the last 2 terms is, but here when you saying the result is expectation x y. So, comma is to be removed, therefore, you can use this result, or show this result, using what we have proved just now.

Question seven consider the joint probability density function of x, y given by f x y of x comma y. So, here it should be x and y as a suffixes are the capital once and small x y, is equal to 2 minus x minus y x between 0 and 1 and y between 0 and 1 0 otherwise. So, part 1 says find the conditional probability density function y given X equal to x, and then in part 2 I want you to find the expected value of y given X equal to x and E y. Then three I want you to find E y, though I want you to show, that E y is actually equal to E a conditional expectation y given x. So, what we have been doing in the lecture also. We have been verifying, you know we have been computing E y independently by first computing the marginal of y, and then its expectation, and secondly, by breaking it up into first the conditional probability a conditional expectation y given x and then we take the expectation again. So, you have to say that the 2 processes redo the same answer.

Question eight; is x y z are 3 random variables and a and b are 2 constants, proved that covariance of x comma a y plus b b is this. So, I had done for you when the constant a was with x, now you please do this, you should not be, because remember. In fact, you can immediately do it, because covariance x comma a y plus b is covariance a y plus b comma x and therefore, from that result, but then I have added up plus b here. So, please work it out, and show that this result is true. Then the correlation coefficient of x comma a y plus b, there will be no, because you see a numerator there will be a from here, and then in the denominator also when you take the variance of a y plus b, a will come out and. So, the a a will cancel and of course, a has to be positive here, because in the variance you will take out a, only if a is positive otherwise you have to take out absolute of a.

So, let x 1 x 2 x 3 be 3 independent random variables, each with variance sigma square. So, there are three independent random variables with the same variance, if we define new random variables. So, here w 1 is x 1 w 2 is root 3 minus 1 upon 2 of x 1 plus 3 minus root 3 upon 2 of x 2 and w 3 is a linear combination of x 2 and x 3 show that a correlation coefficient of between w 1 and w 2, is equal to the correlation coefficient between w 2 and w 3, which is equal to half. While w 1 and w 3 are uncorrelated. So, I just try to I mean the purpose of giving the exercise was that you see that, taking this linear combination some turn out to be correlated, and some pairs turn out to be uncorrelated. So, this is what you have to show.

Now, tenth question is a fair die is successively rolled, and let x and y denote respectively the numbers of rolls necessary to obtain a 6 and a 5. So in fact, the 6 part we discussed took at length and now. So, you are asking for. So, is the number of rolls required, till a 6 shows up and y is the number of rolls required until a five shows up. So, now, find expectation x so; that means, you will write down the probability match functions for x and y, and then compute E x, compute conditional expectation of x given y is equal to 1, and conditional expectation of x given y is equal to 5. So, it should be easy computations; once we have already handled the case, when you had to roll the die till a 6 showed up.