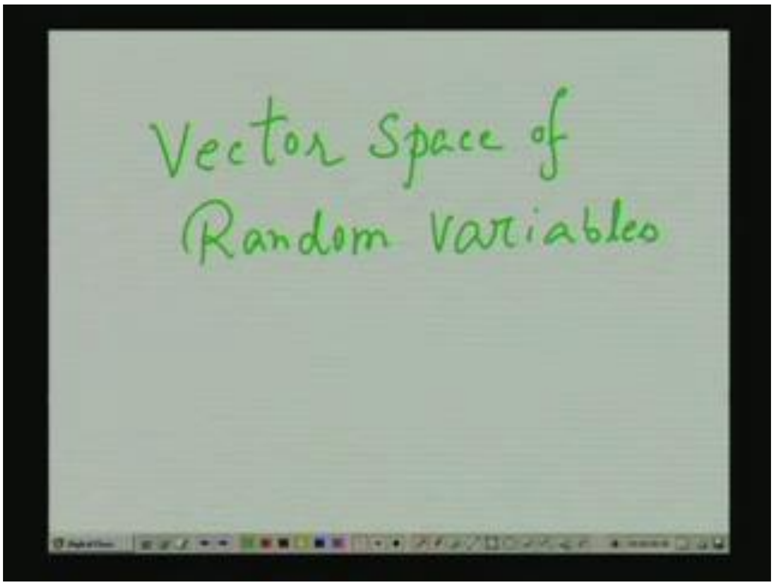
Probability and Random Variables
Prof. Dr. M. Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 16
Vector Space of Random Variables

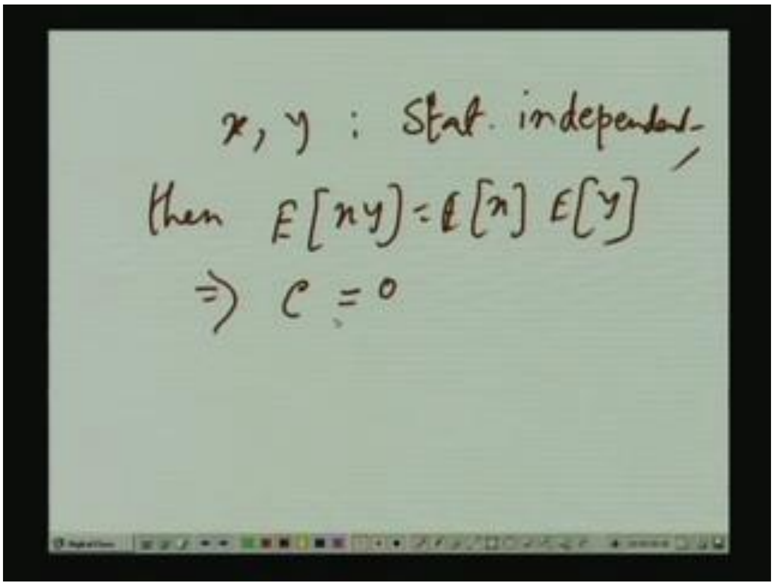
So, today's class, we will be concentrating on a new topic and a very interesting topic, it is called vector space of random variables.

(Refer Slide Time: 00:50)



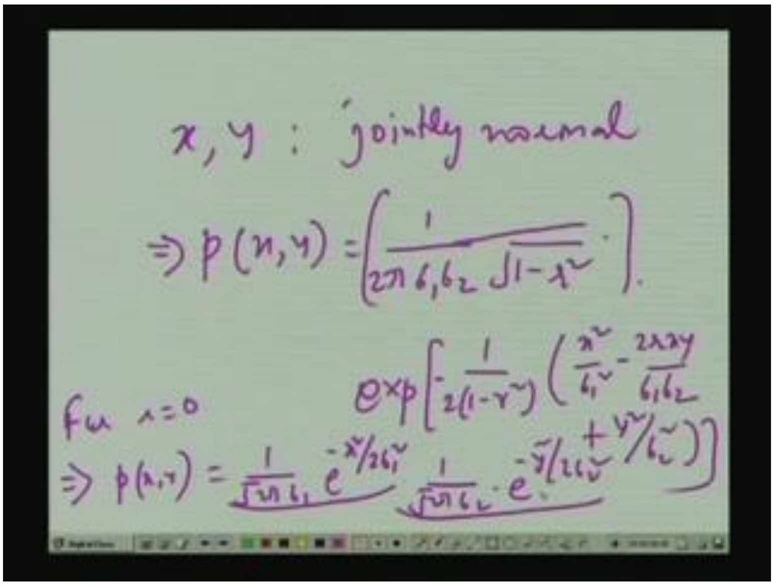
In fact this is a huge topic that cannot be covered just in one lecture. So, we will not make any attempt for that. But before I am going to that, let us just go back to what we were discussing last time, because little bit of that was left.

(Refer Slide Time: 01:40)



Last time, I had said that, if two: x and y are statistically independently – statistically independent; then E xy is nothing but E x E y. This leads to the fact that, covariant C or correlation coefficient, which is C by sigma x sigma y; that is equal to 0. So, if they are statistically independent, it is always true; that is trivially seen. But, if this is given, that is, x and y are two random variables, which are not… which are uncorrelated that they does not necessarily mean that, x and y are also statistically independent; that is, joint density of x and y – p xy is not in general a product of p of x and p of y, but for a particular class of random variables, actually the Gaussian random variables… That is why x and y are mutually jointly Gaussian; then statistical independence means uncorrelatedness as before. But, also uncorrelatedness means statistical independence; that we can see. So, that works for Gaussian random variables.

(Refer Slide Time: 03:08)



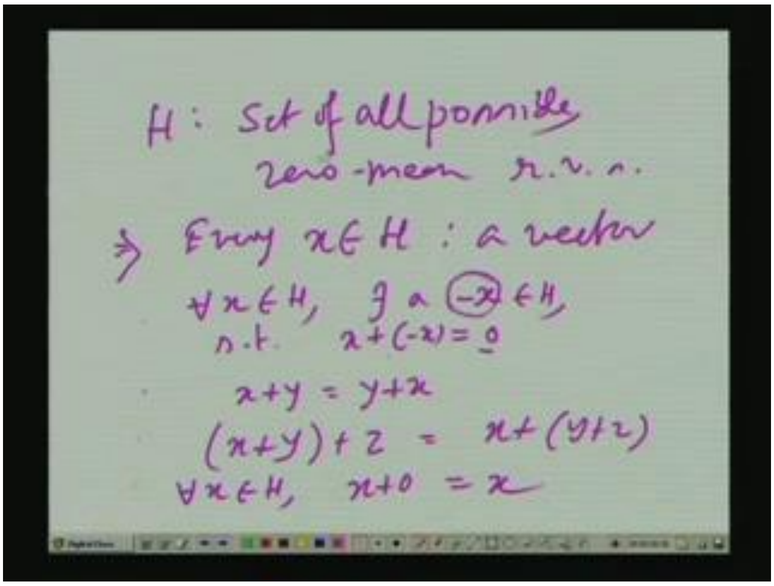
That means suppose jointly Gaussian or sometimes we call jointly normal; then we know what is the probability density. What is the probability density? Just a minute; this thing times exponential minus 1 by 2 into 1 minus r square into pi square by sigma 1 square minus 2 r x y by sigma 1 sigma 2 plus v square sigma 2 square. Now, we have seen earlier… In fact, we have proved that, this r is nothing but the correlation coefficient. So, suppose it is given that x and y are uncorrelated; in that case, r is 0. If r is 0, this term is 0; 1 minus r square plus 1; which is nothing but… So, this becomes nothing but minus 1 by 2 and then x square by sigma 1 square plus y square by sigma 2 square. And this simply becomes 1 by 2 pi sigma 1 sigma 2.

So, for r equal to 0, this leads to a very simple thing; we can write this as a product. One of them is root 2 pi sigma 1 e to the power minus x square by twice sigma 1 square; and again 1 by root 2 pi sigma 2 e to the power minus y square by twice sigma 2 square. So, this is p x; this is p y; which means the joint density is nothing but the probability density of x multiplied by the probability density of y; that is, x and y are statistically independent. So, as I said that, if x and y are statistically independent, they are always uncorrelated; that is true in all cases. But, if x and y are given to be jointly Gaussian or jointly normal, that is, equivalent way of saying; then if they are uncorrelated, then again the reverse is true; that is, they are statistically independent too. But, that is not true for other cases. So, it is not true in general; but in the case of Gaussian random variables, statistically independent leads to uncorrelatedness and uncorrelatedness leads to statistical independence.

Now, we come to this important topic called vector space, rather Hilbert space of random variables. And this cannot be covered just in one lecture, because this topic of vector space as such is a semester long topic. So, I will not try; but I will just tell you the motivation. Actually the random variables can be treated equivalently as though they are some vectors if some infinite dimensional world. See we are all familiar with three-dimensional world, which is spanned by three axes: x, y, z. And we know that, they are orthogonal, but they need not be orthogonal; the angle between them could be anything. But, three mutually independent axes are required and they span the entire three-dimensional world. Any vector can be written as a linear combination of those three basis vectors or elementary vectors: one in x direction, another in y direction, another in z direction. So, we are familiar with these.

We are familiar with other notions, which are valid. They are also like the dot product involving two vectors and all that. But, you know this whole idea of vector space; this three-dimensional vector world can be generalized to an arbitrary n dimensional and later infinite dimensional vector space; where, you do not have any notion of this physical vectors – the position vectors, which are we familiar with in the three-dimensional world. But, there are all the advantages that we are… or all the results that we are familiar with in the conventional three-dimensional vector space. They are all present there. So, in a vector space, actually, in it… And we are defining vector space; we are using vector space over the field of… We say what the field of complex or real numbers; maybe in our case, to make life simple, we are taking vector space over real numbers – means whenever we come cross any scalars, any number; we will assume that, number is just nothing but a real number and you all know all the operations of real numbers; how to add two real numbers, how to subtract and all that. So, I will simply consider the vector space of random variables over the field of real numbers.

(Refer Slide Time: 08:27)



That is, let H be set of all possible maybe zero-mean to make again life simple random variables. Now, this is just a set; it does not become a space. Here every random variable will be called a vector. Every x element of H – a vector. Please do not confuse this vector with the traditional, conventional notion of vector. There is a column vector, row vector or position vector; not at all. This term vector is just an abstract term; it is an element; you can call it an element. But, we call it vector, because there is an analogy between these treatments in the three-dimensional vector world. That is why we say it is a vector. But, actually it is a random variable; there is nothing vector-like about it. So, x is just a random variable; it is not a column vector or row vector or a position vector; it is just a random variable. But, we just call it the ((Refer Slide Time: 09:36)) vector. So, that is in an abstract sense. It becomes a vector space H under certain conditions. What are the conditions? That if you take any vector y, x and any other vector y; that is, two random variables x and y out of H; there must be a rule of addition of these two vectors, so that you get another random variable. So, there must be some rule of addition, so that x and y can be added and you get another element, which must also belong to H.

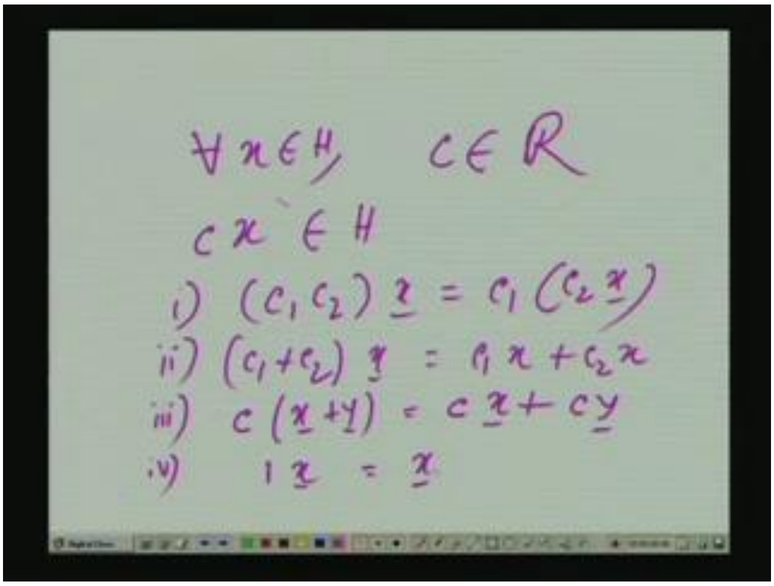
First let us see what can means as rule of addition; and under the rule of addition, if you really add x and y, you get something, which should belong to H; let us first verify that. When I say x plus y; it means every time I conduct this experiment on x and y jointly; I measure the value of x and simultaneously measure the value of y whatever be the values comes up; in those cases, I just add them up. So, I get a real thing value. Obviously, this real thing value changes from experiment to experiment. So, real thing value corresponds to a random variable and that random variable is z; that means every time the value of z is obtained by summing the particular value of x and particular value of y that have come up in that experiment. That is the meaning of this random variable… – the addition of two vectors. In fact, that is how we add two random variables also; that is the usual meaning. But, since the real thing variable, real thing quantity is a random variable and H consists of all possible random variables in the world, z also belongs to H; z does not go somewhere else. So, we say that, it is closed under this vector addition.

This addition should satisfy some basic properties in order for this H to become a vector space. One of them is that, I mean this should be commutative all; which is obvious, because you will be adding the numerical values for x and y to get the real thing value. So, it does not matter whether you add the value of x with y or value of y with x. Then, now, we take z to be some other random variable. So, x plus y plus z – it should be same as – this is called associativity. Then, in this set, there must be a unique zero vector or zero element or zero vector. What is the zero vector? That is, it must be a random variable, so that for all x element of H, x plus 0 – that is, if you add the random variable x with this particular random variable called zero vector, you should get back x. This zero vector is what? It is also a random variable, but it is a random variable, which in all experiments, takes only zero value. We say using probabilistic language, it is a random variable, which takes the value 0 with probability 1.

Obviously, whenever you conduct experiment on x and simultaneously on this random variable, you get some value of x; but zero value always from this, so that value of x plus 0, which will give you the ((Refer Time: 13:01)) value of x only. So, that is why this equality is satisfied. So, this H consists of one such random variable called zero random variable. So, this is all satisfied. And finally, for all x element of H, for each x, there exists a negative of that element of H. What is meant by that? So that when you add x to this particular random variable – this particular vector, this is a notation – minus x; together I say that is a vector; and where x and minus x – these two are added; this should be this zero vector. So, for each x, I should have one such vector, so that when these two are added, you get zero vector.

Now, does it exist? Answer is yes. If there is a random variable x, I can always define a so-called negative version of it; maybe denoted by this. How? Whatever value of x I obtain by an experiment, I will take the negative of that; assign that value to this random variable – minus x. So, whatever I add – x and minus x, the two values cancel each other. So, you will get a zero value. So, resulting random variable always takes zero value or takes zero value with probability 1. So, it must be this zero vector. So, these are the basic things that should be satisfied and they are indeed satisfied. That comes… That is relevant to this addition – addition of two components – any two elements of H; or these elements are also called vectors; where, again I repeat there is nothing is vector like, nothing column vector or row vector like or position vector like; the vector is a just term here. So, here these rules are in the context of the vector addition. That is one side of the story. There is another operation. One operation is that, H has a rule of vector addition involving the vectors or random variables; or whenever you add two random variables, resulting vector or resulting random variable also be should belong to H and that is indeed the case.

(Refer Slide Time: 15:21)

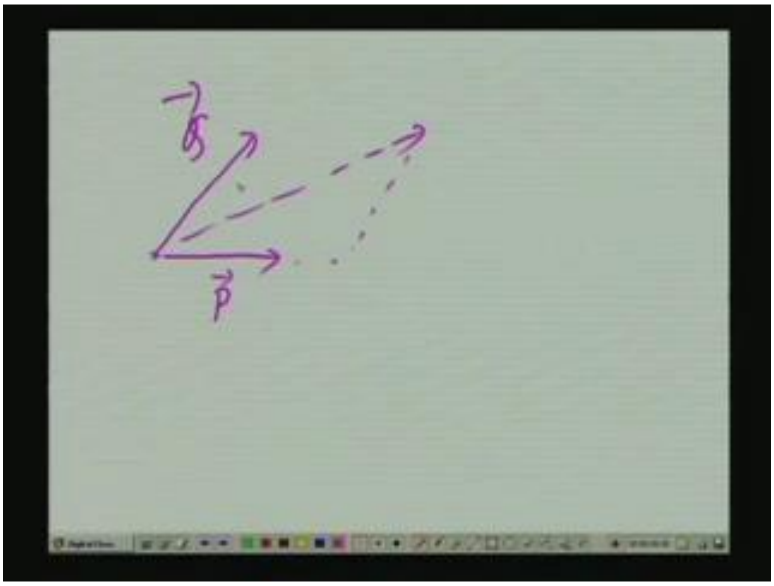


The other operation is scalar multiplication; that is, for any x element of H and take a scalar c element of real number. So, c is just a real number; R is the set of all real numbers; c is… First, cx; whether that cx – I mean I denote by this abstract notation cx; another random variable, which must also belong to H; it should satisfy some basic property.

Now, how we get this? How we found this cx? In practice, I observe x in an experiment; whatever value I get, multiply that by c. So, I get another value. But, that value also is random because every time I measure x, I get new and new values of x. So, cx also takes new and new values. So, this resulting thing also a random variable. So, cx then belongs to H. But, this cx I mean… In the general case, not necessarily in the case of this particular vector space of random variables; in the general case, it means that for any vector element, which is belonging to this vector space H and for giving any scalar c, cx would denote a new vector; but it must also belong to H. We call this new vector as the scalar multiple of x; but the scalar multiple of x must satisfy some basic properties. One – if you have things like this – c 1 c 2; if you break c as c 1 c 2 x, you can as well have c 1 followed by c 2 x; that is, take x multiplied by a scalar c 2; you will get a new vector. On that you can multiply… That again you multiply it by c 1. So, you get another vector. That should be same as taking x multiplying by this c. This should be satisfied. Then, other thing is also very simple. This should be this.

On the other hand, for any x, y, this is trivially satisfied, because in our all experiments, you observe a value of x multiplied by this summation c 1 plus c 2. Whatever you get, you will get the same thing. If you multiply that value of x by c 1 and by c 2 and add those values. Similarly, here if you conduct a joint experiment on x and y, add their values in any experiment multiplied by the c, you will get the same one. If you take the individual value of x – individual value for x multiplied by c; individual value for y multiplied by c and add them. This should be satisfied; and obviously, satisfied. These basic conditions should be satisfied for any space to be vector space. And we are seeing that, these are satisfied for H; where, H is a set of all possible random variables. And 1 x; if you take any vector x and take a scalar one and consider the scalar multiple of x by 1, this is again another vector; this should give you back x. And; obviously, this is satisfied again for H, because when you have an experiment on x, whatever value you observe multiplied by 1, you get the same value, which you had for x. So, these are satisfied. So, these basic conditions are satisfied means H is a vector space. In this case, like you know we had…

(Refer Slide Time: 19:09)

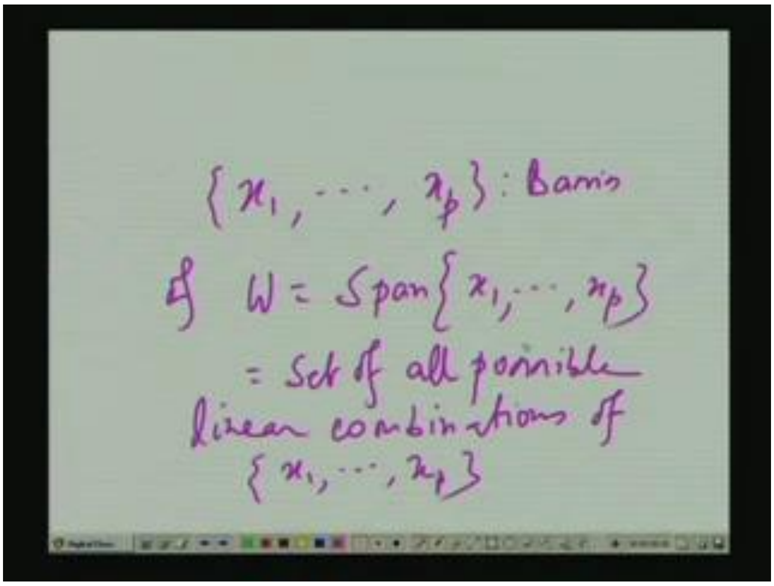


Suppose there is one vector p, another q; and I say this whole two dimensional world is spanned by pq means you consider any point x – any point. This will be some linear combination of p and some linear combination of q. I mean some linear combination of p and q. So, every point of the space is a linear combination of p and q. p and q – they are linearly independent because q cannot be expressed in terms of p; this is pointing in this direction; this is pointing in this direction. Similarly, p cannot be described in terms of q; we say they are linearly independent. In general, if you have more than two dimensions; say another dimension p, q and may be r pointing upward. r, p and q – again they forward linearly independent set, because r cannot be expressed as a linear combination of p and q because any linear combination of p and q will be in this plane, but r is pointing upward.

Similarly, q cannot be expressed as a linear combination of p and r; and p cannot be expressed as a linear combination of q and r. So, they are linearly independent at any point. And what is meant by span of them? Span of them means set of all possible linear combinations of these. For instance, span of p and q means you take all possible linear combinations of p and q. Linear combination is typically of the form some c 1 times p plus c 2 times q; where c 1 and c 2 are scalars. Applying any value for c 1, any value of c 2, you get one particular point and go on, go on. So, these 2-D world is nothing but a set of all possible linear combination of p and q. That we will be calling… We will be saying it as a span of p and q. And here p and q are linearly independent. So, then p and q will be called a basis of this entire two-dimensional world.

Similarly, if it is three-dimensional with the vector r pointing out, p q r form a linearly independent set. What is the span of p q r? It is a set of all possible linear combinations of p q r. Typical linear combinations of this form c 1 p plus c 2 q plus c 3 r; which is a vector pointing the space. And you apply any value of c 1, any value of c 2, any value of c 3, you get some particular point in the 3D space. So, basically, span of p q r means… I mean collection of all points in the space means the whole three-dimensional piece. And p q r becomes a basis, because they are linearly independent. There is no redundancy there. Every vector is needed; p is needed, q is needed, r is needed. None of them can be expressed as a linear combination of the rest; everybody is required.

(Refer Slide Time: 21:53)

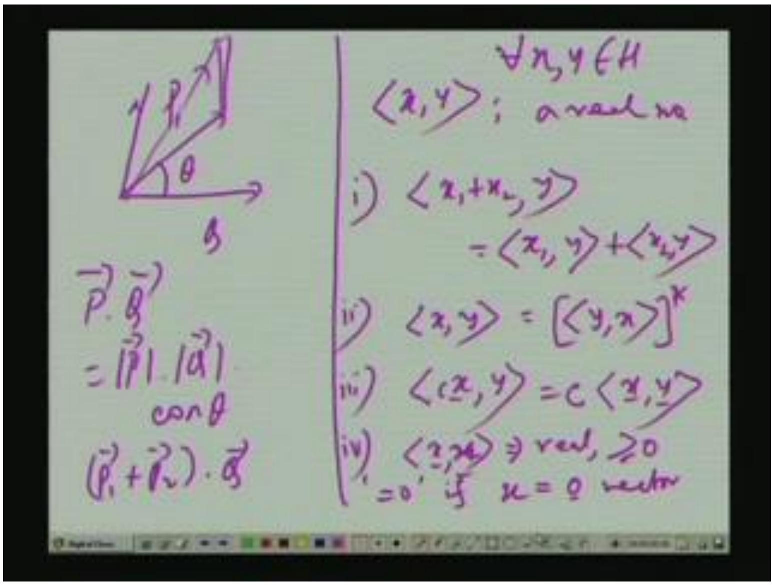


Similarly, here also; here also if you have some vectors x 1 to x p; they are linearly independent if none of them is expressible as a linear combination of the rest. So, x 1 should not be expressible as a linear combination of x 2, x 3 up to x p. Same for x 2; x 2 should not be expressible as a linear combination of x 1, x 2, x 3, x 4 up to x p and so on and so forth. In that case, this becomes a basis of W, which is a span of x 1 up to x p means set of all possible… You can see one thing. If you just take p elements – x 1 to x p. Consider all possible linear combinations of them and form a set. My claim is that set itself becomes a vector space. We will call it a subspace because that is a subset of the mother H. How would you say that this becomes a vector space? You take one combination. That is one element of the span. Take another linear combination of these two. That is secondary element of the span; you add them. So, earlier may be you had c 1 x 1 plus c 2 x 2 plus dot dot plus c p x p. There is one element of W.

Now, we have got d 1 x 1 plus d 2 x 2 plus dot dot dot d p x p. That is another element of W. If you add them, you have within brackets c 1 plus d 1 times x 1 plus c 2 plus d 2 times x 2 plus dot dot dot dot; which is again another linear combination of these elements; so, which must also belong to W, because I am considering set of all possible linear combinations. And you can see that, if you take any linear combination; if you take all the coefficients to be 0; then obviously, you will get zero vector. So, zero vector is present. For any vector, any linear combination if you reverse – I mean reverse the sign of the coefficients; if it is c 1, make it minus c 1; if it is c 2, make it minus c 2; obviously, you get negative of the previous combination. So, negative of a vector also is present. And likewise you can verify that, all those basic ((Refer Slide Time: 24:19)) of a vector space have satisfied by this set of all possible linear combinations of these. And if they are linearly independent, then I say that this becomes a basis. Linearly independent means none of them is expressible as a linear combination of the rest.

Now, I am trying to do things fast and I am skipping many mathematical details. Now, you know that, in a three-dimensional world, there is something more that you are familiar with; that is, a dot product involving two vectors. What does the dot product give you? It takes two vectors and gives you a real number. How? If the one vector is p; another vector is q; then p dot q is nothing but length of p times length of q times cosine theta; where, theta is the angle between p and q.

(Refer Slide Time: 25:09)



That is, if it is theta; then P dot Q is nothing but length of P, length of Q times cosine theta. Now, when you delete a general vector space or the vector space of all possible random variables; obviously, we do not have any angle there, because the angle is a physical; it is a parameter, which occurs in the positional world – in the three dimensional positional world. But, we are not dealing with the vector space of random variables; there is no such angle occurs. But, this dot product actually in our language is called inner product. And even this inner product just should satisfy some basic properties – a definition of inner product. So, depending on the context, you can throw your own definition of inner product. It should satisfy some basic properties. Those are independent of these physical parameters like an angle and all that.

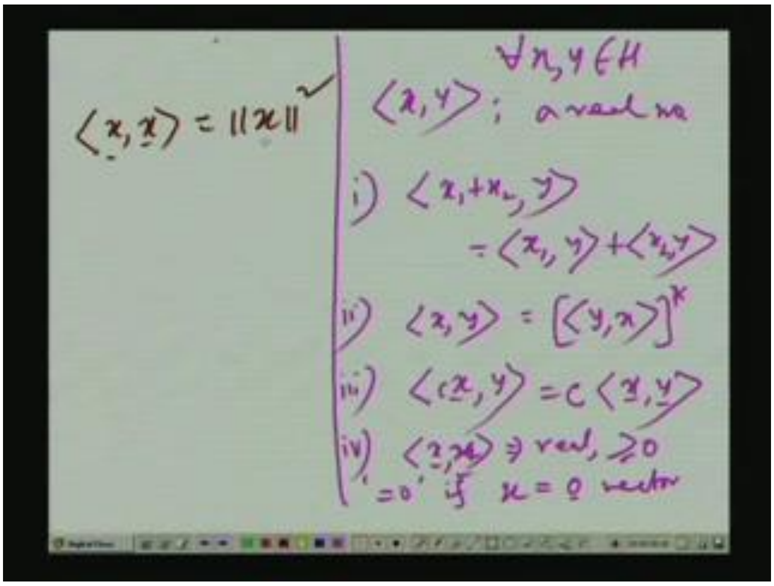
If those basic properties are satisfied, then we can live happily with these. That is what mathematicians found out. In the case of three-dimensional world, this was defined to be the inner product and this was satisfying also those basic ((Refer Slide Time: 26:18)) and probably could live happily with these. Those ((Refer Slide Time: 26:21)) are like these. Inner product – instead of calling it x dot y like P dot Q, we say x comma y bracket. It takes x and y; where, x and y are two random variables; that is, for all x, y belonging to H, what does it do? It gives you a real number; real number because we are considering with real number case. When it satisfies some basic properties… It should satisfy some basic property. If the definition of the inner product satisfies those, this will be acceptable like this definition was acceptable in the three-dimensional vector space – positional vector space.

Those properties are simply… That instead of the vector x, if you have the combination, where x 1 plus x 2 comma y; you should be able to… So, that is, instead of having just one vector, you have x 1 plus x 2. So, first you take x 1, x 2 forming addition. So, this is another vector of H. That vector comma y; its inner product is a real number. This should be same as… If you take the inner product of x 1 with y and x 2 with y; you see this is satisfied here. If you have P 1 plus P 2 dot Q; you can see geometrically P 1 and P 2; this could be P 1; this could be P 2. So, P 1 plus P 2 will be this vector. You can see that, dot Q is nothing but P 1 dot Q plus P 2 dot Q; it can be proved easily using some geometry. So, this should be satisfied.

Secondly, inner product between x and y and y and x – this should be conjugate of each other; y – this is a complex number. But, we are dealing with real case. So, for real number, conjugate has no meaning. So, for our case, x comma y and y comma x – they are same. y star p dot q and q dot p, they are same, because they are in this case also, we are getting real numbers. But, when we are dealing with complex-valued case, x comma y inner product may not be a real number, may be a complex number. Then, y comma x – its conjugate should be equal to x comma y. Number 3 – if instead of x, you have a scalar multiple of x; like instead of P dot Q, I have some c times P dot Q. That should be same as taking c out; here likely you can take c out; within bracket put P dot Q carry out this; then multiply this real number by c; same thing should be valid here also.

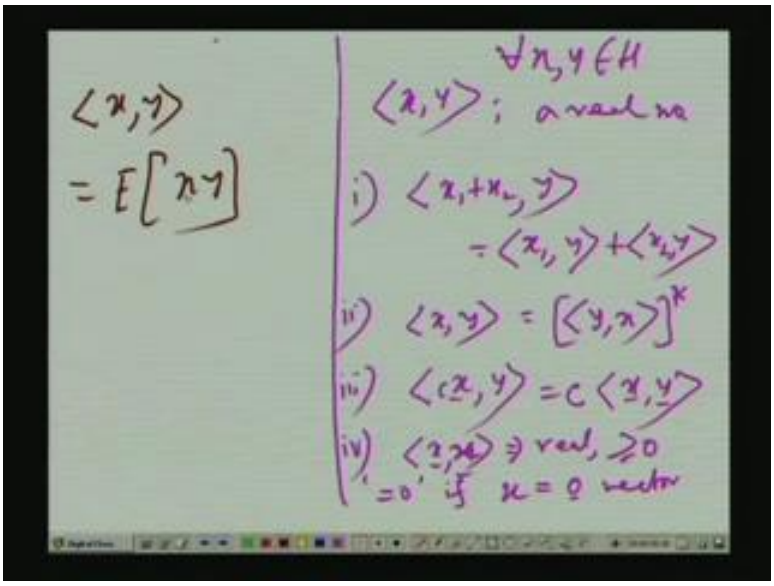
And, last one is very important. If you take the inner product with itself like P dot P, it is nothing but length P square; that is all because theta is 0, cos 0 is 1, length – P square, which is a real number always; and always greater than equal to 0. It is 0 only if p is a zero vector. Similarly, here also, inner product with itself; this is real greater than equal to 0, equal to 0 if x is the zero vector; zero vector because H – the vector space has a zero vector. So, if x is a zero vector, then only the inner product of this with itself will become 0; otherwise, is a real positive quantity. This should be satisfied.

(Refer Slide Time: 30:37)



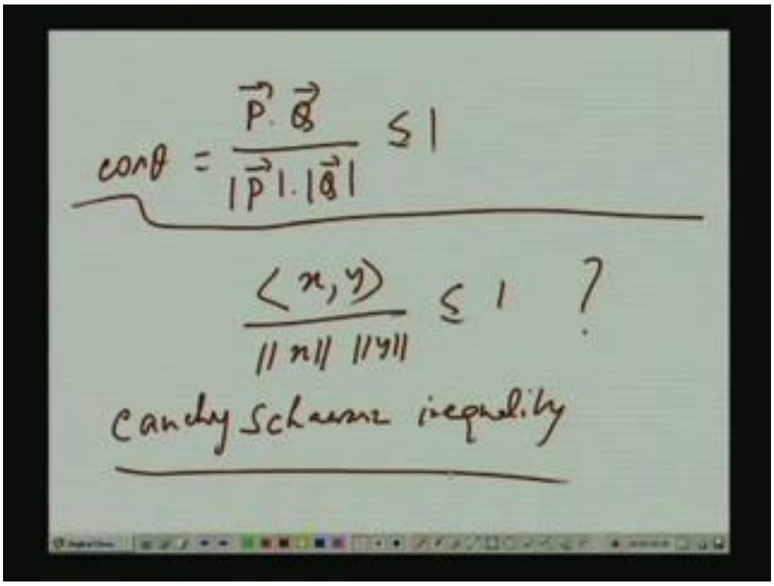
In fact, inner product with itself, which is nothing but in the three-dimensional case square of the length; the length is also called norm – norm of a vector; norm is something similar to length. That is a real number; it is denoted by this – double line, double line. So, this should be square of this – length square. It is norm of x is equivalent to the length because there is no apparent length here for random variable; but is something ((Refer Slide Time: 31:07)) and square of that. This is a real… So, this should be 0 only if x is a zero vector belonging to H. Now, for H, this is a very general definition. These are the conditions that, an inner product definition must satisfy it. Now, in the case of H, I say that, we can define an inner product like this. We can define an inner product like this.

(Refer Slide Time: 31:55)



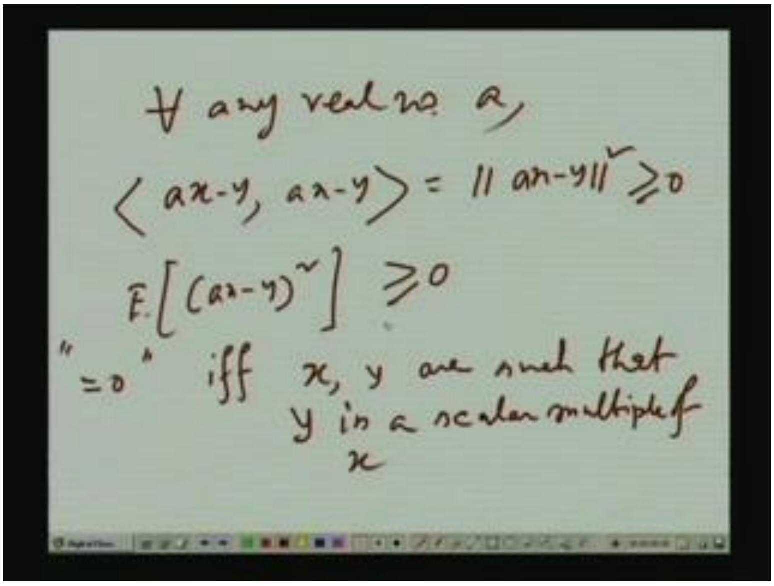
That is the correlation. Take x, y; multiply; take expected value. Now, does it satisfy all the properties? Answer is yes. First, instead of x, if you have x 1 plus x 2 times y, this is E is a linear operator we have seen; it becomes E of x 1 y plus E of x 2 y. So, this is satisfied. Similarly, E of xy or E of yx – they are one of the same; so, satisfied. Similarly, E of cx times y; you can take c out, because c is a scalar; and therefore, not random. It can go out of the expectation and you have c times E xy; this is satisfied. And lastly, x comma x – inner product with itself is nothing but E of x square; which is nothing but the expected power of this random variable or even variance you can say, because the mean is 0. And this is always positive, because x square and then expected value is always real and positive or non negative. And it can be 0 only if x is such; then it always takes zero value. So, its square also is zero value. So, if you take the expected value, that also is 0; that is, x is a zero random variable; which takes zero value with probability 1. This is the correlation.

(Refer Slide Time: 33:32)



Now, you have seen one thing previously that, we can write cos theta in the three-dimensional world as… And this is always less than equal to 1. Our question is do we have this thing valid? Is it so? Answer is yes. This is a famous inequality called Cauchy Schwarz inequality. This will be satisfied. How?

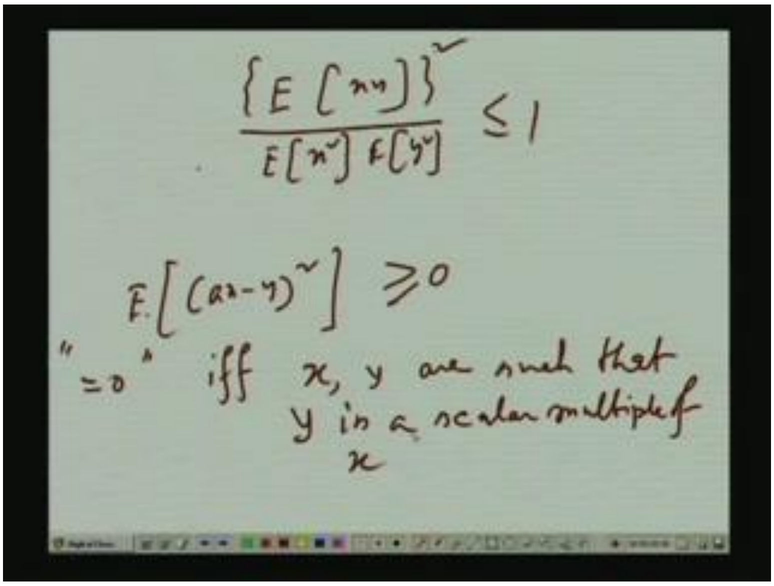
(Refer Slide Time: 35:02)



Let us see how. Take for any real… This is important – any – for any real number a, we can always say… which is nothing but norm square of… This must be greater than equal to 0. In fact, when it is equal to 0; only if x and y are such random variable, so that we can find some scalar number a for which a times x is equal to y. Then only it becomes… It is a 0. And as I told you earlier, norm square or norm can become equal to 0 only if we have a zero vector here. And this can become a zero vector only if y is such that, it is some scalar multiple of x. So, this is equal to 0. But, here we are writing using the general notations of inner product and norm. In our case, actually, it becomes expected value of square. This we always know; it is greater than equal to 0. So, I am rewriting this using our definition of inner product.

Now, equal to 0; here it is greater than equal to 0; equal to 0 means if and only if x, y are such that, y is a scalar multiple of x. Suppose it is not; suppose x, y is not a scalar multiple of x; then let us see whether that is satisfied; whether the triangular infinity is satisfied. You see when it is equal to 0, what happens? Then, after all, we are trying to find out; we are trying to prove that E xy; we are trying to prove that Cauchy Schwarz inequality; we are trying to prove this quantity, this thing.

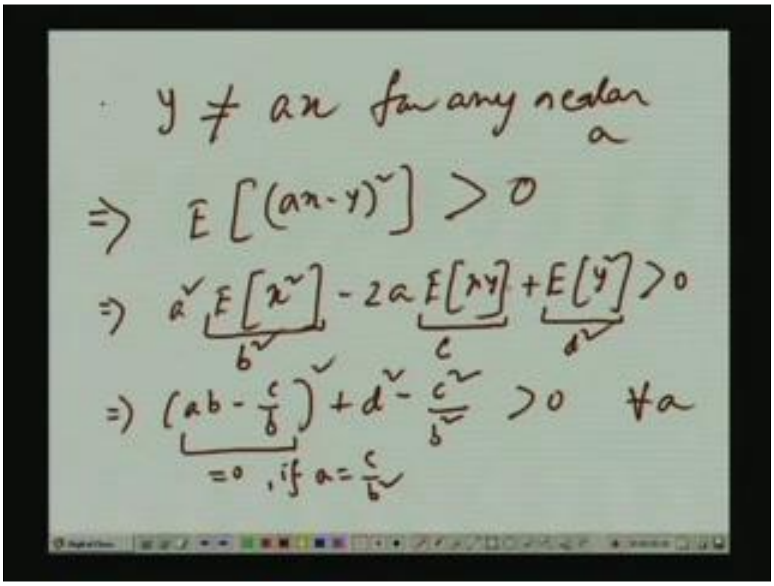
(Refer Slide Time: 38:15)



We are trying to prove that E xy… Maybe you can take a square of it, because if it is less than equal to 1 square; also, is less than equal to 1; this is less than equal to 1. So, I have squared up; otherwise, this square will not be here; it will be a square root. The square root E of x square is the norm of x; square root of E y square is the norm of y. So, norm of x, norm of y; and here is just inner probability between x and y. But, I wanted to get rid of the square root in the denominator. So, I just squared up because if… Before squaring up, it is less than equal to 1; then after squaring up also, it should be less than equal to 1. So, you have to prove this.

Now, in this, in the first case, when x and y are such that, y is a scalar multiple of x; we are delivering this case. This becomes equal to 0. What happens in that case to this ratio? You can see this ratio becomes equal to 1 because y square means a square x square; a square can go out. So, a square E x square E x square. And here a x square; a goes out; E x square; and a whole square. So, a square and a square cancels. And E of x square is doubled because of the square. So, it cancels from numerator and denominator. So, you get equal to 1. So, this quantity when equal to 0; which means when y is a scalar multiple of x, then we have this ratio equal to 1. Now, we consider the case; where, this is not satisfied; that is, y is not a scalar multiple of 1; is a multiple of x. In such case, this is strictly greater than 0.

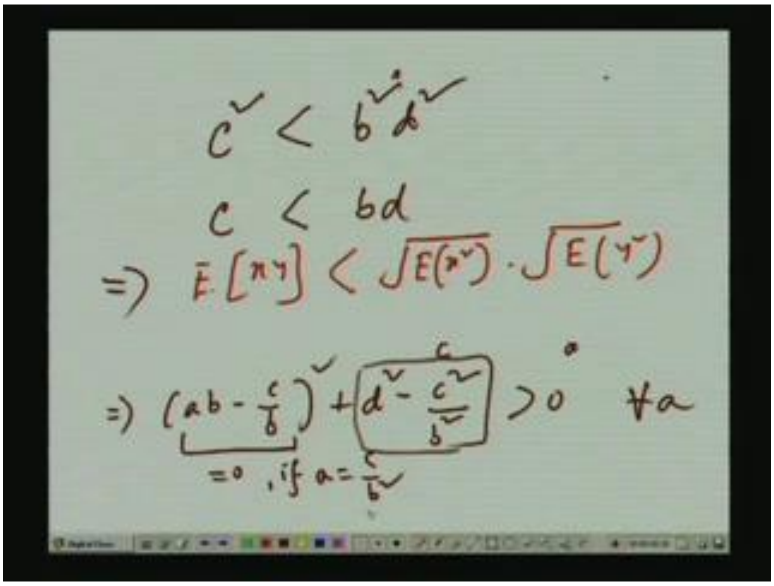
(Refer Slide Time: 40:16)



So, we are considering this case for any scalar a; implies strictly greater than 0. You can write it like a square. Call this quantity b square; this quantity is c; and this quantity is d square. So, a square b square minus 2 ac plus d square greater than 0. You can write like this – ab minus… So, this is ab whole square. So, twice ab into c by b. So, you have to have this c by b whole square; ab minus c by b whole square. So, c square plus b square was introduced. So, it has to be subtracted. So, plus d square minus c square by b square. This should be greater than 0 for any a; that is very important.

Now, when is this is equal to 0? If a is c by b square. Now, c is this inner product or correlation; and b square means just the variance of x; both are real numbers; they are given to us. So, I can always choose a to be a real number like this. For this choice, this quantity is 0; but even then this inequality must be satisfied. So, this must be greater than 0, because this is true for any a. So, this is true for this a also. For this a, this part is 0. So, this must be… This is the quantity; which should always be greater than 0. Okay, what does that mean?

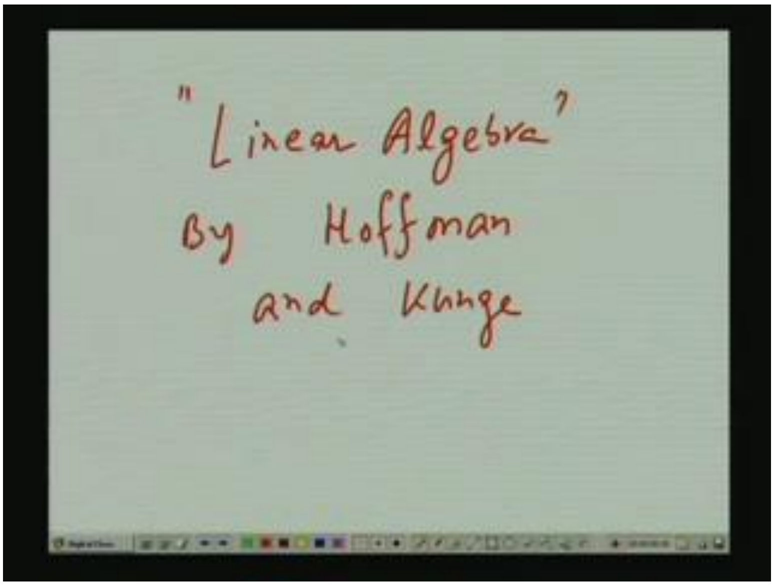
(Refer Slide Time: 43:36)



That c square is less than b square d square. And if you take the positive square root for both sides, c must be less than bd. This implies… What is c? c was… And what is b square root of positive square root… What is d? Again positive square root of this, because square of both sides and you get your result. This is famous inequality called Cauchy Schwarz inequality. This treatment actually is very useful. This universal treatment for carrying out estimation in using random variables; that is, some other some unknown random variable is to be estimated as a linear combination of some observables – some random variables which can be observed as an optimum linear combination. So, what could be that optimum linear combination?

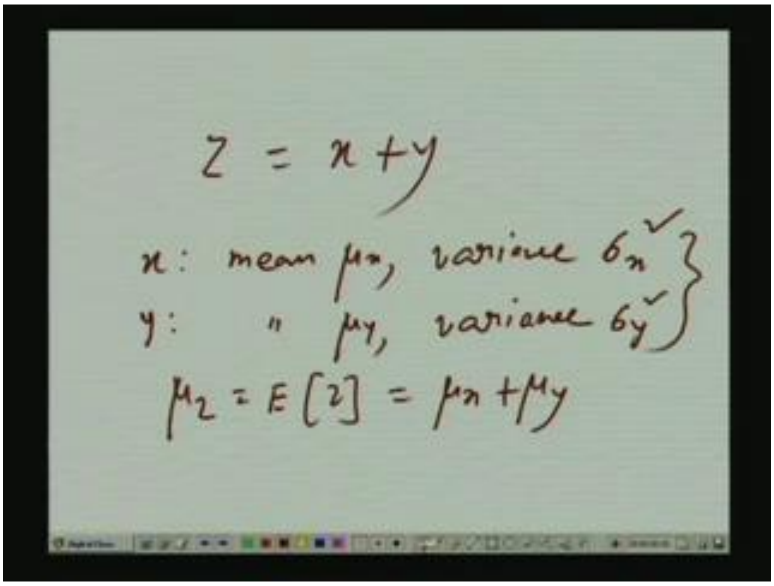
After few lectures, we will consider mean square estimation; there we will see that, such an optimal estimate as a linear combination of a few given random variables is nothing but something called orthogonal projection of that unknown vector into the space spanned by the given set of random variables. So, for that, this is very important. Now, I will tell you again that, this whole vector space theory cannot be covered in one lecture for it to be mathematically accurate and ((Refer Slide Time: 45:24)) lectures. So, if those… If you are interested to know more about it, more about vector space theory, you can consult any book, because there are plenty of books available on this; or maybe you can just consider this book.

(Refer Slide Time: 45:39)



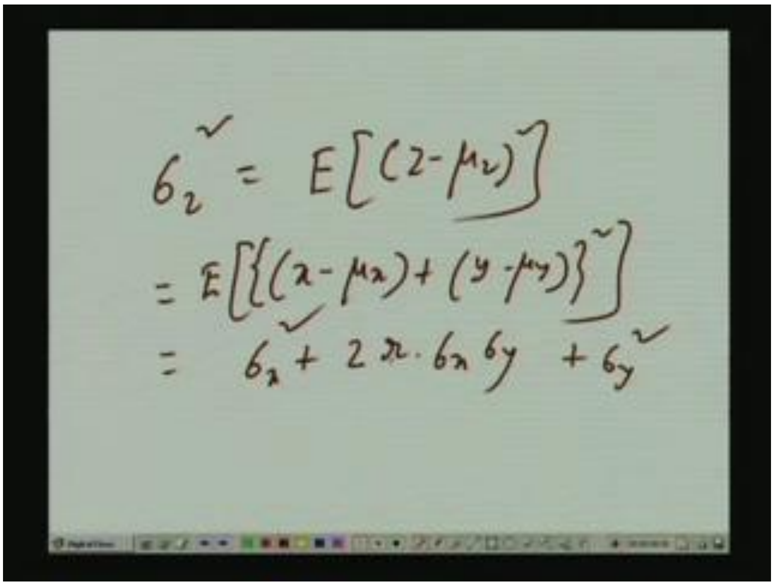
This book has several chapters. Just consider the chapter 1 – vector space by… Before we conclude, I will cover another one small topic.

(Refer Slide Time: 46:15)



That is, suppose z is x plus y; x and y are two random variables; and z is another random variable obtained by summing x and y. x has mean mu x; y has mean… So, what is the mean of z? Obviously, expected value of z is nothing but expected value of x plus y; and you can use the linearity of expectation operator. So, that is nothing but expected value of x plus expected value of y. So, obviously, mu z; which is nothing but E of z is same as mu x plus mu y; that comes trivially… And what happens to the variance?

(Refer Slide Time: 47:24)



Sigma z square is nothing but E of z minus mu z whole square. Now, this you can write as… You replace z by x plus y; mu z by mu x plus mu y. So, this becomes nothing but whole square. So, square of this expected value of that; that will give you the variance of x. Square of this; expected value of that; that will give you the variance of y. And then we have twice E of x minus mu x y minus mu y; which is nothing but the covariance. Covariance – either you can write by c or you can write it as a product of correlation coefficient r times sigma x sigma y. So, if r is given, sigma x is given, sigma y is given; we can find out sigma z square. So, I will stop here today.

But, let me tell you if – I mean in the nutshell, what we are going to do soon. From here we will move to moments; like you remember in the case of single random variable, we defined moments. And moments were used for what? I mean we derived some properties of the moments. And from moments, we marched to characteristic functions. Here also, we marched to what is called joint characteristic function. But, before going from moment to the joint characteristic function, I will again come back to one topic, which I had left out thinking that, I may not need it; but that topic I find is also important; that is, the two functions of two random variables. So far, I considered only one function of two random variables; that is, g of x comma y. But, I will be now considering two such functions; maybe one is f of x comma y, you call it z; another is g of x comma y, you can call it maybe v. So, there are two functions. And obviously, z and v are in general jointly related. So, I have to find out the joint distribution function and joint density of z and y. So, that we will be doing in the next class.

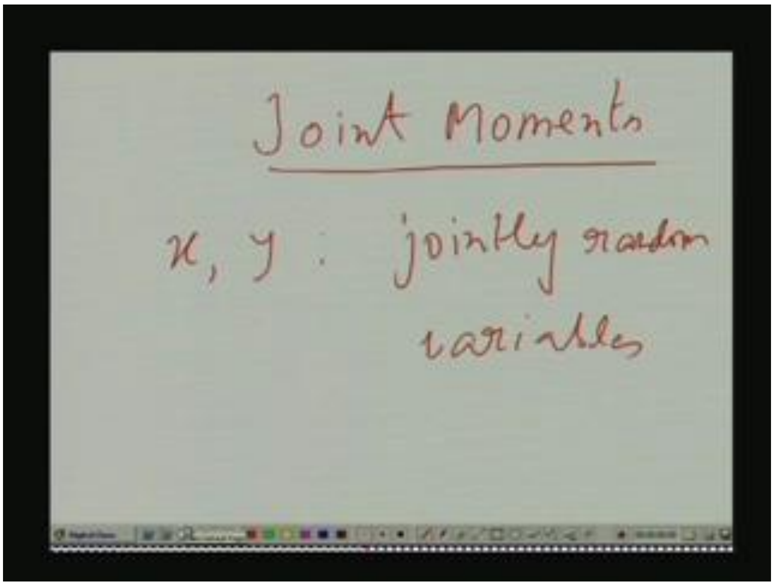
Thank you very much.

Preview of next lecture

Lecture – 17

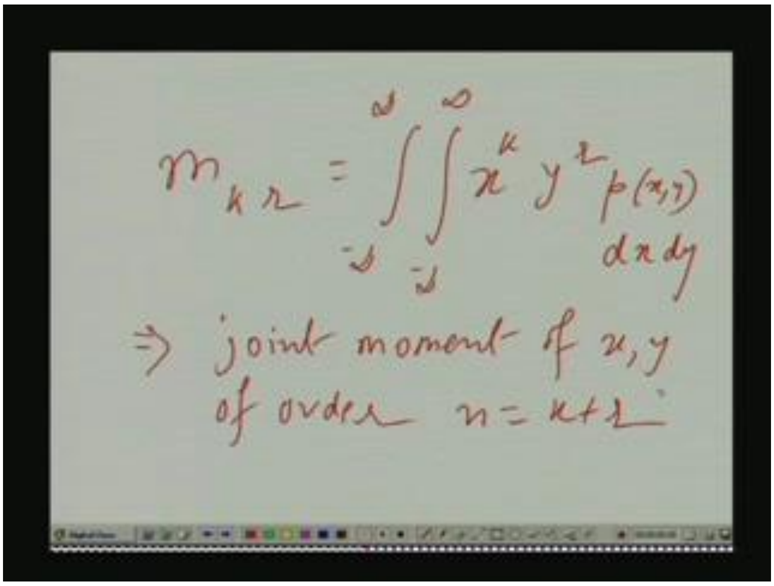
Joint Moments

(Refer Slide Time: 50:28)



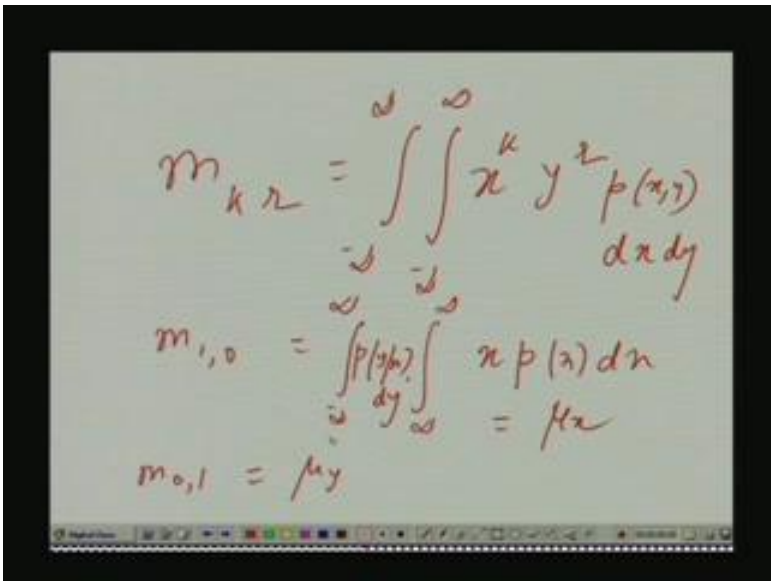
So, today, we will discuss… We will discuss this first and then we will go up to joint characteristic functions. We are given jointly random variable.

(Refer Slide Time: 51:05)



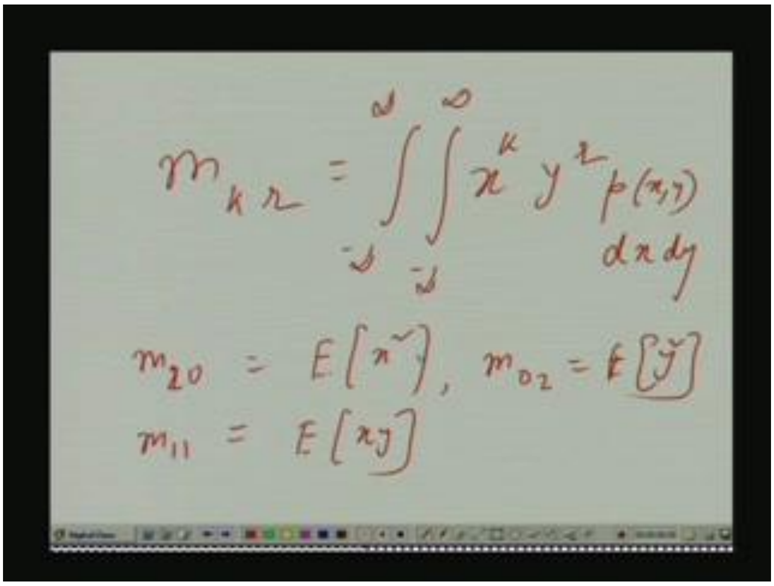
Then, we define the moment m k r as… where p x, y is the probability density – joint probability density of x, y. x to the power k, y to the power r. And this will be called joint moment of x, y of order n is equal to k plus r. So, it is very much similar to the moment that we dealt with for a single random variable case; it is a generalization of that to two variables. Certain things follow easily.

(Refer Slide Time: 52:17)



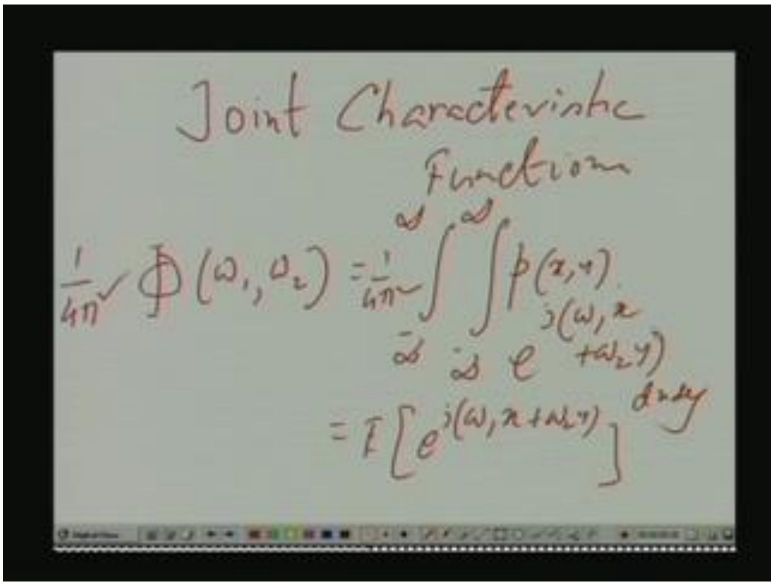
What is m 1 0? That means, x to the power 1; that is x. y to the power 0; that is 1; x p x comma y dx dy. And p x comma y can be written as that is, p x comma y can be written as… I mean you can write the entire thing like this x p x dx. p x comma y will be written like, that is, p x comma y is p x times p of y given x. So, this integral is 1 because condition theory x; total probability of y taking values within minus infinity to infinity; that is equal to 1. And this is the mean value of x. So, m 1 0 is mu x. Similarly, m 0 1 will be mu y; mu y means the mean of y.

(Refer Slide Time: 53:49)



Then, what is m 1 1? Also, you start with m 2 0. Clearly, it will be expected value of x square. Why? y 0 is 1 x square p x comma y dx dy; px comma y will be written as before like p of y by x and then px. This integral will be 1 and we get expected value x square. Similarly, m 0 2 is E of y square; and m 1 1, that is expected value of xy; that is these are all second order moments – second order joint moments.

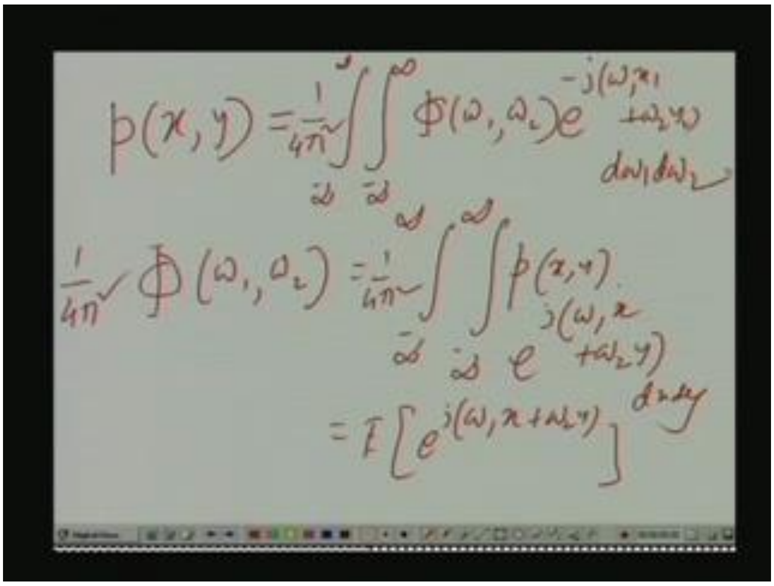
(Refer Slide Time: 54:48)



Similarly, you can have joint central moments. Here you can call it m prime k r. It is actually expected value of omega 1 x plus omega 2 y dx dy. So, given p of x comma y or you can equivalently see also; this is nothing but expected value of e to the power j omega 1 x plus omega 2 y. So, given p x comma y, we can find out y by this formula. And given the characteristic function, we can find out p of x comma y with the inverse formula; that is again obtained just by recalling the forward and backward Fourier transform relations, that is direct and indirect inverse – direct and inverse Fourier transform relations. That is you can see that, if you multiply both sides by 1 by 4 pi square, then this is nothing but inverse Fourier transform of p x comma y and omega 1 comma omega 2. See there is a plus sign here; no minus sign; no minus sign.

You recall in the one variable case. The inverse relation was 1 by 2 pi integral some function of say x e to the power j positive – e to the power plus j omega x dx. Here we have we just have two variables x comma y. So, p x comma y e to the power this 1 by 4 pi square. So, this is again inverse Fourier transform of x comma y at omega 1 comma omega 2. So, p x comma y is nothing but direct Fourier transform of this quantity.

(Refer Slide Time: 57:18)



That is, p x comma y; p x comma y is 1 by 4 pi square… So, this is the definition. And for the given phi omega 1 omega 2, you can also find out the marginal characteristic functions; that is, just phi of omega 1 and phi of omega 2; phi of omega 1 being the characteristic function of x for the random variable x alone; and phi of omega 2 being the characteristic functions of the random variable y alone. And then we will see that, when two variables are independent – statistically independent, then this joint characteristic function becomes just a product of… This is the joint characteristic function – a product of two marginal characteristic functions; then x and y – they are statistically independent. And then we will relate that to convolution and all that. So, I will do it for the next class. So, that is all for today.

Thank you very much.