Data Analysis and Decision Making - I
Prof. Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 36
MLE estimates

A warm welcome to my dear friends and dear students a very good morning, good afternoon, good evening to all of you and this is the Data Analysis and Decision Making-I course under the NPTEL MOOC series. As you know this is a 12 week lecture series under NPTEL MOOC or a course. And 12 weeks basically means this there would be 60 lectures; total a number of hours is 30 hours and each week we have 5 lectures is being for half an hour and I am Raghu Nandan Sengupta from the IME department IIT, Kanpur.

So, this as you see on the screen, this is the 36th lecture; that means, we have just completed 7 weeks we are going to a start the 8th week. Now if you remember we technically started the concept of multivariate statistics in around the end of 29th class or begin of the 30th class. So, we have already covered about 5 lectures or of 1 week.

And we covered the concept of what is a multivariate statistics considering and there are p or k number of random variables. And, what are the actual characteristics of population, what is the significance of the mean, what is of the population, what is the significance of the standard deviation of population, the concept of variance covariance matrix. The concept of correlation coefficient then the concept that when we are not able to find out these characteristics from the population. We absolutely use the sample and have the sample mean have the sample standard deviation, the sample correlation coefficient and the formulas thereof had been discussed.

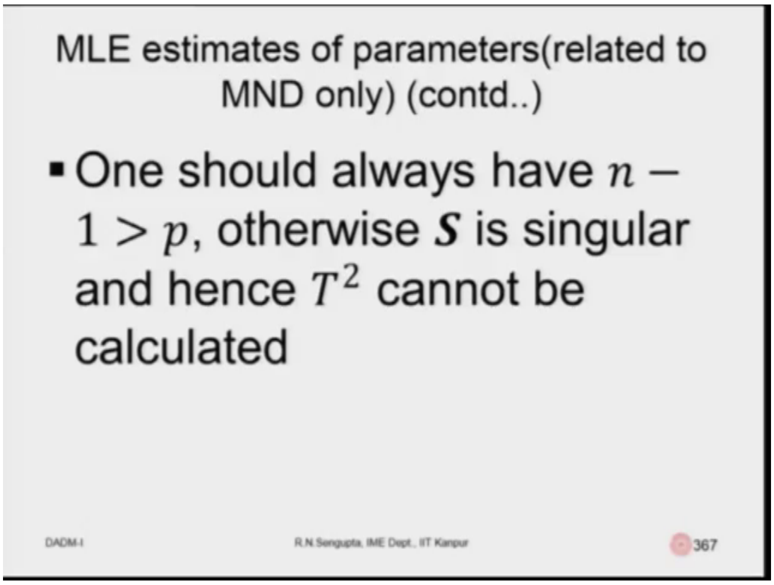
Then we went into the concept of the multinomial distribution gave the pdf of the multinomial, but the pmf of the multinomial distribution my apologies. And I also discussed a very simple example by of using contraceptive in el Salvador, then we went in the concept of other distributions like the multi normal distribution. Then discussed that it is a symmetric distribution and we discussed the wizard distribution student t distribution all in the multivariate case.

And then we went into the discussion very briefly what is a copula. We will come to the copula later on, then we discussed in the hypothesis testing case for the case of the multi variate distributions and discuss, but that was only in the concept of multi normal case. So, not for the other distribution as case and we discussed that how the sample mean and the sample standard deviations are important depending on when the population mean is known or not known, how we basically deduce it by the degrees of freedom, this concept is exactly the same for the univariate case. And we will continue discussing that further on and go into copula and other different type of multivariate statistical methods.

So, considering that considering on the multiple maximum likelihood estimations for the parameter relative the multi normal distribution; so, once we should always see that the n minus 1 n means the some of the set of observation minus 1 because, we are losing 1 degrees of freedom that should be definitely greater than p. P is basically the number of random variables which is which you have otherwise as which is basically the standard deviation matrix of size p cross p.

Because, the principal diagonal would be the variances or the standard error square when we are considering the sample and the off the diagonal element as symmetric; they would basically be the covariances for the population if you considering the population only or they would basically be the covariance's existing between the random variables considering the sample only. So; obviously, they would be replaced and the concept of consistency the unbiasness which I have been discussing time and again would also hold for the multivariate case.

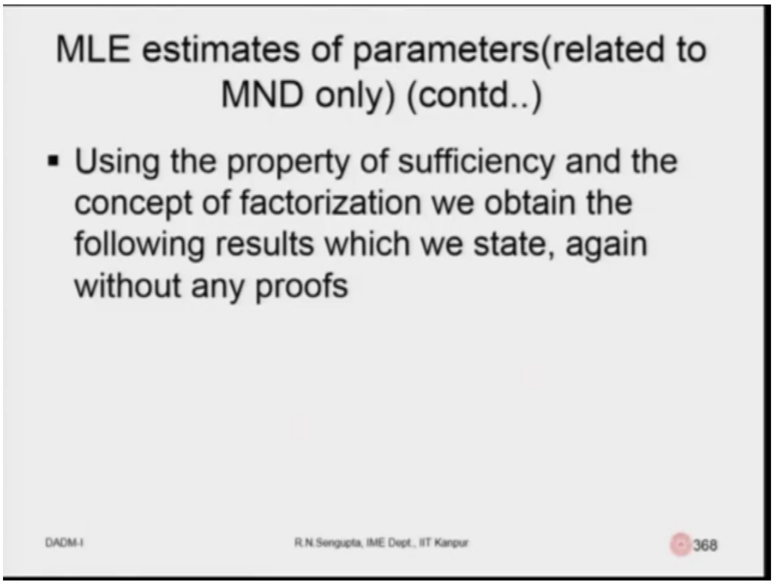
(Refer Slide Time: 04:34)



So, in case if n minus 1 is not greater than p so, an obviously, singularity concept would come and the T square cannot be calculated. Hence we cannot basically proceed with the hypothesis testing for the multi normal case. Now, remember for the normal distribution in the univariate case I am sorry I am repeating in time. And again the for the multi normal distribution case, use the normal distribution and in the case of the point distribution, you had different type of sample statistic. We are know we know about that. Now combining the normal distribution, we can find out the chi square, the t and this and the f distribution and obviously, you had the initial z distribution which is the standardized normal distribution.

So, the counterparts which we have for the chi square would with the wizard distribution we have considered. And we also saw that z and t would be utilized for the univariate case for finding on some characteristics about the mean of the population and chi square and f would be the distribution used to find out something to do with the standard deviation on the ratio the standard deviation for the population. So, we will basically follow the same concept as we do for the univariate case for the multivariate normal distribution case also.

(Refer Slide Time: 05:57)

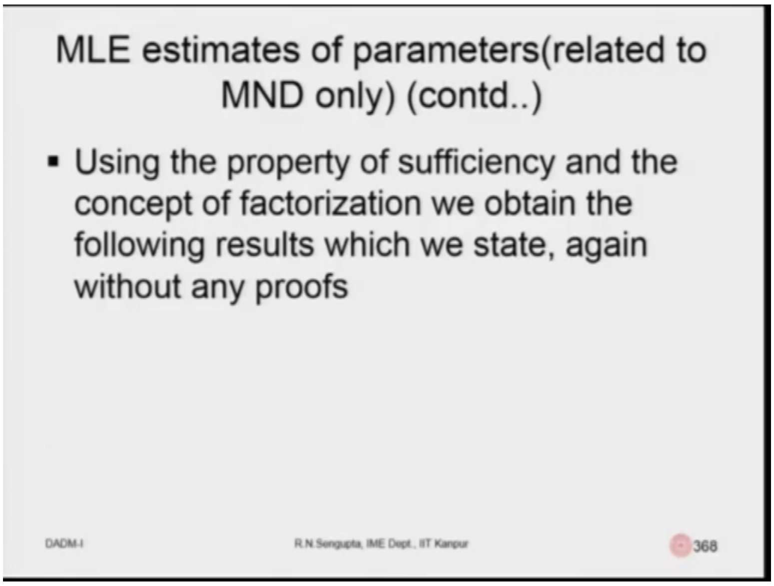


So, now using the concept of sufficiency and concept of factorization, we obtain the following results we state again without proof. So; obviously, in the maximum likelihood estimate what we did and if you remember we also discussed the general methods of moments. So, maximum likelihood estimation the concept basically remains that if you pick up a set of observations which is small and in size which are the observations, you assume that we; obviously, you do not have the values of parameters which is mu and sigma for the multi normal case and you want to some estimate that.

So, we assume the realize values of x 1, x 2, x 3, x 4 considering for the univariate case or x 11, x 12, x 13 or and x 21, x 22, x 23 so on and so forth till x p 1, x p 2, x p 3; obviously, this set of observations would be given such that we will try to maximize the likelihood function. Now likelihood function I am repeating it is basically when the random variables take that realize values, we put them plug them into the functional form of the pdf find out what is the for total probability joint probability and; obviously, we will consider their independent. That independent means for the univariate case their independent of each other from reading to reading for the multivariate case also their independent between each other and they are also independent from reading to reading.

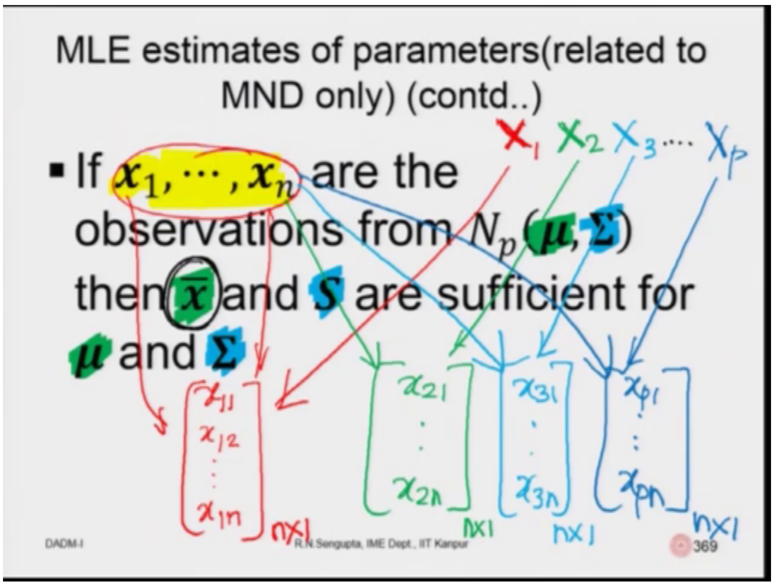
So, once we find out the log likelihood; that means, we converted into a log function because monitoring increasing, then the only things unknown are the parameters for the population we differentiate partially differentiate; that my apology is partially differentiate with respect to the parameters put them to 0 find them. Technically in normal case, they give a closed form solution. So, it will be easy for us to find out the actual sample statistic or the sample estimate which then we will try to prove they are unbiased and consistent such that we can safely use them as the best proxy in case the population parameters are not known.

(Refer Slide Time: 07:59)



So, using the property of sufficiency and concept of factorization, we obtain the following results which we state again without the proof and that would be true what I said what a general statement would be true for the univariate case and the multivariate case.

(Refer Slide Time: 08:10)



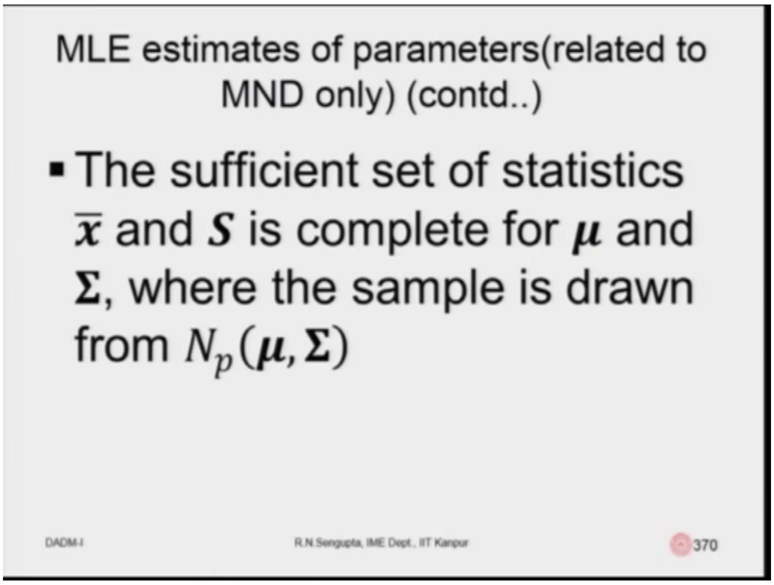
So, x 1 to x n; now x 1 to x n and all of them are vectors of size n cross 1. So, x this x 1 this x 1 we have is basically; so, this one the first one is for the random variable, second number is for the reading. So, this basically means if it is a not vector they basically mean this set which I have. So, actually what I have is this is a bold one X 1 which is this, then let me change the colour for the second variable this again a vector bold one. So, this is basically x 2 1 x 2 n x 3 1 x 3 n and this goes dot dot dot dot till the last one. Let me take a different colour this is X p or X k; whatever it you denote. So, this would be x p 1 x p n. So, each s of size sorry my mistake this colour combination would basic it help us to understand. So, this is for the first variable second variable till the nth pth variables.

So, if we x 1 to x n on the observations from the multivariate normal distribution with the mean of mu bold. So, this is a vector of size p cross 1 and this is the covariance matrix. So, I just highlight it. This is the covariance matrix and this set of values what I would you noted is this. So, basically this will come here, this will come here, this will come here.

So, the variance covariance matrix is again of size p cross p or k cross k and then using the concept of sufficiency and the concept of unbiasness consistency, the values of x bar. This is the x bar means the bold one, the vector one would basically be the corresponding estimate from the sample for each and every population mean for x 1 x 2 x 3 till x p. And, s would basically be the standard error matrix of size p cross p which would basically gives us the best estimate for the case of the population variance covariance.

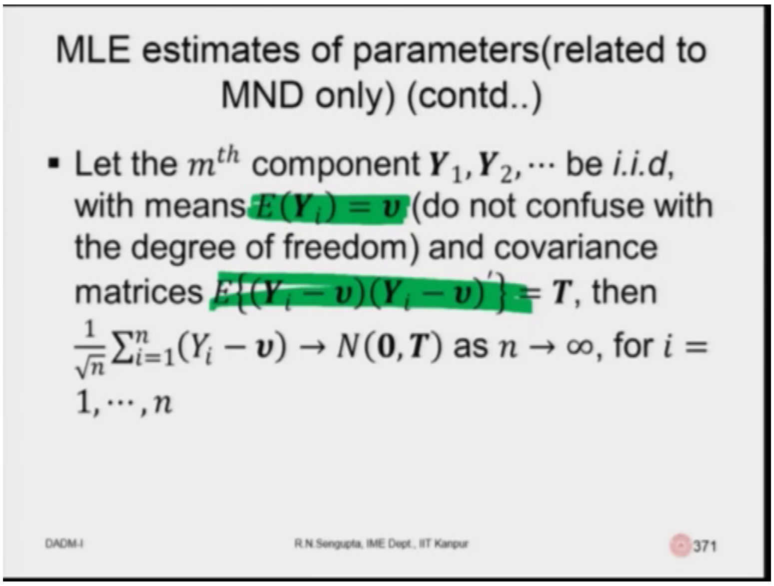
So, this would be the case let me use the highlighter. So, the green one; so, this would be for this and this one would be for this one. If somebody is confused let me remove the colour; so, consider this. So, they are the sufficient and the best estimate for the case of the population mean and the population covariance variance.

(Refer Slide Time: 13:13)



The sufficient set of statistics: which we have considering the sample mean are complete when the sample is basically drawn complete mean in all the statistical properties; if they are drawn from the total population multivariate population of mean mu and variance covariance of sigma.

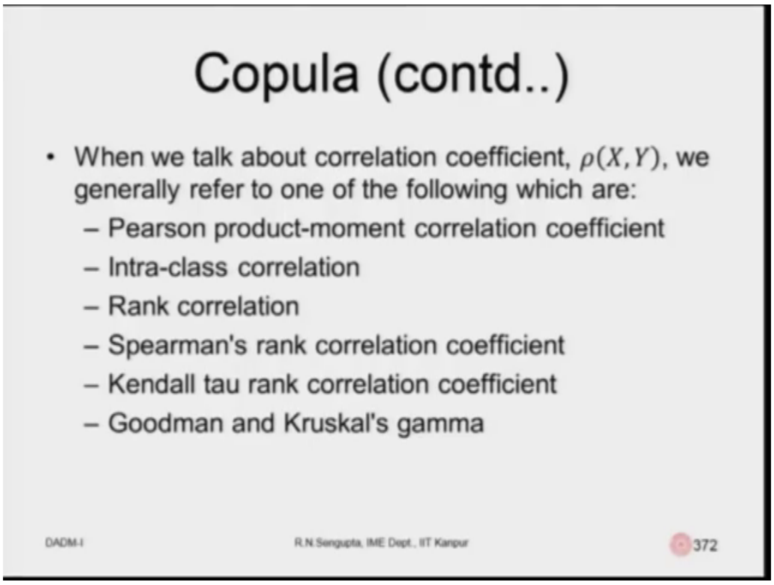
(Refer Slide Time: 13:39)



Now, consider the mth component of the Y 1, Y 2 till the set of observation which we have and they are all i.i.ds. So, in this case consider the expected value of Y i is equal to v. So, do not confuse let us not confuse with the concept of the degrees of freedom. Then the covariance matrix given for Y would be T and then in the asymptotic sense the value would be equal to with a normal distribution with 0 mean and T as the variance.

So, what we mean is that if the expected value of Y i is given as v and then the variance is given by T then in the long run in the asymptotic sense it will basically be a stand. So, called assimilate to the standard normal deviate with 0 mean and T as the variance covariance value.

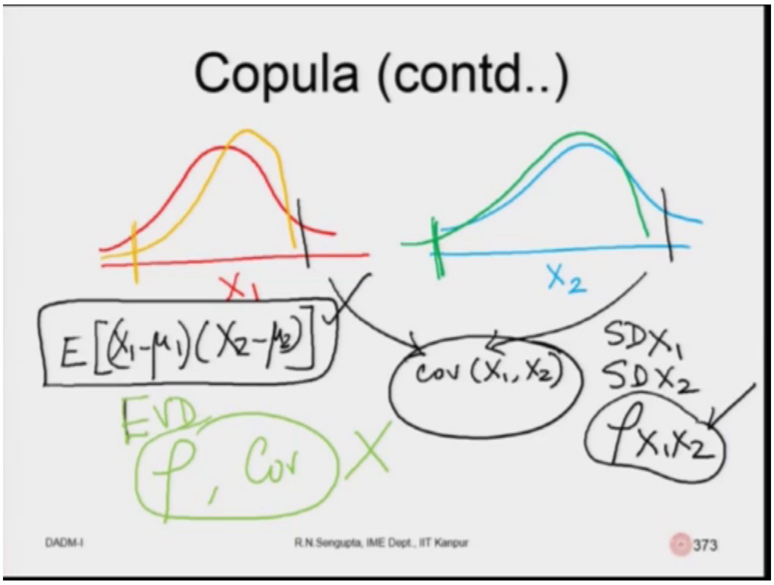
(Refer Slide Time: 14:47)



Now, when we are considering the copula, we will consider that we will talking of the correlation coefficient. So, now, let me come to the covariance correlation coefficient; concept the covariance and the copula. Now consider there are two distributions. Let me draw it let me draw it with the in a fresh blank PPT slide. So, it would be easy for us. I will come to this discussion further on. So, let me remove it just give me few minutes or few seconds ok, now I can draw it.

So, when we consider the correlation coefficient and I will come to this slide later on. So, when we consider the correlation coefficient it is like this. So, this is the diagram which will basically we discussed along with that just the previous.

(Refer Slide Time: 15:49)



So, consider is a normal distribution and there is another normal distribution, it can be more than two also. So, I want to find out the relationship for these. So, this is basically X 2 and this is basically X 1. Now, what I need to do is consider I am trying to find out the relationship of X 1 X 2 at the extremes. So; obviously, in that case covariance between X 1 and X 2 would be sufficient such that from this given the standard deviation of X 1 standard deviation X 2, we can find out the correlation coefficient existing between X 1 and X 2. But what happens if the distributions and; obviously, in that case covariance is fine in case gives us good relationship.

Now, see if I expand the covariance the formula is actually this. So, this is the linear relationship in terms of X 1 and X 2. But what happens if these two distributions, let me use two different colours of orange on a green one and I am interested for the extremes values on to the left or right also. Now, remember here for the exponential distribution, the values fall exponentially bought on to the left on the right. But in the extreme value distribution the corresponding probabilities and those corresponding points which are on a equivalent scale distance on the EVD scale, on the extreme value distribution scale and the normal distributions are unequal.

And the corresponding probabilities would also be unequal such that for a particular value happening in the normal distribution case. And, its probability if we compare that and the same probability when I consider for the extreme value distribution for the same type of value on the EVD scale, the probability in EVD scale would be higher.

Hence, when we try to find out the actual values of the excessive loss; loss I am using in a very general sense then the corresponding probabilities, even if they are little bit higher; they have a catastrophic effect or a huge amount of effect on the overall relationship between X 1 and X 2. That means, per unit change in X 1 and is actual value on X 2 or vice versa in the normal case would be given basically by the correlation coefficient which is fixed for any values of X 1 and X 2 you consider.

But when we consider the extreme value or some non symmetric values or distinctions for which the ends probabilities are higher than a normal case, then the relationship is non-linear. That means, per unit change in X 1 or in X 2 the corresponding change in X 1 and X 2 are definitely not linear, then we use the concept of correlation which means in the case of the normal case this covariance and this correlation coefficient are fine. But in the extreme value case let me use, extreme distribution case correlation coefficient or covariance do not give us the right picture with respect to the relationship between X 1 and X 2.

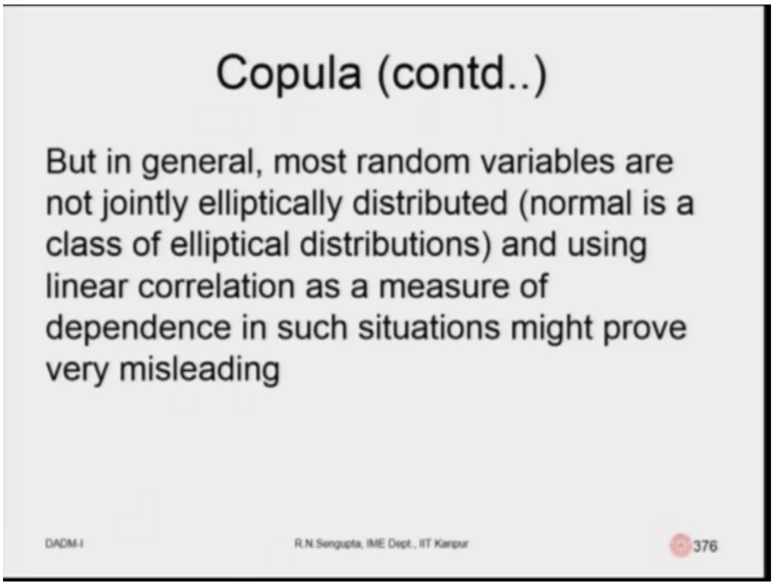
This I will keep repeating it when we come to the copula, but this is just a very brief background which I wanted to mention when you are basically going to study the concept of copula in the multivariate case. Obviously, considering the whole number of sets or topics we have to consider, it may not be possible of us discuss everything, but I thought I will just give you a small brief background about that ok. Let me go back to the initial slide.

So, when we talk about correlation coefficient between X and Y we generally prefer to use on a one of the followings which are the Pearson product- moment correlation coefficient, the intra class correlation factor, the rank correlation depending on what ranks the air and how you basically find out the relationship, the Spearman’s rank correlation coefficient, the Kendall tau and the Goodman’s fit or the Kruskal gamma which are basically used to find out the relationship between two random variables and we consider them to be linear dependent structure.

But as I said with repetition I am again saying that for extreme values, this linear structure at the extreme does not happen. That means, per unit change of X 1 or X 2 would have non-linear effect on X 1 and on X 2 and X 1; that means, X 1 changing one unit would have a non-linear effect on X 2 and similarly one unit changing in X 2 would have a non-linear effect on X 1.

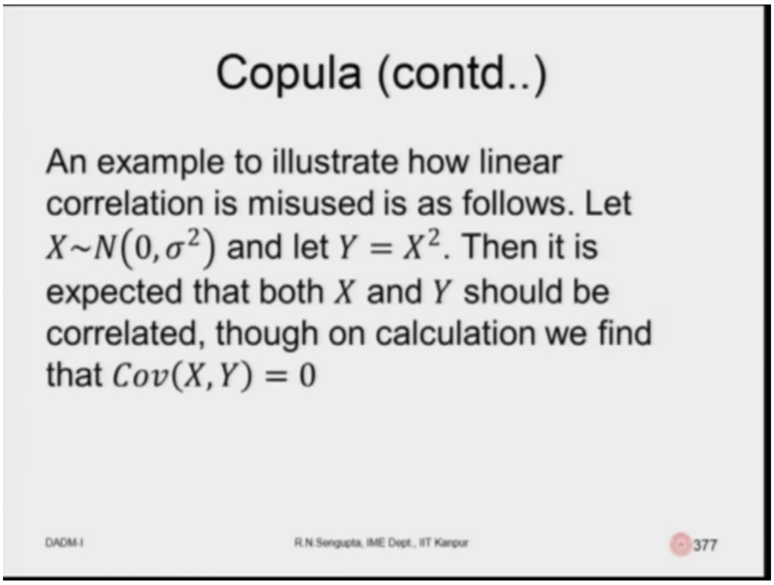
So, this is not there in the case when you consider in the correlation coefficient or the different type of very simple linear correlation coefficient concept. In order to overcome that, we use the basically the copula concept which we copula concept is basically if you remember I I mentioned is a sort of mapping which is happening between the marginal's and the joint distributions in order to give us a good picture that the how the relationship between X 1 and X 2 or if there are more than X 1, X 2 in maybe X 3, X 4. These variables are there and we are basically trying to understand the relationship and the extremes. So, these blank slides I will try to utilize that as it is.

(Refer Slide Time: 22:16)



So, again continuing, but in general most random variables are not jointly elliptical distributed as normal distribution is a case, because normal distribution is class elliptical distribution. And using linear correlation as a measure of dependent structure in such situations might prove very misleading because it would not give us the actual interrelationship between two random variables which are not elliptical. So, hence we use copula.

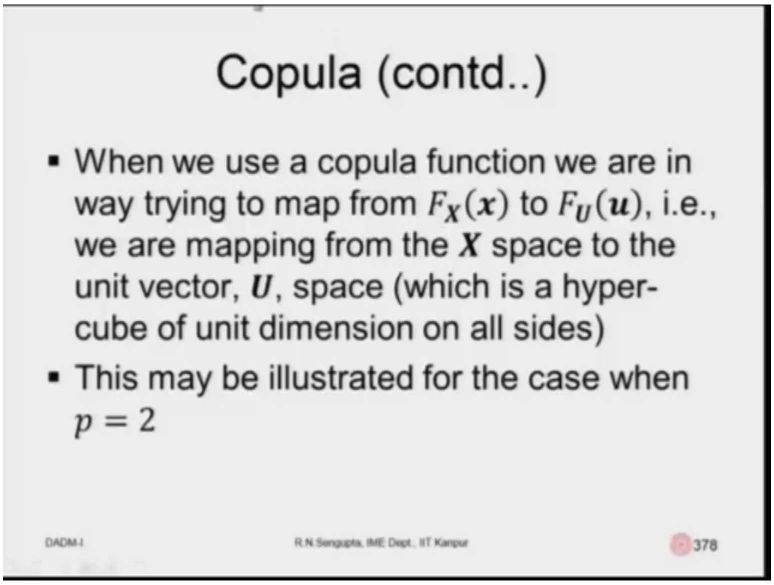
(Refer Slide Time: 22:43)



As an example to illustrate how linear correlation is misused, let us consider as follows. Let us consider x is normally distributed with 0 mean and sigma square a standard deviation or variance and let us find out a random variable Y which is the square of X, X which says that Y is equal X square, then is expected that both an X and X, Y and X and Y should be correlated. But, if we try to actually find out the covariance between X and Y, it will come out to be 0 which means that the actual information which you are trying to get using the covariances on the correlation coefficient would mean that we are not able to portray the relationship between X and Y.

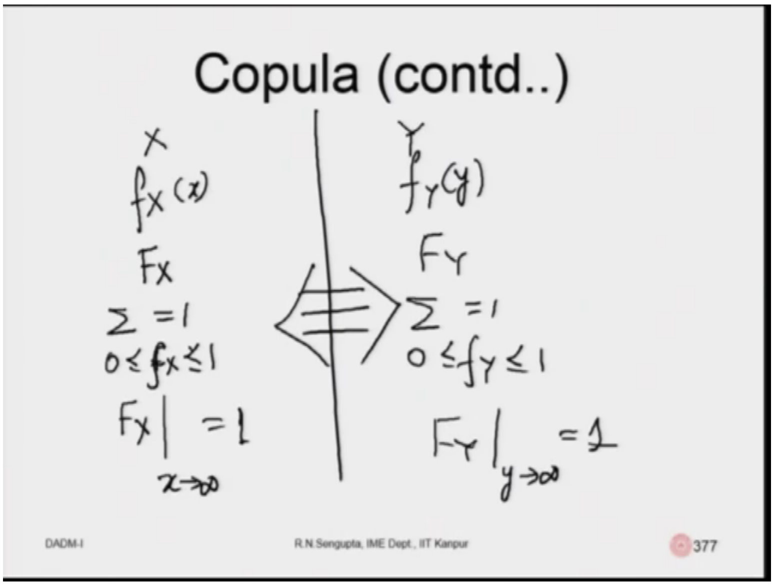
In this case it a Y is basically X square such that it does not give us what is the relationship happening between the two random variables even though they are related in; so, in order to overcome that you basically use the copula function. When we use a copula function we are in a way trying to basically map from the cdf of the actual distribution which is X into the univariate case such that this is a one to one correspondence.

(Refer Slide Time: 24:00)



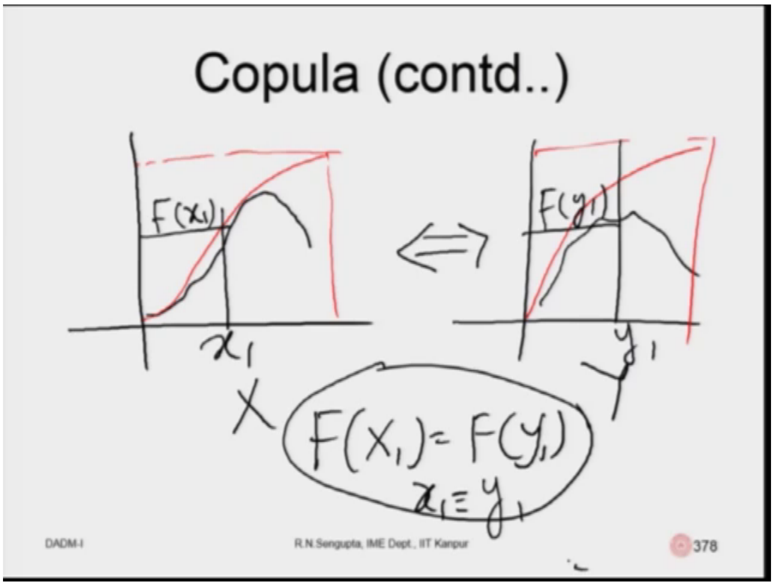
Now, this I will again explain. So, I will come to the second bullet point which is to illustrate for the case when you have p is equal to 2 random variables; obviously, can we increase for the third case also for three dimension. So, what we are doing is like this now background which have did discuss, but I will again consider it again.

(Refer Slide Time: 24:25)



Now, consider there are two random variables X and Y and f of x is the pdf. This would be true for the univariate case also sorry for the discrete case also and this is the pdf of y. So; obviously, all the properties would hold; in the sense that if we are consider capital F of X which is the cdf, capital F of Y then; obviously, it would mean sum of f of x is equal to 1. The values of capital F capital F; so, this is small f let me clarify this. So, sum is 1 sum is are thus the value of pdf of f of y small f of y is between 0 and 1. And F of X natural value starting basically for x tending to positive infinity F of Y y tending to positive infinity considering, then the max maximum value this is 1; obviously, this would be true always remember that. Now, if this is true, we will try to basically bring a one to one correspondence here, what we do is like this.

(Refer Slide Time: 26:02)



Let me draw the pdf of X. So, consider the pdf and consider for whatever this is for Y. So, this is X this is Y. Now consider the cdf function so; obviously, cdf function would be between 0 and 1. So, the cdf value would be this sum cdf value is this which means that in case if we are doing a one to one correspondence between them, then at a certain value of x 1 the F of x 1 which we have capital F of x 1 would have a corresponding value of y 1 such that this would always be true. In case if that is true; obviously, then sorry then the value of x 1 and x 2 would be such that if I add up all the probabilities of x is still the x 1 value add them up and add up all the properties for y till y 1 and add them up both of values are equals such that the total probability cover till that point of x 1 and total probability cover till the y 1 are equivalent.

So, if you are able to use in place of y a univariate distribution case or a very simple distribution which we know then doing a one to one mapping would basically solve our problem. That is what the main crux of the problem or what the essence of how we are going to tackle the copula theory would be. I will discuss that in more details in the 37th class. With this I will end the lecture and I am I know that it is a little bit more theoretical than univariate case, but I will request all the students to please bear with me as I slowly covered the theories and also cover some of the simple concept from the multivariate case.

Thank you very much and have a nice day.