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

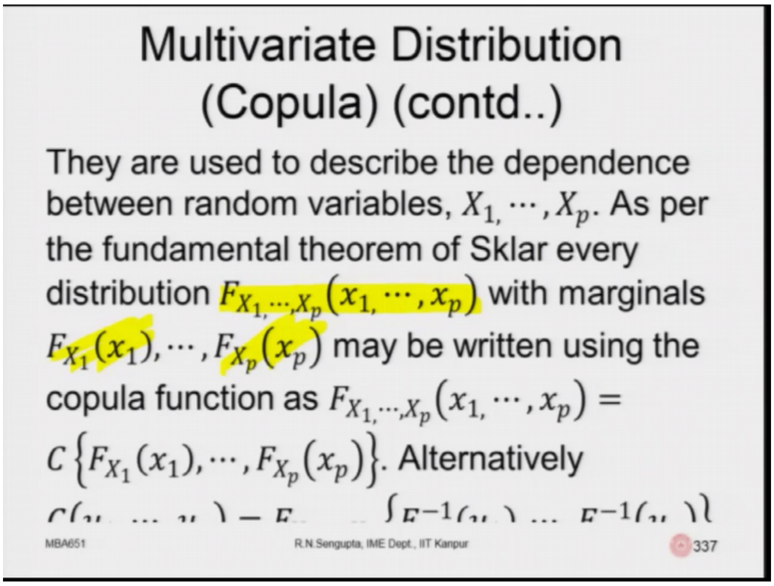
Lecture - 33
Multivariate distribution

Welcome back my dear friends, a very good morning, good afternoon, good evening to all of you. And as you know this is the DADM course which is Data Analysis and it Decision Making and on the NPTEL MOOC series. And this course is for 12 weeks 60 lectures and that is total combined hours being 30 and each week we have 5 lectures each being of half an hour. So, we are in the 33rd lecture and in the 32nd lecture if you remember, we have just discussed very briefly that what is with an example for the multinomial model that contraception example and we do the graph. Later on we went into disk mass mentioning what is the multi normal distribution for the p very p variable or k variable and how it has a one to the, in the formula sense, how easy does a one to one correspondence with the univariate normal distribution.

And then we considered the bivariate normal distribution wrote down the formulas also considered the relationship between the two normal distributions whether they are not related whether they are related positively whether they are related negatively. And we saw the contours, which was looked like one to looks likely lips with a major and minus i axis and one looked like in a circle that the concentric circles, we are taking the slices looking from above. Then we considered the student t-distribution.

So, all these things are background are giving the examples and then how the students t-distribution was exactly for me in the same way as the univariate student t-distribution. Then we are then very briefly, I mentioned about copula theory, I will come to copula theory later on in details and in copula theory I did mention that and the extremes whether high values or low values of x. The relationship at the extremes where the negative or positive between x and y need not be linear, they can be non-linear also. So, in order to basically consider that we consider the copula function which is a function of the univariate distributions u 1 u 3 u 3 till u p and each of them work basically uniform discrete distribution or continuous distribution whatever the case because continuous distribution at the best way between 0 and 1. And copula functions are very heavily used in order to find out the relationship at the extremes. So, continuing the copula function.

(Refer Slide Time: 03:06)



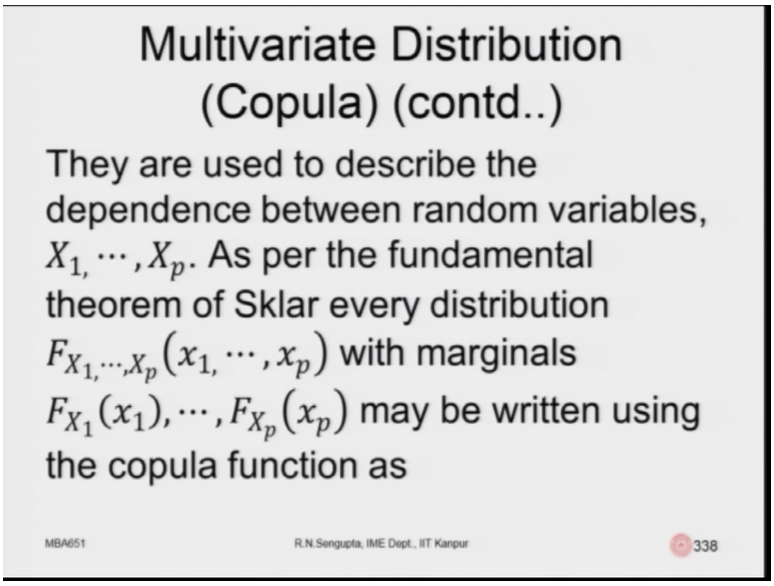
So, they are used to describe the dependence between random variables X 1 to X p as per the fundamental theorem of Sklar every distribution, every joint probability distribution.

So, this is the joint probability distribution because not it is no more running univariate, it is basically the relationship between X 1 to X p and the marginals are given. Marginal means the cope in univariate cases of each as is given. For the X 1 similarly X 2, X 3, X 4 till the pth one. So, all of them are univariate normal for the marginals. So, in that case, the joint probability distribution can be related by a copula function with the variables of the copula functions are the marginal themself. And what we will do is that later on see how the marginals can be used converted using the univariate continuous case between 0 and 1. The concept is intuitively very simple. Now for a univariate uniform continuous distribution between 0 and 1, the sum of the probability is always 1 which is true because the cdf value is 1. What you do is that you map that corresponding value 1 to basically consider the marginal any marginal it is F X 1, F X 2, F X 3 where x 1, x 2, x 3 all are normal distribution so; obviously, that area is also 1.

Now, depending on the value of x or depend x 1 let me use the word x 1 or depending on the value of u 1 you can find a 1 to 1 correspondence between u 1 and x 1. So, you have different realize values of x u 1, you will have basically different realize values of x 1.

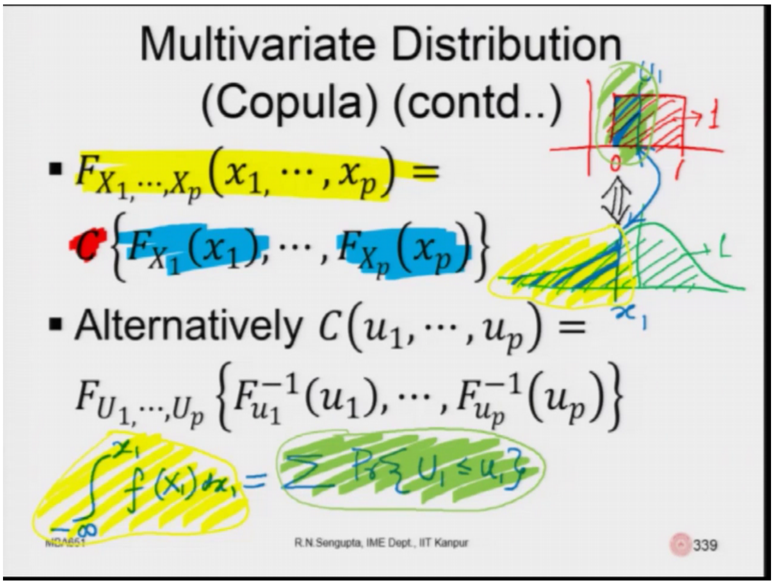
Similarly you can have different realize values of x 2 would have different realize values of u 2 probe. So, underlying fact remains that the cdf value for the normal case from 90 minus infinity to that value of x 2 would be exactly equal to the cdf value sum of the probabilities for the uniform continuous case from 0 to that value which is u 1 on u 2 as the case may be and then you can find out the copula function.

(Refer Slide Time: 05:47)



Now, these copula functions are used to describe dependency between random variables X 1 to X p. As per the fundamental theorem of Sklar every distribution which is the joint-distribution with marginals F X 1 to F X p may be written using the copula function.

(Refer Slide Time: 06:02)



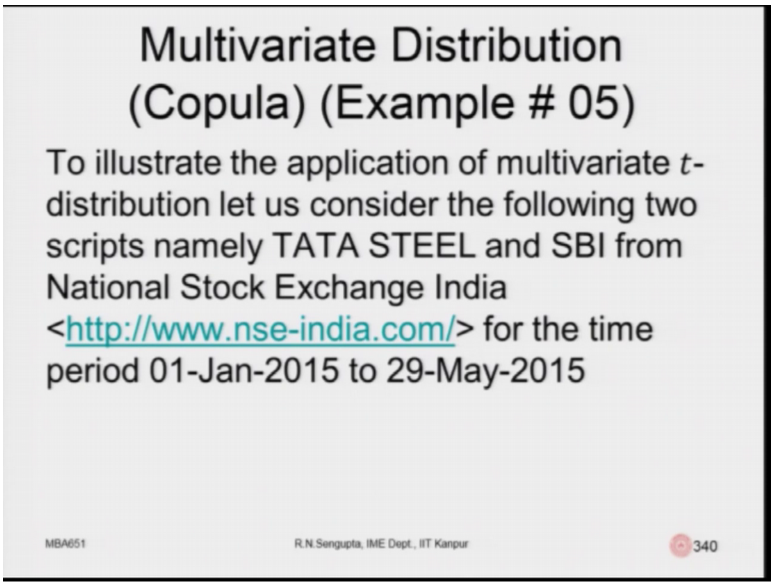
So, this is the copula function which I just mentioned. So, you have the joint distribution of X 1 to X p that is equal to the copula function of the marginal. So, let me use a different colour. So, this is the copula function which you have which you would not need to find and these are the marginals. This is basically a function of the marginals being mapped on to the joint-distribution. So, alternatively copula function of now what you do is that, you convert this X 1 X 2 X 3 distribution corresponding to the fact that the some of the probabilities of all them is equal to 1. You map them on to the univariate case and do the problems accordingly.

So, I will come with the diagrams and of these things later on. So, let me write it in screen. So, consider you have a uniform continuous case between 0 and 1, 0 the axis I have shifted out to the left. So, this cdf value is exactly equal to 1. So, the sum is 1. Now consider in the other case, you have another distribution and that is normal. So, this is normal. So, this distribution value is also 1, I need to do a 1 to 1 correspondence between them. So, 1 to 1 correspondence between these two distribution. So, what I do is then the in, so I will write in and let me use another colour; let me use the blue one yeah. So, the integration from minus infinity two (Refer Time: 08:16) I use a blue colour here in order to highlight. So, consider this value consider a value here. So, these are corresponding values and I am basically trying to map these two values for equivalence. What I mean by equivalence I am going to come to that is colour should be seen yes should this the values of the cdf addition should be exactly equal to this right. So, this is done.

Now, what do we do? And you use this one so; consider this value is I am writing in blue. So, it will the colour difference will make you understand this is x 1. This value is here would be say for example u 1. So, x 1 f of X 1 d x 1 will be equal to summation for probability of u 1; this is the random variables. So, this is a random variable capital, this is capital is less than equal to u 1. So, this part the p use green colour this. So, this part and this part are equivalent. It means if I take that colour the overall formula which is given here and the area which you had here would be exactly same, point 1. Point number 2, this part and this part are again equivalent, but there is equivalence in the sense the total area is same thing.

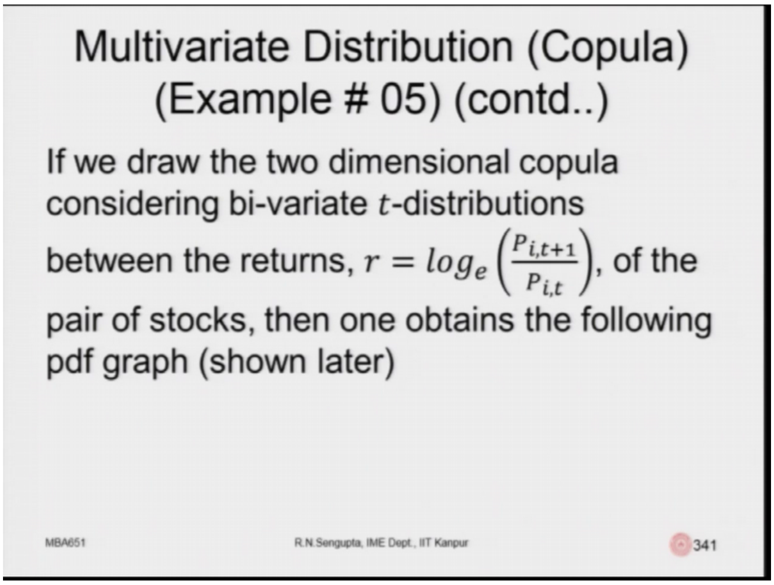
So, this part and this part area wise they are same. So, you use the equivalence and basically find it out. So, that generally we use for generating pseudo random numbers in the computer. So, that is basically the general essence.

(Refer Slide Time: 11:39)



So, very simply to illustrate the application of multivariate t-distribution, let us consider the following two scripts. One is TATA STEEL, one is SBI from the National Stock Exchange date being first to January 2014 to 29th May 2015 and we take the data from NSE India. So, it can be found in Yahoo, Google whatever, but I am taking the end of the day price.

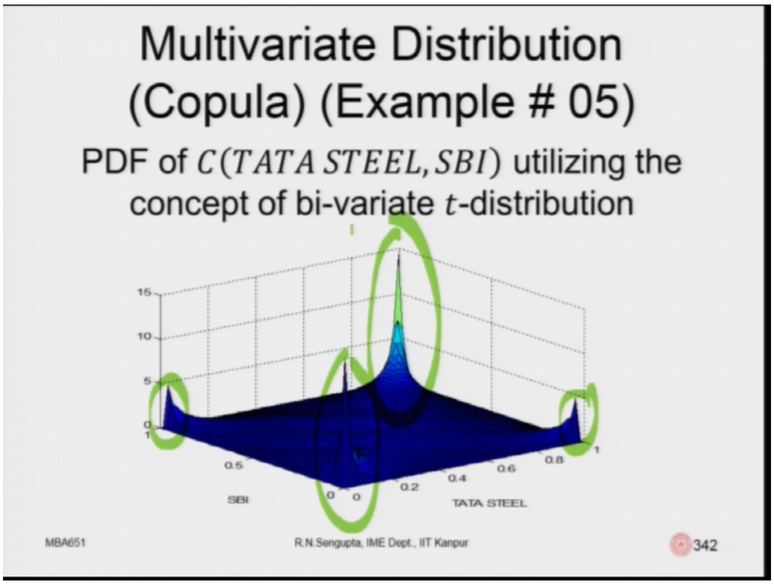
(Refer Slide Time: 11:59)



So, if you draw the two dimensional copula considering a bi variate t-distributions considering the returns, returns would be given by an opinion log base e of the today's price divided by yesterday's price. So, if I am so, that return would be given for yesterday. So, if I want to find out the return for today, it will be the law log base e to the and the ratios of the prices would be tomorrow's price by today's price.

So, that is basically the return of today. So, we find out the returns in this way of all this two pairs of stocks and then based on that one can find. So, before solving that problem let me mention the one can find out the mean values for both the stocks, the standard deviation of both of stocks, the covariance of those stocks, but here we try to basically use the student t-distribution.

(Refer Slide Time: 12:51)

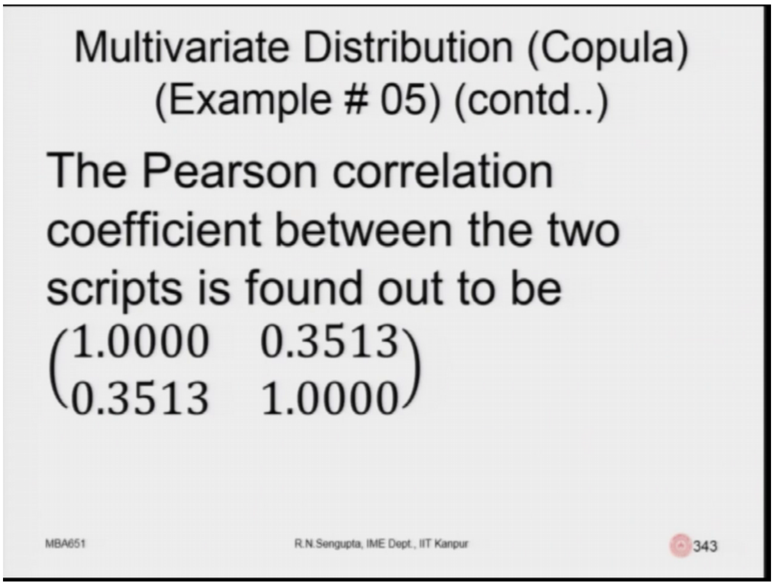


So, we use the copula function using the student t-distribution. So, here now you see very interestingly I am using copula functions also a plus the bi-variate student t-distributions. So, where I am trying to basically use the concept which you have just covered in trying to highlight how they can be used. I am not going to the details of the calculations just the flavour. So, the p d f because student t-distribution if we remember or indium is basically a continuous case and for you; obviously, you would have understood by this now in the chi square f and t and z all are continuous distributions. So, I you along the x axis we plot TATA steel, along y axis you plot SBI and we get the p d f for the student t-distribution on the using the copula functions this is the concept of the bivariate t-distributions using copula.

So, these are the values and if you see that at the extremes when both the in the univariate case; obviously, you have converted into univariate case. Add the extremes means, add the value of 0 for the SBI and 0 for the TATA steel or 0 for the SBI one value. These are the univariate conversions which are talking about so; obviously, the student univariate student t-distribution for SBI would be convert it in univariate continuous case between 0 and 1; similarly for TATA steel. So, I am considering the relationship between SBI and TATA steel at their extremes. So, rather than talking about the extremes I will talk about the point 0 0, 0 1, 1 0, 1 1. So, you see these values at the value of 0 0, this is the relationship as the value of SBI and TATA STEEL. So, we basically in SBI being 1, TATA steel being 0 and TATA steel being 1 SBI being 0 and the final for both the values of 1 1.

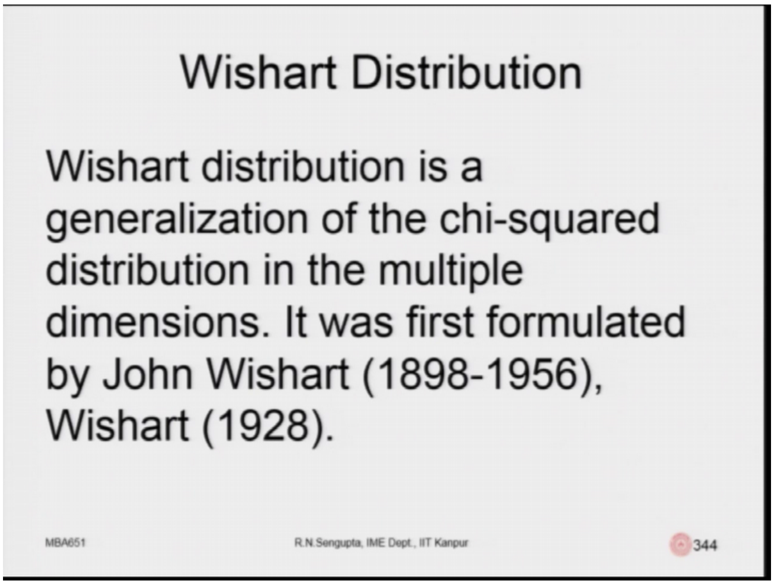
So, you find out the extremes the actual concept of the copula comes into the picture in the right way.

(Refer Slide Time: 15:14)



So, I will mention few coefficient relationship, these are not immediately important we will see that as we proceed. The Pearson correlation coefficient between the two scripts can be found out. So, the Pearson correlation coefficient is exactly like the correlation coefficient which we do so; obviously, the principal diagonal be 1 1 and of the diagonal element values 0.35 which means the correlation coefficient existing between these stocks is about 0.35.

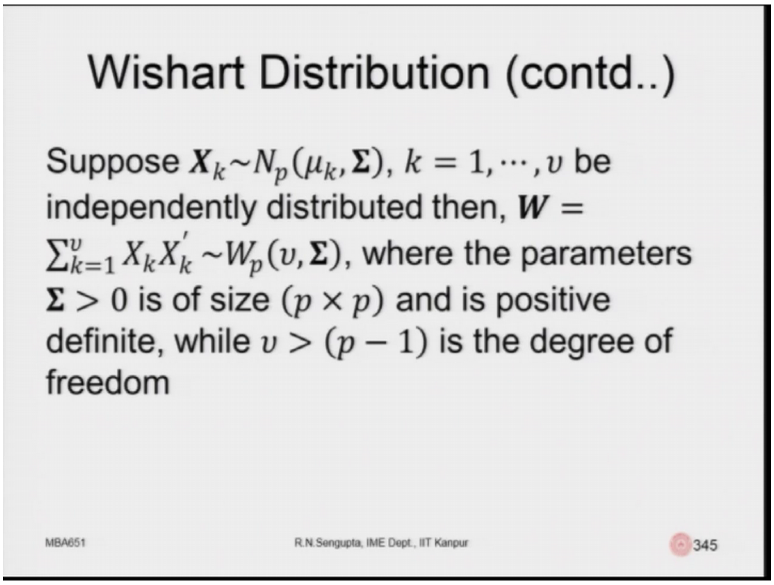
(Refer Slide Time: 15:49)



Now, then we come to the Wishart-distribution is basically a generalized. So, if you remember I am going in a very scientific manner for the univariate normal, then I discussed the p value variate, multivariate normal distribution, then you have initially you had the student t-distribution, then I discussed the multivariate student t-distribution. Initially you had the chi square of all these are the univariate chi square, then I am going to discuss the multivariate chi square which is the Wishart-distribution.

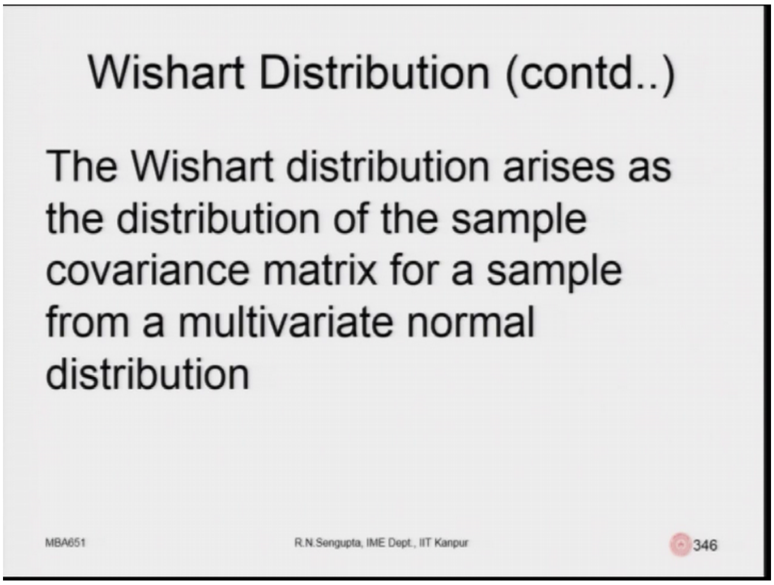
The Wishart-distribution is a generalized of the chi square distribution the multi-dimensional case on a multiple dimension. It was first formulated by John Wishart. So, hence we are basically given the distribution as such.

(Refer Slide Time: 16:41)



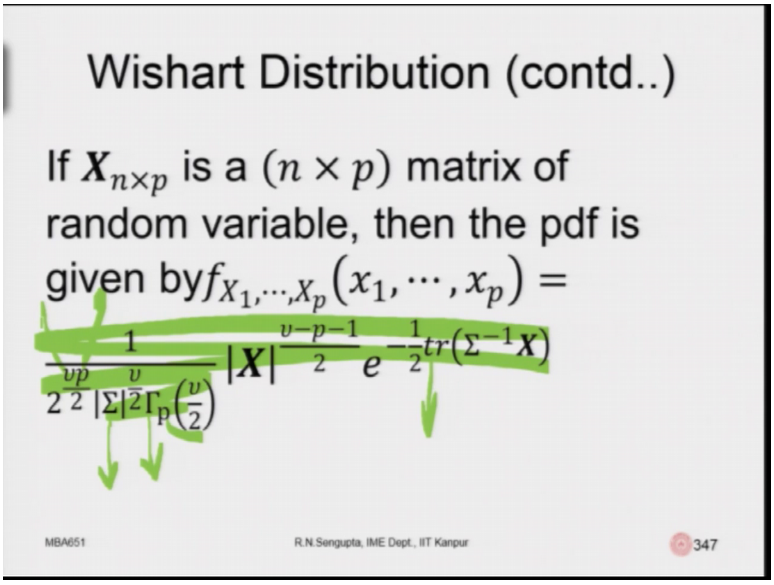
So, suppose X k. So, generally we will consider X k as being 1, 2, 3, 4 to v which is basically the degrees of freedom. So, in that case you will have the Wishart-distribution given as the multiplication of X k and k which are basically normal with certain mean in certain standard deviation which is the covariance matrix variance matrix. So, here the covariance variance would be size of p cross p and it will be a positive definite. When the value of v would definitely be greater than p minus 1 and you can solve the problems accordingly. So, this would be the degrees of freedom.

(Refer Slide Time: 17:29)



The Wishart-distribution arises at the distribution of the sample covariance matrix for a sample from a multivariate normal distribution. So, as if you remember the chi square used to come out for the case when we are going to discuss something to do with the standard deviation. So, it will be chi squared with n degrees of freedom chi square with n minus or degree of freedom. Similarly the Wishart distribution will be rising for the case when you want to basically study something to do with the covariance matrix for the multi normal distributions. In the similar way 1 to 1 simile you can see an similarity.

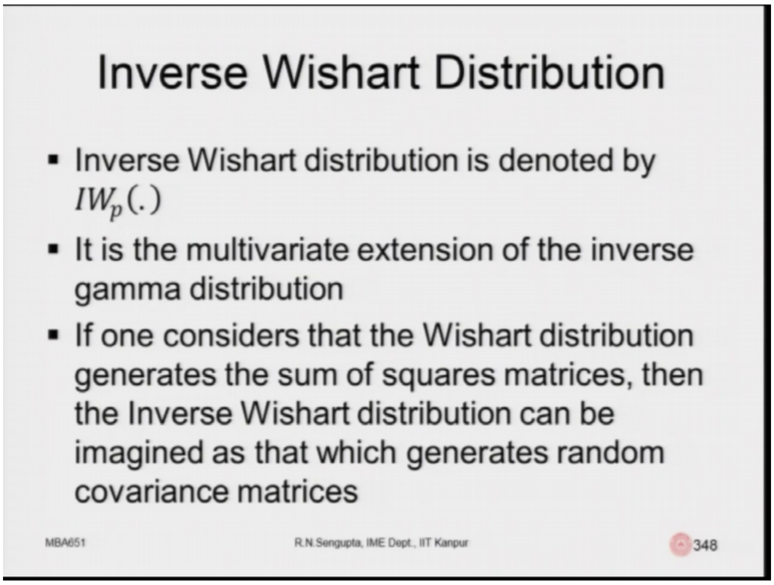
(Refer Slide Time: 18:09)



Then the Wishart-distributions basic it would be basic n if n is a matrix of n cross p where n is the number of readings, p is number of random variables. Then the Wishart-distribution would be given by the formula which is here.

So, this is the trace, t r is the trace and given this formula. These are very simple this p is the number of random variables this is the gamma distribution value. This is the variance covariance matrix, p is b as I discussed is the number of random variables v is basically the degrees of freedom. All these things are known to us.

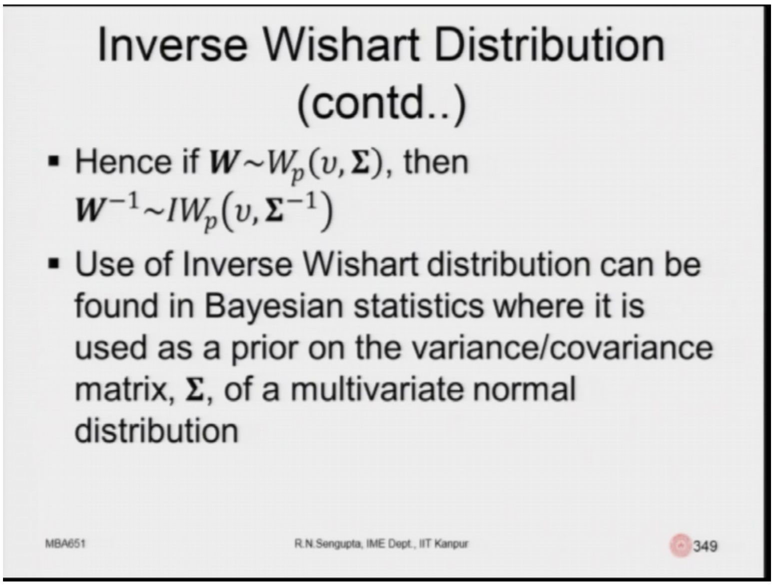
(Refer Slide Time: 18:57)



Now, will consider the inverse Wishart-distribution in an inverse Wishart-distribution is denoted by IW. So, W was for the Wishart-distribution IW would be the inverse Wishart-distribution and the suffix p or k would denote the number of random variables which are independent. It is the multivariate extra it is again in the univariate case you have the simple in inverse gamma function.

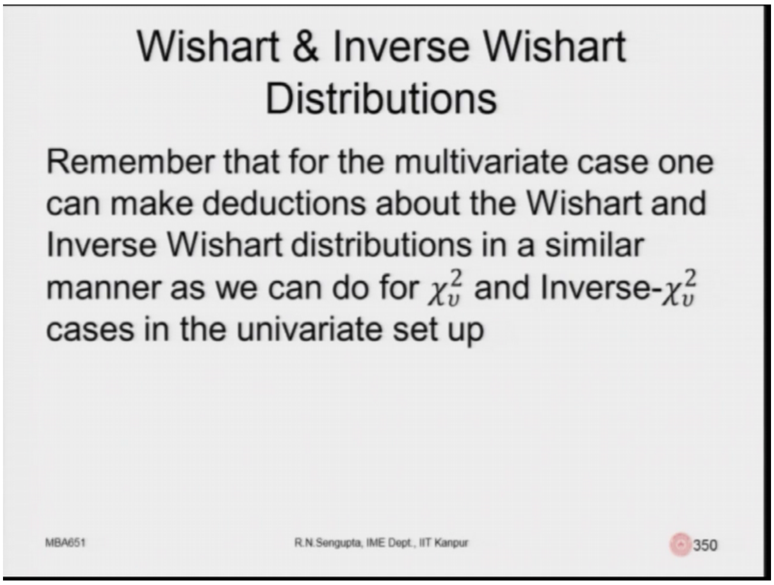
So, the extension of the inverse gamma function in the multivariate case would basically be the inverse Wishart-distribution. If one consider that the Wishart-distribution generates the sum the squares matrices, then the inverse Wishart-distribution can be imagined as that which generates a random covariance matrices based on which we do our studies. So, inverse gamma distribution Wishart inverse Wishart-distribution have a again similarity in the univariate and the multivariate case.

(Refer Slide Time: 19:55)



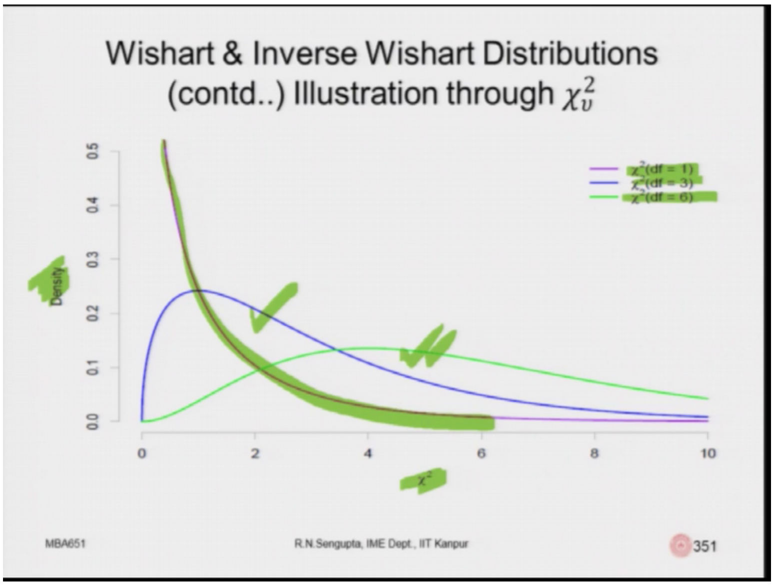
So, in this case if w is Wishart-distribution with degrees of freedom of v and the variance covariance matrix 7 then inverse Wishart-distribution would also have the same degrees of freedom and the other parameter rather being the covariance matrix it covariance covalent matrix into the inverse of the variance covariance matrix. So, user inverse Wishart-distribution can be found in Bayesian statistics where it is used as a prior on the variance covariance matrix of rho of a multivariate normal distribution and based on that we do the studies.

(Refer Slide Time: 20:35)



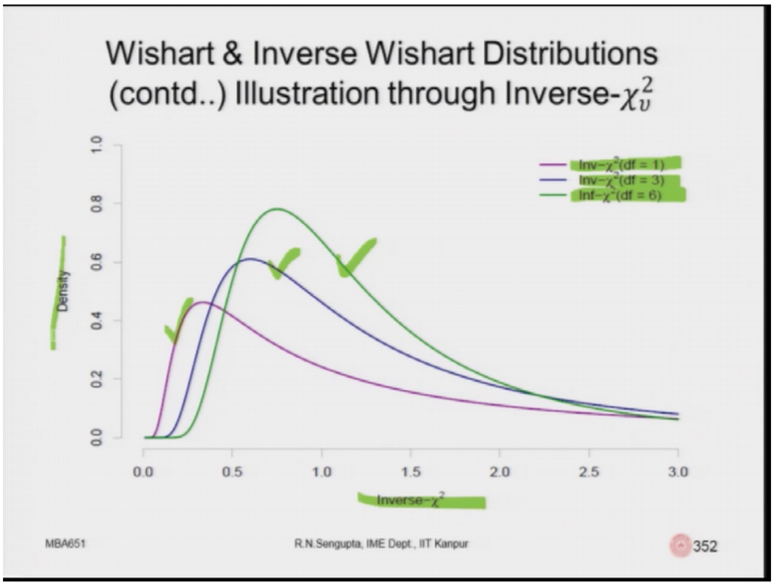
So, remember that for the multivariate case one can make deductions about the Wishart-distributions and inverse Wishart-distributions in a similar manner as we can do for the chi square and inverse chi square which is done in the univariate case. So, Wishart-distributions or the inverse Wishart-distributions would basically be in the counter part from the univariate case would be the chi square and the inverse chi square distribution and will try to basically see that to the diagram.

(Refer Slide Time: 21:11)



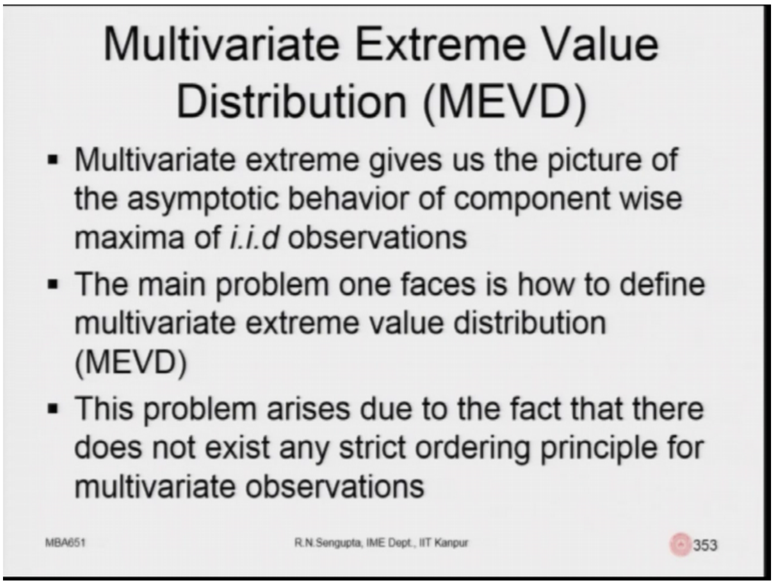
So, this is the Wishart and the inverse Wishart-distributions is being considered. See our illustration is basically through the chi square with v degrees of freedom. So, chi squared with one degrees of freedom is this pink one and this is x value x axis is the chi square, along the y axis here the density, then you have 3 degrees of freedom, 6 degrees of freedom. So, this is for 3 and this is for 6 degrees of freedom based on which we draw. So, this is the univariate chi square so; obviously, you can if you are able to draw for the higher dimension, you will get the Wishart-distribution because you remember the one to one similarity. When we have the inverse for the univariate case, you have the inverse Wishart-distribution.

(Refer Slide Time: 22:09)



Again degrees of freedom being 1 or 3 or 6 using the pink, blue and green line you will basically have the inverse Wishart-distribution. We inverse chi square distribution which is the counterpart in the univariate case for the inverse Wishart-distribution and again you can draw it accordingly. Again along the x axis your inverse chi square and along the y axis you have the density plots.

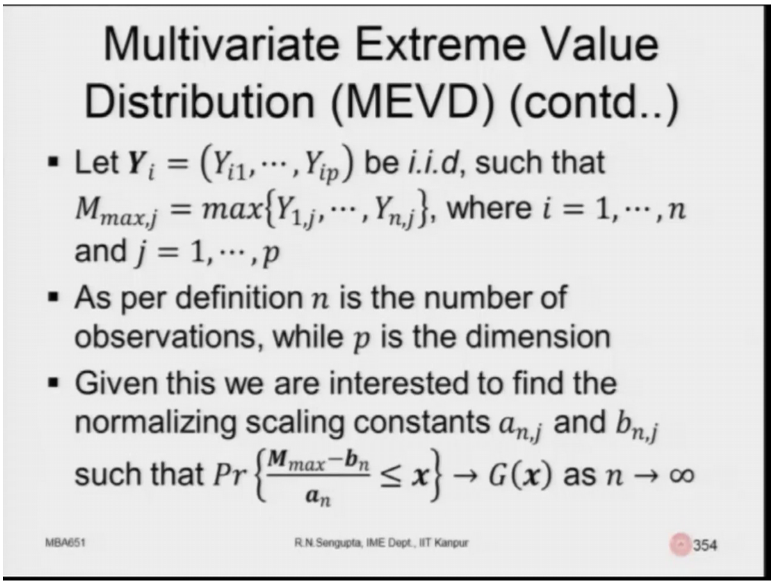
(Refer Slide Time: 22:40)



Now, we will. So, again I am repeating this just a thorough way of trying to basically give you a feel what things are there. So, multivariate extreme values, it gives us the picture of an asymptotic behaviour of the components wise based on the maximum the minimum for a i.i.d. So, the main problem one faces is how to define multivariate extreme value distribution which is MEVD multivariate extreme value distribution. This problem arises due to the fact that there does not exist any strict ordering principle for the multivariate observation.

So, if the problem is if x 1, in univariate case is very simple to basically rank them from the lowest to the highest, but say for example, when I am doing for you multivariate case it may be possible the random variables x 1s are ranked. But the random variables x 1 values need are not possible to rank because they are basically not ordered similarly when you try to do it for the x 2 rank them, it may happen that in the momently moment you do that x 1 is not ranked or ordered. So, trying to basically have an uniform ranking system from for x 1 to x p is not possible in the multi dimension extreme value case.

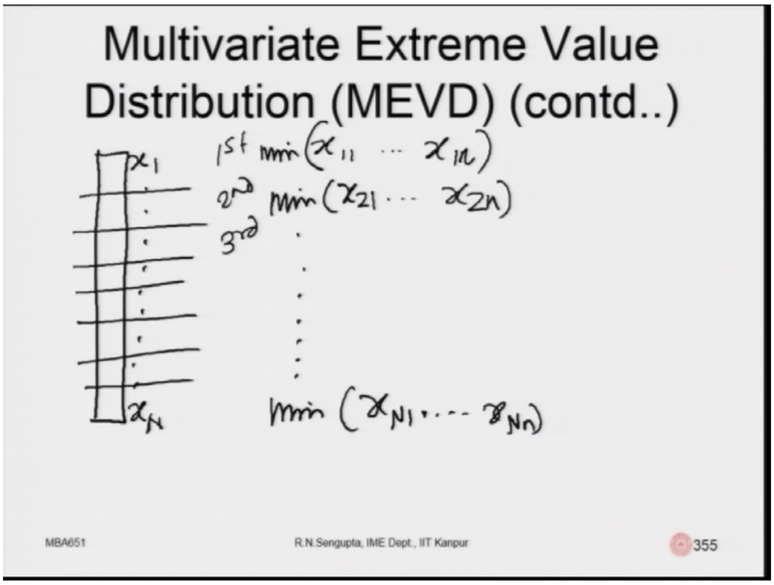
(Refer Slide Time: 24:15)



So, let us assume Y being i.i.d and you want to basically find out the maximum of them and you have basically p number of random variables and each has n number of readings. Now as for the definition is n is the number of observations which I have just mentioned while p is the dimension of that random variables. So, consider p 1, p 2, p 3 till p means p value 1, 2, 3, 4 till 10. So, you have 10 number of random variables each has basically n readings.

So, given this we are interested to find out the maximum. So, what you are trying to do is that in a very simple case without going to the deep the formula consider you have a univariate case. So, this example if I give hopefully it will make things, but clear to you. So, consider your data set is there let me make, try to make it right it is possible.

(Refer Slide Time: 25:15)



So, consider you have our data set and the values are marked x 1 till x see for example, capital N whatever the values forget about that.

Now, what I do is that, I break that into groups same number and consider the groups are given as number first one, second, third and so on and so forth. Now in the extreme value case what I will do? In the first case let the values be the first one be the group number 1, second one with the reading number till say for example, 1 till say for example, n this n and the other value of sample size. Let us not confuse so, what I do is that, I find out the minimum similarly find out I take the minimum of x 2, this is the first 2 is the so called block number. Second one is the reading number till x 2 n. I continue doing it towards the last one give our value x N 1 x Nn. So, once I have the means, I jumble them up consider and keep it in a box. So, in the box what are there? They are the means of them are the minimum of different blocks which have taken.

Now, consider I am able to take them as take the minimum of that minimum and continue doing it is for such different number of n values. So, it may be possible then the minimum of the minimum for each block would actually give you the realized values of the sample of the minimum distribution based on which I am trying to find out the extreme value for the minimum. Similarly if I do for the x maximum value, I repeat it take the blocks find the maximum, for the first block, for the second block third block put it in a box and then find out the maximum the maximum. So, if I do such different type of blocks and I take it one time 2 time, 3 time then; the realize values of the maximum the maximum would actually be from the sample based on which we are trying to find out the extreme value distribution.

So, these type of bootstrapping. So, called bootstrapping other things can be discussed in details if somebody has a good knowledge in statistics and they are very heavily used in data analysis nowadays like the concept of bootstrapping, concept or time series concept of Bayesian analysis; they are being very heavily utilized. So, will cover them later on, but let us proceed slowly and try to basically understand the concept in more in a much better sense. With this I will close this class and hope all of you are finding it interesting and have a nice day and.

Thank you very much.