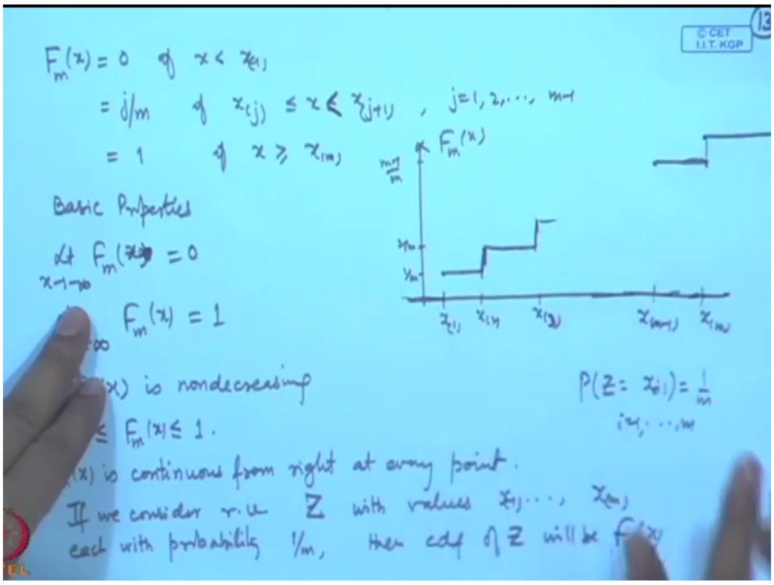
Statistical Methods for Scientists and Engineers
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology - Kharagpur

Lecture - 32
Non parametric Methods - V

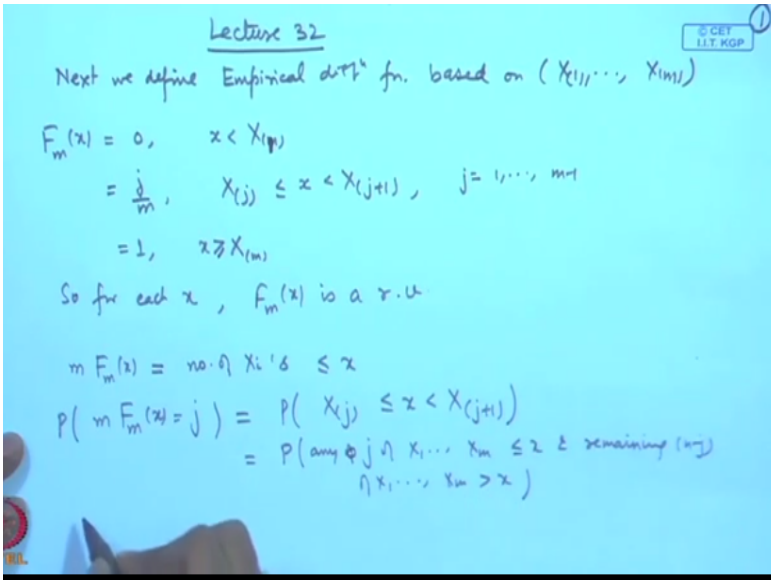
In the previous lecture, I introduced the idea of the tolerance intervals, the coverage probabilities etc and then towards the end I defined what is known as empirical distribution function or the sample distribution function.

(Refer Slide Time: 00:37)



If I have the sample x1, x2, xm and based on that the order statistics x1, x2, xm is defined and if the observed values are taken then based on the observed values we define a step function of this form. As I mentioned this is also the cdf of a discrete random variable, which takes values x1, x2, xm each with probability 1/m. So we get this as the function. Now in place of this small xi suppose I put capital Xi then this will become a random variable.

(Refer Slide Time: 01:22)

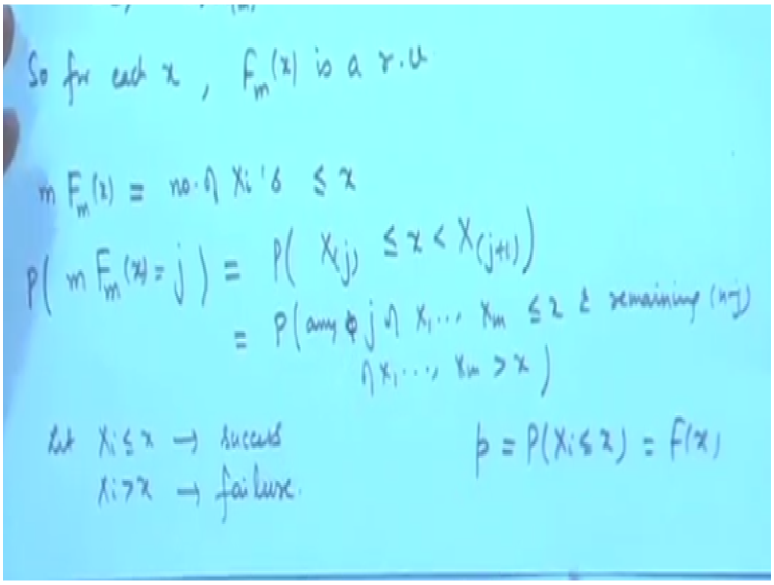


So now let us consider that. Next, we define empirical distribution function based on X1, X2, Xm that is Fmx=0 for x<X1 and it is = j/m if Xj is <= x<Xj+1 for j=1 to m-1 and that is = 1 if Xm that means basically x is >= Xm. Now this has become a random quantity, but still we call it empirical distribution function. So for each x Fmx is a random variable.

We can keep on changing x but still in all the cases this will remain a random variable. Let us analyze this. If I consider say m times Fmx then what are the values, it is = 0 if x<X1, it is = 1 if x is >= X1 but < X2. It is 2 if x is >= X2 but < X3 and so on. That means it is exactly the number of Xi’s that is <= x.

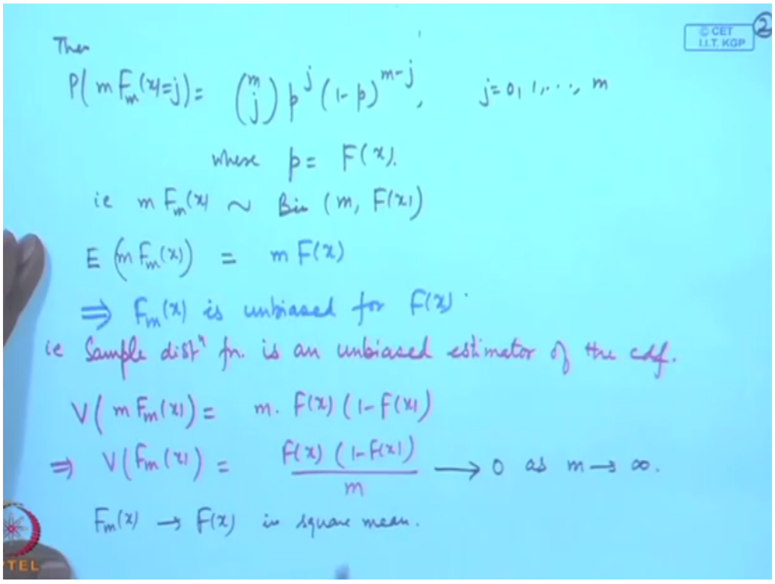
So if I consider the distribution of m times Fmx then that is = probability of Xj <= x<Xj+1. Now we can consider it in the following fashion. Here we can say this is same as probability that any j of X1, X2, Xm are <= x and remaining n-j of X1, X2, Xn they are > x okay.

(Refer Slide Time: 04:00)



So if we consider the event, Xi <= x suppose this is denoting a success and Xi>x to be a failure. Then basically what we are saying is that out of that means because these are i i d because this is actual cdf here.

(Refer Slide Time: 04:38)



Suppose I call it p then this is equal to simply probability of m Fmx=j this will simply become = m C j p to the power j 1-p to the power m-j for j=0, 1, to m where p is nothing but Fx here. So we are actually able to derive the distribution of the empirical distribution function that means we are saying that m Fmx is nothing but binomial m, p that is Fx basically.

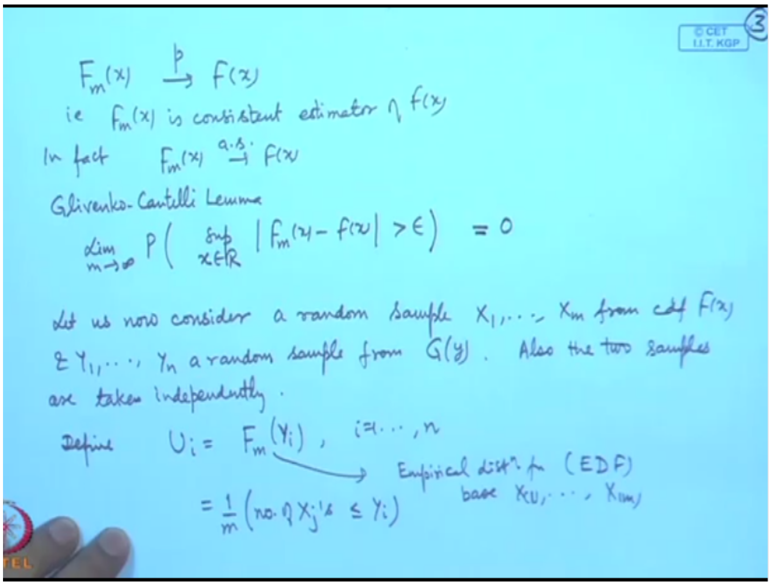
Now based on the binomial distribution you have some simple properties, for example what is expectation of m Fmx=m times Fx so this means Fmx is unbiased for capital Fx. So that is the first property. That means we can say that the sample distribution function is an unbiased estimator of the cdf. So you see the analog with the parametric inference. In the parametric inference, we find out the unbiased estimator of some parametric function here.

In the non-parametric case since the parameter is not there we are only having the form of the cdf that means basically we are saying that the model is not known. Then we can actually estimate the cdf itself by using the empirical distribution function, which is of course that means you are taking m observations here x1, x2, xm and based on that you are constructing the estimate and of course this estimate that we have written it is based on this.

So of course when observed values are there then it is becoming an estimate here and this is actually the estimator and that we are showing that it is an unbiased estimator. Of course, you may feel that see you are taking it as simply an step function and your cdf can be of any form so you may say that this may be too much different than the but basically in the absence of any other information because the parametric form is not there.

Therefore, this is the best that we can do okay. Based on this, then we can have other properties also. For example, what is variance of m Fmx that is m Fx*1-Fx that means variance of Fmx is nothing but Fx*1-Fx/m and of course you can see that this actually goes to 0 as m tends to infinity. So we also have that Fmx converges to Fx in square mean. In square mean this convergence will be there.

(Refer Slide Time: 08:33)



And also since it is unbiased and the variance is going to 0 this is also becoming consistent here that means we are having Fmx converging to Fx in probability. In fact, you have a stronger so basically we can say that Fm is consistent estimator of Fx. In fact, one can prove stronger thing, in fact we can have Fmx converging to Fx almost surely and we have even much more stronger result that is known by Glivenko-Cantelli Lemma.

That is saying that probability of supremum of Fmx-Fx>epsilon. This supremum is taken over all x on the real line. Even this probability goes to 0 as m tends to infinity. So these are some of the very, very you can say strong properties about the empirical distribution function. Now we will develop, you can say theory which will be used for making useful inferences for the various 2 sample problems.

For example, here I have told that we can actually talk about the median or the quantile so we have also discussed a test for the quantile, but many times we will be concerned about 2 populations that means we will be comparing like in the parametric case we had the testing about equality of the means of 2 normal populations, equality of the variances and so on. So similarly in the non-parametric case we may discuss the test about the equality of medians etc.

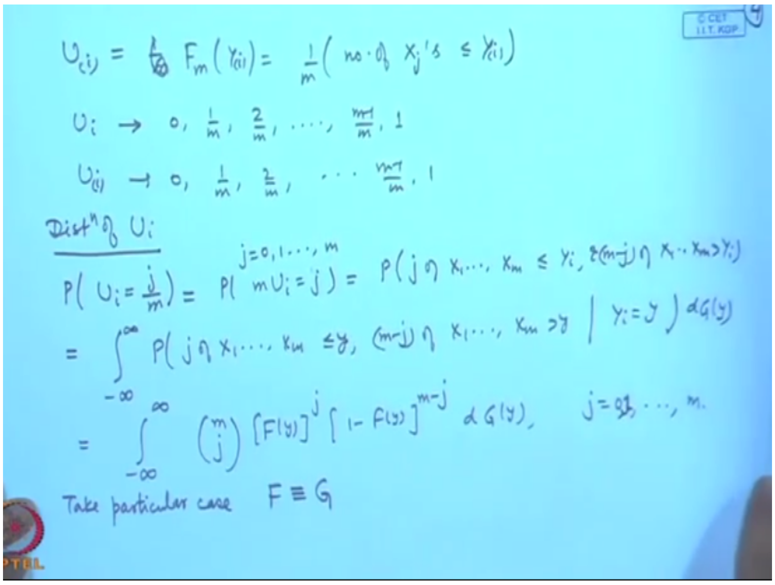
So based on the empirical distribution function, I will construct some procedure which will help in this regard. So let me develop this theory first. Let us now consider a random sample X1, X2, Xm from cdf Fx and say Y1, Y2, Yn a random sample from say Gy so this is different one and also we take the 2 samples to be independent of each other. Also the 2 samples are taken independently.

Now based on this I consider the empirical distribution function here. So Ui I define to be Fm Yi for i=1 to n. Now this Fm is the empirical distribution function. I will use the term EDF actually based on X1, X2, Xm alright, but in the argument I am substituting Yi here. So this is actually then becoming as we have seen what is the interpretation for the m times Fmx, m times Fmx was the number of Xi's, which are <= x.

Therefore, m times Fm Yi will denote the number of Xi's which are <= Yi. So this will become 1/m times number of Xj's which are <= Yi and of course this notation is similar to the one which I used for F of Xi but now it is in a different context so this Ui's are different from that. That was the Xi's were having cdf F, so F of X1, F of X2, F of Xn that was a random sample from uniform 0, 1 that is the continuous distribution.

Now these Ui's I am defining based on the empirical distribution function and okay so the name Ui is taken same but this has some significance that will be clear a little later.

(Refer Slide Time: 13:24)



So if I consider Ui as the 1/m number of Xj's <= Yi then what will be U of i that will be 1/m times well so that is = Fm of Yi that is = 1/m times number of Xj's which are <= Yi okay. Since here this is from 0, 1 to m so what will be the values of Ui's and U (i)'s? It will be 0, 1/m, 2/m etc. So Ui can take values 0, 1/m, 2/m and so on up to m-1/m, 1.

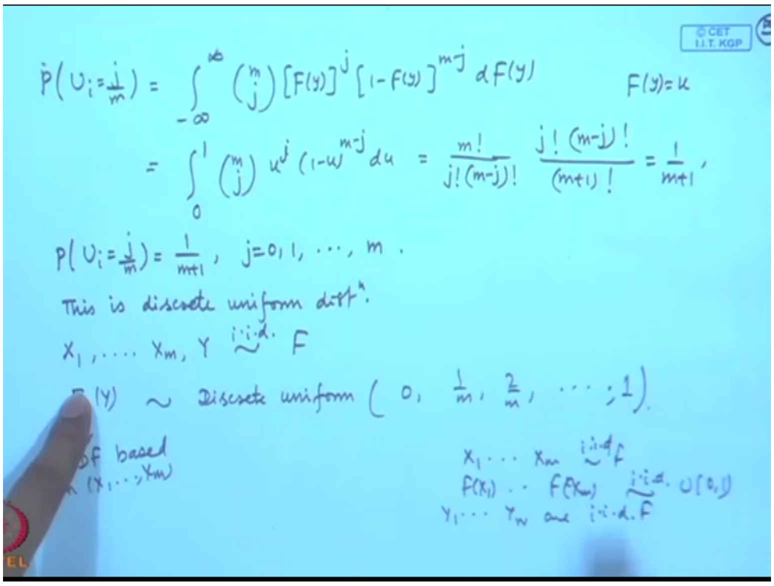
And similarly Ui can take values 0, 1/m, 2/m and so on m-1/m, 1 etc. What is the distribution of Ui? Let us look at this. Distribution of Ui, okay first of all what is the difficulty? Earlier the distribution of Ui was simply uniform distribution because of the probability integral transform. Here I am actually doing an integral transform but it is with respect to the cdf here, so the cdf itself is having the random variables here.

And then this Yi is coming here which is again having random variable so it is having a joint distribution here. So let us derive this thing. What is the probability that Ui=say some number j/m, there j can take values 0, 1 to m okay? So this is equal to probability of m times Ui=j=probability that j of X1, X2, Xm are <= Yi and m-j of X1, X2, Xm are > Yi.

So this you can write as integral probability j of X1, X2, Xm <= y, m-j of X1, X2, Xm > y given that Yi=y says, dGy where capital G was the distribution function of second sample. So this then you can write as now what is happening that given Y this becomes a fixed thing so this is simply coming from the binomial that is m C j Fy to the power j 1-Fy to the power m-j and then dGy.

So actually we are able to determine the distribution of Ui is here. Now we take one particular case, of course I mean if I take for example uniform distributions and here I take say some other distribution say normal distribution etc then this all can be easily written in a closed form. Now let us take one particular case when both the samples are from the same population.

(Refer Slide Time: 17:17)



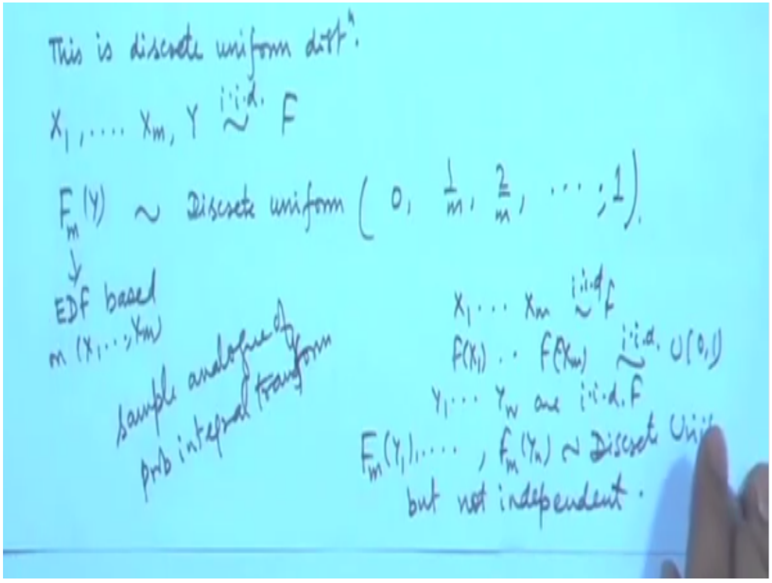
If both the samples are from the same population then this probability of Ui=j/n that is -infinity to infinity m C j Fy to the power j 1-Fy to the power m-j dFy. Now see earlier I had dGy but if G=F then I put it simply as dFy. Now you can simply substitute say Fy=some U if that is happening then this is simply becoming 0 to 1 m C j U to the power j 1-u to the power m-j du, which is simply a beta integral.

So this can be evaluated m factorial this m C j I write as m factorial/j factorial m-j factorial and this is becoming when j factorial m-j factorial/m+1 factorial, which is nothing but 1/m+1. So this is interesting what you are getting probability of Ui=j/m=1/m+1 where j=0, 1 up to m. This is nothing but a discrete uniform distribution. So what we are saying it is something like this.

If I am considering say X1, X2, Xm and Y they are from the same distribution F and if I consider Fm of Y where this is empirical distribution function based on X1, X2, Xm then this is actually having discrete uniform on 0, 1/m, 2/m, and so on. So this if you see, see if I have say X1, X2, Xm from F then if I consider F of X1, F of X2, F of Xm then this is i i d from uniform 0, 1.

Here in place of F I am putting Fm okay that means what I am saying is that if Y1, Y2, Yn they are i i d F.

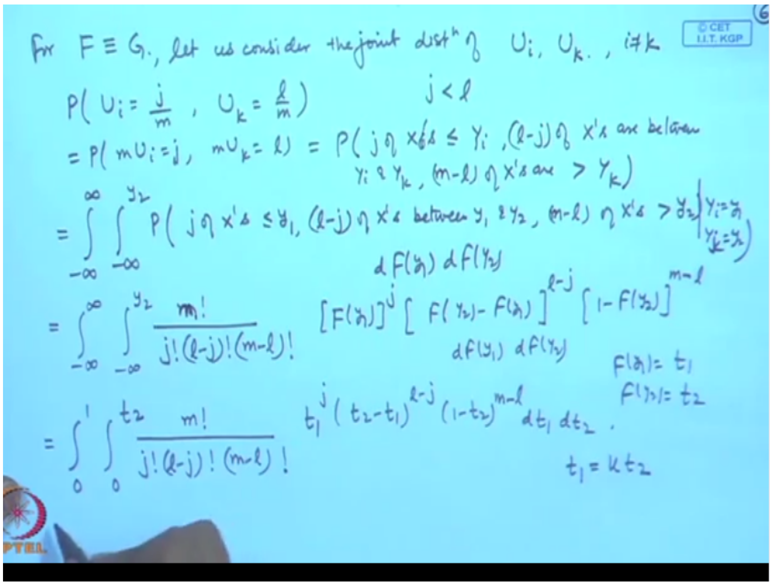
(Refer Slide Time: 20:08)



And if I define Fm of Y1, Fm of Yn then they are discrete uniform, but they are not independent. Certainly, they cannot be independent because all of them are based on the same Xi’s here. So this can be considered as a sample analogue of probability integral transform, that is the first result which I gave in this particular section when we started the non-parametric methods we considered this probability integral transform that if X is having cdf F then Fx is having uniform 0, 1.

So if I have sample X1, X2, Xn then F of X1, F of X2, F of Xn will be a sample from uniform 0, 1 but here if I considered the sample distribution function here and then I define this then this is having discrete uniform but it is not independent. so that is the difference here. So this is quite interesting result here. We will also be interested in the joint distributions for example 2 of them Ui and Uj, so let me discuss this here.

(Refer Slide Time: 21:46)



F=G let me take, let us consider now the joint distribution of say 2 of them say Ui, Uk of course you are taking i != k. So what is the probability of say Ui=something like j/m and Uk=say l/m. Then you take here say j<l because there can be each of the Ui's can take values 0, 1/m, 2/m up to l/m and similarly Uk. So there can be 3 cases j=l, j<l and j>l, so let me take say j<l.

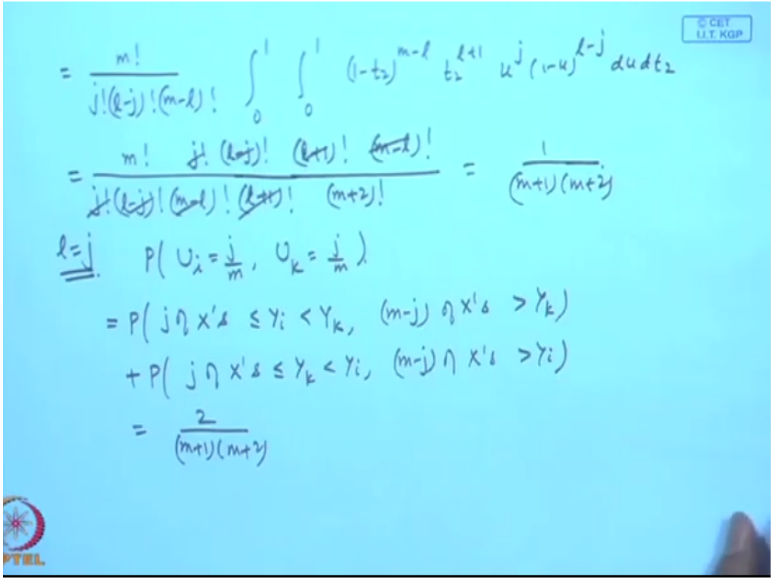
If j<l then the advantage is that you are actually writing it as the probability of mUi=j and mUk=l that means it is actually the probability of j of X's <= Yi. This is I am talking about Ui that is based on F of Yi and then l-j of X's are between Yi and Yk and then m-l of Xs are more than Yk. So this becomes basically a multinomial if we remember our Fx there.

So we do the conditioning now. Now it has to be conditioning on Yi and Yk here. So we write probability of say j of X's <= say y1, l-j of X's between y1 and y2 and m-l of X's > y2. This is conditioned on Yi=y1 and Yk=y2, dFy1 dFy2 see there could be in general case where this will be G but then of course you will not be able to obtain a close form expression for that.

So -infinity to y2 for y1 and -infinity to infinity for y2 here. Now this is simple multinomial so we just write it here -infinity to infinity, -infinity to y2. So this will be m factorial/j l-j and m-l. Then you have F of y1 to the power J F of y2-F of y1 to the power l-j and 1-F of y2 to the power m-l dFy1 dFy2. Now in this one I can make the transformation say F of y1=t1 and F of y2=t2.

So then this will become dt1, dt2 and the range this will become -infinity this will become 0 F of y2 this will become t2, this will become 0 to 1. So this is becoming 0 to 1, 0 to t2 m factorial/j factorial l-j factorial m-l factorial. Here it will become t1 to the power j t2-t1 to the power l-j 1-t2 to the power m-l dt1, dt2. Now this can be simplified. See you can do first time say something like t1=u times t2.

(Refer Slide Time: 26:45)



Then if you do this then this is simplified this one will become = m factorial/j factorial l-j factorial m-l factorial integral 0 to 1 0 to 1 1-t2 to the power m-l t2 to the power l+1 u to the power j 1-u to the power l-j du dt2. So this becomes simply beta term and this is again becoming a beta term. So all of this can be evaluated, m factorial/j factorial l-j factorial m-l factorial this integral will give me j factorial l-j factorial/l+1 factorial.

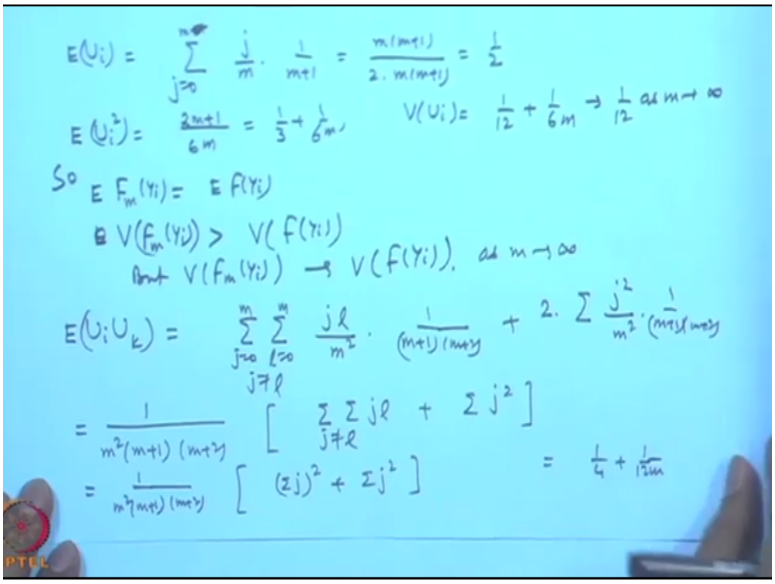
And this integral will give us l+1 factorial m-l factorial and m+2 factorial. So you can see that these terms get cancelled out. You are left with simply 1/m+1*m+2. Now the other case when j>l will be similar, in fact you can see here that this value is not dependent upon j and l here. So only thing that we considered is j<l, so if I considered j>l then in the expressions it will get reversed.

Here it will become l factorial, here it will become j-l factorial and here it will become m-j factorial that is all. So all these things will be suitably interchanged but the final value will still be the same. So now the only other case that I need to consider is j=l, this case if I consider this case then I am saying Ui=j/m and Uk is also = j/m. Then this is equivalent to say j of X's are <= Yi.

Now among i and k, I have to assume something so it can be that Yi<Yk or Yi can be > Yk, so let us take this and then m-j of X's are > Yk itself or it is that reverse of that. That is j of X's is <= Yk is < Yi and then m-j of X's they are > Yi. So basically you can do 2 times that thing so you can actually carry out the calculation, it will become 2/m+1*m+2 okay.

Actually both the terms will have the same expression that is 1/m+1 because basically what is happening in between there is nothing but since it was free from that choice it will be dependent only on this m here. So we are able to derive the complete distribution or complete joint distribution of Ui and Uk. I have considered the case j<l, j>l will give the similar thing and j=l.

(Refer Slide Time: 30:43)



Next, we consider say the moments of this. We consider the moment structure of Ui's okay. So what is expectation of Ui for example? We derive the distribution of Ui as simple, discrete uniform distribution over the range 0 to m so when you consider the mean of this, it is simply becoming = sigma j/m 1/m+1 j=0 to n. So what is this term m*m+1/2? And this is m*m+1 so the mean is 1/2 and you can of course look at higher order moments also.

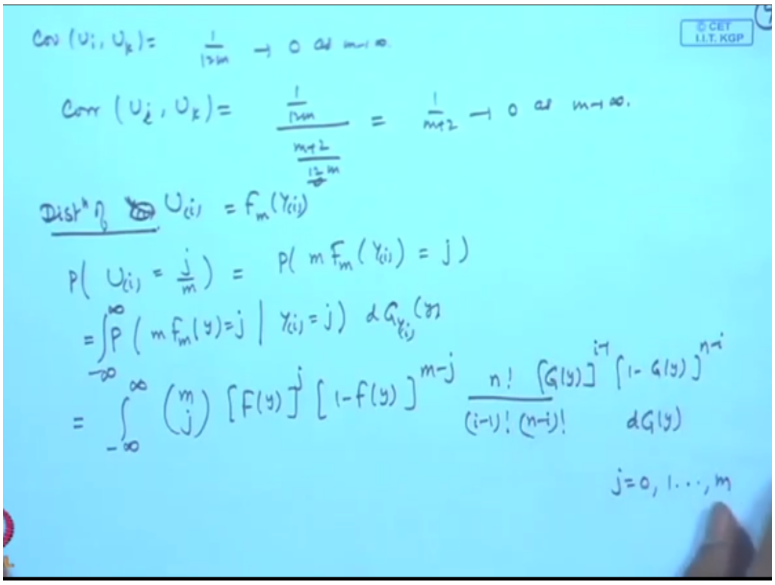
I am not going to write the full details here, you can just check it. It will become = 2m+1/6m basically it is equal 1/3+1/6m and therefore variance of Ui also you can see, variance of Ui=1/12+1/6m that is interesting. See if you remember the uniform 0, 1 there the variance is 1/12, mean is ½. So here mean is still 1/2 but the variance is 1/12+something 1/6m so if m becomes large then this is approximately 1/12 okay.

So this is interesting thing to observe okay. So basically what we are saying is that expectation of Fm Yi=expectation of F of Yi and variance of Fm Yi is actually > variance of F of Yi but variance of Fm Yi converges to variance of F of Yi. So this is the observation that we are having here based on these expressions that we derived here. We can also look at the covariance structure here.

Let us look at for example the product moment UiUk=jl/m square 1/m+1*m+2 this is j=0 to m, l=0 to m where j is != l and when they are equal then it is becoming 2 times sigma j square/m square 1/m+1*m+2. Now you can easily see that this denominator is common so I can keep it outside that is 1/m square*m+1*m+2 and we are left with summation jl j != l+summation j square, which is actually coming out to be.

I can consider it as summation j whole square-the terms which are j square here, so this is then = 1/m square m+1 m+2=sigma j whole square+sigma j square. I have made a mistake here this will become minus this here okay so after simplification this will turn out to be 1/4+1/12m.

(Refer Slide Time: 35:11)



So if I consider here covariance between Ui and Uk that will be = 1/12m, which actually goes to 0 as m tends to infinity and if I look at say correlation between Ui and Uk then that is = 1/12m/m+2/12m and this term is actually = 1/m+2, so that goes to 0 as m tends to infinity. That means the amount of the correlation or correlatedness between ith and kth sample transformed value of the order statistics.

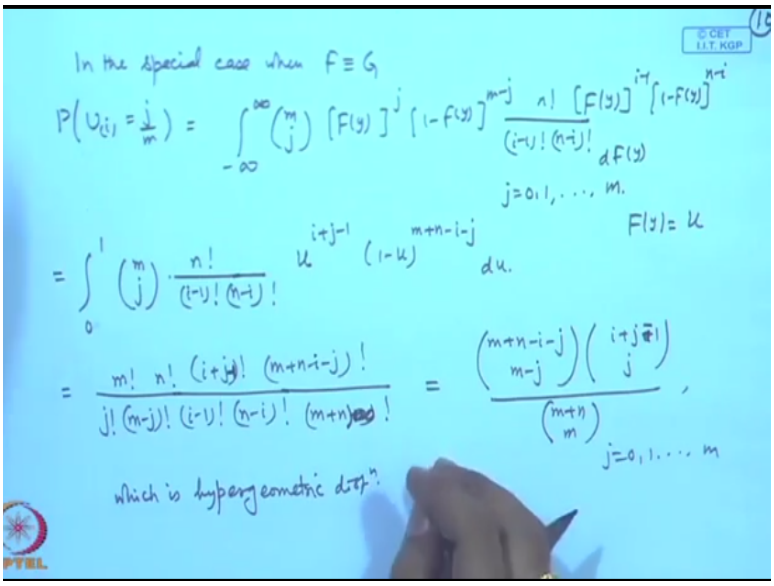
That correlation becomes less and less as the sample size increases. Now this is about you can say sample analogue of the order statistics of any distributions so one we do with the original distribution function and second we do with the empirical distribution function and we are actually able to analyze the distribution completely in these cases. Now on the right hand side in place of Yi values if we consider ordered values then what will happen?

That is distribution of Yi so that means I consider Ui's are actually Fm of Yi okay. In the case of F this was directly the order statistics from the uniform distribution and we were able to derive the distribution as a beta distribution but here what it will give? So let us consider this. Probability of Ui=j/m=probability of m times Fm of Yi=j so that is equal to probability of m times Fmy=j given Yi=j dG Yi y.

Integral from –infinity to infinity. Here the cdf of the ith order statistics from the second population is there and this one is of course known here that is the binomial m C j Fy to the power j 1-Fy to the power m-j and this distribution was known so if we substitute that there I get n factorial/i-1 factorial n-i factorial Gy to the power i-1 1-Gy to the power n-I dGy for j=0, 1, to n.

Because this is the ith order statistics from the second sample, so that was beta i and –i+1 distribution for G so that we are able to get here. So this is the general form of the empirical distribution function transformed value of the ith order statistics from the second sample that is Ui. So this is the general expression. Now we can consider the particular case when G and f are the same.

(Refer Slide Time: 39:12)



Then what happens? F and G are same then this expression can be actually evaluated. So I will show that thing m C j Fy to the power j 1-Fy to the power m-j n factorial/i-1 factorial n-i factorial then this is becoming Fy to the power i-1 1-Fy to the power n-i dFy for j=0, 1, to n. Here all these powers get added up, further you can substitute say Fy=something like u. Then this is becoming integral from 0 to 1, m C j then all this terms will come here.

That is n factorial/i-1 factorial n-i factorial then Fy this power will get added up i+j-1 and then 1-u to the power m+n-i-j du that is = m factorial/j factorial m-j factorial n factorial i-1 factorial n-i factorial then this is becoming i+j-1 factorial m+n-i-j factorial and here it will become m+n factorial. So many of the terms can get adjusted here and you can write it as I am not sure whether all the terms are okay here.

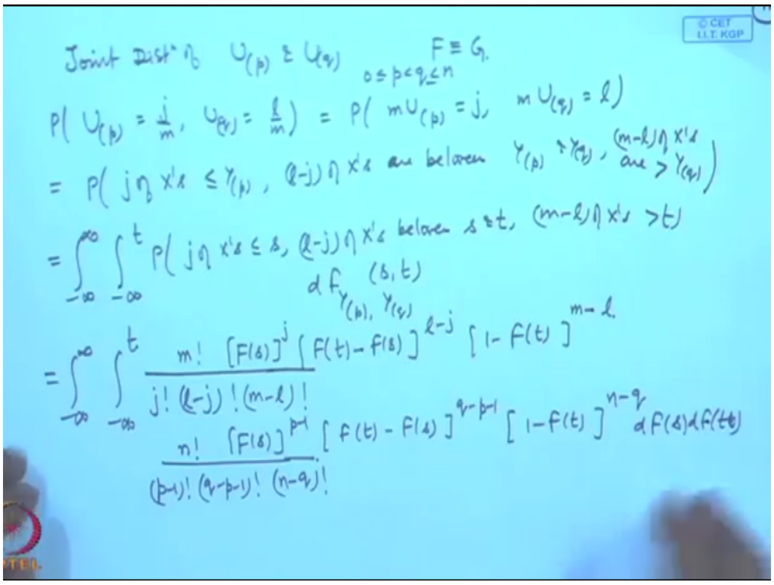
This will become m+n factorial right because m+n+i -i-j so this will get canceled out -1+1 so it will become m+n factorial right, now I think this is correct. So then if you have this then this is becoming m+n-i-j C m-j then you have i+j-1. So i+j-1 C j/m+n C n which is actually hypergeometric distribution. So this is quite interesting, we are able to obtain the distribution of this transforms using the empirical distribution function.

Let me show it again for the convenience. We got the Ui as a discrete uniform distribution on the points 0, 1, to m and the corresponding ordered one that means the empirical distribution function transform value of the ith order statistics that is Yi is coming out to be a hypergeometric distribution here that means here the 2 samples they are added up actually m and n are the respective sample sizes.

So it is becoming m+n where m are X's and n Yi's are there so you can see here. Out of that if we are choosing m things here, then it is becoming so it is dependent upon how many things I am saying j values of X's are <= Yi that is ith one okay. So that is simply playing a role here.

So this is very, very interesting here and we can also talk about the joint distribution of Ui and say Uj kind of thing that is when we are considering 2 different things.

(Refer Slide Time: 44:28)



So let us look at this. So these are leading to very, very interesting observations and we will show later on that we will use it for the inference purpose that means when we do the testing about the locations for the 2 sample problems this will be really used there. So let us consider now joint distribution of say 2 of them say Up and Uq of course I am taking the case when F and G are same.

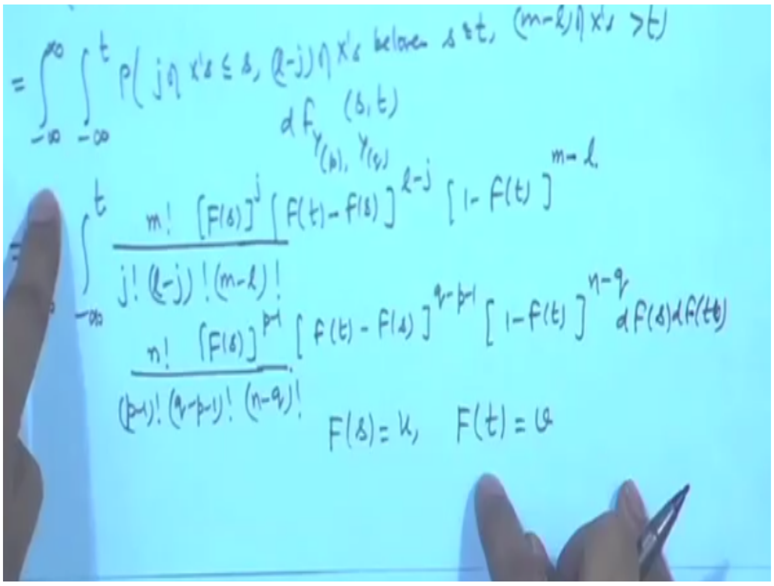
And let us take say 0 p<q here so probability that Up=j/m and Uq=l/m=probability of m times Up=j, m times Uq=l. This will be = probability j of X's are <= Yp, l-j of X's they are <= they are between actually Yp and Yq and m-l of X's are > Yq. So that is equal to double integral probability of j of X's are <= s, l-j of X's are between s and t and m-l of X's are > t dF Yp Yq s, t.

Remember here I am not writing separately this. The reason is that the distributions of Yp and Yq are not independent unlike the case that I discussed a little earlier where the distributions were independent. So there I was able to write separately. You see here, here I was able to write dFy1 dFy2 because this Yi and Yk they were taken to be independent but here they are ordered so they are not independent so I have to write the joint one here.

But this is not a problem because we actually know this so we can write –infinity to infinity –infinity to t this is m factorial/j factorial l-j factorial m-l factorial then you are having the cdf things. So F of s to the power j F of t- F of s to the power l-j and 1-F of t to the power m-l. Now here you have to write the joint distribution of Yp and Yq but that we know so we substitute here.

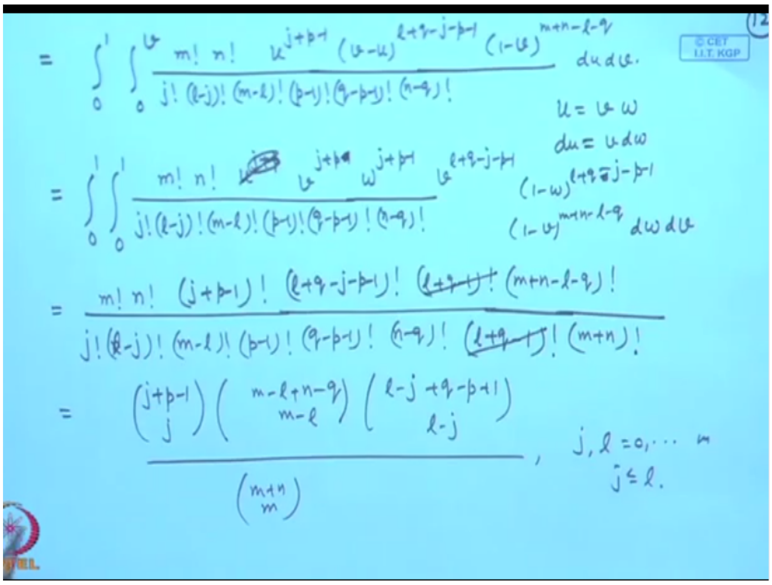
That is n factorial/p-1 factorial q-p-1 factorial n-q factorial Fs to the power p-1 Ft-Fs to the power q-p-1 then 1-Ft to the power n-q dFs dFt where s is from –infinity to t and s is from -infinity to +infinity.

(Refer Slide Time: 48:37)



So now for purpose of evaluation this can be transformed you can consider say Fs=u and Ft=v. In that case, you can see here this will become 0 to v and this will become from 0 to 1 and all other things will be added up so the powers will get added so this can be evaluated in a closed form.

(Refer Slide Time: 49:02)



Let me show you this here. So this is = 0 to 1, 0 to v and all these coefficients will be coming there. Let me write it here, m factorial n factorial/j factorial l-j factorial m-l factorial p-1 factorial q-p-1 factorial n-q factorial. So all these terms will be coming there and then you have the powers, let me see here. You will get u to the power that is Fs, Fs to the power j+p-1 so that is becoming u to the power j+p-1.

Then you have v-u to the power l+q-j-p-1. Then you have 1-v to the power m+n-l-q du dv. So this is actually a bivariate beta integral and this can be easily evaluated in fact I can show you the method of calculation let us put say u=vw that is du=v times dw in the inner integral. In the inner integral if we put this, then when u=0 w is 0 and when u=v w will become = 1.

So basically what is happening is that both are becoming beta integrals 0 to 1 m factorial n factorial j factorial l-j factorial m-l factorial p-1 factorial q-p-1 factorial n-q factorial then you have u to the power j+p so that is becoming v to the power j+p-1 w to the power j+p-1 and then one v is coming here also so I will put here this. Then here you get v-vw so another v is coming out l+q-j-p-1.

And then you have here vw so vs come out so it will become 1-w that is 1-w to the power l+q-j-p-1 and then you have 1-v to the power m+n-l-q dw dv. So now this is becoming m factorial n factorial/j factorial l-j factorial m-l factorial p-1 factorial q-p-1 factorial n-q factorial. Now when you evaluate these beta integrals, you have w to the power j+p-1 so you will get j+p-1 factorial.

Then 1-w to the power l+q-j-p-1 factorial and then you add when you add this j+p will get canceled out. You will get l+q-1 factorial and now let us look at the powers of v and 1-v. So for v it is l+q-1 so l+q-1 factorial 1-v is m+n-l-q then I will get m+n factorial. So here you can see some of terms may get canceled out. For example, I can see here l+q-1 factorial and l+q-1 factorial gets canceled out here.

And the other terms we adjust here. So this becomes j+p-1 C j then this is m-l+n-q C m-l l-j+q-p+1 C l-j and this m+n C n where j and l are from 0 to n j is <= l. So this is something like a bivariate hypergeometric distribution. So again let us compare with the one when ordering was not taken then what we had? We got the distribution as a bivariate discrete uniform distribution here.

This was the distribution of Ui and Uk here. So I had got it as a bivariate discrete uniform and now you can see here that we are getting it as a bivariate hypergeometric so the comparison you can see. You have discrete uniform, well let us consider the sample thing firstly, X1, X2, Xm random sample from F then F of X1, F of X2, F of Xm is a random sample from uniform 0, 1.

So the corresponding if I take ordered observations then it becomes corresponding order statistics from a uniform random sample from 0 to 1. If I replace F by the empirical distribution function and I can see the 2 things F and G that means 2 samples X1, X2, Xm and Y1, Y2 and Yn but if I am taking F=G then what interesting thing that I am observing that if I take unordered ones that is Fm Y1, Fm Y2, Fm Yn then they are identically discrete uniform distributions on 0, 1/m, 0, 2/m up to 1.

But they are not necessarily independent. Now you consider again 2 of them suppose I can see that Xi Xj then the corresponding thing for F of Xi, F of Xj they are independent okay. If I consider F of x (i) F (xj) that is the ordered ones then they are jointly distributed order statistics from uniform distribution and therefore their distributions are derived as something like a bivariate beta kind of thing.

If I am considering the empirical version of that in that case in the first case when I am taking unordered one, I am obtaining a bivariate discrete uniform distribution and now I am obtaining it as a bivariate hypergeometric distribution. So bivariate beta, bivariate discrete uniform and now bivariate hypergeometric distribution that we are getting here.

So we have discussed in detail these applications of the empirical distribution function for some 2 sample cases. In the next class, I will do a few more properties and then we will for example look at the moment structure of this also and then we will look at the application to the testing problems okay. So in the next class I will be trying to cover that.